香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Creative, Controllable & Cognitive Computing Group

Foundation Cures Personalization: Improving Personalized Models' Prompt Consistency via Hidden Foundation Knowledge

Yiyang Cai[1], **Zhengkai Jiang**[2], **Yulong Liu**[1], **Chunyang Jiang**[1]
Wei Xue[1], **Yike Guo**[1], **Wenhan Luo**[1*]
[1] Hong Kong University of Science and Technology (HKUST)
[2] Tencent Hunyuan
https://yiyangcai.github.io/freecure-aigc.github.io/

Background

Definition of Facial Personalization

Given a limited number of images that depict particular identities, facial personalization models generate novel content that reflects these identities through diverse conditions.

Important metrics: identity fidelity (similarity between generated faces and reference faces) & prompt consistency (generated faces follow users' prompts)

Background

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Creative, Controllable & Cognitive Computing Group

Personalized Models often suffer from the "copy and paste" problem: generating faces which are very similar to reference inputs (not controllable when we want to generate novel hairstyle, expressions, accessories).
When we input prompts to control ID' s attributes, they often failed, making the prompt following performance low.

PuLID

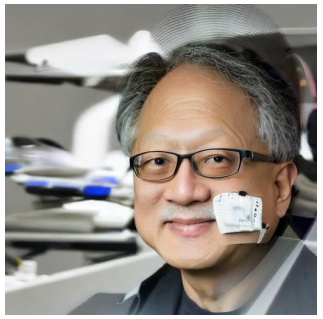
a man with blonde curly hair and blue eyes

PhotoMaker

a woman with angry looking

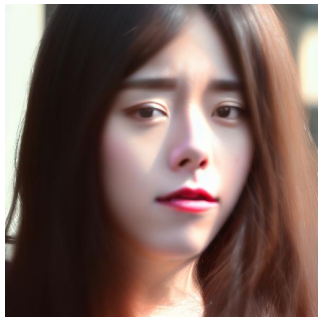
FaceDiffuser

a man wearing a mask



Face2Diffusion

a girl with white curly hair, frowning worriedly



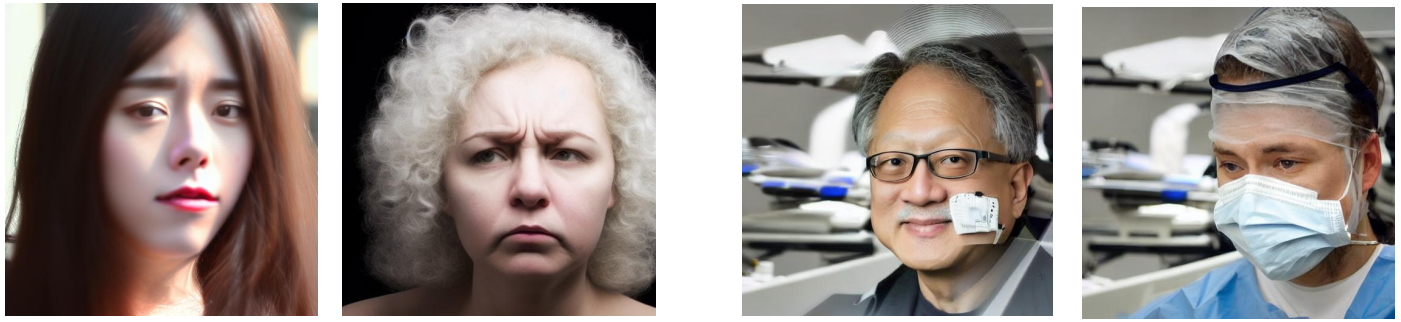
Observation of FreeCure

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Creative, Controllable & Cognitive Computing Group

We discover a fact that when ID conditions are absent, their prompt following ability is still satisfying.

Each pair comes from the same model



Question is: how to combine them together and generate results whose ID fidelity and prompt consistency are both good?

Observation of FreeCure

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Creative, Controllable & Cognitive Computing Group

Our empirical findings reveals a fact that identity embedding override the expression of attribute-related tokens, by visualizing attention maps of several personalization models

