Finding Tiny Faces in the Wild with Generative Adversarial Network

Cheng Guan

July 16, 2018

1. GAN

In this section, the author introduce their method. Firstly, they give a description on the classical GAN network. Then, the whole architecture of their method as shown in Fig. 1.The objective function is defined as shown in Eq.1.

$$\begin{aligned}\mathcal{L}(G, D) =& \mathbb{E}_{x\sim P_{data}(x)}\left[\log D_\theta(x)\right] + \\ & \mathbb{E}_{z\sim P_z(z)}\left[\log\left(1 - D_\theta\left(G_w(z)\right)\right)\right]\end{aligned} \tag{1}$$

z is the random noise and x denotes the real data, and denote the parameters of G and D respectively. Here, G tries to minimize the objective function and adversarial D tries to maximize it as shown in Eq.2:

$$\arg\min_G \max_D \mathcal{L}_{GAN}(G, D) \tag{2}$$

Similar to [1], the authors design a generator network G_{w_G} which is optimized in an alternative method along with a discriminator network D_{θ_D} to solve the small face super-resolution and classification problem, which is defined as shown in Eq.3:

$$\begin{aligned}\arg\min_{w_G}\max_{\theta_D} & \mathbb{E}_{(I^{HR},y)\sim P_{train}(I^{HR},y)}\left[\log D_{\theta_D}\left(I^{HR}, y\right)\right] + \\ & \mathbb{E}_{(I^{LR},y)\sim P_G(I^{LR},y)}\left[\log\left(1 - D_{\theta_D}\left(G_{w_G}\left(I^{HR}, y\right)\right)\right)\right]\end{aligned} \tag{3}$$

where I^{LR} denotes face candidates with low-resolution, I^{HR} represents the face candidates with high-resolution, and y is the label (*i.e.* face or non-face).

2. Network Architecture

Generator network. As shown in Fig. 1, the authors adopt a deep CNN architecture which has shown effectiveness for image super-solution. There are two fractionally-strided convolutional layers [2] (*i.e.* de-convolutional layer) in the network, and each de-convolutional layer consists of learned kernels which perform up-sampling a low-resolution image to a 2× super-resolution image.

Discriminator network. The authors use VGG19 [3] as their backbone network in the discriminator. To avoid too many down-sampling operations for the small blurry faces, they remove the max-pooling from the "conv5" layer. Moreover, the authors replace all the fully connected layer (*i.e.* f_{c6}, f_{c7}, f_{c8}) with two parallel fully connected layers $f_{c_{GAN}}$ and $f_{c_{clc}}$. The input is the super-resolution image, the output of f_c GAN branch is the probability of the input being a real image, and the output of the $f_{c_{clc}}$ is the probability of the input being a face.

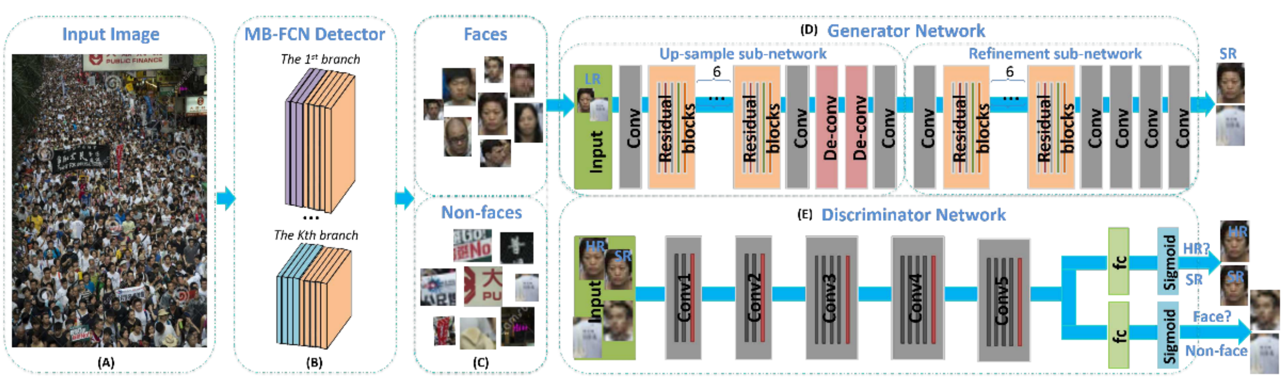

Figure 1. The pipeline of the proposed tiny face detector system.

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] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015. 1

[3] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1