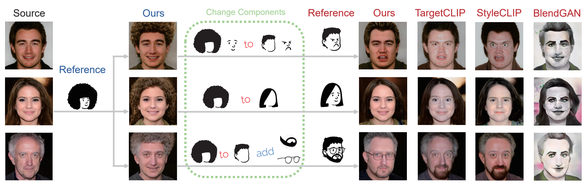
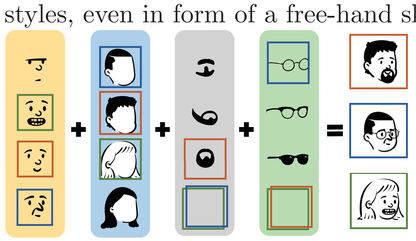
Bring Clipart to Life

Nanxuan Zhao [1], Shengqi Dang [2], Hexun Lin [2], Yang Shi [2], Nan Cao [2]
[1]Adobe Research [2]Tongji University
nanxuanzhao@gmail.com, dangsq123@tongji.edu.cn, hexunlin@ieee.org, shiyang1230@gmail.com, nan.cao@gmail.com

Abstract



We propose a new interaction method by guiding the editing with abstract clipart, composed of a set of simple semantic parts, allowing users to control across face photos with simple clicks. However, this is a challenging task given the large domain gap between colorful face photos and abstract clipart with limited data. To solve this problem, we introduce a framework called *ClipFaceShop* built on top of StyleGAN. The key idea is to take advantage of $\mathcal{W}+$ latent code encoded rich and disentangled visual features, and create a new lightweight selective feature adaptor to predict a modifiable path toward the target output photo. Since no pairwise labeled data exists for training, we design a set of losses to provide supervision signals for learning the modifiable path. Experimental results show that *ClipFaceShop* generates realistic and faithful face photos, sharing the same facial attributes as the reference clipart. We demonstrate that *ClipFaceShop* supports clipart in diverse styles, even in form of a free-hand sketch.



Examples of mix-and-match clipart, ©OpenPeeps. Users can easily create the reference clipart by combining the components through simple clicks.

References

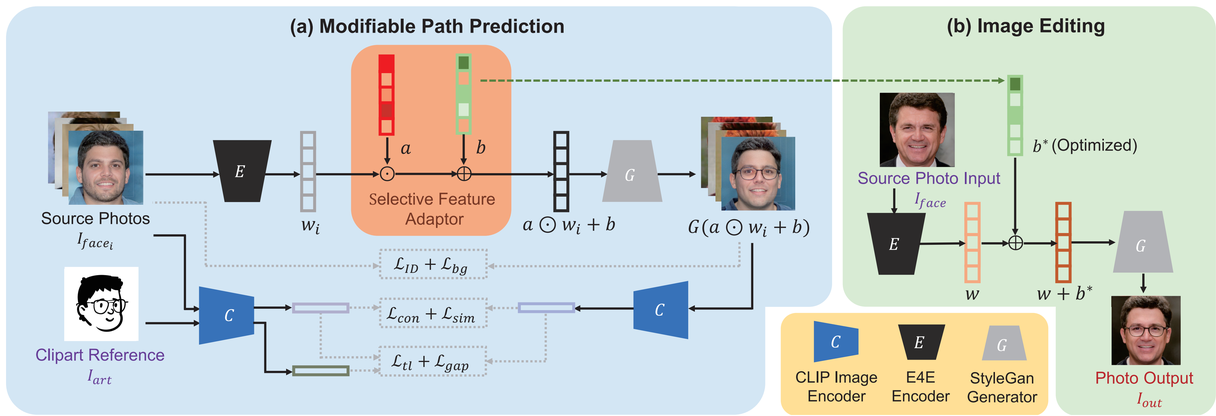
[1] Pablo Stanley. Openpeeps Dataset. In https://www.openpeeps.com/.

[2] Patashnik, Or and Wu, Zongze and etc. On Styleclip: Text-driven manipulation of stylegan imagery. In ICCV, 2021.

[3] Chefer, Hila and Benaim, Sagie and etc. On Image-based clip-guided essence transfer. In ECCV, 2022.

[4] Liu Mingcong and Li Qiang and etc. On Blendgan: implicitly gan blending for arbitrary stylized face generation. In NeurIPS, 2021.

