Multi-label Learning of Part Detectors for Heavily Occluded Pedestrian Detection

Chunluan Zhou Junsong Yuan
School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore
czhou002@e.ntu.edu.sg, jsyuan@ntu.edu.sg

1. Effectiveness of cost-sensitive learning

In the paper, we learn the K part detectors in a cost-sensitive way (See Section 3.2 of the paper for details). If a pedestrian example x_i is misclassified by a part detector d_k, a cost c_{ik} would be incurred which is defined as follows

$$c_{ik} = \begin{cases} O_{ik} & O_{ik} \geq 0.7; \\ I_{ik} & O_{ik} < 0.7 \text{ and } I_{ik} \geq 0.5; \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We always assign a label of 1 to a pedestrian example for all part detectors, *i.e.* $y_{ik} = 1$ for all $1 \leq k \leq K$, but associate different misclassification costs with the K part detectors as defined in Eq. (1). We call this label assignment strategy cost-sensitive labeling (CSL). To demonstrate the effectiveness of CSL, we conduct an experiment in which we assign hard labels to a pedestrian example for the K part detectors. Specifically, for a pedestrian example x_i, we set $y_{ik} = 1$ if $O_{ik} \geq 0.7$ or $I_{ik} \geq 0.5$ and set $y_{ik} = -1$ otherwise. $c_{ik} = 1$ for all i and k. We call this label assignment strategy hard labeling (HL). We learn the K part detectors jointly with CSL and HL respectively and compare the part detectors learned with CSL and those learned with HL on three subsets of the Caltech dataset [2], *Reasonable*, *Partial* and *Heavy*. Tables 1 and 2 show the experimental results using channel and CNN features respectively. We can see that for both types of features, the part detectors learned with CSL have much better performance than their counterparts learned with HL, which shows the importance of proper labeling to joint learning of part detectors.

Method	*Reasonable*	*Partial*	*Heavy*
P1-HL	20.9	44.2	79.5
P1-CSL	**17.0**	**34.2**	**70.5**
P4-HL	31.4	53.2	73.6
P4-CSL	**17.8**	**35.5**	**67.9**
P6-HL	22.6	44.0	82.0
P6-CSL	**17.3**	**34.1**	**70.7**
P11-HL	22.4	46.3	83.0
P11-CSL	**16.9**	**35.2**	**70.8**
Avg. Imp.	+7.1	+12.2	+9.6

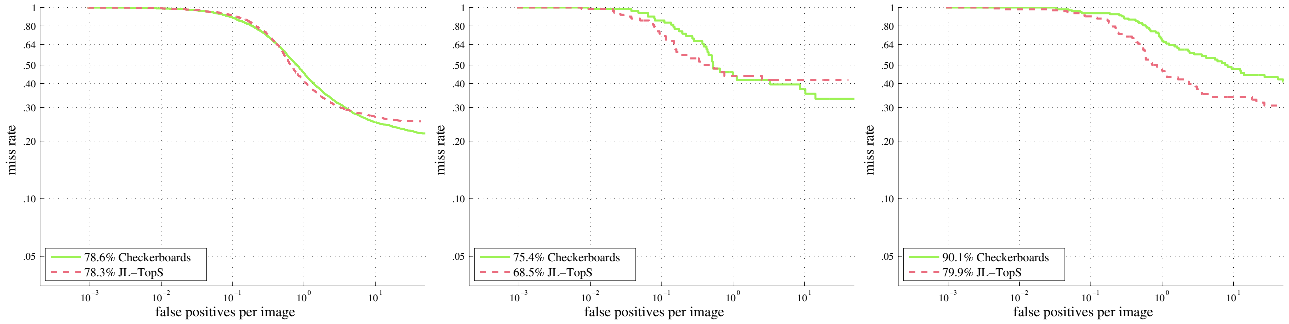
Table 1. Comparison of cost-sensitive labeling (CSL) and hard learning (HL) using channel features. P1, P4, P6 and P11 are four typical parts shown in Figure 2 of the paper. The last row shows the average improvements on the three subsets brought by cost-sensitive labeling.

