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

Motivation

- Personalized portrait synthesis has extensive applications in the social entertainment domain. Currently, many commercial products have integrated this feature, such as TikTok, Epik, and Miaoya Camera

- Existing personalized portrait synthesis methods can be classified into two categories:
  - **Tuning-based**: DreamBooth, LoRA
  - **Tuning-free**: IP-Adapter, InstantID, PuLID

Motivation

- Personalized portrait synthesis has extensive applications in the social entertainment domain. Currently, many commercial products have integrated this feature, such as TikTok, Epik, and Miaoya Camera

- Existing personalized portrait synthesis methods can be classified into two categories:
  - **Tuning-based**: DreamBooth, LoRA
    - These methods can always produce photorealistic images, but rely on a time-consuming fine-tuning
  - **Tuning-free**: IP-Adapter, InstantID, PuLID

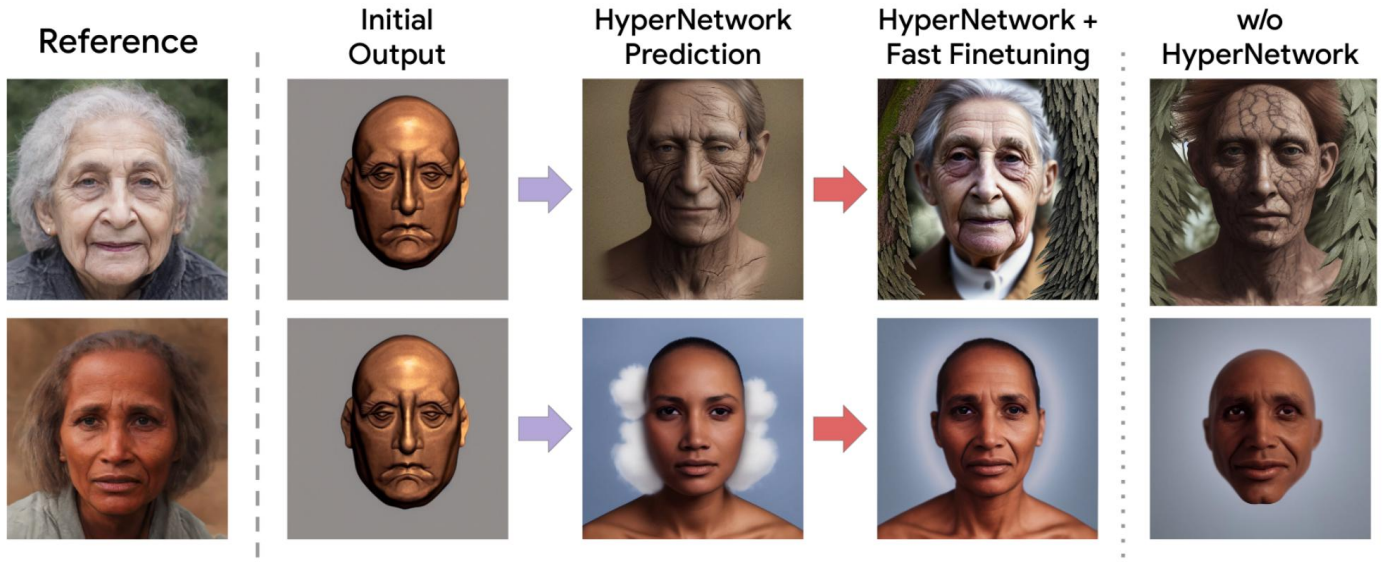
Motivation

- Personalized portrait synthesis has extensive applications in the social entertainment domain. Currently, many commercial products have integrated this feature, such as TikTok, Epik, and Miaoya Camera

- Existing personalized portrait synthesis methods can be classified into two categories:
  - **Tuning-based**: DreamBooth, LoRA
    - These methods can always produce photorealistic images, but rely on a time-consuming fine-tuning
  - **Tuning-free**: IP-Adapter, InstantID, PuLID
    - These methods usually introduce an extra cross attention module to inject identity information without online fine-tuning, however the generated images suffer from oversaturation, lack of naturalness and details

Motivation

- HyperDreambooth advances the development of portrait LoRA, but its zero-shot results are not good enough, thus still requires some fine-tuning steps



