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

Chaonan Song

May 1, 2018

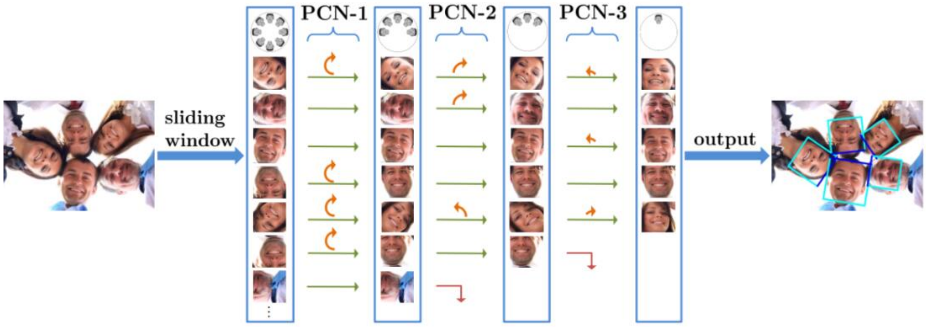

Figure 1: An overview of our proposed progressive calibration networks (PCN) for rotation-invariant face detection. Our PCN progressively calibrates the RIP orientation of each face candidate to upright for better distinguishing faces from non-faces. Specifically, PCN-1 first identifies face candidates and calibrates those facing down to facing up, halving the range of RIP angles from [180°, 180°] to [90°, 90°].Then the rotated face candidates are further distinguished and calibrated to an upright range of [45°, 45°] in PCN-2, shrinking the RIP ranges by half again. Finally, PCN-3 makes the accurate final decision for each face candidate to determine whether it is a face and predict the precise RIP angle.

Starting today I will introduce PCN specific implementation method.

1 Progressive Calibration Networks (PCN)

1.1 Overall Framework

The proposed PCN detector is diagrammed in Figure 1. Given a picture, candidate faces are obtained by sliding window principles and image pyramid principles. Each candidate window passes through the detector in stages.At each stage of the PCN, the detector deletes the face with a low face recognition degree while regressing the bounding box of the remaining candidate faces, and calibration the RIP angle of each candidate faces. After each step, the detector use N-MS merge the faces that have highly overlapped face as most existing methods do.

1.2 PCN-1 in 1^{st} stage

There are three goals for each input window x in PCN-1: face or non-face classification, bounding box regression, and calibration, formulated as follows:

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

where F_1 is the detector of the first stage structured with a small CNN. The f is face confidence score, t is a vector represent the predict of bounding box regression, and g is orientation score. The first 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.

The second objective aims to regressing accurate bounding box, as blow:

$$L_{reg}(t, t^*) = S(t - t^*), \tag{3}$$

where t and t^* represents the results of predicted and ground-truth regression respectively, S is the robust smooth l_1 loss define in [1]. The bounding box regression consist of three terms:

$$\begin{aligned} t_w &= w^*/w, \\ t_a &= (a^* + 0.5w^* - a - 0.5w)/w^*, \\ t_b &= (b^* + 0.5w^* - b - 0.5w)/w^*, \end{aligned} \tag{4}$$

where a, b and w denote the box's top-left coordinate and width. a and a^* for the box and ground-truth box respectively.

The third objective aims to predict the coarse orientation of the face candidate in a binary classification manner as follows:

$$L_{cal} = y \log g + (1 - y) \log(1 - g), \tag{5}$$

where y equals 1 if x is facing up, and equals 0 if facing down.

Overall the objective of PCN-1 in the first stage is defined as:

$$\min_{F_1} L = L_{cls} + \lambda_{reg} \cdot L_{reg} + \lambda_{cal} \cdot L_{cal}, \tag{6}$$

where λ_{reg}, λ_{cal} are parameters to balance different loss.

After optimizing Eq (6), the PCN-1 can be used to filter all windows to get a small number of face candidates.The remaining candidate faces are first updated to the new bounding box of PCN-1 regression, and then the updated candidate face is rotated according to the roughly predicted RIP angle. The predicted RIP angle in first stage i.e. θ_1 can be calculated by:

$$\theta_1 = \begin{cases} 0^\circ, & g \geq 0.5 \\ 180^\circ, & g < 0.5 \end{cases} \tag{7}$$

Specifically, $\theta = 0^\circ$ means that the face candidate is facing up and does not need to rotate, and $\theta = 180^\circ$ means that the face candidate is inverted and needs to be rotated facing up. As a result, the range of RIP angle decreased from $[-180^\circ, 180^\circ]$ to $[-90^\circ, 90^\circ]$.

During training phase, three kinds of data are employed: positive samples, negative samples, and suspected samples. positive samples are those windows with IoU over 0.7; negative samples are those windows with IoU lower than 0.3; suspected samples are those windows with IoU between 0.4 and 0.7. For positive and suspected samples, if their RIP angles are in the range of $[-65^\circ, 65^\circ]$, we define them as facing up, and if in $[-180^\circ, -115^\circ] \cup [115^\circ, 180^\circ]$, we define them as facing down. Samples with RIP angles outside of the above range do not contribute to calibration training. down.

References

[1] Ross Girshick. Fast r-cnn. *In The IEEE International Conference on Computer Vision (ICCV)*, December 2015.