Real-Time Rotation-Invariant Face Detection with Progressive Calibration Networks

Chaonan Song

May 5, 2018

1 Experiments

In the following part, I will describe the implementation of PCN. Then, In order to better evaluate the effectiveness of PCN and analyze PCN's accuracy and calculation speed in depth, they use the world's most challenging two data sets: FDDB and WIDER FACE.

1.1 Implementation Details

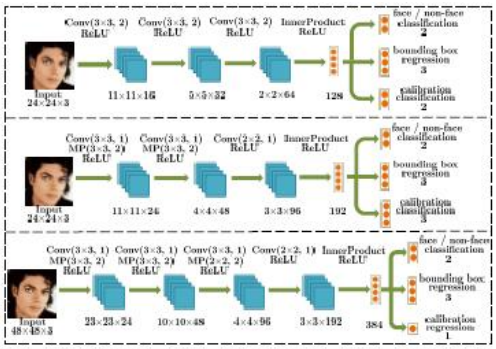

Figure 1: The detailed CNN structures of three stages in our proposed PCN method. Conv, MP, InnerProduct, and ReLU mean convolution layer, max pooling layer, inner product layer, and relu layer, respectively. (k ×k, s) represents that the kernel size is k and the stride is s.

Their network architecture is shown in Figure 1. The PCN consists of three CNNs from largest to smallest. They use the WIDER FACE training set for training, they adjusted the noted faces to squares. The network is optimized by back-propagating Stochastic Gradient Descent (SGD) with a maximum iteration setting of 10^5. They adopt the step strategy in Caffe [4] to adjust learning rate. The learning rate for the first iteration was 7×10^4 and then revised to 10^{-3}, then reduced to 10^{-4} again. Weight Decay is set to 5×10^{-4} and Momentum Decay is set to 0.9. All layers are initialized using a Gaussian distribution with a standard deviation of 0.01 in order to achieve stable convergence. In each small branch, the proportion of positive, negative, and suspect samples is approximately 2:2:1. The experiments run on desktop computer that have 3.4GHz CPU and GTX Titan X GPU.

1.2 Methods for Comparison

They compared the three methods mentioned earlier. 1) *Data Augmentation*: Training with data augmentation Advanced detection data models including Faster R-CNN [7], RFCN [2], and SSD500 [6], randomly rotating training images throughout the full range of $[-180°, 180°]$. They use base networks is VGG16 [9],VGGM [1], and ResNet-50 [3]. They also use the same networks as their PCN to implement cascade CNN [5] face detector, and training with Data Augmentation. 2) *Divide-and-Conquer*: They implemented an upright face detector based on cascade CNN and run this detector four times on images rotated to $0°$, $90°$, $-90°$, and $180°$ to form a rotational invariant face detector. 3) *Rotation Router* [8]

is implemented for comparison, in which the router network first estimates the orientation of faces in the range of up, down, left, right and rotate it to upright respectively. The routing network of *Rotation Router* shares the same structure as the PCN uses WIDER FACE for training and adjusts the annotation surface to square.

References

[1] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *Computer Science*, 2014.

[2] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-fcn: Object detection via region-based fully convolutional networks. 2016.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, june 2016.

[4] Jia, Yangqing, Shelhamer, Evan, Donahue, Jeff, Karayev, Sergey, Long, and Jonathan. Caffe: Convolutional architecture for fast feature embedding. *In ACM International Conference on Multimedia (MM)*, pages 675–678, 2014.

[5] Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua. A convolutional neural network cascade for face detection. In *In The IEEE Conference on Computer Vision and Pattern Recognition*, pages 5325–5334, 2015.

[6] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. In *In European Conference on Computer Vision(ECCV)*, pages 21–37, 2016.

[7] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *In Advances in Neural Information Processing Systems(NIPS)*, pages 91–99, 2015.

[8] H. A. Rowley, S. Baluja, and T. Kanade. Rotation invariant neural network-based face detection. In *In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 38–44, 1998.

[9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *Computer Science*, 2014.