РУБРИКА: ГЕОЛОГИЯ И ГЕОЛОГО-РАЗВЕДОЧНЫЕ РАБОТЫ

# Повышение точности сегментационных моделей с применением генеративного ИИ на примере интерпретации данных аэрофотосъемки

С. В. Зайцев, Д. Г. Семин, Г. С. Григорьев, М. А. Лекомцев (Nedra Digital (ООО «НЕДРА»))

В статье исследуется применение генеративных моделей искусственного интеллекта для повышения точности интерпретации аэрофотоснимков в задачах геологоразведки. Основное внимание уделено адаптации моделей Stable Diffusion XL (SDXL) с использованием модулей LoRA и ControlNet, а также сравнительному анализу с моделью Flux. Актуальность исследования обусловлена высокой стоимостью сбора новых полевых данных и ограниченностью существующих датасетов, что снижает эффективность моделей компьютерного зрения при сегментации геологических объектов. Целью работы является разработка методики автоматизированной генерации изображений с высокой визуальной и структурной достоверностью для расширения обучающей выборки и повышения точности моделей сегментации. В исследовании использовались синтетические изображения, сгенерированные на основе масок, с последующим дообучением моделей сегментации. Для оценки качества применялись метрики FID, KID, LPIPS и IoU. Полученные результаты показали, что добавление синтетических данных позволяет увеличить точность сегментации до 10.7 % по метрике IoU, а в случае редких объектов, таких как болота, прирост достигает 21.2 %. Модель SDXL с ControlNet и LoRA продемонстрировала баланс между качеством генерации и вычислительными затратами, а модель Flux показала лучшие визуальные характеристики при повышенных требованиях к ресурсам. Практическая значимость работы заключается в ускорении подготовки обучающих данных и возможности адаптации метода к различным типам геологических изображений. Перспективы включают расширение методики на другие типы данных, включая данные лазерного сканирования.

**Ключевые слова:** генеративные модели, Stable Diffusion XL, LoRA, ControlNet, аэрофотоснимки, геологоразведка, сегментация, синтетические данные.

# Improving the Accuracy of Segmentation Models Using Generative AI: A Case Study on Interpreting Aerial Photographic Data

S. V. Zaitsev, D. G. Semin, G. S. Grigoryev, M. A. Lekomtsev (“Nedra Digital” (LLC “Nedra”))

Modern geological exploration faces significant challenges, including the high cost of field data collection, limited availability of labeled datasets, and the need for rapid and accurate interpretation of aerial photographic (AP) data. Traditional visual interpretation relies heavily on manual labeling and expert involvement, which are time-consuming, subjective, and constrained by dataset completeness. A key limitation in applying computer vision (CV) models to geological tasks is the scarcity of annotated data, especially for rare or small-scale objects such as wetlands, water bodies, and geomorphological features. Collecting and annotating such data requires extensive fieldwork and domain expertise, making it difficult to train robust segmentation models.

In this context, generative artificial intelligence (AI) models offer a promising solution. Specifically, diffusion-based models like Stable Diffusion XL (SDXL), enhanced with LoRA (Low-Rank Adaptation) and ControlNet modules, enable the generation of high-fidelity synthetic images from semantic masks and text prompts. This approach allows for the expansion of training datasets without additional fieldwork, thereby improving model generalization and accuracy.

This study investigates the use of SDXL with LoRA and ControlNet for generating realistic synthetic aerial images to augment training data for geological object segmentation. We also conduct a comparative analysis with the Flux model, evaluating performance in terms of image quality, structural fidelity, and computational efficiency. Synthetic images were used to fine-tune segmentation models, and their impact was assessed using standard metrics: FID, KID, LPIPS, and IoU.

The results demonstrate that incorporating synthetic data increases segmentation accuracy by up to 10.7 % in IoU, with gains reaching 21.2 % for rare classes such as wetlands. SDXL with ControlNet and LoRA provides an optimal balance between generation quality and computational cost, while Flux achieves superior visual realism at higher resource demands.

