NEURAL INFORMATION PROCESSING SYSTEMS

阿里巴巴 Alibaba.com

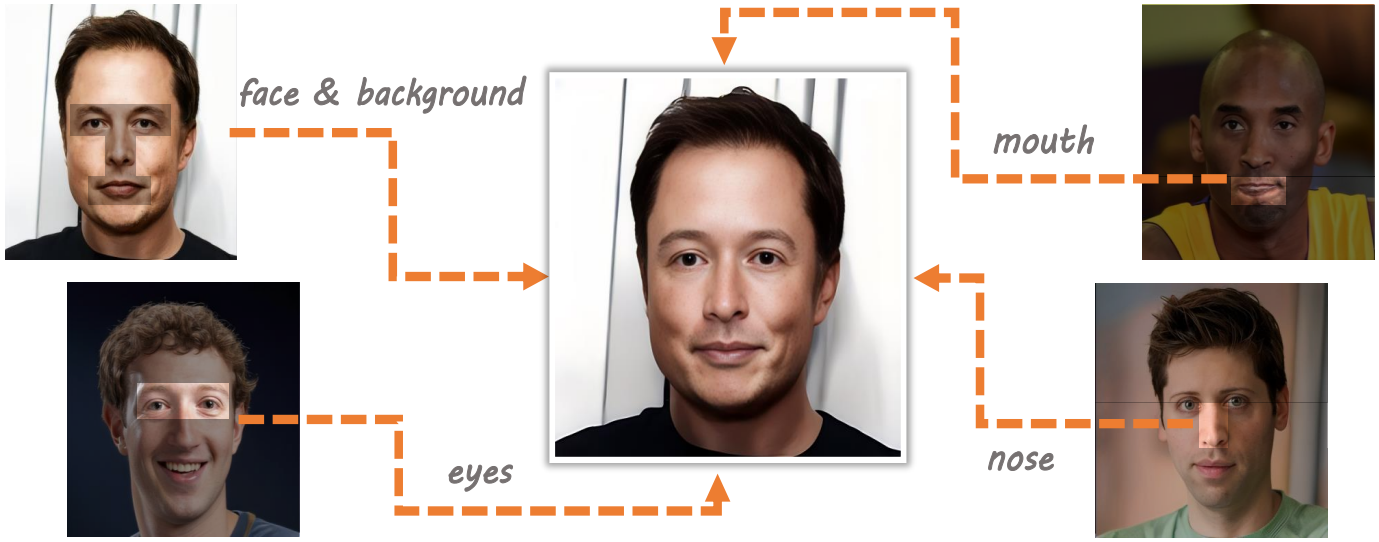
FuseAnyPart: Diffusion-Driven Facial Parts Swapping via Multiple Reference Images

Zheng Yu*, Yaohua Wang*, Siying Cui,
Aixi Zhang, Wei-Long Zheng, Senzhang Wang

Motivation

Facial parts swapping aims to selectively transfer regions of interest from the source image onto the target image while maintaining the rest of the target image unchanged.

Traditional full face swapping lakes the ability to independently swap specific facial features (such as the eyes, nose, and mouth), thereby limiting flexibility in character design and facial modification.

