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

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.

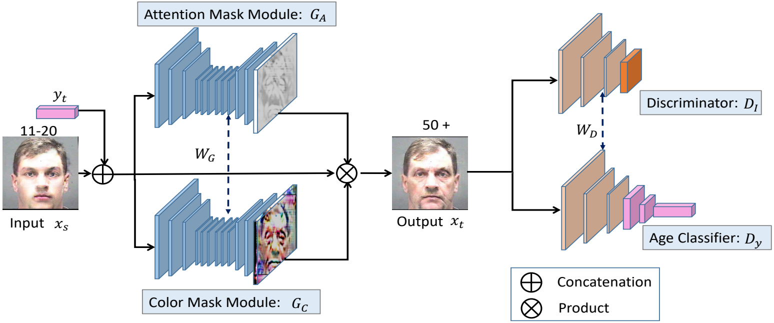
Pixel-wise Loss

$$Loss = ||X_{in} - X_{out}||_2^2 \tag{1}$$

- 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

Framework of AcGANs



- AcGANs: **A**ttention **C**onditional **GANs** with an attention mechanism.
- $\boldsymbol{x}_t = G(\boldsymbol{x}_s, \boldsymbol{y}_t) = (\boldsymbol{I} - \boldsymbol{A}) \otimes \boldsymbol{C} + \boldsymbol{A} \otimes \boldsymbol{x}_s$

Objective Function of AcGANs

- **Adversarial Loss** $\mathcal{L}_{adv}(G, D_I, \boldsymbol{x}_s, \boldsymbol{y}_s)$ is:

$$\begin{aligned}\mathcal{L}_{adv} =& \mathbb{E}_{\boldsymbol{x}_s \sim \mathbb{P}_{x_s}}[D_I(\boldsymbol{x}_s)] - \mathbb{E}_{\boldsymbol{x}_s \sim \mathbb{P}_{x_s}}[D_I(G(\boldsymbol{x}_s, \boldsymbol{y}_t))] \\ & - \lambda_{gp} \mathbb{E}_{\widetilde{\boldsymbol{x}} \sim \mathbb{P}_{\widetilde{x}}}\left[(\|\nabla_{\widetilde{\boldsymbol{x}}} D_I(\widetilde{\boldsymbol{x}})\| - 1)^2\right],\end{aligned} \tag{2}$$

  where λ_{gp} is a penalty coefficient.

- **Attention Loss** $\mathcal{L}_{att}(G, \boldsymbol{x}_s, \boldsymbol{y}_t)$ is:

$$\begin{aligned}\mathcal{L}_{att} =& \lambda_{TV} \mathbb{E}_{\boldsymbol{x}_s \sim \mathbb{P}_{x_s}}\left[\sum_{i,j}^{h,w}[(\boldsymbol{A}_{i+1,j} - \boldsymbol{A}_{i,j})^2 + (\boldsymbol{A}_{i,j+1} - \boldsymbol{A}_{i,j})^2]\right] \\ & + \mathbb{E}_{\boldsymbol{x}_t \sim \mathbb{P}_{x_t}}\left[\|\boldsymbol{A}\|_2\right],\end{aligned} \tag{3}$$

  where $\boldsymbol{A} = G_A(\boldsymbol{x}, \boldsymbol{y}_t)$ and $\boldsymbol{A}_{i,j}$ is the i, j entry of $\boldsymbol{A}$. Besides, λ_{TV} is a penalty coefficient.

Objective Function of AcGANs

- **Age Classification Loss** $\mathcal{L}_{cls}(G, D_y, \boldsymbol{x}_s, \boldsymbol{y}_t, \boldsymbol{y}_s)$ is:

$$\mathcal{L}_{cls} = \mathbb{E}_{\boldsymbol{x}_s \sim \mathbb{P}_{x_s}}\Big[\ell(D_y(G(\boldsymbol{x}_s, \boldsymbol{y}_t)), \boldsymbol{y}_t) + \ell(D_y(\boldsymbol{x}_s), \boldsymbol{y}_s)\Big], \tag{4}$$

  where $\boldsymbol{y}_s$ is the label of input image $\boldsymbol{x}_s$, $\ell(\cdot)$ corresponds to a softmax loss.

