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.

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

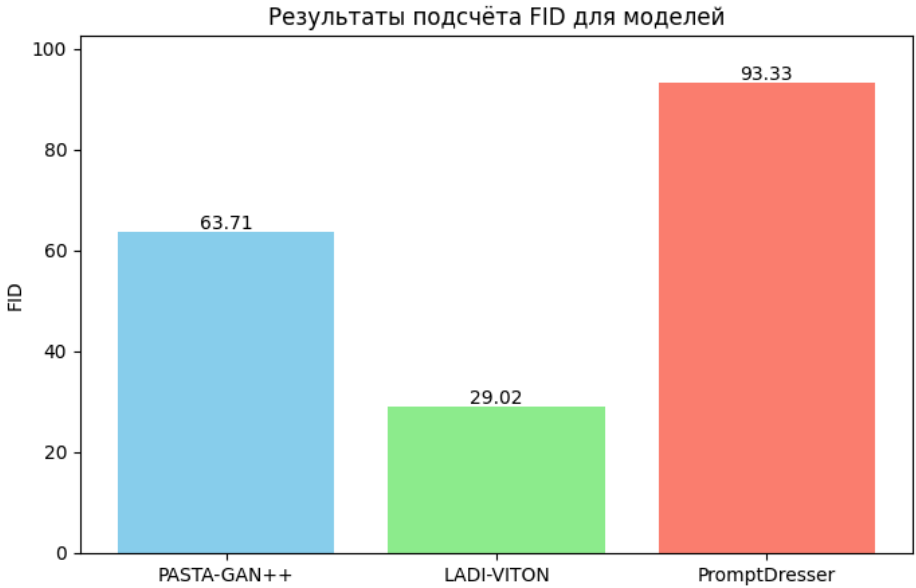
LPIPS

`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:

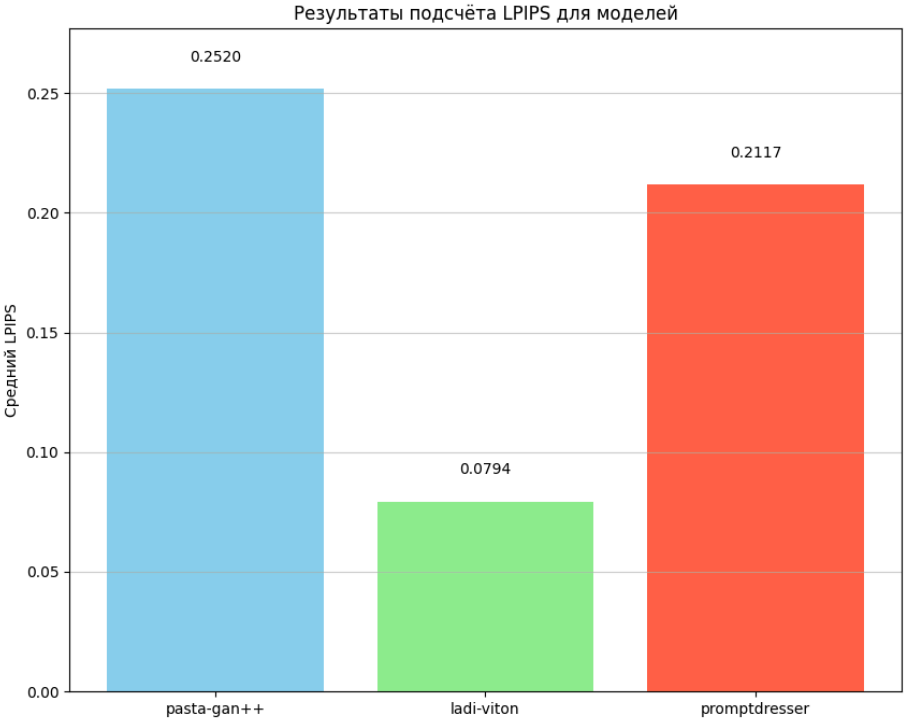
$$d_l(x, y) = \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot \left(\hat{F}(x)_{hw} - \hat{F}(y)_{hw}\right)\|_2^2$$

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

Подсчёт метрик: FID



Подсчёт метрик: LPIPS



Спасибо за внимание.

Список использованных источников

1. Понкин И.В. et al. Компьютерное зрение: концепт, функционально-целевое назначение, структура, регуляторика // International Journal of Open Information Technologies. 2024. Vol. 12. Pp. 57–67.
2. Маркин Е.И., Зупарова В.В., Мартышкин А.И. Исследование возможности применения нейронных сетей для восстановления изображения лица в системах распознавания // Труды Института системного программирования РАН. 2022. Vol. 34. Pp. 117–126.
3. Николенко С.И. Глубокое обучение. Москва: Издательство <<Диалектика>>, 2018.
4. Generative Adversarial Networks [Electronic resource]. URL: https://arxiv.org/abs/1406.2661.
5. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks [Electronic resource]. URL: https://arxiv.org/abs/1703.10593.
6. PASTA-GAN++: A Versatile Framework for High-Resolution Unpaired Virtual Try-on [Electronic resource]. URL: https://arxiv.org/abs/2207.13475.

Список использованных источников (ii)

7. Semantic Image Synthesis with Spatially-Adaptive Normalization [Electronic resource]. URL: https://arxiv.org/abs/1903.07291.
8. Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models // Advances in Neural Information Processing Systems. 2020. Vol. 33. Pp. 6840–6851.
9. Sohl-Dickstein J. et al. Deep Unsupervised Learning using Nonequilibrium Thermodynamics // Proceedings of the 32nd International Conference on Machine Learning (ICML). 2015. Vol. 37. Pp. 2256–2265.
10. LaDI-VTON: Latent Diffusion Textual-Inversion Enhanced Virtual Try-On [Electronic resource]. URL: https://arxiv.org/abs/2305.13501.
11. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields [Electronic resource]. URL: https://arxiv.org/abs/1812.08008.
12. NumPy [Electronic resource]. URL: https://numpy.org/.
13. OpenCV - Open Computer Vision Library [Electronic resource]. URL: https://opencv.org/.
14. PyTorch [Electronic resource]. URL: https://pytorch.org/.

Список использованных источников (iii)

15. CUDA Toolkit - Free Tools and Training | NVIDIA Developer [Electronic resource]. URL: https://developer.nvidia.com/cuda-toolkit.
16. Reviewing FID and SID Metrics on Generative Adversarial Networks [Electronic resource]. URL: https://arxiv.org/abs/2402.03654.
17. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric [Electronic resource]. URL: https://arxiv.org/abs/1801.03924.
18. 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.