

Personalized Anime Character Generation

Команда №73

Состав команды: Гончаров Антон, Чайчук Михаил

Куратор: Неудачина Ева

Репозиторий:

<https://github.com/HSE-AI-Master-s-Team73-1st-Year-Project/Team73-Annual-Project>

Постановка задачи и целей



Задача: разработать модель, которая будет генерировать аниме-версию переданного изображения с возможностью дополнительной опоры на текстовый промпт.

Цели:

- Выбор модели + сбор данных
- Предобработка данных и их анализ
- Разработка модели-бейзлайна
- Создание прототипа сервиса (MVP)
- Улучшение результатов, создание модели на основе более современных статей
- Тестирование полученной модели, внедрение в сервис
- ТГ бот (опционально)

Описание данных

[Multi-Modal-CelebA-HQ](#): 30000 изображений лиц знаменитостей в разрешении 1024x1024

10 текстовых описаний на картинку + 1 описание, сгенерированное нами с помощью модели [blip2](#). В описаниях основной акцент сделан на конкретные черты внешности.

19812



22735



16916



7981



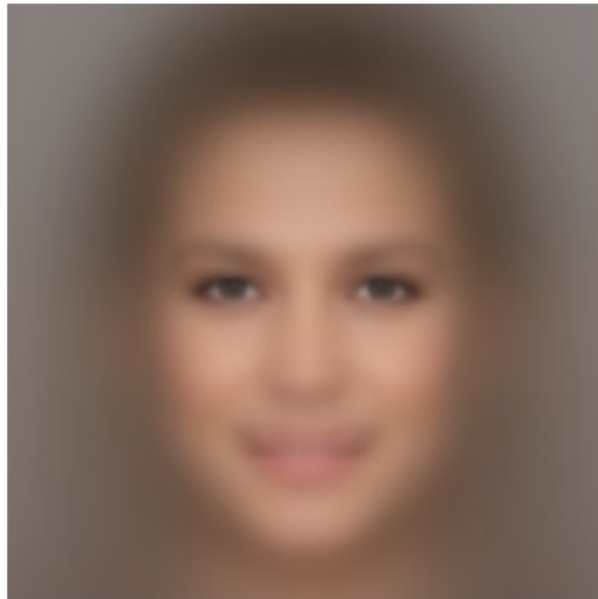
22706



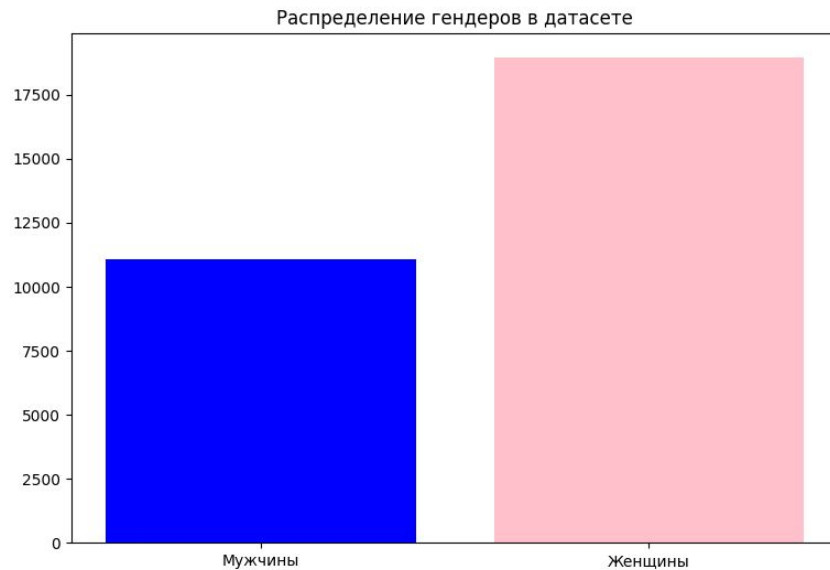
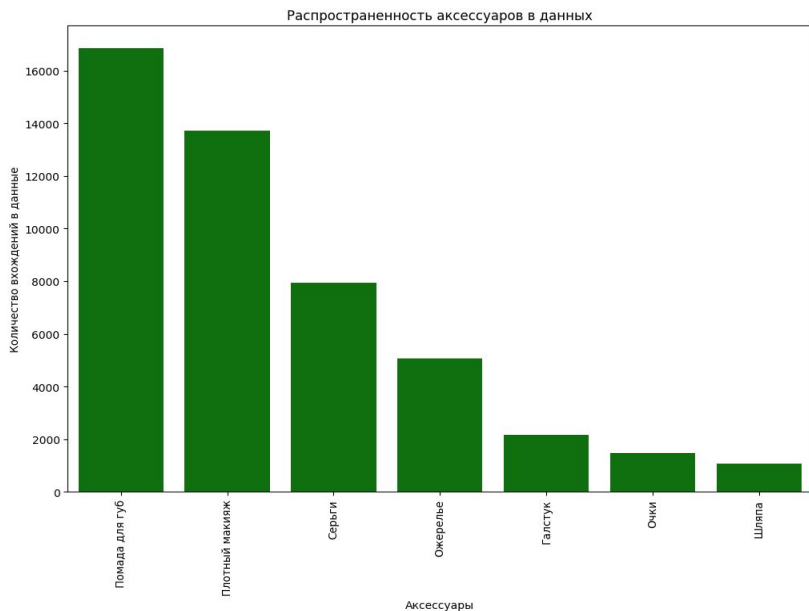
EDA (изображения)

