北京大学 PEKING UNIVERSITY

HyperLoRA: Parameter-Efficient Adaptive Generation for Portrait Synthesis

Mengtian Li, JinShu Chen, Wanquan Feng, Bingchuan Li, Fei Dai,
Songtao Zhao, Qian He
Intelligent Creation, Bytedance

CVPR 2025 (Highlight)

Presenter: Chenyu Niu
2025.09.21

Background

Personalized portrait synthesis

Reference Set

Wizard hat

Pink wavy hair

Chinese clothing

Snow

Beach

Cyberpunk

北京大学 PEKING UNIVERSITY

Key Challenges

- **Fidelity**
  - Extract useful ID embeddings to guide generation process
  - Coarse-grained improvements is insufficient for ID-preserving generation
- **Editability**
  - Embedding ID disrupts the behavior of the original model
  - Hard to edit ID attributes with text prompt
- **Speed**

Background

Personalized portrait synthesis

Reference Set

Existing Methods

Tuning-Based

- LoRA
- DreamBooth

Tuning-Free

- IP-Adapter

Wizard hat

Pink wavy hair

Chinese clothing

Snow

Beach

Cyberpunk

北京大学 PEKING UNIVERSITY

Low-Rank Adaptation (LoRA)

Given

- a pre-trained autoregressive language model $P_\Phi(y|x)$ parametrized by Φ
- a downstream task to fine-tune on: $Z = \{(x_i, y_i)\}_{i=1,\dots,N}$

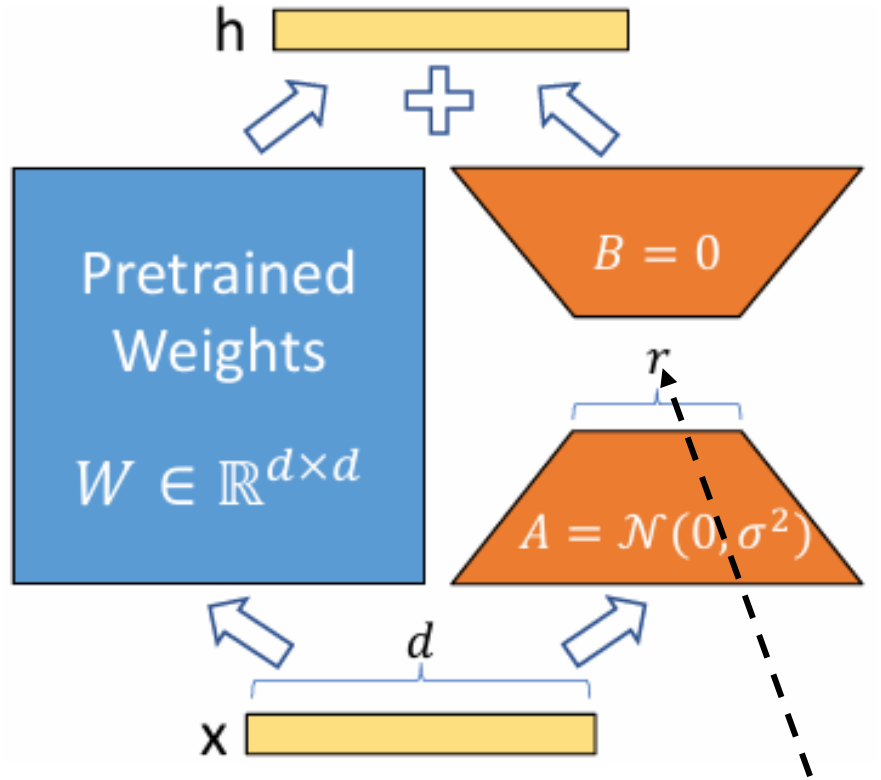
Full fine-tuning:

$$\max_{\Phi} \sum_{(x,y)\in\mathcal{Z}} \sum_{t=1}^{|y|} \log\left(P_\Phi(y_t|x, y_{<t})\right) \qquad |\Delta\Phi| = |\Phi_0|$$

Reparameterized:

$$\max_{\Theta} \sum_{(x,y)\in\mathcal{Z}} \sum_{t=1}^{|y|} \log\left(p_{\Phi_0+\Delta\Phi(\Theta)}(y_t|x, y_{<t})\right) \qquad |\Theta| \ll |\Phi_0|$$

Low-Rank Adaptation (LoRA)



Intrinsic rank during adaptation

Pre-trained weight matrix: $W_0 \in \mathbb{R}^{d\times k}$

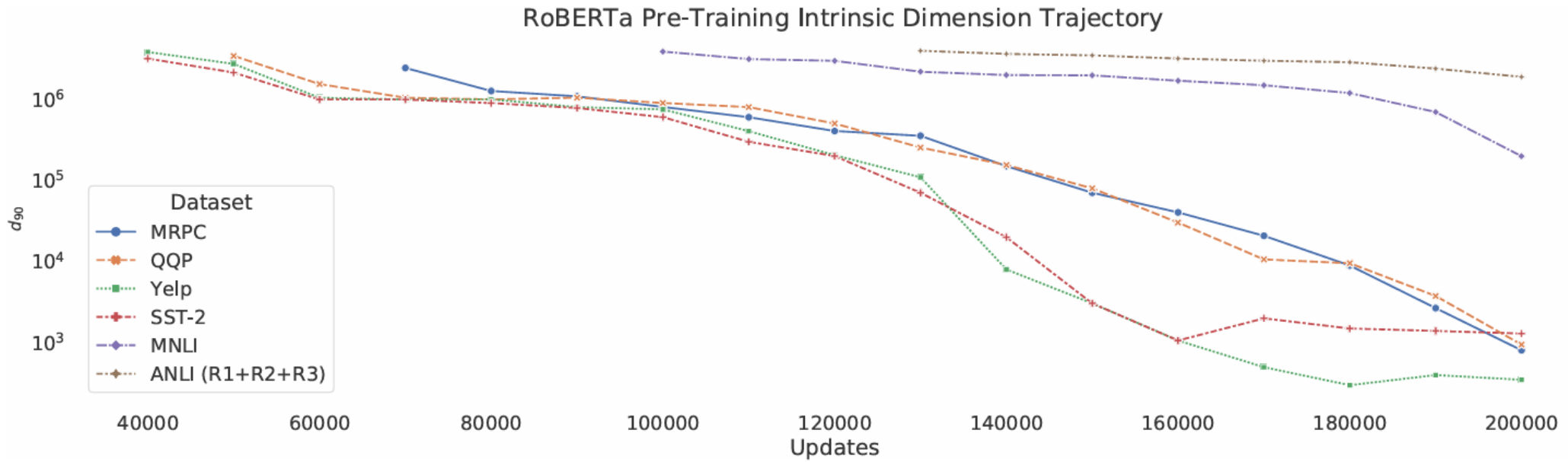
Low-rank-parametrized update matrices:

$W_0 + \Delta W = W_0 + BA$, where

$B \in \mathbb{R}^{d\times r}, A \in \mathbb{R}^{r\times k}, r \ll \min(d, k)$

北京大学 PEKING UNIVERSITY

Low-Rank Adaptation (LoRA)

