Look globally, age locally: Face aging with an attention mechanism

Haiping Zhu, Zhizhong Huang, Honming Shan, and Junping Zhang

School of Computer Science, Fudan University, China
hpzhu14@fudan.edu.cn

April 17, 2020
Outline

1 Background

2 Proposed Method

3 Experiments

4 Conclusion
Facial Aging

Many challenges: the intrinsic complexity of aging in nature and the insufficient labeled aging data;
Auto-encoder V.S Conditional GANs;
cGANs-based methods are more powerful;
Existing cGANs-based methods usually use pixel-wise loss to preserve identity consistency and keep background information.
Pixel-wise Loss

\[ \text{Loss} = \|X_{in} - X_{out}\|_2^2 \]  

- The input facial image and the output facial image are different if their age are different.
- Pixel-wise loss can keep the background information, but make the generated image more blurry.
- How to keep the background information and preserve identity consistency but without using the pixel-wise loss? **Attention Mechanism!**
Outline

1. Background
2. Proposed Method
3. Experiments
4. Conclusion
**Framework of AcGANs**

- **AcGANs**: Attention Conditional GANs with an attention mechanism.
- \( x_t = G(x_s, y_t) = (I - A) \otimes C + A \otimes x_s \)
Adversarial Loss \( \mathcal{L}_{adv}(G, D_I, x_s, y_s) \) is:

\[
\mathcal{L}_{adv} = \mathbb{E}_{x_s \sim P_{x_s}}[D_I(x_s)] - \mathbb{E}_{x_s \sim P_{x_s}}[D_I(G(x_s, y_t))] - \lambda_{gp}\mathbb{E}_{x \sim P_{x}}[(\|\nabla_x D_I(\tilde{x})\| - 1)^2],
\]

where \( \lambda_{gp} \) is a penalty coefficient.

Attention Loss \( \mathcal{L}_{att}(G, x_s, y_t) \) is:

\[
\mathcal{L}_{att} = \lambda_{TV}\mathbb{E}_{x_s \sim P_{x_s}} \left[ \sum_{i,j}^{h,w} [(A_{i+1,j} - A_{i,j})^2 + (A_{i,j+1} - A_{i,j})^2] \right] + \mathbb{E}_{x_t \sim P_{x_t}}[\|A\|_2],
\]

where \( A = G_A(x, y_t) \) and \( A_{i,j} \) is the \( i,j \) entry of \( A \). Besides, \( \lambda_{TV} \) is a penalty coefficient.
Objective Function of AcGANs

- **Age Classification Loss** $\mathcal{L}_{cls}(G, D_y, x_s, y_t, y_s)$ is:
  \[
  \mathcal{L}_{cls} = \mathbb{E}_{x_s \sim P_{x_s}} \left[ \ell(D_y(G(x_s, y_t)), y_t) + \ell(D_y(x_s), y_s) \right],
  \]
  (4)

  where $y_s$ is the label of input image $x_s$, $\ell(\cdot)$ corresponds to a softmax loss.

- **Final loss** is:
  \[
  \mathcal{L} = \lambda_{adv}\mathcal{L}_{adv}(G, D_I, x_s, y_t) + \lambda_{att}\mathcal{L}_{att}(G, x_s, y_t) \\
  + \lambda_{cls}\mathcal{L}_{cls}(G, D_y, x_s, y_t, y_s),
  \]
  (5)

  where $\lambda_{adv}$, $\lambda_{att}$ and $\lambda_{cls}$ are the hyper-parameters that control the relative importance of every loss term.
Outline

1. Background
2. Proposed Method
3. Experiments
4. Conclusion
Attention Results on Morph

- A is Attention Mask; C is Color Mask
Generative results on Morph
The proposed AcGANs can better preserve the details unrelated to face aging.
Quantitative Comparison results

- Estimated age distributions (in years) on Morph. Generic means that the mean value of each group is computed in the ground truth, while the number in brackets indicates the differences from generic mean age.

<table>
<thead>
<tr>
<th>Estimated Age Distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group</td>
</tr>
<tr>
<td>Generic</td>
</tr>
<tr>
<td>CAAE [4]</td>
</tr>
<tr>
<td>IPCGANs [5]</td>
</tr>
<tr>
<td>AcGANs</td>
</tr>
</tbody>
</table>

- Face verification results on Morph. The top is the verification confidence by AcGANs and the bottom is the verification rate for all methods.

<table>
<thead>
<tr>
<th>Verification Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group</td>
</tr>
<tr>
<td>10-20</td>
</tr>
<tr>
<td>21-30</td>
</tr>
<tr>
<td>31-40</td>
</tr>
<tr>
<td>41-50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Verification Rate (Threshold = 73.975, FAR = 1e-5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group</td>
</tr>
<tr>
<td>CAAE [4]</td>
</tr>
<tr>
<td>IPCGANs [5]</td>
</tr>
<tr>
<td>AcGANs</td>
</tr>
</tbody>
</table>
Outline

1. Background
2. Proposed Method
3. Experiments
4. Conclusion
Conclusion

- Proposed a cGAN-based model that can generate clear facial images with a high age accuracy and identity consistency.
- Proposed model is **simple**. It consists of only a generator and a discriminator sub-networks and it can be learned without additional pre-trained models.
- Using an **attention mechanism** to keep the background information and preserve the identity information, rather than using **pixel-wise loss**.
- GAN framework is difficult to train, no clear objective function to track.
- Code: [https://github.com/JensonZhu14/AcGAN](https://github.com/JensonZhu14/AcGAN)
The End! (Q&A)

Code: https://github.com/JensonZhu14/AcGAN