S³FD: Single Shot Scale-invariant Face Detector

Shifeng Zhang, Xiangyu Zhu, Zhen Lei, Hailin Shi, Xiaobo Wang and Stan Z. Li

CBSR & NLPR, Institute of Automation, Chinese Academy of Sciences, University of Chinese Academy of Sciences

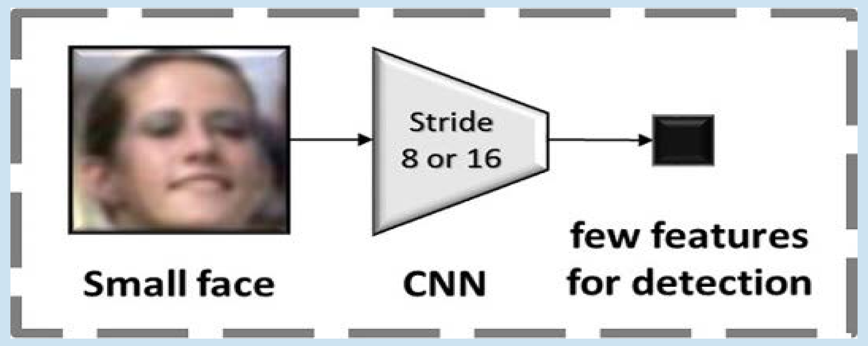
Problem

Though tremendous strides have been made in object detection, one of the remaining open challenges is detecting small objects, since **anchor-based detectors deteriorate dramatically as the objects become smaller**. In this work, we analyze this problem in three aspects and propose three corresponding solutions to solve it.