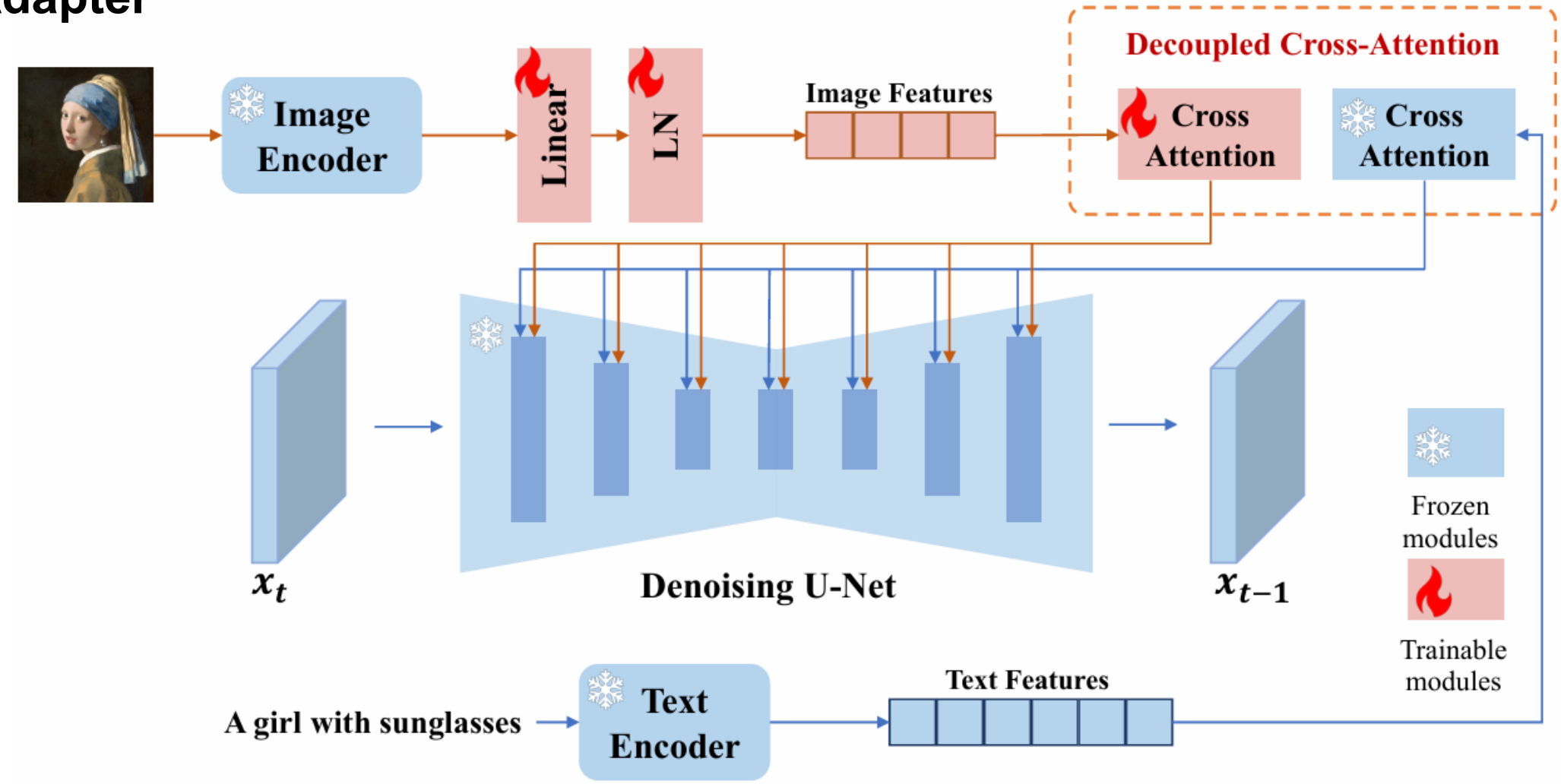


Learned over-parametrized models in fact reside on a low intrinsic dimension

Hypothesize that the change in weights during model adaptation also has a low “intrinsic rank”

Intrinsic Dimensionality Explains the Effectiveness of Language Model Fine-Tuning, arXiv:2012.13255

IP-Adapter



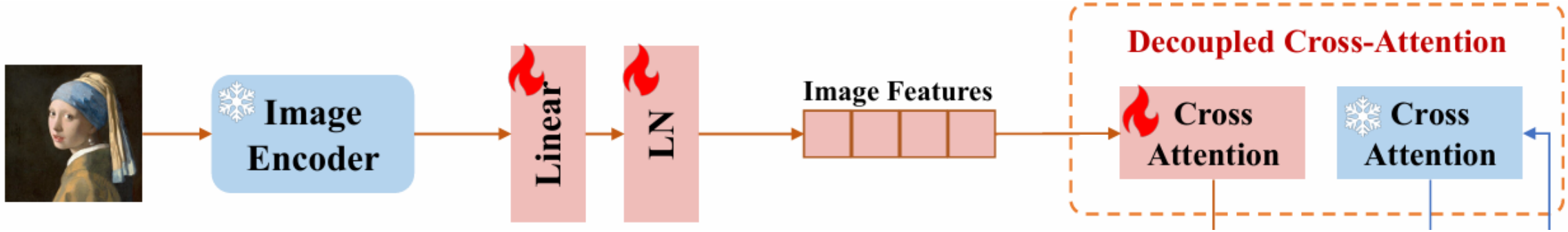
IP-Adapter: Text Compatible Image Prompt Adapter for Text-to-Image Diffusion Models, arXiv: 2308.06721

北京大学 PEKING UNIVERSITY

IP-Adapter

Image encoder

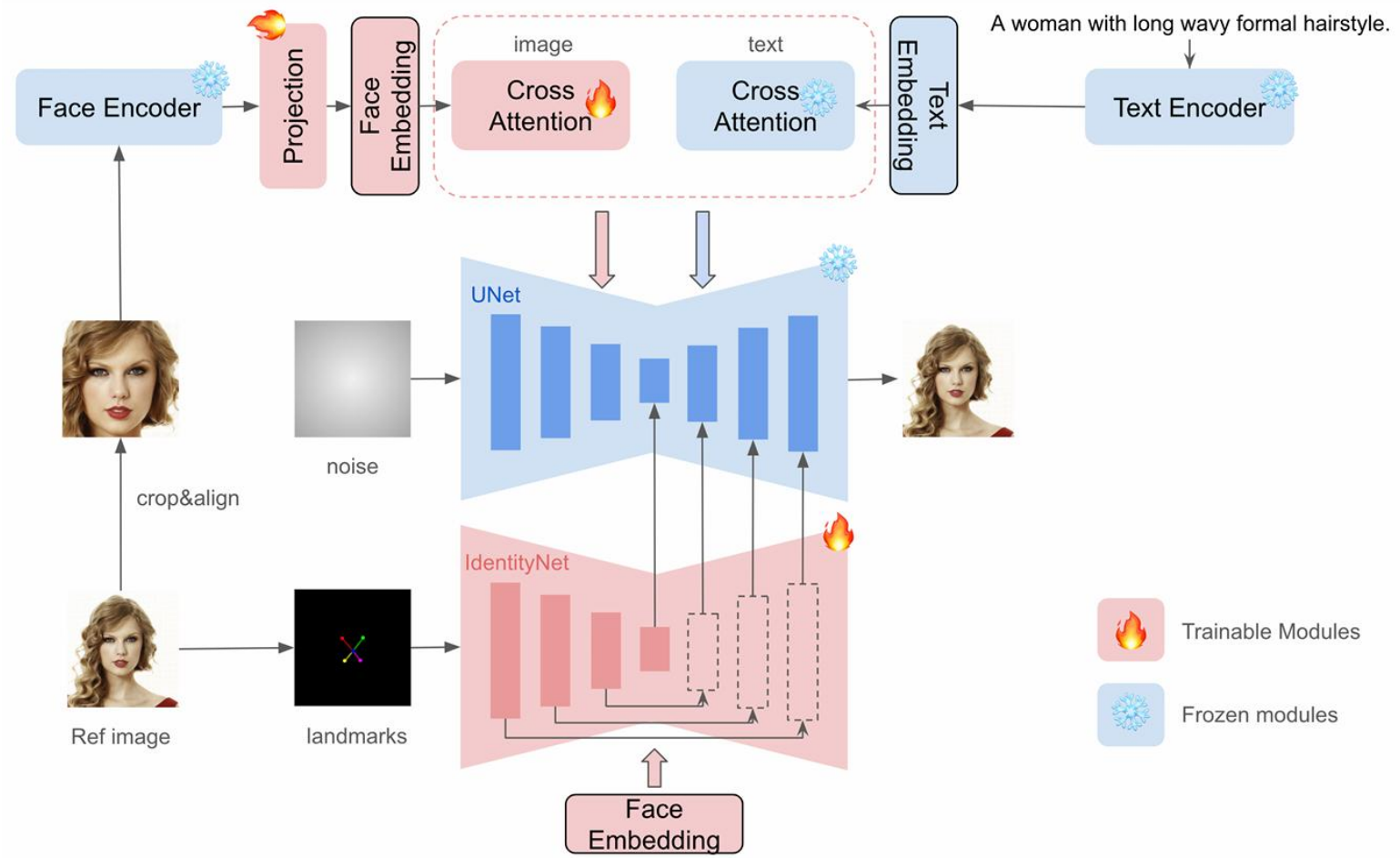
- Frozen CLIP image encoder model
- Trainable projection network



