



DOI: 10.22363/2312-8143-2025-26-3-266-272

EDN: YICUJW

Научная статья / Research article

Генерация реалистичных изображений нефтегазовой инфраструктуры на космических снимках с использованием диффузионных моделей

В.К. Лобанов^{ID}, М.С. Кондрашина^{ID}✉, Ш.М. Гаджиев^{ID}, М.Ш. Сокибеков^{ID}

Российский университет дружбы народов, Москва, Российская Федерация

✉ 1132236536@rudn.ru

История статьи

Поступила в редакцию: 27 апреля 2025 г.

Доработана: 27 июня 2025 г.

Принята к публикации: 5 июля 2025 г.

Заявление о конфликте интересов

Авторы заявляют об отсутствии конфликта интересов.

Аннотация. В рамках исследования была изучена возможность применения методов машинного обучения, в частности генеративных моделей, для семантического редактирования космических снимков. Основное внимание уделено актуальной архитектуре на основе диффузионных моделей, способной генерировать целевые объекты непосредственно на спутниковых изображениях. Однако ввиду специфики выбранной предметной области — генерации реалистичных изображений объектов нефтегазовой инфраструктуры (таких как трубопроводы) были обнаружены существенные недостатки стандартной модели в части реализма и соответствия окружающему контексту. Для решения данной проблемы проводилось дообучение нейронной сети. Цель дообучения — улучшение качества визуализации проектных решений, связанных с трубопроводами. Предложен и детально описан методический подход к формированию специализированного обучающего набора данных. На основе реальных трасс трубопроводов в QGIS созданы пространственно привязанные векторные слои; сгенерирован набор тайлов космических снимков с точными аннотациями границ труб. Результаты экспериментального дообучения модели продемонстрировали значимое улучшение качества генерируемых изображений объектов нефтегазовой инфраструктуры на космических снимках по сравнению с показателями исходной, неадаптированной модели. Разработанная дообученная модель обеспечивает высокореалистичную генерацию трубопроводов, корректно интегрируя их в существующий ландшафт на снимке. Визуальное сравнение результатов до и после дообучения подтверждает устранение артефактов и достижение необходимого уровня детализации. Работа подтверждает эффективность подхода с формированием предметно-ориентированных датасетов и дообучением для решения специфических задач визуализации в ДЗЗ.

Ключевые слова: дистанционное зондирование Земли, нейронные сети, генеративные модели, машинное обучение, семантическое редактирование

Вклад авторов

Лобанов В.К. — общая концепция исследования, научное руководство; Кондрашина М.С. — написание текста; Гаджиев Ш.М. — написание изначального текста, визуализация; Сокибеков М.Ш. — написание текста. Все авторы ознакомлены с окончательной версией статьи и одобрили ее.

© Лобанов В.К., Кондрашина М.С., Гаджиев Ш.М., Сокибеков М.Ш., 2025



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

Финансирование

Статья подготовлена при поддержке Министерства науки и высшего образования Российской Федерации в рамках выполнения государственного задания по соглашению № 075-03-2024-059 (FSSF-2024-0005).

Для цитирования

Лобанов В.К., Кондрашина М.С., Гаджиев Ш.М., Сокибеков М.Ш. Генерация реалистичных изображений нефтегазовой инфраструктуры на космических снимках с использованием диффузионных моделей // Вестник Российского университета дружбы народов. Серия: Инженерные исследования. 2025. Т. 26. № 3. С. 266–272. <http://doi.org/10.22363/2312-8143-2025-26-3-266-272>

Generating Realistic Images of Oil and Gas Infrastructure in Satellite Imagery Using Diffusion Models

Vasily K. Lobanov^{ID}, Mariia S. Kondrashina^{ID}✉,
Shamil M. Gadzhiev^{ID}, Maksad Sh. Sokibekov^{ID}

RUDN University, Moscow, Russian Federation

✉ 1132236536@rudn.ru

Article history

Received: April 27, 2025

Revised: June 27, 2025

Accepted: July 5, 2025

Conflicts of interest

The authors declare that there is no conflict of interest.

