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

Chaonan Song

May 3, 2018

To summarize the contents of the previous section, I made a table as follows

PCN	PCN-1 in 1^{st} stage
Objective 1	Face or non-face classification: aims for distinguishing faces from non-faces with softmax loss
Objective 2	Bounding box regression: attempts to regress the fine bounding box
Objective 3	calibration: aims to predict the coarse orientation of the face candidate in abinary classification manner

Table 1: PCN-1 in 1^{st} stage

1 Progressive Calibration Networks (PCN)

1.1 PCN-2 in 2^{nd} stage

Similar to PCN-1, PCN-2 in second stage will more accurately distinguish between face and non-face, regressing the bounding box, and calibration the face candidates. The different between them is the coarse orientation in PCN-2 is a ternary classification of the RIP angle range, i.e. [−90°, −45°], [−45°, 45°], [45°,90°]. Rotation calibration is conducted with the predicted RIP angle in the second stage:

$$id = \arg\max_i g_i,$$
$$\theta_2 = \begin{cases} -90°, & id = 0 \\ 180°, & id = 1 \\ 90°, & id = 2 \end{cases} \tag{8}$$

where g0, g1, and g2 are the predicted ternary orientation classification scores. Accordingly ,the face candidates are rotate in −90°, 0°, 90°. After the second stage, the RIP angle are decreased from [−90°, 90°] to [−45°, 45°]. In the training phase of second stage, The original image rotates uniformly within [−90°, 90°] and filter the negative samples that via the training of PCN-1. Than in order to calibration the positive samples, negative samples and suspected samples are called 0, 1, 2, respectively. The samples those RIP angle over this range above are not contribute to the calibration.

1.2 PCN-3 in 2^{rd} stage

After the second stage, the RIP angle of all face candidates are calibrate to the range [−45°, 45°]. In the third stage PCN is as easy to detect human faces as most existing face detectors, and returns to the bounding box. Because the RIP angle in the previous phase has been reduced to a very small range, PCN-3 can return to the exact RIP direction, rather than rough. In the end, the RIP angle of face candidates i.e. θ_{RIP} can be obtained by accumulating the angles of all stages as below:

$$\theta_{RIP} = \theta_1 + \theta_2 + \theta_3 \tag{9}$$

There are some examples of calculating the RIP angle, show in Figure 1. The RIP angle regression is in a coarse-to-fine cascade regression style like [1, 2].

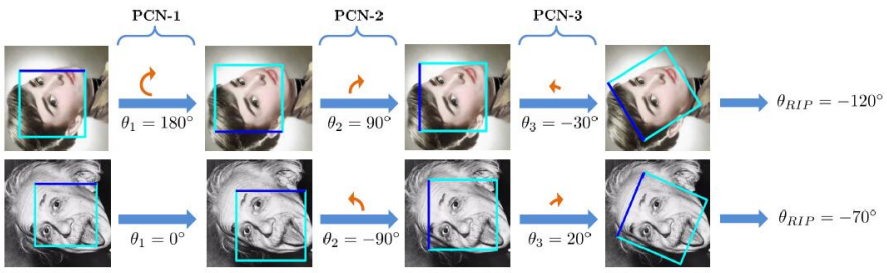

Figure 1: Figure 4. The RIP angle is predicted in a coarse-to-fine cascade regression style. The RIP angle of a face candidate, i.e. θRIP , is obtained as the sum of predicted RIP angles from three stages, i.e. $\theta_{RIP} = \theta_1 + \theta_2 + \theta_3$. Particularly, θ_1 only has two values, $0°$ or $180°$, θ_2 only has three values, $0°$, $90°$ or $-90°$, and θ_3 is a continuous value in the range of $[45°, 45°]$.

References

[1] Piotr Dollar, Peter Welinder, and Pietro Perona. Cascaded pose regression. *IEEE*, 238(6):1078–1085, 2010.

[2] Jie Zhang, Shiguang Shan, Meina Kan, and Xilin Chen. Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment. In *European Conference on Computer Vision*, 2014.