The proposed methodology enables faster preparation of training datasets and can be adapted to various types of geological imagery. Its practical significance lies in reducing reliance on costly field campaigns and improving model performance in data-scarce scenarios. Future work includes extending the approach to other geospatial data types, such as LiDAR point clouds.

**Keywords:** generative AI, diffusion models, geological exploration, large language models, aerial survey.

## Введение

Одним из наиболее перспективных направлений в развитии искусственного интеллекта (ИИ) в последние годы стали **диффузионные генеративные модели** [1], которые демонстрируют высокое качество синтеза изображений за счет итеративного восстановления сигнала из шума. Эти модели, включая Stable Diffusion и ее модификации [2], работают на основе стохастических процессов, приближая сложные распределения данных через обучение генерации изображений в несколько этапов. В отличие от генеративно-состязательных сетей (GAN), диффузионные модели обеспечивают стабильность обучения, более высокую детализацию и контролируемость генерации [3], особенно в сочетании с дополнительными модулями управления, такими как **ControlNet**. Это делает их особенно привлекательными для задач, где требуется воспроизведение сложных структур, характерных для изображений, используемых в геологоразведке, например для аэрофотосъемки (АФС).

Современные задачи геологоразведки (ГРР) требуют высокой точности при интерпретации пространственно-геологических объектов, представленных на АФС. Такие изображения широко используются для анализа водных объектов, геоморфологических структур и других особенностей поверхности на всех стадиях планирования ГРР. Однако традиционные методы интерпретации, основанные на ручной разметке или классических алгоритмах компьютерного зрения (CV), ограничены трудоемкостью, субъективностью анализа и высокой зависимостью от полноты и разнообразия обучающих выборок [4, 5].

Одной из ключевых проблем в применении моделей CV в геологоразведке является дефицит размеченных данных, особенно для редких и структурно сложных объектов. Полноценный сбор и аннотирование новых полевых данных занимают от нескольких месяцев до полугода, что значительно тормозит внедрение ИИ-решений. Использование генеративных моделей позволяет синтезировать изображения, соответствующие структурным и визуальным особенностям реальных объектов, сокращая зависимость от полевых данных.

Особый интерес представляют диффузионные модели, такие как **Stable Diffusion XL (SDXL)**, которые обеспечивают высокую визуальную достоверность синтетических изображений. В сочетании с адаптерами **LoRA (Low-Rank Adaptation)** [6] и **ControlNet** [7] эти модели могут быть настроены на ограниченных датасетах, обеспечивая генерацию изображений, структурно соответствующих входным маскам. Это открывает возможности для генерации изображений с разметкой «по маске» и стилизации под конкретные геологические условия.

Целью настоящего исследования является разработка и оптимизация подхода на базе SDXL, LoRA и ControlNet для автоматизированной генерации синтетических аэрофотоснимков геологических объектов и повышения точности их сегментации. В работе также рассматривается сравнение с моделью **Flux**, обладающей более высоким уровнем детализации, но требующей значительных вычислительных ресурсов.

Научная новизна заключается в адаптации архитектур диффузионных моделей к специфике геологических задач. В отличие от существующих решений, данная методика ориентирована на воспроизведение структурной геометрии объектов, таких как водоемы и болота, с последующим дообучением моделей сегментации.

## Материалы и методы

Разработанная методика генерации новых изображений на основе небольшого числа опорных кадров и текстовых описаний включает следующие шаги (рис. 1):

1. Обучение адаптера LoRA на доступной части изображений для переноса стиля генерируемых изображений. Требуется не более десятка размеченных изображений с текстовым описанием (будущий промпт для генерации).

2. Использование полученного на предыдущем этапе адаптера LoRA и готового ControlNet адаптера для генерации изображений на основе входных масок: генерация изображений сразу с разметкой, что упрощает процесс дообучения моделей.

