Attributes Aware Face Generation with Generative Adversarial Networks

Zheng Yuan    Jie Zhang    Shiguang Shan    Xilin Chen

Institute of Computing Technology,
Chinese Academy of Sciences
Task: Attribute to Facial Image

- blond hair
- female
- mouth slightly open
- arched eyebrows
- heavy makeup
Related Work

• Text to image
  ➢ StackGAN++, AttnGAN, MirrorGAN, etc.
  ➢ the input is different: text vs attribute
  ➢ can not well embed the attribute label

• Attribute to image
  ➢ Attribute2sketch2face, Lu et al., Wang et al., etc.
  ➢ the generated images are always low resolution
  ➢ do not consider the relationship between different attributes
Method: AFGAN

- Attributes Aware Face Generation with Generative Adversarial Networks
Method

- **AEM: Attribute Embedding Module**
  - convert the input face attributes into global and local features respectively
  - two path embedding: well reflect their meanings of the input attribute

\[
S_{local} = \text{embed}_1 \ast S_{attr} + \text{embed}_2 \ast (1 - S_{attr})
\]
Method

- **AEM: Attribute Embedding Module**
  - convert the input face attributes into global and local features respectively
  - self attention layer: model the relationships between different attributes

\[
\begin{align*}
  f(x) &= W_f \ast S_{\text{local}} \\
  s_{ij} &= f(x_i)^T g(x_j) \\
  \beta_{ij} &= \frac{\exp(s_{ij})}{\sum_{i=1}^{N} \exp(s_{ij})} \\
  h(x) &= W_h \ast S_{\text{local}} \\
  S_{\text{local}}' &= \sum_{i=1}^{N} \beta_{ij} h(x_i) \\
  S_{\text{local}}' &= (S_{\text{local}}'_1, S_{\text{local}}'_2, \cdots, S_{\text{local}}'_N) \in \mathbb{R}^{C \times N}
\end{align*}
\]
Method

• SIGM: Stacked Image Generation Module
  ➢ gradually generate faces with more details through a three-stage generator
  ➢ can generate images with high resolution

$h_0 = F_0(z, F^{ca}(S_{global}))$

$h_i = F_i(h_{i-1}, F_i^{attn}(S_{local}', h_{i-1}))$

$x_i = G_i(h_i)$
Method

- **SCM**: Similarity Constrain Module
  - encode the generated images with a pretrained model: $i_{\text{local}}$ and $i_{\text{global}}$
  - calculate the matching degree between attribute features and image features
  - the generated images can match the input attributes well

\[
R(c_i, S_{\text{local}'}^{\prime}) = \frac{c_i^T S_{\text{local}'}^{\prime}}{\|c_i\| \|S_{\text{local}'}^{\prime}\|}
\]

\[
S_{\text{global}} = \sum_{i=1}^{289} \alpha_{ij} i_{\text{local}j}
\]

\[
R_{\text{local}}(Q, D) = \log \left( \sum_{i=1}^{N} \exp(\gamma_2 R(c_i, S_{\text{local}'})) \right) \right)^{\frac{1}{\gamma_2}}
\]

\[
R_{\text{global}}(Q, D) = \frac{i_{\text{global}}^T S_{\text{global}}}{\|i_{\text{global}}^T\| \|S_{\text{global}}\|}
\]
Formulation

- Generator
  - Overall objective function
    \[ \mathcal{L} = \mathcal{L}_G + \mathcal{L}_{SCM} \]

  - In SIGM
    \[ \mathcal{L}_G = \sum_{i=0}^{2} \mathcal{L}_{G_i} \]
    \[ \mathcal{L}_{G_i} = -\frac{1}{2} \mathbb{E}_{x_i \sim p_{G_i}} \left[ \log (D_i (x_i)) \right] - \frac{1}{2} \mathbb{E}_{x_i \sim p_{G_i}} \left[ \log (D_i (x_i, S_{global})) \right] \]
    unconditional conditional

  - In SCM
    \[ \mathcal{L}_{SCM} = \mathcal{L}_{1\text{local}} + \mathcal{L}_{2\text{local}} + \mathcal{L}_{1\text{global}} + \mathcal{L}_{2\text{global}} \]
    \[ \mathcal{L}_{1\text{local}} = -\sum_{i=1}^{M} \log P_{local} (D_i | Q_i) \]
    \[ \mathcal{L}_{2\text{local}} = -\sum_{i=1}^{M} \log P_{local} (Q_i | D_i) \]
    \[ \mathcal{L}_{1\text{global}} = -\sum_{i=1}^{M} \log P_{global} (D_i | Q_i) \]
    \[ \mathcal{L}_{2\text{global}} = -\sum_{i=1}^{M} \log P_{global} (Q_i | D_i) \]
Formulation

