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

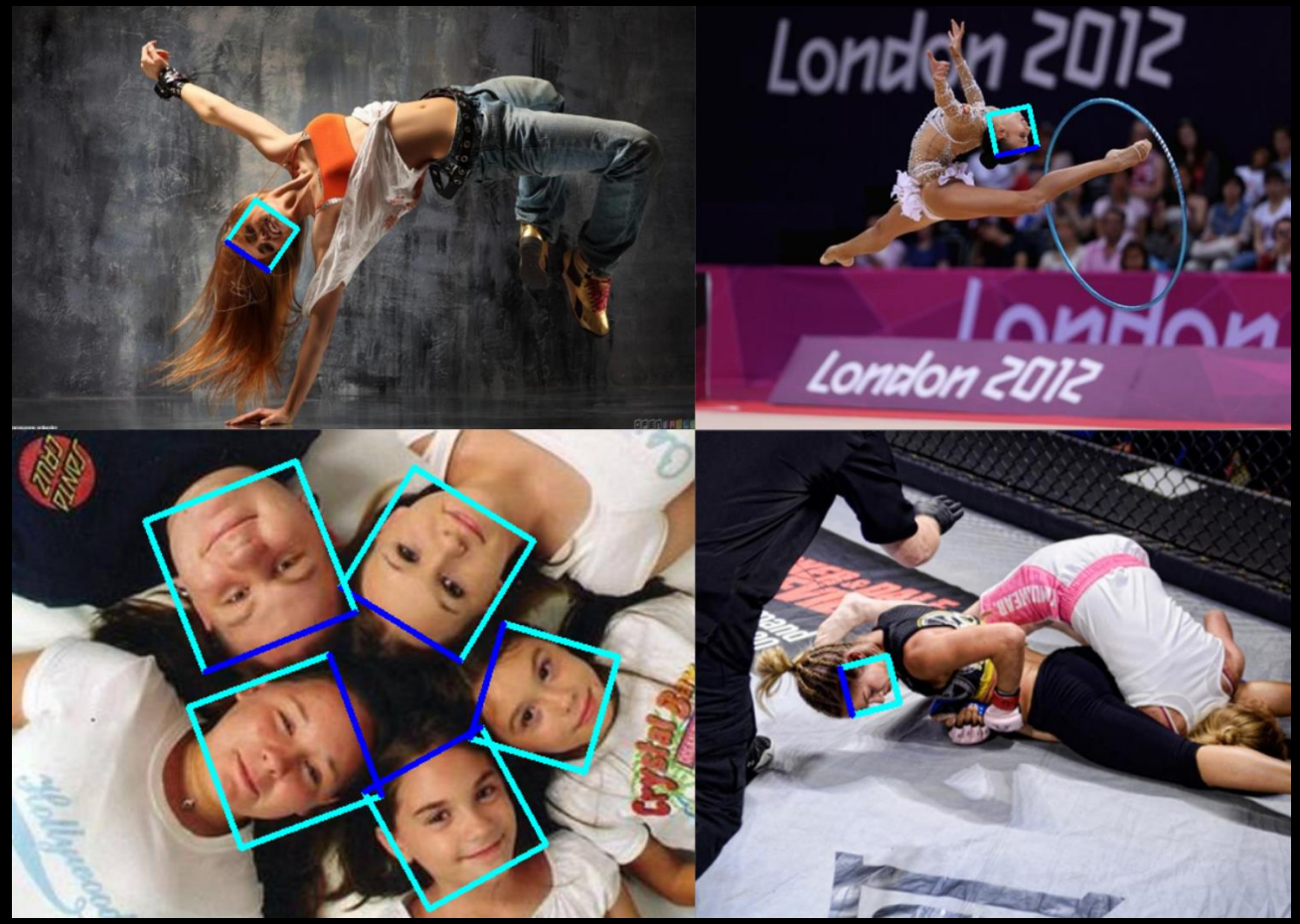
CVPR 2018

Xuepeng Shi, Shiguang Shan, Meina Kan, Shuzhe Wu, Xilin Chen

Rotation-Invariant Face Detection

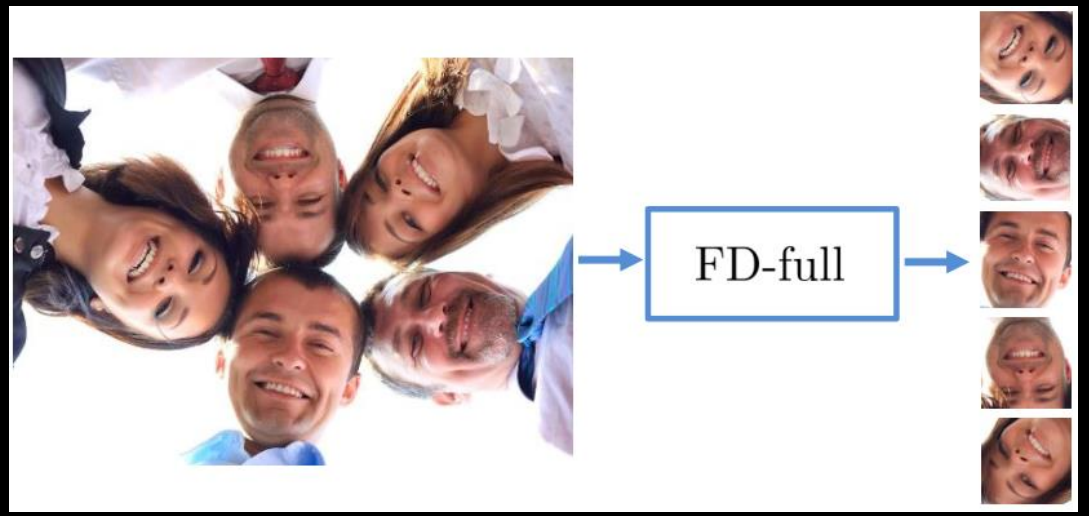
