Imperfect ImaGANation: Implications of GANs Exacerbating Biases on Facial Data

Niharika Jain*, Alberto Olmo*, Sailik Sengupta, Lydia Manikonda, Subbarao Kambhampati

Motivation

• Wide adoption of GANs as a seemingly trustworthy data augmentation technique.
• Practitioners possibly unaware of Mode Collapse causing exacerbation of biases.

The Mode Collapse Problem

The diversity of the generated distribution is much lower than that of the training set due to the non-infinite capacity of the generator nor discriminator.

Evaluation

• Gathering and preprocessing of engineering professor headshots dataset from 47 U.S. universities
• Training and generation of new headshot distributions from 4 unconditional and 1 conditional GANs

Results

Human annotation

<table>
<thead>
<tr>
<th>Feminine</th>
<th>Masculine features</th>
<th>Non-white</th>
<th>White color</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original</td>
<td>20%</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>DCGAN</td>
<td>6.57%</td>
<td>92.67%</td>
<td></td>
</tr>
<tr>
<td>ProGAN</td>
<td>18%</td>
<td>82%</td>
<td></td>
</tr>
</tbody>
</table>

MS Face API

<table>
<thead>
<tr>
<th>Feminine</th>
<th>Masculine</th>
<th>Can't tell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>16.5%</td>
<td>80.4%</td>
</tr>
<tr>
<td>DCGAN</td>
<td>11.8%</td>
<td>87.1%</td>
</tr>
<tr>
<td>ProGAN</td>
<td>14.8%</td>
<td>82.9%</td>
</tr>
</tbody>
</table>