Abstract. This study investigated the feasibility of applying machine learning methods, specifically generative models, for semantic editing of satellite imagery. The research focused on an architecture based on diffusion models capable of generating desirable objects directly on satellite images. However, significant shortcomings were identified in the standard model with regard to realism and relevance to the surrounding context, given the specific nature of the chosen subject area, namely the generation of realistic images of oil and gas infrastructure objects (such as pipelines). To address this limitation, fine-tuning of the neural network was performed. The objective of the fine-tuning was to enhance the quality of visualizing pipeline-related design solutions. A methodological approach for creating training dataset was proposed and described in detail. Based on actual pipeline routes, spatially referenced vector layers were created in QGIS, and a set of satellite image tiles with precise pipeline boundary annotations was generated. The results of the experimental fine-tuning demonstrated a significant improvement in the quality of generated images depicting oil and gas infrastructure objects in satellite imagery compared to the original, non-adapted model. The developed fine-tuned model enables highly realistic pipeline generation, effectively integrating them into the existing landscape within the image. Visual comparison of results before and after fine-tuning confirms the elimination of artifacts and the achievement of the required level of detail. This work demonstrates the effectiveness of the approach involving the creation of specific datasets and fine-tuning for solving specialized visualization tasks in remote sensing.

Keywords: earth remote sensing, neural network, generative models, machine learning, semantic editing

Authors' contribution

Lobanov V.K. — research concept, supervision; Kondrashina M.S. — writing; Gadzhiev Sh.M. — writing, visualization; Sokibekov M.Sh. — writing. All authors read and approved the final version of the article.

Funding

This paper has been supported by the Ministry of Science and Higher Education of the Russian Federation under Agreement No. FSSF-2024-0005.

For citation

Lobanov VK, Kondrashina MS, Gadzhiev ShM, Sokibekov MSh. Generating realistic images of oil and gas infrastructure in satellite imagery using diffusion models. *RUDN Journal of Engineering Research*. 2025;26(3):266–272. (In Russ.) <http://doi.org/10.22363/2312-8143-2025-26-3-266-272>

Введение

Использование космических снимков позволяет решать многие задачи. Поэтому разработка моделей и методов обработки изображений является чрезвычайно актуальной.

Цель исследования — оценка использования нейронных сетей для визуализации прокладки трубопроводов. Для решения поставленной нами задачи изучены возможности модели, предложенной авторами статьи “Tackling Few-Shot Segmentation in Remote Sensing via Inpainting Diffusion Model” [1].

Применение данной модели для прокладки трубопровода показало ее недостаточную эффективность. Для улучшения результатов использования модели провели дообучение на основе сформированного обучающего набора.

Актуальность подхода связана с применением современных нейронных сетей и решаемой проблемы визуализации нефтегазовой инфраструктуры на космических снимках.

1. Методы

1.1. Архитектура нейронной сети

Проанализируем принцип работы модели, предложенной в [1] и состоящей из нижеперечисленных «блоков».

Stable Diffusion — это генеративная диффузионная модель, предназначенная для создания высококачественных изображений из текстовых описаний или на основе других изображений. Модель работает с пространством скрытых признаков, за счет этого способна достигать высокой эффективности при работе с изображениями. Обучение модели происходит на задаче «обратного шума» — она учится постепенно восстанавливать изображение из шума. При генерации *Stable Diffusion* «движется в обратном порядке» от случайного шума к содержательному изображению [2; 3].

Основные компоненты *Stable Diffusion*:

1. **Автокодировщик** (VAE — Variational Autoencoder) — используется для перевода изображений в пространство признаков и обратно

и состоит из кодировщика (encoder), который сжимает изображение до латентного представления и декодировщика (decoder), восстанавливающего изображение из полученного кодировщиком представления [4].

Использование автокодировщика позволяет значительно ускорить обучение и генерацию по сравнению с работой с обычными пиксельными изображениями за счет снижения размерности данных и извлечения признаков.

2. **UNet** — это основной модуль, обрабатывающий скрытое пространство признаков.

