Analyzing and Combating Attribute Bias for Face Restoration

Zelin Li, Dan Zeng, Xiao Yan, Qiaomu Shen, Bo Tang
Face restoration: Recover high-quality (HQ) faces from low-quality (LQ) faces
- Super-resolution, Denoise, Deblur, etc
Previous Problem: Over-smooth
Background

- Face restoration: Recover high-quality (HQ) faces from low-quality (LQ) faces
  - Super-resolution, Denoise, Deblur, etc
- Previous Problem: Over-smooth
Face restoration: Recover high-quality (HQ) faces from low-quality (LQ) faces

- Super-resolution, Denoise, Deblur, etc

Leveraging generative prior
Face restoration: Recover high-quality (HQ) faces from low-quality (LQ) faces
  - Super-resolution, Denoise, Deblur, etc
Leveraging generative prior
Attribute Bias

- Key face attributes are dramatically different

**Input**

```
16×16
```

**Target**

```
32×
Age: 17, female
```
attribute bias

- Key face attributes are dramatically different

Input: 16×16

<table>
<thead>
<tr>
<th>Input</th>
<th>VQFR</th>
<th>GFPGAN</th>
<th>GPEN</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age: 48, male</td>
<td>Age: 57, male</td>
<td>Age: 54, male</td>
<td>Age: 17, female</td>
<td></td>
</tr>
</tbody>
</table>
Analysis on Attribute Bias

- Attribute information loses as image resolution decreases

(a) Age Confidence.

(b) Gender Confidence.
Analysis on Attribute Bias

Attribute information loses as image resolution decreases

(a) Age Confidence.

(b) Gender Confidence.
Analysis on Attribute Bias

- Attribute information loses as image resolution decreases

(a) Age Confidence.

(b) Gender Confidence.
Analysis on Attribute Bias

➤ Attribute bias enlarges as image resolution decreases
Analysis on Attribute Bias

Attribute bias enlarges as image resolution decreases.
Attribute bias enlarges as image resolution decreases
Analysis on Attribute Bias

- Affected by the training data distribution

---

(a) Mean confidence of male (test)

(b) Probability distribution of age
Challenges

- The causes are inevitable
  - Lack of attribute information
  - Training data prior
Challenges

- The causes are inevitable
  - Lack of attribute information – Degradation is unavoidable
  - Training data prior – Hard to collect a large dataset with balanced attribute distribution
Challenges

➢ The causes are inevitable
  - Lack of attribute information – Degradation is unavoidable
  - Training data prior – Hard to collect a large dataset with balanced attribute distribution

➢ Attribute information is obtainable
  - Witness description, Actor profile
Challenges

- The causes are inevitable
  - Lack of attribute information – Degradation is unavoidable
  - Training data prior – Hard to collect a large dataset with balanced attribute distribution

- Attribute information is obtainable
  - Witness description, Actor profile

- Given attribute information, do image restoration
  - Align attribute & image information
  - Let the attribute have a fine-grained impact on the restoration
Methodology

x
Methodology

$X \rightarrow \text{Encoder} \rightarrow \text{Linear}$
Methodology

$x$

Encoder

Linear

$\alpha_0$, $\alpha_1$, $\beta_0$, $\beta_5$, $\beta_9$

Gender representation

Male, Female

Age representation

0-2, ..., 20-30, ..., 70+
Methodology

Encoder

Decoder

Linear

Adaptive feature fusion module

\( \alpha_0 \alpha_1 \beta_0 \beta_5 \beta_9 \)

Gender representation

Age representation

\( x \)

\( \tilde{y} \)
Methodology

Encoder

Decoder

Adaptive feature fusion module

\[ x \rightarrow \alpha_0, \alpha_1 \rightarrow \beta_0, \beta_5, \beta_9 \rightarrow \text{Linear} \rightarrow \text{Style Block} \rightarrow \text{Style Block} \rightarrow \text{Style Block} \rightarrow y \rightarrow \hat{x} \]

\[ L_{deg}, L_{res}, L_{att} \]

