It takes two to tango: Cascading off-the-shelf face detectors

Siqi Yang, Arnold Wiliem and Brian C. Lovell
The University of Queensland, Australia

Why do we care about false alarms?

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

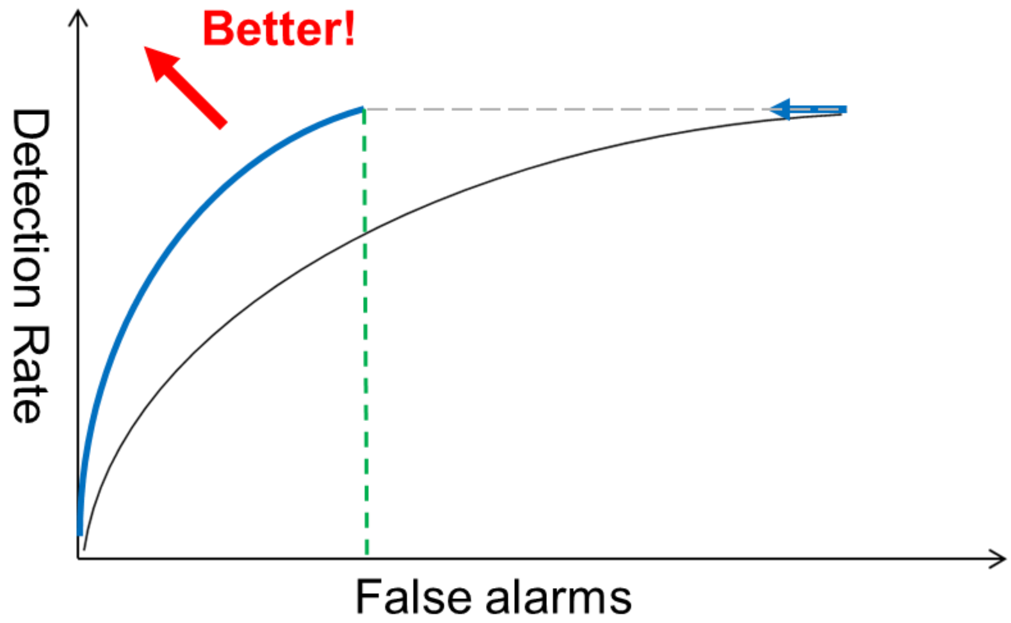
- In a surveillance video, the majority of video frames are occupied by the background (i.e., non-faces) , which increases the probability of generating false positives

- When too many false alarms are raised, users will try to turn off the security system
- Waste computations：face detection is an initial step
- All face detectors generate false positives

Photo credit: http://www.a3con.com.br/ler-noticia/portaria-virtual-segura

Reducing FPs can improve detectors

Aim: reducing false positives + maintaining true positive rate

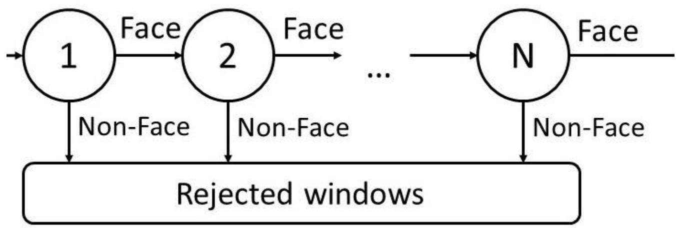


Contributions

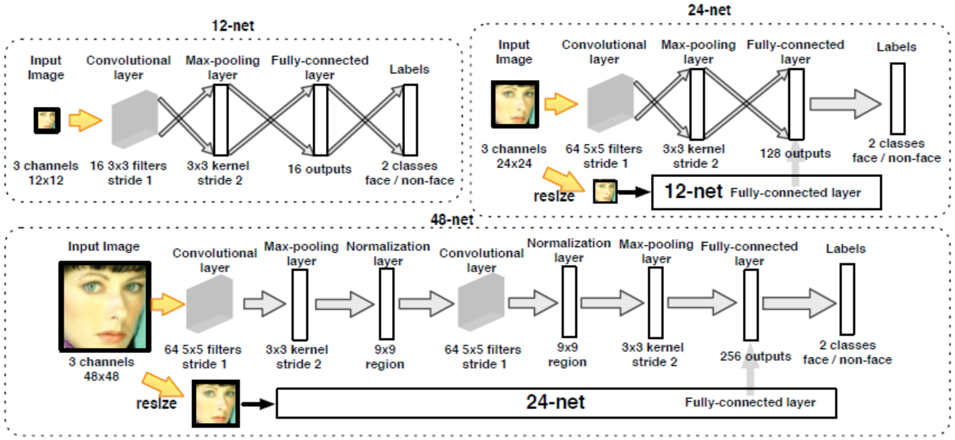
- Propose a set of cascade properties useful to find the best pair of face detectors to cascade
- Performed evaluation of 12 pairs of recent face detectors
- Found a pair of face detector that achieves significantly lower positive rate with competitive detection rate. This pair runs five times faster than the recent state-of-the-art detector