Full rotation-in-plane (RIP) angles

Large appearance variations



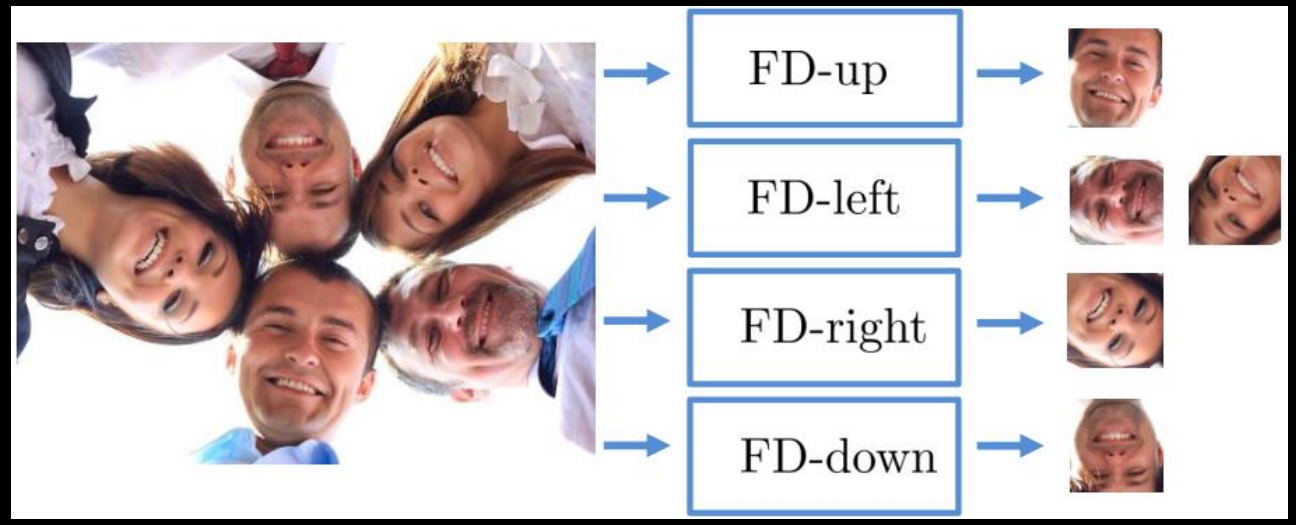
Existing Methods [1/3]

- Data Augmentation
  - To ensure the accuracy, a large neural network is usually required



