Appendix

A. Face Structure Scale

As illustrated in Table 4 of the main paper, the introduction of the face structure branch has indeed enhanced ID similarity. However, it has also compromised some aspects of prompt consistency. In practical implementation, a scaling factor β (see Fig. I(a)) can be incorporated to modulate the intensity of the face structure branch, thereby enabling a flexible trade-off between face similarity and prompt consistency. Fig. II provides a visualization of the impact of different β values on the generated outcomes. As expected, a larger β value leads to increased facial shape and texture fidelity (attributed to the detailed spatial information in the face structure branch), but also results in decreased facial editability (*e.g.*, expression editing, due to the insufficient disentanglement of face structure features). Empirically, a lower β value (*e.g.*, 0.1) can be selected to achieve a balance between prompt consistency and ID alignment.
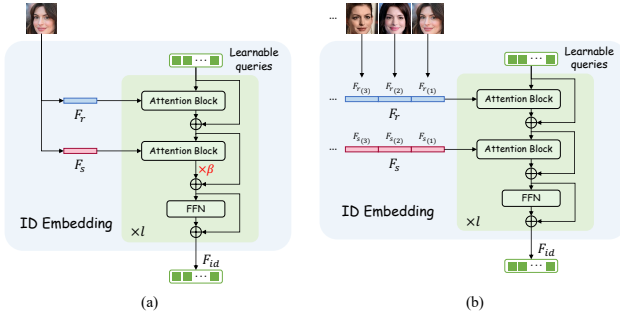

Figure I. **Illustrations of (a) face structure scale β and (b) single-ID multi-reference embedding in the inference phrase.**

Number of reference images	Face Sim. ↑ (%)	CLIP-T ↑ (%)	FID ↓	LAION-Aes ↑
1	55.5	**26.1**	123.4	6.42
2	58.5	26.1	**122.8**	6.41
4	**61.2**	25.9	124.7	**6.44**

Table I. **Performance of using multi-reference images.**

B. Single-ID Multi-Reference Embedding

Owing to the scalability of the designed architecture, our ID embedding module can be seamlessly adapted to cater to the context of embedding multiple references for a single ID. We only need to extract the intrinsic ID features $F_{r(j)}$ and face structure features $F_{s(j)}$ for each reference (j is the reference index) and concatenate them as the new intrinsic ID and face structure features, as shown in Fig. I(b).

Table I provides a quantitative assessment of the advantages of using multi-reference images. Here, we continue to use the 15 IDs described in Sec. 4.2 for evaluation, but instead of using the same reference image to compute face similarity, we use a different image of the same ID as the target. The findings demonstrate that multi-reference images can improve ID similarity without compromising prompt consistency, FID, and aesthetic scores.

C. Visualization of Multi-ID Routing Map

Fig. IV and Fig. V display the routing map of each cross-attention layer of the U-Net under different diffusion steps. It is observed that, with the utilization of routing regularization loss, the ID router can discern different IDs at earlier time steps and in a more pronounced manner, particularly within the cross-attention layers of the U-Net's decoder.
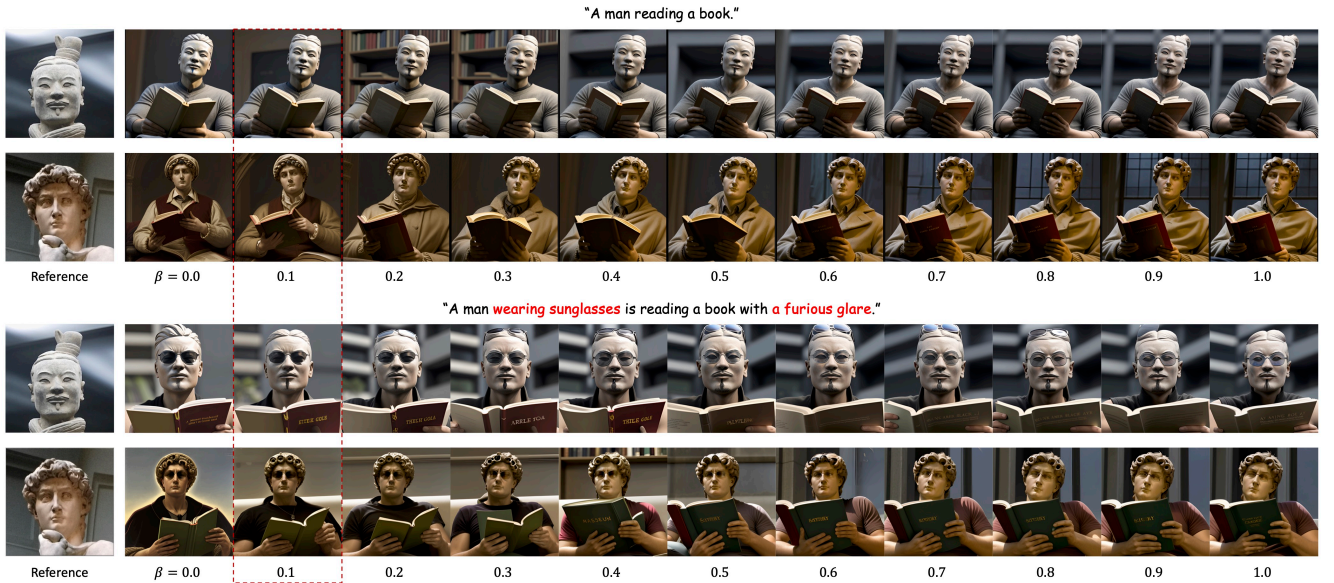

