UniPortrait: A Unified Framework for Identity-Preserving Single- and Multi-Human Image Personalization

Junjie He, Yifeng Geng, Liefeng Bo

Tongyi Lab, Alibaba Group

Introduction

Text-to-Single-ID Personalization

"Blonde with hair beaten into loose waves. Tattoos on both arms and multiple bracelets on each wrist. Green lace blouse with high-waisted jeans. Yellow walls."

"The man was sitting at a table with a laptop on it. With one hand over his head, he looked at the screen with a worried expression on his face."

"An old woman, white hair and a wrinkled face, a serious expression. Wearing a black coat and a straw hat, she stood in a snowy field with a white sky background."

Stylized Portrait Synthesis

Identity-Preserving Style Transfer With ControlNet

Text-to-Multi-ID Personalization
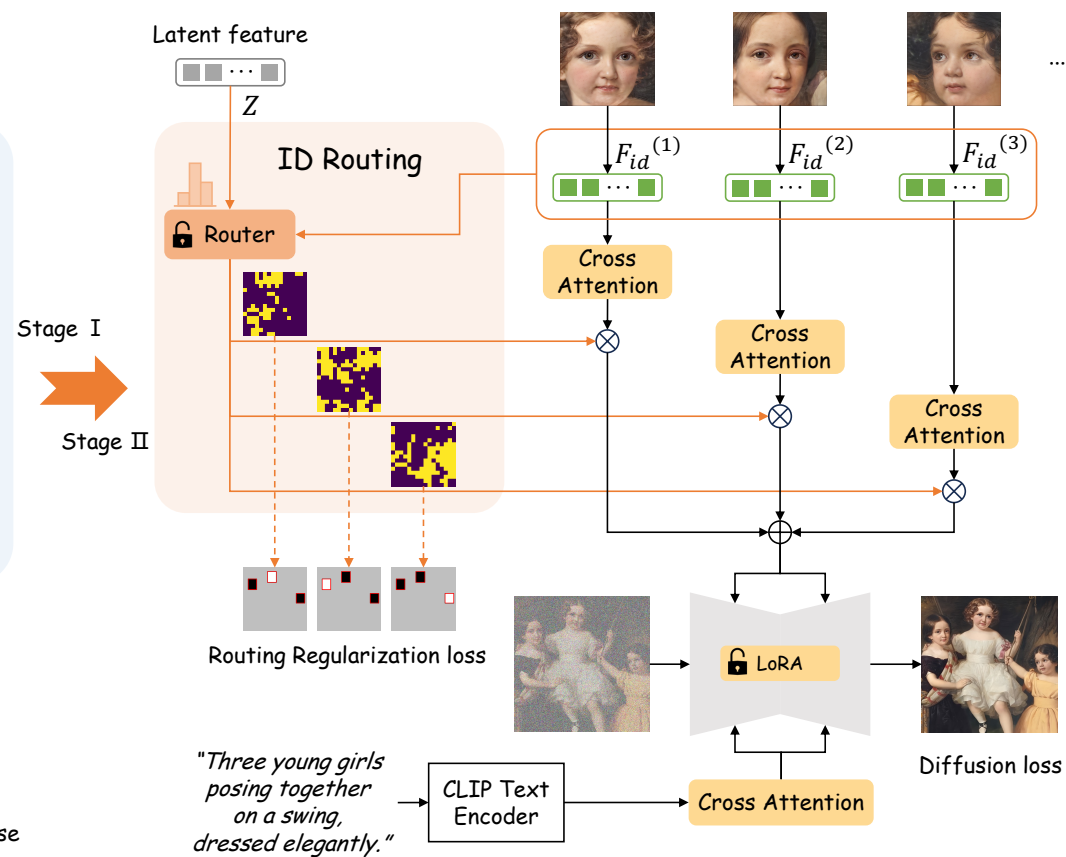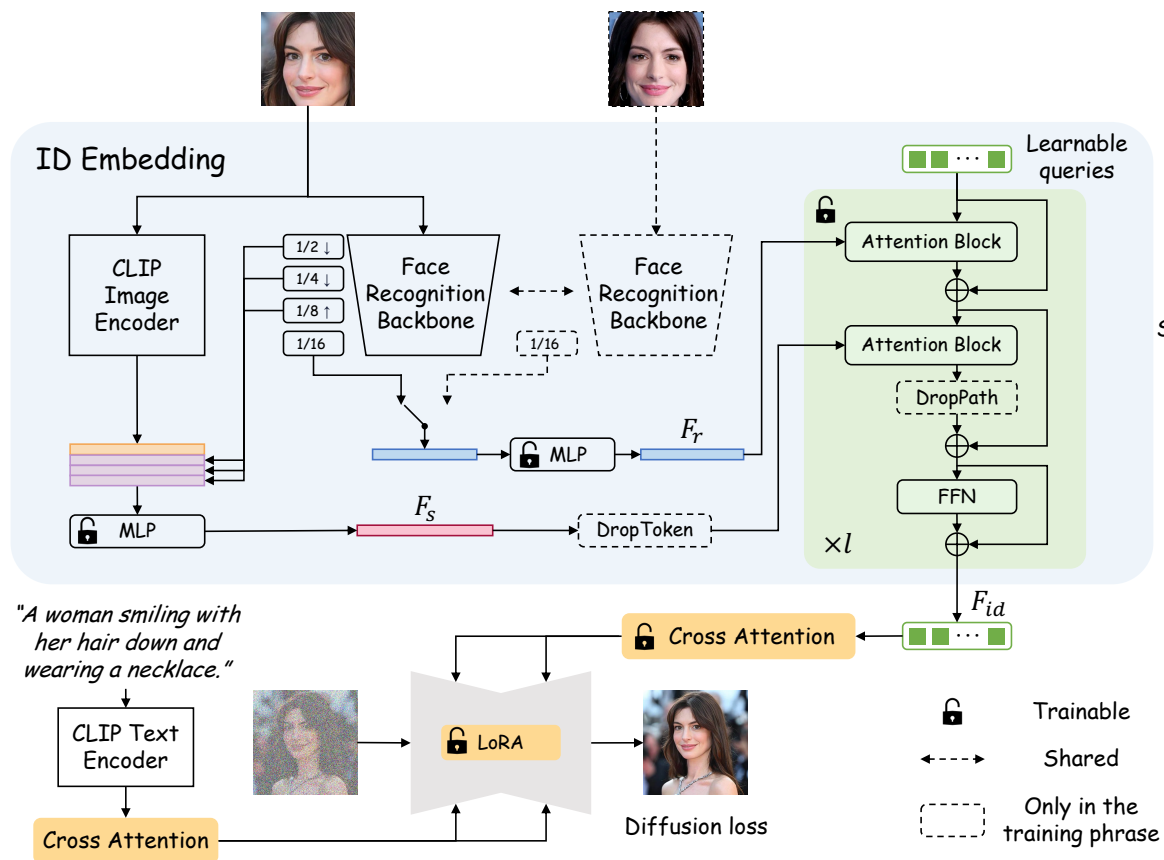

