# Литературный обзор статьи "StyleGAN-NADA: CLIP-Guided Domain Adaptation of Image Generators"

## **Введение**

Статья "StyleGAN-NADA: CLIP-Guided Domain Adaptation of Image Generators" представляет подход, позволяющий адаптировать генеративные модели (такие как StyleGAN) к новым доменам, используя исключительно текстовые подсказки, без необходимости иметь обучающие изображения для нового домена. Основная идея заключается в использовании модели CLIP для описания семантических различий между доменами и переноса этих различий в генеративную модель.

**Основные концепции и подход**

**1. StyleGAN2**

StyleGAN2 — одна из продвинутых моделей для генерации изображений. Она генерирует изображения через латентное пространство W+, предоставляя возможность точного управления стилем и структурой изображения. StyleGAN-NADA использует предобученные веса StyleGAN2 как основу для адаптации.

**2. CLIP**

CLIP — мощная модель от OpenAI, которая связывает текстовые и визуальные представления. CLIP позволяет задать домен для генерации изображений с помощью текстовых описаний. Вместо использования размеченных данных модель опирается на текстовые подсказки (prompts), такие как "пейзаж в стиле Ван Гога".

**3. Проблема доменной адаптации**

Адаптация StyleGAN к новому домену — сложная задача. Основная трудность заключается в сохранении исходной структуры изображения при изменении стилистических характеристик.

**4. Предлагаемое решение**

StyleGAN-NADA решает проблему адаптации домена с помощью:

* **CLIP Loss**:
  + **Глобальный CLIP Loss** используется для выбора слоёв генератора, наиболее чувствительных к изменению домена.
  + **Directional CLIP Loss** минимизирует расстояние между исходными и целевыми представлениями.
* **Адаптивное замораживание слоёв**:
  + Метод замораживает часть слоёв генератора, чтобы избежать нежелательных изменений структуры изображения.

**Методология**

1. **Определение значимых слоёв**
   * Исследователи используют CLIP для анализа изменений латентного кода W+ на уровне слоёв.
   * Слои, которые демонстрируют наибольшие изменения при оптимизации, становятся основными кандидатами для обучения.
2. **Оптимизация латентного кода**
   * Латентный код W+ оптимизируется с использованием текстовых подсказок, чтобы обеспечить лучшее соответствие целевому домену.
3. **Двухэтапный процесс обучения**
   * На первом этапе замораживаются веса всех слоёв генератора, и оптимизация проводится только в латентном пространстве W+.
   * На втором этапе снимается заморозка с выбранных слоёв, и они дообучаются под целевой домен (используется два генератора).
4. **Использование CLIP**
   * CLIP предоставляет направление в текстово-визуальном пространстве, определяющее изменение домена.
   * Это позволяет задавать такие стили, как "карандашный набросок" или "в стиле Модильяни", без необходимости парных данных.

**Результаты**

1. **Качество адаптации**
   * StyleGAN-NADA демонстрирует качественные изменения домена без существенной потери структуры исходных изображений.
2. **Гибкость**
   * Метод позволяет адаптировать изображения к различным стилям и доменам, включая художественные стили, жанры и даже фантазийные концепты ("человек превращается в оборотня").
3. **Эффективность**
   * Показано, что метод превосходит многие подходы, которые требуют парных данных или сложных процедур регуляризации.

**Связь с предыдущими работами**

1. **GAN-based Domain Adaptation**
   * Ранее доменная адаптация для GAN требовала большого количества размеченных данных (например, CycleGAN, StarGAN). StyleGAN-NADA устраняет эту необходимость, используя только текстовые подсказки.
2. **CLIP-Guided Approaches**
   * До StyleGAN-NADA CLIP уже применялся для задач редактирования изображений, например, в StyleCLIP. Однако StyleGAN-NADA делает шаг вперёд, интегрируя CLIP для полной адаптации домена.
3. **Работы с латентным пространством**
   * Методы оптимизации латентного пространства, такие как e4e и pSp, вдохновили авторов на использование оптимизации W+ для изменения доменов.

**Преимущества и ограничения**

**Преимущества:**

* **Отсутствие парных данных:** Метод не требует исходных и целевых изображений для обучения.
* **Гибкость:** Поддержка широкого диапазона текстовых описаний.
* **Эффективность:** Сохранение структуры изображения при изменении домена.

**Ограничения:**

* **Ограничение CLIP:** Текстовые подсказки должны быть достаточно специфичными для достижения желаемого результата.
* **Высокая вычислительная сложность:** Процесс обучения требует значительных ресурсов, особенно при обучении слоёв генератора.

**Заключение**

StyleGAN-NADA предлагает интересный подход к адаптации доменов изображений, сочетая возможности StyleGAN2 и CLIP. Это делает возможным создание новых стилей и преобразований изображений без необходимости в парных данных. Несмотря на вычислительные ограничения, метод открывает широкие перспективы для художественных и практических приложений.