Архитектура UNet: симметричная сеть с блоками *downsampling* (понижающей дискретизации) и *upsampling* (повышающей дискретизации) [5].

В каждом блоке используются следующие составляющие: остаточные связи (*residual*), внимание (*attention*) и нормализация (*group normalization*).

На вход модуля поступает зашумлённое пространство признаков, время шага в процессе диффузии (*timestep*) и текстовая подсказка (через *cross-attention*).

3. **Text Encoder** (CLIP Text Encoder) — разработанный OpenAI текстовый кодировщик CLIP (Contrastive Language — Image Pretraining) (ViT-B/32), преобразующий текстовую подсказку в вектор признаков. Эти признаки используются в UNet через *cross-attention* механизмы, чтобы направлять генерацию изображения [6].

4. **Scheduler** (DDIM/DDPM) — управляет процессом добавления и удаления шума. Во время генерации работает в режиме «обратного процесса»: начиная со случайного шума, модель постепенно приближает его к изображению.

Принцип работы этого метода заключается в следующем:

1. Текст → Вектор признаков: текст обрабатывается кодировщиком CLIP.

2. Латентное пространство: генерация начинается со случайного шума в латентном пространстве.

3. Процесс диффузии: UNet итеративно очищает шум, направляясь вектором признаков текста.

4. VAE Decoder: финальное латентное представление декодируется обратно в изображение.

RemoteCLIP — использовалась для обработки спутниковых и аэрофотоснимков (дистанционного зондирования) [7] и представляет собой адаптированную версию CLIP [6].

Модель обучается сопоставлять спутниковые изображения с текстовыми описаниями. Изображения и тексты кодируются в общее признаковое пространство, где близкие по смыслу пары находятся рядом.

RemoteCLIP содержит следующие основные компоненты:

1. Кодировщик изображения (image encoder) обычно используют ResNet [8] или Vision Transformer (ViT) [9]. Кодировщик преобразует изображение в векторные представления, с целью уменьшения объема данных для передачи используется квантование или компрессия.

2. Текстовый кодировщик (Text Encoder) (на сервере / в облаке) — использует трансформер (Transformer) для получения вектора признаков текстовой подсказки.

3. Коммуникационный интерфейс передает векторы признаков от клиента к серверу. Важными задачами являются сжатие, безопасность и стабильность передачи.

4. Fusion Module (на сервере) — сопоставляет визуальный и текстовый вектор признаков, вычисляя косинусное сходство.

Взаимодействие: Stable Diffusion + RemoteCLIP. На первом этапе Stable Diffusion используется в режиме дорисовки (inpainting) для заполнения области изображения (с заданной маской) синтетическим объектом.

Вместо текстового запроса подается визуальный образец (изображение с примером объекта нужного класса).

На следующем этапе RemoteCLIP применяется в качестве кодировщика изображения, чтобы получить вектор признаков объекта из референсного изображения.

Полученные векторы признаков передаются в Stable Diffusion как условие генерации, т.е. модель понимает, какой объект необходимо «нарисовать» в области маски.

1.2. Формирование обучающего набора

По маршрутам существующих трубопроводов в QGIS (свободная бесплатная географическая информационная система с открытым кодом) были построены векторные слои. Затем для формирования обучающей выборки для маршрутов трубопроводов были получены космические снимки и векторные данные с точным расположением труб, которые использовались для дообучения модели.

Формирование обучающего набора начинается с подготовки слоев в QGIS и преобразования координат. Загружается векторный линейный слой (маршрут трубопровода) и растровая подложка, после чего определяется необходимость перехода между системами координат. Если требуется, то все линейные объекты из слоя преобразуются к системе координат подложки.

Далее выполняется создание фрагментов карты (тайлов) — квадратных зон с центрированием вдоль линейного объекта. Количество и расположение рассчитывается с учетом процента перекрытия соседних тайлов. Каждый полученный участок добавляется в память в виде полигона с уникальным идентификатором.