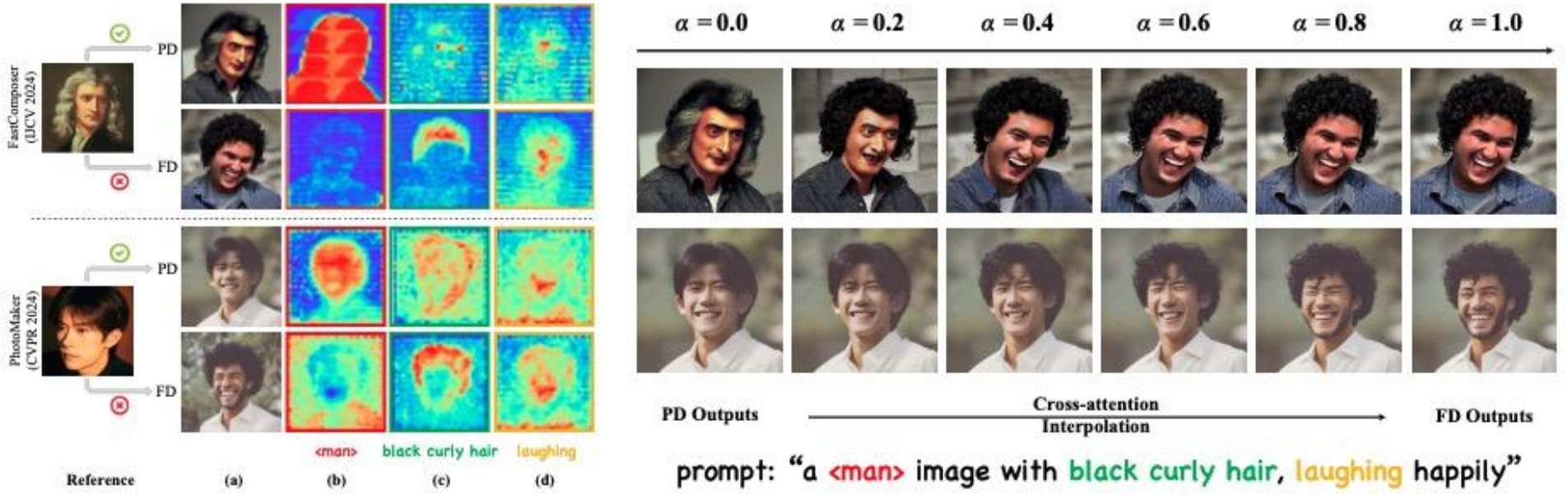


Figure 2: **Analysis on cross-attention maps of facial personalization models. Left**: token-wise attention map visualization. **Right**: interpolation experiment on PD and FD's cross-attention maps.

Is it feasible to mitigate the erosion of prompt consistency in personalization models while keeping their trained cross-attention modules unaffected?

Method

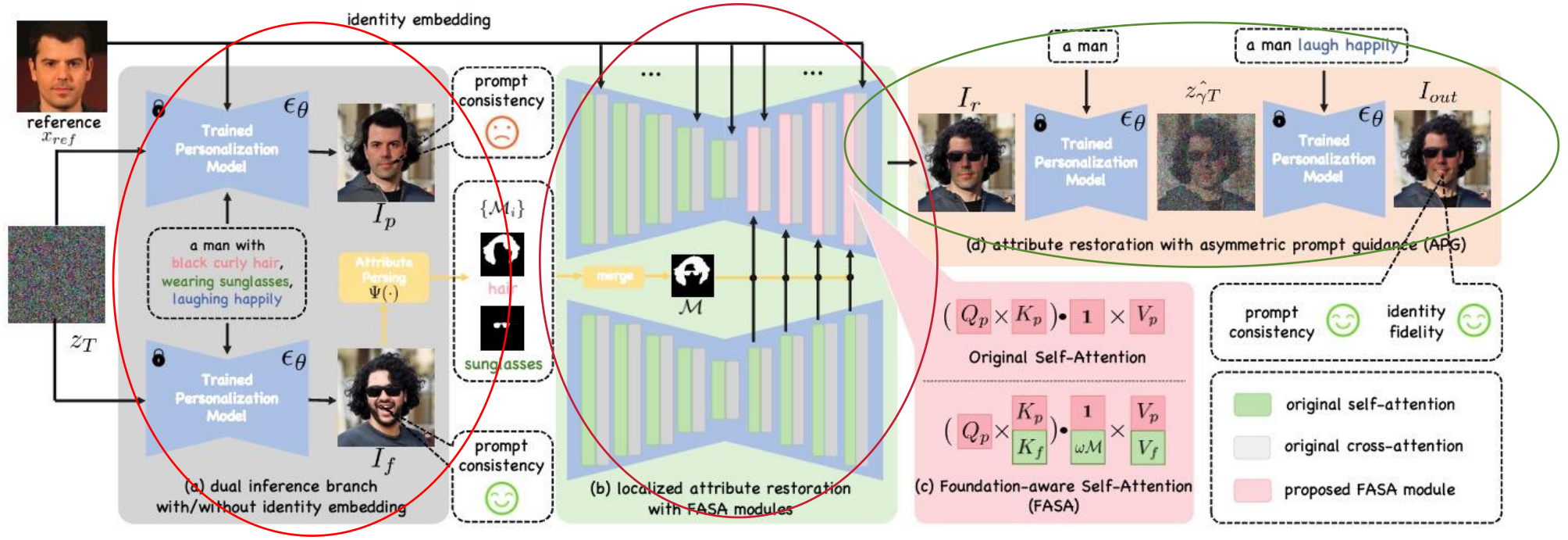
香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Creative, Controllable & Cognitive Computing Group

To keep the original ability of identity preservation (cross-attention adapters), we modify sefl-attention modules in personalization models to improve their prompt consistency.

Use mask and new self-attention module to merge two denoising process

Use a inversion process to deal with abstract attributes such as expression



Generate two results with/without ID

Method

Foundation-aware self-attention (FASA):

UNet version:

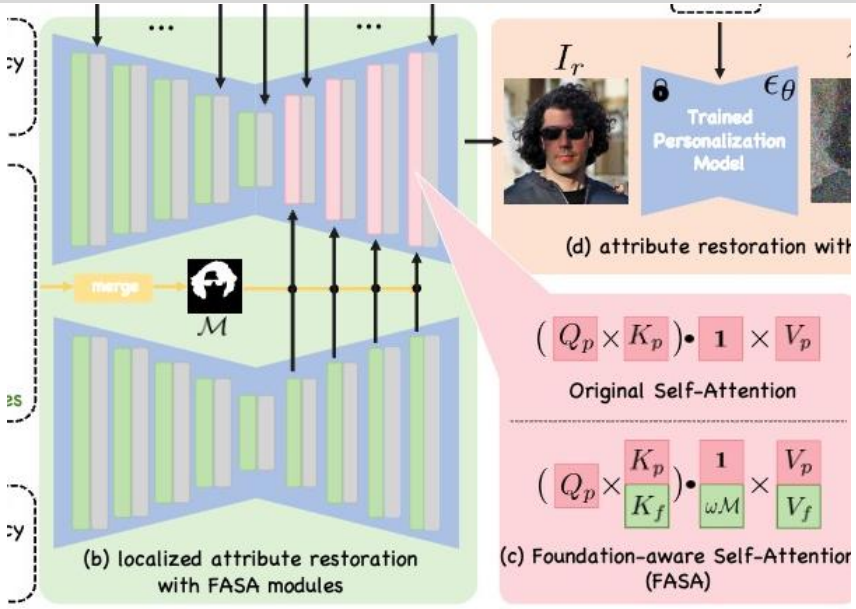
$$\text{FASA}(\mathcal{KQV}_p, \mathcal{KQV}_f) = \text{Softmax}(\frac{[\mathbf{1}, \omega\mathcal{M}] \odot Q_p\hat{K}^T}{\sqrt{d}})\hat{V}.$$

FLUX version (similar to OmniControl):

$$\text{FASA}_{flux}(\mathcal{KQV}_p, \mathcal{KQV}_f) = \text{Softmax}(\frac{\mathcal{M}(\omega)_{flux} \odot Q_p\hat{K}^T}{\sqrt{d}})\hat{V},$$

$$\mathcal{M}(\omega)_{flux} = \begin{pmatrix} \mathbf{1}_{l_1\times l_1} & \mathbf{1}_{l_1\times l_2} & \omega\mathcal{M}_{l_1\times l_1} \\ \mathbf{1}_{l_2\times l_1} & \mathbf{1}_{l_2\times l_2} & \mathbf{0}_{l_2\times l_1} \end{pmatrix}$$

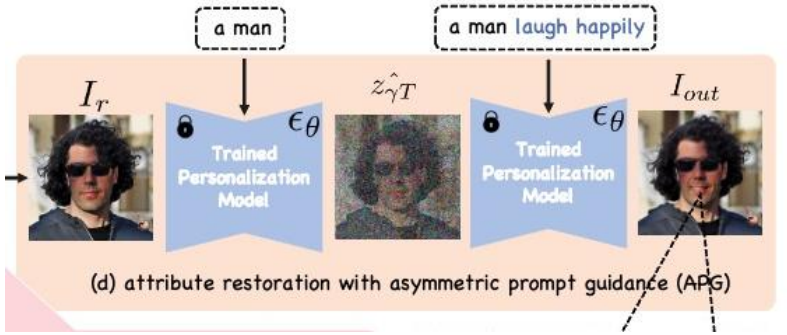
Semantic masks can ensure that only areas related to target attribtues are enhanced.



Method

APG

Followed the idea of diffusion inversion, we use a template prompt to convert images to noised ones and use the original prompt in the denoising process, which enhances abstract attributes such as expressions.



Results

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Creative, Controllable & Cognitive Computing Group

Table 1: **Main quantitative evaluation results.** With FreeCure, the mainstream personalization models' prompt consistency is highly enhanced on critical quantitative metrics.