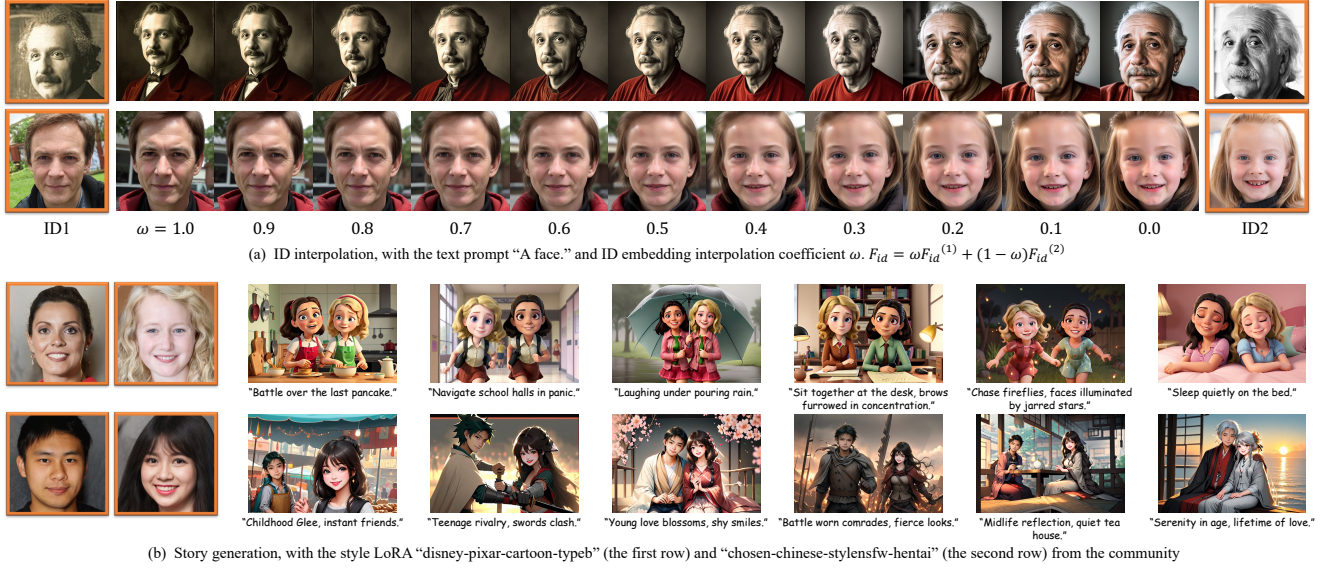Figure II. **The effect of the face structure scale.**

(a) ID interpolation, with the text prompt "A face." and ID embedding interpolation coefficient ω. $F_{id} = \omega F_{id}^{(1)} + (1 - \omega)F_{id}^{(2)}$

(b) Story generation, with the style LoRA "disney-pixar-cartoon-typeb" (the first row) and "chosen-chinese-stylensfw-hentai" (the second row) from the community

Figure III. **More applications of UniPortrait.**

Model / LoRA	URL
Realistic Vision V4.0	https://huggingface.co/SG161222/Realistic_Vision_V4.0_noVAE
ToonYou	https://civitai.com/models/30240/toonyou
disney-pixar-cartoon-typeb	https://civitai.com/models/75650/disney-pixar-cartoon-typeb
chosen-chinese-stylensfw-hentai	https://civitai.com/models/95643/chosen-chinese-stylensfw-hentai

