WIDER FACE DETECTION (TRACK 1)

Vishweswar Mohapatra
Lebbar Mohammed Abdul Rahman
Aajad Chauhan
Baizel Kurian Varghese
Dr. Tapas Badal

LEADINGINDIA.AI NATIONWIDE AI SKILLING & RESEARCH INITIATIVE

BENNETT UNIVERSITY TIMES OF INDIA GROUP

Abstract

- **Face detection and alignment in unconstrained environment are challenging due to various poses, illuminations and occlusions.**
- **We propose a deep cascaded multi-task framework which boost up the detection performance.**
- **In particular, our framework leverages a cascaded architecture with three stages of carefully designed deep convolutional networks to predict face and landmark location in a coarse-to-fine manner.**
- **Our method achieves superior accuracy over the state-of-the-art techniques on the challenging WIDER FACE benchmarks for face detection while keeps real time performance.**

Introduction

FACE detection is essential to many face applications, such as face recognition and facial expression analysis. However, the large visual variations of faces, such as occlusions, large pose variations and extreme lightings, impose great challenges for these tasks in real world applications. Recently, convolutional neural networks (CNNs) achieve remarkable progresses in a variety of computer vision tasks, such as image classification and face recognition. Inspired by the significant successes of deep learning methods in computer vision tasks, several studies utilize deep CNNs for face detection.

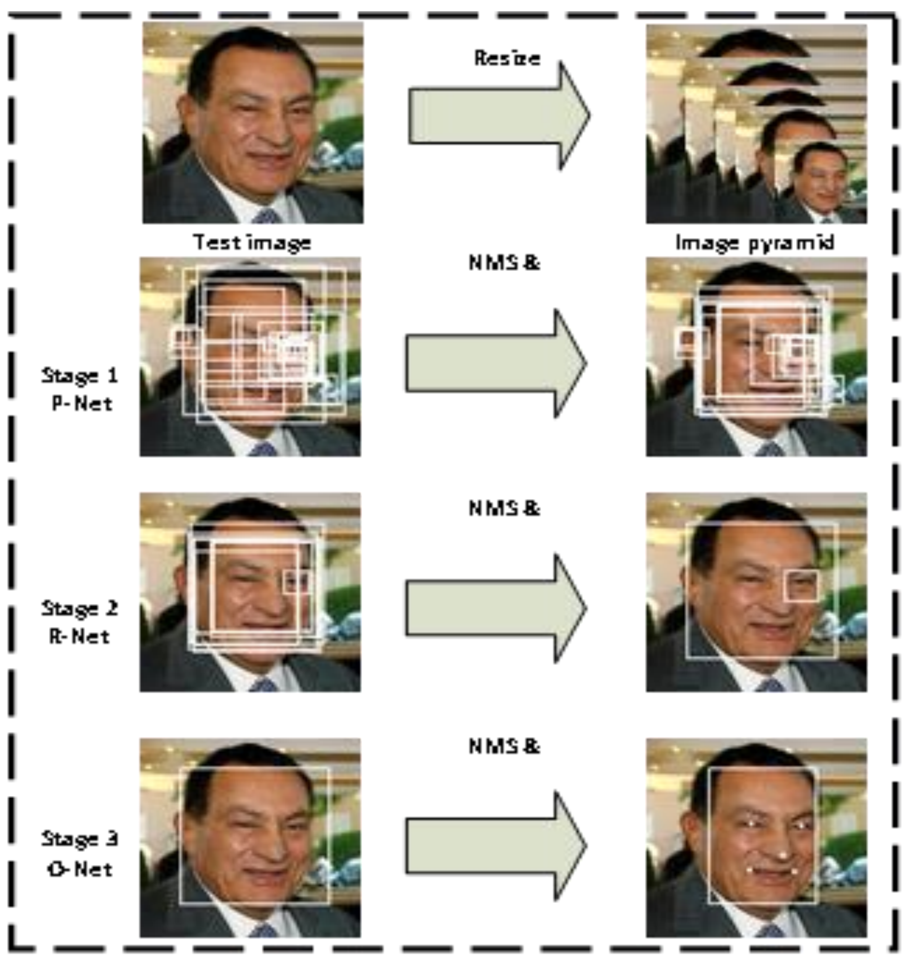


Fig. 1. Pipeline of our cascaded framework that includes three-stage multi-task deep convolutional networks. Firstly, candidate windows are produced through a fast Proposal Network (P-Net). After that, we refine these candidates in the next stage through a Refinement Network (R-Net). In the third stage, The Output Network (O-Net) produces final bounding box and facial land- marks position.