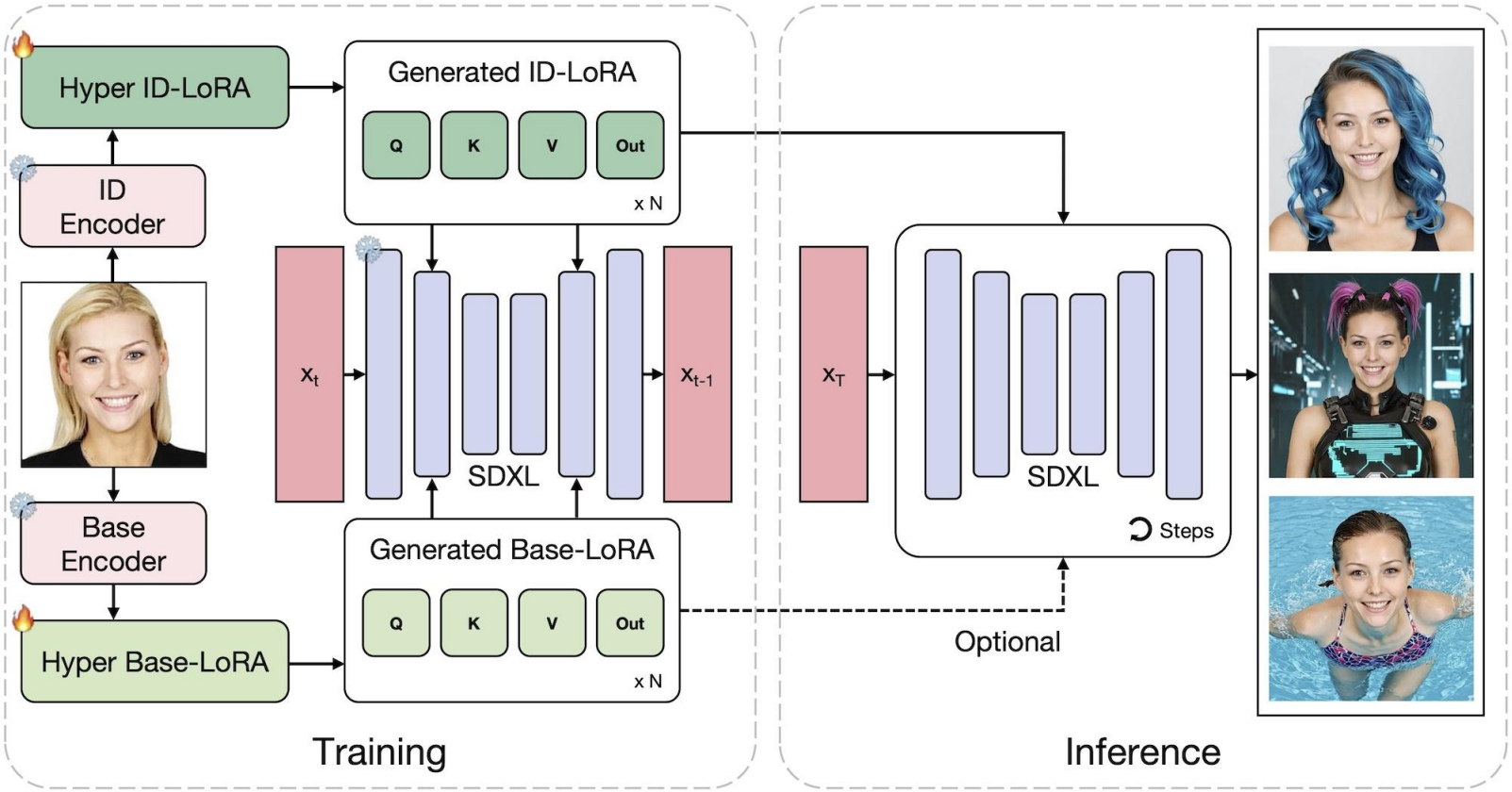
- Existing research has not yet proposed a LoRA-based method that enables zero-shot generation

Contribution

➢ We grant LoRA the zero-shot capability for portrait synthesis

- We propose the first **zero-shot** portrait synthesis method based on LoRA, which is trained in an **end-to-end manner**
- We introduce a well-designed model architecture and training scheme, to facilitate **efficient training** and **ID decoupling**
- Our method can produce highly **photorealistic** and **detailed** images while ensuring **fidelity**, **editability** and **inference speed**

