



<http://dx.doi.org/10.35596/1729-7648-2025-23-5-93-98>

УДК 004.89

МОБИЛЬНОЕ ПРИЛОЖЕНИЕ ДЛЯ ГЕНЕРАЦИИ ИЗОБРАЖЕНИЙ ПОМЕЩЕНИЯ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ STABLE DIFFUSION

К. И. УЛАСОВЕЦ, В. А. ЧУЙКО, Е. И. КОЗЛОВА

Белорусский государственный университет (Минск, Республика Беларусь)

Аннотация. На современном рынке существует множество веб-сервисов, предназначенных для генерации изображений интерьеров с помощью методов генеративного дизайна. В то же время высокая стоимость и ограничения по использованию некоторых коммерческих решений ограничивают их доступность для широкого круга пользователей. В статье предложено экономически выгодное решение без значительных потерь в функциональности и качестве. Представлены результаты разработки мобильного приложения для генерации изображений помещения (интерьеров) по фотографии и текстовому описанию пользователя. Для генерации изображений применена генеративно-сопоставительная нейронная сеть. Описан процесс преобразования текстового запроса в изображение. Приведены результаты тестирования приложения.

Ключевые слова: генерация изображений, диффузионная модель, латентная диффузия, зашумленные данные, сверточная нейронная сеть.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Для цитирования. Уласовец, К. И. Мобильное приложение для генерации изображений помещения с использованием модели Stable Diffusion / К. И. Уласовец, В. А. Чуйко, Е. И. Козлова // Доклады БГУИР. 2025. Т. 23, № 5. С. 93–98. <http://dx.doi.org/10.35596/1729-7648-2025-23-5-93-98>.

MOBILE APPLICATION FOR GENERATING ROOM IMAGES USING STABLE DIFFUSION MODEL

KSENIYA I. ULASAVETS, VLADISLAV A. CHUYKO, ALENA I. KAZLOVA

Belarusian State University (Minsk, Republic of Belarus)

Abstract. Numerous web services designed for generating interior images using generative design methods currently exist on the market. However, the high cost and limitations of some commercial solutions limit their accessibility to a wide range of users. This article proposes a cost-effective solution without significant sacrifices in functionality and quality. The article presents the results of developing a mobile application for generating images of rooms (interiors) based on a photograph and a user-supplied text description. A generative adversarial neural network was used to generate the images. The process of converting a text query into an image is described, and application testing results are provided.

Keywords: image generation, diffusion model, latent diffusion, noisy data, convolutional neural network.

Conflict of interests. The authors declare no conflict of interests.

For citation. Ulasavets K. I., Chuyko V. A., Kazlova A. I. (2025) Mobile Application for Generating Room Images Using Stable Diffusion Model. *Doklady BGUIR*. 23 (5), 93–98. <http://dx.doi.org/10.35596/1729-7648-2025-23-5-93-98> (in Russian).

Введение

Сегодня на рынке веб-сервисов, предназначенных для создания изображений интерьеров с помощью методов генеративного дизайна [1], предлагается широкий выбор вариантов как для пользователей-профессионалов, так и для любителей. В то же время высокая стоимость и лимиты по использованию некоторых коммерческих решений ограничивают их доступность для широкого круга пользователей.

Мобильное приложение для генерации изображений помещений (интерьеров) может предоставить пользователю интуитивно понятный инструмент визуализации помещения, а именно – сгенерированное изображение на основе фотографии и текстового запроса. В статье представлены результаты разработки и тестирования мобильного приложения для генерации изображений помещения. Требования к приложению: совместимость с различными мобильными устройствами, работающими на платформе Android, начиная с версии Android 7 (Nougat), соответствие функционала дизайна функционалу приложения.

При создании мобильного приложения использовался язык программирования Kotlin. Выбор обусловлен требованием к совместимости приложения с различными устройствами, наличием инструмента Multiplatform, системой типизации, позволяющей избегать ошибок, связанных с NullPointerException, возможностью функционального программирования. В качестве основной платформы, для которой разрабатывалось приложение, была выбрана Android, как доступная широкому кругу пользователей. Выбор минимальной версии Android – Android 7 API 24 – позволяет скачать и использовать данное приложение более 97 % пользователей Android [2].

