CrossID: Cross-Supervised Spatio-Temporal Gated Fusion for Personalized Portrait Generation

Dongxu Yue Qixin Yan* Shiao Yang Xiaoqiang Zhou Hao Liu Zhihai He† Chun Yuan†

* Project leader † Corresponding authors

Tsinghua University

Tencent 腾讯

NEURAL INFORMATION PROCESSING SYSTEMS

NeurIPS 2026

Overview

Preserve who the person is, while changing how they appear. CrossID combines **cross-supervised training** with a **GatedFaceAdapter** to retain facial details and reduce copy-paste artifacts under diverse text prompts.

11M	0.9M	2,000
annotated portraits	paired I2V clips	test cases / method



Single- and multi-person generation with flexible control.

Motivation

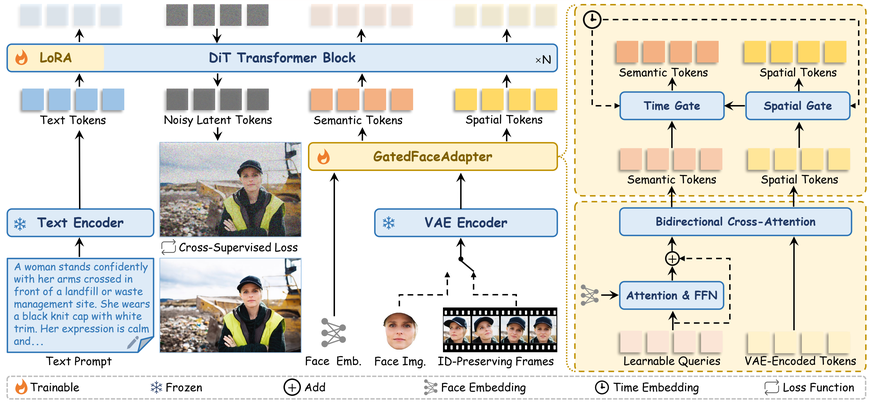
Self-pair training couples identity to the target appearance. A face crop from the target makes copying an easy solution, limiting edits to pose, expression and gaze.



Our cross-pairs keep identity but vary pose and expression, so the model must learn more than a spatial copy.

$$\mathrm{ID}(v) = \mathrm{ID}(x_0), \quad v \neq x_{\text{crop}}$$

Method: CrossID



Dual-branch gated fusion

ArcFace semantic tokens and VAE detail tokens interact through **bidirectional cross-attention**. Time and spatial gates regulate their contribution.

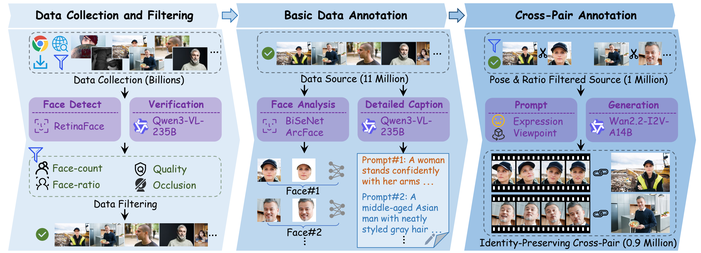
Cross-supervised training

Stage 1: learn facial reconstruction with self-pairs. **Stage 2:** freeze the adapter and train LoRA using I2V frames as conditions and real photos as targets.

Results



CrossID-11M



11M high-quality portraits with detailed annotations. **0.9M** are paired with identity-preserving I2V clips. Video frames provide diverse conditions, while real photographs remain the training targets.

CrossID-Bench

100 identities × 20 prompts. Higher is better for all metrics.

Method	Face Sim	CLIP-T	NI	SC	IQ
IP-Adapter	0.5021	0.2014	3.109	2.803	2.955
FastComposer	0.5580	0.2061	3.070	2.475	1.860
PhotoMaker	0.4723	0.2381	4.678	4.295	4.235
InstantID	0.6246	0.1998	4.347	3.811	2.708
PuLID-FLUX	0.6274	0.2422	4.287	4.094	4.001
InfiniteYou	0.5769	0.2496	4.609	4.060	3.587
DreamO	0.6253	0.2381	3.542	3.865	3.450
XVerse	0.5685	0.2318	3.792	4.068	4.021
WithAnyone	0.5646	0.2309	**4.827**	4.305	4.104
CrossID (Ours)	**0.6291**	**0.2508**	4.733	**4.362**	**4.316**

Bold: best. Underlined: second best. NI / SC / IQ are scored 1–5.
NI: natural integration · SC: semantic correctness · IQ: image quality