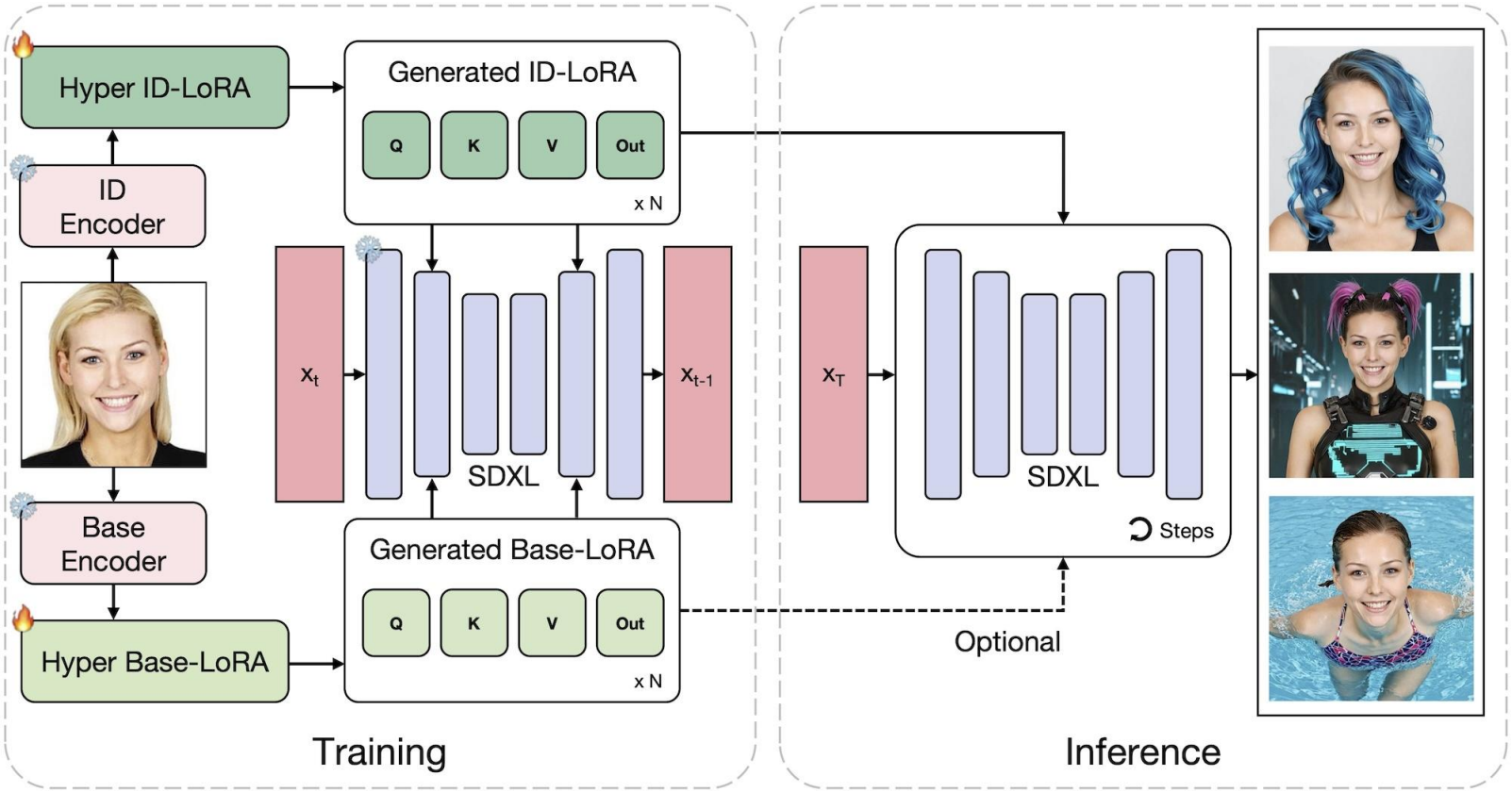
Method



We decompose the HyperLoRA into a **ID-LoRA** and a **Base-LoRA**

- **ID-LoRA**: learn facial identity
- **Base-LoRA**: learn others, such as layout, clothing and hairstyle

