Finding Tiny Faces in the Wild With Generative Adversarial Network

Wenjie Niu

June 10,2018

Abstract

[2]Face detection techniques have been developed for decades, and one of remaining open challenges is detecting small faces in unconstrained conditions. The reason is that tiny faces are often lacking detailed information and blurring. In this paper, we proposed an algorithm to directly generate a clear high-resolution face from a blurry small one by adopting a generative adversarial network (GAN). Toward this end, the basic GAN formulation achieves it by super-resolving and refining sequentially (e.g. SR-GAN and cycle-GAN). However, we design a novel network to address the problem of super-resolving and refining jointly. We also introduce new training losses to guide the generator network to recover fine details and to promote the discriminator network to distinguish real vs. fake and face vs. non-face simultaneously. Extensive experiments on the challenging dataset WIDER FACE demonstrate the effectiveness of our proposed method in restoring a clear high-resolution face from a blurry small one, and show that the detection performance outperforms other state-of-the-art methods.

1. Introduction

Face detection is a fundamental and important problem in computer vision, since it is usually a key step towards many subsequent face-related applications, including face parsing, face verification, face tagging and retrieval, etc. Face detection has been widely studied over the past few decades and numerous accurate and efficient methods have been proposed for most constrained scenarios. Recent works focus on faces in uncontrolled settings, which is much more challenging due to the significant variations in scale, blur, pose, expressions and illumination. A thorough survey on face detection methods can be found in [6].

Modern face detectors have achieved impressive results on the large and medium faces, however, the performance on small faces is far from satisfactory. The main difficulty for small face (e.g. 10×10 pixels) detection is that small faces lack sufficient detailed information to distinguish them from the similar background, e.g. regions of partial faces or hands. Another problem is that modern CNN-based face detectors use the down-sampled convolutional (conv) feature maps with stride 8, 16 or 32 to represent faces, which lose most spatial information and are too coarse to describe small faces. To detect small faces, [5] directly up-samples images using bi-linear operation and exhaustively searches faces on the up-sampled images. However, this method will increase the computation cost and the inference time will increase significantly too. Moreover, images are often zoomed in with a small upscaling factors (2 at most) in [5], otherwise, artifacts will be generated. Besides, [1], [3], [4], [7] use the intermediate conv feature maps to represent faces at specific scales, which keeps the balance between the computation burden and the performance. However, the shallow but fine-grained intermediate conv feature maps lack discrimination, which causes many false positive results. More importantly, these methods take no care of other challenges, like blur and illumination.

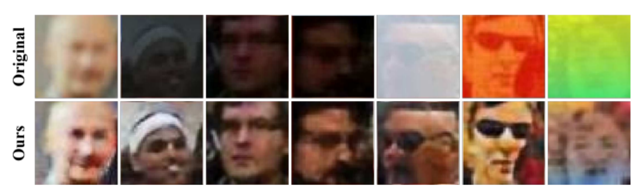

Figure 1. Some examples of the clear faces generated by our generator network from the blurry ones. The top row shows the small faces influenced by blur and illumination, and the bottom row shows the clearer faces generated by our method. The lowresolution images in the top row are re-sized for visualization.

Influence of the GAN. Table 1 (the 1 st and the 5 th row) shows the detection performance (AP) of the baseline detector and our method on WIDER FACE validation set. Our baseline detector is a multi-branch RPN face detector with skip connection of feature maps, and please refer to [1] more detailed information . From Table 1 we observe that the performance of our detector outperforms the baseline detector by a large margin (1.5% in AP) on the Hard subset. The reason is that the baseline detector performs the down-sampling operations (i.e. convolution with stride 2) on the small faces. The small faces themselves contain limited information, and the majority of the detailed information will be lost after several convolutional operations. For example, the input is a 1616 face, and the result is 11 on the C4 feature map and nothing is reserved on the C5 feature map. Based on those limited features, it is normal to get the poor detection performance. In contrast, our method first learns a super-resolution image and then refines it, which solves the problem that the original small blurry faces lack detailed information and blurring simultaneously. Based on the super-resolution images with fine details, the boosting of the detection performance is inevitable.

Table 1. Performance of the baseline model trained with and without GAN, refinement network, adversarial loss and classification loss on the WIDER FACE invalidation set.

Method	Easy	Medium	Hard
Baseline[1]	0.932	0.922	0.858
w/o Refinement Network	0.940	0.929	0.863
w/o adv loss	0.935	0.925	0.867
w/o clc loss	0.936	0.927	0.865
Ours(Baseline+MES+adv+clc)	**0.944**	**0.933**	**0.873**

Influence of the refinement network. From Table 1 (the 2 nd and 5 th row), we see that the AP performance increases by 1% on the Hard subset by adding the refinement sub-network to the generator network. Interestingly, the performances of Easy and Medium subset also have an improvement (0.4%). We visualize the reconstructed faces from the generator network and find that our refinement network can reduce the influence of illumination and blur as shown in Figure 1. In some cases, the baseline detector fails to detect the faces if those faces are heavily blurred or illuminated. However, our method reduces influence of such attributions and can find these faces successfully. Here, we would like to note that our framework is not specific and any off-the-shelf face detectors can be used as our baseline.

References

[1] Y. Bai and B. Ghanem. Multi-branch fully convolutional network for face detection. *arXiv preprint arXiv:1707.06330*, 2017. 1, 2

[2] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem. Finding tiny faces in the wild with generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 2018. 1

[3] H. Jiang and E. Learnedmiller. Face detection with the faster r-cnn. In *IEEE International Conference on Automatic Face and Gesture Recognition*, May 2017. 1

[4] S. Wan, Z. Chen, T. Zhang, B. Zhang, and K. Wong. Bootstrapping face detection with hard negative examples. *arXiv preprint arXiv:1608.02236*, 2016. 1

[5] X. Xu, D. Sun, J. Pan, Y. Zhang, H. Pfister, and M. H. Yang. Learning to super-resolve blurry face and text images. In *IEEE International Conference on Computer Vision*, Oct. 2017. 1

[6] S. Zafeiriou, C. Zhang, and Z. Zhang. A survey on face detection in the wild: Past, present and future. *Computer Vision and Image Understanding*, 2015. 1

[7] C. Zhu, Y. Zheng, K. Luu, and M. Savvides. CMS-RCNN: Contextual multi-scale region-based cnn for unconstrained face detection. *arXiv preprint arXiv:1606.05413*, 2016. 1