"Three men in a coffee shop."

"Two women, retro comic style artwork, comic book cover, symmetrical, vibrant."

Contributions

- We propose UniPortrait, an innovative human image personalization framework that unifies single- and multi-ID customization with high face fidelity and controllability

- We propose a novel, efficient ID embedding module with a decoupling strategy, which embeds detailed face identity information while maintaining good editability

- We introduce the ID routing mechanism, which addresses the identity blending issue in multi-ID customization yet without compromising each identity integrity, generated image diversity, and prompt design flexibility

Method

"A woman smiling with her hair down and wearing a necklace."

"Three young girls posing together on a swing, dressed elegantly."

ID Embedding

CLIP Image Encoder

Face Recognition Backbone

Face Recognition Backbone

F_r

F_s

Learnable queries

Attention Block

DropPath

Attention Block

FFN

$\times l$

F_{id}

CLIP Text Encoder

Cross Attention

LoRA

Diffusion loss

Trainable

Shared

Only in the training phrase

Stage I

Stage II

Latent feature

Z

ID Routing

Router

$F_{id}^{(1)}$

$F_{id}^{(2)}$

$F_{id}^{(3)}$

Cross Attention

Cross Attention

Cross Attention

Routing Regularization loss

LoRA

Diffusion loss

CLIP Text Encoder

Cross Attention

Training

- Single-ID training stage: We only introduce the ID embedding module; the training regimen is limited to images that feature a singular ID. Only the parameters within the ID embedding module and the U-Net's LoRA are subjected to training in this stage. The training loss is the original diffusion loss

- Multi-ID fine-tuning stage: We introduce the ID routing module. We fix all the parameters in the ID embedding module and only fine-tune the parameters of the ID router and LoRA module. The loss function encompasses the original diffusion loss and the routing regularization loss
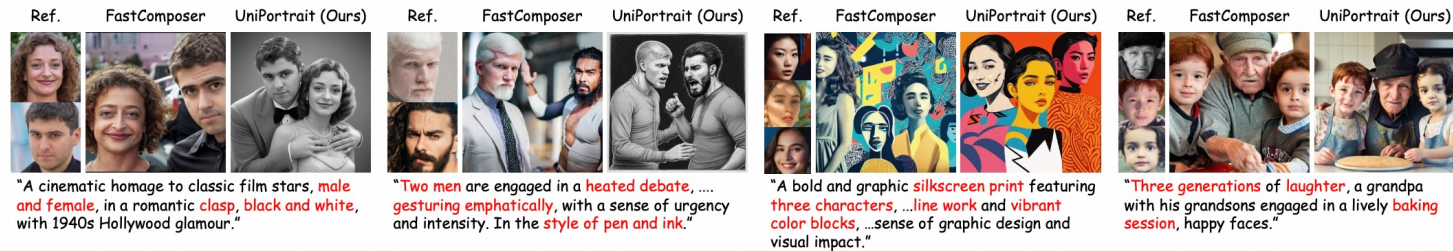
Experiments

Method	Arch.	Cond. Type	Multi-ID?	Face Sim. ↑ (%)	CLIP-T ↑ (%)	FID ↓	LAION-Aes ↑
SD v1-5 [37]	SD15	-	-	4.4	27.7	-	6.60
PortraitBooth[†] [35]	SD15	Face Img	✗	65.7	24.5	-	-
IP-Adapter[‡] [51]	SD15	Face Img	✓✗	68.4	24.7	139.5	6.43
FastComposer [49]	SD15	Face Img	✓	50.8	24.1	134.5	6.17
UniPortrait (ours)	SD15	Face Img	✓	71.1	26.1	**123.4**	6.42
PhotoMaker [27]	SDXL	Face Img	✗	41.7	**26.9**	136.1	6.01
PuLID [10]	SDXL	Face Img	✗	59.0	25.8	128.4	**6.57**
InstantID* [45]	SDXL	Face Img& Landmark	✗	**77.9**	24.1	163.5	6.15

Single-ID personalization

Method	Arch.	Face Sim. ↑ (%)	CLIP-T ↑ (%)	FID ↓	LAION-Aes ↑	BS ↓ (%)
SD v1-5 [37]	SD15	1.6	28.9	-	6.25	20.3
Custom Diffusion [21]	SD15	10.4	25.9	157.1	5.82	45.8
W+ adapter [26]	SD15	8.0	24.5	138.6	5.73	49.2
IP-Adapter[‡] [51]	SD15	7.7	25.4	153.2	5.85	52.7
IP-Adapter[‡] (w/ Router)	SD15	60.5	26.6	143.7	5.86	7.9
FastComposer [49]	SD15	38.0	25.5	156.6	5.81	12.9
UniPortrait (ours)	SD15	**67.3**	**27.4**	**139.5**	5.89	**7.6**
MuDI [18]	SDXL	44.2	26.9	159.2	5.87	36.8
PuLID [10]	SDXL	28.1	27.0	140.2	**6.16**	55.4
StoryMaker [55]	SDXL	57.4	24.1	180.4	5.44	31.4

Multi-ID personalization

Visualizations

	Reference	FastComposer	PhotoMaker	InstantID	FlashFace	IP-Adapter-FaceID-PlusV2	UniPortrait (Ours)

"A woman eating a healthy parfait and yogurt, cartoon."