3. Дообучение базовой модели на новом датасете, в том числе сравнение качества сегментирования объектов по классам.


***Рисунок 1**. Процесс создания синтетической выборки для дополнения обучающего датасета*

*Данные и их подготовка*

Для обучения базовой модели сегментации водного слоя и валидации были собраны датасеты из открытых источников с высоким качеством разметки и реальные данные АФС. Всего 12 наборов аэрофотоснимков: 2 набора реальных лицензионных участков (ЛУ), 7 наборов АФС из открытых источников, 3 набора спутниковых снимков.

Все наборы снимков были приведены к общему формату разметки (бинарные маски) и к общему размеру изображения (1280 × 1280 пикс.), в снимках с изображениями высокого разрешения и мелкими деталями изображения были разбиты на части.

Данные были разбиты на обучающую (12 187 снимков), валидационную (671 снимок) и 2 тестовые выборки (671 и 80 снимков соответственно). Во второй тестовый набор полностью вошел один ЛУ, на котором в дальнейшем тестировался сценарий генерации изображений и валидации модели сегментации. Остальные части набора данных были разбиты в соотношении 90/5/5 (обучающий/валидационный/тестовый).

## Архитектура моделей

В качестве основной архитектуры использовалась **Stable Diffusion XL (SDXL)** — латентная диффузионная модель с улучшенной структурой декодирования. Для адаптации модели под специфические геологические объекты применялись:

- **LoRA (Low-Rank Adaptation)** — модуль, позволяющий дообучать только малую часть параметров, что особенно полезно при ограниченном объеме данных;

- **ControlNet** — модуль управления генерацией на основе входных масок, интегрированный в процесс итеративного восстановления изображения.

Дополнительно проводилось сравнение с моделью **Flux**, которая продемонстрировала лучшие результаты в задачах с высокими требованиями к детализации, но при существенно большей ресурсоемкости (время генерации изображения — ~28 с. против ~3 с. у SDXL).

В качестве базовой сегментационной модели использовалась модель **YOLOv11m**.

## Обучение и настройка гиперпараметров

Обучение моделей проводилось на вычислительном кластере с графическими ускорителями NVIDIA A100 (40 ГБ). Были проведены эксперименты по оптимизации гиперпараметров адаптера LoRA и ControlNet с использованием байесовской оптимизации для поиска оптимальных гиперпараметров.

Основные гиперпараметры адаптера LoRA, влияющие на результат:

1. **Rank (ранг)**: определяет размерность низкоранговой матрицы. Более высокий ранг может улучшить качество адаптации, но увеличивает вычислительные затраты.

2. **Alpha (коэффициент масштабирования)**: влияет на силу адаптации. Большие значения alpha могут привести к более сильным изменениям в модели.

Результаты оптимизации гиперпараметров LoRA (рис. 2):

1. Использовать learning rate = 1e-5 для обучения модели SDXL с адаптером LoRA.

2. Установить batch size = 32/64 для баланса между скоростью и стабильностью обучения.

3. Применять адаптер LoRA с рангом = 64 и alpha = 32/64 для эффективной адаптации модели.

Полученные изображения демонстрируют высокую реалистичность (рис. 2).

***Рисунок 2**. Результат генерации SDXL + LoRA с гиперпараметрами: lr = 1e-5, rank = 64, batch_size = 128, alpha = 64, epochs = 1000*

Аналогичным образом была проведена оптимизация гиперпараметров для ControlNet.

Основные параметры адаптера, влияющие на результат:

- **Network Dim (размерность сети)**: определяет сложность адаптера.
- **Cond Emb Dim (размерность условного эмбеддинга)**: влияет на способность модели обрабатывать условия (в нашем случае бинарные маски).
- **ControlNet Multiplier**: регулирует силу влияния ControlNet на генерацию.
- **Network (LoRA) Multiplier**: регулирует силу влияния LoRA на генерацию.