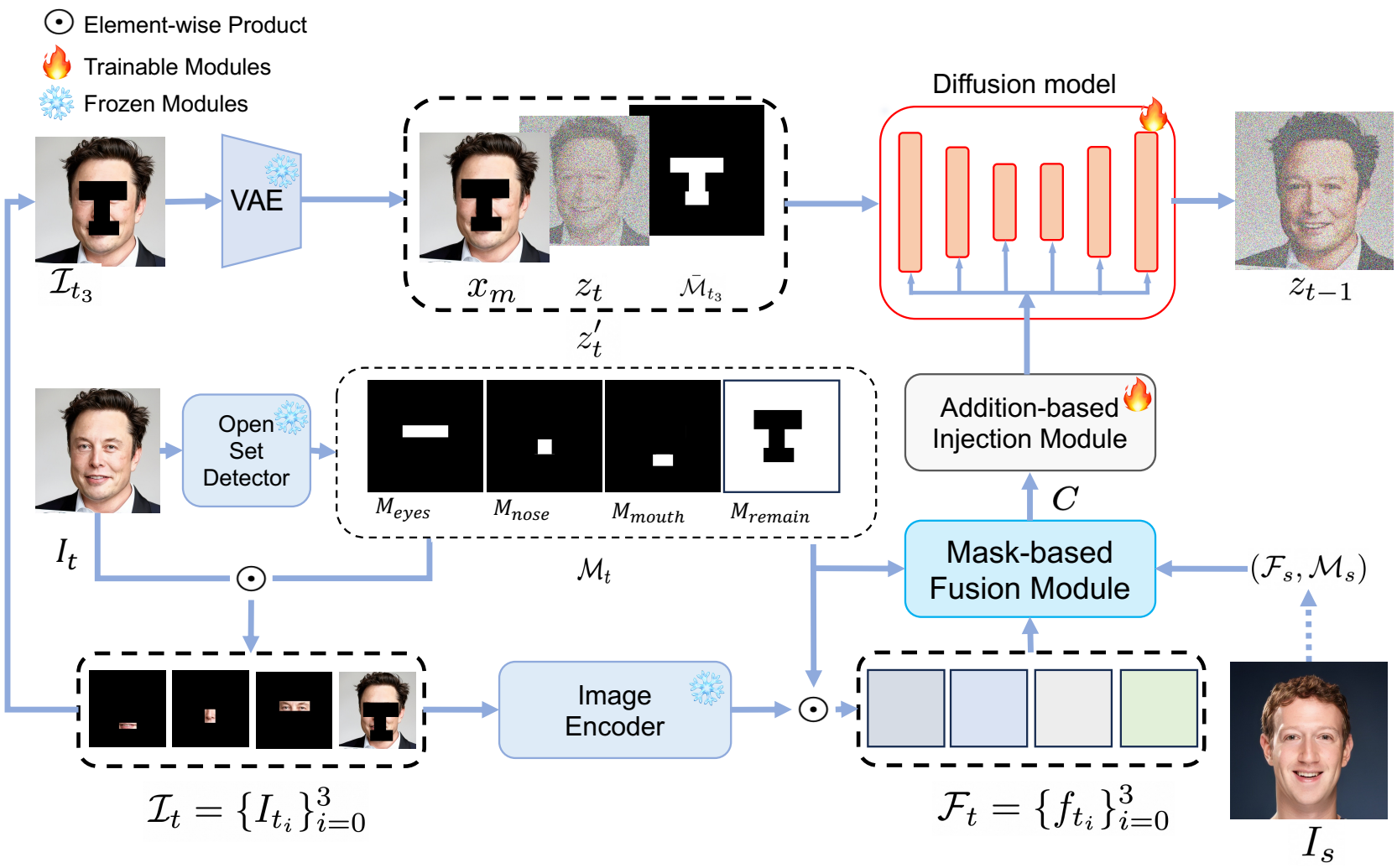


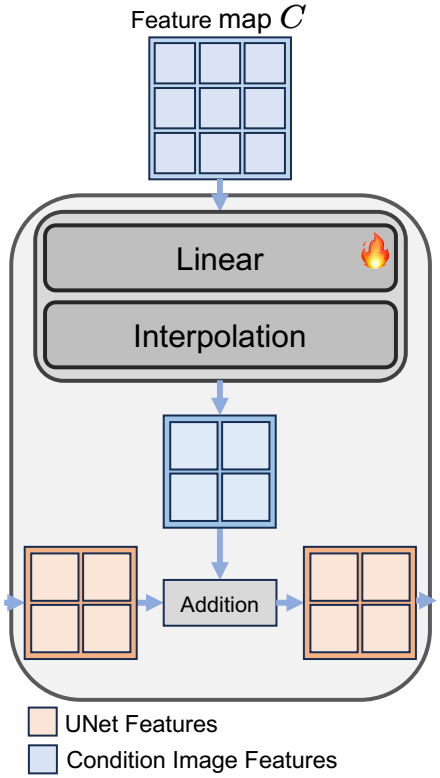
Results of facial parts swapping using FuseAnyPart

Challenges

- The main challenge in facial parts swapping lies in the fusion mechanism and multiple source images complicates the process.
- Popular methods that use cross-attention adapters struggle to accurately align fine-grained facial features.
- With multiple references increasing computational demands, efficient fusion becomes essential for performance.

