ICCV2015 Santiago - Chile December 7-13, 2015

Convolutional Channel Features

Bin Yang, Junjie Yan, Zhen Lei, Stan Z. Li
Center for Biometrics and Security Research, Chinese Academy of Sciences
{bin.yang, jjyan, zlei, szli}@nlpr.ia.ac.cn

Motivation

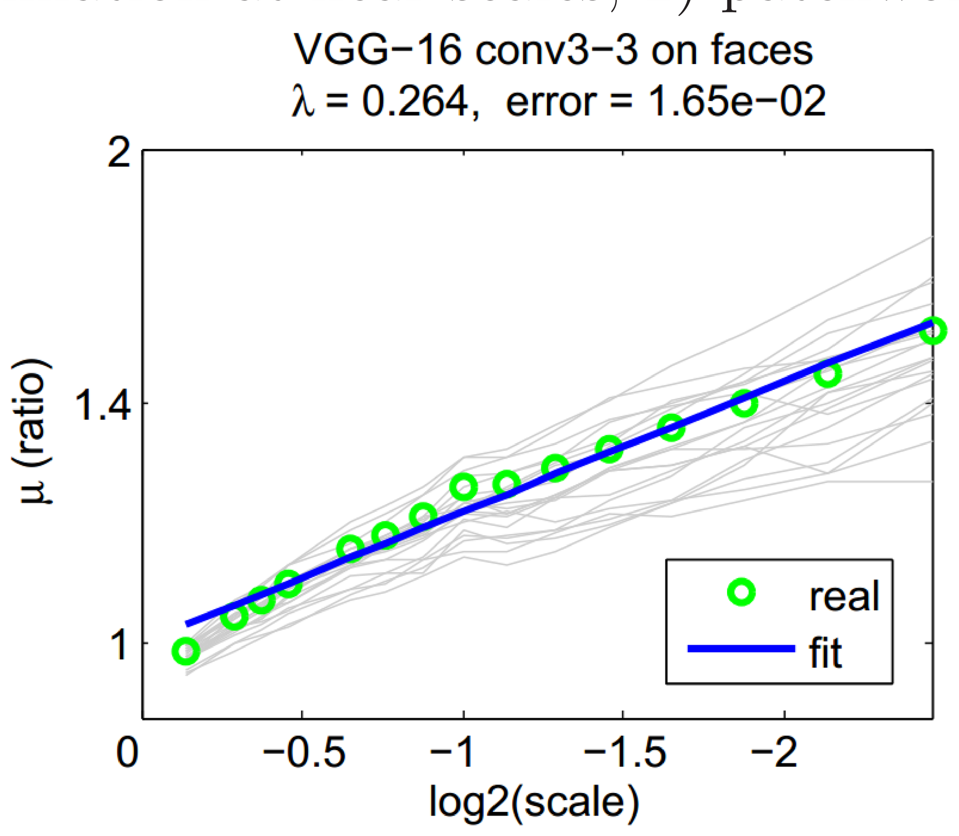
(1) Channel features with boosting approach and fine-tuned CNN features with SVM/MLP are the two most popular methods on various detection tasks.
(2) The multiple channel features approach achieves excellent results in various tasks with very few parameters. However, the performance saturates as we add more and more hand-crafted filters (from tens to thousands) on the channel features.
(3) The fine-tuned CNN feature with SVM/MLP has recently achieved state-of-the-art performance on challenging tasks due to its great representation capacity and end-to-end learning. However, currently CNN model is often accompanied with huge computation complexity, and the model size is usually large (e.g, more than 500MB for widely used VGG net).