- The matching degree between generated images and input attributes

\[ p^{local}(Q_i | D_i) = \frac{\exp(\gamma_3 R^{local}(Q_i, D_i))}{\sum_{j=1}^{M} \exp(\gamma_3 R^{local}(Q_j, D_i))} \]

\[ p^{local}(D_i | Q_i) = \frac{\exp(\gamma_3 R^{local}(Q_i, D_i))}{\sum_{j=1}^{M} \exp(\gamma_3 R^{local}(Q_i, D_j))} \]

\[ p^{global}(Q_i | D_i) = \frac{\exp(\gamma_3 R^{global}(Q_i, D_i))}{\sum_{j=1}^{M} \exp(\gamma_3 R^{global}(Q_j, D_i))} \]

\[ p^{global}(D_i | Q_i) = \frac{\exp(\gamma_3 R^{global}(Q_i, D_i))}{\sum_{j=1}^{M} \exp(\gamma_3 R^{global}(Q_i, D_j))} \]

- Discriminator

\[ \mathcal{L}_D = \sum_{i=0}^{2} \mathcal{L}_{D_i} \]

\[ \mathcal{L}_{D_i} = -\frac{1}{2} \mathbb{E}_{x_i^{gt} \sim p_{data_i}} \left[ \log D_i(x_i^{gt}) \right] - \frac{1}{2} \mathbb{E}_{x_i \sim p_{G_i}} \left[ \log (1 - D_i(x_i)) \right] \]

unconditional

\[ -\frac{1}{2} \mathbb{E}_{x_i^{gt} \sim p_{data_i}} \left[ \log D_i(x_i^{gt}, S^{global}) \right] \]

conditional

\[ -\frac{1}{2} \mathbb{E}_{x_i \sim p_{G_i}} \left[ \log (1 - D_i(x_i, S^{global})) \right] \]
## Experiment

- **Dataset:** CelebA
- **Attribute:** 18 attributes selected from full 40 attributes

<table>
<thead>
<tr>
<th>No.</th>
<th>Attribute</th>
<th>No.</th>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5 o'clock Shadow</td>
<td>9</td>
<td>Eyeglasses</td>
</tr>
<tr>
<td>1</td>
<td>Arched Eyebrows</td>
<td>10</td>
<td>Gray Hair</td>
</tr>
<tr>
<td>2</td>
<td>Bags Under Eyes</td>
<td>11</td>
<td>Male</td>
</tr>
<tr>
<td>3</td>
<td>Bald</td>
<td>12</td>
<td>Mouth Slightly Open</td>
</tr>
<tr>
<td>4</td>
<td>Bangs</td>
<td>13</td>
<td>Narrow Eyes</td>
</tr>
<tr>
<td>5</td>
<td>Black Hair</td>
<td>14</td>
<td>No Beard</td>
</tr>
<tr>
<td>6</td>
<td>Blond Hair</td>
<td>15</td>
<td>Pale Skin</td>
</tr>
<tr>
<td>7</td>
<td>Brown Hair</td>
<td>16</td>
<td>Pointy Nose</td>
</tr>
<tr>
<td>8</td>
<td>Bushy Eyebrows</td>
<td>17</td>
<td>Smiling</td>
</tr>
</tbody>
</table>
Experiment

• The attention maps in SCM module

➢ the response areas of different attributes are consistent with common sense
Experiment

• The generated face images of three stages in SIGM module
  ➢ the faces generated in the three stages are consistent
  ➢ more details in higher resolution images
Experiment

• The generated face images in ablation study of AFGAN model
  ➢ the quality of the images generated by AFGAN is clearer and more realistic

• The quantitative evaluation of generated images

<table>
<thead>
<tr>
<th>Setting</th>
<th>Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>AttnGAN</td>
<td>0.902</td>
</tr>
<tr>
<td>AFGAN w/o AEM</td>
<td>0.924</td>
</tr>
<tr>
<td>AFGAN w/o SCM</td>
<td>0.940</td>
</tr>
<tr>
<td>AFGAN(ours)</td>
<td><strong>0.955</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>BRISQUE↓</th>
<th>IS↑</th>
<th>FID↓</th>
<th>MS-SSIM↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>AttnGAN</td>
<td>62.843</td>
<td>5.124</td>
<td>40.254</td>
<td>0.398</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>—</td>
<td>2.2</td>
<td>43.8</td>
<td>—</td>
</tr>
<tr>
<td>AFGAN(ours)</td>
<td><strong>35.979</strong></td>
<td>5.853</td>
<td>36.607</td>
<td><strong>0.347</strong></td>
</tr>
</tbody>
</table>
Thank you!

zheng.yuan@vipl.ict.ac.cn

Institute of Computing Technology,
Chinese Academy of Sciences