Decoupled Cross-Attention

$$\mathbf{Z}^{new} = \text{Softmax}(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}})\mathbf{V} + \text{Softmax}(\frac{\mathbf{Q}(\mathbf{K}')^\top}{\sqrt{d}})\mathbf{V}'$$

InstantID

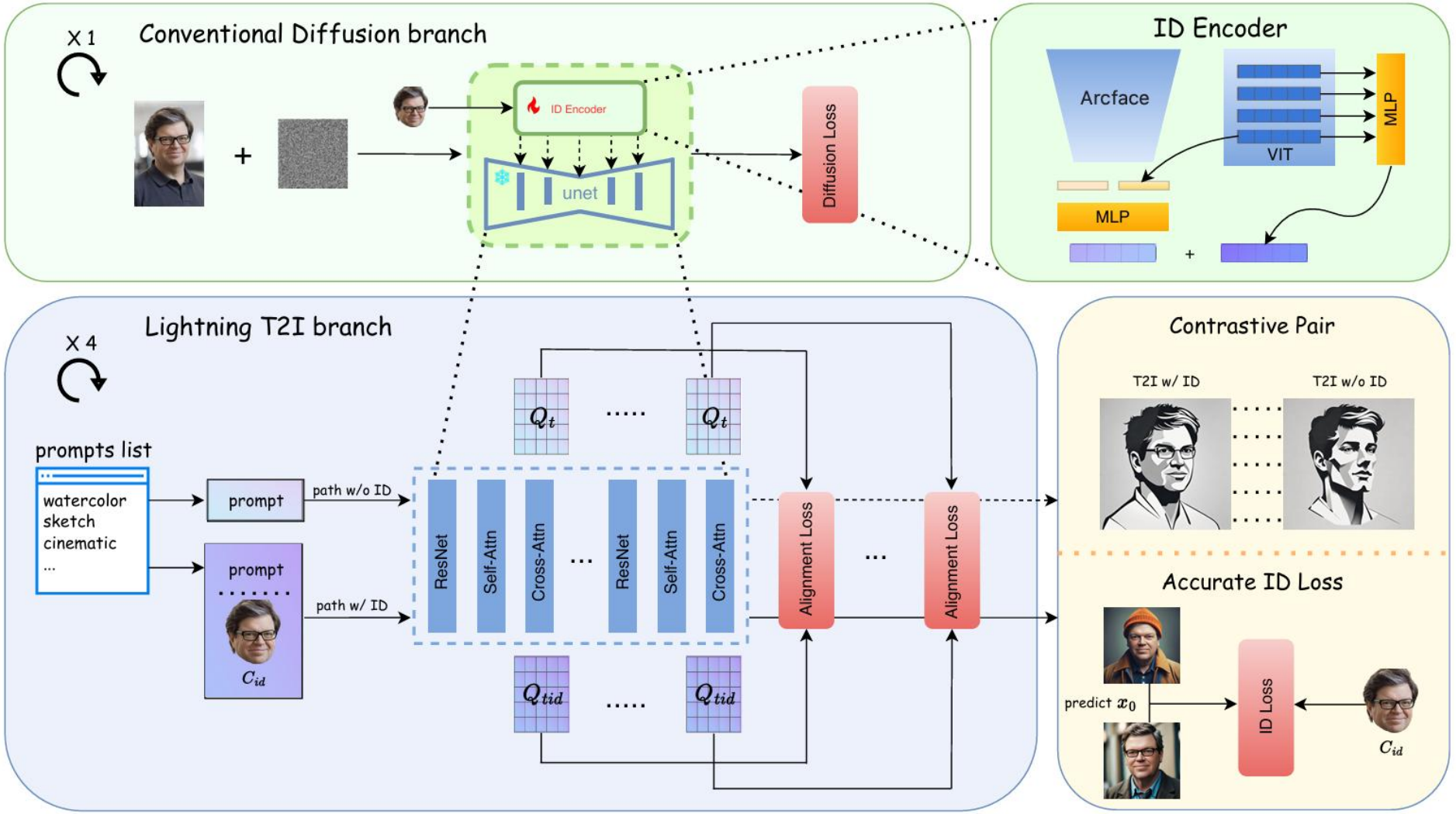


- Adopt a face encoder instead of CLIP encoder
- IdentityNet for enhancing identity feature injection

InstantID: Zero-shotIdentity-Preserving Generation in Seconds, arXiv: 2401.07519

Background

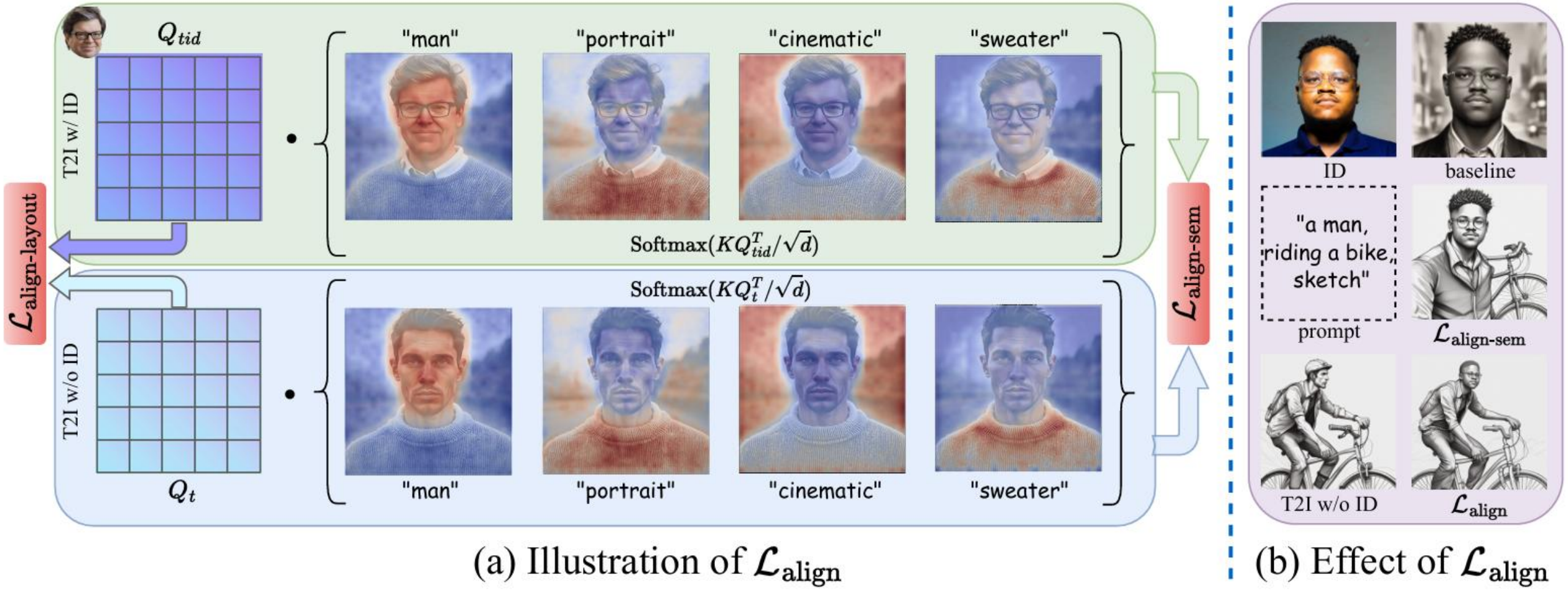
PuLID



PuLID: Pure and Lightning ID Customization via Contrastive Alignment, arXiv: 2404.16022