Proposed Method

In this project, we propose a new framework to integrate face detection tasks using unified cascaded CNNs by multi-task learning. The proposed CNNs consist of three stages. In the first stage, it produces candidate windows quickly through a shallow CNN. Then, it refines the windows by rejecting a large number of non-faces windows through a more complex CNN. Finally, it uses a more powerful CNN to refine the result again and output five facial landmarks positions.

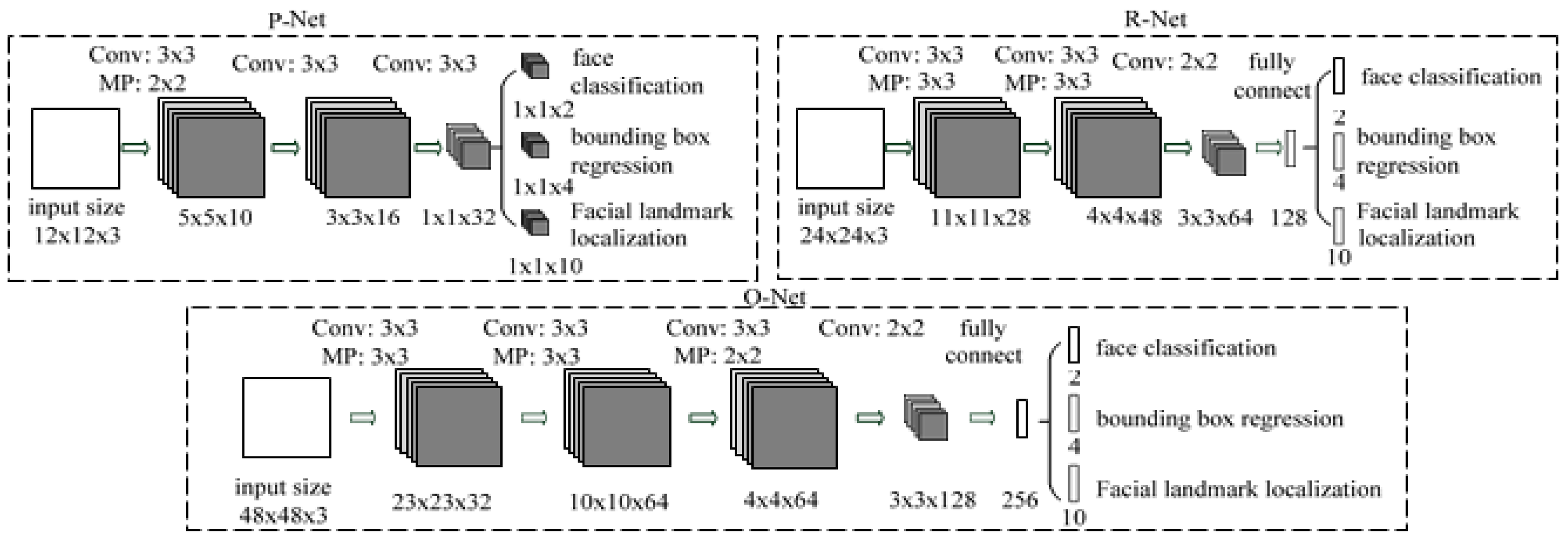


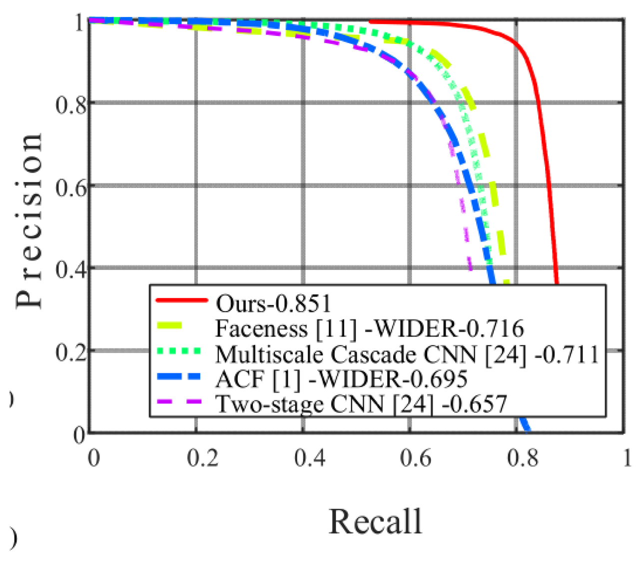
Fig. 2. The architectures of P-Net, R-Net, and O-Net, where "MP" means max pooling and "Conv" means convolution. The step size in convolution and pooling is 1 and 2, respectively.

