

# Body-Lightning ID Diffusion

Даниил Иванович Казачков  
Научный руководитель: А. В. Филатов

Кафедра интеллектуальных систем ФПМИ МФТИ  
Специализация: Интеллектуальный анализ данных

2025

# BoLID: Body-Lightning ID Diffusion

**вход:** Ref Image + TextPrompt + Image Prompt (optional)  
**выход:** Output Image

$$L = \mathcal{L}_{\text{rec}} + \mathcal{L}_{\text{id}} + \mathcal{L}_{\text{CLIP}}, \quad (1)$$

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

$$\mathcal{L}_{\text{id}} = \sum_{k=1}^{12} \|p_{\text{gen},k} - p_{\text{ref},k}\|^2, \quad (3)$$

$$\mathcal{L}_{\text{CLIP}} = 1 - \cos(\text{CLIP}(x_{\text{gen}}), \text{CLIP}(\tau)), \quad (4)$$

где

$x_0$  исходная картинка

промпты

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

$t$  шаг диффузионного  
процесса

$c_t, c_i$  текстовый и визуальный

$p_{\text{gen},k}, p_{\text{ref},k}$  — координаты  $k$ -й точки

## 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).

Качество генерации будем проверять следующими метриками: FID в пространстве признаков InceptionV3, LPIPS, CLIPScore, где LPIPS оценивает, насколько два изображения похожи с точки зрения восприятия.

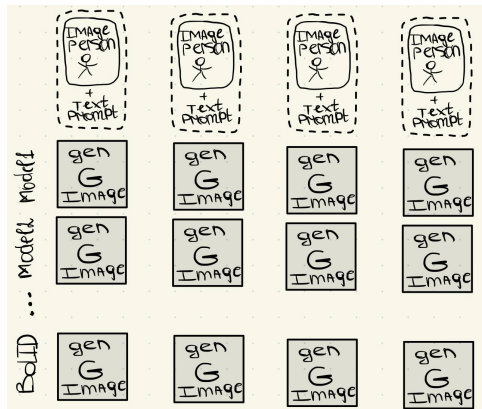
$$FID = ||\mu_p - \mu_q||^2 + Tr(\mathbf{\Sigma}_p + \mathbf{\Sigma}_q - 2(\mathbf{\Sigma}_p \mathbf{\Sigma}_q)^{1/2}) \quad (5)$$

$$ID_{sim} = \cos(\mathbf{e}_{gen}, \mathbf{e}_{ref}),$$

$$CLIP_{score} = \cos(CLIP(x_{gen}), CLIP(\tau)).$$

# Вычислительный эксперимент

На данном этапе подготовлен код обучения модели, с инференсом которой возникли проблемы.



(a) Визуальное сравнение

	FID	LPIPS	ID similarity	POSE Accuracy	CLIP Score
Model1	—	—	—	—	—
Model2	—	—	—	—	—
BoLiD(ours)	—	—	—	—	—

(b) Численное сравнение

## 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

,