ELEGANT: Exchanging Latent Encodings with GAN for Transferring Multiple Face Attributes

Taihong Xiao, Jiapeng Hong & Jinwen Ma

Department of Information Sciences, School of Mathematical Sciences and LMAM, Peking University

xiaotaihong@pku.edu.cn, jphong@pku.edu.cn, jwma@math.pku.edu.cn

Introduction

There are many approaches to transferring face attributes. However, most of them suffer from one or more following limitations:

1. Incapability of generating image by exemplars;
2. Being unable to transfer multiple face attributes simultaneously;
3. Low quality of generated images, such as low-resolution or artifacts.

Here are our intuitions to address the above three limitations.

1. To generate images by exemplars, a model must receive a reference for conditional image generation.
2. For manipulating multiple attributes simultaneously, the latent encodings of an image can be divided into different parts, where each part encodes information of a single attribute.
3. To improve the quality of generated images, we adopt the idea of residual learning and multi-scale discriminators.

Model

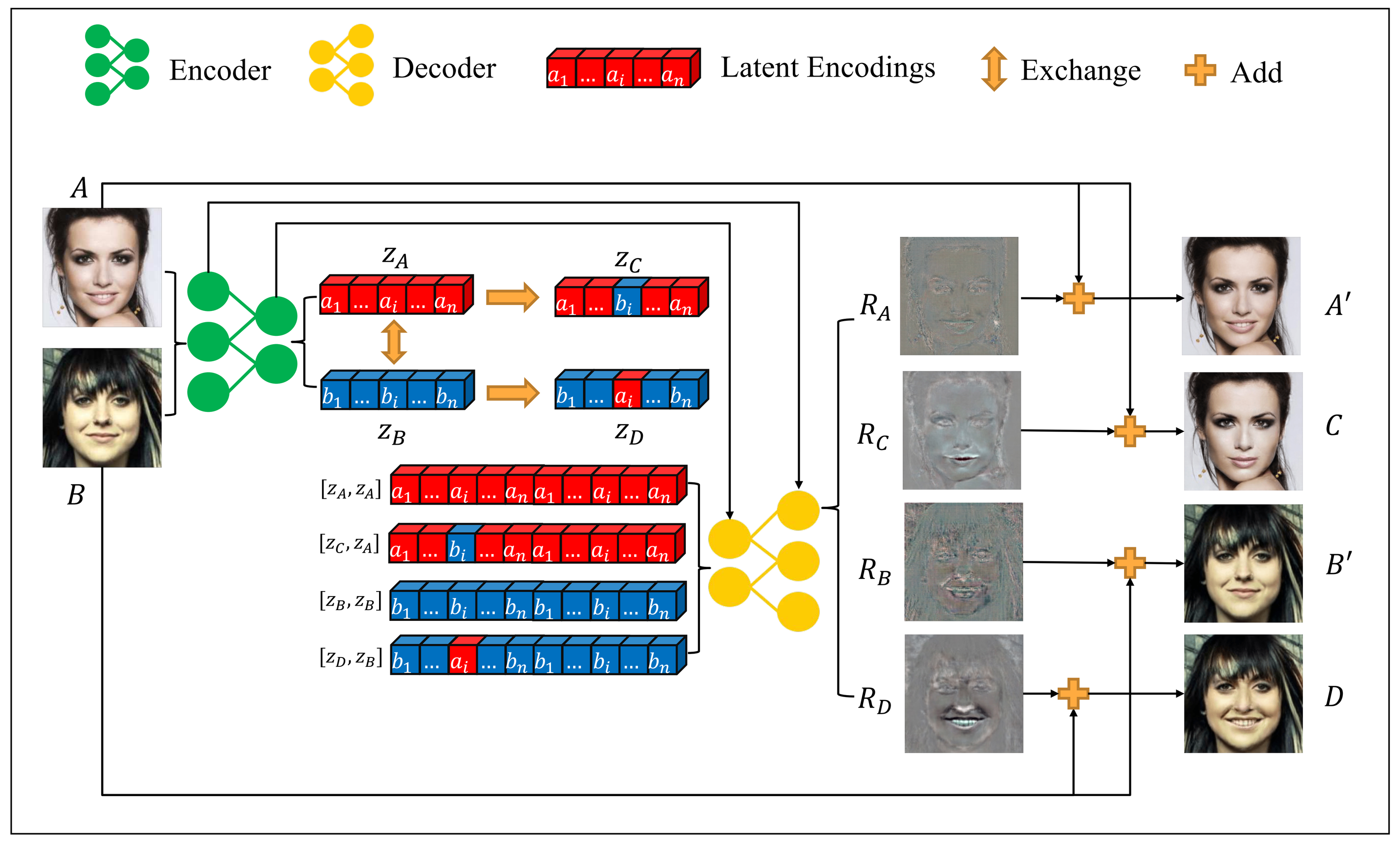


Figure 1: The ELEGANT model architecture.

- Two input images batches of opposite attribute. Formally, the attribute labels of A and B are required to be in this form $Y^A = (y_1^A, \dots, 1_i, \dots, y_n^A)$ and $Y^B = (y_1^B, \dots, 0_i, \dots, y_n^B)$.
- An encoder was used to obtain the disentangled latent encodings for different attributes,

$$z_A = \mathrm{Enc}(A) = [a_1, \dots, a_i, \dots, a_n], \qquad z_B = \mathrm{Enc}(B) = [b_1, \dots, b_i, \dots, b_n] \tag{1}$$

where a_i and b_i are the smiling feature tensors.

- Exchanging latent encodings so as to swap face attribute,

$$z_C = [a_1, \dots, b_i, \dots, a_n], \qquad z_D = [b_1, \dots, a_i, \dots, b_n] \tag{2}$$