Table II. **URLs of the used models and LoRAs in this paper.**

	IP-Adapter	Fast-Composer	Uni-Portrait	Photo-Maker	Instant-ID	Fast-Composer	Uni-Portrait
Face Sim. ↑ (%)	72.0	62.6	<u>78.6</u>	54.9	**83.6**	49.7	**77.3**
CLIP-T ↑ (%)	24.2	24.2	<u>25.8</u>	**26.1**	23.5	25.9	**26.3**
LAION-Aes ↑	<u>6.11</u>	6.02	**6.26**	5.86	5.95	5.74	**5.86**

Table III. **Quantitative results on FFHQ.** The last two columns display results for multi-ID personalizations, while the remaining columns show those for single-ID personalizations.

Number of IDs	0	1	2	3	4
Speed (s) (512×512)	1.780	1.940	2.179	2.323	2.386

Table IV. **Generation speed.** We test the model generation speed (including ID embedder and router) on a V100 card with a 20-step DDIM sampling.

D. More Applications

In Fig. III(a), we investigate the UniPortrait's capacity for identity interpolation among different characters. We extract the ID embedding $F_{id}^{(n)}$ for each identity (n is the identity index) and perform linear interpolation on them to achieve ID interpolation. The results demonstrate that UniPortrait can create a meaningful semantic transition between different IDs, or even different states of the same ID.

In Fig. III(b), we demonstrate the potential of UniPortrait to generate stories with consistent IDs.

E. More Example Generations

Fig. VI-IX present more personalized human image results of UniPortrait, demonstrating the outstanding performance of our approach in single- and multi-ID customization again.

F. Open-Sourced Text-to-Image Models

Table II shows the URLs of the utilized open-sourced text-to-image models and LoRAs in this paper.

G. Quantitative Results on FFHQ

We replace 15 CelebA test IDs in our test set with 15 randomly selected FFHQ IDs and report the results in Table III. The conclusions on FFHQ are consistent with those on CelebA.

H. Generation Speed

Table IV presents our model's generation speed under different numbers of input IDs. Each additional ID adds 16 tokens of cross-attention computation to the original diffusion model. Overall, the additional computational load for our ID personalization method is not significant.

I. Example Face Image Source

Most of the example face images used in this paper come from the Pexels, Unsplash, Pixabay, and Wikipedia websites. We thank the owners of these images for sharing their

Figure IV. **Visualization of the routing map within each cross-attention layer of the U-Net without the routing regularization loss.**

Figure V. **Visualization of the routing map within each cross-attention layer of the U-Net with the routing regularization loss.**

valuable assets. We also thank the StyleGAN2 authors for sharing their high-quality synthesized face images, which constitute another important part of the example ID images we used.

J. Ethics Considerations

The utilization of human image personalization technologies presents potential social risks, with the misuse of such technologies for deepfakes being a significant concern. In order to address these risks, the implementation of ethical guidelines and responsible usage practices is imperative. At present, the synthesized outcomes display certain visual anomalies that may facilitate the identification of deepfakes.
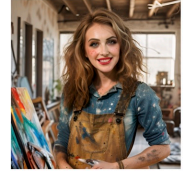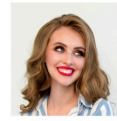
"A cinematic noir image of a detective, sharp contrasts and moody lighting suggesting an untold story."
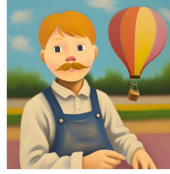
"A whimsical impressionist portrait of a balloon seller, painted with soft, hazy brushstrokes that capture the essence of a breezy spring day."
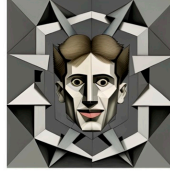
"A woman wearing a pastel-colored dress against a blank, white wall."

"Finely crafted image of a bohemian artist with a wild look of inspiration, untamed auburn curls, paint-splattered overalls, brush in hand, surrounded by chaotic canvases in a sunlit loft."

"Popmart Blind Box, cartoon."

"A man, Cubist face, Picasso-style, trapped within a labyrinth of sharp angles and intersecting planes."

"Felt style crafting a portrait of a women, with a soft woolen texture and fibers that lend a cozy, tactile sensation to the image."

"A serene elven princess with delicate features and long, flowing silver hair, her hands gently resting on a harp, a serene expression, wearing a crystal-embellished gown, in a moonlit forest clearing."