Methodology





$$Z' = Z + \lambda \cdot \textbf{Inter}(\textbf{Linear}(C))$$

Experiments

Table 1: **Quantitative Comparisons on FF++.** We report Fréchet inception distance, eye similarity, nose similarity, mouth similarity and Mean Square Error and show that our method achieves SoTA or competitive results compared with existing methods. FacePartsSwap is essentially a cut & paste method, rather than a generative one, and thus has a higher FPSim. Therefore, we only present its results here and do not include it in the quantitative comparisons.

Methods	FID ↓	FPSim-E↑	FPSim-N↑	FPSim-M ↑	MSE↓
StableDiffusion [24]	18.57	0.3080	0.2215	0.2127	1.66
IP-Adapter [36]	69.35	0.2865	0.2066	0.1886	13.72
FacePartsSwap [5]	44.23	0.3269	0.2190	0.2220	24.40
E4S [15]	30.61	0.2764	**0.4047**	0.1903	3.03
Diffswap [38]	12.07	0.2461	0.1967	0.1731	**0.15**
Ours	**10.54**	**0.3186**	0.2234	**0.2196**	0.77

Qualitative results

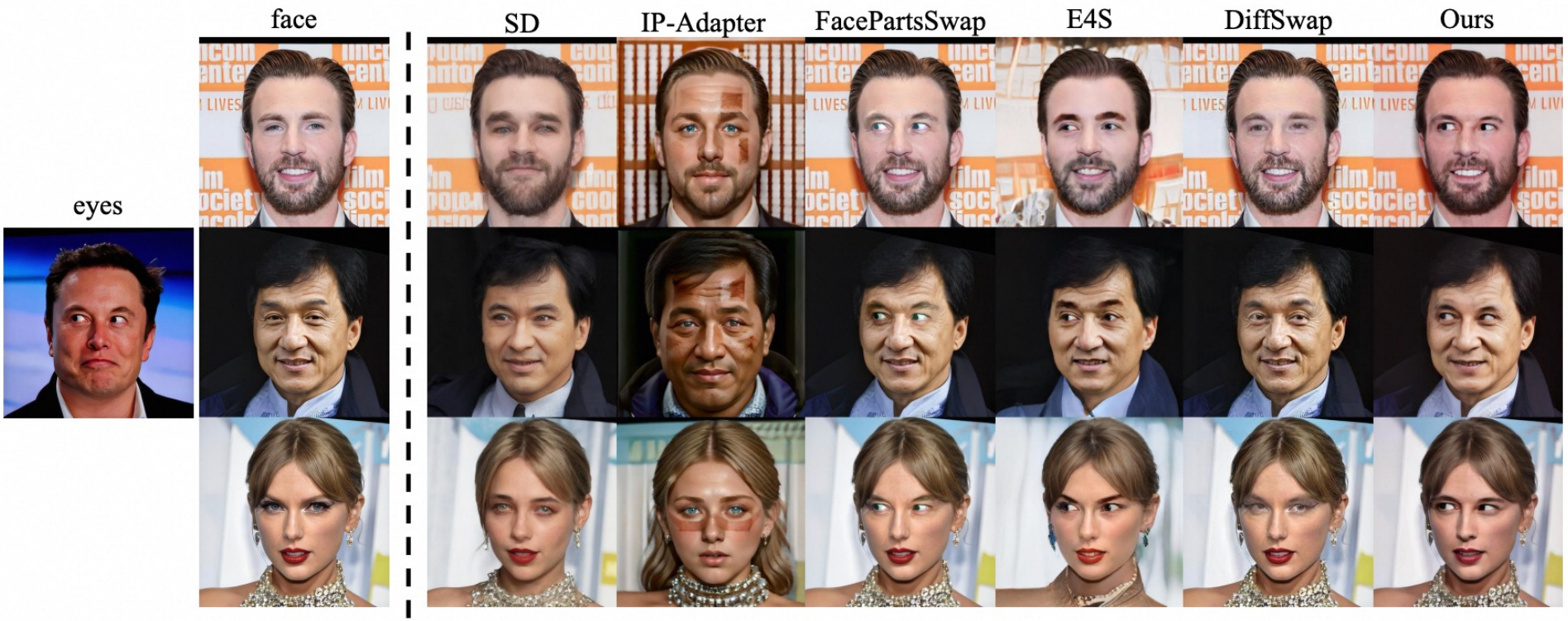


Figure 3: **Qualitative comparison of eyes swapping.** Our method produces high-fidelity results that maintain the consistency of facial features while ensuring a natural appearance.

Qualitative results

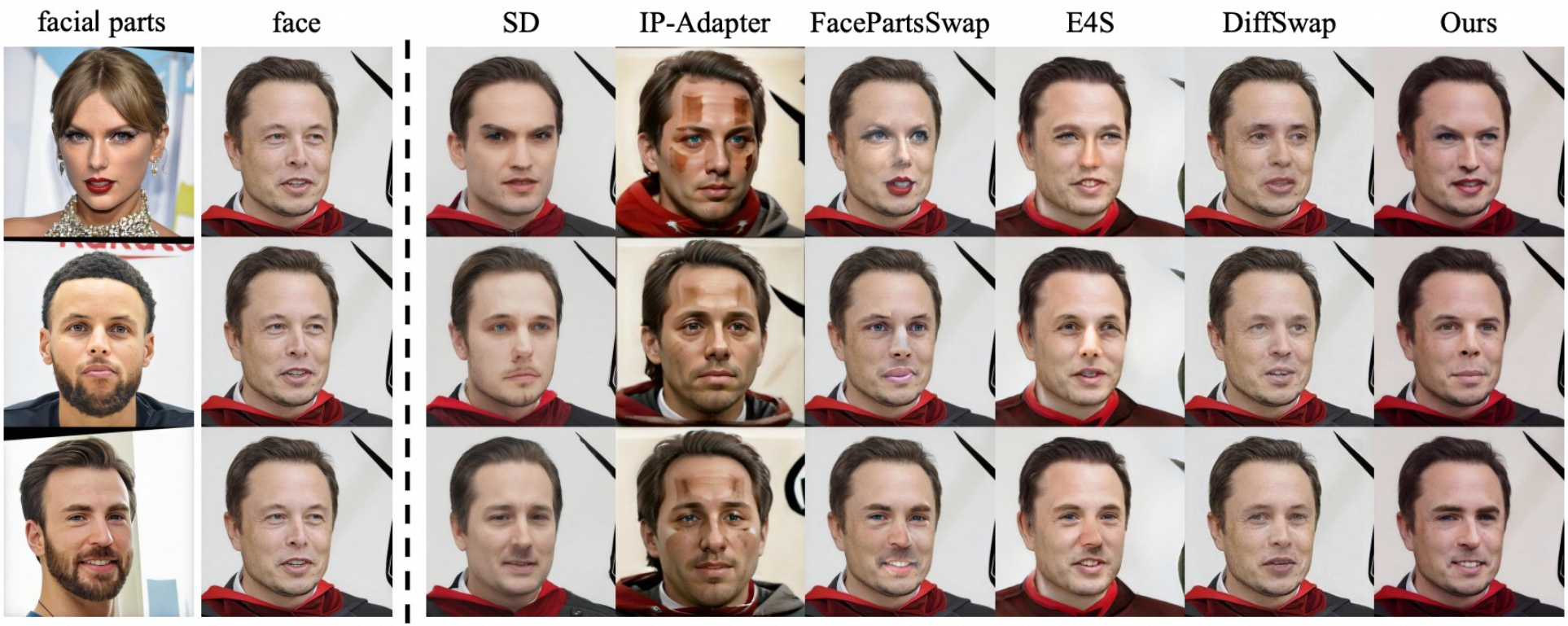


Figure 4: **Qualitative comparison of multiple facial parts swapping with a single reference face.** Our method can naturally replace multiple facial parts of one face with those of another and better preserve both the characteristics and the facial part shape. More results are presented in Fig. 10.