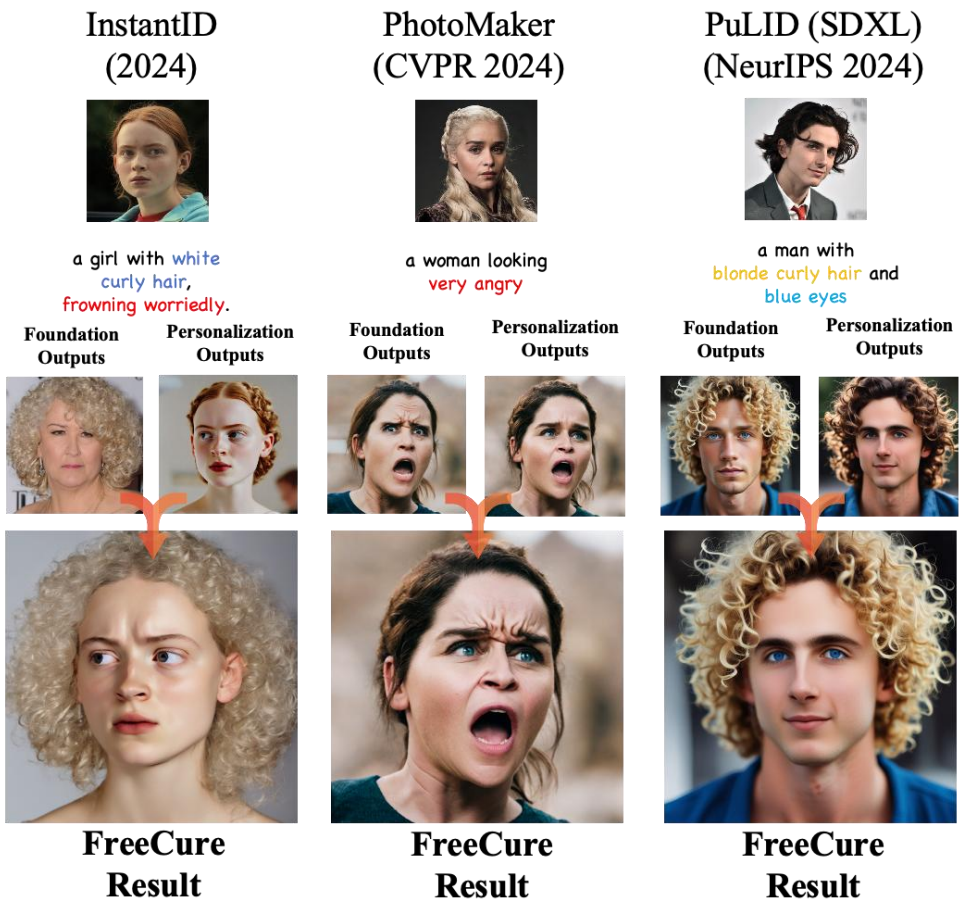
Method	PC(%) ↑	IF(%) ↑	Face Div. (%) ↑	PC × IF (hMean) ↑
FastComposer	18.14	**43.19**	38.92	25.55
FastComposer + **FreeCure**	**21.02** (+15.91%)	41.02 (-5.02%)	**41.01**(+5.37%)	**27.80** (+8.82%)
Face-Diffuser	20.67	**58.34**	40.82	30.52
Face-Diffuser + **FreeCure**	**22.48** (+8.76%)	57.51 (-1.42%)	**41.95**(+2.77%)	**32.32** (+5.90%)
Face2Diffusion	21.92	**39.98**	43.51	28.31
Face2Diffusion + **FreeCure**	**23.26** (+6.12%)	39.23 (-1.88%)	**44.29**(+1.79%)	**29.20** (+3.15%)
InstantID	21.89	**63.94**	48.98	32.61
InstantID + **FreeCure**	**23.62** (+7.90%)	62.01(-3.02%)	**51.82** (+5.80%)	**34.21** (+4.91%)
PhotoMaker	23.04	**51.84**	47.29	31.90
PhotoMaker + **FreeCure**	**24.91** (+8.11%)	50.15 (-3.26%)	**48.52** (+2.60%)	**33.28** (+4.34%)
PuLID (SDXL)	25.16	**58.23**	42.12	35.14
PuLID (SDXL) + **FreeCure**	**26.05** (+3.55%)	56.95 (-2.20%)	**43.72**(+3.80%)	**35.74** (+1.74%)
PuLID (FLUX)	22.42	**74.97**	43.91	34.52
PuLID (FLUX) + **FreeCure**	**24.78** (+10.53%)	72.61 (-3.15%)	**46.09**(+4.96%)	**36.95** (+7.04%)
InfiniteYou	23.77	**79.71**	44.28	36.62
InfiniteYou + **FreeCure**	**25.25** (+6.23%)	77.13 (-3.24%)	**46.82**(+5.74%)	**38.05** (+3.90%)