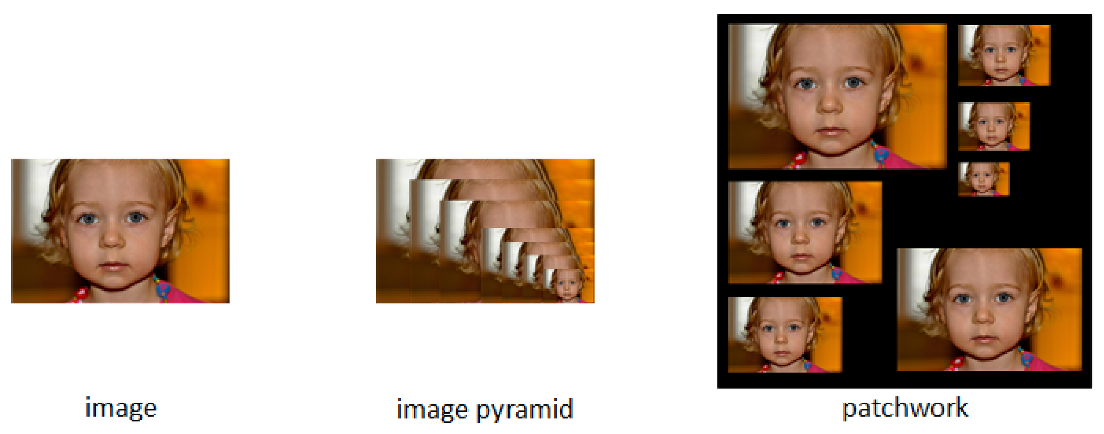
Contribution

(1) We prove that the low-level feature representation in pre-trained CNN model can be used as a new type of channel features and can generalize well to diverse tasks, without even fine-tuning to each domain.
(2) We prove that the high-level connections (convolutional and fully-connected layers) in CNN model can be replaced with a boosting forest model on some specific tasks.
(3) We achieve state-of-the-art results on Caltech pedestrian detection, AFW face detection, BSDS500 edge detection and VOC2007 object proposal generation.

Acceleration

Given a test image, we first compute its CCF feature pyramid composing of multiple scales, and then apply the boosting model on the feature pyramid in a sliding-window style to get detections at multiple scales and locations. We adopt two techniques to accelerate the feature pyramid computation: 1) feature approximation at near scales; 2) patchwork.

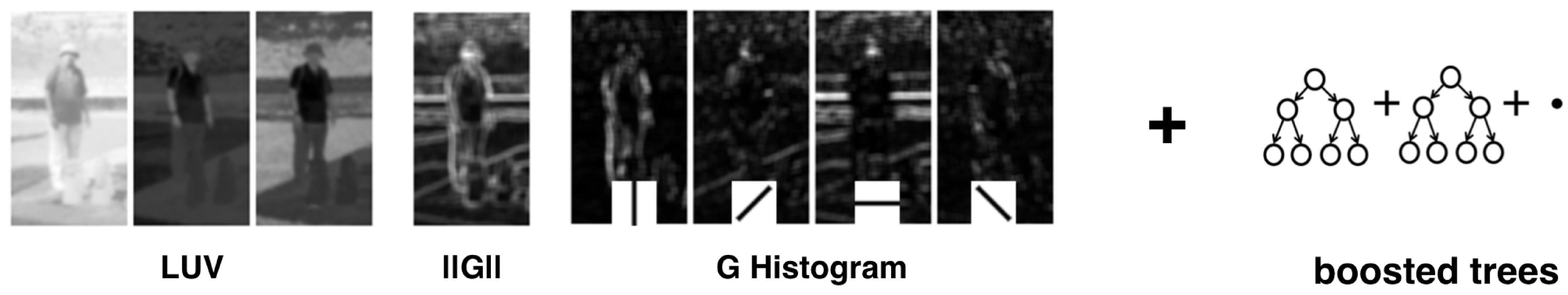




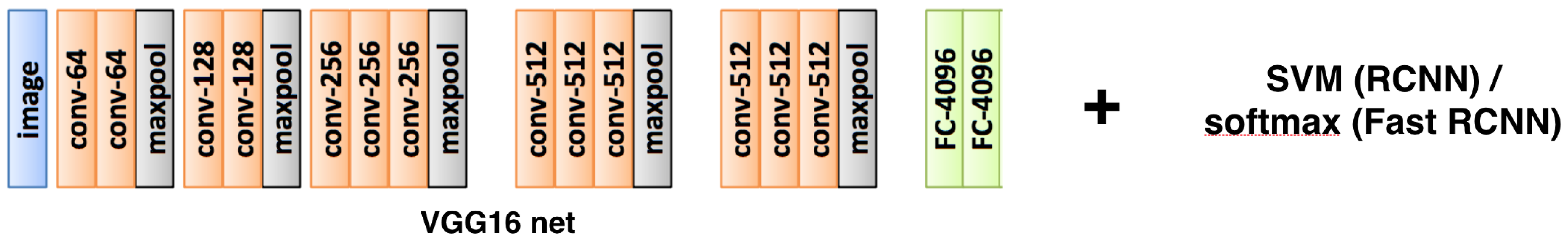
Solution

In practice, we always desire for better performance and lower computation/storage cost. This motivates us to build a bridge between the above two approaches and gain benefits from their respective advantages at the same time.

