# IdentityStory: Taming Your Identity-Preserving Generator for Human-Centric Story Generation

Donghao Zhou[1]*, Jingyu Lin[2]*, Guibao Shen[3], Quande Liu[4], Jialin Gao[1], Lihao Liu[5], Lan Du[2],
Cunjian Chen[2†], Chi-Wing Fu[1], Xiaowei Hu[6†], Pheng-Ann Heng[1]

[1]CUHK, [2]Monash, [3]HKUST (GZ), [4]Kling, [5]Amazon, [6]SCUT

Page   Code

Task and Motivations

We study **human-centric story generation**, aiming to enable consistent generation of human characters **with only text as input**.

Character: "middle-aged Middle Eastern man with short, black hair and olive skin"

Recently, identity-preserving generators have shown great capabilities in subject-driven generation. Therefore, **can we tame identity-preserving generators to achieve human-centric story generation**?

Key Observations

Averaging   Ours

Face-Sim 49.5%   Face-Sim 54.0%

(a)   (b)
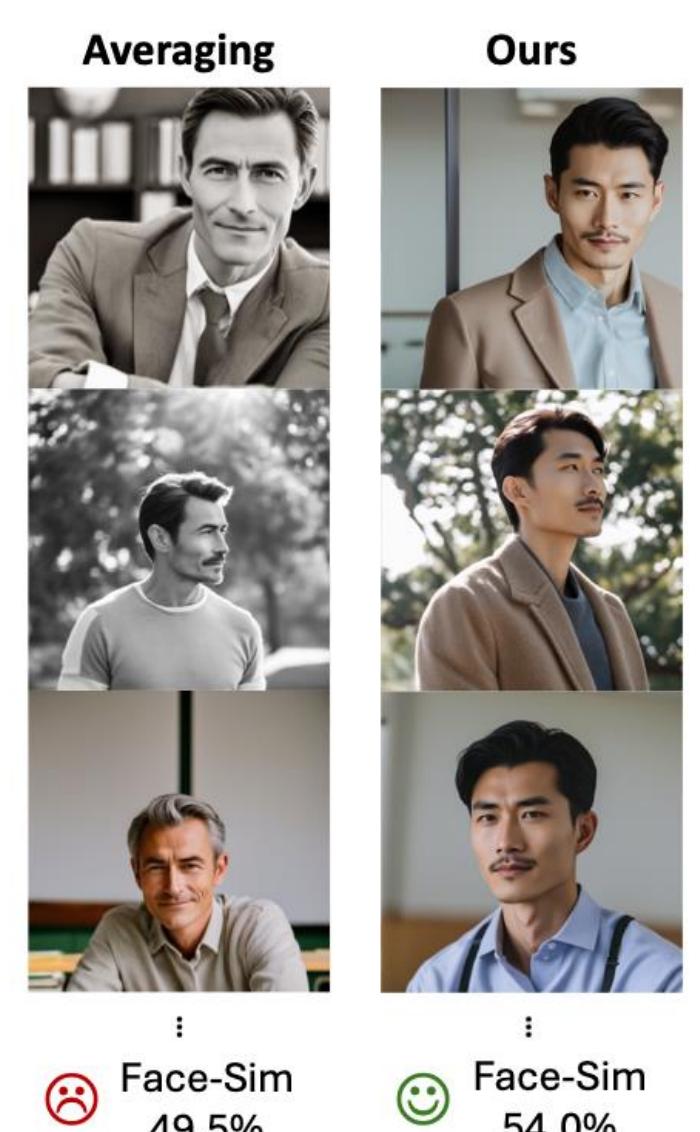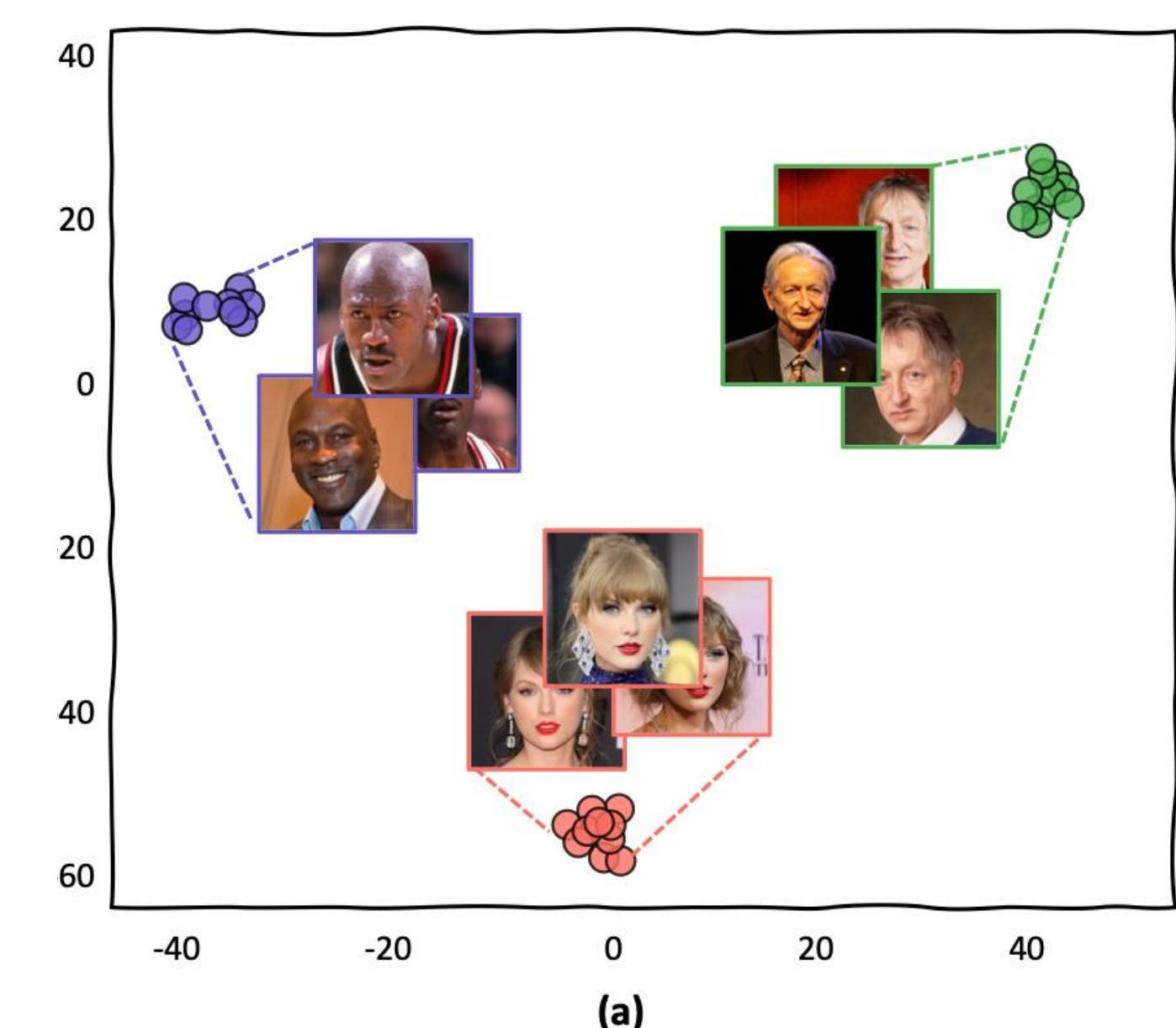
Identity-preserving generators **possess a well-constructed identity space**!

Vanilla SDXL   PhotoMaker   InstantID

"A surreal digital artwork of a young Black girl with curly hair and bright eyes, standing on a rainbow bridge high above the clouds"

"A hyper-realistic illustration of a middle-aged Latina woman with dark, curly hair and strong cheekbones, wearing an apron in a cozy home kitchen"