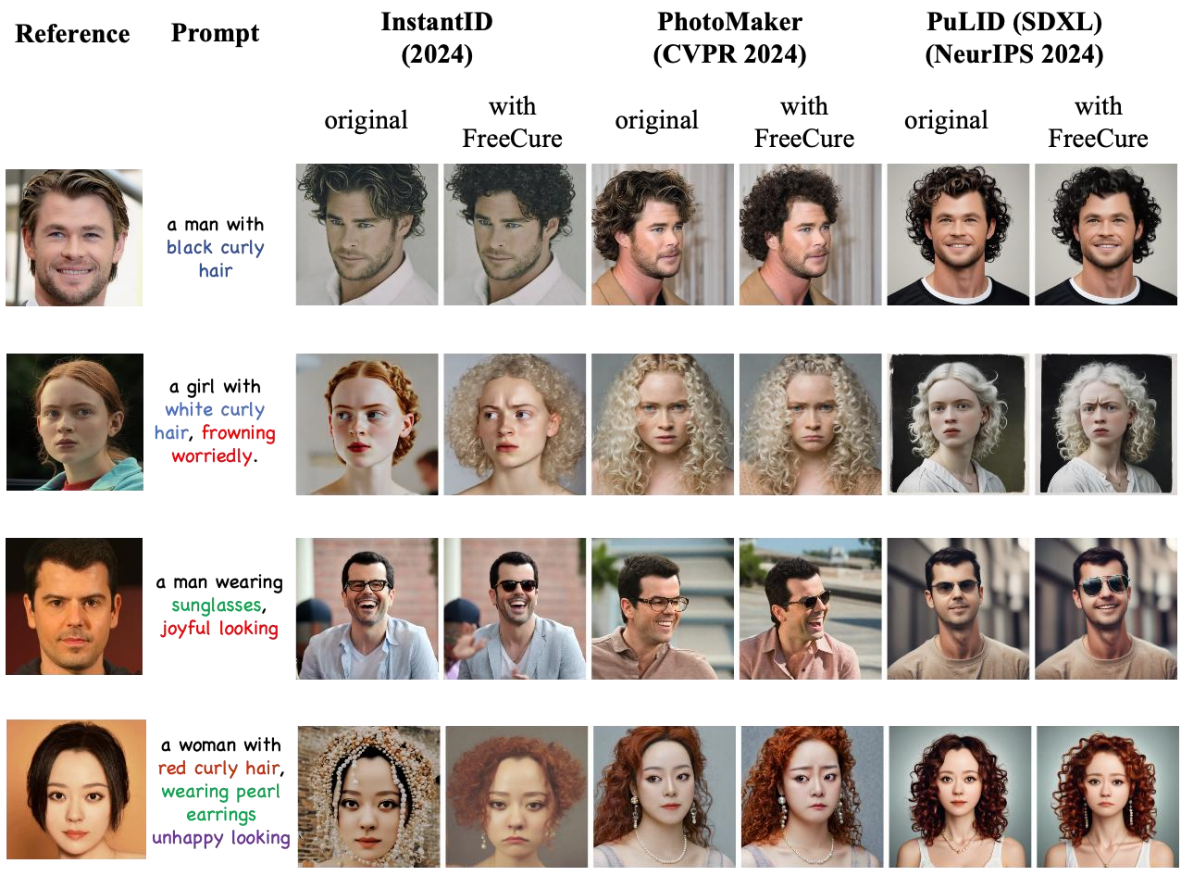
Results

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Creative, Controllable & Cognitive Computing Group

FreeCure can be applied on several personalization models, no matter they are trained based on Stable Diffusion or FLUX



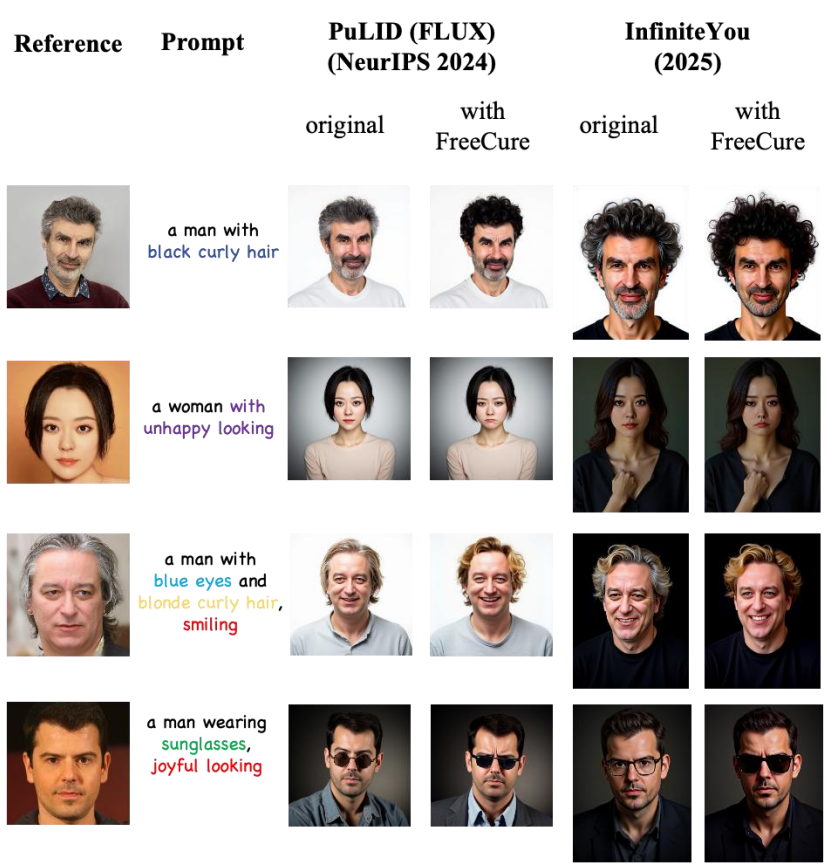


FreeCure

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Creative, Controllable & Cognitive Computing Group

FreeCure can be applied on several personalization models, no matter they are trained based on Stable Diffusion or FLUX