- Средние значения цвета изображения: красный (~105-155), зеленый (~85-130), синий (~75-115).
- Усредненное изображение имеет более женские черты лица
- Высокая отчетливость глаз.

Среднее изображение датасета



EDA (текст)



Метрики качества



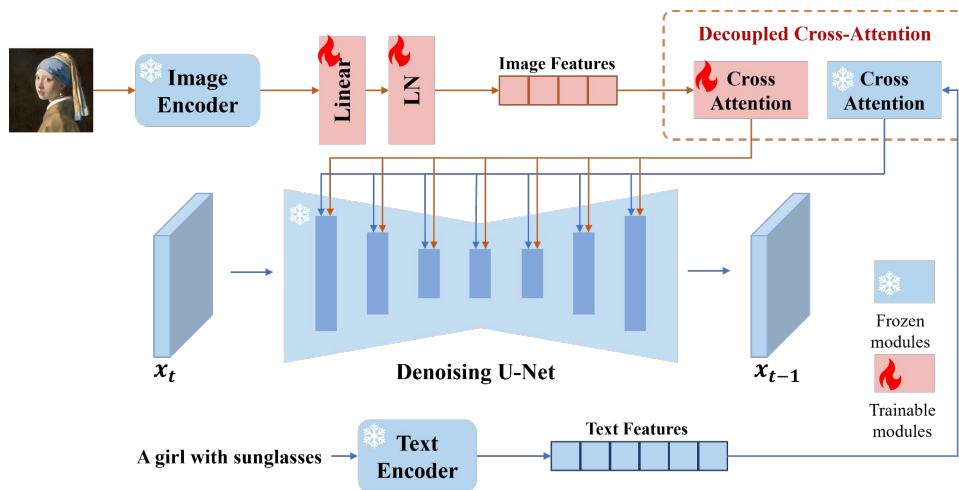
Проблема: На данный момент нет общепризнанной и качественной метрики для задач, подобных нашей.

На основе релевантных статей мы выбрали:

- **FID** (Freschet Inception Distance) между сгенерированными изображениями и изображениями из [Anime Face Dataset](#).
- **CLIP-I** - косинусный коэффициент между эмбедингами картинки-промпта и сгенерированной картинки.
- **CLIP-T** - CLIP-score между сгенерированной картинкой и текстом-описанием картинки-промпта.
- **Визуальная оценка**

Бейзлайн-модель. IP-Adapter.

В качестве бейзлайна мы решили реализовать **IP-Adapter** — метод файнтюнинга диффузионных генеративных моделей, предложенный в статье [IP-Adapter: Text Compatible Image Prompt Adapter for Text-to-Image Diffusion Models](#)



Бейзлайн-модель. IP-Adapter.