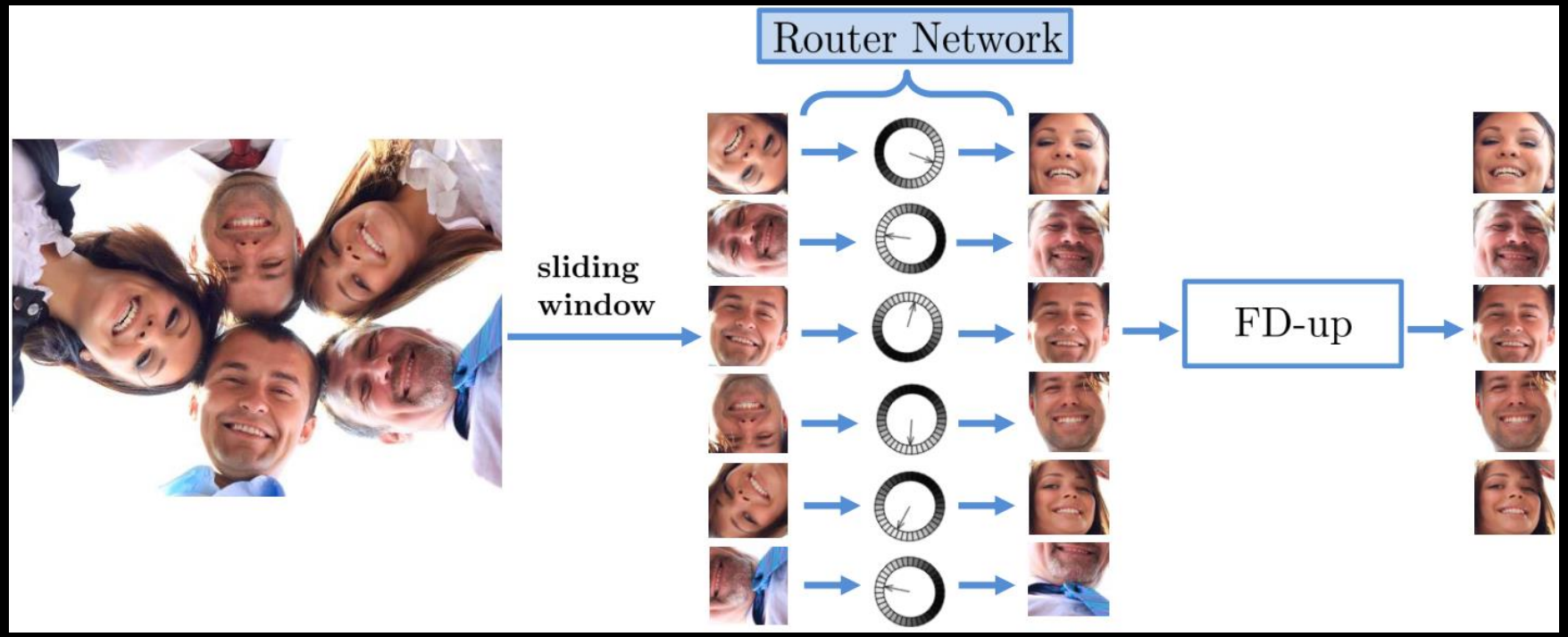
Existing Methods [2/3]

- Divide-and-Conquer [Huang, TPAMI 2007]
  - The overall time-cost largely increases
  - More false alarms are easily introduced



Existing Methods [3/3]

- Rotation Router [Rowley, CVPR 1998]
  - Precisely estimating the RIP angles of faces is quite challenging