Методы обработки и генерации изображений

Генерация изображений помещений производится на основе модели Stable Diffusion – генеративной нейронной сети с открытым исходным кодом, разработанной компанией Stability AI и представленной в публичный доступ в августе 2022 г. Модель предназначена для создания детализированных изображений на основе текстовых запросов. Stable Diffusion относится к классу моделей глубокого обучения, называемых диффузионными моделями. Диффузионные модели относятся к категории генеративных моделей, предназначенных для синтеза новых данных, обладающих сходными характеристиками с данными, использованными для обучения модели. В случае Stable Diffusion данными являются изображения.

Диффузионные нейросетевые модели работают на основе двух взаимобратных процессов – прямой и обратной диффузии. Процесс прямой диффузии представляет собой добавление шума к исходным данным на протяжении нескольких этапов. Конечная цель прямой диффузии – достижение состояния, в котором данные становятся полностью зашумленными и неразличимыми от чистого шума.

Процесс обратной диффузии является противоположным по отношению к процессу прямой диффузии. Обратная диффузия начинается с полностью зашумленных данных и заключается в пошаговом удалении шума с целью восстановления исходной структуры данных. Процесс предполагает наличие модели, обученной предсказывать шум, который необходимо удалить. Итог процесса обратной диффузии – получение данных, максимально приближенных к исходным. На рис. 1 представлен процесс прямой и обратной диффузии.



Рис. 1. Процесс прямой и обратной диффузии
Fig. 1. Forward and reverse diffusion process

Однако сами описанные выше процессы не формируют напрямую механизм стабильной диффузии, поскольку с точки зрения вычислительной эффективности скорость их работы в пространстве изображения значительно ниже, чем у классических алгоритмов машинного обучения. Модель стабильной диффузии применяется как раз для решения проблемы скорости.

Стабильная диффузия – это модель латентной диффузии, когда вместо работы в многомерном пространстве изображение сжимается в латентном (скрытом) пространстве, т. е. процессы прямой и обратной диффузии происходят в скрытом пространстве. Для изображения размером $3 \times 512 \times 512$ px скрытое пространство имеет размеры $4 \times 64 \times 64$ px, что в 48 раз меньше. Меньший размер пространства позволяет обрабатывать больше чисел, что обуславливает ускорение работы алгоритма на основе стабильной диффузии.

Архитектура модели стабильной диффузии состоит из двух компонент – модуля кодирования текста и модуля генерации изображения (рис. 2).

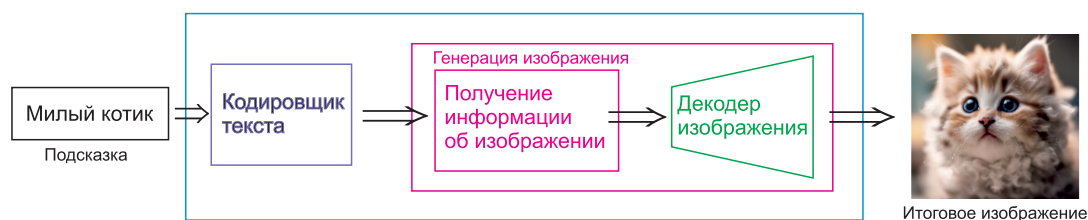


Рис. 2. Структура модели стабильной диффузии
Fig. 2. Structure of the stable diffusion model

Модуль кодирования текста используется для создания его векторных представлений, направляющих процесс генерации изображений. Вместо разработки нового кодировщика Stable Diffusion интегрирует предварительно обученную на парах «изображение–текст» модель CLIP (Contrastive Language-Image Pre-Training), способную использовать естественный язык для создания соответствующих векторных представлений текста [3].

Созданная пользователем текстовая подсказка (описание) сначала обрабатывается токенизатором, разделяющим текст на токены, представляющие собой слова или фразы, на которых токенизатор был предварительно обучен [4]. Затем токены проходят процессы создания взаимосвязей между словами и получения векторных представлений. Далее данные передаются в преобразователь текста, подготавливающий информацию для ввода шума в предсказатель [5].

Модуль генерации изображения включает в себя кодер и декодер. Кодер преобразует входное изображение в параметры распределения в скрытом пространстве, представляющие собой средние значения и стандартные отклонения для каждого измерения. Декодер затем использует эти параметры для генерации нового изображения, которое должно быть похоже на исходное [4].