Method



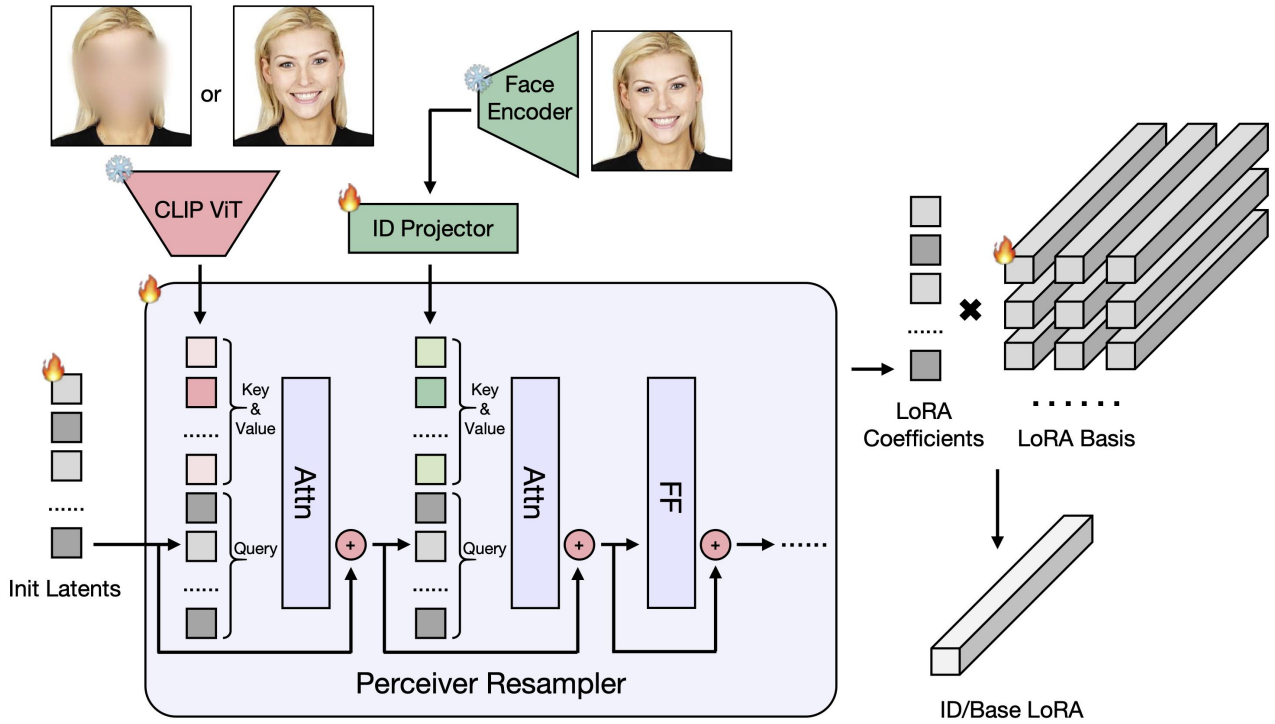
Such a design has the following benefits:

- Prevent irrelevant features leaking to ID-LoRA
- Mitigate the impact of training data on image quality

Method

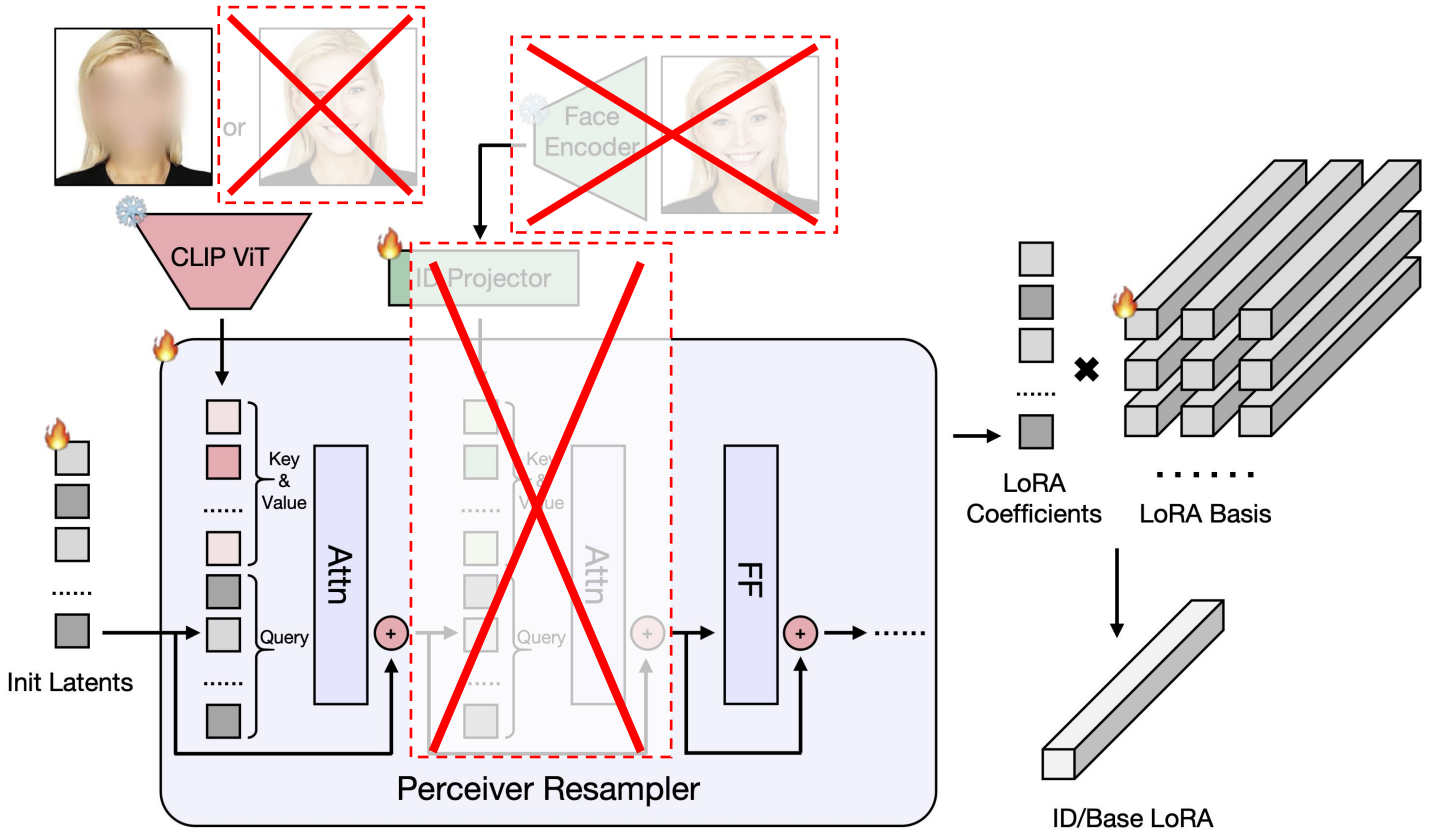
To make HyperLoRA training feasible, we represent a LoRA matrix **M** with a linear combination of ***K* = 128** dim LoRA basis, and leverage Perceiver Resampler to predict the coefficients **α** and **β**