Gender representation: Male, Female

Age representation: 0-2, 70+

\[ F_{enc}(i) \rightarrow \text{Conv} \rightarrow \text{Conv} \rightarrow \text{Conv} \rightarrow F_{dec}(i) \rightarrow \text{Conv} \rightarrow \text{Conv} \rightarrow F_{dec}(i+1) \]
Methodology:
- Determine the attribute weights
- Base latent vector + weighted sum of attribute representations
- Update the attribute representations through back-propagation (in the training phase)

Benefits:
- Fine-grained control by weights
- Cheap to extend to more attributes
Two-objective Optimization

- **Image Restoration**
  - **Reconstruction Loss**
    \[ \mathcal{L}_{rec} = \lambda_{L_{pix}} \| \hat{y} - y \|_1 + \mathcal{L}_{per} \]
    \[ \mathcal{L}_{per} = \| \phi(\hat{y}) - \phi(y) \|_1 + \lambda_{style} \| \text{Gram}(\phi(\hat{y})) - \text{Gram}(\phi(y)) \|_1 \]
  - **Adversarial Loss**
    \[ \mathcal{L}_{adv,D} = \mathbb{E}_{\hat{y}}[\text{Softplus}(D(\hat{y}))] + \mathbb{E}_y[\text{Softplus}(-D(y))] \]
    \[ \mathcal{L}_{adv,G} = \mathbb{E}_{\hat{y}}[\text{Softplus}(-D(\hat{y}))] \]

- **Attribute Consistency**
  - **Attribute Loss**
    \[ \mathcal{L}_{att} = \text{CE}(a, P(a|\hat{y})) \]

\( \phi \): Pretrained VGG-16
\( D \): Discriminator
\( \text{Gram} \): Gram matrix
\( \hat{y} \): Restored image
\( y \): Target image
For loss calculation:
- \( Y \) is degraded to \( X \) to form pair
- FR model estimate \( \hat{Y} \) based on \( X \)
- Calculate losses based on \( \hat{Y}, Y \) and attribute

Model only trained with paired image and attribute label
Pseudo-pair Strategy

\[ x = \left( (y \otimes k)_{+r} + n_\sigma \right)_{JPEG_q} \]

\[ x = \left( (y \otimes k)_{+r} + n_\sigma \right)_{JPEG_q} \]

- \( \hat{y} \) is degraded to \( \hat{x} \) by the same degradation model
- Calculate losses based on \( \hat{x}, x \) and attribute (\( L_{deg} = |x - \hat{x}| \))
- Straight-through estimator for gradient calculation
Metrics

- **Information fidelity:**
  - PSNR, SSIM

- **Image quality:**
  - NIQE, FID

- **Attribute Bias:**

\[
SSIM(x, y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)}
\]

\[
PSNR = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}}\right)
\]

**Age/Gender error:**

\[
1 - \frac{\sum_{y \in D} \mathbb{I}(\mathbf{C}(\hat{y}) \neq \mathbf{C}(y))}{|D|}
\]

**Legend:**

- \(\mathbf{C}\) : Pretrained Classifier
- \(c_1, c_2\) : Constants
- \(\text{MAX}_I^2\) : 255² or 1²
## Performance Comparison

