LaTeX Joint Training of Cascaded CNN for Face Detection

Hongwei Qin Junjie Yan Xiu Li Xiaolin Hu

Abstract

Cascade has been widely used in face detection, where classifier with low computation cost can be firstly used to shrink most of the background while keeping the recall. The cascade in detection is popularized by seminal Viola-Jones framework and then widely used in other pipelines, such as DPM and CNN. However, to our best knowledge, most of the previous detection methods use cascade in a greedy manner, where previous stages in cascade are fixed when training a new stage. So optimizations of different CNNs are isolated. In this paper, we propose joint training to achieve end-to-end optimization for CNN cascade. We show that the back propagation algorithm used in training CNN can be naturally used in training CNN cascade. We present how jointly training can be conducted on naive CNN cascade and more sophisticated region proposal network (RPN) and fast R-CNN. Experiments on face detection benchmarks verify the advantages of the joint training

1. ExperimentsofjointlytrainedfasterR-CNN

Benchmark	Separate	Joint
AFW	97.0%	98.7%
FDDB	89.7%	91.2%

Table 1. Comparison of training methods of RPN + F-RCNN

As shown in Table. 1, with our presented RPN + FR-CNN (fast R-CNN) joint training pipeline, the AP (average precision) on AFW is 98.7%, compared to the baseline result 97.0% trainedwith4-stagetrainingmethodproposed in[21]. OnFDDB,therecall(1000falsepositives)is 91.2% and 89.7%. For the F-RCNN branch, the final joint training loss decreases 64% compared to separate training. In joint RPN + F-RCNN, the detection results mostly have much higher confidence scores than separate training results, which have lower confidence scores because of FRCNN domination in convolution layers.

2. Jointloss

Each branch has a face v.s. non-face classfication loss and a bounding-box regression loss. Adding them with loss weights, we get the joint loss function:

$$L_{point} = \lambda_1 L_{x12} + \lambda_2 L_{x24} + \lambda_3 L_{x48}$$

3. AFW results

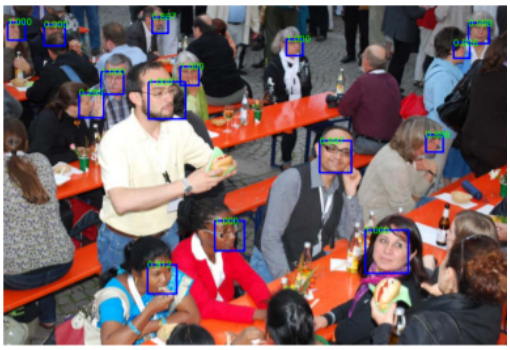

Figure 1. Qualitative results of FaceCraft on AFW.

Examples of detection results are shown in Fig. 4. [1] In our test results, non-frontal face bounding-box centred on the nose, which isconsistentwithourtraininggroundtruthshowninFig.3. While in AFW ground-truth, nose is on the bounding-box edge [2].

References

[1] M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool. Face detection without bells and whistles. *In Computer Vision–ECCV*, 13(1):720–735, 2014.

[2] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. *In Computer Vision andPatternRecognition(CVPR),*, 12(1):2879–2886, 2012.