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

Chaonan Song

May 7, 2018

1 Experiments

1.1 Benchmark Datasets

Multi-Oriented FDDB [3] include 5,171 labeled faces, those faces collected from 2,845 news headlines. FDDB is challenging in some means that the performance of faces have great changes in view, skin color, facial expression, illumination, occlusion, resolution, etc. However, most faces are upright because it collected from news headlines. In order to better evaluate the performance of Rotation-Invariant detector, FDDB faces are rotate in $-90°$, $90°$, $180°$ respectively, forming a multi-oriented version of FDDB.In this task, initial FDDB are called FDDB-up, and others are called FDDB-left, FDDB-light, FDDB-down according to their rotated angles. In order to evaluate the performance of Rotation-Invariant comprehensively, detectors are evaluated respectively on multi-oriented FDDB. They use official evaluation tools to obtain the ROC curves. They ignore the RIP angle of detection box, and simply use horizontal boxes for evaluation. Rotation WIDER FACE [8] include faces have high level in scale, pose and occlusion. They manual select some images that contain rotation faces from WIDER FACE to obtain a rotation subset with 370 images and 987 rotation flat, as shown in Figure 1. Due to WIDER FACE test sets do not provide the ground-truth faces, they manually annotate the faces in this subset following the WIDER FACE annotation rules.

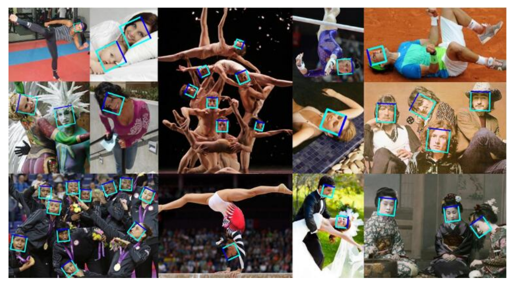

Figure 1: Our PCNs detection results on rotation WIDER FACE

1.2 Evaluation Results

1.2.1 Results of Rotation Calibration

For their PCN, the direction classification for the first and second phase are 95% and 96% respectively. The third stage calibration error is $8°$. In [7], The orientation classification accuracy of router network is 90%, which shows that their progressive calibration mechanism can achieve better orientation classification accuracy. The mean error of the router network in continuous angle regression manner is quite large, due to the fine regression task is too challenging, as shown in Figure 2.

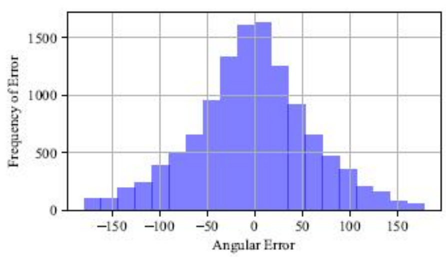

Figure 2: Histogram of angular error in the router network (in degrees)

1.2.2 Results on Multi-Oriented FDDB

They evaluate the rotation-invariant face detectors mentioned above in terms of ROC curves on multi-oriented FDDB, shown in Figure 3. As the picture shows, Their PCN performed well, defeating all other methods, and the performance of the PCN far outperformed the cascaded CNN with little extra time cost, thanks to a robust and accurate coarse-to-fine calibration process.

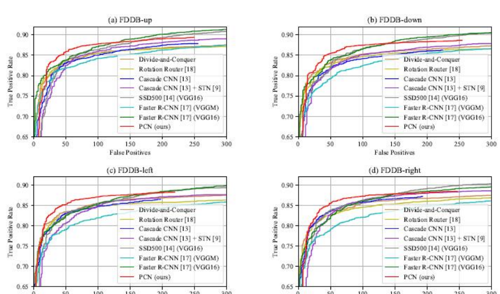

Figure 3: Histogram of angular error in the router network (in degrees)

1.2.3 Results on Rotation WIDER FACE

Moreover, their PCN are compared with the existing methods on the more challenging rotation WIDER FACE, as shown in Figure 4. their PCN achieves quite promising performance, which demonstrates the effectiveness of PCN. Some detection results can be viewed Figure 1.

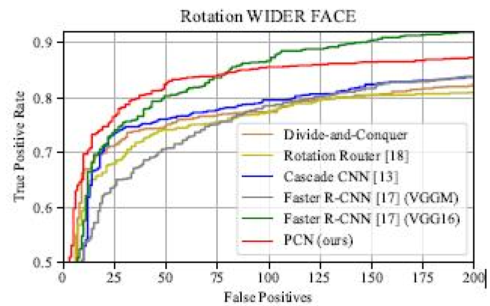

Figure 4: Histogram of angular error in the router network (in degrees)

1.2.4 Speed and Accuracy Comparison

Their PCN aims at accurate rotation-invariant face detection with low time cost. They compare PCNs speed with other rotation-invariant face detectors on standard 640 × 480 VGA images with 40 × 40 minimum face size with low time cost. The speed results and the recall rate at 100 false positives on multi-oriented FDDB are shown in Table **??**. As can be seen, their PCN can run with almost the same speed as Cascade CNN, and PCN runs much faster than Faster R-CNN (VGG16), SSD500(VGG16), and R-FCN (ResNet-50) with better detection accuracy, demonstrating the their PCN are best in both accuracy and speed.

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 [7]	85.4	84.7	84.6	84.5	84.8	12FPS	15FPS	2.5M
Cascade CNN [4]	85.0	84.2	84.7	85.8	84.9	31FPS	67FPS	4.2M
Cascade CNN [4] + STN [2]	85.8	85.0	84.9	86.2	85.5	16FPS	30FPS	4.7M
SSD500 [5] (VGG16)	86.3	86.5	85.5	86.1	86.1	1FPS	20FPS	95M
Faster R-CNN [6] (VGGM)	84.2	82.5	81.9	82.1	82.7	1FPS	20FPS	350M
Faster R-CNN [6] (VGG16)	87.0	86.5	85.2	86.1	86.2	0.5FPS	10FPS	547M
R-FCN [1] (ResNet-50)	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] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-fcn: Object detection via region-based fully convolutional networks. 2016.

[2] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks. *In Neural Information Processing Systems (NIPS)*, pages 2017–2025, 2015.

[3] Vidit Jain and Erik Learned-Miller. Fddb: A benchmark for face detection in unconstrained settings. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst, 2010.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] Shuo Yang, Ping Luo, Change Loy Chen, and Xiaoou Tang. Wider face: A face detection benchmark. In *Computer Vision and Pattern Recognition*, pages 5525–5533, 2016.