- **Quantitatively comparison on CelebA-HQ (Synthesized LQs):**

<table>
<thead>
<tr>
<th>CelebA-HQ</th>
<th>NIQE</th>
<th>FID</th>
<th>PSNR</th>
<th>SSIM</th>
<th>Age error (%)</th>
<th>Gender error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilinear</td>
<td>16.59</td>
<td>213.3</td>
<td><strong>23.25</strong></td>
<td><strong>0.6951</strong></td>
<td>60.90</td>
<td>14.37</td>
</tr>
<tr>
<td>AACNN</td>
<td>4.132</td>
<td>59.80</td>
<td><strong>22.65</strong></td>
<td>0.6052</td>
<td>43.60</td>
<td>2.930</td>
</tr>
<tr>
<td>PULSE</td>
<td><strong>3.765</strong></td>
<td>65.90</td>
<td>20.81</td>
<td>0.5695</td>
<td>68.90</td>
<td>14.16</td>
</tr>
<tr>
<td>PSFRGAN</td>
<td>3.982</td>
<td>55.88</td>
<td>21.71</td>
<td>0.6173</td>
<td>50.66</td>
<td>3.500</td>
</tr>
<tr>
<td>GFPGAN</td>
<td>3.800</td>
<td>53.87</td>
<td>20.72</td>
<td>0.6001</td>
<td>50.50</td>
<td>4.160</td>
</tr>
<tr>
<td>GPEN</td>
<td>3.877</td>
<td><strong>47.39</strong></td>
<td>22.12</td>
<td>0.6152</td>
<td>47.70</td>
<td>3.333</td>
</tr>
<tr>
<td>VQFR</td>
<td><strong>3.350</strong></td>
<td>52.09</td>
<td>20.48</td>
<td>0.5699</td>
<td>50.60</td>
<td>3.666</td>
</tr>
<tr>
<td>Ours</td>
<td>4.411</td>
<td><strong>48.63</strong></td>
<td>21.71</td>
<td><strong>0.6247</strong></td>
<td><strong>22.30</strong></td>
<td><strong>1.467</strong></td>
</tr>
</tbody>
</table>

The best, second results are in **red** and **blue** respectively.
Quantitatively comparison on IMDB&COX (Captured LQs):

<table>
<thead>
<tr>
<th>IMDB</th>
<th>NIQE↓</th>
<th>FID↓</th>
<th>Age error (%)↓</th>
<th>Gender error (%)↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSFRGAN</td>
<td>4.316</td>
<td>39.04</td>
<td>50.72</td>
<td>1.790</td>
</tr>
<tr>
<td>GFPGAN</td>
<td>4.133</td>
<td><strong>32.30</strong></td>
<td>48.30</td>
<td>1.699</td>
</tr>
<tr>
<td>GPEN</td>
<td>4.719</td>
<td>54.13</td>
<td>49.25</td>
<td><strong>1.155</strong></td>
</tr>
<tr>
<td>VQFR</td>
<td><strong>3.540</strong></td>
<td>33.40</td>
<td>50.29</td>
<td>1.631</td>
</tr>
<tr>
<td>Ours</td>
<td>4.482</td>
<td>39.41</td>
<td><strong>26.57</strong></td>
<td><strong>0.929</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>COX</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PSFRGAN</td>
<td><strong>4.521</strong></td>
<td>88.94</td>
<td>61.60</td>
<td>21.19</td>
</tr>
<tr>
<td>GFPGAN</td>
<td>5.036</td>
<td>82.52</td>
<td>59.50</td>
<td><strong>20.59</strong></td>
</tr>
<tr>
<td>GPEN</td>
<td>4.713</td>
<td>84.01</td>
<td>64.30</td>
<td>22.59</td>
</tr>
<tr>
<td>VQFR</td>
<td><strong>4.190</strong></td>
<td><strong>70.00</strong></td>
<td>59.40</td>
<td>22.90</td>
</tr>
<tr>
<td>Ours</td>
<td>5.238</td>
<td>80.45</td>
<td><strong>29.50</strong></td>
<td><strong>8.80</strong></td>
</tr>
</tbody>
</table>

The best, second results are in red and blue respectively.
Human Interaction
Summarization

- **Attribute bias problem**: facial attributes (e.g., age and gender) of the restored faces could be dramatically different from the target faces.

- Owe it to two leading causes: *the lack of attribute information* and *bias in training data*.

- Propose **DebiasFR**, which faithfully preserves input attribute information and produces quality HQ faces. Supports human interaction attribute adjustment.
Thanks

Presenter: Zelin Li
12132338@mail.sustech.edu.cn