"A portrait of a space pirate with a Laser Light eye-patch, cutting through the galactic darkness."
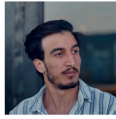
"A comic book style mage with vibrant energy crackling from his fingertips, detailed with bold outlines and strong shadows."
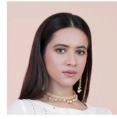
"An elegant portrait of a celestial mage, with flowing silver hair and intricate runic tattoos that glow with a soft luminescence."

"A traditional woodblock style image of a tea master, with clean-lines and deliberate simplicity honoring Japanese ukiyo-e."

"A dwarven man in Pop Art style, with a pattern of bright, repeating motifs and bold lines capturing his industrious spirit."

"A gritty street art mural depicting a revolutionary, with expressive, aerosol-sprayed lines and poignant urban symbolism."
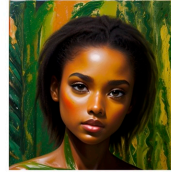
"Acrylic texture portrait of a jungle goddess, with thick, impasto brushstrokes that give the image dimension and vibrant life."

"A woman sitting in front of a vanity, with a large makeup assortment in front of her."

"A girl wearing blue clothes. Wearing white roses, jewel-like eyes, blue light, acrylic, super dreamy, silver-white hair, watercolor rendering, unreal rendering, background is watercolor embellishment, blue and white."
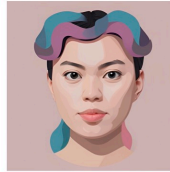
"Vector illustration of a portrait in Flat Style, pastel color palette."
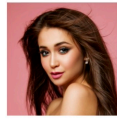
"Art Nouveau portrait in the style of Mucha, encompassing a nymph with flowing hair, entwined with a tapestry of ornate botanical and astronomical motifs."

"Human head plaster sculpture, Vaporwave, pink background, studio lighting."

"A delicate artwork rendered in colored pencils portrays a young girl, her eyes shimmering with innocence and a hint of untold stories."

"An older woman knitting a blanket in a rocking chair."

"Portrait capturing the Tyndall effect, where a celestial being's translucence reveals beams of light scattering through her figure, against a backdrop of deep blues and soft whites."
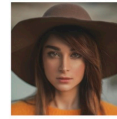
"Golden flames and indigo smoke envelope a hooded assassin, a visual metaphor for combustion chemistry, digital art illustrating the fierce beauty and transient nature of chemical reactions."

"An enigmatic man with charismatic charm and classic, timeless appeal, reclining in an opulent, vintage car. He exudes sophistication and allure in a finely tailored, dapper suit with classic accessories, emanating timeless refinement amidst opulent luxury."

"A charismatic man with intense features and a confident smile, engaging in a contemplative activity at a serene, remote cabin. He's dressed in cozy, rugged winter wear, equipped with a cozy scarf and stylish hat."

"An alluring woman with a fierce gaze and modern bob hairstyle, posing against a dramatic, industrial urban backdrop. She's wearing a sleek, edgy black ensemble with avant-garde accessories."
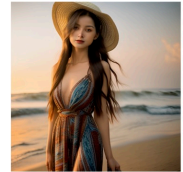
"A striking woman with long flowing hair and captivating eyes, standing on a beach at sunset. She wears a flowing, bohemian-style maxi dress and a wide-brimmed hat, exuding a carefree and ethereal aura."

Figure VI. **Text-to-single-ID personalization examples.**

"A mesmerizing woman with an enigmatic beauty and a sleek, modern hairstyle, posing on a grand stairway in a lavish, historic mansion. She's adorned with a breathtaking couture gown and sparkling, heirloom jewelry, exuding grace and sophistication against the backdrop of opulent luxury."

"A man with a rugged beard and a leather jacket, riding a vintage motorcycle with a determined expression, against a backdrop of a desert highway with a setting sun."

"A little girl with pig-tails and a backpack, smiling and skipping while holding a bunch of balloons, in a bustling city street with colorful buildings and street vendors."

"A middle-aged man with a bald head and glasses, sitting at a desk in a cozy home office, surrounded by books and with a look of deep concentration as he writes with a fountain pen."

"A woman with a sleek bun and a business suit, walking confidently with a briefcase, in a modern city with skyscrapers and bustling professionals."

"A man with a fedora and a trench coat, smoking a pipe with a thoughtful expression, in a dimly lit jazz club with sultry music and a vintage vibe."

"A girl with a messy bun and overalls, running happily through a field of wildflowers, with a carefree and joyful expression."