Background

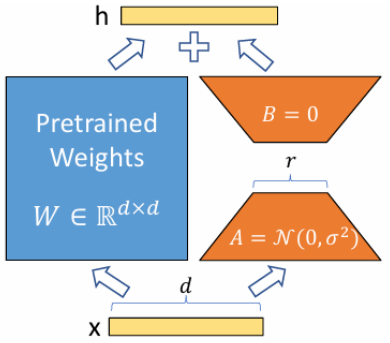
PuLID

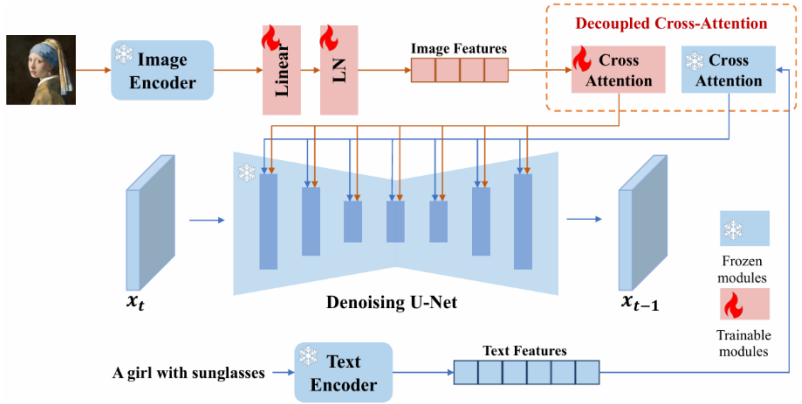


(a) Illustration of $\mathcal{L}_{\text{align}}$

(b) Effect of $\mathcal{L}_{\text{align}}$

Existing Disadvantages





Key challenge in personalized portrait synthesis:

- **Fidelity**
- **Editability**
- **Speed**

LoRA

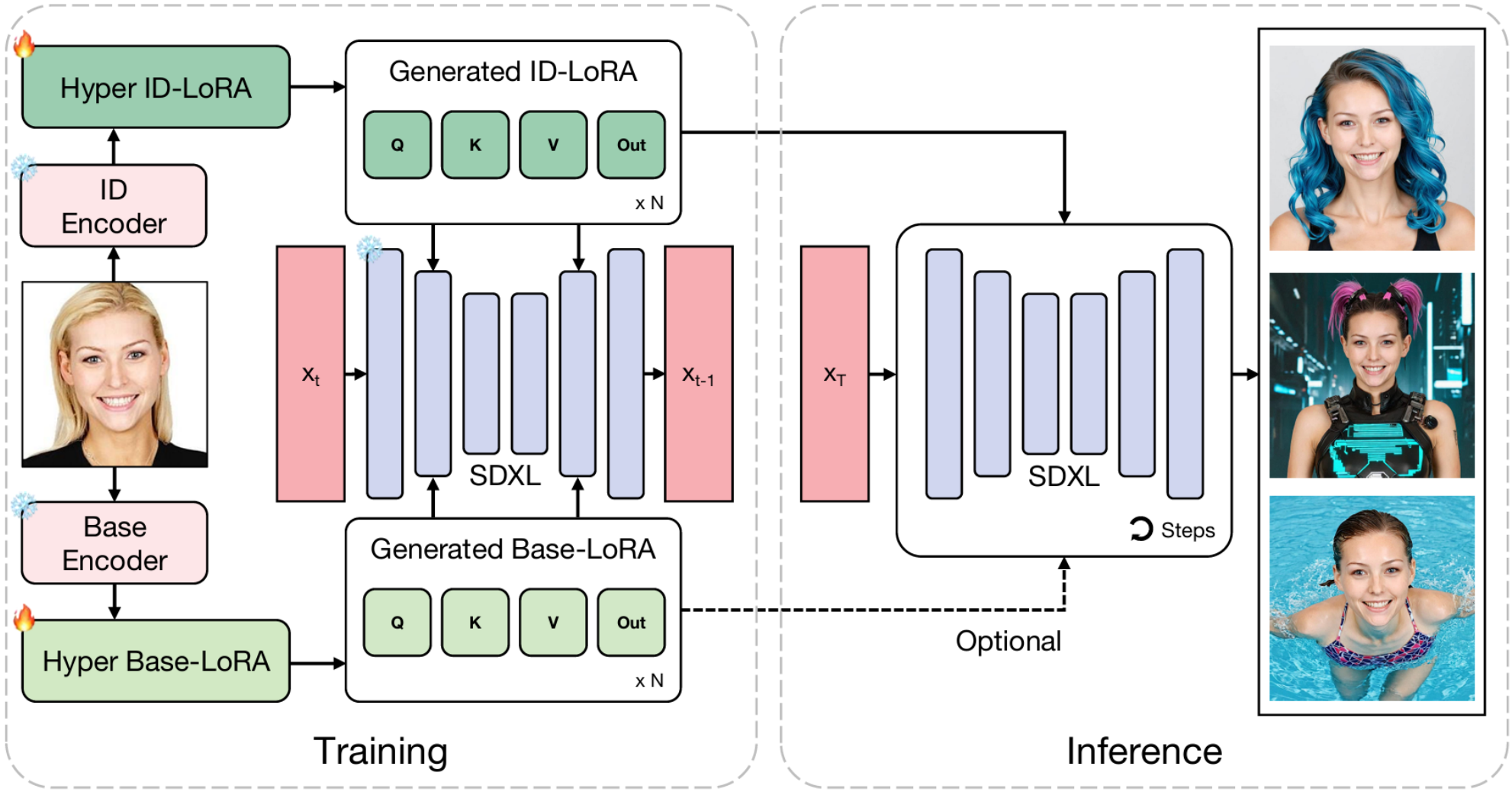
- High computing cost
- Unstable training
- Lengthy training process

IP-Adapter

- Quality degradation
- Unnatural facial features
- Fail to recover fine-grained details

北京大学 PEKING UNIVERSITY

Overview of HyperLoRA



北京大学 PEKING UNIVERSITY

Decouple LoRA

ID Encoder

Hyper ID-LoRA

Specific identity information

Base Encoder

Hyper Base-LoRA

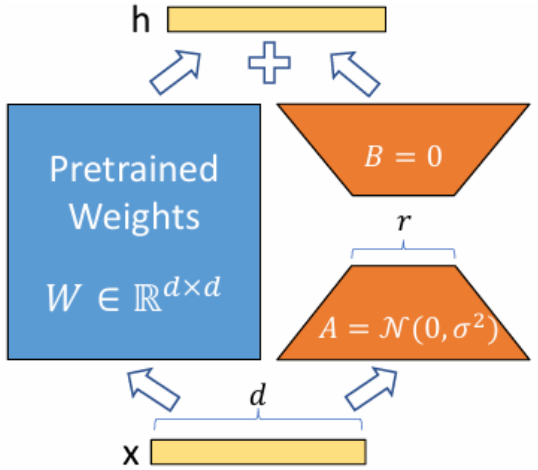
Irrelevant features
(clothes, background…)

