

Генеративное компьютерное зрение

Саратовский государственный университет им. Н. Г.

Чернышевского

Кафедра математической кибернетики и компьютерных наук

Студент 211 группы Пицик Х.Н.

Научный руководитель: доцент Филиппов Б.А.

Постановка задачи

Цель работы — проведение анализа современных подходов к решению задачи VITON (Virtual Try-On, виртуальная примерочная).

Выполненные задачи:

- Теоретический анализ каждого из рассматриваемых подходов, явное описание функций потерь и требований к набору данных;
- Составление собственного набора данных с помощью ручной разметки и вспомогательных инструментов (например, OpenPose);
- Эмпирический анализ полученных результатов, введение метрик качества для более объективной оценки качества, подведение итогов.

Используемые технологии

Основной язык программирования — **Python**

Ускорение операций по обработке данных — **NumPy**

Обработка изображений — **PIL, OpenCV**

Визуализация статистических данных — **Matplotlib**

Применение алгоритмов классического машинного обучения — **scikit-image**

Построение архитектур глубокого обучения — **PyTorch**

Оптимизация взаимодействия графического процессора с тензорами данных — **CUDA**

Рассматриваемые решения

PASTA-GAN++ — в качестве основы используются состязательные порождающие сети, доступно обучение без парных изображений, разделение одежды на нормализованные патчи для удачного наложения в результирующую позу.

LaDI-VITON — первая модель на основе латентных диффузионных моделей для виртуальной примерки, эффективное поддержание деталей и текстур одежды, отображение визуальных характеристик в пространство CLIP.

PromptDresser — использование латентных диффузионных моделей, генерация масок одежды на основе сгенерированных с помощью LMM текстовых описаний, что позволяет более гибко настраивать отображение.

Анализ результатов

Технические характеристики

В силу разности требований к вычислительным мощностям рассматриваемых решений, был проведён запуск на нескольких устройствах. Технические характеристики:

	PASTA-GAN++	LaDI-VITON	PromptDresser
GPU	GeForce GTX 1060 3GB	GeForce RTX 2060 8GB	GeForce GTX 1070 8GB
CPU	Intel Core i5-6400	AMD Ryzen 7 2700 AM4	Intel Core i5-8400
RAM	16GB DDR4	16GB DDR4	32GB DDR4

Эмпирический анализ



Figure 1: PASTA-GAN++,
320x512



Figure 2: LaDI-VITON,
384x512



Figure 3: PromptDresser,
768x1024

Метрики качества

FID

FID (Fréchet Inception Distance) оценивает сходство между изображениями путем оценивания их статистических признаков в признаковом пространстве:

$$FID(\mu_r, \Sigma_r, \mu_g, \Sigma_g) = \|\mu_r - \mu_g\|_2^2 + \text{Tr}\left(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{\frac{1}{2}}\right)$$

