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. Inability 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.

Model

- Two input images batches of opposite attribute. Formally, the attribute labels of A and B are required to be in this form $y_A = [a_1, \ldots, a_n]$ and $y_B = [b_1, \ldots, b_n]$.
- An encoder was used to obtain the disentangled latent encodings for different attributes, $z_A = \text{Enc}(A) = [a_1, \ldots, a_n]$ and $z_B = \text{Enc}(B) = [b_1, \ldots, b_n]$.
- Exchanging latent encodings so as to swap face attribute, $z_A = [a_1, \ldots, a_n, b_1, \ldots, b_n]$ and $z_B = [b_1, \ldots, b_n, a_1, \ldots, a_n]$.
- 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.

Loss Functions

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

$\mathcal{L}_{adv} = -\mathbb{E}_{y_A\sim p_{data}}[\mathbb{E}_{y_B\sim p_{data}}[\log(D_1(y_A, y_B))] + \mathbb{E}_{y_B\sim p_{data}}[\log(1 - D_2(y_B, y_A))]]$

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

$\mathcal{L}_{adv} = -\mathbb{E}_{y_A\sim p_{data}}[\mathbb{E}_{y_B\sim p_{data}}[\log(D_1(y_A, y_B))] + \mathbb{E}_{y_B\sim p_{data}}[\log(1 - D_2(y_B, y_A))]]$

- As for the Enc and Dec, there are two types of losses. The first type is the reconstruction loss, $\mathcal{L}_{\text{reconstruction}} = ||A - A'|| + ||B - B'||$ and the second type is the standard adversarial loss

$\mathcal{L}_{\text{adv}} = -\mathbb{E}_{y_A\sim p_{data}}[\mathbb{E}_{y_B\sim p_{data}}[\log(D_2(y_A, y_B))] + \mathbb{E}_{y_B\sim p_{data}}[\log(1 - D_1(y_B, y_A))]]$

The total loss for the generator is

$\mathcal{L}_G = \mathcal{L}_{\text{reconstruction}} + \mathcal{L}_{\text{adv}}$

Experiments

- Multi-scale discriminators: two discriminators having identical network structures whereas operating at different image scales. $D_2$ has a smaller receptive field compared with $D_1$. Therefore, $D_2$ 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.

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

<table>
<thead>
<tr>
<th>Method</th>
<th>bangs</th>
<th>smiling</th>
<th>mustache</th>
<th>eyeglasses</th>
<th>male</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNIT</td>
<td>115.41</td>
<td>37.94</td>
<td>120.25</td>
<td>109.32</td>
<td>131.33</td>
</tr>
<tr>
<td>CycleGAN</td>
<td>27.81</td>
<td>33.22</td>
<td>27.32</td>
<td>14.56</td>
<td>33.49</td>
</tr>
<tr>
<td>StarGAN</td>
<td>59.68</td>
<td>71.94</td>
<td>83.36</td>
<td>99.03</td>
<td>57.68</td>
</tr>
<tr>
<td>DNA-GAN</td>
<td>79.27</td>
<td>76.89</td>
<td>37.04</td>
<td>32.43</td>
<td>120.33</td>
</tr>
<tr>
<td>ELEGANT</td>
<td>30.71</td>
<td>31.12</td>
<td>35.71</td>
<td>24.88</td>
<td>37.51</td>
</tr>
</tbody>
</table>

Figure 1: The ELEGANT model architecture.

Figure 2: Face image generation by exemplars.

Figure 3: Interpolation of different types of bangs.

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

Scan for source code.
https://github.com/Prinsphield/ELEGANT