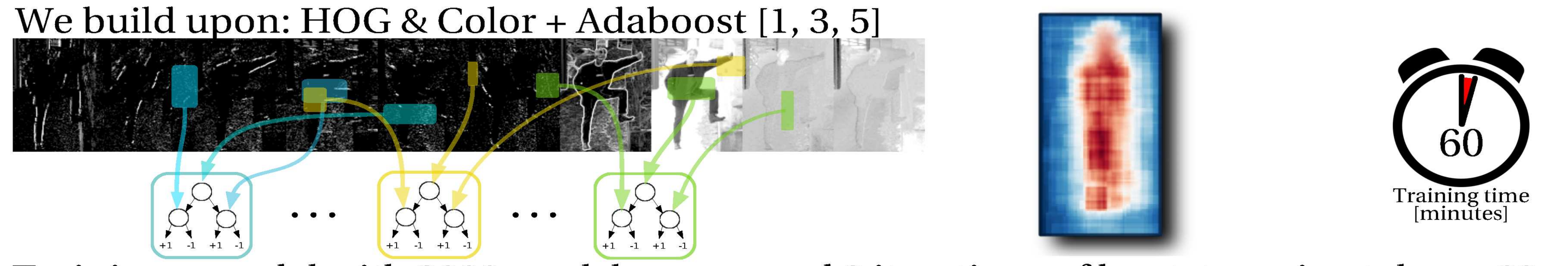
Handling Occlusions with Franken-classifiers

Markus Mathias[1], Rodrigo Benenson[2], Radu Timofte[3], and Luc Van Gool[1,3]

[1] ESAT-PSI-VISICS/IBBT, Katholieke Universiteit Leuven, Belgium [2] Max-Planck-Institut für Informatik, Saarbrücken, Germany [3] D-ITET/CVL, ETH Zürich, Switzerland

Background

We build upon: HOG & Color + Adaboost [1, 3, 5]



Training a model with 2000 weak learners and 3 iterations of bootstrapping takes ~ 60 min

Motivation

- Occlusions occur frequently
- 97% of occlusion cases are covered by a small subset of occlusion types [2]
- Approaches:
  - Most top performing methods do not handle occlusions explicitly; occlusions from the image border by border extrapolation
  - Train joint model for object + occluder
  - Train occlusion specific models, usually only 3~5

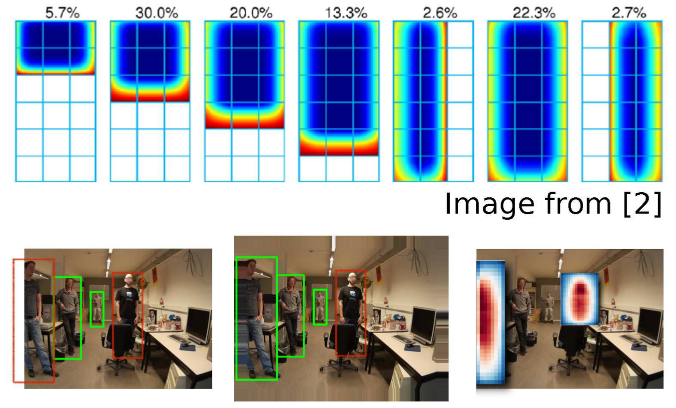
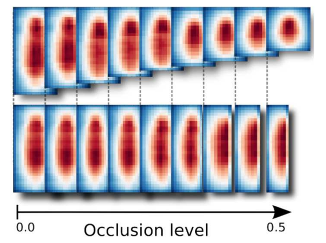


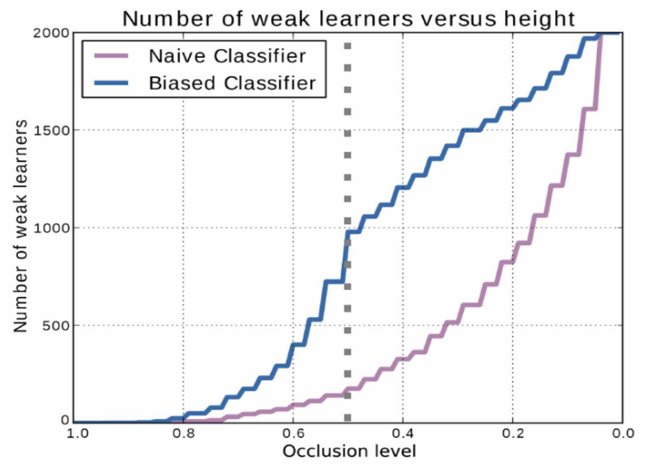
Image from [2]