"A woman with a flowy bohemian dress and flower crown, playing an acoustic guitar with a dreamy expression, in a lush forest clearing with dappled sunlight."

"A woman with a flowing scarf and ethereal attire, meditating serenely in a tranquil yoga studio with natural light and calming decor."

"A man with slicked-back hair and a tailored suit, striking a confident pose at a modern art gallery with avant-garde paintings and sculptures."

"A black woman with long curly red hair, freckles on her cheeks, and a mischievous smile, standing by a lake with a fishing pole in her hand, wearing a wide-brimmed straw hat, a plaid button-up shirt, and denim shorts."

"A man with a sleek undercut, in a graffiti-covered alleyway, wearing a graphic tee and baggy jeans, with a confident smirk on his face."

"A man with a slicked-back pompadour, strutting down a city street in a sharply tailored suit and designer sunglasses, with an air of confidence and the hustle and bustle of urban life swirling around him."

"A woman with a sleek bob and cat-eye glasses, sipping coffee at a sidewalk cafe, wearing a turtleneck sweater and high-waisted jeans, with a contemplative gaze and a vintage handbag slung over her shoulder."
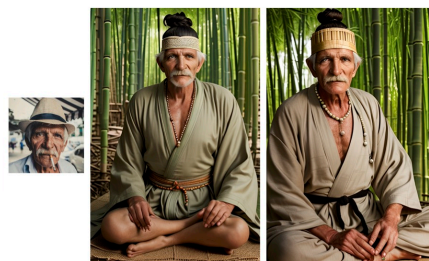
"A man with a topknot and a full beard, meditating in a tranquil bamboo forest, wearing a robe, with a serene expression and the soft rustle of leaves in the background."

Figure VII. **Text-to-single-ID personalization examples.**

"A man and a woman share a laughter-filled dance beneath a canopy of string lights."

"A pair of women lost in conversation sits on a park bench, autumn leaves falling around them."

"A man and a woman locked in a poignant embrace beside a rain-streaked window."

"A comforting hug between a mother and her teenage son, full of unspoken love and support."

"A jovial duo of men cooking together, surrounded by spices and culinary chaos."

"Claymation Whimsy: Two clay figures, lovingly crafted and animated, stroll through a miniature world of vibrant flowers and quirky creatures, capturing the playful nature of their bond."

"Two people delicately folding paper cranes together, their faces displaying concentration and determination as they focus on the intricate folds, with a hint of satisfaction as they complete each step."

"Pixel Art: Generate a retro-inspired pixel art composition portraying two men, using blocky, pixelated graphics and limited color palettes to evoke a nostalgic and vintage aesthetic."
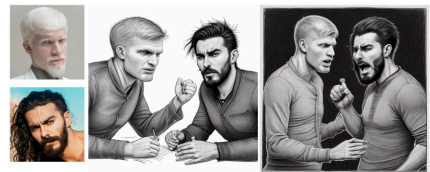
"Two men are engaged in a heated debate, with animated and passionate expressions on their faces. They are gesturing emphatically, with a sense of urgency and intensity. In the style of pen and ink."

"A comic strip depicting two people playfully teasing each other, with mischievous grins and animated gestures. In the style of comic art."

"A minimalist illustration of two people sitting side by side, lost in thought with serene expressions on their faces. In the style of minimalism."

"A pointillism artwork of two people standing in a crowded city street, with expressions of wonder and awe as they take in the sights and sounds. In the style of pointillism."

"A watercolor painting of two people sitting on a park bench, deep in conversation with a look of concentration on their faces. In the style of watercolor."

"A line drawing of two people sharing a warm embrace, with big smiles on their faces and their arms wrapped tightly around each other. In the style of pen and ink."

"An impressionist painting of two people enjoying a leisurely picnic, with relaxed expressions and dappled sunlight on their faces. In the style of impressionism."

"A textile artwork of two women holding hands and strolling through a field of flowers, with joyful expressions and flowing movements. In the style of textile art."

"The characters are dancing in a crowded ballroom, with the woman in a flowing gown and the man in a sharp suit. Their expressions show joy and excitement as they move gracefully across the dance floor. In the style of vintage poster art."

"The characters are engaged in a heated argument, with the woman gesturing angrily and the man looking exasperated. Their expressions are intense and their body language is confrontational. In the style of comic book illustration."

"The characters are cooking together in a cozy kitchen, sharing laughs and exchanging loving glances. Their expressions are playful and affectionate as they enjoy each other's company. In the style of food illustration."

"The girls are sitting on a porch swing, watching the sunset and holding hands. They enjoy the tranquil moment. In the style of landscape painting."