После этого для каждого полученного полигона проверяется, пересекает ли его хотя бы одна из трубопроводных линий. Если пересечение обнаружено, вычисляются прямоугольные границы (bbox) в пиксельных координатах относительно тайла. Координаты полигонов bbox сохраняются в текстовых файлах в формате YOLO для дальнейшего использования в задаче обнаружения объектов.

На этапе сохранения для каждого фрагмента карты формируется временный файл GeoTIFF в пределах границ тайла, который затем преобразуется в формат JPEG с помощью инструментов GDAL с последующим удалением исходного файла. Файл с текстовыми координатами bbox сохраняется под тем же названием, что и изображение.

Полученный набор для дообучения имеет следующую рекомендованную структуру [10], представленную на рис. 1.

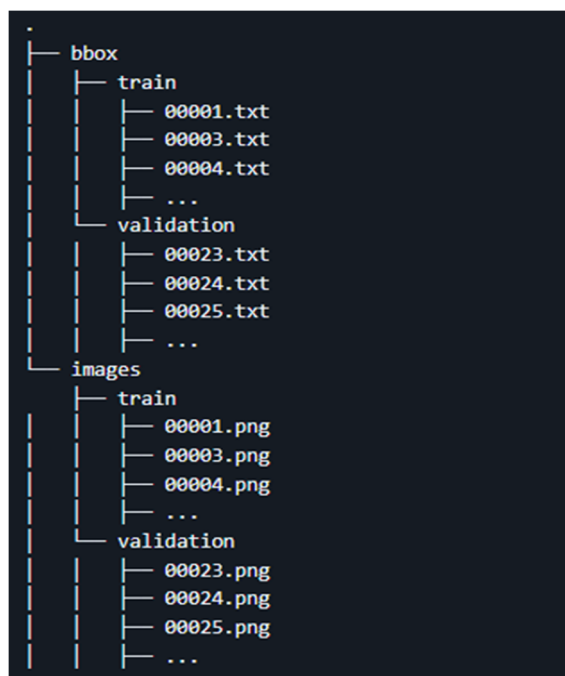


Рис. 1. Структура обучающего набора данных:
bbox — вектор объекта;
images — соответствующее изображение
Источник: выполнено Ш.М. Гаджиевым

Figure 1. Training dataset structure:
bbox — object bounding box coordinates;
images — corresponding image
Source: by Sh.M. Gadzhiev

2. Результаты и обсуждение

2.1. Дообучение

Дообучение требовалось провести из-за отсутствия в начальном обучающем наборе класса с трубами. Следовательно, модель была не способна их генерировать. Для генерации необходимы изначально изображение, маска области и непосредственно объект, который мы будем вставлять. При подаче в качестве референса участка трубопровода созданная труба выглядела как белая линия (рис. 2, а, б).

Во время дообучения можно заметить, как с каждой эпохой модель справлялась с задачей генерации труб лучше и лучше. В качестве метрики была выбрана сглаженная ошибка валидации (экспоненциальное среднее). Приведем сравнение величины на 5-й и 9-й итерации, где значение уменьшилось с 0,86 до 0,77 (рис. 3 и 4).

После дообучения модели на основе обучающей выборки с учетом вектора координат трубы на новый класс — «трубы» модель стала лучше встраивать трубопровод в космический снимок. Результат представлен на рис. 5.

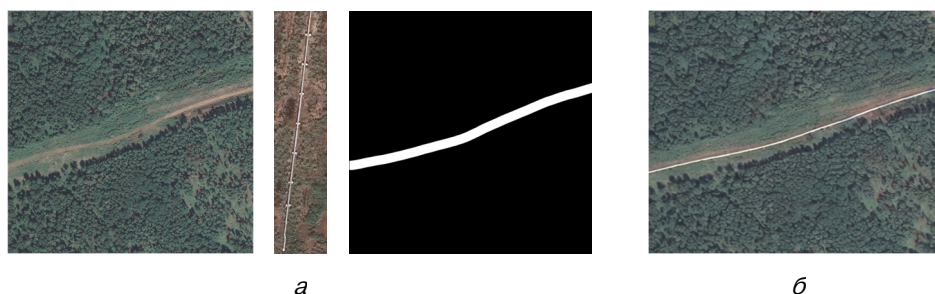


Рис. 2. Применение оригинальной модели для генерации трубопровода:
а — подаваемые параметры; б — полученный результат
Источник: выполнено Ш.М. Гаджиевым