(a)   (b)
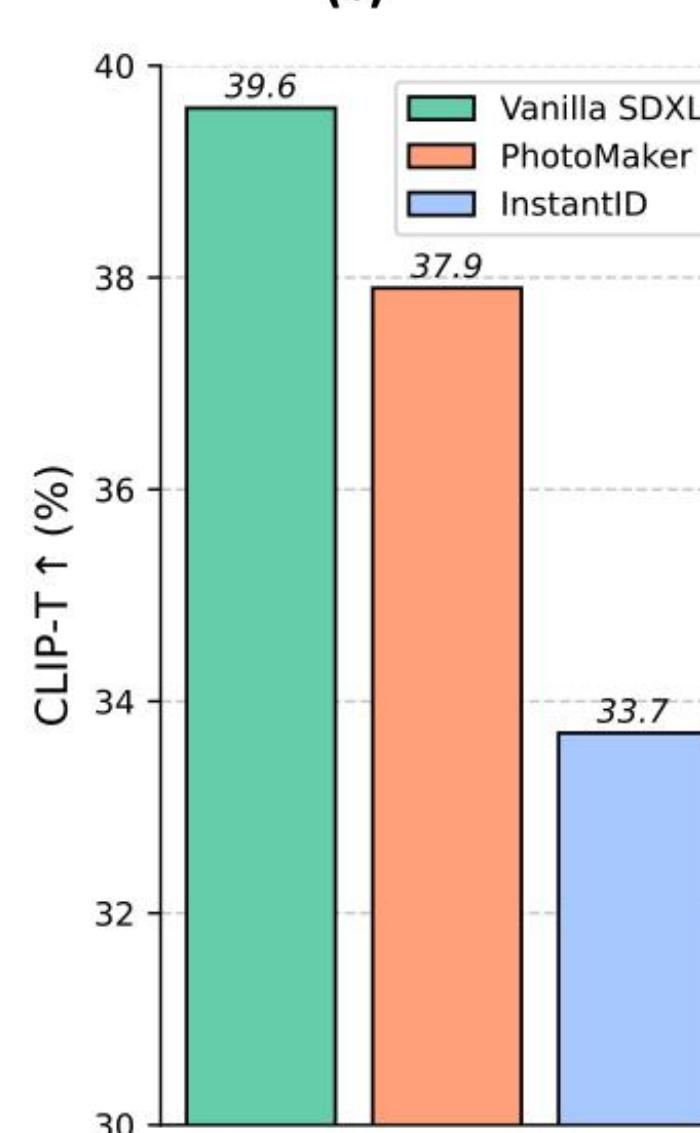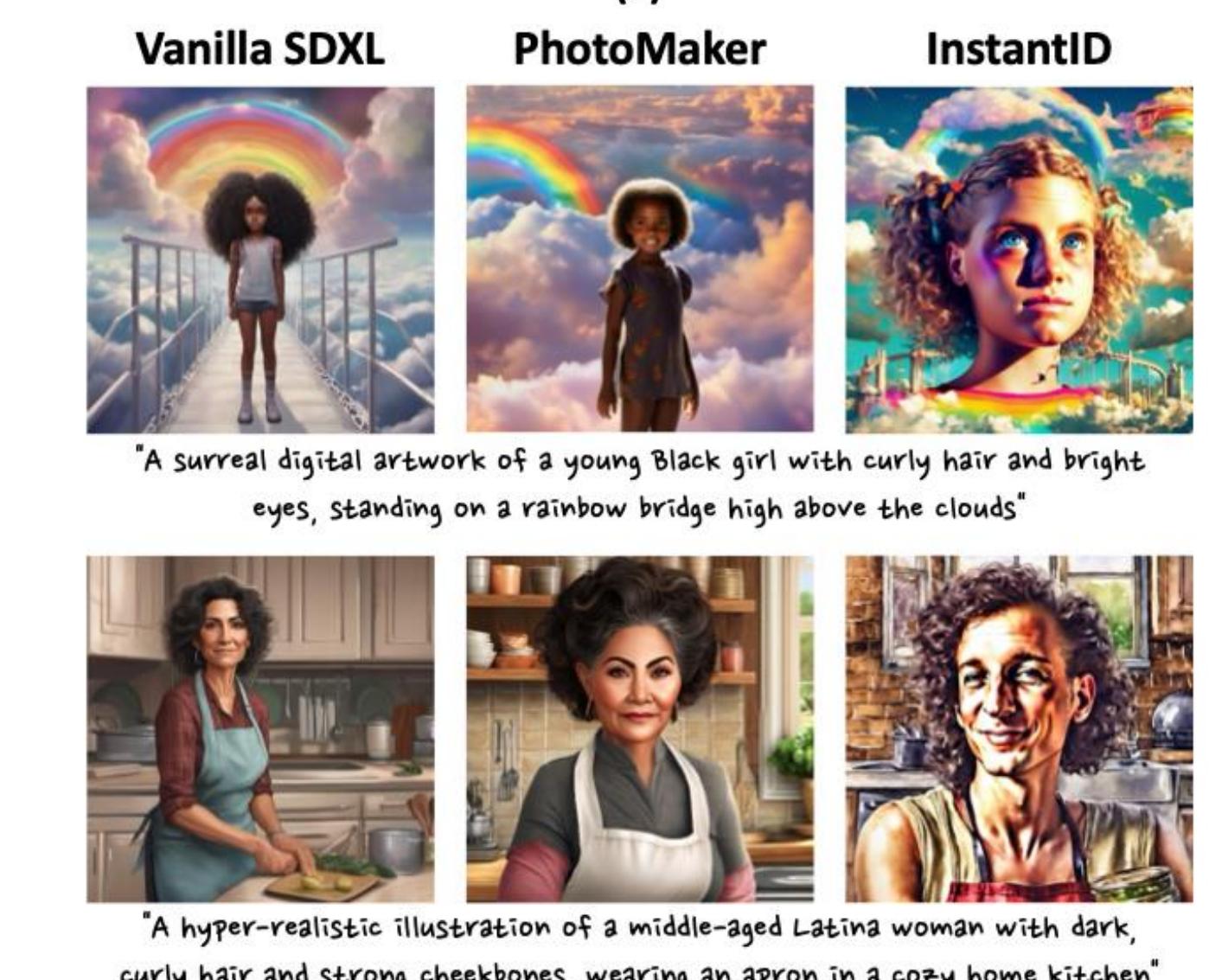
Identity-preserving generators **struggle with text alignment**!