We train 16+8+8+1=33 occlusion specific classifiers for different types and levels of occlusion



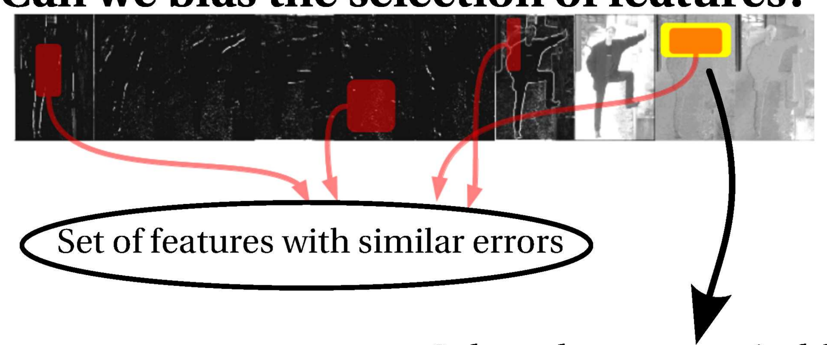
Key insight

- Feature pool huge, many features with similar errors
- AdaBoost algorithm is greedy: worse choices at one point may not affect quality
- Boosting compensates for bad choices

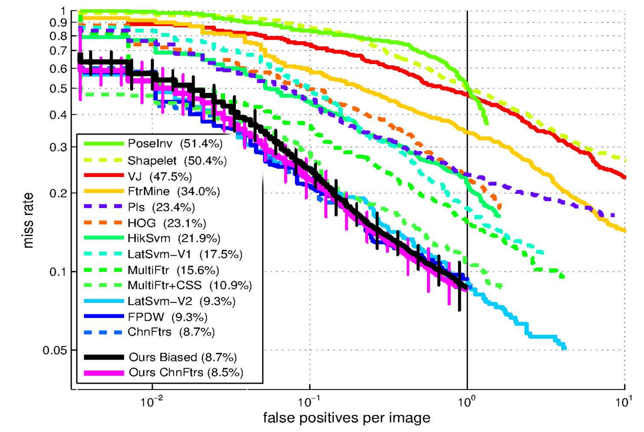


Biasing changes the feature distribution

Can we bias the selection of features?