Основным важным выводом является гипотеза, что должно всегда выполняться неравенство Network Dim < Cond Emb Dim с соотношением ((Network Dim / Cond Emb Dim) <= 0.5) для выполнения наилучшей генерации изображений. Также заметно, что при ControlNet Multiplier > 1.0 появляются артефакты генерации, но совпадение маски высокое. Отлично видно различие качества генерации при выполнении (рис. 3) и невыполнении условий соотношения гиперпараметров.

***Рисунок 3.*** *Маска и результат генерации с гиперпараметрами. Сверху: lr = 1e-5 network_dim = 512 batch = 64 cond_emb_dim = 256 epoch = 1000 (не выполняется условие Network Dim < Cond Emb Dim), снизу: lr = 1e-5 network_dim = 128 batch = 64 cond_emb_dim = 256 epoch = 1000 (условие Network Dim < Cond Emb Dim выполнено)*

## Результаты оптимизации гиперпараметров ControlNet

1. Использовать learning rate = 1e-5 для обучения модели SDXL с адаптером ControlNet.
2. Установить batch size = 64/128 для баланса между скоростью и стабильностью обучения.
3. Количество эпох 750/1000.
4. Применять адаптер ControlNet с выполнением неравенства ((Network Dim / Cond Emb Dim) <= 0.5): оптимально Network Dim = 256/512, Cond Emb Dim = 1024.
5. Использовать константное значение control_net_multipliers = 1.0 и попробовать варианты lora_multipliers = [0.8, 0.85, 0.9] для минимизации эффекта размытия.
6. Параметр steps при генерации поддерживать в диапазоне [40, 60], но этот параметр не является ключевым, поэтому его можно зафиксировать на значении 40.

*Оценка точности сегментации*

Для оценки качества генерации и последующей сегментации использовались следующие метрики:

- **FID (Fréchet Inception Distance)** — сравнение распределений признаков реальных и сгенерированных изображений;
- **KID (Kernel Inception Distance)** — менее чувствительная к объему выборки альтернатива FID;
- **LPIPS (Learned Perceptual Image Patch Similarity)** — оценка перцептивного сходства;
- **IoU (Intersection over Union)** — стандартная метрика точности сегментации.

При использовании метрик для оценки качества сгенерированных изображений в данной работе руководствовались следующими правилами:

1. **Использовать комбинацию метрик.**
2. **Учитывать размер выборки:**

- для FID рекомендуется не менее 5000 изображений для надежных результатов,
- для KID можно использовать меньшие выборки (50 изображений).

3. **Отслеживать динамику метрик** во время обучения или оптимизации модели.
4. **Количественные метрики дополнять качественной экспертной оценкой** для выявления конкретных проблем, которые могут не отражаться в числовых показателях.

## Оптимизация промптов для улучшения результатов генерации изображений

Один из параметров, который наиболее влияет на улучшение качества генерации изображений, — это описание генерируемого изображения (промпт). Анализ существующих промптов показывает, что они имеют недостаточную детализацию, отсутствуют указания на стиль или контекст, неоднозначно сформулированы.

Автоматическая генерация улучшенных текстовых промптов с помощью LLM позволяет повысить релевантность и детализацию сгенерированных

изображений, а также обеспечить более точное соответствие метрикам качества.

1. Выбраны несколько открытых моделей:

- GPT-neo-1.3B,
- Phi-2,
- Phi-4-reasoning-plus,
- Mistral-7B-v0.1,
- Qwen3-8B.

2. Генерация улучшенных промптов. Использованы следующие подходы:

- добавление ключевых слов: включение терминов, таких как detailed, sharp focus, vibrant colors;
- контекстуализация: указание на окружение, освещение, ракурс (например, scene with soft morning light);
- структурирование промпта: разделение на основной объект, фон, стиль и технические параметры.

