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


Xuepeng Shi
xuepeng.shi@vipl.ict.ac.cn

Shiguang Shan
shiguang.shan@vipl.ict.ac.cn



## Abstract

*Rotation-invariant face detection, i.e. detecting faces with arbitrary rotation-in-plane (RIP) angles, is widely required in unconstrained applications but still remains as a challenging task, due to the large variations of face appearances. Most existing methods compromise with speed or accuracy to handle the large RIP variations.To address this problem more efficiently, we propose Progressive Calibration Networks (PCN) to perform rotation-invariant face detection in a coarse-to-fine manner. PCN consists of three stages, each of which not only distinguishes the faces from non-faces, but also calibrates the RIP orientation of each face candidate to upright progressively. By dividing the calibration process into several progressive steps and only predicting coarse orientations in early stages, PCN can achieve precise and fast calibration. By performing binary classification of face vs. non-face with gradually decreasing RIP ranges, PCN can accurately detect faces with full 360° RIP angles. Such designs lead to a real-time rotationinvariant face detector. The experiments on multi-oriented FDDB and a challenging subset of WIDER FACE containing rotated faces in the wild show that our PCN achieves quite promising performance. A demo of PCN can be available at* `https://github.com/Jack-CV/PCN`.


## 1. Introduction

Face detection serves as an important component in computer vision systems which aim to extract information from face images. Practical applications, such as face recognition and face animation, all need to quickly and accurately detect faces on input images in advance. Same as many other vision tasks, the performance of face detection has been substantially improved by Convolutional Neural Network(CNN) [[1],[2]].The CNN-based detectors enjoy the natural advantage of strong capability in non-linear feature learning. However, most works focus on designing an effective detector for generic faces without considerations for specific scenarios, such as detecting faces with full rotation-in-plane (RIP) angles as shown in Figure 1. They become less satisfactory in such complex applications. Face detection in full RIP, i.e. rotation-invariant face detection, is quite challenging, because faces can be captured almost from any RIP angle, leading to significant divergence in face appearances. An accurate rotationinvariant face detector can greatly boost the performance of subsequent process, e.g. face alignment and face recognition.

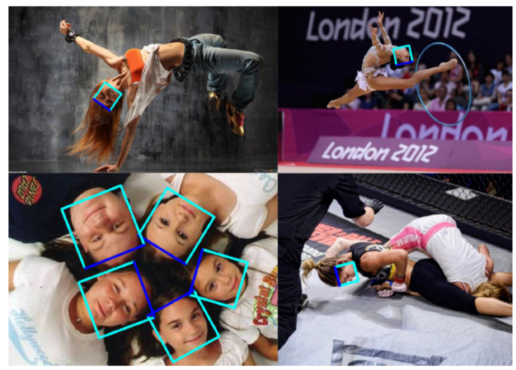


Figure 1. Many complex situations need rotation-invariant face detectors. The face boxes are the outputs of our detector, and the blue line indicates the orientation of faces.

### 1.1. PCN-1 in 1st stage

For each input window $x$, PCN-1 has three objectives: face or non-face classification, bounding box regression, and calibration, formulated as follows:

$$[f, g, t] = F_1(x) \tag{1}$$

where $F_1$ is the detector in the first stage structured with a small CNN. The $f$ is face confidence score, $t$ is a vector representing the prediction of bounding box regression, and $g$ is orientation score. The first objective, which is also the primary objective, aims for distinguishing faces from non-faces with softmax loss as follows:

$$L_{cls} = y log f + (1 - y) log(1 - f) \tag{2}$$

where $y$ equals 1 if $x$ is face, otherwise is 0.

| Method | Recall rate at 100 FP on FDDB | | | | | Speed | | Model Size |
|---|---|---|---|---|---|---|---|---|
| | Up | Down | Left | Right | Ave | CPU | GPU | |
| Divide-and-Conquer | 85.5 | 85.2 | 85.5 | 85.6 | 85.5 | 15FPS | 20FPS | 2.2M |
| Rotation Router | 85.4 | 84.7 | 84.6 | 84.5 | 84.8 | 12FPS | 15FPS | 2.5M |
| Cascade CNN | 85.0 | 84.2 | 84.7 | 85.8 | 84.9 | 31FPS | 67FPS | 4.2M |
| Cascade CNN | 85.8 | 85.0 | 84.9 | 86.2 | 85.5 | 16FPS | 30FPS | 4.7M |
| SSD500 | 86.3 | 86.5 | 85.5 | 86.1 | 86.1 | 1FPS | 20FPS | 95M |
| Faster R-CNN | 84.2 | 82.5 | 81.9 | 82.1 | 82.7 | 1FPS | 20FPS | 350M |
| Faster R-CNN | 87.0 | 86.5 | 85.2 | 86.1 | 86.2 | 0.5FPS | 10FPS | 547M |
| R-FCN | 87.1 | 86.6 | 85.9 | 86.0 | 86.4 | 0.8FPS | 15FPS | 123M |
| PCN (ours) | 87.8 | 87.5 | 87.1 | 87.3 | 87.4 | 29FPS | 63FPS | 4.2M |

Table 1. Speed and accuracy comparison between different methods. The FDDB recall rate (%) is at 100 false positives.

## References


[1] K. He J. Dai, Y. Li and J. Sun. R-FCN. Object detection via region-based fully convolutional networks. *Journal of Foo*, 12(1):234–778, 2016.

[2] P. Welinder P. Dollar and P. Perona. Cascaded pose regression. *Journal of Foo*, 13(1):234–778, 2010.