Efforts to reduce false positives

- Cascade structure



Viola and Jones (2001)



Cascade CNN (Li et al., 2015)

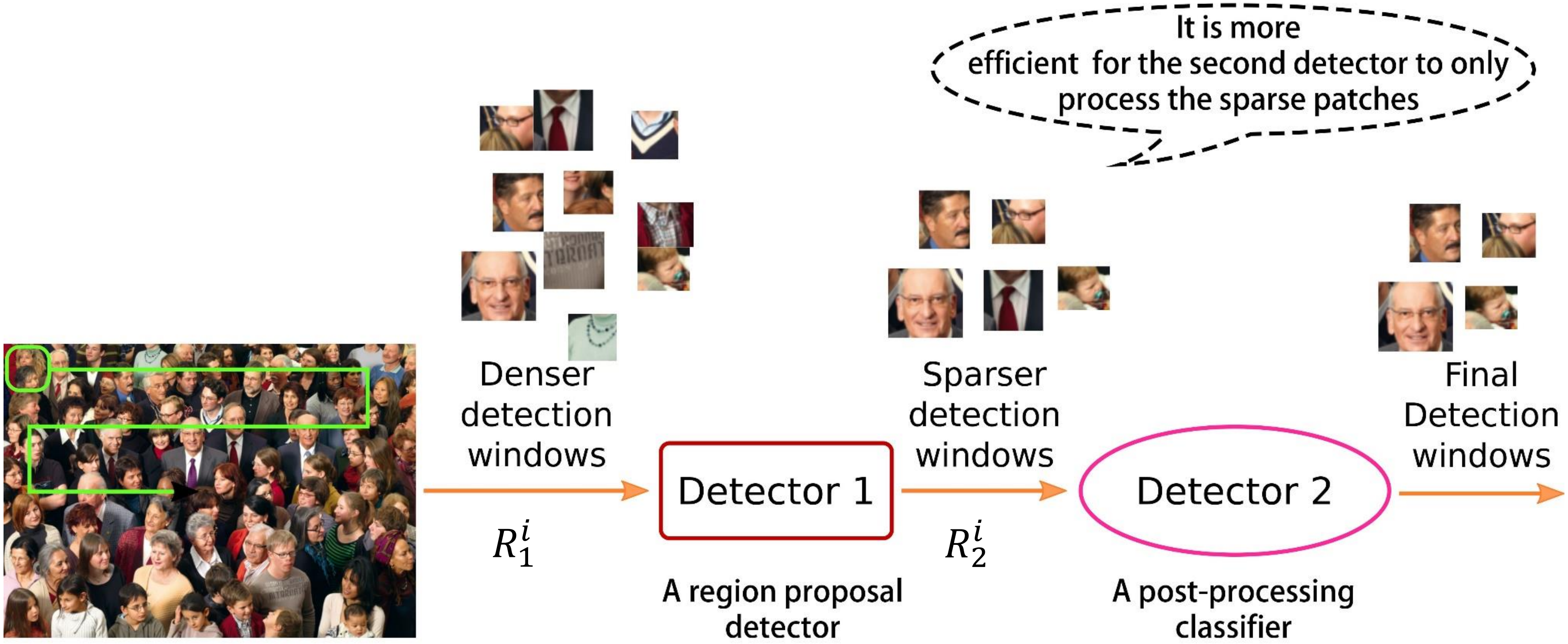
- Bootstrapping or hard negative mining
  – Online Hard Example Mining (OHEM), (Shrivastava et al., 2016)

Shortcomings:

- Due to the features, classifiers and training samples, every face detector has its own theoretical limits
- The effort to train a new face detection model is enormous, e.g., large training data and some face detectors do not provide open source training codes.

[1] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in CVPR, 2001.
[2] Li et al. "A convolutional neural network cascade for face detection." *in CVPR* 2015.
[3] Shrivastava et al."Training region-based object detectors with online hard example mining." IN CVPR, 2016.