Select the most suitable feature rather than the best feature



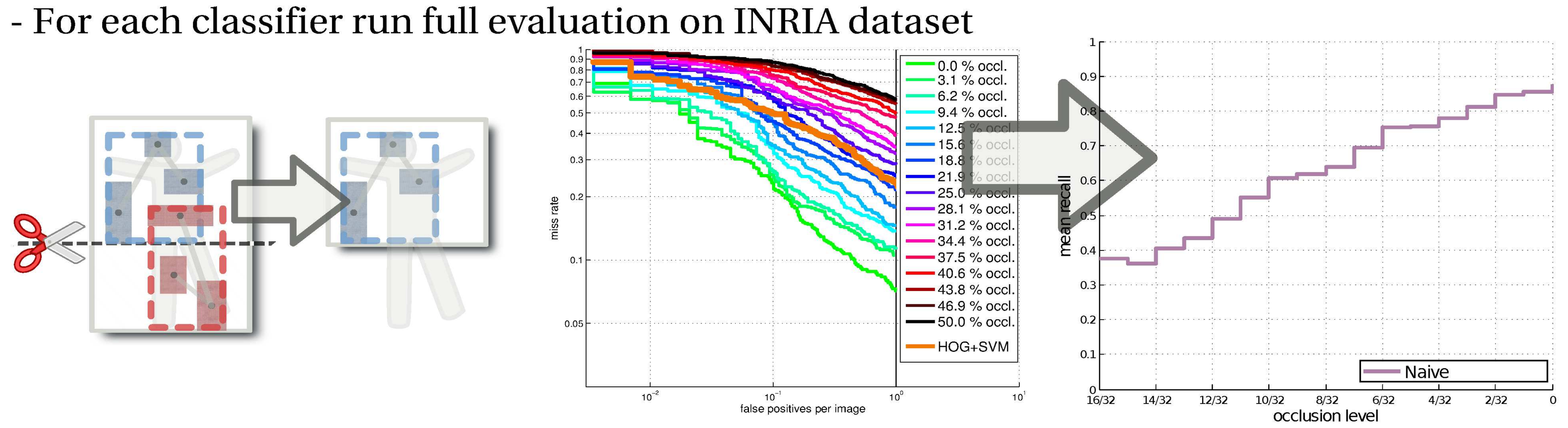
Biasing does not decrease quality

References

[1] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. CVPR 2005
[2] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian Detection: An Evaluation of the State of the Art. TPAMI, 2011
[3] P. Dollár, Z. Tu, P. Perona, and S. Belongie. Integral channel features. BMVC 2009
[4] R. Benenson, M. Mathias, R. Timofte and L. Van Gool. Pedestrian detection at 100 frames per second. CVPR 2012
[5] R. Benenson, M. Mathias, R. Timofte, Tinne Tuytelaars, and L. Van Gool. Seeking the strongest rigid detector. CVPR 2013

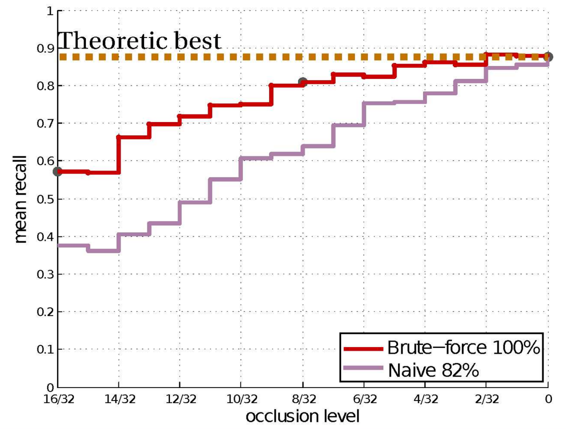
A. Naive approach: Cutting classifiers

- For each classifier run full evaluation on INRIA dataset

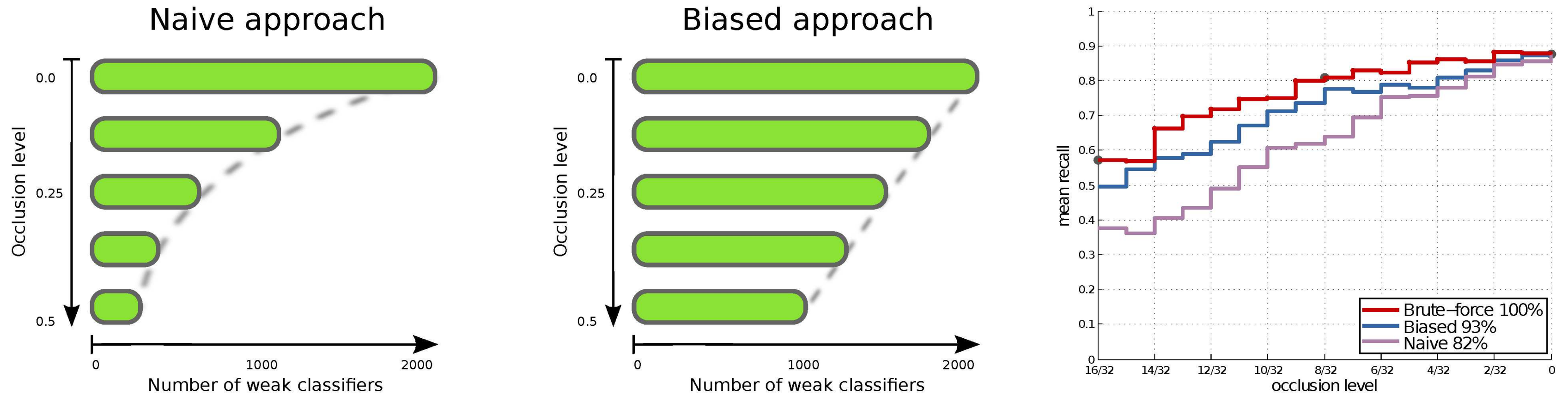


B. Brute force classifiers

- Train classifier for each possible occlusion level
- Theoretic best outcome: half pedestrian classifier performs as good as the full pedestrian classifier

