

Правительство Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
**"Национальный исследовательский университет
Высшая школа экономики"**

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

**Проект 1515. Исследование современных технологий
генерации изображений по фото и текстовому описанию**

Выполнил:
Колодин Матвей

Руководитель:
Сластников Сергей Александрович

Москва, 2024

Содержание

1	Наименование темы проекта	3
2	Актуальность проекта	3
3	Назначение результатов проекта	3
4	Цель разработки/исследований	3
5	Требования к результату проекта	4
6	Содержание и сроки этапов выполнения проекта — календарный план (мероприятия, задачи, сроки, исполнители, ожидаемые результаты)	4
7	Требования к отчетным материалам, предъявляемым к защите проекта	6

1 Наименование темы проекта

Проект посвящен следующей теме: **"Исследования применимости технологий генерации изображений для создания и обогащения датасетов в области компьютерного зрения"**.

2 Актуальность проекта

В последние годы технологии искусственного интеллекта, в частности генерация изображений, достигли значительного прогресса. К примеру, в 2022 году картина сгенерированная с помощью MidJourney выиграла первый приз в конкурсе Colorado State Fair's annual art competition.

Прогресс в области генерации изображений открыл новые горизонты возможностей. Одним из потенциальных применений данного прогресса может являться автоматическая генерация датасетов, что может существенно повлиять на развитие и эффективность моделей машинного зрения, а также на скорость формирования датасетов. Проект актуален для понимания ограничений и потенциала данных технологий.

3 Назначение результатов проекта

В ходе активной разработки проекта, посвященного созданию и оптимизации технологий для распознавания людей на изображениях с верхнего ракурса, снятых камерами с эффектом 'рыбьего глаза', было выявлено ключевое ограничение: отсутствие доступных размеченных датасетов, соответствующих специфике поставленной задачи. Данная проблема обусловлена уникальностью условий съемки, что делает невозможным применение существующих датасетов, находящихся в открытом доступе, из-за чего приходилось собирать и размечать фотографии в ручную. В связи с этим возникла идея использования нейросетевых технологий для генерации изображений.

Результаты проекта позволят понять, насколько эффективно можно использовать сгенерированные искусственным интеллектом изображения для улучшения качества систем распознавания людей, а также сравнение скорости и эффективности генерации данных с помощью ИИ с традиционным методом ручного фотографирования и разметки.

4 Цель разработки/исследований

Учитывая, что размеченные датасеты для такой узкоспециализированной задачи отсутствуют в интернете, проект также направлен на разработку методологии для создания качественных датасетов, приближенных к условиям реального мира.

Результаты этой работы помогут не только конкретной компании в решении специфической задачи распознавания людей, но и внесут значимый вклад в область искусственного интеллекта и машинного зрения. Это исследование позволит понять потенциал и ограничения применения современных технологий генерации изображений, а также оценить практическую применимость подобных подходов в разработке систем компьютерного зрения.

Результаты проекта позволят ответить на следующие вопросы:

1. Возможно ли повысить качество модели, дообучив ее на сгенерированных данных?
2. Если да, то можно ли достичь тех же показателей, что и на данных собранных в ручном режиме?
3. Оценка трудоемкости расширения датасета данным способом.

5 Требования к результату проекта

Можно выделить следующие требования:

1. Исследования качественных и количественных ограничений инструментов генерации изображений
2. Подготовка публикации или выступления на конференции по тематике ИИ

6 Содержание и сроки этапов выполнения проекта — календарный план (мероприятия, задачи, сроки, исполнители, ожидаемые результаты)

По плану проекта выделены следующие ключевые этапы:

1. Поиск источников и докладов по данной тематике
 - **Период деятельности:** с 29.03.23 по 1.05.23
 - **Результат:** в течение всего времени поиска не было найдено проектов по аналогичной или схожей тематике. Возможно, по состоянию на начало 2024 года ситуация изменилась, но при поиске каких-либо докладов и статей не было найдено.
2. Изучение текущих возможностей нейросетей
 - **Период деятельности:** с 29.03.23 по текущее время.

- **Результат:** с течением времени возможности генерации изображений, количество опций, функций, а также само качество генерируемых изображений менялось.

На примере сети Midjourney можно видеть, что ее 6 версия показывает поистинне удивительные результаты по уровню реалистичности, детальности и проработанности.

За время ее существования были добавлены функции 'Upscaling', 'Variations', 'Zoom Out' и т.д.; параметры 'Stylize' и 'Quality'; общие улучшения кластера, на котором произошло обучение и кодовой базы.

Поэтому было важно отслеживать новые возможности в области генерации изображений на протяжении всего времени выполнения задачи.

3. Подготовка промтов (текстовых описаний генерируемых изображений)

- **Период деятельности:** с 26.11.23 по 12.12.23
- **Результат:** были получены файлы с содержимым генерируемых сцен. Всего получилось 150 текстовых описаний на двух языках - русском и английском.

4. Изучение влияния тегов на результат генерации, подготовка шаблона тегов

- **Период деятельности:** с 7.10.23 по 10.11.23
- **Результат:** был подготовлен шаблон из тегов генерации, который показывал наибольшее улучшение качества генерируемых изображений. Отобор проводился по многим параметрам, после чего проводился качественный анализ на практике - изображения генерировались, а после чего сравнивались между собой.

5. Подготовка программного обеспечения для удобства генерации запросов к нейросетям

- **Период деятельности:** с 1.11.23 по 24.11.23
- **Результат:** было разработано приложение, позволяющее комфортно создавать запросы для генерации изображений. Более детально его возможности описаны в отчетности.

6. Генерация изображений

- **Период деятельности:** с 1.01.24 по 8.01.24

- **Результат:** был получен датасет сгенерированных, а в последующем обработанных изображений в количестве 1020 штук.
Также был получен код генерации изображений через API Kadinsky, что кратно ускоряло получение результатов.

7. Разметка изображений

- **Период деятельности:** с 9.01.24 по 11.01.24
- **Результат:** при помощи Grounded-Segment-Anything был довольно оперативно размечен полученный датасет. В последующем разметка была переведена в формат Pascal VOC.

8. Обучение и сравнение результатов. Получение ответов на поставленные вопросы

- **Период деятельности:** с 14.01.24 по 19.01.24
- **Результат:** Получены модели, а также сравнительные данные их качества.

7 Требования к отчетным материалам, предъявляемым к защите проекта

Требуется наличие следующих материалов для защиты проекта:

1. Титульный лист
2. Список исполнителей
3. Реферат
4. Датасет из обработанных и необработанных изображений
5. Разметка датасета из обработанных изображений
6. Обученные модели и их метрики