

Body-Lightning ID Diffusion

Даниил Иванович Казачков
Научный руководитель: А. В. Филатов

Кафедра интеллектуальных систем ФПМИ МФТИ
Специализация: Интеллектуальный анализ данных

2025

О модели. Вход и выход. Метрика

вход: Ref Image + TextPrompt + Image Prompt (optional)

выход: Output Image

В качестве метрики буду использовать аналогичную той, что использовалась в Stable Diffusion модели. Это естественная функция для проверки того, насколько хорошо диффузионная модель "расшумляет" данные.

$$L = \mathbb{E}_{x_0, \epsilon, c_t, c_i, t} \left[\|\epsilon - \epsilon_\theta(x_t, c_t, c_i, t)\|^2 \right], \quad (1)$$

где

x_0 незашумленные данные

ϵ случайный шум из распределения $\mathcal{N}(\mathbf{0}, \mathbf{I})$

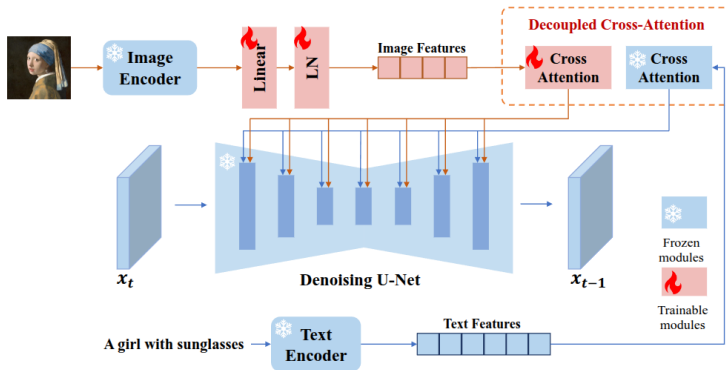
c_t, c_i дополнительное условие на текст (text) и картинку (image)

t шаг (time step) диффузионного процесса

Постановка задачи

Цель: проверить гипотезу о том, что подходы в работах PuLID, InstantID, IP-Adapter распространяются не только на аватары лиц, но и на ростовые аватары.

Задача: улучшить качество генерируемых аватаров, путем обуславливания картинки всем телом человека, причем человек может находиться в любой позе (сидеть, лежать и т.д., т.е. не обязательно стоять строго прямо).



Источники:

- ▶ IP-Adapter: Text Compatible Image Prompt Adapter for Text-to-Image Diffusion Models
- ▶ InstantID : Zero-shot Identity-Preserving Generation in Seconds
- ▶ PuLID: Pure and Lightning ID Customization via Contrastive Alignment

,