

ConsistentID

- ConsistentID : Portrait Generation with Multimodal Fine-Grained Identity Preserving
- <https://arxiv.org/abs/2404.16771>
- arxiv

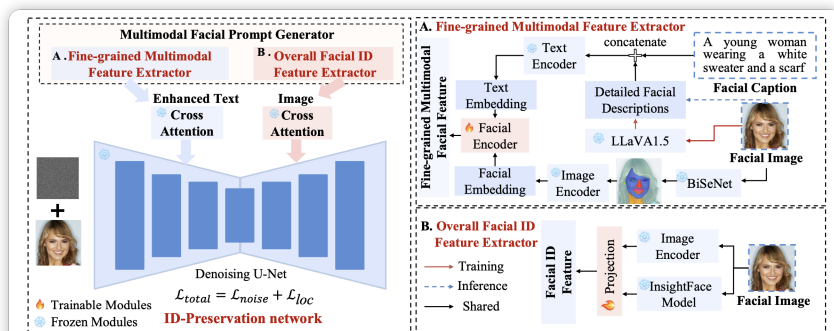


Fig. 2: The overall framework of our proposed ConsistentID. The framework comprises two key modules: a multimodal facial ID generator and a purposefully crafted ID-preservation network. The multimodal facial prompt generator consists of two essential components: a fine-grained multimodal feature extractor, which focuses on capturing detailed facial information, and a facial ID feature extractor dedicated to learning facial ID features. On the other hand, the ID-preservation network utilizes both facial textual and visual prompts, preventing the blending of ID information from different facial regions through the facial attention localization strategy. This approach ensures the preservation of ID consistency in the facial regions.

prompt token eyes mouth nose embedding
embedding prompt IP-Adapter
token mask cross-attention
map

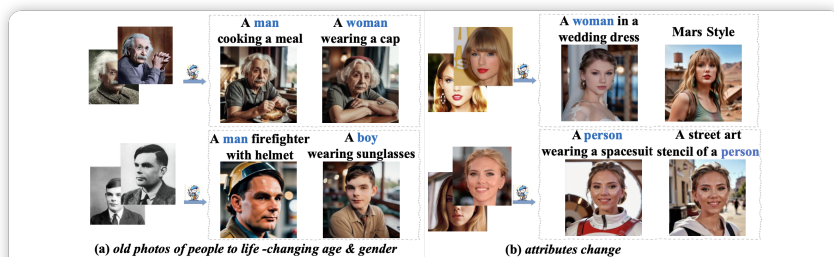


Fig. 1: Given some images of input IDs, our ConsistentID can generate diverse personalized ID images based on text prompts using only a single image.

- CLIP-I;CLIP-T;DINO;
- 3090/bs16
- <https://github.com/JackAllLab/ConsistentID>

Newer

Older

2024-10-23
FlashFace

2024-10-22
CapHuman