Figure 2. Application of the original model for pipeline generation:
а — input parameters; б — obtained result
Source: by Sh.M. Gadzhiev

```
Average Peak memory 22746.30MiB
Epoch 5, global step 1949: val/loss_simple_ema reached 0.08618 (best 0.08618),
Epoch 6: 8% 40/474 [00:24<04:22, 1.65it/s, loss=0.153, v_num=0, train/loss_
Data shape for DDIM sampling is (1, 4, 64, 64), eta 1.0
```

Рис. 3. Значение потерь на 5-й эпохе
Источник: выполнено Ш.М. Гаджиевым

Figure 3. Loss value at epoch 5
Source: by Sh.M. Gadzhiev

```
Average Peak memory 22747.53MiB
Epoch 9, global step 3249: val/loss_simple_ema reached 0.07785 (best 0.07785), saving model to "/
Epoch 9: 100% 474/474 [06:13<00:00, 1.27it/s, loss=0.11, v num=0, train/loss simple step=0.0445,
```

Рис. 4. Значение потерь на 9-й эпохе
Источник: выполнено Ш.М. Гаджиевым

Figure 4. Loss value at epoch 9
Source: by Sh.M. Gadzhiev

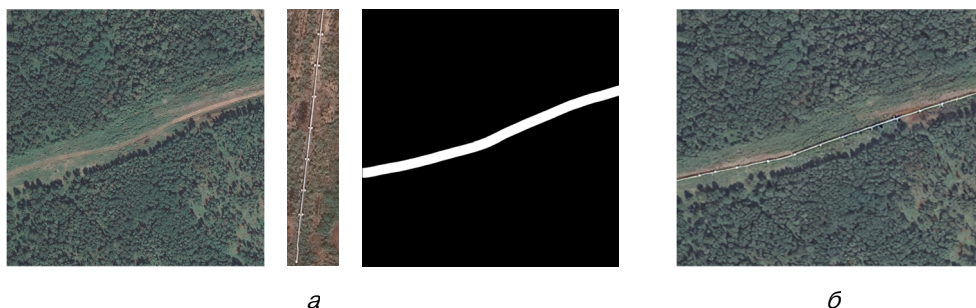


Рис. 5. Применение дообученной модели для генерации трубопровода:
а — подаваемые параметры; б — полученный результат
Источник: выполнено Ш.М. Гаджиевым

Figure 5. Application of the fine-tuned model for pipeline generation:
a — input parameters; б — obtained result
Source: by Sh.M. Gadzhiev

Заключение

В исследовании рассматривалась проблема визуализации объектов нефтегазовой индустрии на снимках дистанционного зондирования Земли. В ходе работы был изучен и доработан подход на основе диффузионных моделей. Для улучшения результатов изучаемой модели было осуществлено дообучение на основе сформированного набора данных. После нашего дообучения модели повысилось качество снимков, полученных в ходе генерации реалистичных изображений нефтегазовой инфраструктуры на космических снимках по сравнению с моделью, в референсной статье.

Список литературы / References

1. Immanuel SA, Cho W, Heo J, Kwon D. Tackling Few-Shot Segmentation in Remote Sensing via Inpainting Diffusion Model. *ICLR 2025 Machine Learning for Remote Sensing (ML4RS) Workshop*. 2025. <https://doi.org/10.48550/arXiv.2503.03785>
2. Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-Resolution Image Synthesis with Latent