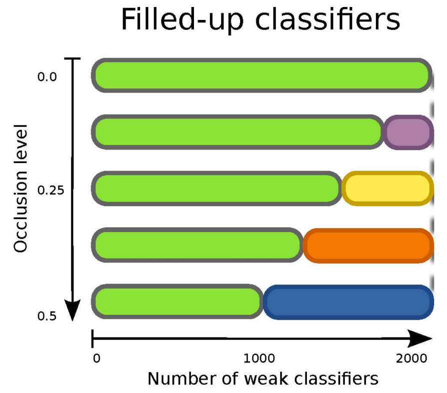


C. Biased approach: Cutting biased classifiers



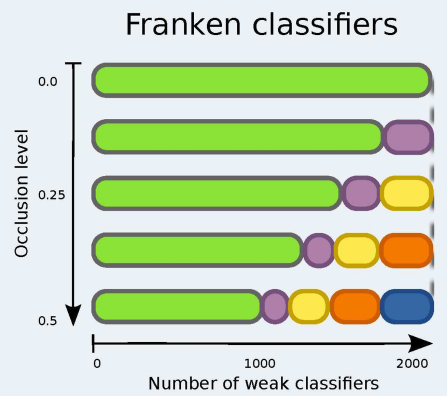
D. Fill-up classifiers

- Train a Biased classifier as basis
- Number of weak learners as proxy for quality
- Train additional features for each occlusion level



E. Franken-classifiers

- Train a Biased classifier as basis
- Train Fill-up classifiers in a recursive way
- Classifier of the last (most drastic) occlusion level will potentially have weak classifiers from all previous occlusion levels, hence the name Franken-classifiers



Contributions

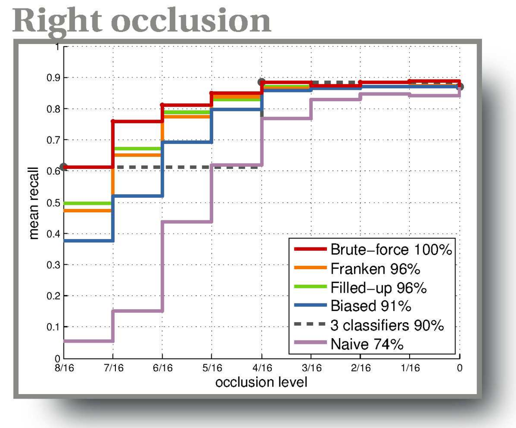
- We show that the Integral Channel Features Detector family is brittle under **occlusion** ([3,4,5])
- Concept of **biasing AdaBoost** training biased on a spatial criterion
- Franken classifiers with **sub-linear** cost for training and testing (to number of classifiers)
- Franken classifiers are **generic** and provide **state-of-the-art results** under occlusion

Results

How good is each occlusion specific classifier?