Qualitative results

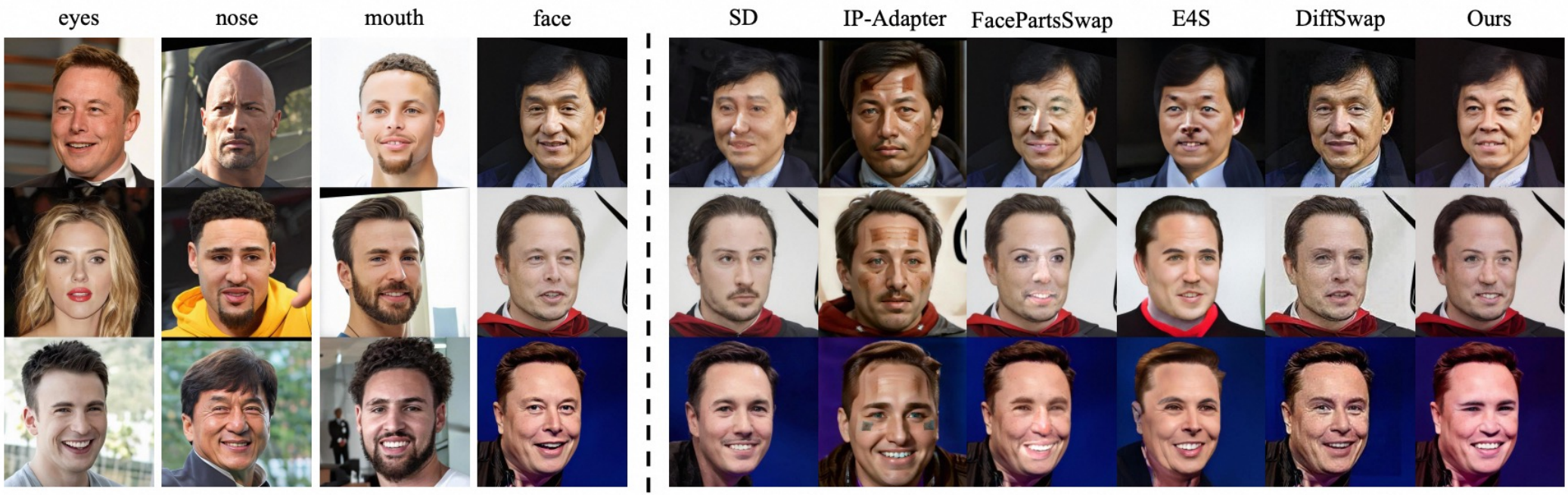
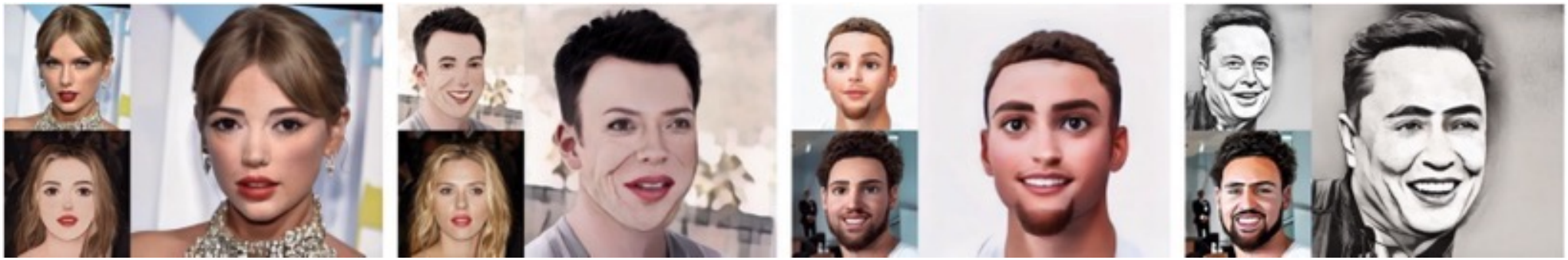
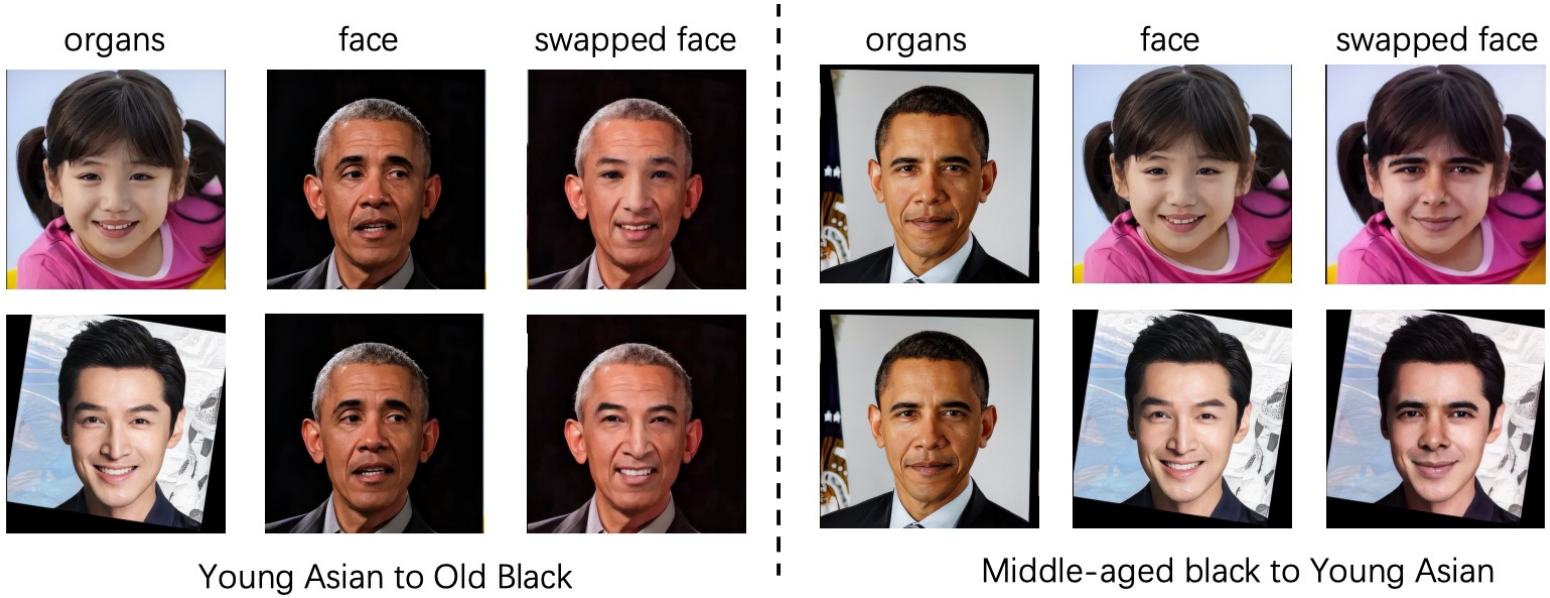


Figure 5: **Qualitative comparison of multi swapping with multiple reference faces.** Our method remains robust to different appearances of various reference facial parts.

Qualitative results





NEURAL INFORMATION PROCESSING SYSTEMS

阿里巴巴 Alibaba.com

FuseAnyPart: Diffusion-Driven Facial Parts Swapping via Multiple Reference Images

Zheng Yu*, Yaohua Wang*, Siying Cui,
Aixi Zhang, Wei-Long Zheng, Senzhang Wang