- Diffusion Models. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 June 18–24; New Orleans, LA, USA. *IEEE*. 2022:10674–10685. <https://doi.org/10.1109/CVPR52688.2022.01042>
3. Panboonyuen T, Charoenphon C, Satirapod C. SatDiff: A Stable Diffusion Framework for Inpainting Very High-Resolution Satellite Imagery. *IEEE Access*. 2025;13:51617–51631. <https://doi.org/10.1109/ACCESS.2025.3551782>
4. Kingma DP, Welling M. Auto-Encoding Variational Bayes (Version 11). *International Conference on Learning Representations (ICLR)*. 2014. <https://doi.org/10.48550/ARXIV.1312.6114>
5. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention MICCAI*. 2015;9351:234–241. <https://doi.org/10.48550/arXiv.1505.04597>
6. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. Learning Transferable Visual Models from Natural Language Supervision. *Proceedings of the 38th International Conference on Machine Learning, PMLR*. 2021;139:8748–8763. <https://doi.org/10.48550/ARXIV.2103.00020>
7. Liu F, Chen D, Guan Z, Zhou X, Zhu J, Ye Q, et al. RemoteCLIP: A Vision Language Foundation Model for Remote Sensing. *IEEE Transactions on Geoscience*

and Remote Sensing. 2024;62:1–16. <https://doi.org/10.1109/TGRS.2024.3390838>

8. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016; 770–778. <https://doi.org/10.48550/ARXIV.1512.03385>

9. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An Image is Worth

16×16 Words: Transformers for Image Recognition at Scale. *International Conference on Learning Representations (ICLR 2021)*. <https://doi.org/10.48550/ARXIV.2010.11929>

10. Immanuel SA, Cho W, Heo J, Kwon D. Tackling Few-Shot Segmentation in Remote Sensing via Inpainting Diffusion Model. *ICLR 2025 Machine Learning for Remote Sensing (MLRS) Workshop*. 2025. <https://doi.org/10.48550/arXiv.2503.03785>

Сведения об авторах

Лобанов Василий Константинович, старший преподаватель кафедры механики и процессов управления, инженерная академия, Российский университет дружбы народов, Российская Федерация, 117198, г. Москва, ул. Миклухо-Маклая, д. 6; eLIBRARY SPIN-код: 7266-5340; ORCID: 0000-0001-8163-9663; e-mail: lobanov_vk@pfur.ru

Кондрашина Мария Сергеевна, магистрант кафедры механики и процессов управления, инженерная академия, Российский университет дружбы народов, Российская Федерация, 117198, г. Москва, ул. Миклухо-Маклая, д. 6; ORCID: 0009-0008-8526-9143; e-mail: 1132236536@rudn.ru

Гаджиев Шамиль Магомедминович, магистрант кафедры механики и процессов управления, инженерная академия, Российский университет дружбы народов, Российская Федерация, 117198, г. Москва, ул. Миклухо-Маклая, д. 6; ORCID: 0009-0006-1570-4133; e-mail: 1132236511@rudn.ru

Сокибеков Максад Ширинбекович, магистрант кафедры архитектура, реставрация и дизайн, инженерная академия, Российский университет дружбы народов, Российская Федерация, 117198, г. Москва, ул. Миклухо-Маклая, д. 6; ORCID: 0009-0009-0261-7374; e-mail: 1032185455@rudn.ru

About the authors

Vasily K. Lobanov, Senior Lecturer of the Department of Mechanics and Control Processes, Academy of Engineering, RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation; eLIBRARY SPIN-code: 7266-5340; ORCID: 0000-0001-8163-9663; e-mail: lobanov_vk@pfur.ru

Mariia S. Kondrashina, Master student of the Department of Mechanics and Control Processes, Academy of Engineering, RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation; ORCID: 0009-0008-8526-9143; e-mail: 1132236536@rudn.ru

Shamil M. Gadzhiev, Master student of the Department of Mechanics and Control Processes, Academy of Engineering, RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation; ORCID: 0009-0006-1570-4133; e-mail: 1132236511@rudn.ru

Maksad Sh. Sokibekov, Master student of the Department of Architecture, Restoration and Design, Academy of Engineering, RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation; ORCID: 0009-0009-0261-7374; e-mail: 1032185455@rudn.ru