Motivation and Goal

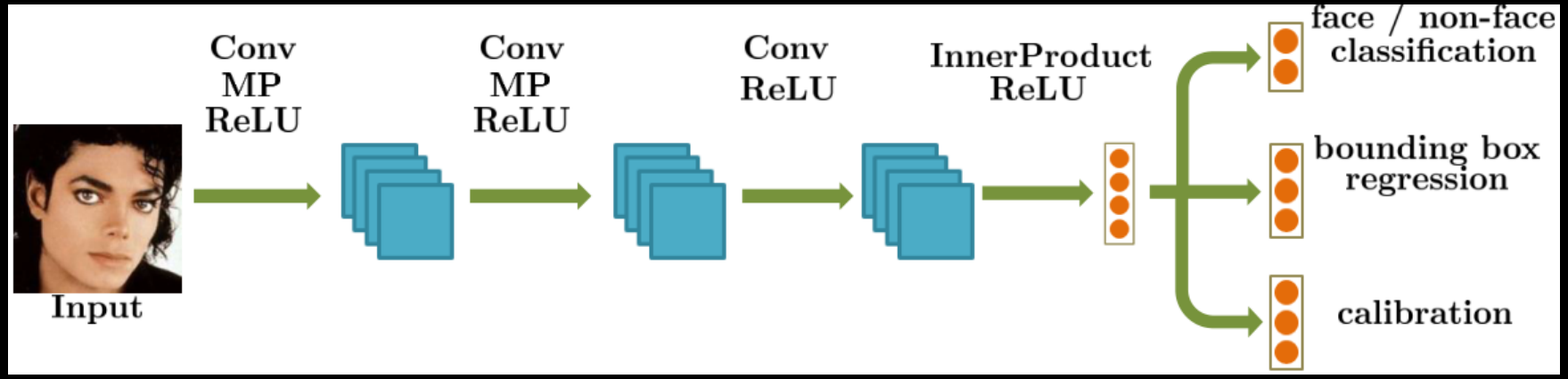
- Most existing methods compromise with speed or accuracy to handle the large RIP variations

- Our PCN aims at accurate rotation-invariant face detection with low time-cost

Progressive Calibration Networks [1/2]

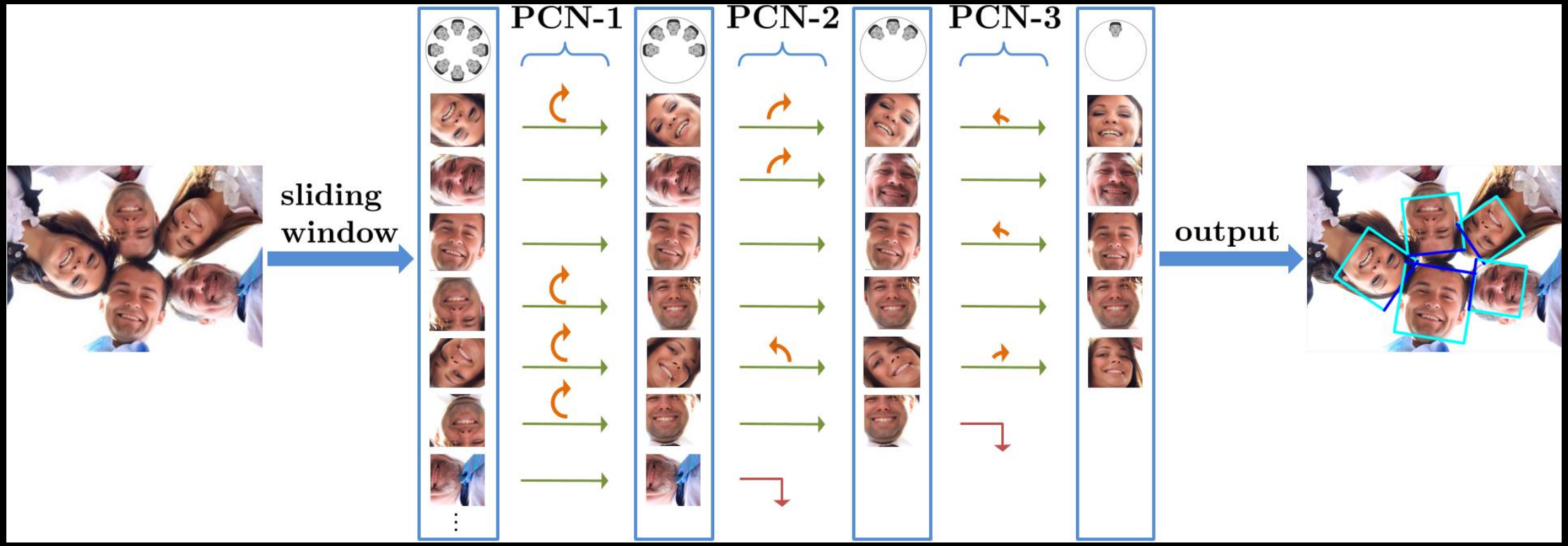
- A sequence of multi-task CNNs
  - Face / non-face classification + bounding box regression + calibration

$$\min_{F} L = L_{classification} + \lambda_{bbox\_reg} \cdot L_{bbox\_reg} + \lambda_{calibration} \cdot L_{calibration}$$