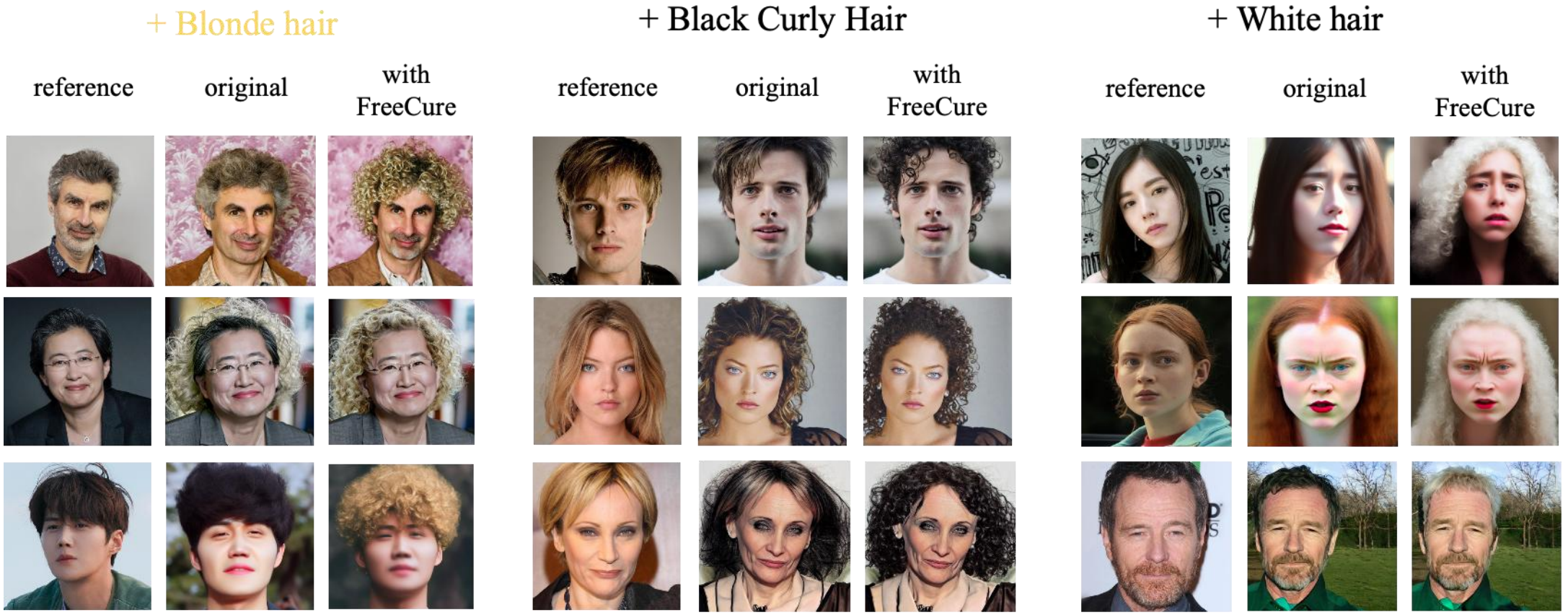


FreeCure

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Creative, Controllable & Cognitive Computing Group

FreeCure can handle different facial attributes' enhancement

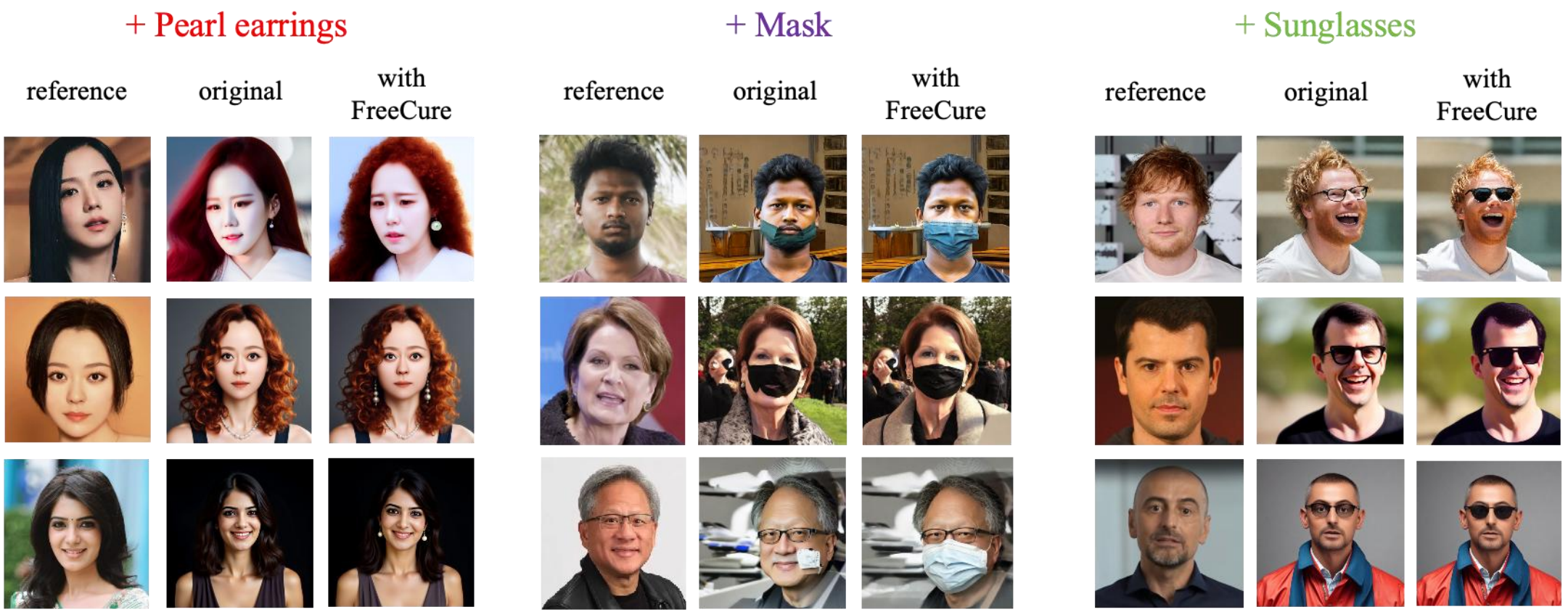


FreeCure

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Creative, Controllable & Cognitive Computing Group

