# **Отчет по проекту: Имплементация подхода StyleGAN-NADA**

**1. Введение**

**1.1 Описание проекта**

Целью данного проекта является реализация подхода, описанного в статье "StyleGAN-NADA: CLIP-Guided Domain Adaptation of Image Generators". Этот подход позволяет адаптировать генератор изображений StyleGAN2 к новым доменам, используя текстовые подсказки и текстово-визуальную модель CLIP. Реализация ориентирована на создание гибкой системы для преобразования изображений в различные стили и домены без использования парных данных.

**1.2 Обоснование проекта**

Адаптация домена является важной задачей в области генеративных моделей. Предложенный метод StyleGAN-NADA позволяет:

* Устранять необходимость в размеченных данных.
* Использовать текстовые описания для управления доменами.
* Сохранять структуру и содержание исходных изображений.

Проект направлен на изучение и проверку эффективности подхода, а также на создание инструмента, который может быть применён в художественных, развлекательных и научных задачах.

**2. Обзор подхода**

**2.1 Основные компоненты**

* **StyleGAN2**: Генеративная модель изображений, используемая для генерации исходного и целевого доменов.
* **CLIP**: Модель, связывающая текстовые и визуальные представления, которая предоставляет текстовые направления для изменения домена.
* **Латентное пространство W+**: Пространство, в котором производится оптимизация для адаптации изображения к новому домену.

**2.2 Ключевые шаги**

1. **Определение значимых слоёв**: Использование глобального CLIP loss для выбора слоёв генератора, чувствительных к изменениям домена.
2. **Оптимизация латентного пространства**: Адаптация латентного кода W+ для генерации изображений в новом домене.
3. **Обучение слоёв генератора**: Дообучение значимых слоёв генератора с использованием направлений, заданных CLIP.

**2.3 Потери**

* **Глобальный CLIP Loss**: Используется для анализа изменений латентного кода.
* **Directional CLIP Loss**: Минимизирует разницу между исходным и целевым представлениями.

**3. Реализация**

**3.1 Среда разработки**

* **Язык программирования**: Python
* **Фреймворки**: PyTorch, Gradio
* **Среда**: Google Colab

**3.2 Структура проекта**

1. **Модели**:
   * Загрузка предобученной модели StyleGAN2.
   * Интеграция CLIP для анализа доменов.
2. **Функции**:
   * Оптимизация латентного кода W+.
   * Замораживание и дообучение слоёв генератора.
   * Генерация примеров
   * Генерация примеров из пресетов (заранее обученные генераторы)
   * Стилизация изображения из файла
3. **Интерфейс**:
   * Ноутбук для демонстрации работы с моделью.
   * Веб-приложения для взаимодействия с моделью.

**4. Результаты**

**4.1 Примеры преобразований**

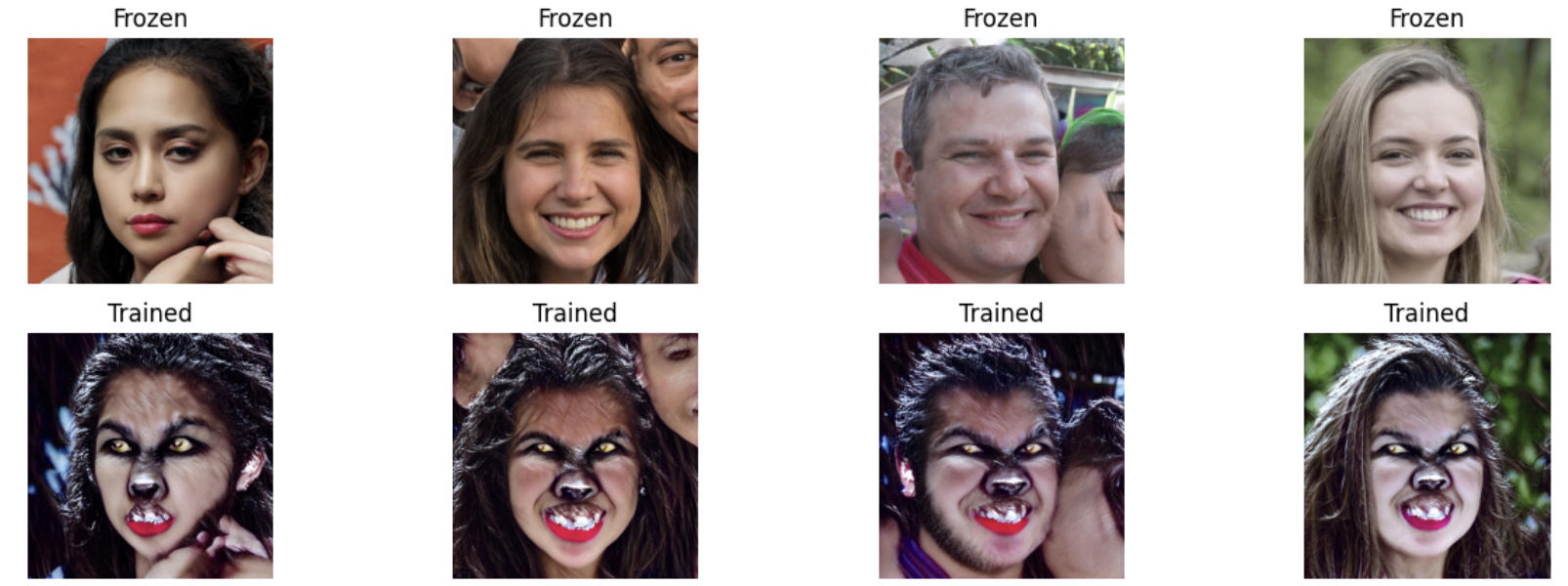
* **Исходный домен**: Photo
* **Целевой домен**: Sketch