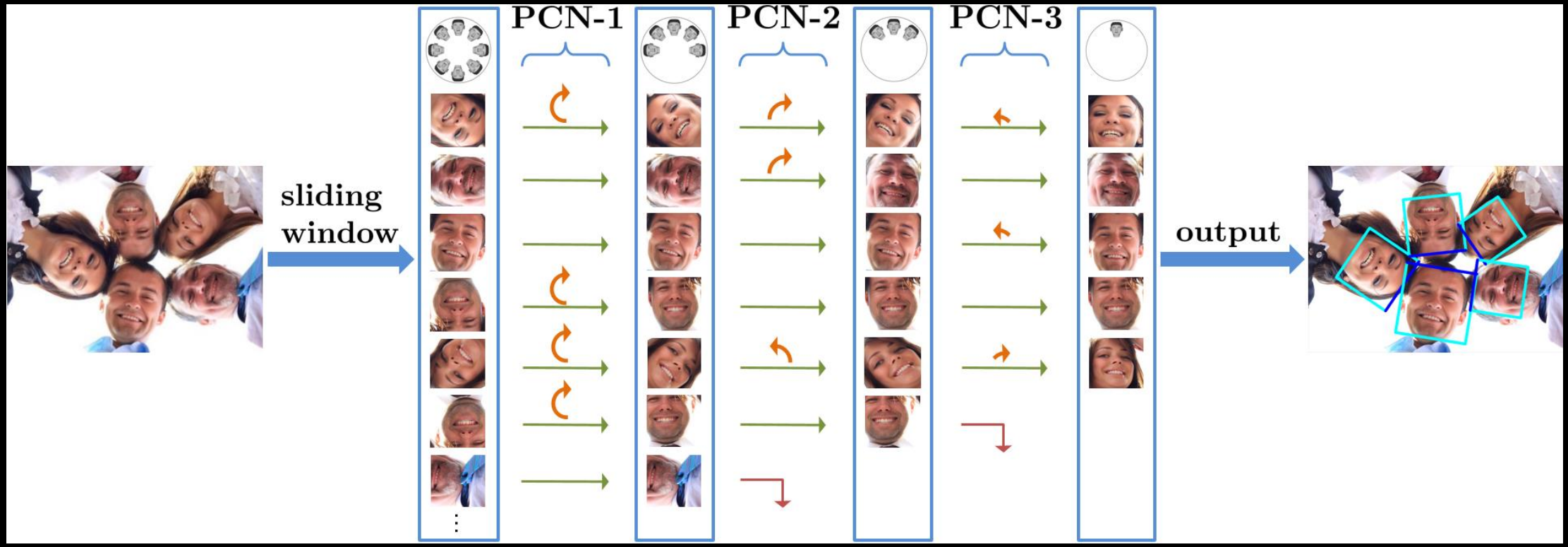
Progressive Calibration Networks [2/2]

- Calibrates the RIP orientation to upright progressively
- Only coarse calibrations are conducted in early stages



Two Cascade Structures in PCN [1/2]