- Residual learning: the input of the decoder is the concatenation (or difference) of the source and target encodings, and the output of the decoder is the residual image.

$$\mathrm{Dec}([z_A, z_A]) = R_A,\ A' = A + R_A \qquad \mathrm{Dec}([z_C, z_A]) = R_C,\ C = A + R_C \tag{3}$$

$$\mathrm{Dec}([z_B, z_B]) = R_B,\ B' = B + R_B \qquad \mathrm{Dec}([z_D, z_B]) = R_D,\ D = B + R_D \tag{4}$$

- Multi-scale discriminators: two discriminators having identical network structures whereas operating at different image scales. D_1 has a smaller receptive field compared with D_2. Therefore, D_1 is specialized in guiding the Enc and Dec to produce finer details, whereas D_2 is adept in handling the holistic image content so as to avoid generating grimaces.

Loss Functions

- The multi-scale discriminators D_1 and D_2 receive the standard adversarial loss

$$\begin{aligned} L_{\mathrm{D}_1} = & -\mathbb{E}(\log(\mathrm{D}_1(A|Y^A))) - \mathbb{E}(\log(1 - \mathrm{D}_1(C|Y^C))) \\ & -\mathbb{E}(\log(\mathrm{D}_1(B|Y^B))) - \mathbb{E}(\log(1 - \mathrm{D}_1(D|Y^D))) \end{aligned} \tag{5}$$

$$\begin{aligned} L_{\mathrm{D}_2} = & -\mathbb{E}(\log(\mathrm{D}_2(A|Y^A))) - \mathbb{E}(\log(1 - \mathrm{D}_2(C|Y^C))) \\ & -\mathbb{E}(\log(\mathrm{D}_2(B|Y^B))) - \mathbb{E}(\log(1 - \mathrm{D}_2(D|Y^D))) \end{aligned} \tag{6}$$

$$L_{\mathrm{D}} = L_{\mathrm{D}_1} + L_{\mathrm{D}_2} \tag{7}$$

- As for the Enc and Dec, there are two types of losses. The first type is the reconstruction loss,

$$L_{reconstruction} = ||A - A'|| + ||B - B'|| \tag{8}$$

and the second type is the standard adversarial loss

$$\begin{aligned} L_{adv} = & -\mathbb{E}(\log(\mathrm{D}_1(C|Y^C))) - \mathbb{E}(\log(\mathrm{D}_1(D|Y^D))) \\ & -\mathbb{E}(\log(\mathrm{D}_2(C|Y^C))) - \mathbb{E}(\log(\mathrm{D}_2(D|Y^D))) \end{aligned} \tag{9}$$