- Модель реализована на основе кода от разработчиков.
- Обучение производилось с гиперпараметрами по умолчанию, предложенными авторами, в течение 100 эпох.
- Несколько проведенных экспериментов заключались в подборе оптимального разрешения изображений для обучения, а также наиболее подходящей модели-энкодера.
- Лучший результат дало применение ip-адаптера к специально дообученной на аниме модели

Бейзлайн-модель. Метрики.

По итогу получены следующие метрики:

- **FID: 145.1**
- **CLIP-T Score: 22.5**
- **CLIP-I Score: 98.7**



Бейзлайн-модель. Визуальная оценка.



Сервис. FastAPI



По согласованию с куратором **мы поменяли изначальное ТЗ, сместив акцент с дообучения на инференс разных моделей и чекпоинтов ip-адаптера.**

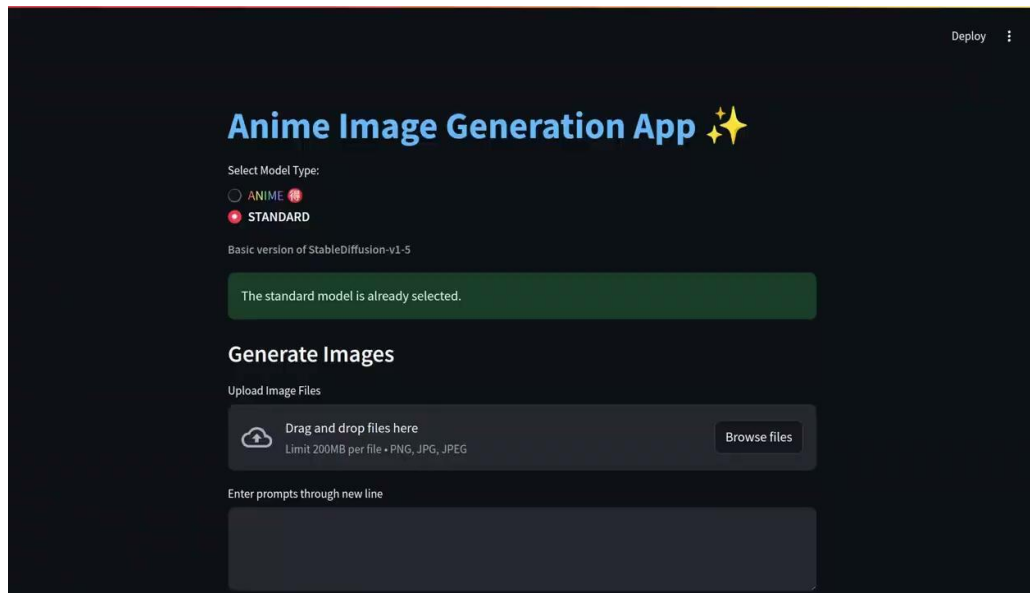
Реализовано:

- **4 POST метода**, которые позволяют генерировать изображения по загруженным картинкам, загрузить новый чекпоинт ip-адаптера, сменить действующий чекпоинт и тип модели.
- **3 GET метода** для получения списков доступных моделей, чекпоинтов и текущих параметров
- **2 DELETE метода** для удаления загруженных чекпоинтов

В приложение также добавлены логирование и контейнеризация

Сервис. Streamlit

Streamlit-приложение позволяет воспользоваться всеми возможностями реализованного API и является удобным инструментом для генерации изображений при помощи ip-адаптера, а также тестирования разных чекпоинтов.



Распределение работы



Гончаров Антон:

- EDA
- Подбор релевантных статей, выбор метрик и планирование экспериментов
- Streamlit-приложение
- Docker

Чайчук Михаил:

- Подбор датасета и его предобработка
- Реализация кода для обучения бейзлайна и подсчета метрик, запуск экспериментов
- FastAPI-сервис
- Тестирование всего проекта

Цели на будущее



- Имплементировать [PuLID](#) модель для лучшей генерации изображений. Возможно попробовать еще какие-либо архитектуры.
- При необходимости, дополнительная обработка датасета или поиск другого.
- Улучшить сервисы на основе новой полученной модели.
- Создание ТГ бота.

Спасибо за внимание!