Methodology

We propose **IdentityStory**, an innovative framework consisting of **Iterative Identity Discovery** and **Re-denoising Identity Injection**.


(a) Iterative Identity Discovery

(b) Re-denoising Identity Injection
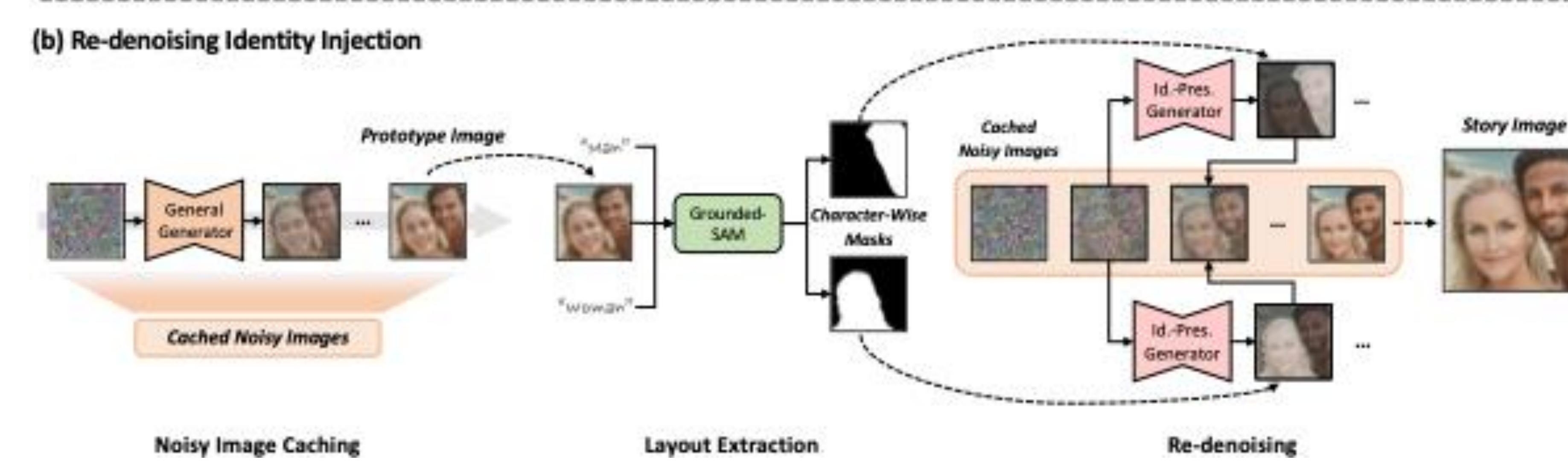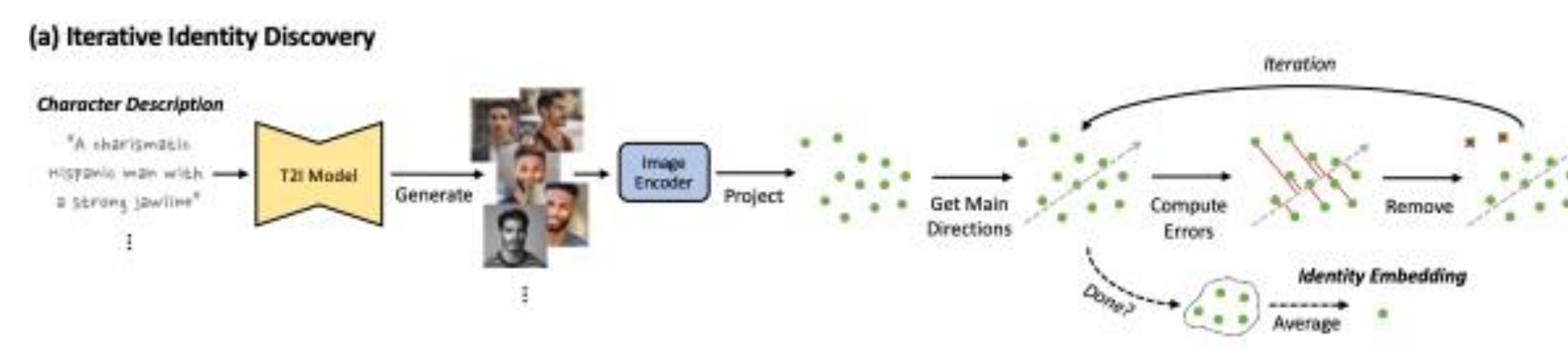
Two key design choices of Re-denoising Identity Injection:
**Sweet-Spot Timestep** (left) and **Progressive Masking Strategy** (right)