В контексте диффузионных моделей процесс диффузии включает в себя постепенное добавление шума к данным и последующее его удаление. Этот процесс состоит из множества итераций, на каждой из которых применяется заранее предсказанный моделью шум.

Для успешного выполнения процесса обратной диффузии используется модель, например, сверточная нейронная сеть U-Net [6], предсказывающая разность между зашумленными данными и данными на предыдущем шаге диффузии, позволяя таким образом постепенно «очищать» данные от шума и восстанавливать изображение. Процесс преобразования текстового запроса в изображение в этом приложении включает следующие этапы:

- 1) текстовая подсказка и изображение в качестве входных данных [3];
- 2) кодирование входного изображения в скрытое пространство, текстовая подсказка делится на токены;
- 3) добавление гауссовского шума к скрытому представлению посредством прямой диффузии;
- 4) предсказатель шума U-Net принимает текстовую подсказку и зашумленное изображение и отнимает от него предсказанный шум в скрытом пространстве [7]. Этот процесс протекает столько раз, сколько шагов было указано при разработке алгоритма;
- 5) декодирование вариационным автокодировщиком скрытого представления в новое изображение, аналогично описанному в [6].

Проектирование интерфейса пользователя

Для выполнения поставленной задачи необходимо разработать графический интерфейс пользователя мобильного приложения. В качестве интерфейса выступают файлы, содержащие

XML-разметку, и классы представления. Интерфейс должен предоставлять пользователю возможность вводить текстовые запросы, просматривать разделы мобильного приложения, содержащие настройку для создания изображения, а также сохранять изображения в галерею.

При запуске приложения пользователь попадает на главную страницу. Раздел Img2Img – это раздел настроек генерации. Он содержит форму ввода данных для генерации изображения, которые будут отправлены на сервер. Детальные настройки генерации выставлены при разработке приложения, т. е. пользователю необходимо только ввести желаемый запрос в поле Prompt и нажать кнопку Generate. Форма для заполнения настроек генерации изображения представлена на рис. 3, а.

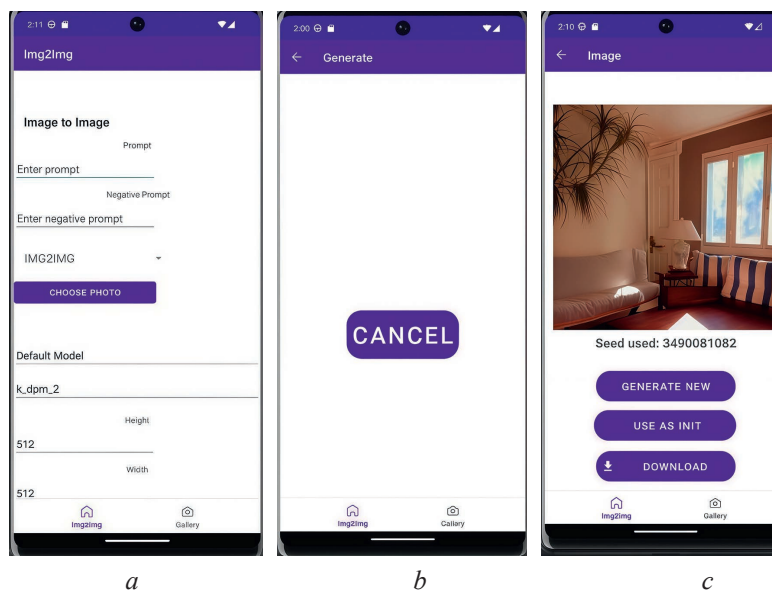


Рис. 3. Вид страниц приложения: а – главная страница Img2Img; б – страница «ожидания»; с – результат генерации

Fig. 3. The application pages: a – img2Img home page; b – “waiting” page; c – generation result

Работа приложения включает следующие этапы:

- ввод пользователем желаемого запроса в поле Prompt (например, interior design, 4K, high resolution, photorealistic, stylish modern style office, blue walls, white desks, laptop и т. д.);
- ввод в поле negative prompt – дополнение к запросу, уточняющее то, что пользователь не хочет видеть на картинке (например, window, door, low resolution, banner, logo, watermark, text, deformed, blurry, out of focus, surreal, ugly, beginner и т. д.);
- выбор на устройстве фотографии, на основе которой будет получено изображение интерьера;
- пользователь выбирает модель обработки изображения на основе его предварительного анализа (Default model, waifu_diffusion, Realistic Vision и т. д.);
- выбор пользователем:
 - метода обработки входного шума (k_dpm_2, k_dpm_adaptive, k_euler_a, lcm и т. д.);
 - размеров изображения (512×512 px, 512×768 px и т. д.);
 - количества шагов генерации (от 1 до 40);
 - seed (0 до 4294967295);
- в поле prompt strength пользователем вводится величина CFG Scale (принимает значения от 0 до 30) – это величина соответствия текстовому запросу. Чем больше значение, тем ближе результат к запрошенному, но вместе с тем и более зашумленный;
- в поле Image strength вводится значение от 0 до 1. Данная величина отвечает за то, насколько сгенерированное изображение похоже на исходное. Чем ближе величина к 1, тем меньше сгенерированное изображение будет походить на исходное;
- в поле AI API key вводится API-ключ, который пользователь получает по запросу на сайте Stable Horde.

После нажатия на кнопку Generate на сервер отправляются данные, необходимые для генерации изображений. Далее пользователь видит страницу «ожидание» (рис. 3, б), которую можно использовать для монетизации приложения, например, ввести в нее рекламу, пока сервер не закончит обрабатывать запрос. Эта страница не позволяет пользователю отправлять новые запросы

на сервер во время обработки предыдущего запроса. У пользователя есть возможность отменить запрос в случае необходимости.

После получения ответа от сервера пользователю виден результат генерации изображения, пример которого представлен на рис. 3, с. На данной странице у пользователя есть возможность:

- вернуться на начальную страницу для генерации нового изображения, нажав на кнопку Generate new;
- вернуться на главную страницу для генерации изображений и использовать полученное изображение в качестве входных данных, нажав на кнопку Use as init;
- скачать изображение на свое устройство, нажав на кнопку Download.

Пользователю в любой момент доступен раздел Gallery, представляющий собой галерею изображений. Также ему открыт доступ к просмотру ранее сгенерированных изображений. Каждому изображению и текстовой информации соответствует один блок. Пример страницы с галереей представлен на рис. 4.

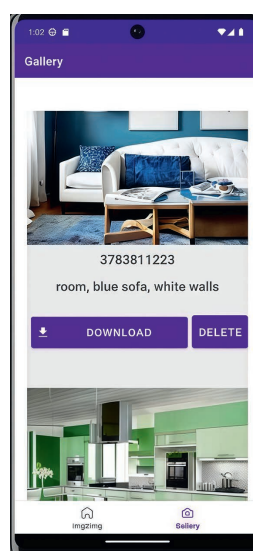


Рис. 4. Галерея изображений
Fig. 4. Image gallery

В каждом блоке пользователю доступны просмотр изображений, смена и запроса к изображению, а также возможность скачать изображение на свое устройство, нажав кнопку Download, или удалить его из галереи с помощью кнопки Delete.

Заключение

1. Представлены этапы и результаты разработки мобильного приложения для генерации изображений помещений на основе модели Stable Diffusion. Приложение разработано на Kotlin для платформы Android, обеспечивает интуитивно понятный интерфейс и позволяет создавать интерьеры по текстовым запросам. Тестирование разработанного приложения показало успешную генерацию в популярных стилях (скандинавский, модерн), однако с артефактами при замене мелких деталей (ручки шкафов, узоры на текстиле).

2. Приложение можно использовать в качестве дополнительного помощника (вдохновителя/визуализатора), а также для генерации изображений для сцен виртуальной реальности. Приложение предназначено для работы на любом устройстве с системой Android (начиная с Android 7), время генерации изображения по запросу составляет от 20 с до 5 мин в зависимости от нагрузки на сервер Stable Horde, организованный как кооперативная платформа, где производительность системы напрямую зависит от количества активных участников.

Список литературы

1. Методы генеративного дизайна [Электронный ресурс] / А. Я. Пахтаева // Ноэма. Режим доступа: <https://cyberleninka.ru/article/n/metody-generativnogo-dizayna>. Дата доступа: 12.06.2025.

