Active Learning for Deep Detection Neural Networks (Supplementary Materials)

Hamed H. Aghdam[1], Abel Gonzalez-Garcia[1], Joost van de Weijer[1], Antonio M. López[1,2]
Computer Vision Center (CVC)[1] and Computer Science Dpt.[2], Univ. Autònoma de Barcelona (UAB)
{haghdam,agonzalez,joost,antonio}@cvc.uab.es

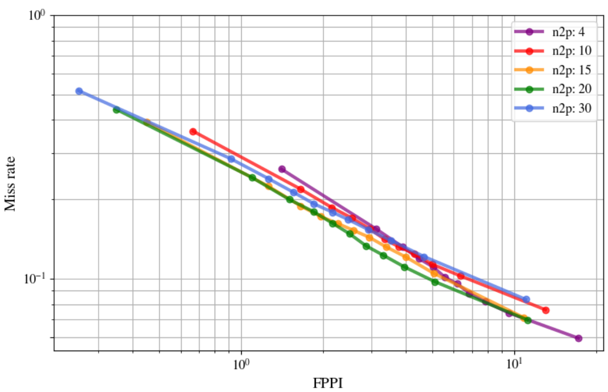

Figure 1. Pedestrian detection performance of our base neural network on the Caltech Pedestrian dataset, using different N2Ps (negative to positive ratios).

1. Caltech Pedestrian dataset

Lower-bound error. $\mathcal{X}_u$ is the set of unlabeled images that must be partially labeled by following our active learning method. In our experiments, $\mathcal{X}_u$ corresponds to either Caltech Pedestian dataset or BDD100K. This section focuses on the former, next section in the later.

In order to estimate a lower-bound error for our active learning method, we trained our detection network on the Caltech Pedestrian dataset using all labeled training frames and evaluated on its test set. Specifically, the network is trained using different negative-to-positive (N2P) ratios. For each N2P, we trained the network three times. Figure 1 illustrates the mean false positive per image (FPPI) *vs*. the miss rate [1].

We observe that the N2P affects the overall performance of the network. The minimum FPPI is greater than one when the N2P is fixed to 4. Moreover, as the N2P increases, the minimum FPPI is reduced. However, the maximum FPPIs are comparable when the N2P is greater than 10. Another way to compare these curves is to study their miss rates at $FPPI = 10^0$. This way, the network trained using

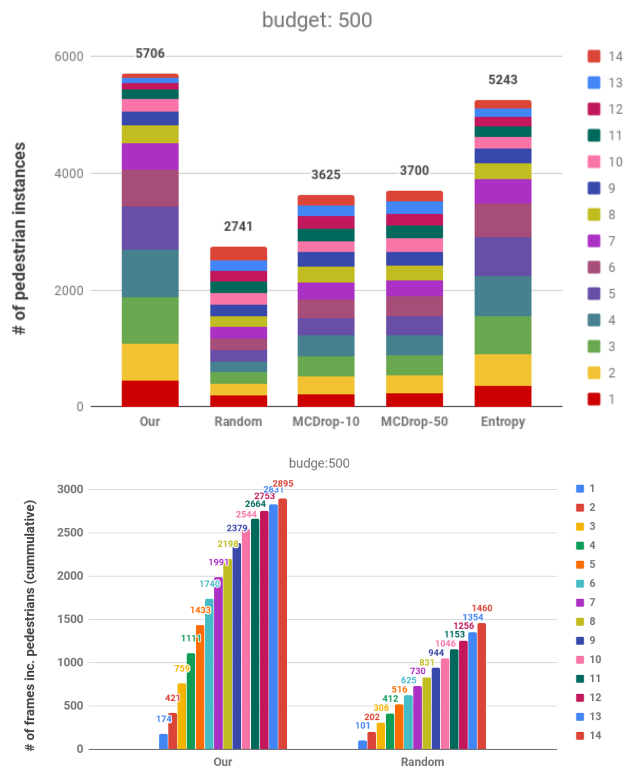

Figure 2. Statistics of $\mathcal{X}_{al}$ at each cycle in terms of number of pedestrian instances (top) and number of images containing at least one pedestrian instance (bottom) when $\mathcal{X}_u$ is the Caltech Pedestrian dataset. Bars correspond to cycles.

N2P=15 produces the best results (lower miss rate) [1].