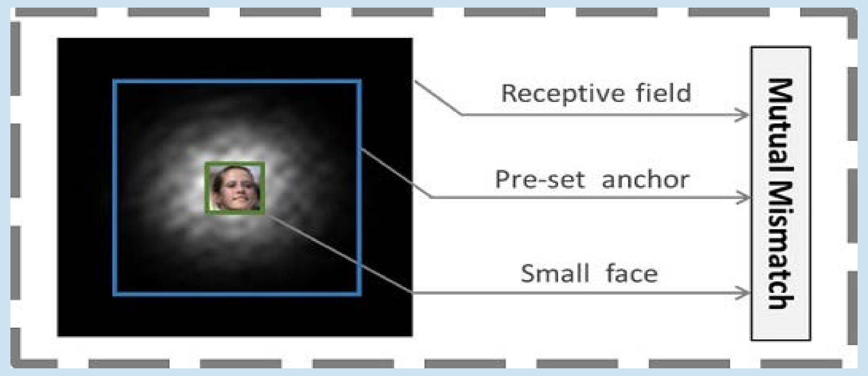
Reasons

Reason1: Biased framework

- Small faces have few features at detection layer

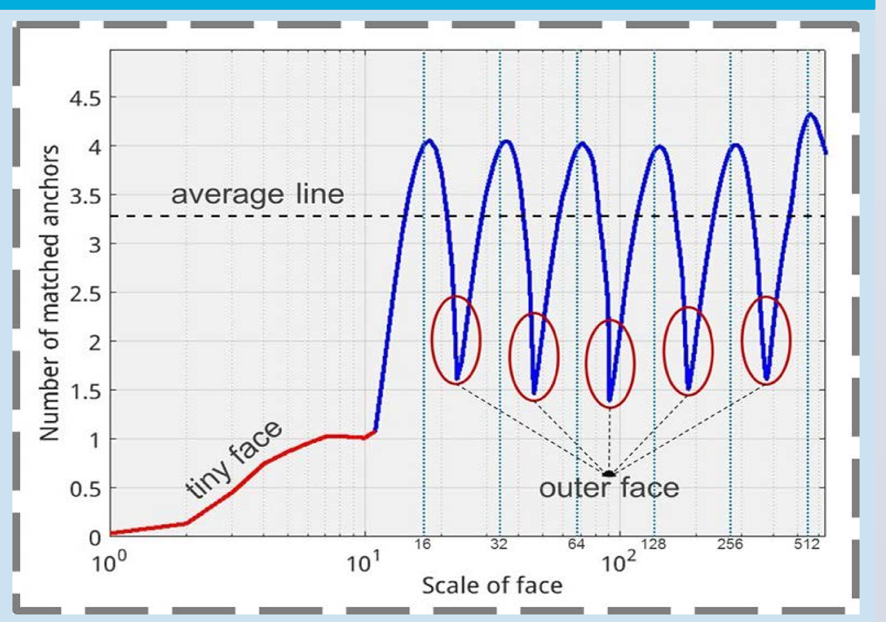


- Small face, anchor scale and receptive field are mutual mismatch



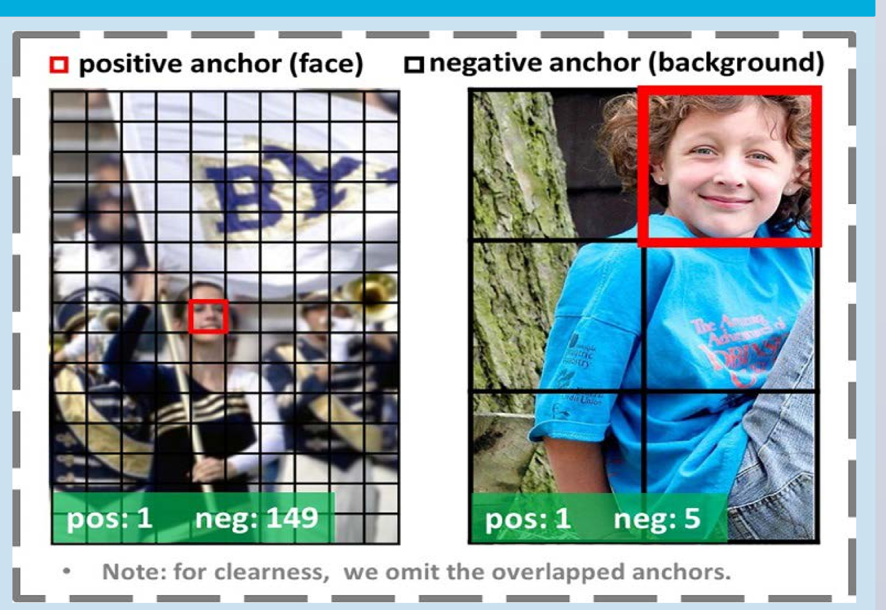
Reason2: Anchor matching strategy

- Some faces can not match enough anchors, such as tiny and outer face



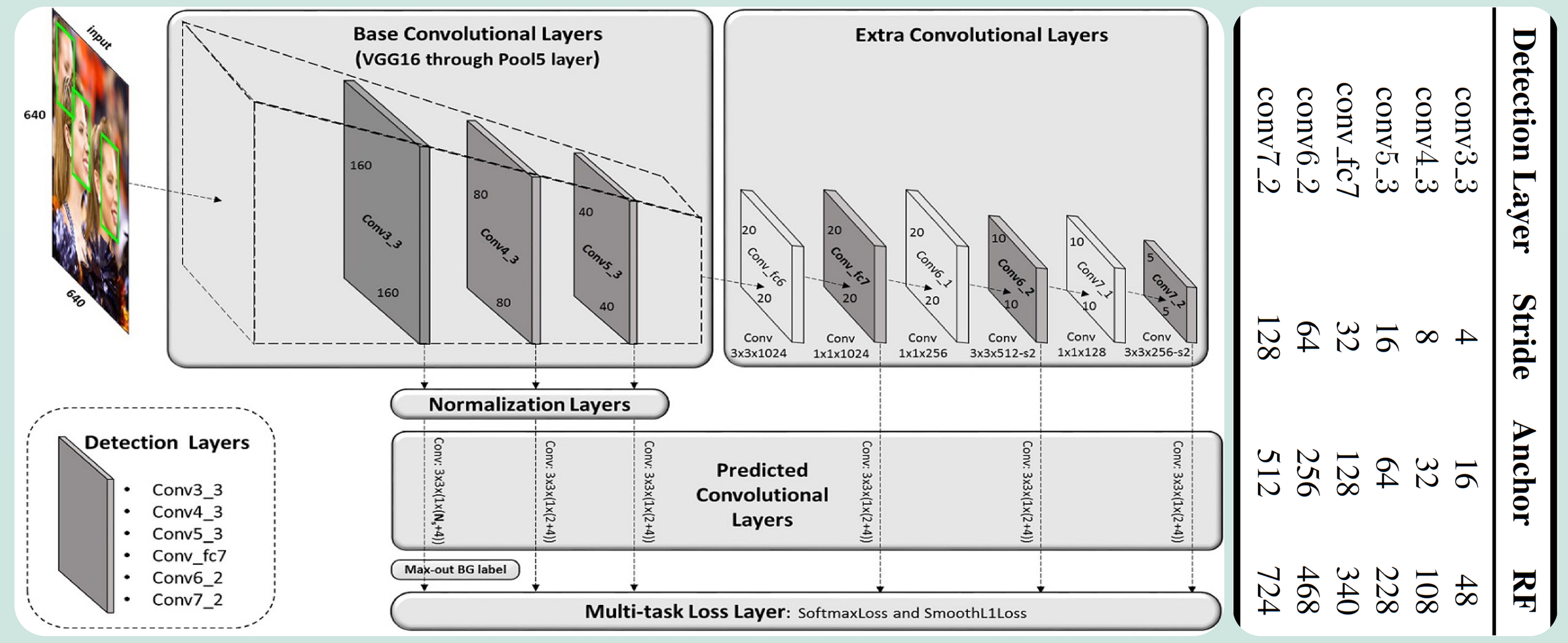
Reason3: Background from small anchors

- Detecting small faces needs to tile plenty of small anchors, which may bring about many false positive faces



Solutions

