Uniface: A Unified Network for Face Detection and Recognition

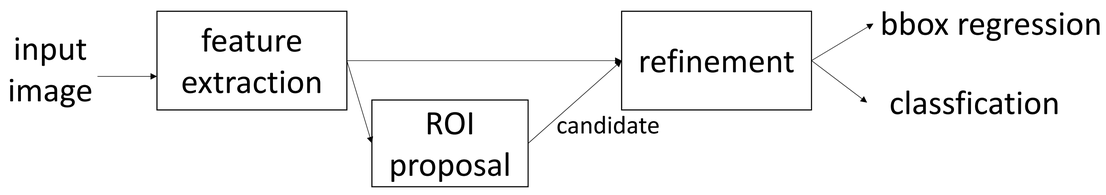
Zhouyingcheng Liao[1], Peng Zhou[1], Qinlong Wu[2] and Bingbing Ni[1]

Shanghai Jiao Tong University[1]
China Mobile Suzhou Software Technology[2]

Motivations

Current face recognition pipeline

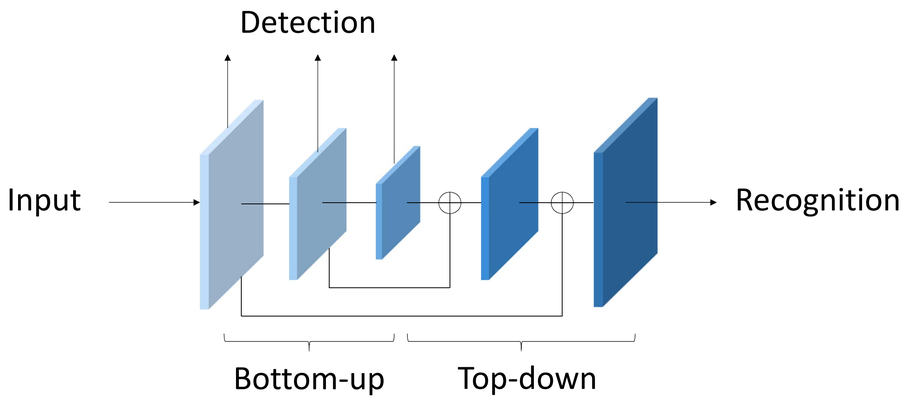
Current object detection pipeline



- Separate face detection and recognition could be computationally wasting
- Inspired by Faster RCNN, we design a unified network for simultaneous face detection and recognition

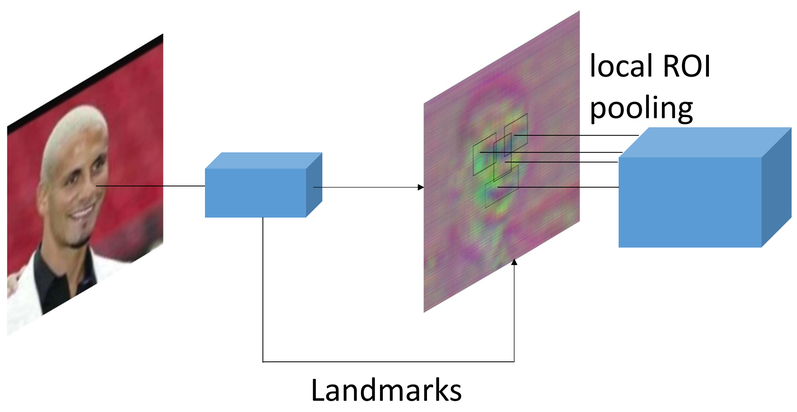
Methods

- **A bottom-up/top-down structure for joint detection and recognition**



High features contain richer semantic information but lower resolution. Thus we use a bottom-up/top-down structure to get features with both rich semantic and high resolution.

- **Landmark attention mechanism for refined features**



Face alignment has great impact on the accuracy of face recognition. We design a simple mechanism called landmark attention to act as face alignment in feature level.