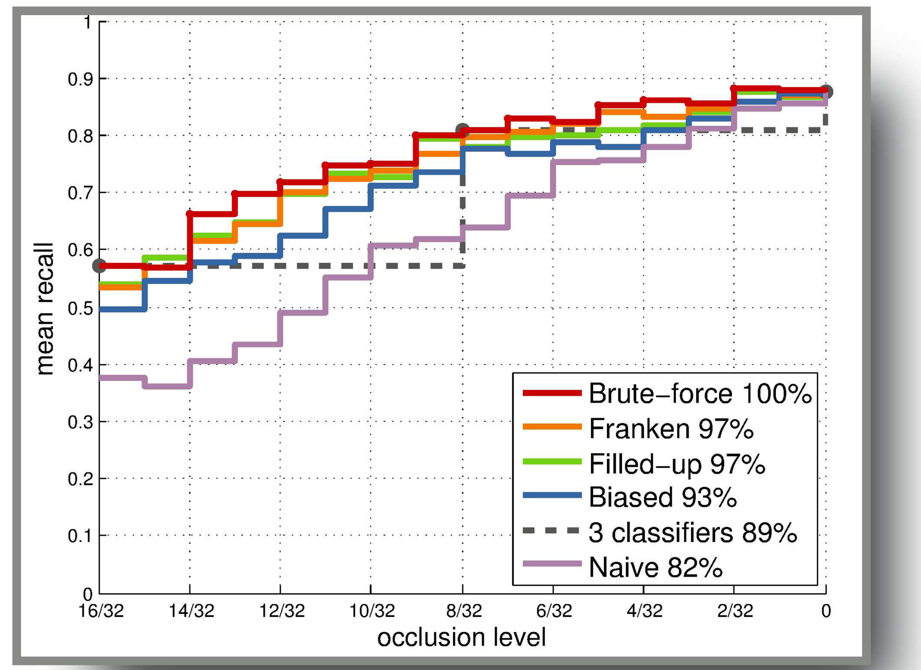
- Usage e.g. on the image border, depth discontinuities, around cars...
- More efficient than border extrapolation

Right occlusion



Type	Number of classifiers	Quality	Relative training time	Training time in minutes
Brute-force	17	100%	100%	1 088
5 classifiers	5	94%	29%	320
3 classifiers	3	89%	18%	192
Franken	1+16	**97%**	**8%**	64+28
Filled-up	1+16	97%	11%	64+54
Biased	1+16	93%	6%	64
Naive	1+16	82%	6%	64

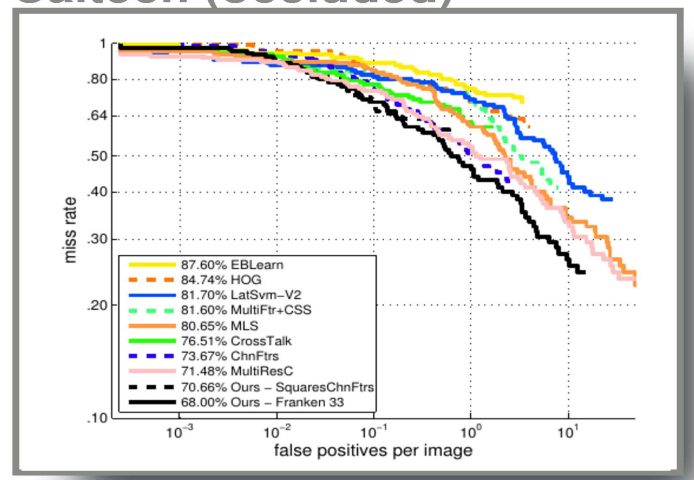
Bottom occlusion



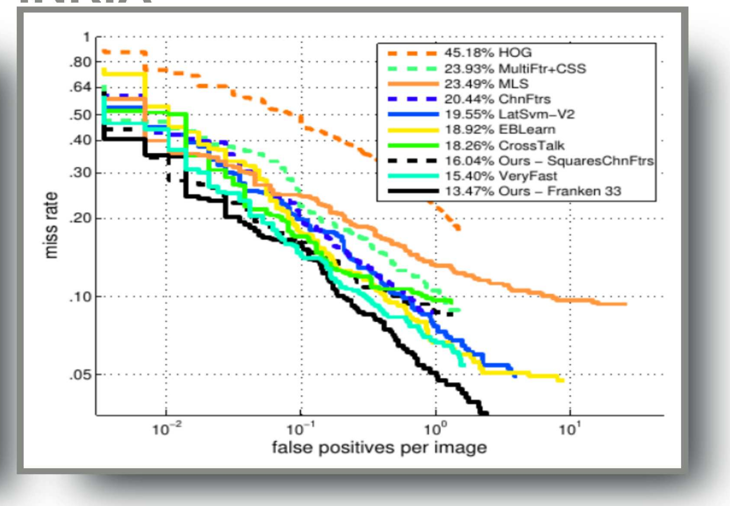
How good do all 33 classifiers perform jointly?

- Different types of occlusion classifiers share at least 1000 features
- Joint non-maximum suppression possible
- Sublinear in number of occlusion specific classifiers

Caltech (occluded)



INRIA



ETH