北京大学 PEKING UNIVERSITY

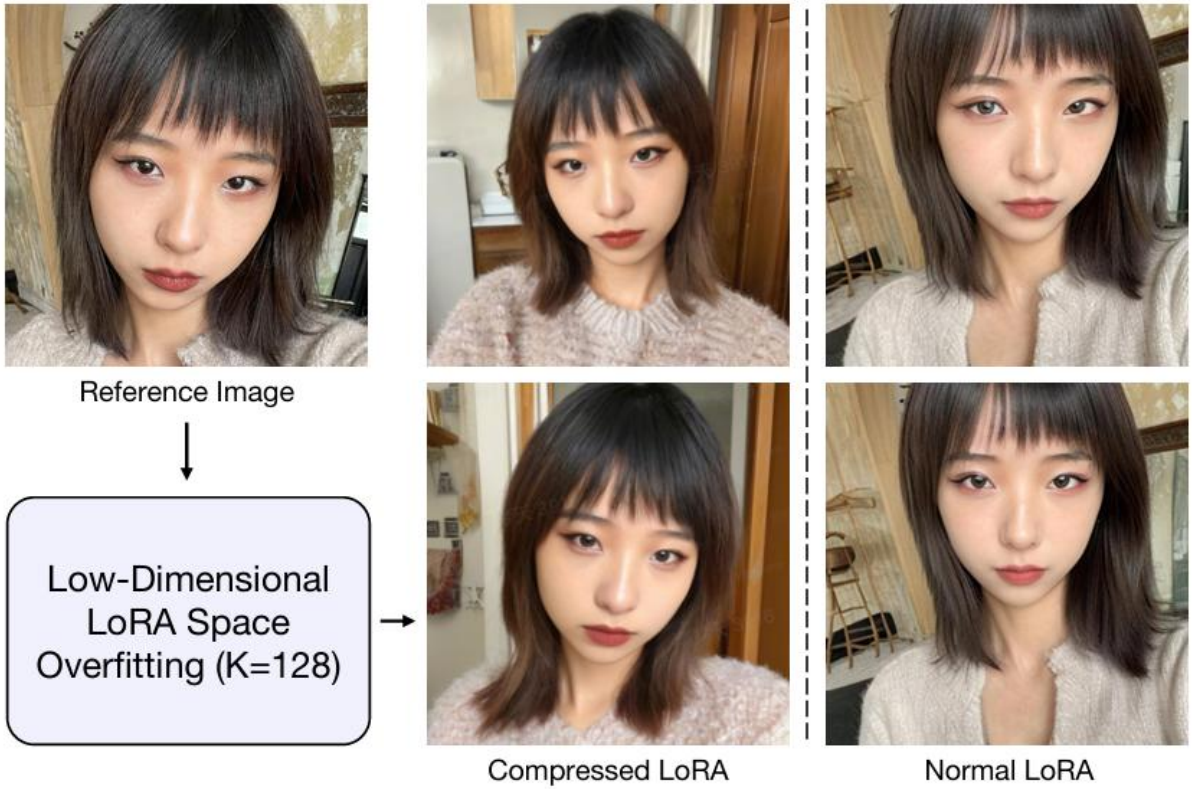
Decouple LoRA

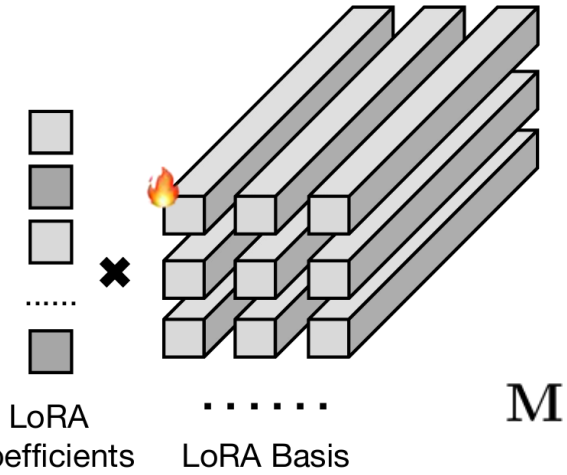
北京大学 PEKING UNIVERSITY

Low Dimensional Linear LoRA Space



total parameters count
$N \times 8 \times d \times r$

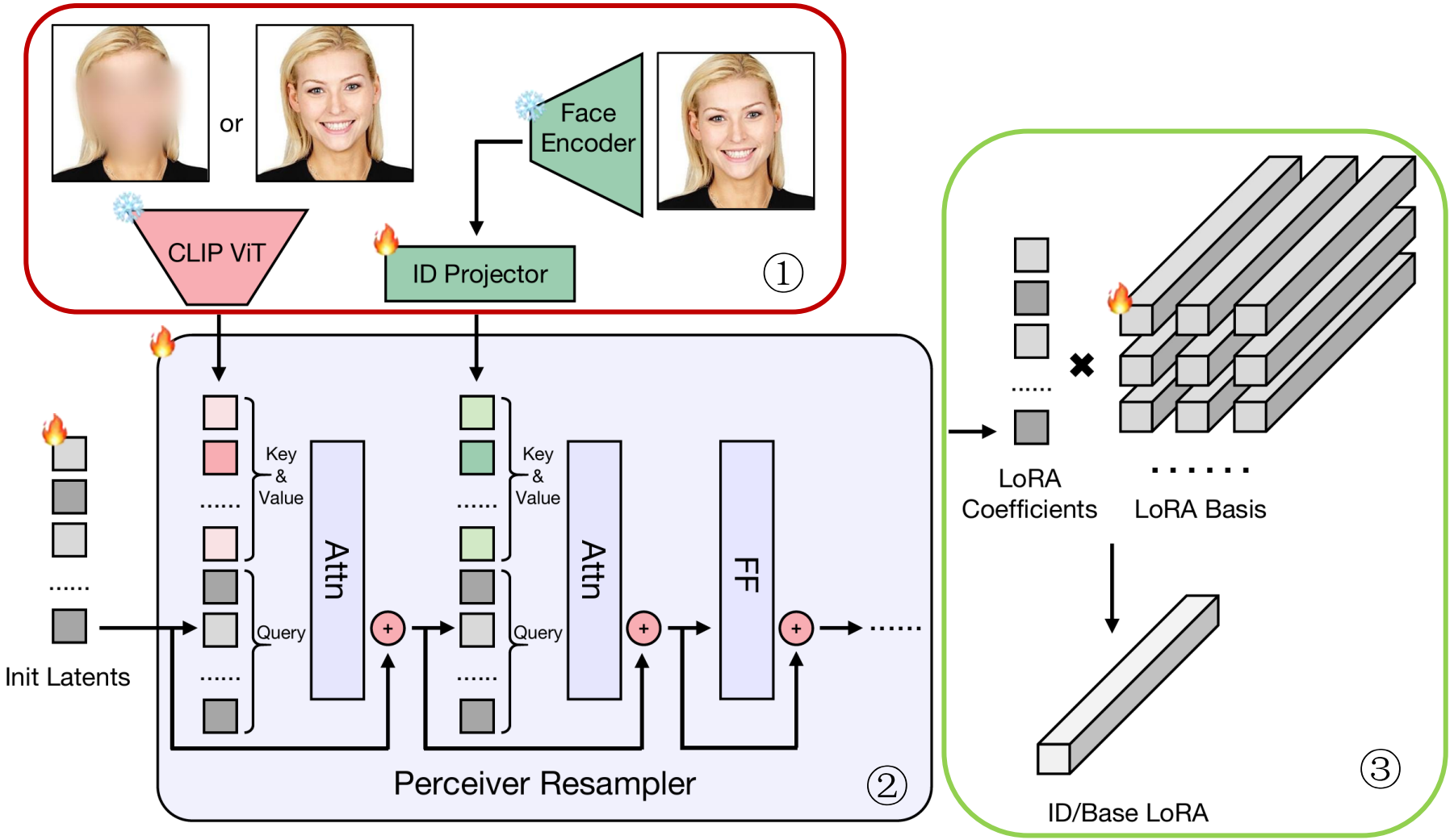




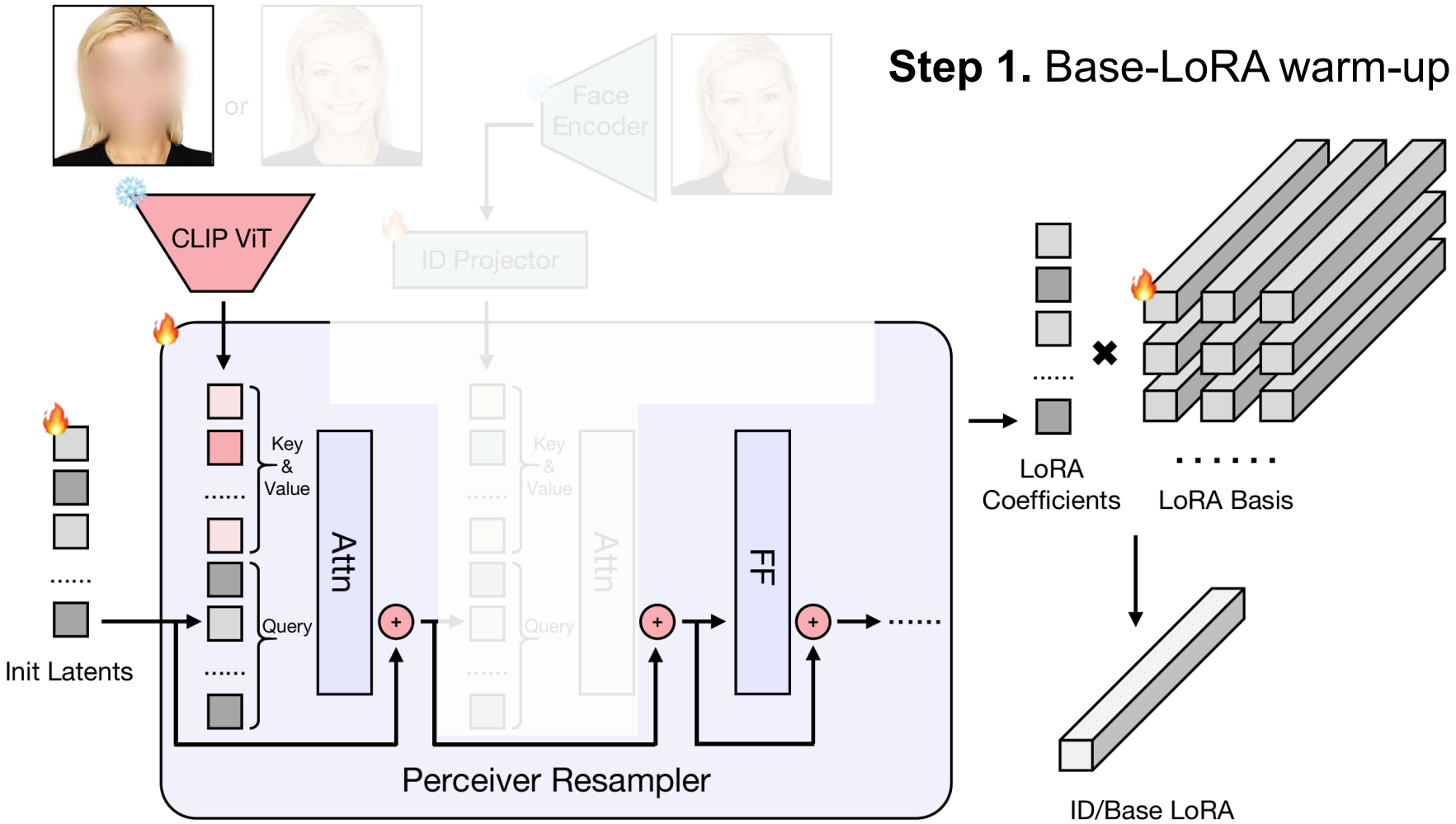
total degrees of freedom count
$N \times 8 \times K$