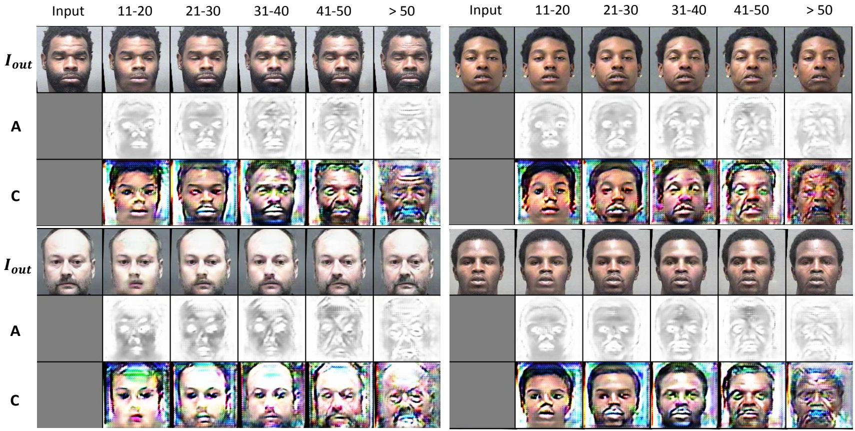
- **Final loss** is:

$$\begin{aligned}\mathcal{L} =& \lambda_{adv}\mathcal{L}_{adv}(G, D_I, \boldsymbol{x}_s, \boldsymbol{y}_t) + \lambda_{att}\mathcal{L}_{att}(G, \boldsymbol{x}_s, \boldsymbol{y}_t) \\ &+ \lambda_{cls}\mathcal{L}_{cls}(G, D_y, \boldsymbol{x}_s, \boldsymbol{y}_t, \boldsymbol{y}_s),\end{aligned} \tag{5}$$

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

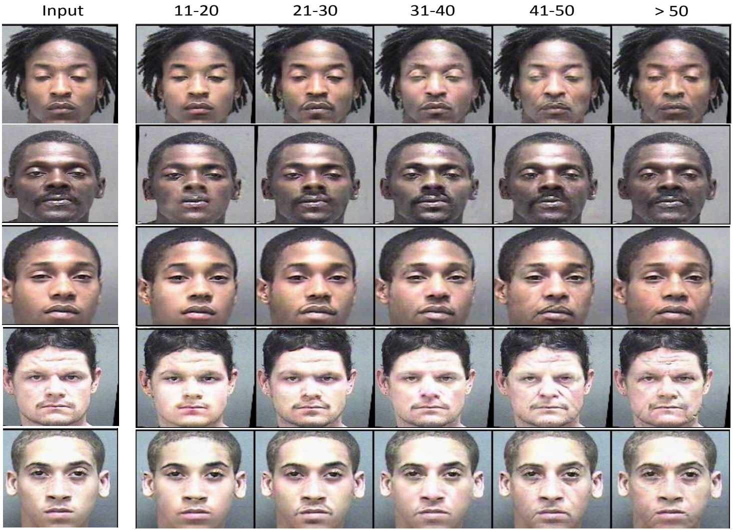
Outline

Attention Results on Morph

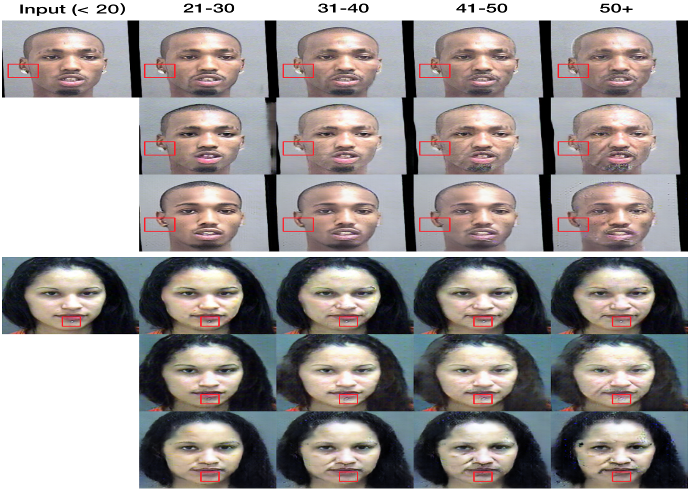


- **A** is Attention Mask; **C** is Color Mask

Generative results on Morph



More detailed results



- 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.

Estimated Age Distributions				
Age group	21-30	31-40	41-50	50+
Generic	25.12	35.43	44.72	54.88
CAAE [4]	24.31(0.81)	31.02(4.41)	39.03(5.69)	47.84(7.04)
IPCGANs [5]	22.38(2.74)	27.53(7.90)	36.41(8.31)	46.42(8.46)
AcGANs	**25.92(0.80)**	**36.49(1.06)**	**40.59(4.13)**	**47.88(7.00)**

	21-30	31-40	41-50	50+
	Verification Confidence			
10-20	95.36	94.78	94.74	93.44
21-30	-	95.37	95.28	94.11
31-40	-	-	95.65	94.72
41-50	-	-	-	95.26
	Verification Rate (Threshold = 73.975, FAR = 1e-5)			
CAAE [4]	99.38	97.82	92.72	80.56
IPCGANs [5]	**100**	**100**	**100**	**100**
AcGANs	**100**	**100**	**100**	**100**

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

Outline

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 machanism** to keep the backgound 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

The End! (Q&A)

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