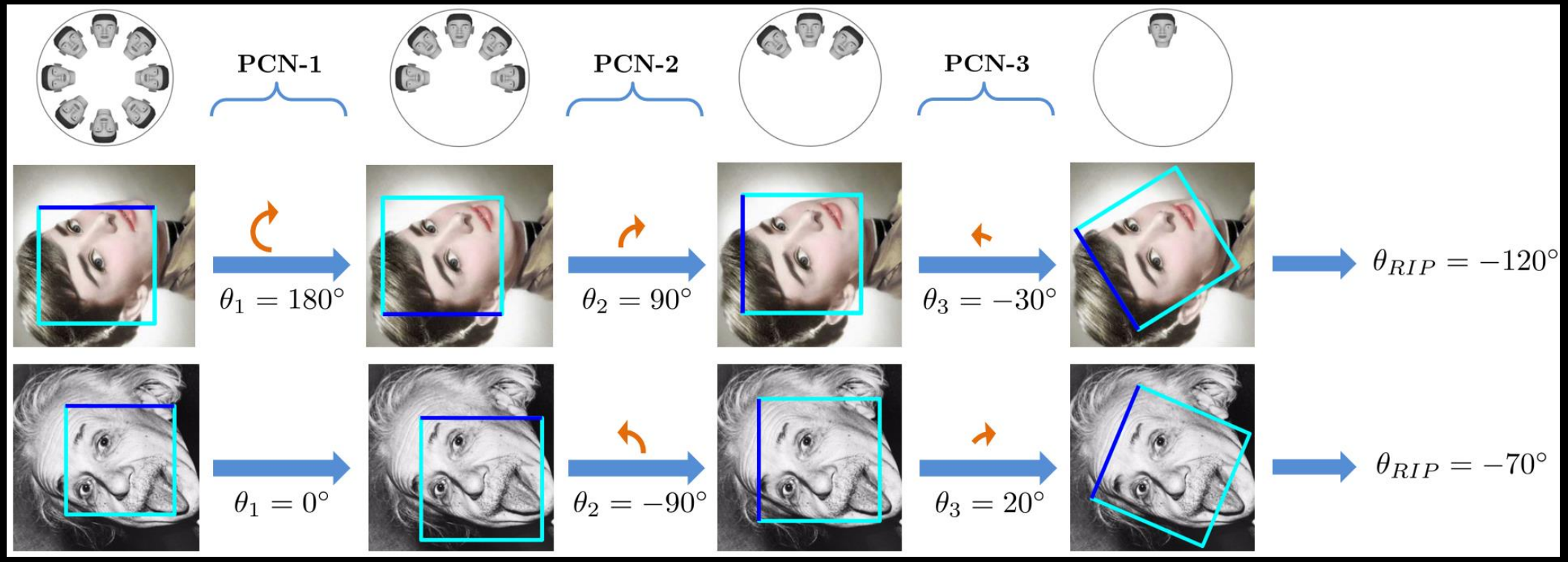
- Cascade classification [Li, CVPR 2015] [Zhang, SPL 2016]
  - Significantly improve the speed of detection



Two Cascade Structures in PCN [2/2]

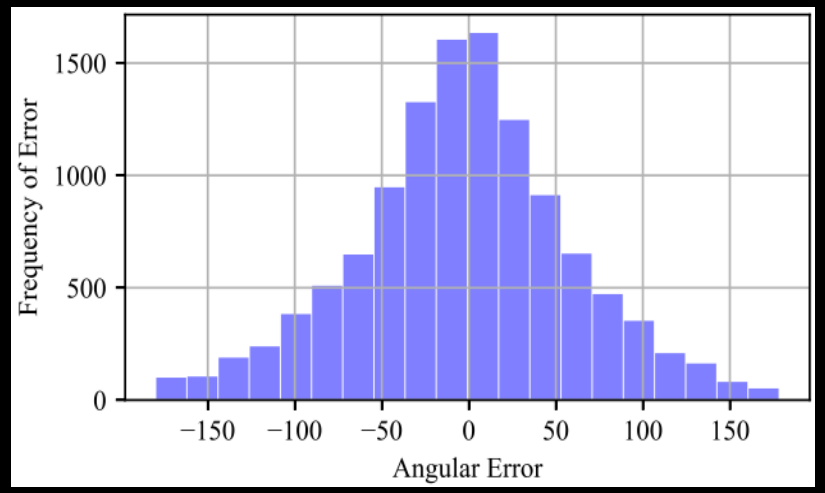
- Cascade calibration
  - Binary classification + ternary classification + regression

$$\theta_{RIP} = \theta_1 + \theta_2 + \theta_3, \theta_1 = 0°/180°, \theta_2 = -90°/0°/90°$$

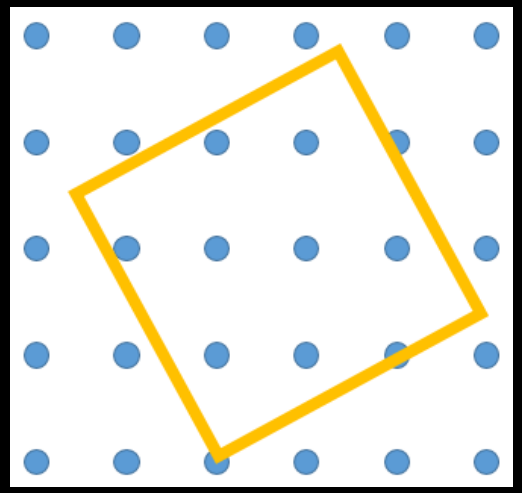


Cascade Calibration [1/2]