Timestep = 50   Timestep = 0

Injected Identity

w/o Progressive Masking Strategy   w/ Progressive Masking Strategy

Experiments

Qualitative Comparison

Methods	Venues	Text Alignment		Character Consistency		Image Quality	
		CLIP-T↑ (%)	CLIP-T-C↑ (%)	CLIP-I↑ (%)	Face-Sim↑ (%)	Q-Align-Gen↑	Q-Align-Aes↑
ConsiStory [51]	SIGGRAPH'24	**35.5**	30.1	78.2	17.1	3.75	4.71
StoryDiffusion [69]	NeurIPS'24	34.0	30.7	85.2	27.1	3.58	4.20
Story-Adapter [35]	arXiv'24	34.3	29.1	76.6	23.9	3.65	4.42
1Prompt1Story [31]	ICLR'25	34.9	29.7	79.8	23.5	4.16	4.81
IdentityStory (Ours)	-	35.4	31.1	85.8	55.5	4.25	4.92

Quantitative Comparison

Methods	Text Align.↑ (%)	Char. Consis.↑ (%)	Img.Qual.↑ (%)
ConsiStory [51]	6.5	6.5	8.5
StoryDiffusion [69]	10.3	11.0	14.3
Story-Adapter [35]	5.0	5.2	3.8
1Prompt1Story [31]	9.0	10.8	8.7
IdentityStory (Ours)	69.2	66.5	64.7