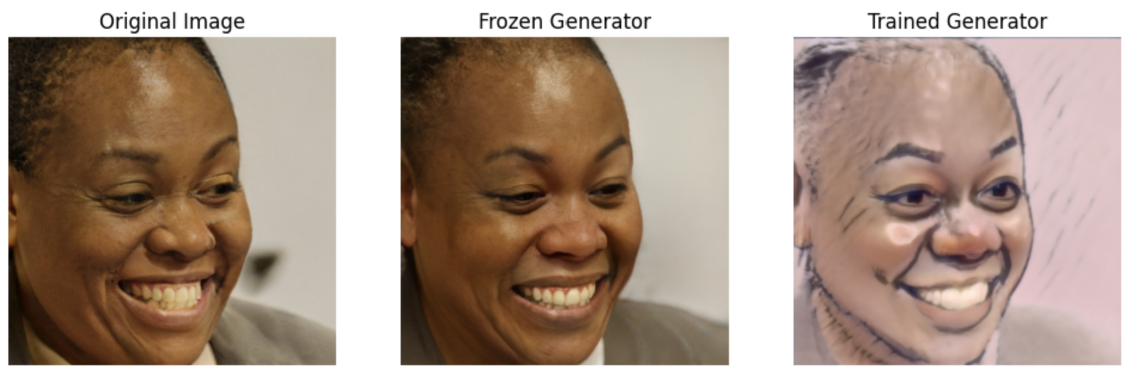
* **Исходный домен**: Photo
* **Целевой домен**: Modigliani Painting