Methodology



- **Selective Feature Adaptor**

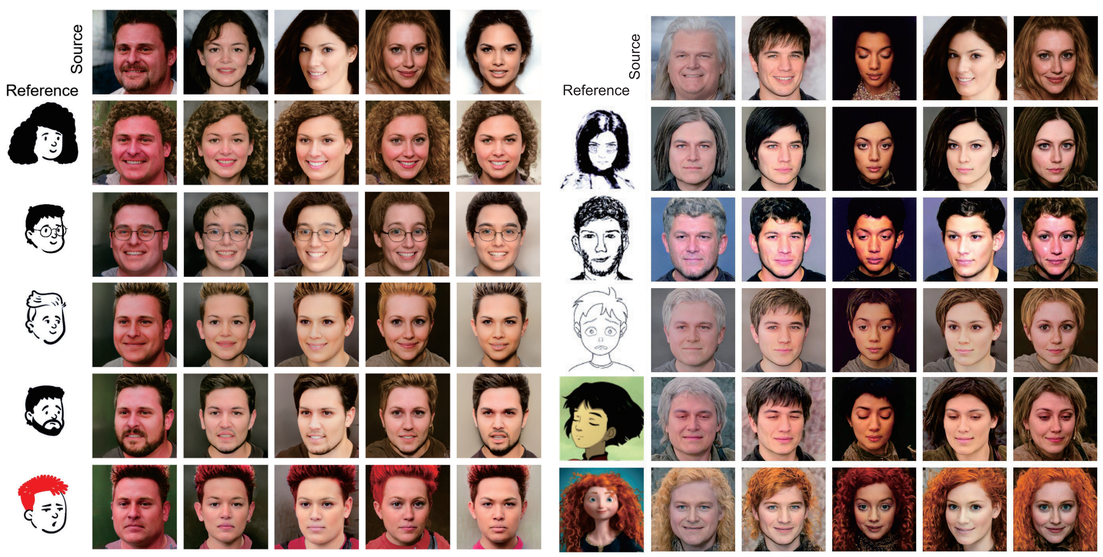
$$\mathcal{A}(w) = a \odot w + b, \quad I_{out} = G(w + b).$$

- **Training Objectives**

$$L = \lambda_{con}\mathcal{L}_{con} + \lambda_{tl}\mathcal{L}_{tl} + \lambda_{sim}\mathcal{L}_{sim} + \lambda_{ID}\mathcal{L}_{ID} + \lambda_{gap}\mathcal{L}_{gap} + \lambda_{bg}\mathcal{L}_{bg} + \lambda_{L_2}(\|a\|_2 + \|b\|_2)$$

To learn the modifiable path b and domain adaption operator a, we design a set of losses. They can be generally categorized into two groups, one for facial attributes transfer, and the other for identity and background preservation.

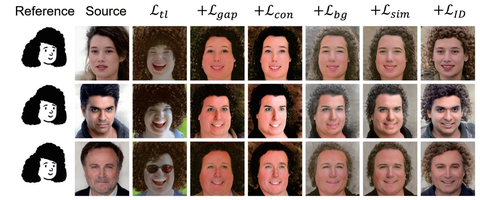
Results



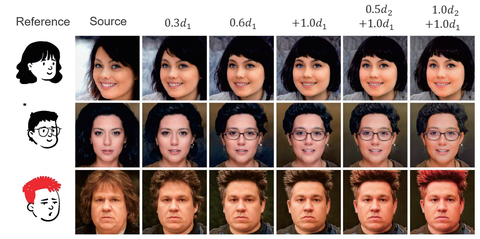
Our model can transfer the facial attributes and accessories not only at the semantic level, but at the appearance/style level of those in the reference clipart.

More Results

- **The effectiveness of different loss terms**

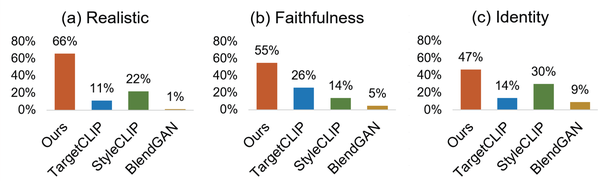


- **The effectiveness of parameters**



- **Different frames of the same person**

Comparison



Model	$ID\downarrow$	$FID\downarrow$	$SIM\uparrow$	$Color\uparrow$	$Beard_{Acc}\uparrow$	$Glass_{Acc}\uparrow$	$Hair_{Acc}\uparrow$	$Smile_{Acc}\uparrow$
BlendGAN	0.8418	105.14	0.66	0.52	0.50	0.50	0.51	0.46
StyleCLIP	**0.5452**	54.54	**0.77**	0.42	0.59	0.52	0.48	0.62
TargetCLIP	0.6642	**49.35**	0.68	0.60	**0.78**	0.53	0.46	0.64
Ours	0.5750	74.92	0.68	**0.65**	0.64	**0.56**	**0.58**	**0.80**