"A man clad in cyberpunk fashion: neon accents, reflective sunglasses, and a leather jacket with glowing circuit patterns. He stands stoically amidst a soaked cityscape."

"An 8-bit pixel art representation of a skateboarder, constructed from carefully arranged blocks of vibrant pixels. The background is snowy mountains and a blue sky."

Reference   UniPortrait (Ours)

"A couple eating a candlelit dinner at a romantic restaurant."

"A photo of three friends laughing."

"A winter wonderland, four Popmart blind boxes, cartoon."

Ref.   FastComposer   UniPortrait (Ours)

"A cinematic homage to classic film stars, male and female, in a romantic clasp, black and white, with 1940s Hollywood glamour."

"Two men are engaged in a heated debate, ... gesturing emphatically, with a sense of urgency and intensity. In the style of pen and ink."

"A bold and graphic silkscreen print featuring three characters, ...line work and vibrant color blocks, ...sense of graphic design and visual impact."

"Three generations of laughter, a grandpa with his grandsons engaged in a lively baking session, happy faces."

Experiments

Reference "happy" "sad" "angry" "surprised" "confused" "disgusted"

(a) Face editing, with text prompts

"Blue eyes, red lips, blush." "A child." "A man."
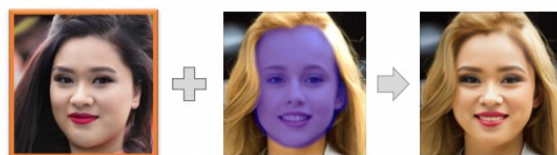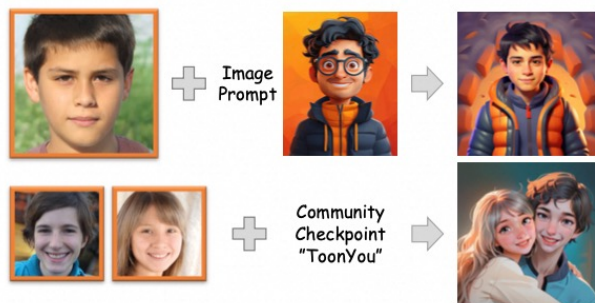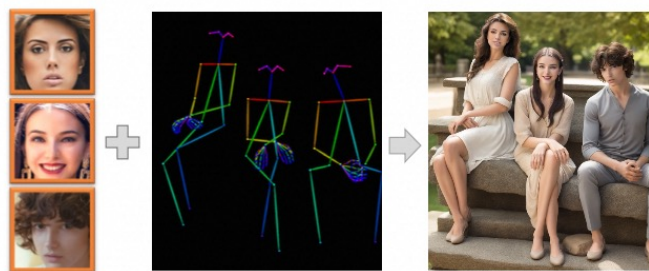
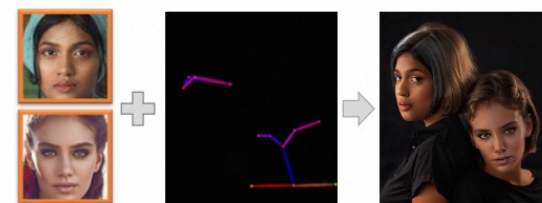(b) Face swap, with ControlNet-Inpainting

(c) ID-preserving style transfer, with IP-Adapter and ControlNet-Edge

Image Prompt

(d) Stylized generation, with IP-Adapter or Community Checkpoint

Image Prompt

Community Checkpoint "ToonYou"

(e) Pose Control, with ControlNet-Pose

"Best quality, masterpiece, two women and a man sitting on a stone bench."

"The woman on the left has short brown hair and is wearing a black shirt. The woman on the right has blonde hair styled back and is also dressed in a black shirt. Her head is resting on the shoulder of the other woman. "

Conclusion

Codes and pretrained models are available at

- https://github.com/junjiehe96/UniPortrait