2. Павленко, Е. Ю. Статический анализ безопасности Android-приложений / Е. Ю. Павленко, Г. Ю. Игнатиев, П. Д. Зегжда // Проблемы информационной безопасности. Компьютерные системы. 2017. № 4. С. 73–79.
3. Diffusion Models: A Comprehensive Survey of Methods and Applications / L. Yang [et al.] // ACM Computing Surveys (CSUR). 2023.
4. Ho, J. Classifier-Free Diffusion Guidance / J. Ho, T. Salimans // arXiv:2207.12598. 2022.
5. Deep Unsupervised Learning Using Nonequilibrium Thermodynamics / J. Sohl-Dickstein [et al.] // International Conference on Machine Learning. 2015.
6. High-Resolution Image Synthesis with Latent Diffusion Models / R. Rombach [et al.] // Conference on Computer Vision and Pattern Recognition. 2022.
7. Weng, L. What are Diffusion Models? [Electronic Resource] / L. Weng. Mode of access: <https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>. Date of access: 10.06.2025.

Поступила 19.06.2025

Принята в печать 23.09.2025

References

1. Pakhtaeva A. Ya. (2021) Generative Design Methods. *Nohma*. Available: <https://cyberleninka.ru/article/n/metody-generativnogo-dizayna> (Accessed 12 June 2025) (in Russian).
2. Pavlenko E. Yu., Ignatiev G. Yu., Zegzhda P. D. (2017) Static Analysis of Android Application Security. *Information Security Issues. Computer Systems*. (4), 73–79 (in Russian).
3. Yang L., Zhang Z., Song Y., Hong Sh., Xu R., Zhao Y., et al. (2023) Diffusion Models: A Comprehensive Survey of Methods and Applications. *ACM Computing Surveys (CSUR)*.
4. Ho J., Salimans T. (2022) Classifier-Free Diffusion Guidance. *arXiv:2207.12598*.
5. Sohl-Dickstein J., Weiss E., Maheswaranathan N., Ganguli S. (2015) Deep Unsupervised Learning Using Nonequilibrium Thermodynamics. *International Conference on Machine Learning*.
6. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. (2022) High-Resolution Image Synthesis with Latent Diffusion Model. *Conference on Computer Vision and Pattern Recognition*.
7. Weng L. (2022) *What are Diffusion Models?* Available: <https://lilianweng.github.io/posts/2021-07-11-diffusion-models/> (Accessed 10 June 2025).

Received: 19 June 2025

Accepted: 23 September 2025

Вклад авторов

Уласовец К. И. осуществила разработку приложения, подготовила рукопись статьи.
Чуйко В. А. разработал концепцию работы, провел критический анализ модели и текста статьи.
Козлова Е. И. выполнила критический анализ содержания статьи, подготовила окончательный ее вариант для публикации.

Authors' contribution

Ulasavets K. I. developed the application and prepared the manuscript of the article.
Chuyko V. A. developed the concept of the work, conducted a critical analysis of the model and text of the article.
Kazlova A. I. performed a critical analysis of the article's content and prepared its final version for publication.

Сведения об авторах

Уласовец К. И., студ., Белорусский государственный университет (БГУ)

Чуйко В. А., магистр физ.-мат. наук, ст. преп. каф. интеллектуальных систем, БГУ

Козлова Е. И., канд. физ.-мат. наук, доц., зав. каф. интеллектуальных систем, БГУ

Information about the authors

Ulasavets K. I., Student, Belarusian State University (BSU)

Chuyko V. A., M. Sci. (Phys. and Math.), Senior Lecturer at the Department of Intelligent Systems, BSU

Kazlova A. I., Cand. Sci. (Phys. and Math.), Associate Professor, Head of the Department of Intelligent Systems, BSU

Адрес для корреспонденции

220064, Республика Беларусь,
Минск, ул. Курчатова, 5
Белорусский государственный университет
Тел.: +375 29 853-07-96
E-mail: Vchuyko@bsu.by
Чуйко Владислав Александрович

Address for correspondence

220064, Republic of Belarus,
Minsk, Kurchatova St., 5
Belarusian State University
Tel.: +375 29 853-07-96
E-mail: Vchuyko@bsu.by
Chuyko Vladislav Aleksandrovich