Per cycle comparison. In the main submission, we compared our method and its variants of MC-Dropout and binary entropy with the guided random selection at specific cycles. In Figure 5, we compared these method at all cycles. We see how our active learning method and the guided

[1]The high value for N2P also depends on our implementation which is available at www.gitlab.com/haghdam/deep_active_learning

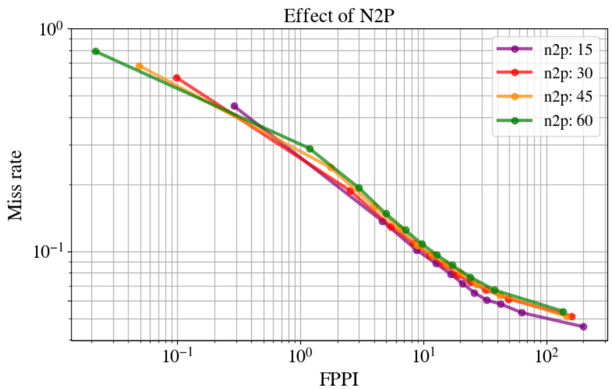

Figure 3. Pedestrian detection performance of our base neural network on the BDD100K dataset, using different N2Ps (negative to positive ratios).

random one select images which give rise to detectors of similar accuracy at 1st cycle. However, starting from the 2nd cycle, our method selects images which turn out in a more accurate detector.

Statistics of $\mathcal{X}_{al}$. We showed in our experiments that the number of pedestrian instances selected by our method is higher than for the guided random method. At each cycle, we also computed the number of frames in $\mathcal{X}_{al}$ that contains at least one pedestrian instance, both for our active learning method and guided random. Figure 2 shows the results.

At the end of cycle 14th, 2895 out of 7K frames (41$\%$ of frames) have at least one pedestrian instance when $\mathcal{X}_{al}$ is selected using our method. In contrast, 1460 out of 7K frames (21$\%$) contain pedestrian instances using the guided random method.

2. BDD100K dataset

Lower-bound error. Figure 3 illustrates the performance of our network on the BDD100K dataset using different N2Ps. The results show that our method is less accurate on the BDD100K dataset compared to the Caltech Pedestrian dataset. We think this is due to the fact that our network is too lightweight for BDD100K complexity. Thus, our immediate future work is to use a network with higher capacity for this case.

Per cycle comparison. For BDD100K, Figure 6 compares the detection performance based on the images selected by our active learning method *vs*. the ones selected by the guided random method, at each cycle. The results indicate that our method performs slightly better than the random selection. However, the improvement is not as significant as for the Caltech Pedestrian dataset. We think this is because for this dataset it is required a more complex network architecture able to reduce the bias.

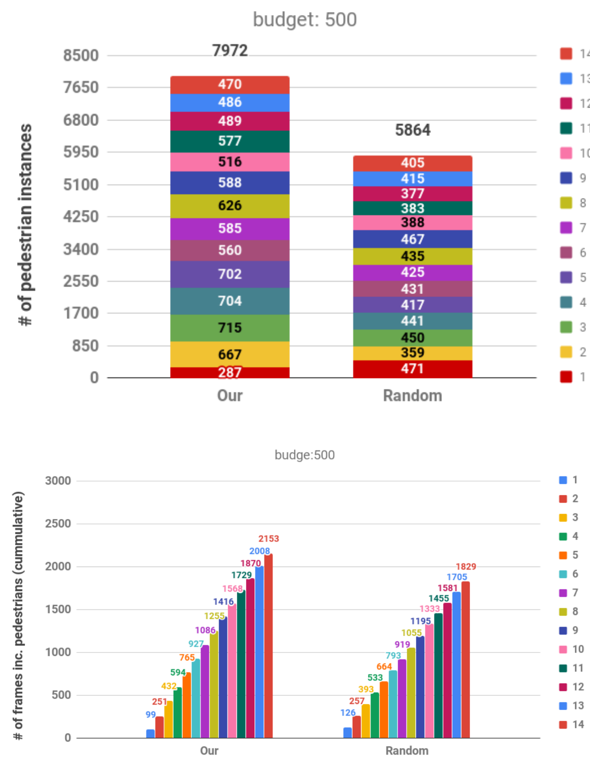

Figure 4. Statistics of $\mathcal{X}_{al}$ at each cycle in terms of number of pedestrian instances (top) and number of images containing at least one pedestrian instance (bottom) when $\mathcal{X}_u$ is the BDD100K dataset. Bars correspond to cycles.