3. Автоматизация процесса:

- разработан скрипт для автоматической генерации улучшенных промптов для LLM:

```
instructions = [
    f"Improve the following description: {description}",
    f"Rephrase this description in a natural and vivid way: {description}",
    f"Add more details to make this description engaging: {description}",
    f"Create a new, detailed description based on this: {description}"
];
```

- с помощью NLP-метрик BLEU и ROUGE выбираем лучший промпт, который подадим на вход LLM;
- генерируем новое описание картинки с помощью выбранного промпта.

Сравнение изображений, сгенерированных с оптимизацией различными LLM моделями, представлено на рисунке 4.

***Рисунок 4.*** *Генерация изображений при различной оптимизации промптов: верхний ряд — исходный датасет (LPIPS = 0.65), средний ряд — Microsoft_Phi-2 (LPIPS = 0.77), нижний ряд — Microsoft_Phi-4-reasoning-plus (LPIPS = 0.7)*

## Результаты и обсуждение

*Оценка точности сегментации с использованием сгенерированных изображений*

Базовая модель сегментации — **YOLOv11m** [8], предобученная на открытых наборах данных, — использовалась для задачи выделения водных объектов. Для оценки точности применялась метрика IoU при пороге уверенности 0.3. Для понимания качества работы модели производилась оценка **по классам**: отдельно анализировалась точность сегментации **водного слоя** и **болот**, как часто недопредставленных классов.

*Обучение модели для генерации изображений на основе масок*

Для улучшения качества синтезируемых изображений и исключения процесса разметки данных для модели SDXL был обучен адаптер ControlNet, использующий изображение, текстовое описание и маску. Генерация новых данных с использованием масок позволила создать около 1000 изображений, включенных в обучающий датасет. Это привело к улучшению метрики IoU на тестовом наборе данных на 10.7 %.

*Использование генерации для редких классов объектов*

Основное преимущество использования генеративных моделей выявлено в случае работы с редкими объектами, такими как болота. В стандартных обучающих выборках подобного рода данных крайне мало, что значительно снижает точность предсказаний для этих объектов при использовании базовой модели. Дообучение стилистике болот и генерация 350 дополнительных их изображений с разметкой, включенных в обучающую выборку, позволило повысить точность сегментации болотистых территорий на 21.2 % (рис. 5).