$$\mathbf{M} = \mathbf{M}_{base} + \mathbf{M}_{id} = \sum_{k=1}^{K} \beta_k \cdot \mathbf{M}_{base}^k + \sum_{k=1}^{K} \alpha_k \cdot \mathbf{M}_{id}^k$$

Network Structure of HyperLoRA



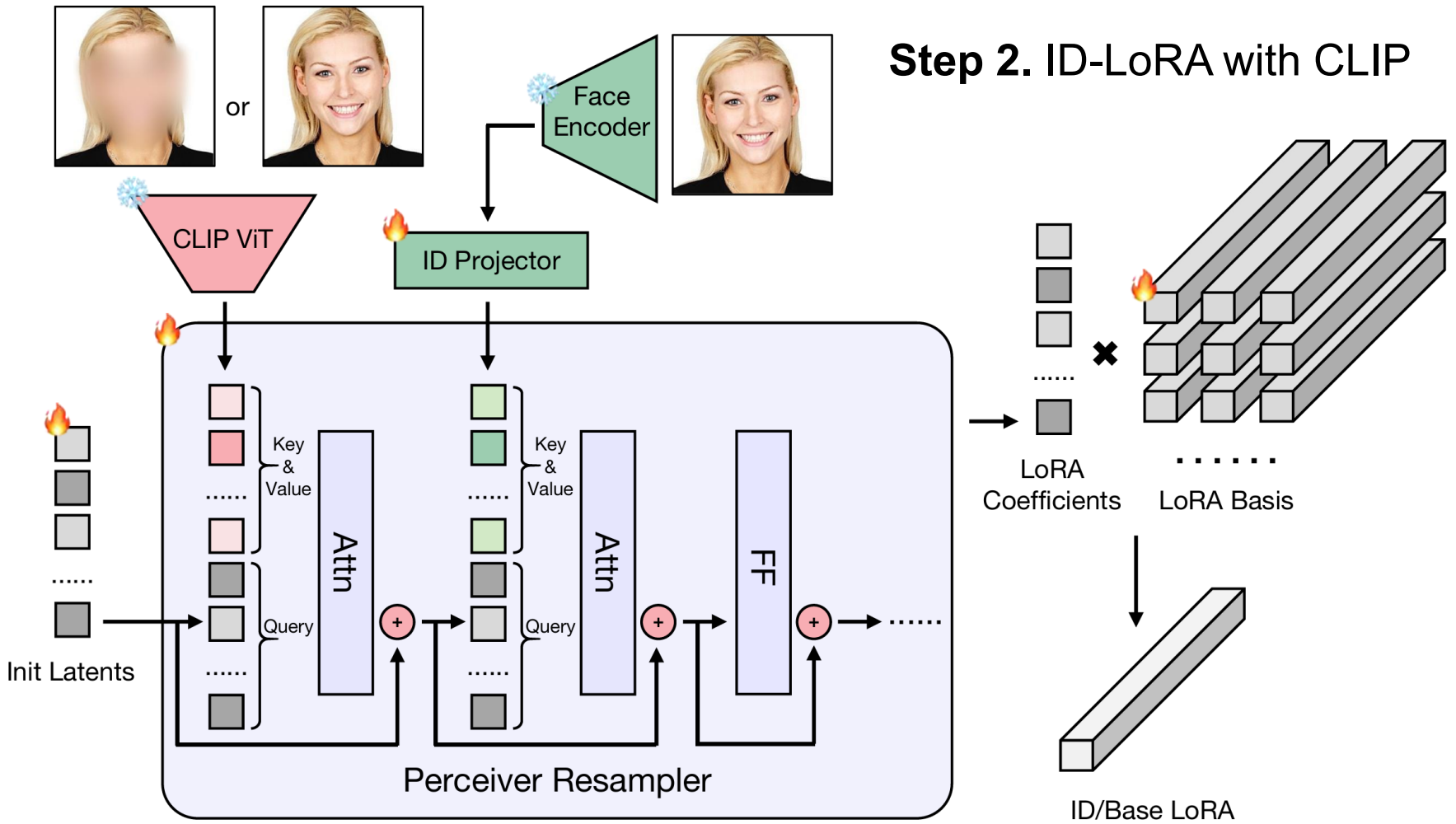
Multi-Stage Training Process



Step 1. Base-LoRA warm-up

- Train ID-LoRA from scratch is hard
- Train non-facial information into the Base-LoRA

