



PortraitBooth

home
papers
study
life
About me

- PortraitBooth: A Versatile Portrait Model for Fast Identity-preserved Personalization
- <https://arxiv.org/abs/2312.06354>
- CVPR 2024

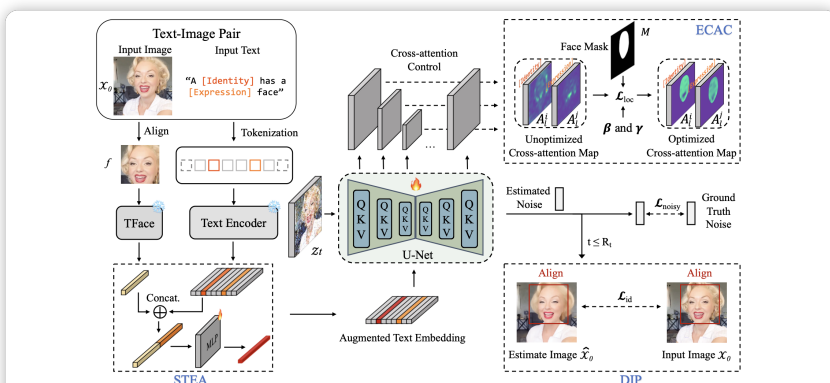


Figure 3. **Overview framework of PortraitBooth.** PortraitBooth extracts the face f from the input image x_0 , and augments the subject's features using TFace for improved identity representation. The diffusion model is trained to generate images with enhanced conditioning, incorporating emotion-aware cross-attention for expression editing and dynamic identity preservation to maintain identity. During the testing phase, we only need to input a single image and the corresponding prompt to achieve rapid, robust identity preservation and diverse expression editing capabilities. A_i^t, A_j^t represents the cross-attention map corresponding to the i -th and j -th token at the l -th cross-attention layer, respectively. β and γ represent the maximum values of the cross-attention map for the identity token and expression token respectively, while R_t indicates the timing for identity preservation.

PortraitBooth is a versatile portrait model for fast identity-preserved personalization. It is trained on a large dataset of human faces and text prompts. The model takes an input image and a text prompt as input and generates a personalized portrait. The model is designed to be fast and efficient, allowing for real-time personalization. The model is trained using a combination of supervised and unsupervised learning. The supervised learning is used to learn the mapping from input images and text prompts to personalized portraits. The unsupervised learning is used to learn the distribution of human faces and text prompts. The model is evaluated on a variety of tasks, including identity preservation, expression editing, and multi-subject generation. The results show that PortraitBooth achieves state-of-the-art performance on these tasks.

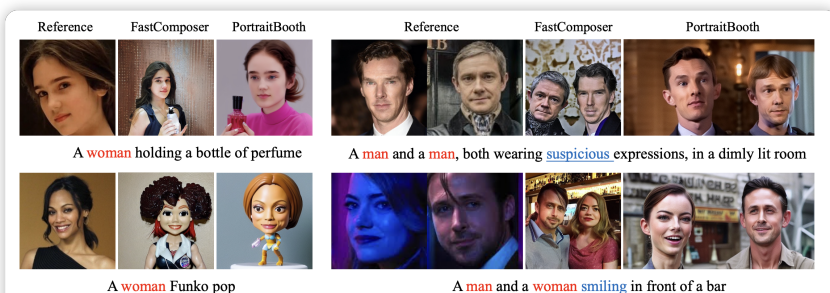


Figure 1. **Qualitative comparison of PortraitBooth and FastComposer** on action, style, expression editing, multi-subject generation, and identity preservation, all without any test-time tuning.

- CelebV-T (reference) RAM Caption
- ID FaceNet CLIP
- 6 V100/bs2
- <https://portraitbooth.github.io>

Newer

Older

2024-10-21	2024-10-19
CAFE	Cones

leicheng © 2022-2025

Archive RSS feed GitHub Email QR Code

Made with Montaigne and bigmission