Method	*Reasonable*	*Partial*	*Heavy*
P1-HL	10.8	22.9	66.6
P1-CSL	**9.9**	**17.2**	**50.5**
P4-HL	19.7	38.8	52.2
P4-CSL	**10.4**	**18.6**	**48.6**
P6-HL	11.9	27.0	73.1
P6-CSL	**10.2**	**18.3**	**51.6**
P11-HL	12.2	28.4	74.6
P11-CSL	**10.0**	**17.4**	**51.0**
Avg. Imp.	+3.5	+11.4	+13.7

Table 2. Comparison of cost-sensitive labeling (CSL) and hard learning (HL) using CNN features. P1, P4, P6 and P11 are four typical parts shown in Figure 2 of the paper. The last row shows the average improvements on the three subsets brought by cost-sensitive labeling.

2. Experiments on CUHK

To further demonstrate the effectiveness of our approach for occlusion handling, we test it on the CUHK dataset [3] which is specially collected for evaluating occlusion handling approaches for pedestrian detection. The dataset consists of 1063 images from Caltech, ETHZ, TUD-Brussels, INRIA, Caviar and other sources. Each image contains at least one partially occluded pedestrian. We perform training on Caltech and use CUHK for testing. As for Caltech, evaluation is carried out on three subsets, *Reasonable*, *Partial* and *Heavy*, and detection performance is summarized by log-average miss rate. For channel features, we compare

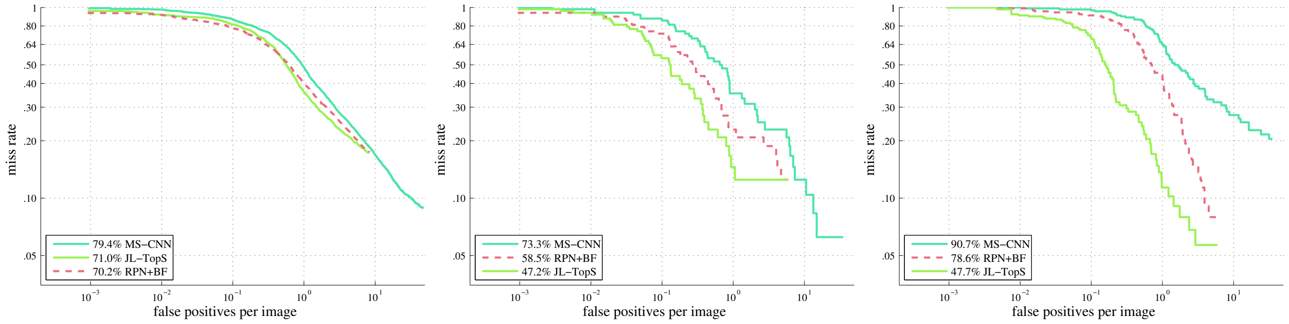


(a) *Reasonable*

(b) *Partial*

(c) *Heavy*

Figure 1. Results on CUHK using channel features.



(a) *Reasonable*

(b) *Partial*

(c) *Heavy*

Figure 2. Results on CUHK using CNN features.

our approach with Checkerboards [5]. Figure 1 shows the results of both approaches. Our approach (JL-TopS) outperforms Checkerboards on *Reasonable*, *Partial* and *Heavy* by 0.3%, 6.9% and 10.2% respectively. The performance improvements of our approach over Checkerboards on *Partial* and *Heavy* are significant. For CNN features, we compare our approach with RPN+BF [4] and MS-CNN [1]. The same RPN is used in RPN+BF and our approach for proposal generation and feature extraction. Figure 2 shows the results of the three approaches. MS-CNN performs poorly on this dataset. Our approach (JL-TopS) performs slightly worse than RPN+BF on *Reasonable* but outperforms it on *Partial* and *Heavy* by 11.3% and 10.9% respectively, which shows the advantage of our approach for occlusion handling.

References

[1] Z. Cai, M. Saberian, and Vasconcelos. A unified multi-scale deep convolutional neural network for fast object detection. In *European Conference on Computer Vision (ECCV)*, 2016.

[2] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI)*, 2012.

[3] W. Ouyang and X. Wang. A discriminative deep model for pedestrian detection with occlusion handling. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[4] L. Zhang, L. Lin, X. Liang, and K. He. Is faster r-cnn doing well for pedestrian detection? In *European Conference on Computer Vision (ECCV)*, 2016.

[5] S. Zhang, R. Benenson, and B. Schiele. Filtered channel features for pedestrian detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.