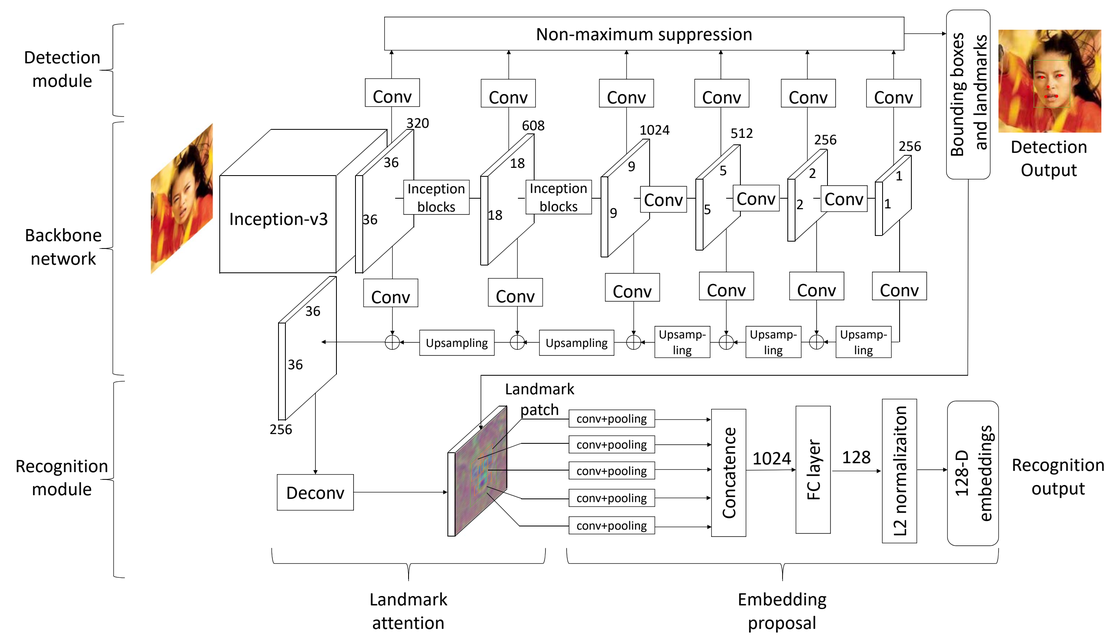
- **Loss functions**

$$L = \frac{1}{N}(\lambda_1 L_{cls} + \lambda_2 L_{bb} + \lambda_3 L_{lm} + L_{id})$$

Face detection, landmark localization and face recognition are trained together in our network. The total loss is a weighted sum of each loss.

1. The classification loss is a cross entropy loss between face areas and background areas.
2. The bounding box loss and the landmark loss are smooth L1 loss between predicted boxes and anchor boxes.
3. For the identification(recognition) loss, we use a facenet model as a teacher network and we calculate L2 loss between embedding of the teacher network and our network.

Network architecture



Experiments

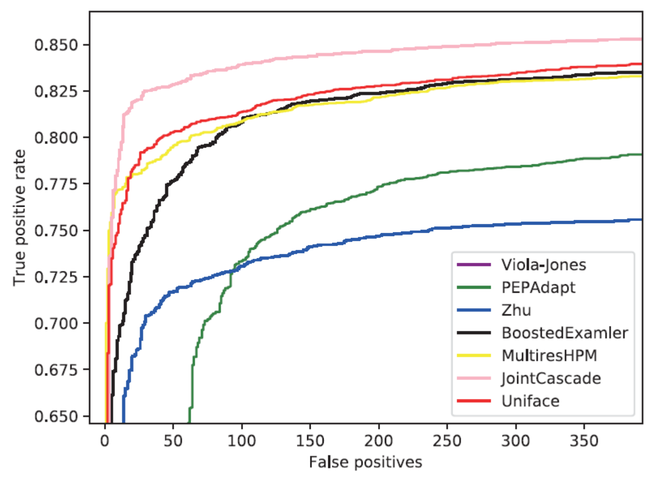
- **Results on LFW**

(Our model doesn't require external face detector)

Methods	Images	Aligned	Networks	Accuracy
DeepFace	4M	3D	4	97.35%
DeepID	203K	2D	60	97.45%
DeepID2	203K	2D	25	**99.15%**
DeepID2+	290K	2D	25	**99.47%**
Facenet	260M	**No**	1	98.87%
Facenet	260M	Yes	1	**99.63%**
Ours	1M	**No**	1	**98.98%**

- **Results on FDDB**

(Our input size is only 300x300)



- **Inference speed**

Model	Task	GPU	Speed/FPS
Faceness	Face Detection	Titan Black	20
MTCNN	Face detection and landmark localization	Titan Black	99
All-in-one	Face detection, attribute analysis and recognition	Titan X	0.286
Faster-RCNN	Object detection	Titan X	7
Ours	Face detection, landmark localization and face recognition	Titan X	**120**

Contributions

1. We adopt the bottom-up/top-down architecture for combining face detection and recognition. It brings the network the capability of multi-scale face recognition.
2. We introduce an attention mechanism called landmark attention for face recognition, with which the network could get more refined attention to the face area.
3. We present a single-network model, i.e. Uniface network for simultaneous face detection, landmark localization and recognition. It achieves the accuracy of 99.0% on LFW and 85.1% on FDDB, while it runs with the speed of 120 FPS (pre-process time excluded).