Multi-Stage Training Process



Step 2. ID-LoRA with CLIP

3 settings

	1	2	3
Trigger words	√	×	×
Base LoRA	√	×	√
ID LoRA	√	√	√

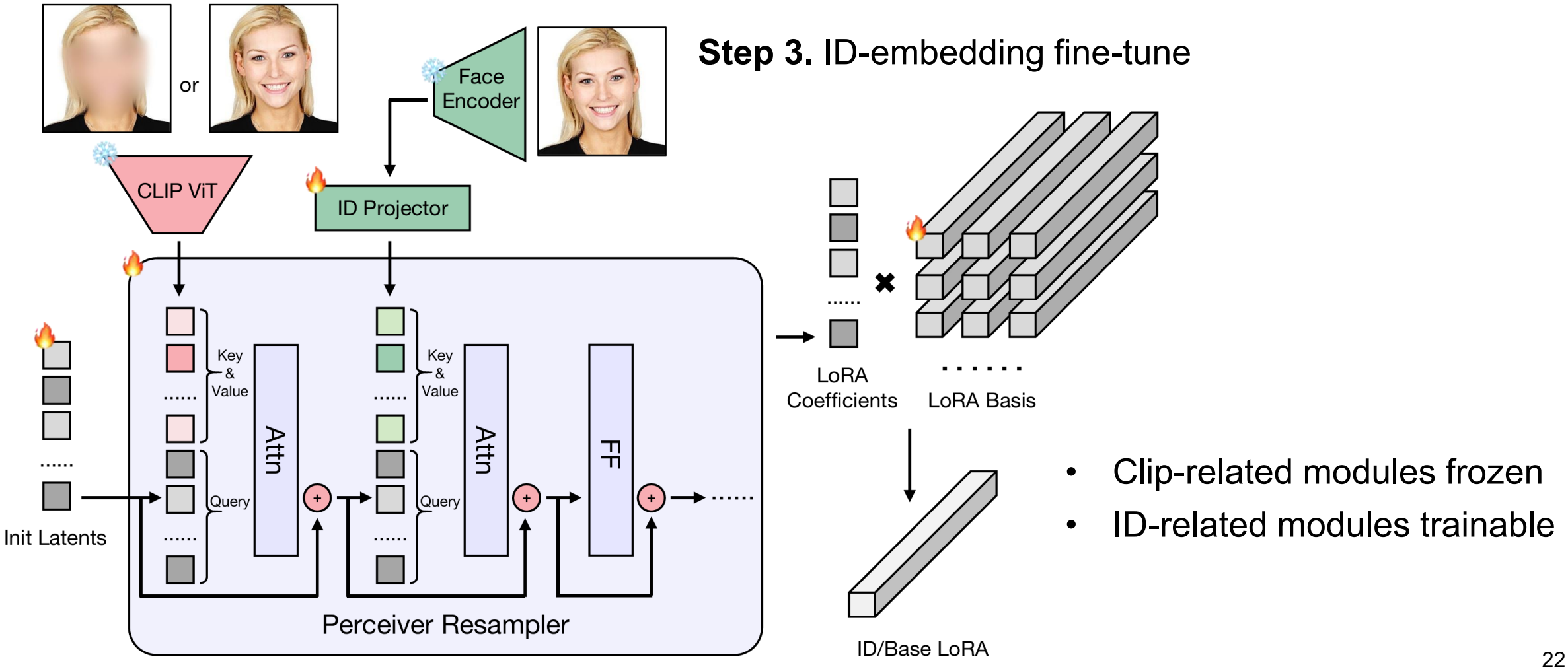
北京大学 PEKING UNIVERSITY

Multi-Stage Training Process

Step 2. ID-LoRA with CLIP

	Trigger words	Base-LoRA	ID-LoRA	Training target
Settings 1 (90%)	√	√	√	Reconstruct the target image
Settings 2 (5%)	×	×	√	Noise prediction align with foundation model
Settings 3 (5%)	×	√	√	Noise prediction align with Base-LoRA

Multi-Stage Training Process



Step 3. ID-embedding fine-tune

- Clip-related modules frozen
- ID-related modules trainable

北京大学 PEKING UNIVERSITY

Quantitative Comparison

	Fidelity		Editability	
	CLIP-I↑	ID Sim.↑	CLIP-I↑	CLIP-T↑
IP-Adapter	0.764	0.566	0.725	0.244
InstantID	0.734	**0.681**	0.688	0.237
PuLID	0.771	0.613	**0.805**	**0.259**
Arc2Face	0.786	0.643	-	-
Ours (Full)	**0.853**	<u>0.678</u>	0.710	0.243
Ours (ID)	<u>0.831</u>	0.625	<u>0.748</u>	<u>0.252</u>

北京大学 PEKING UNIVERSITY

Quantitative Comparison

Table 2. Inference speed of different methods (milliseconds).

Method	IP-Adapter	InstantID	PuLID	HyperLoRA
Preprocess	2996	758	236	1143
Inference	6148	8037	6616	4327

- **Preprocess:**

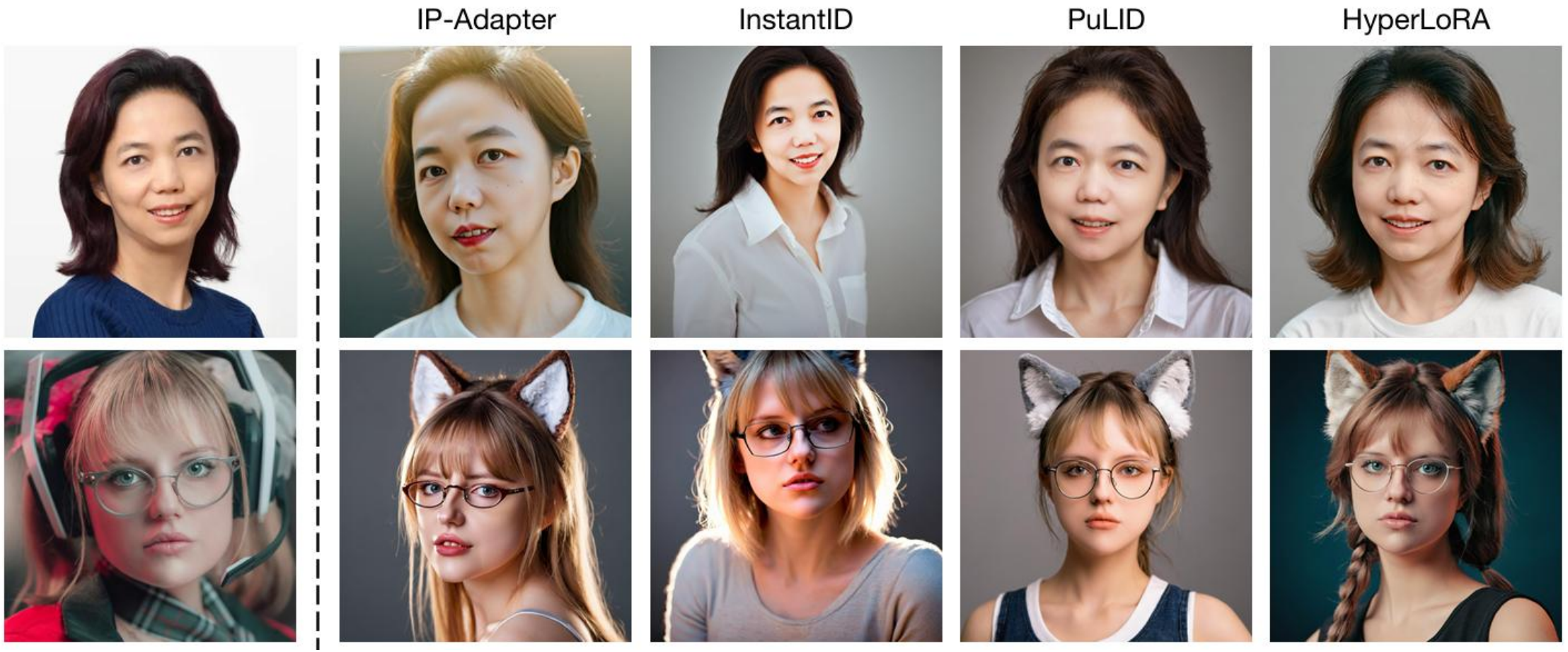
  HyperLoRA needs to generate full LoRA weights while adapters only predict ID embeddings

