Filtered Channel Features for Pedestrian Detection

Shanshan Zhang, Rodrigo Benenson, Bernt Schiele
MPI Informatics, Saarbrücken, Germany

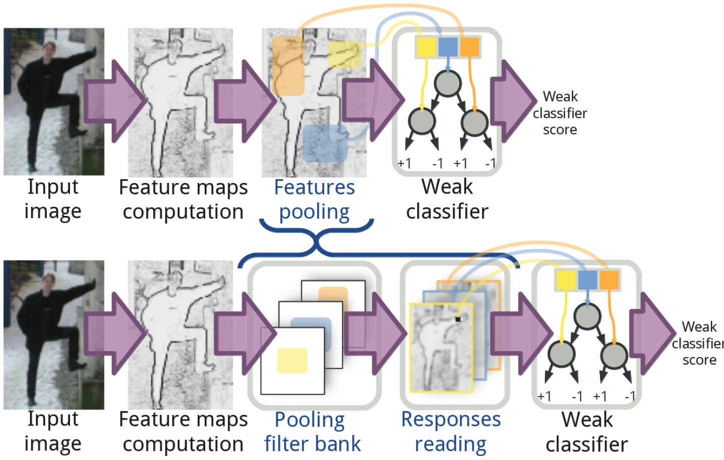

Figure 1: Integral channel feature detectors pool features via sums over rectangular regions. We can equivalently re-write this operation as convolution with a filter bank followed by single pixel reads. We aim to answer: *What is the effect of selecting different filter banks?*

Pedestrian detection is an active research area, with 1000+ papers published in the last decade, and well established benchmark datasets. It is considered a canonical case of object detection, and has served as playground to explore ideas that might be effective for generic object detection. Although many different ideas have been explored, and detection quality has been steadily improving [2], arguably it is still unclear how effective parts, components, and features learning are for this task.

Current top performing pedestrian detection methods all point to an intermediate layer (such as max-pooling or filtering) between the low-level feature maps and the classification layer [5, 6, 7]. In this paper we explore the simplest of such intermediary: a linear transformation implemented as convolution with a filter bank. We propose a framework for filtered channel features (see figure 1) that unifies multiple top performing methods [1, 3, 5, 7], and that enables a systematic exploration of different filter bank families.

It has been shown that using extra information at test time (such as context, stereo images, optical flow, etc.) can boost detection quality. In this paper we focus on the "core" sliding window algorithm using solely HOG+LUV features. We consider context information and optical flow as add-ons, included in the experiments section for the sake of completeness and comparison with existing methods. Using only HOG+LUV features we already reach top performance on the challenging Caltech and KITTI datasets, matching results using optical flow and significantly more features.

The main contributions of this paper can be summarized as follows:

- We point out the link between `ACF` [4], `ChnFtrs` [3], `SquaresChnFtrs` [1, 2], `InformedHaar` [7], and `LDCF` [5]; and introduce the "filtered channel features detectors" as a generalization of these methods.
- We provide extensive experiments to enable a systematic analysis of the filtered integral channels, covering aspects not explored by related work. We report the summary of 65+ trained models.
- We show that top detection performance can be reached on Caltech and KITTI using HOG+LUV features only. We report the best known results on Caltech.

Figure 2 presents our key results on the Caltech test set. For proper comparison, only methods using the same training set should be compared. We include for comparison: `Roerei` [1] the best known method trained without any Caltech images, and top performers `Katamari` [2] and `SpatialPooling+` [6]. Our experiments show that, with the proper filter bank, filtered channel features reach top detection quality.

Our results show that multiple filter bank families reach similar results; this indicates that expanding the feature channels via filtering is the key step for improving detection quality, while selecting the "perfect" filters is a secondary concern. Our results indicate that competitive results (over Caltech and KITTI datasets) can be obtained using only HOG+LUV features. When optical flow information is added we set the new state of the art for the Caltech dataset, reaching 17.1% MR (93% recall at 1 false positive per image).

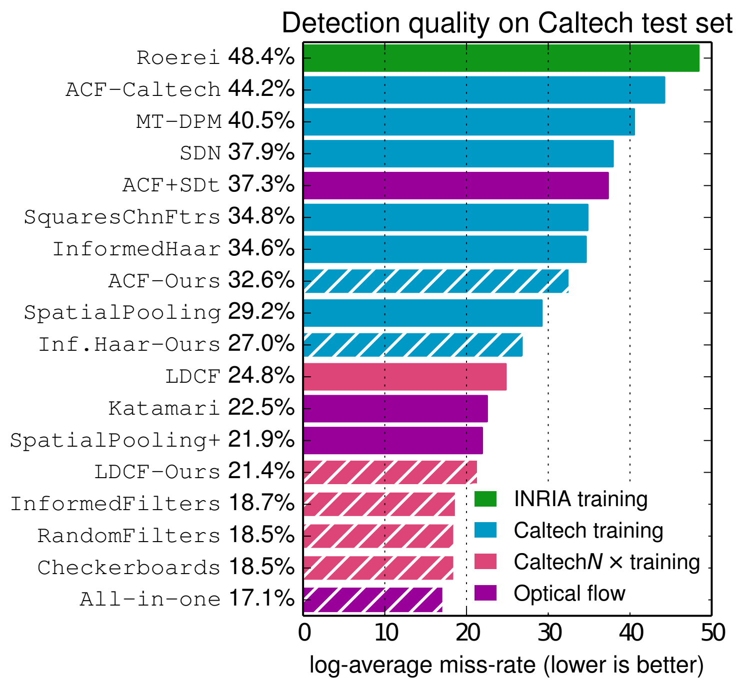

Figure 2: Some top-performing methods for Caltech test set, and our results (highlighted with white hatch, lower is better). Methods using optical flow are trained on original Caltech except our `All-in-one` which uses Caltech10x. Caltech$N\times$ indicates Caltech10x for all methods but the original `LDCF` [5].

[1] R. Benenson, M. Mathias, T. Tuytelaars, and L. Van Gool. Seeking the strongest rigid detector. In *CVPR*, 2013.

[2] R. Benenson, M. Omran, J. Hosang, , and B. Schiele. Ten years of pedestrian detection, what have we learned? In *ECCV, CVRSUAD workshop*, 2014.

[3] P. Dollár, Z. Tu, P. Perona, and S. Belongie. Integral channel features. In *BMVC*, 2009.

[4] P. Dollár, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. *PAMI*, 2014.

[5] W. Nam, P. Dollár, and J. Hee Han. Local decorrelation for improved detection. In *NIPS*, 2014.

[6] S. Paisitkriangkrai, C. Shen, and A. van den Hengel. Strengthening the effectiveness of pedestrian detection with spatially pooled features. In *ECCV*, 2014.

[7] S. Zhang, C. Bauckhage, and A. B. Cremers. Informed haar-like features improve pedestrian detection. In *CVPR*, 2014.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.