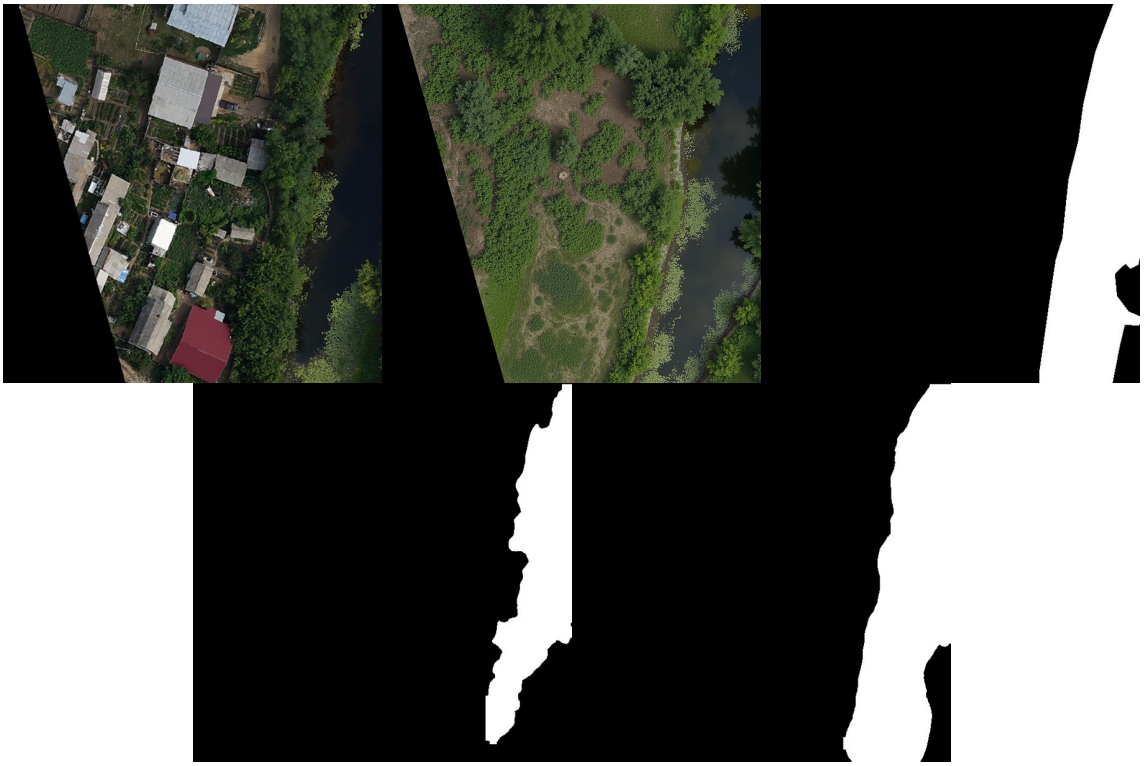

***Рисунок 5.*** *Пример повышения качества сегментации базовой модели (слева направо, сверху вниз): реальное изображение АФС из тестовой выборки, сгенерированное изображение, истинная маска болота, результат сегментации базовой модели, результат сегментации модели, дообученной на сгенерированном датасете*

*Сравнение диффузионных моделей Flux и SDXL*

Было проведено сравнение моделей SDXL и Flux с точки зрения качества генерации, скорости работы и применимости (рис. 6).

***Рисунок 6.*** *Примеры генерации изображений SDXL с ControlNet и Flux по одной и той же маске (слева направо, сверху вниз): реальное изображение АФС из тестовой выборки, маска водного объекта, генерация по маске модели Flux, генерация по маске модели SDXL*

Сравнивая полученные метрики для двух моделей, видим:

1. **FID**: SDXL показывает немного лучший результат (164.26 против 167.23), что свидетельствует о том, что общее распределение признаков сгенерированных изображений немного ближе к реальным. Однако разница составляет всего ~1.8 %, что не является существенным преимуществом.

2. **KID**: Flux демонстрирует лучший результат (0.0423 против 0.0447), что предполагает более высокое качество генерации на уровне локальных деталей. Преимущество составляет ~5.7 %.

3. **LPIPS**: Flux значительно превосходит SDXL (0.603 против 0.665), указывая на то, что изображения, сгенерированные Flux, перцептивно ближе к реальным. Разница составляет ~9.3 %, что является наиболее заметным различием между моделями.

Проведенный анализ показывает, что обе модели демонстрируют сопоставимое качество генерации, с некоторыми различиями в сильных сторонах:

- **SDXL** имеет небольшое преимущество в глобальном сходстве набора изображений с реальными (FID);

- **Flux** показывает лучшие результаты в точности воспроизведения локальных деталей (KID) и перцептивном сходстве отдельных изображений с реальными (LPIPS).

Учитывая, что LPIPS наиболее точно коррелирует с человеческим восприятием качества изображений, можно сделать вывод, что **Flux** в целом генерирует изображения, которые субъективно оцениваются как более реалистичные. Flux показал более высокий уровень детализации, но при этом требует значительных вычислительных ресурсов и имеет ограничения в лицензировании. SDXL, напротив, обеспечил более быстрое время генерации (3 секунды против 28 секунд у Flux), что делает его более подходящим вариантом для практического использования в коммерческих проектах. Полноценное сравнение различных моделей, обсуждаемых в статье, представлено в таблице 1.