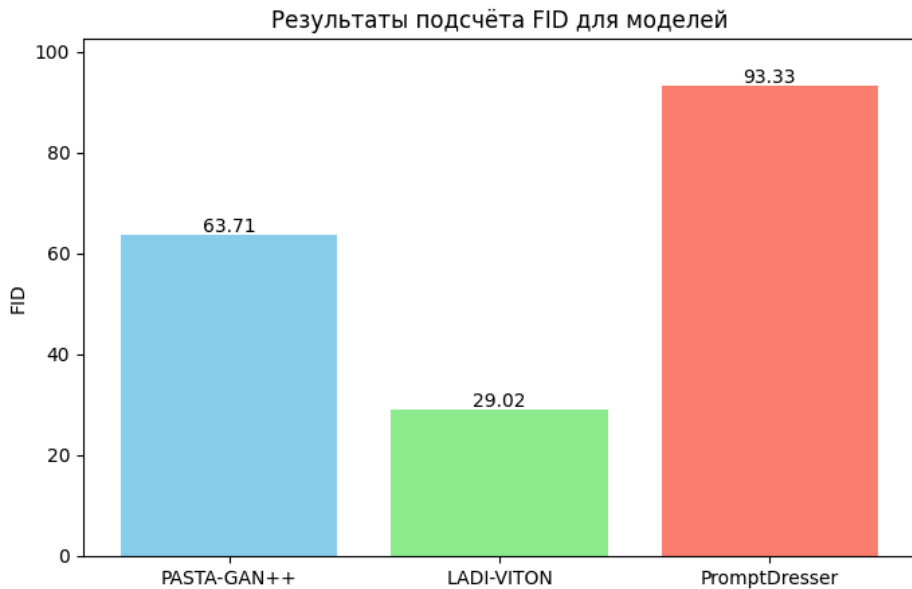
LPIPS

LPIPS (Learned Perceptual Image Patch Similarity). Для каждой пары соответствующих карт признаков вычисляется евклидово расстояние, взвешенное специальными обучаемыми коэффициентами для каждого слоя, а затем они суммируются:

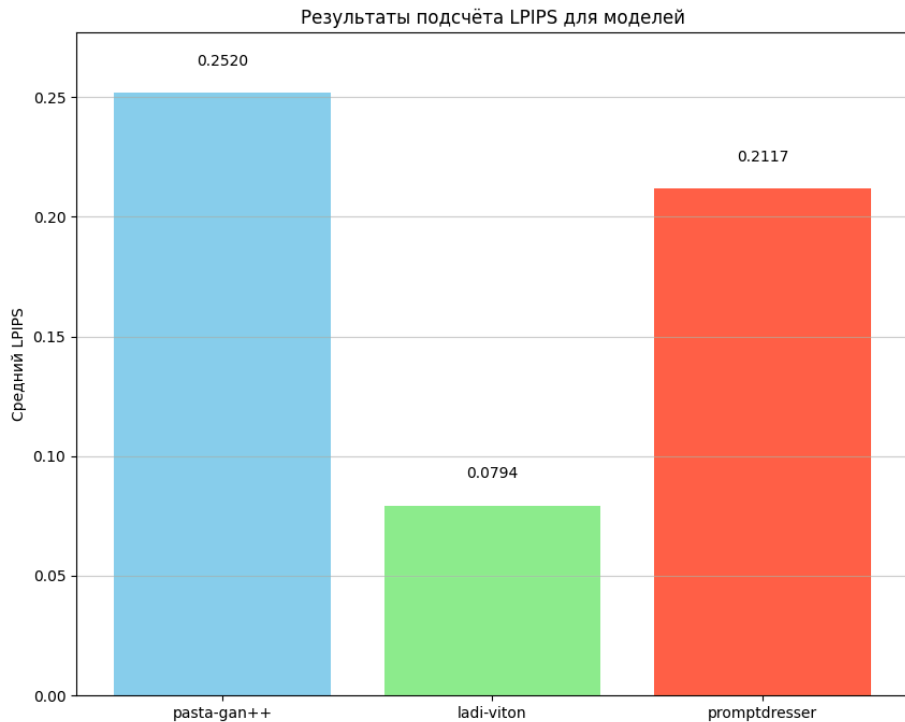
$$d_l(x, y) = \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (\hat{F}(x)_{hw} - \hat{F}(y)_{hw})\|_2^2$$

$$LPIPS(x, y) = \sum_{l \in L} d_l(x, y),$$

Подсчёт метрик: FID



Подсчёт метрик: LPIPS



Список использованных источников

1. Николенко С.И. Глубокое обучение. Москва: Издательство <<Диалектика>>, 2018.
2. PASTA-GAN++: A Versatile Framework for High-Resolution Unpaired Virtual Try-on [Electronic resource]. URL: <https://arxiv.org/abs/2207.13475>.
3. LaDI-VTON: Latent Diffusion Textual-Inversion Enhanced Virtual Try-On [Electronic resource]. URL: <https://arxiv.org/abs/2305.13501>.
4. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields [Electronic resource]. URL: <https://arxiv.org/abs/1812.08008>.
5. Reviewing FID and SID Metrics on Generative Adversarial Networks [Electronic resource]. URL: <https://arxiv.org/abs/2402.03654>.
6. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric [Electronic resource]. URL: <https://arxiv.org/abs/1801.03924>.
7. PromptDresser: Improving the Quality and Controllability of Virtual Try-On via Generative Textual Prompt and Prompt-aware Mask [Electronic resource]. URL: <https://arxiv.org/pdf/2412.16978>.

Спасибо за внимание.