- **Inference:**

  HyperLoRA does not introduce extra attention modules while adapters do

北京大学 PEKING UNIVERSITY

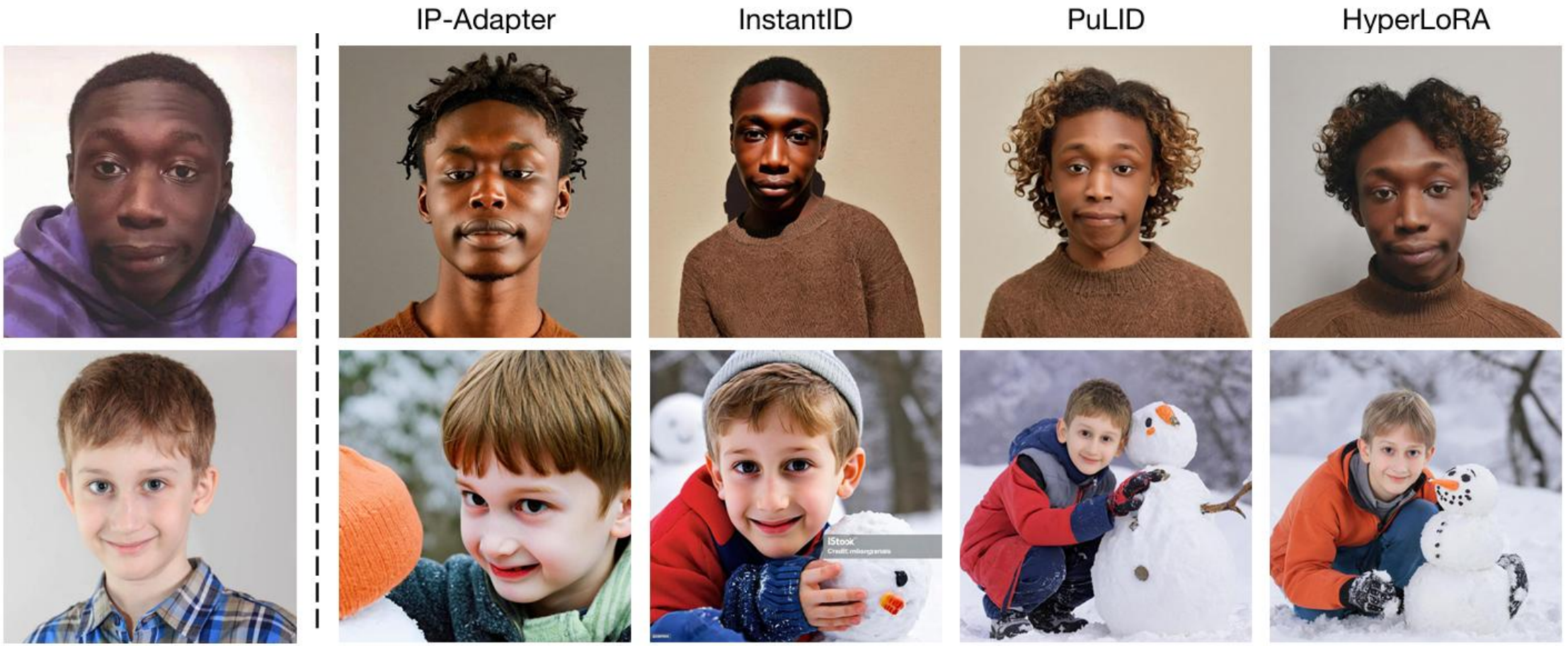
Qualitative Comparison



Top: white shirt. **Bottom:** wolf ears.

Experiments

Qualitative Comparison

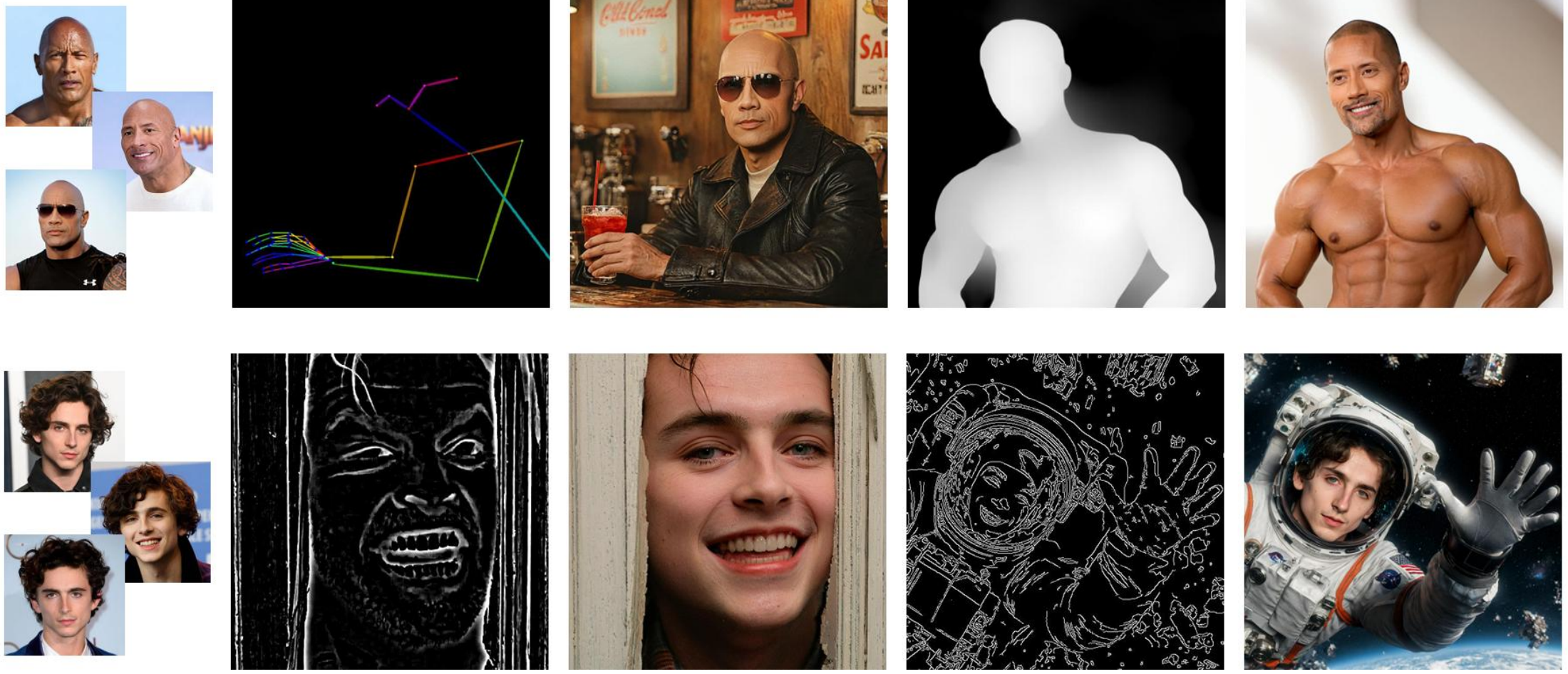


Top: wavy hair and brown sweater.

Bottom: play with snowman.

Experiments

Inference with ControlNet



Interpolability

Figure 12. ID interpolation with HyperLoRA.

Interpolability

Figure 10. Multiple input images contribute to a higher ID similarity and a more realistic and aesthetic synthesized portrait.

Conclusion

北京大学 PEKING UNIVERSITY

- **Fidelity**
  - Can take multiple reference images as input achieving better results
  - ID-LoRA learns and only learns face-relevant features
- **Editability**
  - Base-LoRA is optional while generation, allowing flexible editing strategy
  - Easy to adjust or interpolate LoRA coefficients to generate diverse output
  - Compatible with other condition control modules (e.g. ControlNet)
- **Speed**
  - Compared to adapters, HyperLoRA does not introduce additional attention modules
  - Compared to traditional LoRA, HyperLoRA does not need lengthy training process given reference images

北京大学 PEKING UNIVERSITY

Thanks for listening!

Presenter: Chenyu Niu
2025.09.21