***Таблица 1.*** *Критерии сравнительной оценки используемых моделей*

| **Критерий** | **Качество/ реализм** | **Нативное разрешение** | **Лицензия** | **Объем параметров** | **Ключевые ограничения** |
|---|---|---|---|---|---|
| **Stable Diffusion 1.5** | среднее | 512 × 512 | CreativeML OpenRAIL-M (коммерчески свободна) | ~0,98 млрд | деградация детализации при апскейле; ограниченное понимание сложных промптов |
| **Stable Diffusion 2.1** | выше SD 1.5, но ниже SDXL | 768 × 768 | CreativeML OpenRAIL++ | ~0,86 млрд | недостаточная детализация при 1024 px; заметные артефакты текста |
| **Flux 1.0 (dev)** | наилучшее среди сравниваемых | 1024 × 1024 | **Non-Commercial** | ~12 млрд | запрещено коммерческое использование; высокие требования к VRAM |
| **Stable Diffusion XL 1.0** | **высокое** | 1024 × 1024 | CreativeML OpenRAIL-M | ~3,5 млрд | несколько выше потребление GPU-памяти по сравнению с SD 1.x/2.x |

*Генерация псевдо-3D-данных*

Часто для сегментации сложных объектов недостаточно использования только АФС, и специалистами привлекаются данные из воздушного лазерного сканирования (ВЛС). ВЛС несет информацию о расстоянии до объекта и интенсивности отражения. Для аугментации такого рода датасетов необходимо генерировать дополнительно слой интенсивности глубины отражения от объекта.

Применение модели Depth Anything V2 [9], основанной на DPT и DINOv2 [10], позволило генерировать карты глубины на основе синтетических изображений. Это открывает путь к созданию наборов псевдо-3D-данных, пригодных для построения цифровых моделей рельефа, геоморфологического анализа и пространственного планирования в геологоразведке. Процесс обучения модели аналогичен представленному на рисунке 1, однако в обучающей выборке дополнительно присутствуют данные ВЛС на одной площади с АФС.

Для количественной оценки качества модели, предсказывающей карты глубины, используется метрика Accurancy Under threshold (d1) — эта метрика показывает долю пикселей, где отношение между предсказанным и истинным значением глубины находится в пределах заданного порога. То есть чем выше значения метрики, тем лучше модель предсказывает глубину. Например, значение 95 % означает, что 95 % пикселей имеют ошибку менее уровня threshold. В нашем случае threshold был выбрал 50 %.

Представленная на рисунке 7 способность модели восстанавливать глубину открывает широкие возможности по генерации сложных датасетов для сегментации объектов, где требуется наличие объемных атрибутов, например при задачах определения категории рубки леса и определении древостоя.