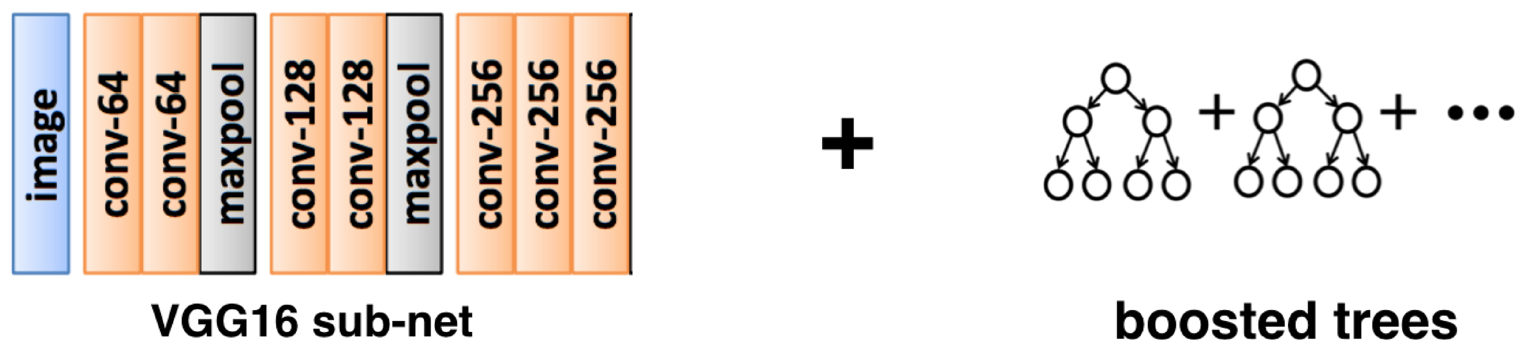
Approach I: Channel Features with boosting