Statistics of $\mathcal{X}_{al}$. We also computed the statistics of $\mathcal{X}_{al}$ for each cycle on the BDD100K dataset. Figure 4 illustrates the results. Similar to the Caltech Pedestrian datasets, our method selects frames with more pedestrian instances compared to the random selection. Moreover, the number of frames containing at least one pedestrian instance is higher using our method.

References

[1] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *TPAMI*, 2012. 1

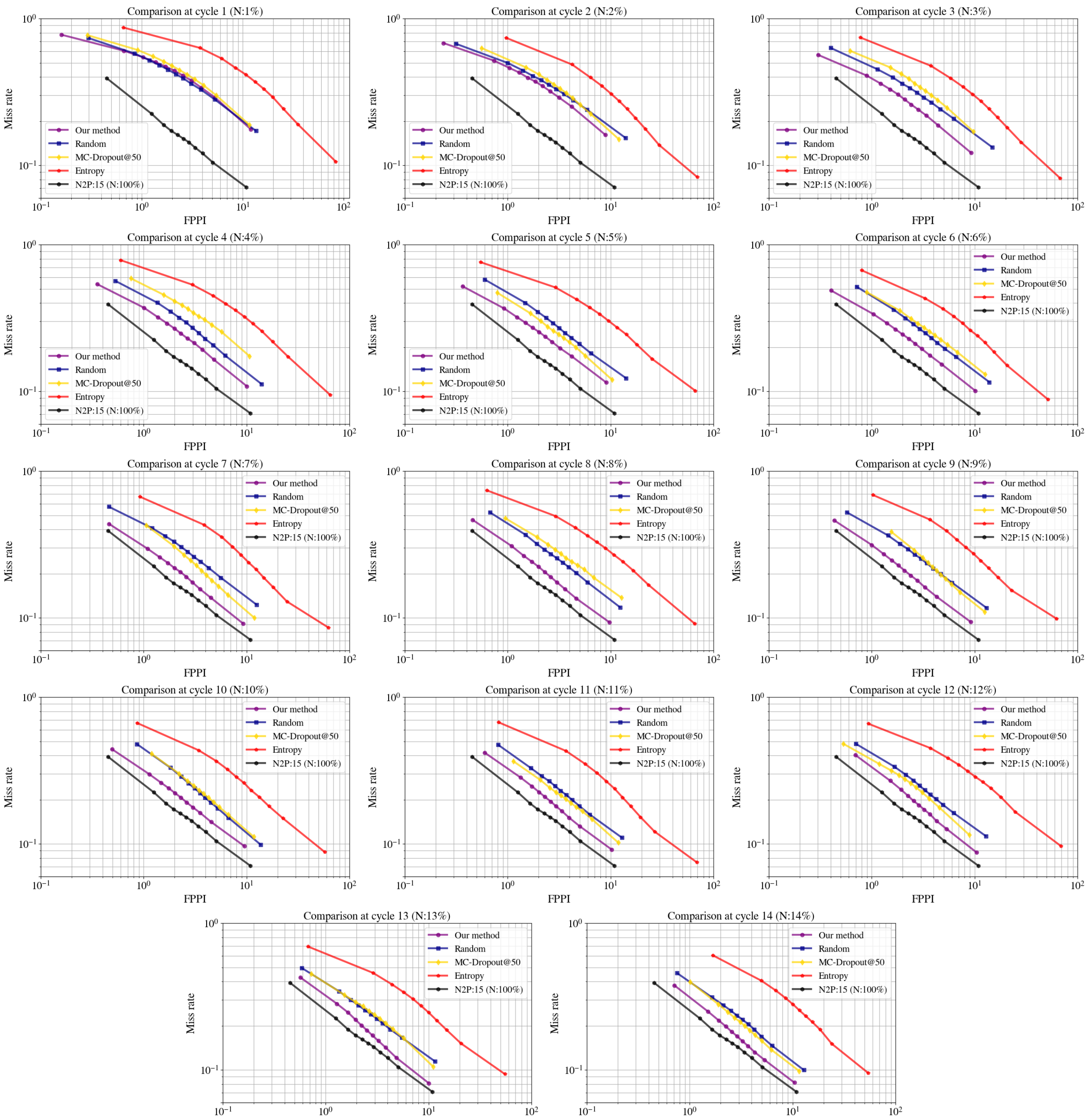

Figure 5. Comparing our method and its MC-Dropout and Entropy variants with the guided random selection at each cycle, for Caltech Pedestrian dataset.