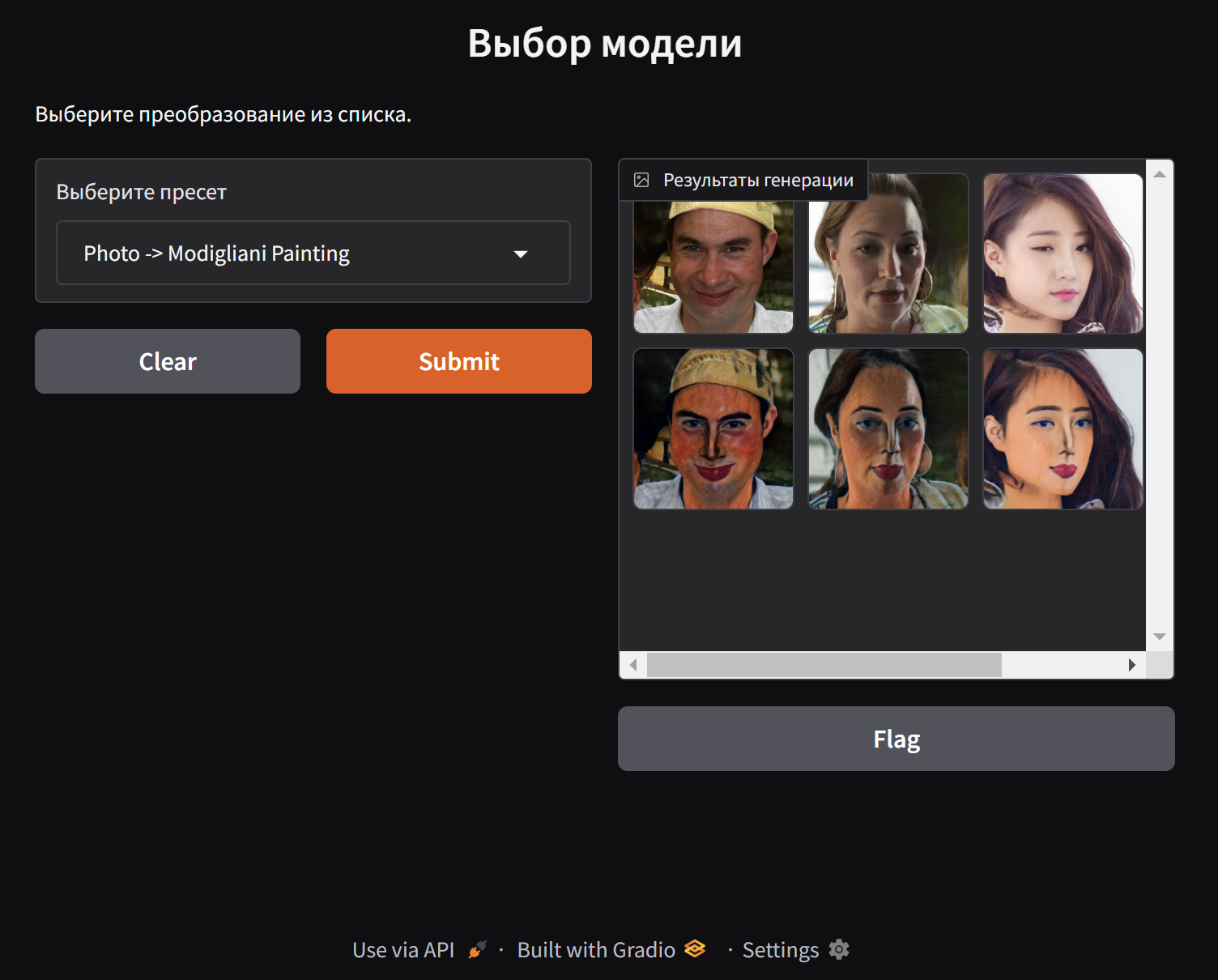
* **Исходный домен**: Human
* **Целевой домен**: Werewolf



**4.2 Пример преобразования изображения из файла (инверсия)**



**4.3 Пример работы веб-демо**

****

**5. Обсуждение**

**5.1 Что сделано:**

* Имплементация подхода адаптации домена по текстовому запросу
* Генерация примеров изображений со сдвигом домена
* Создание готовых пресетов для быстрой генераций изображений со сдвигом домена
* Веб-демо с готовыми пресетами
* Стилизация изображения из файла с помощью инверсии и обученного генератора

**5.1 Что не получилось:**

* Время полного цикла обучения порядка 5 минут, что выше времени, продемонстрированном в оригинальной имплементации.
* По некоторым запросам не удалось получить хорошего результата. Например, Human -> Clown:



* А по некоторым запросам, например Photo -> Anime получается не тот результат, который привычно видеть для Anime.
* В некоторых случаях стиль изображения очень трудно описать текстовым запросом.
* Инверсия изображения из файла в латентное представление вызывает дополнительные сложности для задачи стилизации.

**6. Заключение**

Проект продемонстрировал эффективность подхода StyleGAN-NADA для адаптации изображений к новым доменам. Реализация позволяет исследовать широкий спектр преобразований с минимальными данными. Однако вычислительные требования и зависимость от текста подсказки остаются основными вызовами, которые требуют дальнейшего исследования.

**7. Будущая работа**

* Улучшение качества адаптации путём доработки текстовых подсказок.
* Оптимизация вычислительных требований.
* Исследование альтернативных методов)

**8. Ссылки**

1. [StyleGAN-NADA: CLIP-Guided Domain Adaptation of Image Generators](https://arxiv.org/abs/2108.00946)
2. [StyleGAN2](https://github.com/NVlabs/stylegan2)
3. [CLIP](https://github.com/openai/CLIP)