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



Method

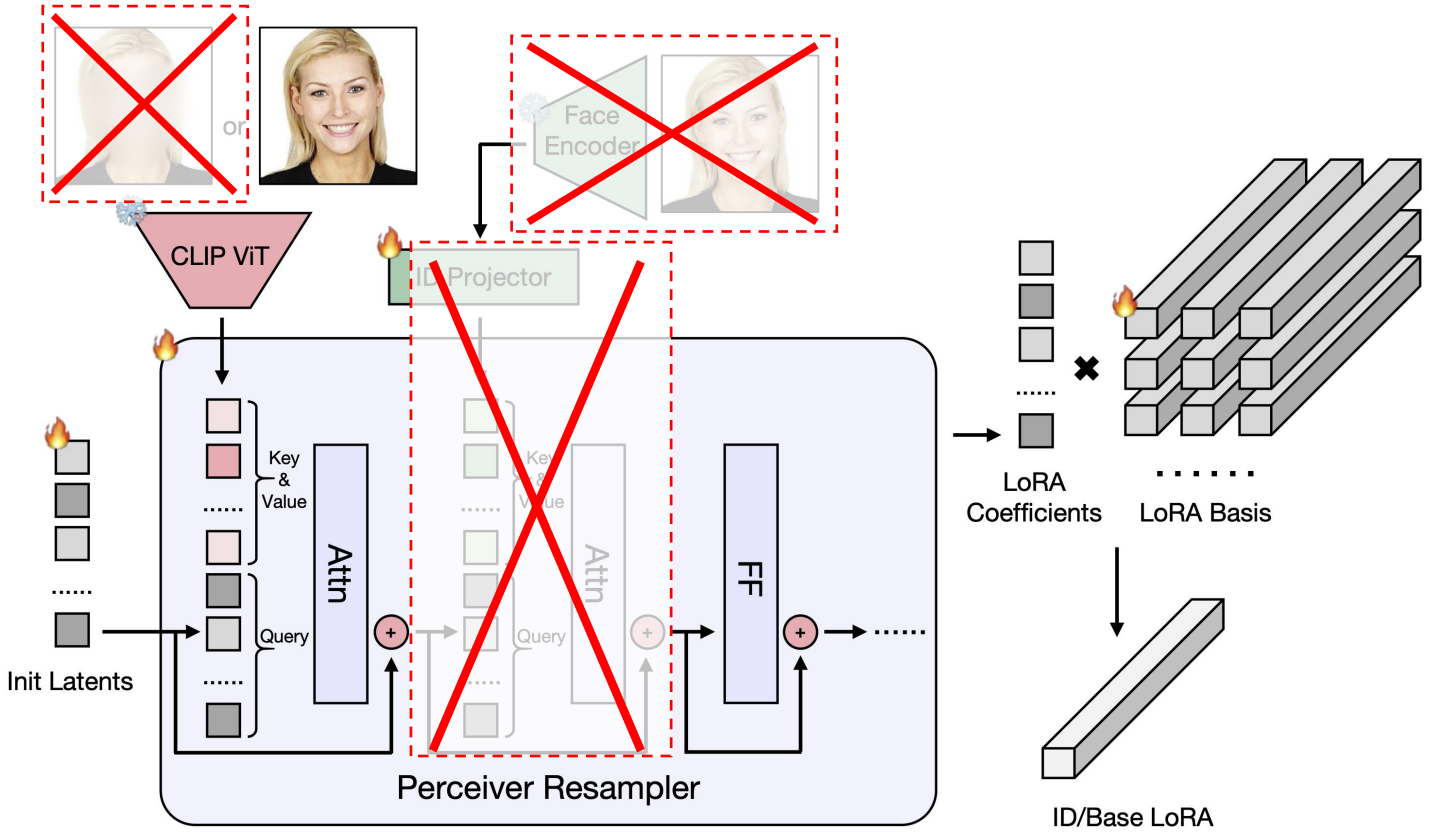
3-stage training scheme: (1) Train the Base-LoRA