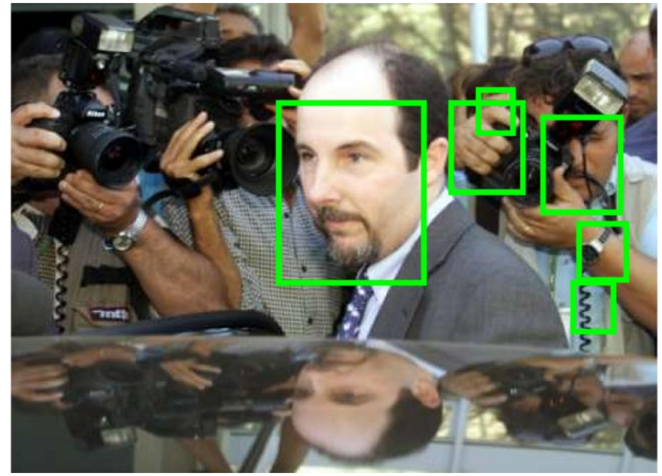
Two-stage Cascade Framework



Proposition

The cardinality of the set of input regions of the second detector is always far smaller than the cardinality of that of the first detector, $|R_2^i| \ll |R_1^i|$

NPD (Liao et al., 2016)



HeadHunter (Mathias et al., 2014)

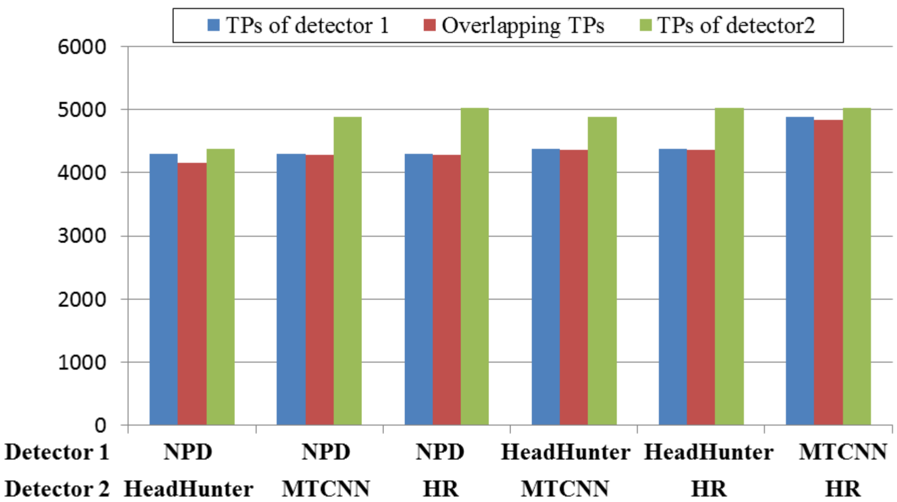
MTCNN (Zhang et al., 2016)

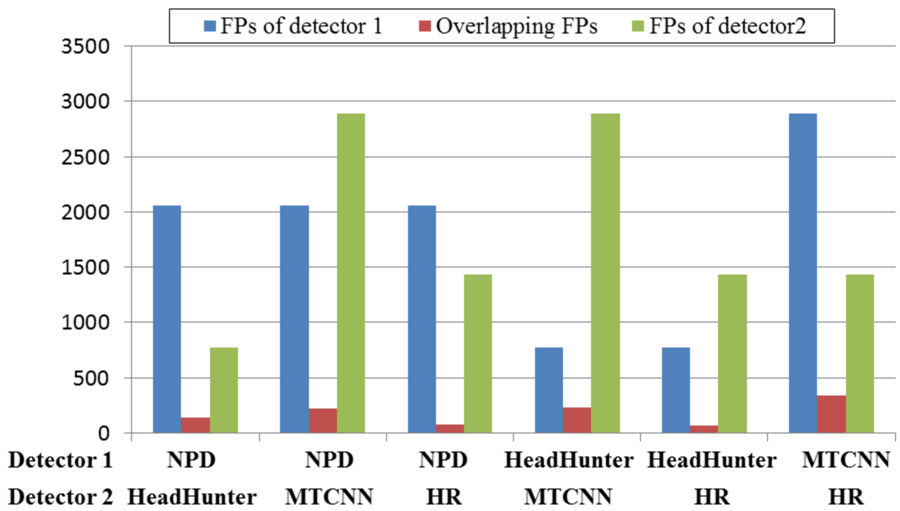
HR (Hu et al., 2017)

[1]S. Liao, A. K. Jain, and S. Z. Li. “A fast and accurate unconstrained face detector”. In PAMI, 2016..
[2] M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool. “Face detection without bells and whistles.” In ECCV, 2014.
[3] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. “Joint face detection and alignment using multitask cascaded convolutional networks.” In SPL, 2016.
[4] P. Hu and D. Ramanan. “Finding tiny faces.” In CVPR, 2017.

