Celebrity and You Synthesis Generator

Zhu Chen, zhc047@stanford.edu
Facebook Inc.

Introduction

As the development of image morphing, computer generated images can be synthesized smoothly. Naturally, it leads us to explore real photo synthesis. The project is about generating pictures of real human faces synthesized with celebrities' faces, with user adjustable degree of synthesis. It would be an interesting feature to add to photoshop applications if users can combine their faces with celebrities' to get a picture that looks like them but somewhat different. The inputs to our neural net are two real photos. We then train the neural net to output two latent vectors that can be used by the StyleGAN generator to generate images similar to the real photos. Finally, we interpolate these two latent vectors and use the interpolated latent vector to generate the synthesized image.

Related Work

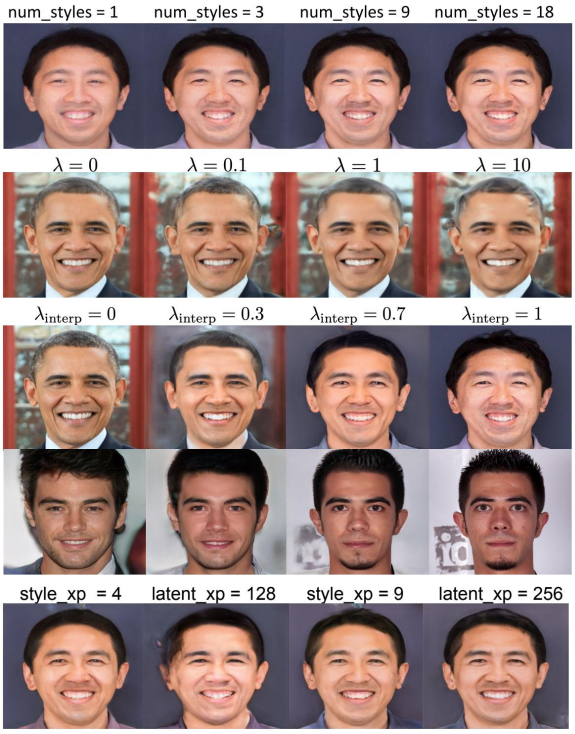
Among recent advances in GAN architectures after first proposal by Ian Goodfellow et al., Karras et al. introduced a style-based generator architecture which can separate high-level attributes on human faces, using same technique by Huang and Belongie in real time style transfer: adaptive instance normalization. Rameen Abdal et al. proposed to embed real images to StyleGAN latent space in an optimization setting.

Dataset

The datasets we use are two existing ones:

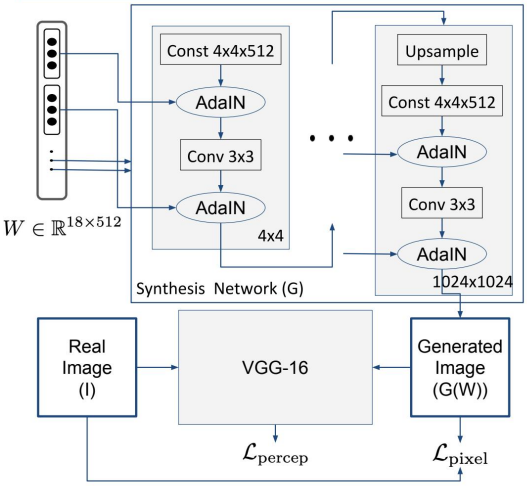
- CelebFaces Attributes Dataset (CelebA-HQ): It consists of more than 202,599 celebrity images at 1024 x 1024 resolution. It includes face images taken from different angles and in different backgrounds. Each is annotated with 40 attributes.
- Flickr-Faces-HQ (FFHQ): This dataset consists of 70,000 human face images each at 1024 x 1024 resolution. The faces are varied in terms of gender, image background, angle, etc.

Experiments and Analysis



Styles	1	3	9	18
Loss	10504	8835	6737	**6037**
PPL (x10^7)	1.0997	1.2537	0.9501	**0.8592**

Model Architecture



$$\mathcal{L}(G(W), I) = \mathcal{L}_{\text{percep}}(G(W), I) + \lambda\mathcal{L}_{\text{pixel}}(G(W), I)$$
$$= \sum_{i=1}^{4} \|A_i(G(W)) - A_i(I)\|_2 + \lambda\|G(W) - I\|_2$$

$$W^* = \arg\min_W \mathcal{L}(G(W), I),\ W_s = \lambda_{\text{interp}} W_p^* + (1 - \lambda_{\text{inpterp}}) W_c^*$$

Discussion and Future Work

We embedded two face photos into StyleGAN's latent vectors, and interpolated them to generate a synthesized image. Using only perceptual loss can reproduce the real image better than with pixel wise loss due to more freedom to optimize. Synthesis on styles has smoother results than on latent space because choosing cross point on latent space might separate two entities which control one feature. Future, we can restrict the distribution of latent space so that linear interpolation is more robust. Another direction might be using facial landmarks.

[1] Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems. 2014.
[2] Karras et al. "A style-based generator architecture for generative adversarial networks." arXiv preprint arXiv:1812.04948 (2018).
[3] Huang, Xun, and Serge Belongie. "Arbitrary style transfer in real-time with adaptive instance normalization." ICCV. 2017.
[4] Abdal et al. "Image2StyleGAN: How to Embed Images Into the StyleGAN Latent Space?." arXiv preprint arXiv:1904.03189 (2019).