- Blur the facial region of original input image
- Bypass the Face Encoder (AntelopeV2) branch
- Learn ID-irrelevant features

Method

3-stage training scheme: (2) Train the ID-LoRA with CLIP feature while bypassing Face Encoder branch

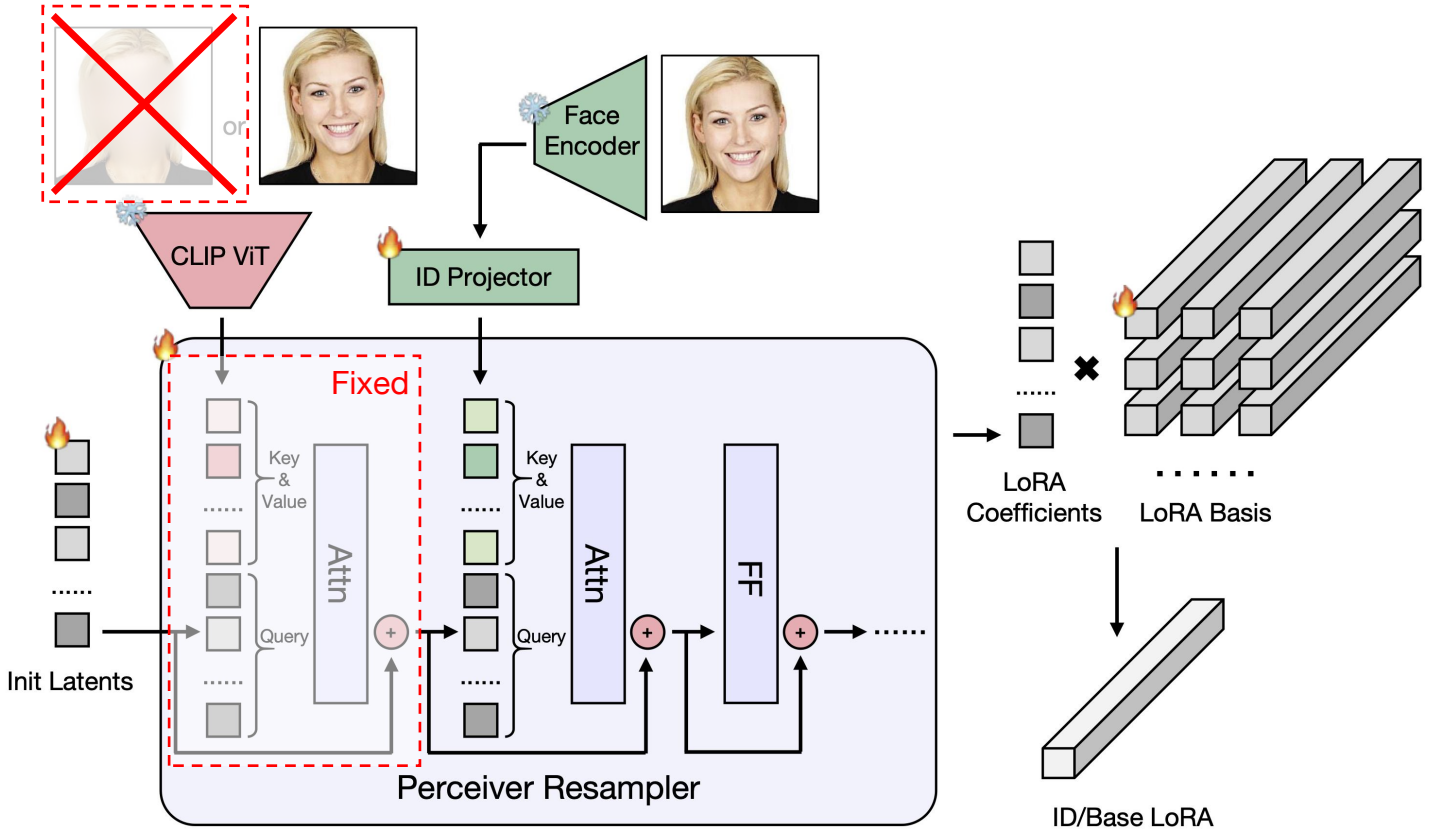


Associate ID-LoRA with some rare words (a.k.a. trigger words), to reduce the impact on other meaningful concepts

- with trigger words, both Base-LoRA and ID-LoRA enabled => reconstruct the target image
- without trigger words, only ID-LoRA enabled => align with the output of base model
- without trigger words, both Base-LoRA and ID-LoRA enabled => align with the output of Base-LoRA

Method

3-stage training scheme: (3) Train the ID-LoRA with ID feature and fixing CLIP ViT branch



- Training with CLIP ViT feature facilitates fast convergence
- Switching to ID embeddings for the subsequent training alleviates the structural constraints on face