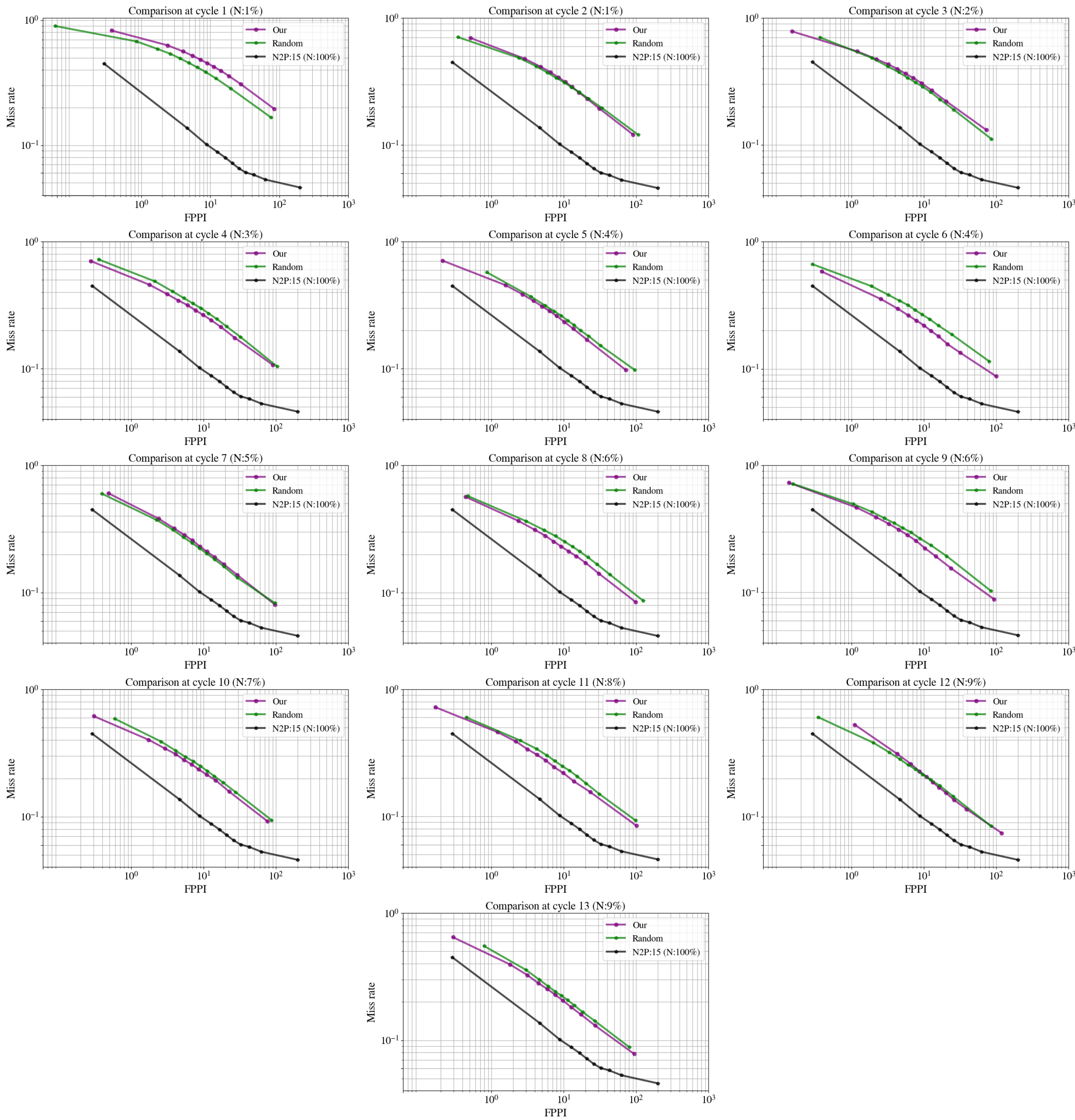

Figure 6. Comparing our method with random selection at each cycle, for BDD100K.