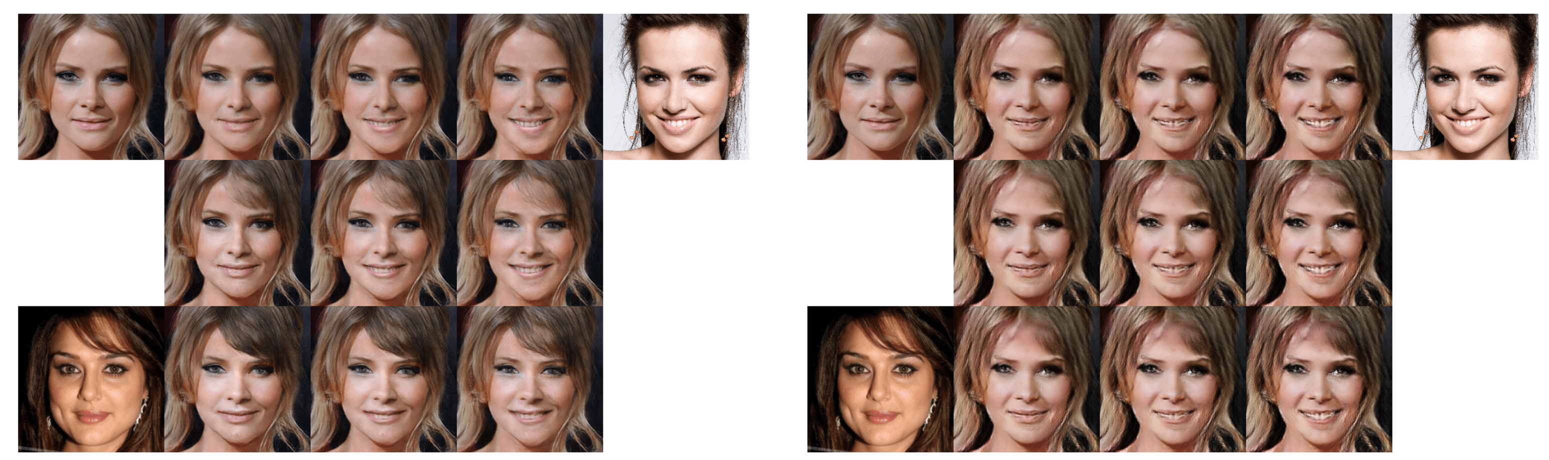
The total loss for the generator is

$$L_G = L_{reconstruction} + L_{adv}. \tag{10}$$

Experiments

Figure 2: Face image generation by exemplars.

Figure 3: Interpolation of different types of bangs.



(a) Bangs and Smiling

Figure 4: Multiple Attributes Interpolation. Compared with DNA-GAN, ELEGANT preserves the face id without the annihilation operation, and generate visually better images.

Table 1: FID of Different Methods with respect to five attributes. The $+$ ($-$) represents the generated images by adding (removing) the attribute.

FID	bangs		smiling		mustache		eyeglasses		male	
	+	−	+	−	+	−	+	−	+	−
UNIT	135.41	137.94	120.25	125.04	119.32	131.33	111.49	139.43	152.16	154.59
CycleGAN	**27.81**	33.22	**23.23**	**22.74**	43.58	55.49	**36.87**	**48.82**	60.25	**46.25**
StarGAN	59.68	71.07	51.36	78.87	99.03	176.18	70.40	142.35	70.14	206.21
DNA-GAN	79.27	76.89	77.04	72.35	126.33	127.66	75.02	75.96	121.04	118.67
ELEGANT	30.71	**31.12**	25.71	24.88	**37.51**	**49.13**	47.35	60.71	**59.37**	56.80

Scan for source code.

https://github.com/Prinsphield/ELEGANT