Experimental Result

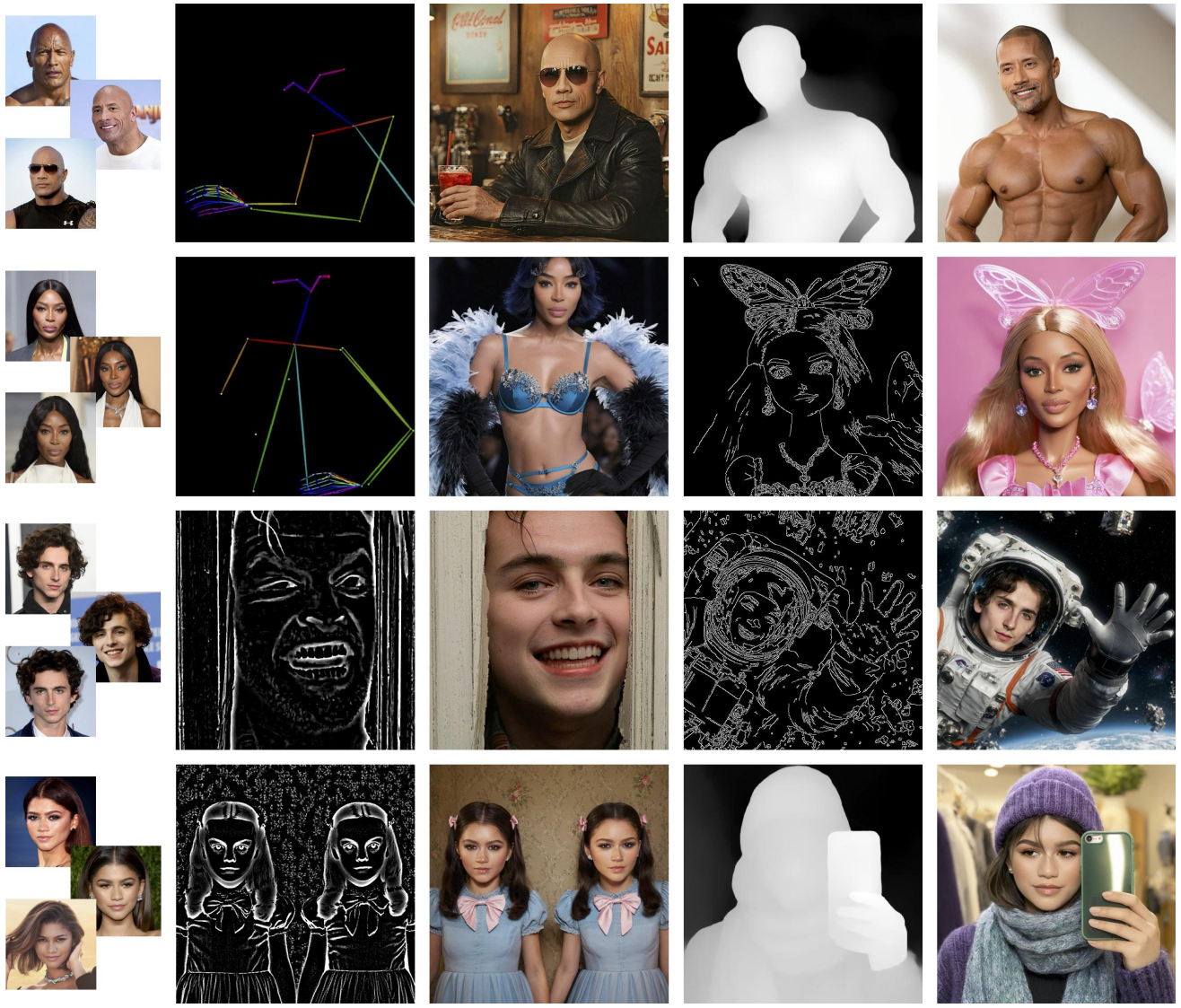
└ Comparison with previous methods

Experimental Result

└ T2I with HyperLoRA

Experimental Result

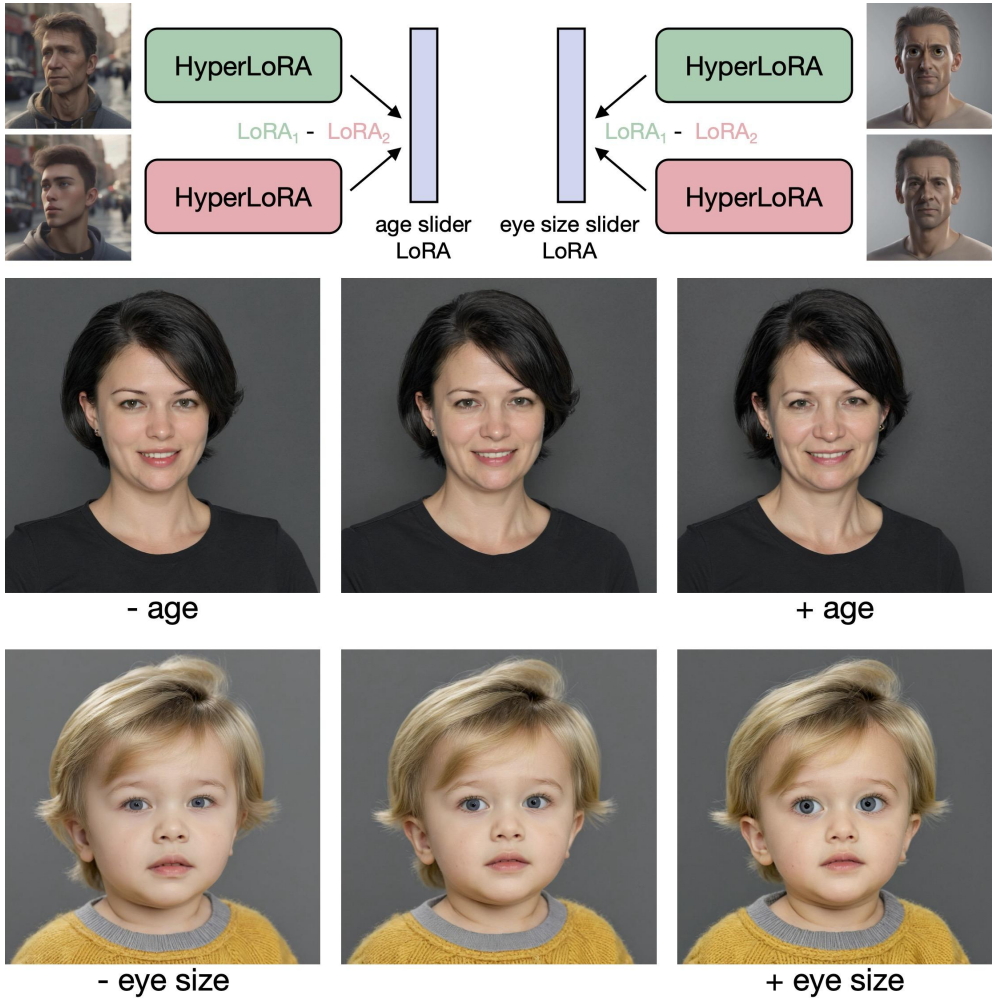
└ HyperLoRA is compatible with ControlNet



Experimental Result

└ Concept Slider LoRA

HyperLoRA exhibits the similar properties as StyleGAN. For example, a concept slider LoRA can be created by the LoRAs generated from a pair of images



Thank you for watching

Code and model available!

https://github.com/bytedance/ComfyUI-HyperLoRA

https://huggingface.co/bytedance-research/HyperLoRA