FreeCure can handle different facial attributes' enhancement

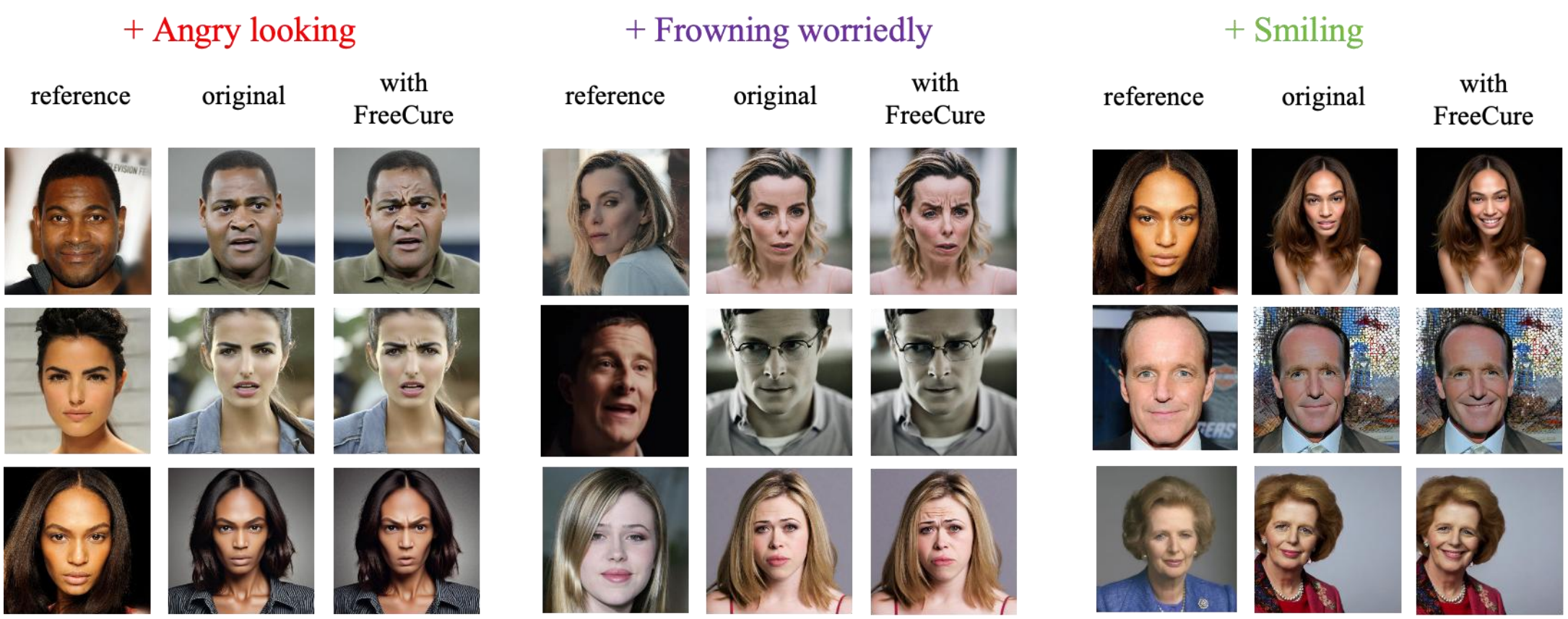


FreeCure

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Creative, Controllable & Cognitive Computing Group

FreeCure can handle different facial attributes' enhancement



FreeCure

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Creative, Controllable & Cognitive Computing Group

FreeCure can handle multi-attribute enhancement

+ Sunglasses; + Smiling

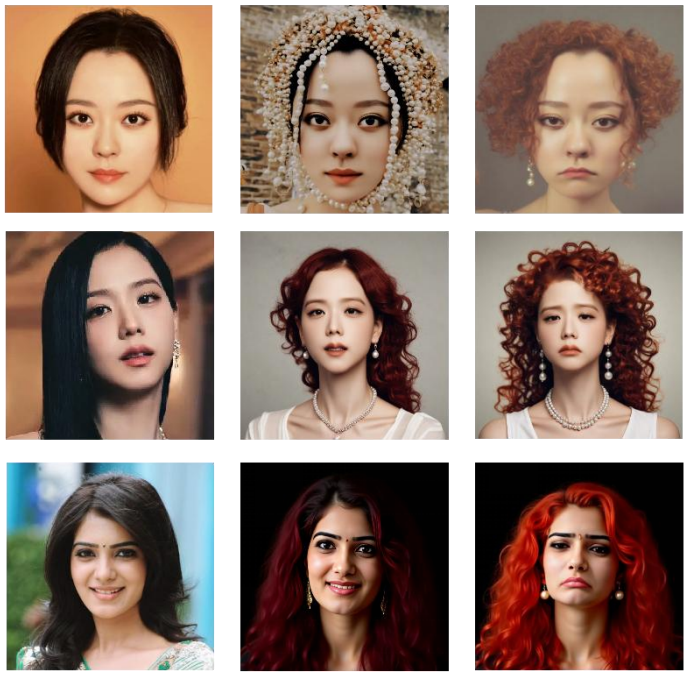
reference | original | with FreeCure

+ Blonde curly hair; + Blue eyes; + Smiling

reference | original | with FreeCure

+ Pearl earrings; + Red curly hair; + Unhappy looking

reference | original | with FreeCure



香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
Creative, Controllable & Cognitive Computing Group

Robustness Analysis

Difference initial noise:

FASA's performance is not constrained to initial noise, when we apply different noises for FD and PD process, the performance can still be significant.