Correlation and Diversity

- The overlapping of true and false positives





Only a small number of false positives are detected by both detectors, whereas a majority of true positives overlap

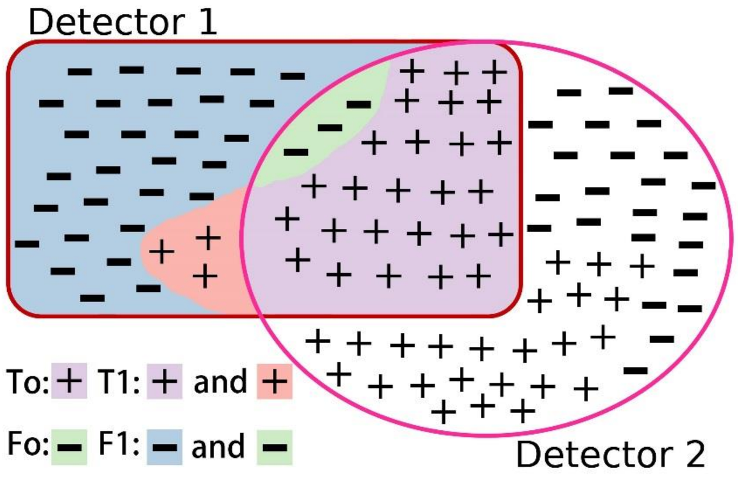
Correlation and Diversity

- Evaluation metrics
  - Correlation of true positives:

$$c^{T}_{2\to 1} = \frac{|T_o|}{|T_1|} \ ,$$

  - Diversity of false positives:

$$d^{F}_{2\to 1} = 1 - \frac{|F_o|}{|F_1|} \ ,$$



Cascade Properties

1. Correlation of true positives:

$$c^{T}_{2\rightarrow 1} \approx 1$$

2. Diversity of false positives:

$$d^{F}_{2\rightarrow 1} \approx 1$$

3. Detector runtime:

   Faster detector in the first stage to achieve an overall fast speed

Experiments

Method	CPU time (SPF*)			TPR (FPPI#=0.1)
	1st stage	2nd stage	total time	
VJ [24]	0.271	-	0.271	0.462
NPD [15]	0.678	-	0.678	0.801
NPD-HeadHunter	0.678	988	988.678	0.810
NPD-MTCNN	0.678	0.073	0.751	**0.841**
NPD-HR	0.678	2.678	3.356	**0.841**
HeadHunter [18]	1961	-	1961	0.834
HeadHunter-NPD	1961	0.404	1961.404	0.819
HeadHunter-MTCNN	1961	0.116	1961.116	**0.889**
HeadHunter-HR	1961	3.648	1964.648	**0.889**
MTCNN [30]	0.355	-	0.355	0.919
MTCNN-NPD	0.355	0.220	0.575	0.843
MTCNN-HeadHunter	0.355	456	456.355	0.882
MTCNN-HR	0.355	3.496	3.851	**0.930**
HR [4]	17.687	-	17.687	**0.943**
HR-NPD	17.687	0.170	17.857	0.839
HR-HeadHunter	17.687	794	811.687	0.886
HR-MTCNN	17.687	0.076	17.763	0.930

* SPF–Seconds Per Frame # FPPI–False Positives Per Image

Table 2: The correlation of true positives $c^T_{2\to1}$.

	Detector 2			
Detector 1	NPD [15]	HeadHunter [18]	MTCNN [30]	HR [4]
NPD [15]	1	0.9683	0.9970	0.9967
HeadHunter [18]	0.9487	1	0.9959	0.9961
MTCNN [30]	0.8755	0.8926	1	0.9900
HR [4]	0.8523	0.8694	0.9640	1

Table 3: The diversity of false positives $d^F_{2\to1}$.

	Detector 2			
Detector 1	NPD [15]	HeadHunter [18]	MTCNN [30]	HR [4]
NPD [15]	0	0.9339	0.8916	0.9645
HeadHunter [18]	0.8236	0	0.7030	0.9170
MTCNN [30]	0.9228	0.9207	0	0.8826
HR [4]	0.9491	0.9554	0.7636	0

Conclusions

- Two-stage cascade framework allowed us to improve the existing face detector with minimal effort
- Not all face detectors can be paired and with the proposed cascade properties, we could find the best pair of detectors