Methods	CLIP-T↑ (%)	CLIP-I↑ (%)	Face-Sim↑ (%)	Q-Align-Gen↑
PhotoMaker [29]	34.6	84.1	47.9	3.84
IdentityStory (Ours)	**35.4**	**85.8**	**55.5**	**4.25**
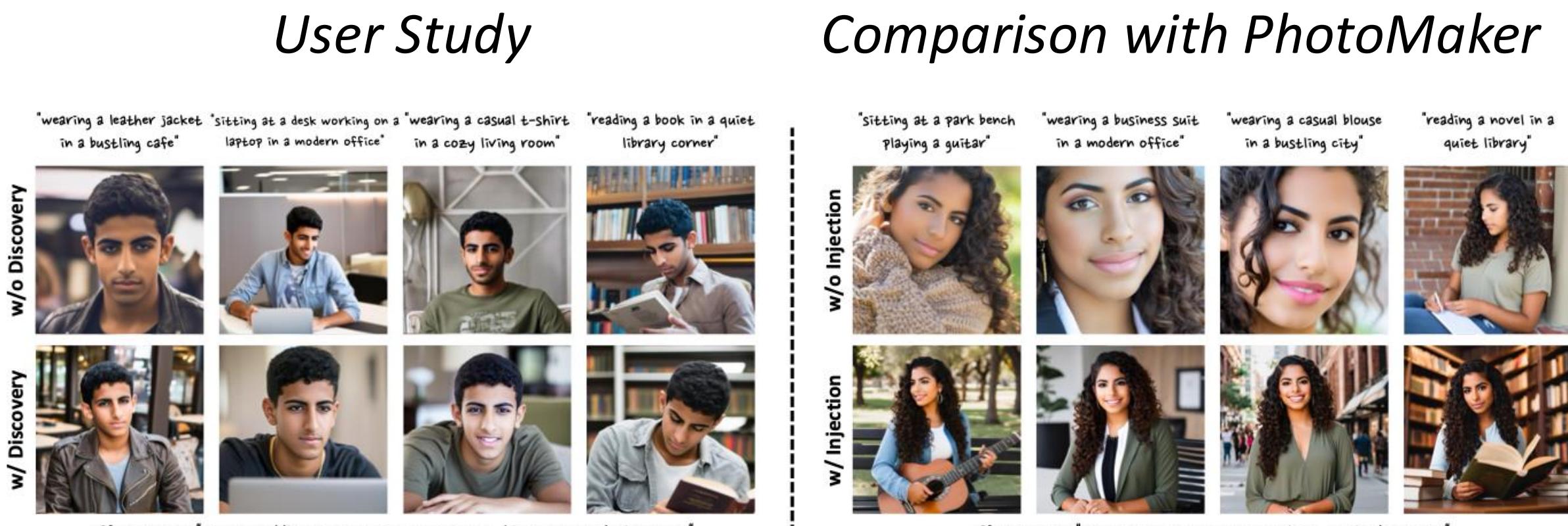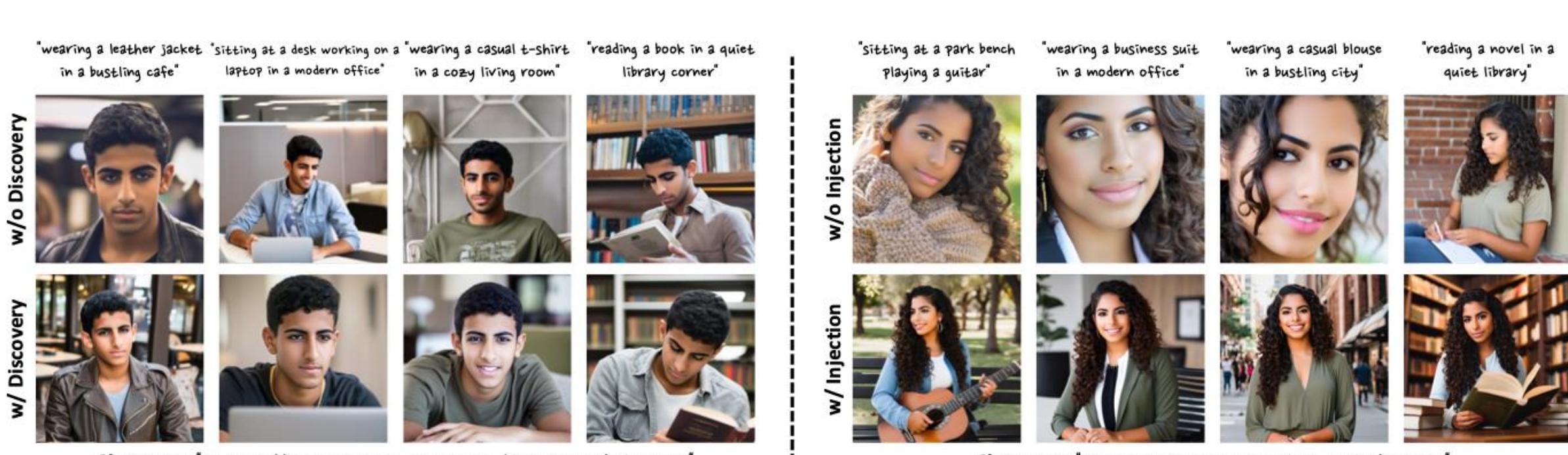
User Study   *Comparison with PhotoMaker*