Experimental Results and Discussion

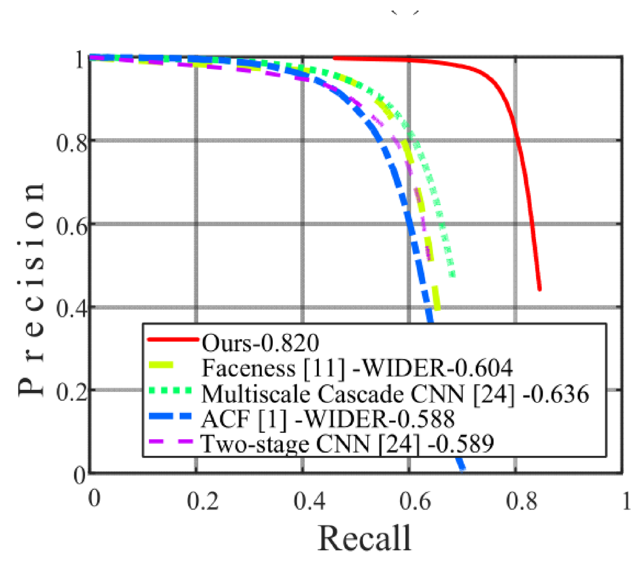
Total training data are composed of 3:1:1:2 (negatives/ positives/ part face/ landmark face) data. The training data collection for each network is described as follows:

1) **P-Net**: We randomly crop several patches from WIDER FACE to collect positives, negatives and part face.
2) **R-Net**: We use the first stage of our framework to detect faces from WIDER FACE to collect positives, negatives and part face.
3) **O-Net**: Similar to R-Net to collect data but we use the first two stages of our framework to detect faces and collect data.

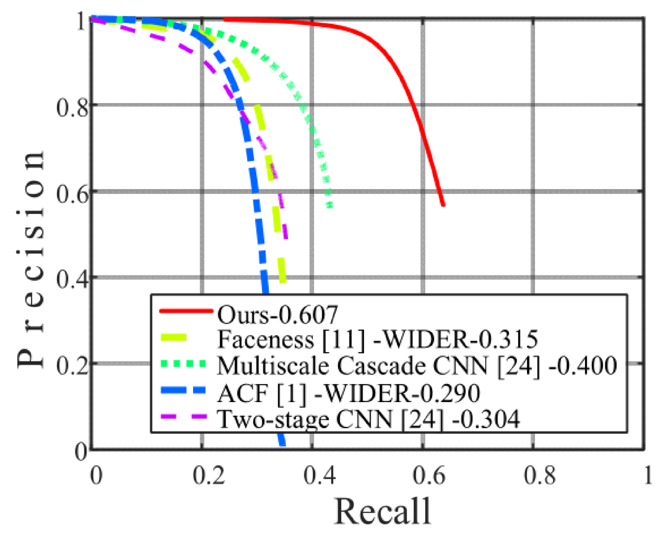
After training our data using this MTCNN Model for 10 epochs in P-Net and 30 epochs in R-Net and 50 epochs for O-Net we got the accuracy of 85.1% in easy dataset, 82% in medium dataset and 60.7% on hard dataset.



(a) Easy Dataset



(b) Medium Dataset



(c) Hard Dataset

Fig 3. (a-c) These are the results of our trained model in comparison with other models on WIDER Face Dataset

Fig. 4. Test Images of the model with best accuracy

Conclusions

Our algorithm can detect between 90.5% and 99.8% of faces in a set of 32, 203 total images, with an acceptable number of false detections. Depending on the application, the system can be made more or less conservative by varying the arbitration heuristics or thresholds used. The system has been tested on a wide variety of images, with many faces and unconstrained backgrounds.

References

- https://github.com/ipazc/mtcnn/tree/master/mtcnn
- https://github.com/wangbm/MTCNN-Tensorflow
- Joint Face Detection and Alignment Using Multi-Task Cascaded Convolutional Networks

Team No. - 12