Approach II: Convolutional Neural Networks

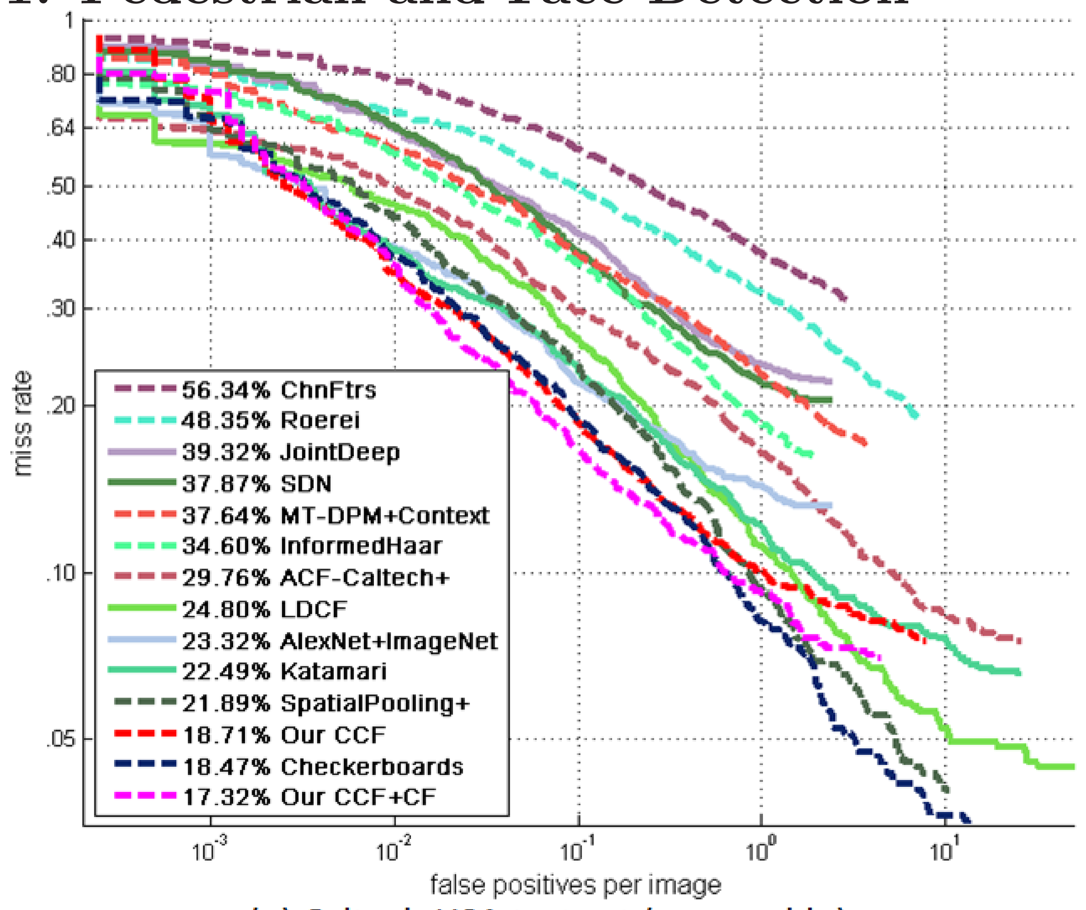


Our solution: Convolutional Channel Features

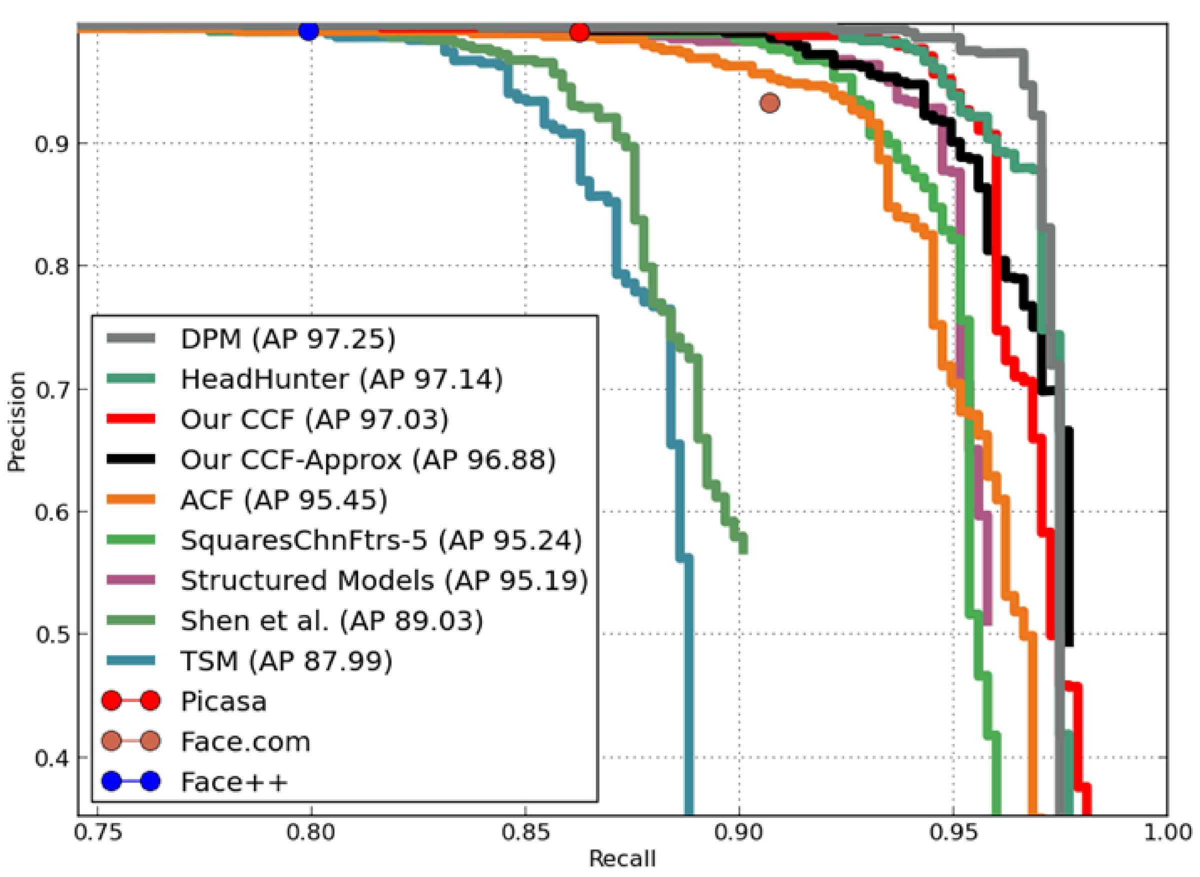


Experiments

1. Pedestrian and Face Detection



(a) Caltech-USA test set (reasonable)



(b) AFW test set

2. Edge Detection and Object Proposal Generation

Method	ODS	OIS	AP
Human	0.80	0.80	-
DeepNet	0.738	0.759	0.758
SE	0.739	0.759	0.792
SE+ms	0.746	0.767	0.803
MCG	0.747	**0.779**	0.759
DeepEdge	**0.753**	0.772	0.807
CCF	0.741	0.761	0.808
CCF+ms	0.744	0.767	**0.809**

Table 1: Evaluation results of edge detection on BSDS500 dataset. Three standard metrics are used, which are fixed contour threshold (ODS), per-image best threshold (OIS), and average precision (AP).

Method	AUC	N@50%	N@75%	Recall
BING	0.20	-	-	29%
Objectness	0.27	584	-	68%
Sel. Search	0.40	199	1434	87%
CPMC	0.41	111	-	65%
EdgeBoxe	0.46	108	800	87%
CCF	**0.48**	**89**	**649**	**88%**

Table 2: Evaluation results of object proposal generation on PASCAL VOC 2007 test set with IoU threshold of 0.7. Metrics are Area Under Curve (AUC), number of proposals needed to reach 50% and 75% recall and maximum recall rate.

Conclusion

In this paper, we propose an integrated method called Convolutional Channel Features (CCF) by combining the low-level CNN features and boosting forest model together. CCF benefits from the rich representative capacity of CNN as well as the efficiency in inference and learning from boosting forest model. The proposed method achieves state-of-the-art performances in diverse detection tasks, showing potentials for use in mobile and embedded devices with small model size.