- Why cascade?
  - Easier
- Why coarse classification?
  - More accurate

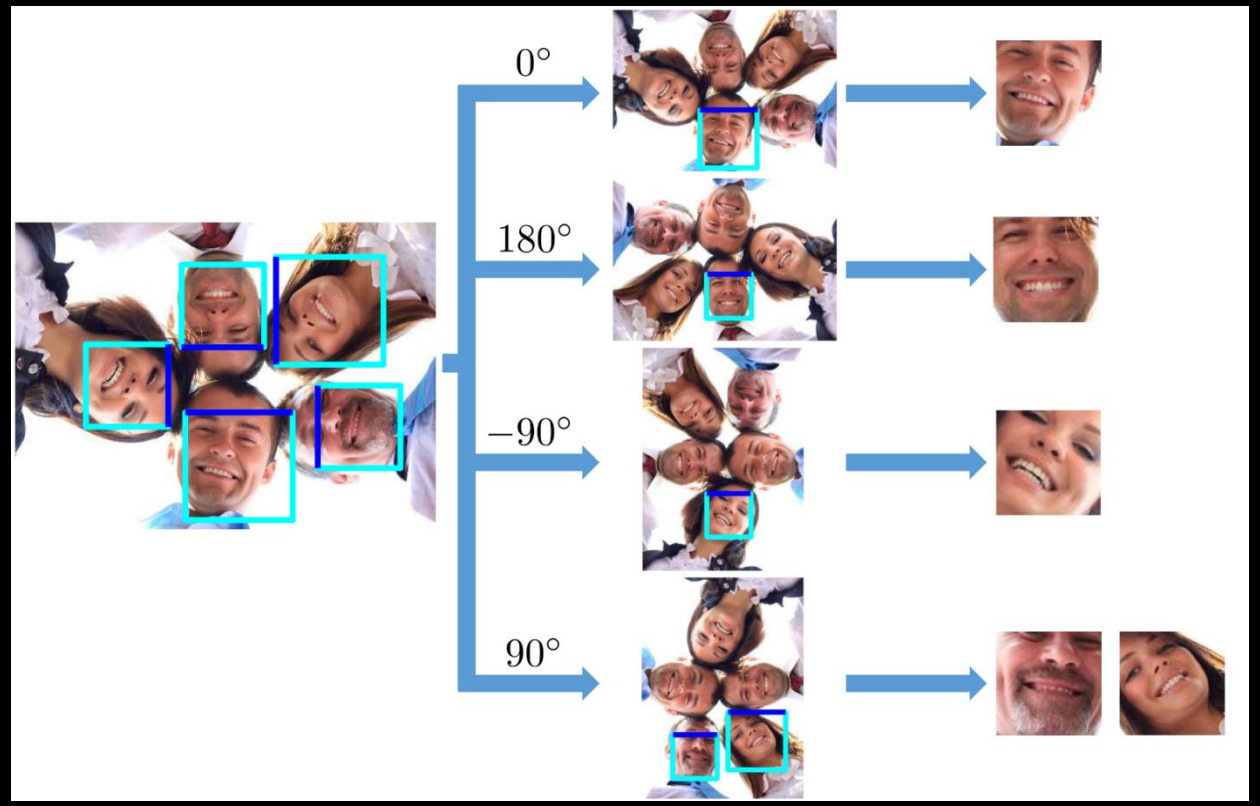


- More efficient
  (Interpolation VS Flipping)



Cascade Calibration [2/2]

- Flip original image three times in advance (1ms - 2ms)
- Reduce the time-cost of calibration dozens of times