"The characters are standing on a stage, holding microphones and singing a duet with passion and emotion. Their expressions are powerful and emotive as they pour their hearts into the music. In the style of concert poster art."
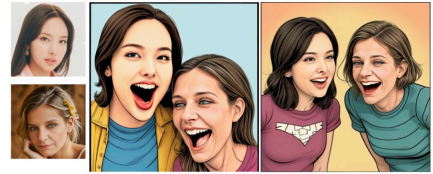
Figure VIII. **Text-to-multi-ID personalization examples.**

"A pair of siblings sitting back to back, with crossed arms and stern expressions, clearly in the midst of a disagreement. In the style of woodblock print."

"Two women engaged in a lighthearted conversation, with animated expressions and infectious laughter. In the style of cartoon illustration."

"A pair of friends laughing uproariously, in the style of a comic book illustration."

"Two people facing each other, gazing into each other's eyes with expressions of deep love and connection, in the style of abstract ink painting."

"A bold and graphic silkscreen print featuring three characters with stylized and vibrant expressions, highlighted by expressive line work and vibrant color blocks, evoking a sense of graphic design and visual impact."

"A vintage-inspired engraving of three woman with dignified and stoic expressions, rendered with meticulous cross-hatching and intricate line work, evoking a sense of time-honored tradition and historical gravity."

"A vibrant and spirited comic book-style illustration of three individuals with animated and exaggerated expressions, featuring larger-than-life emotions and dynamic action that convey a sense of lightheartedness and humor."
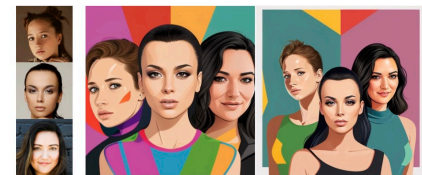
"A modern and vibrant vector illustration of three characters with bold and dynamic expressions, featuring clean lines and vivid colors to convey a sense of energy and liveliness."
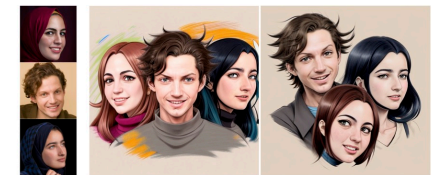
"A lively and expressive anime/manga-style illustration of three characters with exaggerated and emotive expressions, drawn in a bold and stylized manner to convey a sense of youthful energy and emotion."

"A vivid and urban-inspired street art depiction of three male with bold and defiant expressions, featuring graffiti elements and vibrant colors to convey a sense of energy and dynamism."

"A raw and expressive sketch of three individuals with candid and unfiltered expressions, featuring loose and dynamic lines to convey a sense of spontaneity and authenticity."

"A moody and atmospheric film noir scene featuring three characters with intense and brooding expressions, captured in dramatic lighting and shadowy contrasts for a sense of mystery and intrigue."

"A detailed and tactile ceramic portrait of three figures with serene and contemplative expressions, rendered in three-dimensional relief with subtle glazes and textures to convey a sense of depth and tactility."

"Pastel drawing: Three friends enjoying a day at the beach, with one person looking relaxed, one looking cheerful, and one looking contemplative, capturing the essence of a carefree summer day."

"Three man rendered in a meticulous and detailed miniature style, with one person wearing a gentle and warm smile, one person appearing thoughtful and contemplative, and one person exuding a sense of quiet strength and resilience."

"Embroidery portrait: Three women captured in a delicate and intricate embroidery style, with one person wearing a bright and infectious smile, one person exuding a sense of quiet determination, and one person displaying a warm and welcoming expression."

"Three women rendered in a retro pixel art style, with one person wearing a playful and mischievous grin, one person exuding a sense of quiet determination, and one person displaying a warm and friendly expression."

"Three men are depicted with simple, clean lines, their expressions distilled to the essential features, evoking a sense of purity and understated emotion. In the style of minimalist vector art."

"Three men walking through a bustling city street, completely engrossed in conversation. Their expressions are animated and engaged, with exaggerated hand gestures and bright eyes. In the style of colored pen and ink."

"Three people are meditating together, with serene and peaceful expressions on their faces. They are sitting, with their eyes closed and a sense of calm surrounding them. In the style of ink wash painting."

"Two young women wear a leather jacket and a bright floral dress, standing together, one smiling and giving a thumbs up, in an urban alleyway with graffiti-covered walls."

Figure IX. **Text-to-multi-ID personalization examples.**