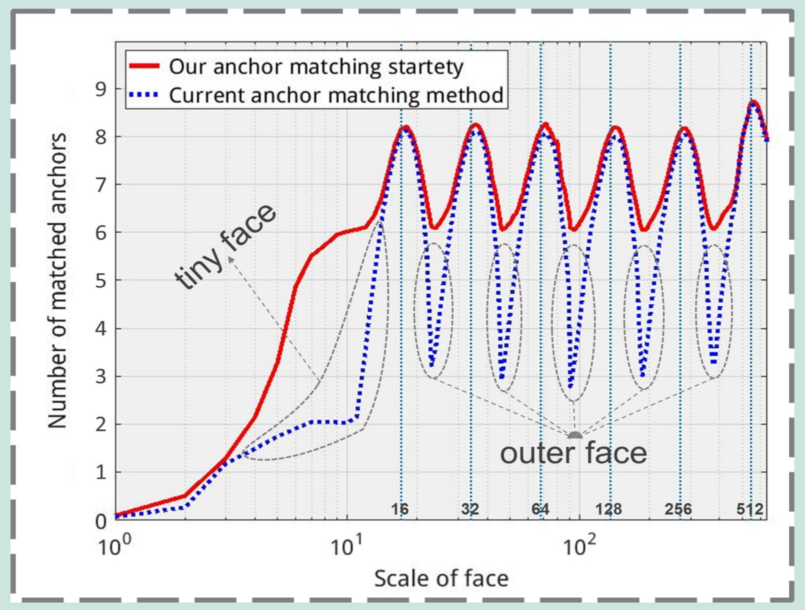
Solution1: Scale-equitable framework



- **Anchor locations**: To ensure various scales of faces have adequate features for detection, we associate anchors with layers whose stride size gradually double from 4 to 128
- **Anchor scales**: To make anchors match the effective receptive field and keep them have the same density, we design scales from 16 to 512 pixels based on effective receptive field theory and equal-proportion interval principle

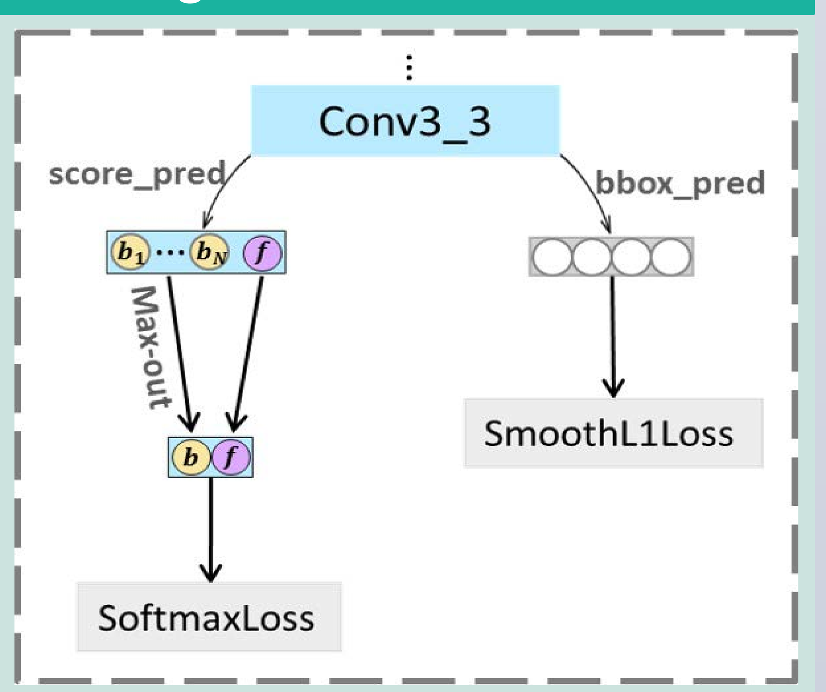
Solution2: Scale compensation anchor matching strategy