Visualization of FASA:

When we select a pixel (in latent space) in personalization denoising query. Selecting its row (include two key matrices) and convert its value into heat map:

1. if this pixel belongs to region of target attributes, attention scores are higher at foundation branch (meaning attribute information flows from foundation to personalization)

2. if this pixel belongs to region other than target attributes, attention scores are higher at original personalization branch (meaning stay unchanged)

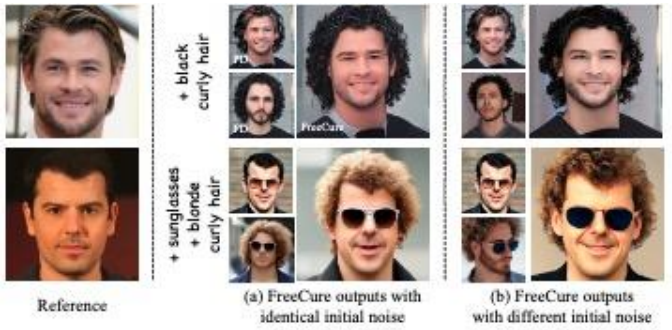


Figure 7: **Performance of FASA w/ and w/o identical initial noises**. FASA can precisely enhance attributes even if PD and FD produce faces with different locations, sizes, and angles.

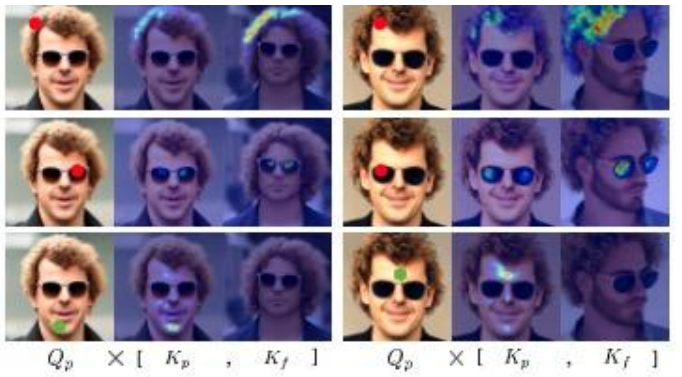
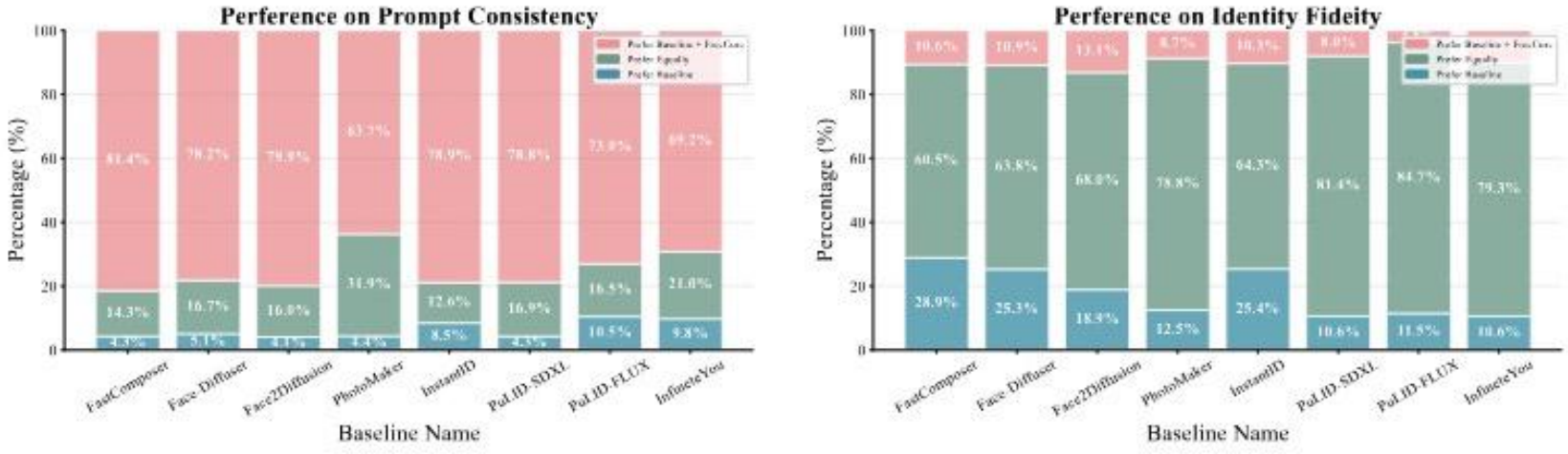


Figure 8: **Visualization of the FASA maps** for attribute related area (red points) and non-attribute related area (green points).

User study

By applying FreeCure on different baselines, users can have significant greater chance to prefer their prompt concsistency.

Meanwhile, their preference on identity fidelity between baselines with/without FreeCure does not change much, indicating FreeCure does not undermine model’ s original performance in identity fidelity.



(a) User preference on prompt consistency

(b) User preference on identity fidelity

Figure 10: **User study of FreeCure**. The preference ratio indicate that FreeCure can improve prompt consistency without undermining identity fidelity of different personalization models.

FreeCure

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Creative, Controllable & Cognitive Computing Group

More info about FreeCure:

Original paper (to be updated): https://arxiv.org/abs/2411.15277

Code: https://github.com/YIYANGCAI/FreeCure

Project Page: https://yiyangcai.github.io/freecure-aigc.github.io/

Foundation Cures Personalization: Improving Personalized Models' Prompt Consistency via Hidden Foundation Knowledge

Yiyang Cai[1], **Zhengkai Jiang**[2], **Yulong Liu**[1], **Chunyang Jiang**[1]
Wei Xue[1], **Yike Guo**[1], **Wenhan Luo**[1*]
[1] Hong Kong University of Science and Technology (HKUST)
[2] Tencent Hunyuan
https://yiyangcai.github.io/freecure-aigc.github.io/