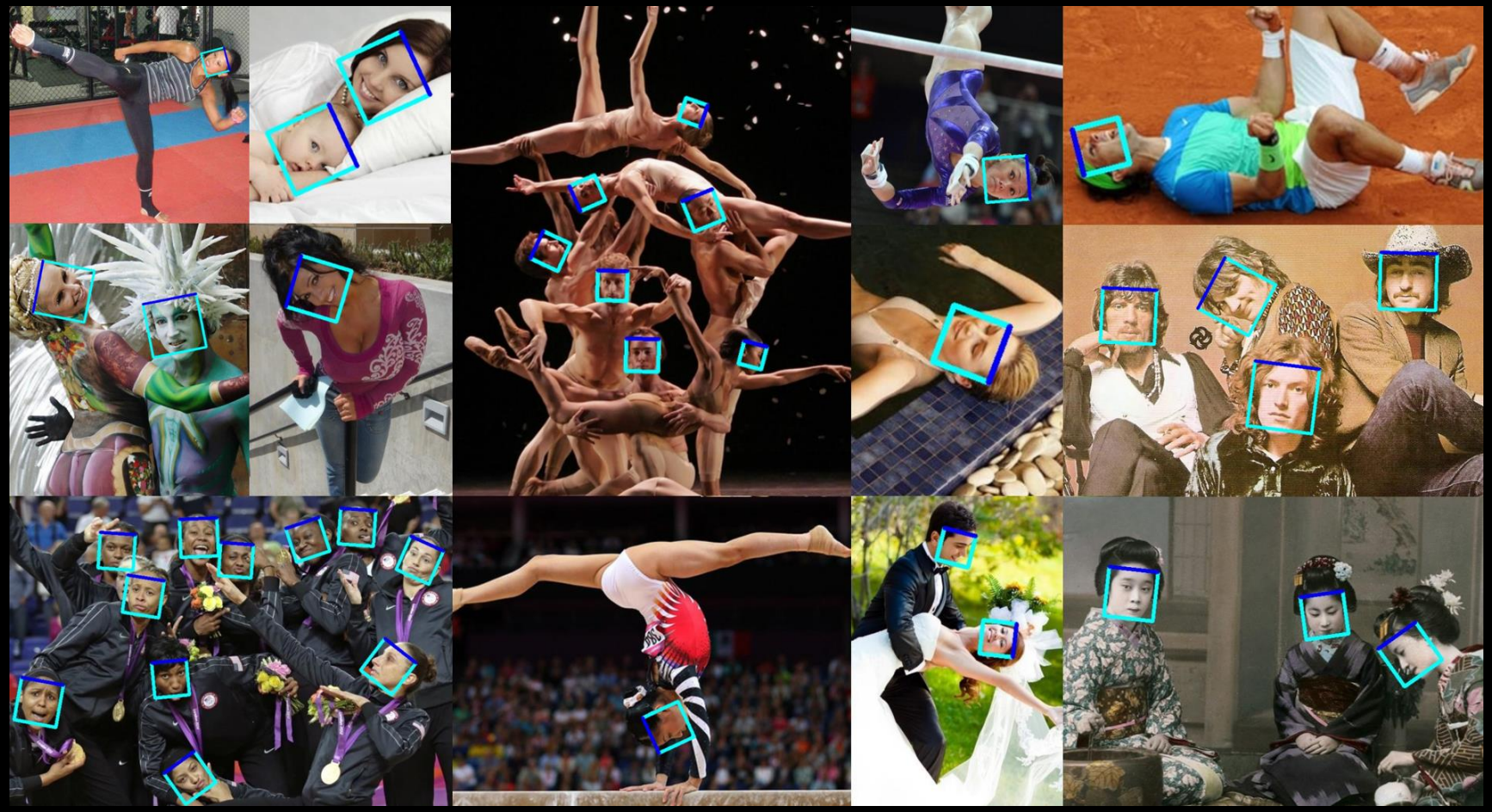
Results

- 640×480 VGA images with 40×40 minimum face size
- A desktop computer with 3.4GHz CPU, GTX Titan X
- PCN can run in real-time on both CPU and GPU

Method	Recall rate at 100 FP on FDDB					Speed		Model Size
	Up	Down	Left	Right	Ave	CPU	GPU	
SSD500 (VGG16)	86.3	86.5	85.5	86.1	86.1	1FPS	20FPS	95M
Faster R-CNN (VGGM)	84.2	82.5	81.9	82.1	82.7	1FPS	20FPS	350M
Faster R-CNN (VGG16)	87.0	86.5	85.2	86.1	86.2	0.5FPS	10FPS	547M
Cascade CNN	84.3	83.7	84.2	83.8	84.0	**31FPS**	**67FPS**	2M
PCN (ours)	**88.1**	**87.6**	**87.2**	**87.6**	**87.6**	29FPS	63FPS	2M

Progressive Calibration Networks

- Fast, accurate, robust



Thanks!