- For these faces that can not match enough anchors, we increase their matched anchors via an extra step that relaxes the matching threshold then selects the top N anchors



Solution3: Max-out background label

- For each of the small anchors on the lowest layer, we predict N_m scores for background label and then choose the highest as its final score, so as to reduce the false positive rate of small faces



Results

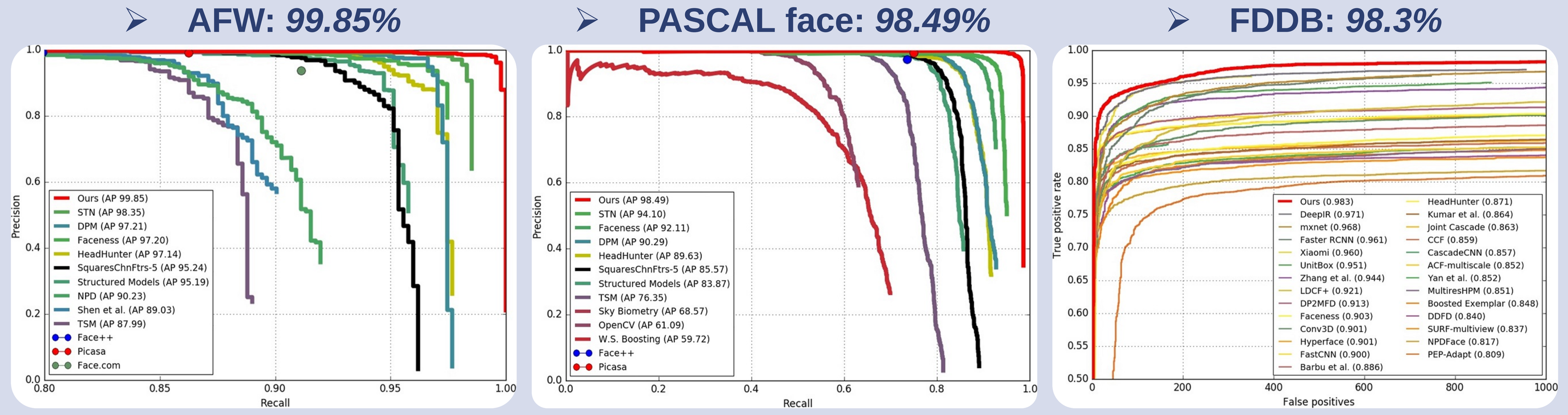
Model analysis

- Scale-equitable framework is crucial
- Scale compensation anchor matching strategy is better
- Max-out background label is promising

