Learning Face Age Progression: A Pyramid Architecture of GANs

Cheng Guan

July 20, 2018

1. Generator

To achieve aging effects while simultaneously maintaining person-specific information, a compound critic is exploited, which incorporates the traditional squared Euclidean loss in the image space, the GAN loss that encourages generated faces to be indistinguishable from the training elderly faces in terms of age, and the identity loss minimizing the input-output distance in a high-level feature representation which embeds the personalized characteristics. As shown in Fig. 1.

Synthesizing age progressed faces only requires a forward pass through G. The generator network is combination of encoder and decoder. With the input young face, it first exploits three strided convolutional layers to encode it to a latent space, capturing the facial properties that tend to be stable w.r.t. the elapsed time, followed by four residual blocks [2] modeling the common structure shared by the input and output faces, similar to the settings in [3]. Rather than using the max-pooling and upsampling layers to calculate the feature maps, the authors employ the 3×3 convolution kernels with a stride of 2, ensuring that every pixel contributes and the adjacent pixels transform in a synergistic manner. All the convolutional layers are followed by Instance Normalization and ReLU non-linearity activation. Paddings are added to the layers to make the input and output have exactly the same size.

2. Discriminator

The system critic incorporates the prior knowledge of the data density of the faces from the target age cluster, and a discriminative network D is thus introduced, which outputs a scalar $D(x)$ representing the probability that x comes from the data. The distribution of the generated faces P_g (They denote the distribution of young faces as $x \sim P_{young}$, then $G(x) \sim P_g$) is supposed to be equivalent to the distribution P_{old} when optimality is reached. Supposing that we follow the classic GAN [1], which uses a binary cross entropy classification, the process of training D amounts to minimizing the loss:

$$\begin{aligned} \mathcal{L}_{GAN\_D} = & -\mathbb{E}_{x \in P_{young}(x)} \log\left[1 - D\left(G\left(x\right)\right)\right] \\ & -\mathbb{E}_{x \in P_{old}(x)} \log\left[D\left(x\right)\right] \end{aligned} \tag{1}$$

It is always desirable that G and D converge coherently;however, D frequently achieves the distinguishability faster in practice, and feeds back vanishing gradients for G to learn, since the JS divergence is locally saturated. Recent studies, (i.e.) the Wasserstein GAN, the Least Squares GAN [4], and the Loss-Sensitive GAN [5], reveal that the

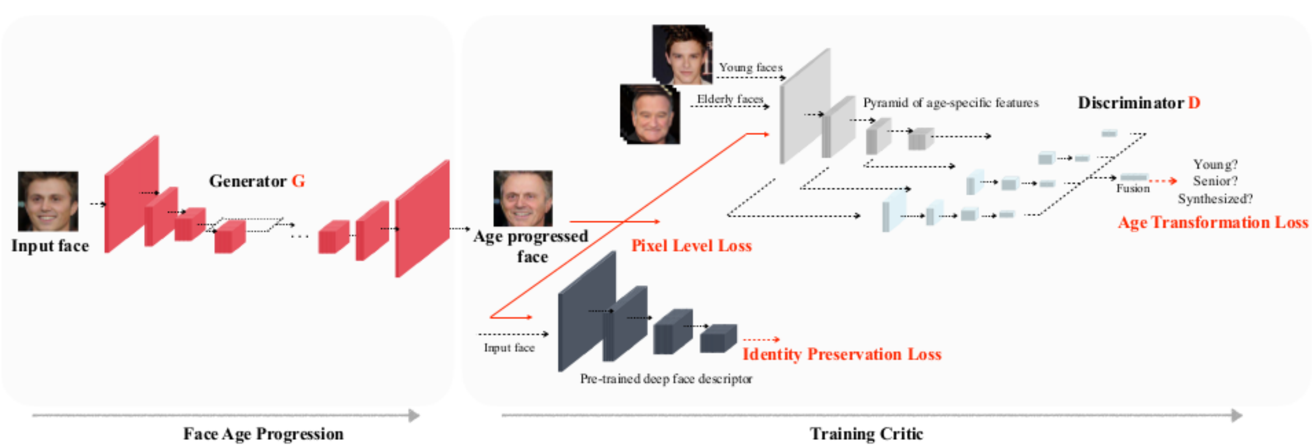

Figure 1. Framework of the proposed age progression method.

most fundamental issue lies in how exactly the distance between sequences of probability distributions is defined.

3. Objective

Besides the specially designed age-related GAN critic and the identity permanence penalty, a pixel-wise L2 loss in the image space is also adopted for further bridging the input-output gap, *e.g.*, the color aberration, which is formulated as shown in Eq.2:

$$\mathcal{L}_{pixel} = \frac{1}{W \times H \times C} \left\| G\left(x\right) - x \right\|_2^2 \tag{2}$$

where x denotes the input face and W, H, and C correspond to the image shape.

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *NIPS*, 2014. 1

[2] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 1

[3] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, 2016. 1

[4] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley. Least squares generative adversarial networks. In *ICCV*, 2017. 2

[5] G.-J. Qi. Loss-sensitive generative adversarial networks on lipschitz densities. *arXiv preprint arXiv:1701.06264*, 2017. 2