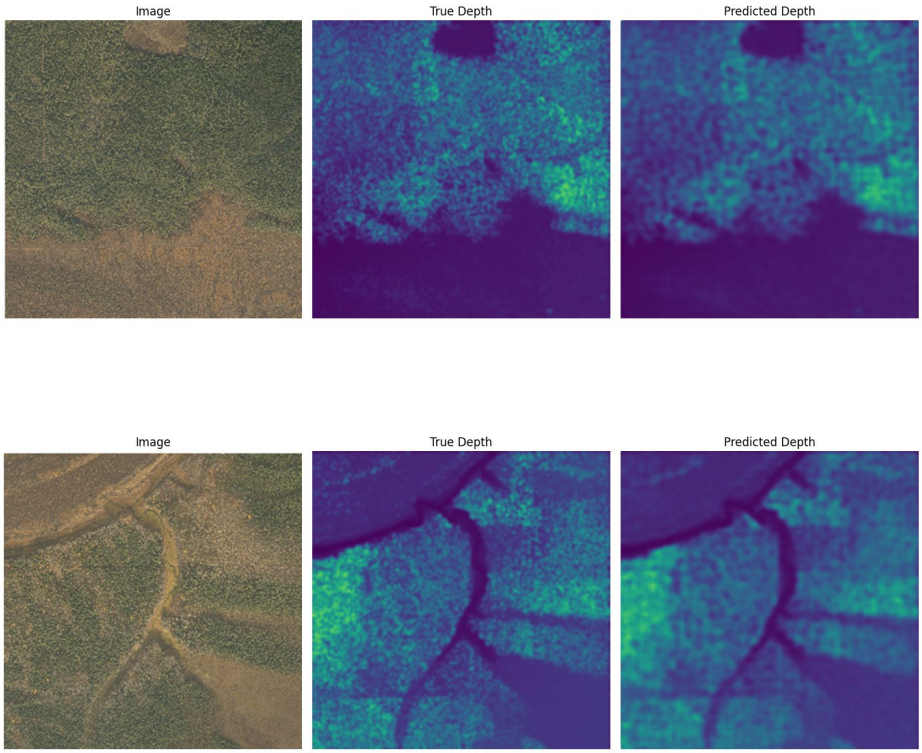


***Рисунок 7**. Пример карты глубины, полученной на основе сгенерированного изображения с метрикой d1 > 0.8*

## Выводы

Разработанная методика генерации синтетических изображений с готовыми масками для обучения CV-моделей позволяет в короткие сроки увеличивать точность моделей сегментации: генерация одного изображения занимает не более 10 секунд. Возможность создания изображений под необходимую стилистику позволяет увеличивать точность определения уникальных объектов, слабо представленных в обучающей выборке, что на практике позволяет не выполнять дополнительные измерения или разметку для расширения датасетов при сильном дисбалансе классов. Это обеспечивает быстрое добавление объектов в бизнес-сценарии работы CV-моделей.

Основные выводы по результатам проведенных исследований:

1. оптимизирована архитектура SDXL + LoRA + ControlNet для применения в геологоразведке;

2. модель Flux обеспечивает лучшее визуальное качество, но требует больших вычислительных ресурсов;

3. добавление синтетических изображений улучшает сегментацию (до + 21 % по IoU) на примере данных сегментации болот на АФС;

4. подход существенно сокращает сроки подготовки данных в сравнении с ручной разметкой;

5. впервые показана применимость диффузионных моделей к геологическим изображениям;

6. продемонстрирована возможность генерации псевдо-3D-данных на основе синтетических аэрофотоснимков.

Дальнейшим развитием представленного подхода является тестирование методики на более сложных объектах АФС (деревья), расширение на сложные геологические изображения (например, сейсмические кубы, карты атрибутов сейсморазведки и т. д.) для расширения потенциальных бизнес-сценариев применения методики. Также в ближайшее время планируется проведение экспериментов по оценке точности включения трехмерных сгенерированных данных в обучающую выборку.

## Список литературы / References


1. Rombach R. High-Resolution Image Synthesis with Latent Diffusion Models / Rombach R. et al. // CVPR, 2022.

2. Goodfellow I. Generative Adversarial Nets / Goodfellow I. et al. // NeurIPS, 2014.

3. Dhariwal P. Diffusion Models Beat GANs / Dhariwal P., Nichol A. // NeurIPS, 2021.

4. Zaytsev S. V. Multimodal Generative Models in Geodata Interpretation / Zaytsev S. V. et al. // Conf. on AI in O&G, 2024.

5. Richter S. Playing for Data / Richter S. et al. // ECCV, 2016.

6. Hu E. J. LoRA: Low-Rank Adaptation of Large Language Models / Hu E. J. et al. // arXiv:2106.09685.

7. Zhang L. ControlNet: Conditional Control for Diffusion Models / Zhang L. et al. // arXiv:2302.05543.

8. YOLO Documentation. — URL: https://docs.ultralytics.com/ru.

9. Depth Anything V2. Internal Report, 2025.

10. Carion N. Vision Transformers for Dense Prediction / Carion N. et al. // ECCV, 2020.