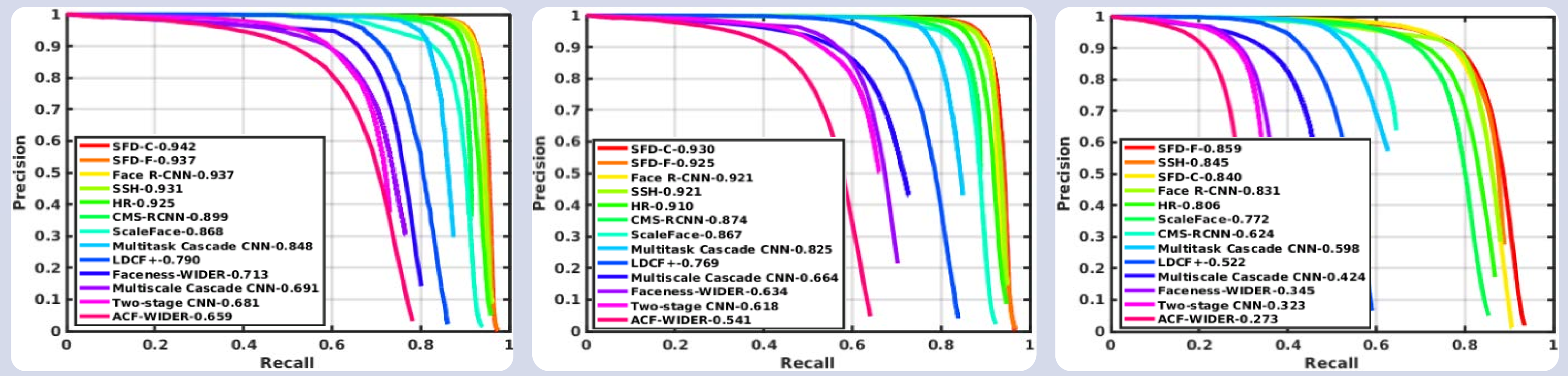
mAP(%) Methods \ Subsets	Easy	Medium	Hard
RPN-face	91.0	88.2	73.7
SSD-face	92.1	89.5	71.6
S³FD(F)	92.6	91.6	82.3
S³FD(F+S)	93.5	92.0	84.5
S³FD(F+S+M)	**93.7**	**92.4**	**85.2**

Evaluation on benchmark

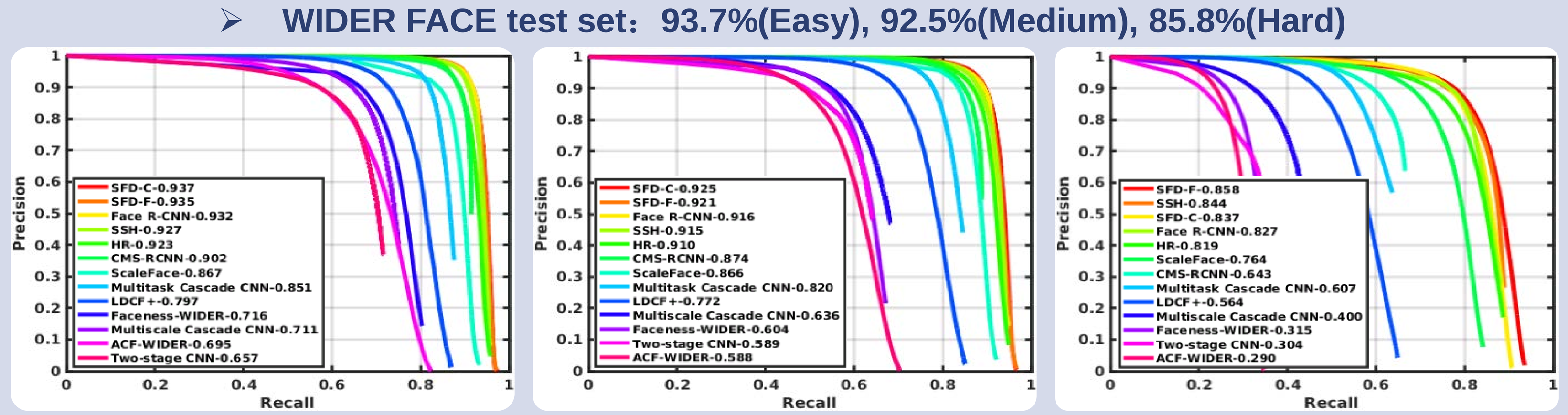
- AFW: *99.85%*
- PASCAL face: *98.49%*
- FDDB: *98.3%*



- WIDER FACE validation set: 94.2%(Easy), 93.0%(Medium), 85.9%(Hard)



- WIDER FACE test set: 93.7%(Easy), 92.5%(Medium), 85.8%(Hard)



Conclusion

We propose a scale-equitable framework, a scale compensation matching strategy and a max-out background label to solve the problem of finding small faces. Consequently, our detector achieves state-of-the-art detection performance on all the common face detection benchmarks, including the AFW, PASCAL face, FDDB and WIDER FACE datasets. Besides it can run at 36 FPS on a NVIDIA Titan X (Pascal) for the VGA-resolution images.

ICCV17
International Conference on Computer Vision 2017

WeChat: 365947064
E-mail: shifeng.zhang@nlpr.ia.ac.cn
Homepage
http://www.cbsr.ia.ac.cn/users/sfzhang