w/o Discovery   w/o Injection

w/ Discovery   w/ Injection

Character: "young Middle Eastern man with short, black hair and olive skin"   Character: "young Latina woman with long, curly dark hair"

Ablation Study

ControlNet   RealVisXL

AnimagineXL

(a)   (b)
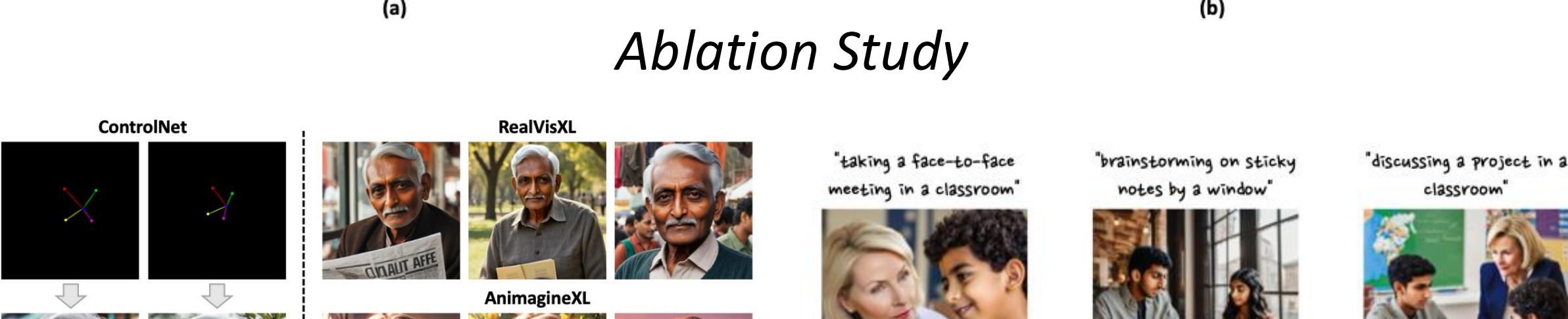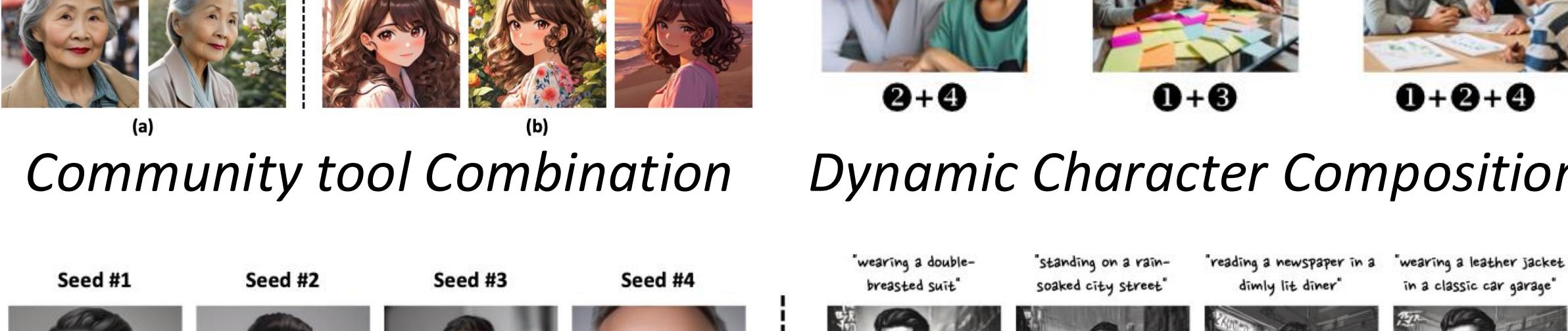
Community tool Combination   *Dynamic Character Composition*

Seed #1   Seed #2   Seed #3   Seed #4

"a professional man"

"a cute boy"

"a gentle man"

"a beautiful woman"
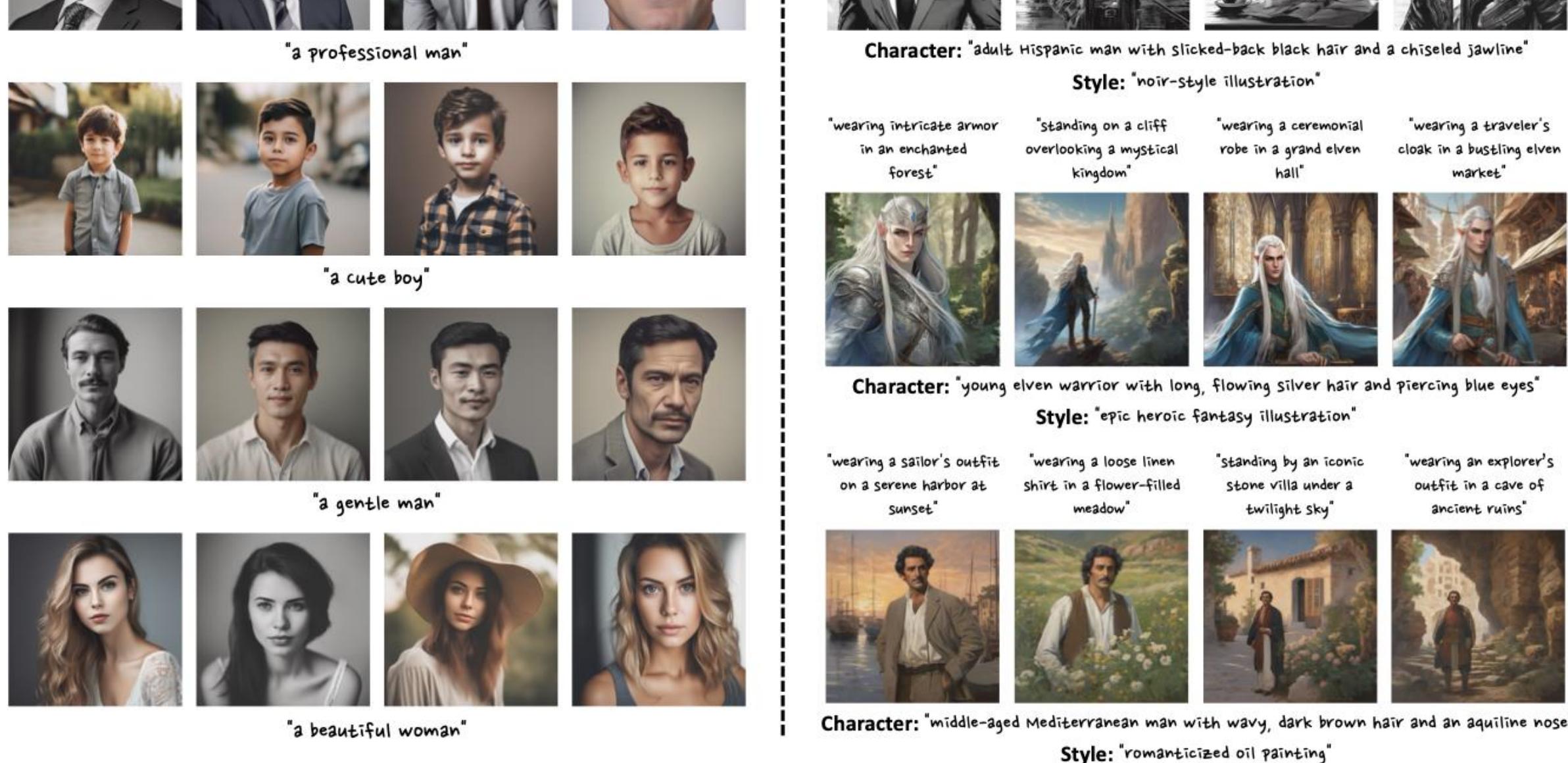
Seed Variation & Stylized Story Generation