Supplementary Material: Few-Shot Supervised Prototype Alignment for Pedestrian Detection on Fisheye Images

Thaddäus Wiedemer[1,2,3], Stefan Wolf[2,3,4], Arne Schumann[3,4], Kaisheng Ma[1], Jürgen Beyerer[3,2]
[1]Tsinghua University, [2]Karlsruhe Institute of Technology,
[3]Fraunhofer IOSB, [4]Fraunhofer Center for Machine Learning
thaddaeus.wiedemer@uni-tuebingen.de, stefan.wolf@iosb.fraunhofer.de

A. Further Ablation Studies

In this section, we list the results of additional experiments that validate our choice of parameters. The training and evaluation settings are the same as in our ablation studies.

SPA-class thresholds. Table 1 lists the performance of SPA-class for different values of the foreground/background thresholds t_f and t_b. The exact values do not impact performance significantly. We choose $t_f = 0.75$ and $t_b = 0.25$ for the experiments.

Bounding box alignment variants. Table 2 is the full version of table 2 in the main paper, showing different methods for aligning features in the bounding box head. While aligning all four dimensions x, y, w, h together performs only marginally better than aligning each one individually, we still use the combination to avoid a bias.

Combining bounding box adaptation heads. When aligning bounding box features along all four offset dimensions, the naive implementation is to simply use four individual SPA heads on the same features. However, as shown in Table 3, this leads to interference between heads and decreases the performance compared to the individual heads. We experimented with reducing each head's influence on the main network by dampening gradients from each head by 0.25 or training heads on a round-robin schedule (in each epoch, only one head is active). The best overall performance is achieved by instead using a single SPA head for all dimensions. Compared to individual heads, this implements only a single FC layer for feature embeddings, which are then used to construct the losses along all bounding box offset dimensions. Thus, we choose the combined head for the experiments.

Combining multiple prototype-based adaptation heads. Both GPA and SPA can be used directly on the RoI features after pooling, which are shared by the classification and localization head. Alternatively, they can be implemented on features specific to either head after the respective FC layers. For SPA, alignment on shared features and in the classification head is performed by class, while alignment on bounding box features is performed by each offset dimension. GPA can only align by class. In addition to the final method SPA, comprising SPA-class and SPA-bbox, we also tested a variant "SPA all" that additionally aligns the shared features. Similarly, we re-implemented GPA, which aligns shared features and class-specific features, and compared it to "GPA all", which additionally aligns features in the bounding box regressor. The performance of all models is shown in figure Fig. 2. SPA performs best; additional alignment of the shared features only decreases performance.

Size and number of samples for adversarial adaptation. In Table 4, the results for varying size s and number n of samples in the adversarial adaptation head for alignment on `neck 0` are shown. The results indicate better alignment for larger sample sizes, while the number of samples is only relevant for smaller sample sizes. In our experiments, we choose $n = 32$ and $s = 35$, but the results here indicate that fewer and smaller samples should work well when resources are constrained.

Combining multiple adversarial adaptation heads. We also evaluated the combination of multiple adversarial adaptation heads applied to the backbone and the FPN. The results are shown in Table 5. The results show that adversarial adaptation is more effective on the FPN than the backbone due to the features of the FPN carrying semantically richer information better suited for alignment. The results on the combinations of FPN feature maps on PIROPO-20a are inconclusive. Therefore, we evaluated the best combinations (layer 0, layers 0 and 4, and layers 0, 1, and 4) additionally using all subsets of PIROPO as shown in Figure 3. The results are best for `neck 0`. Thus, adaptation on this layer is chosen for the final model.

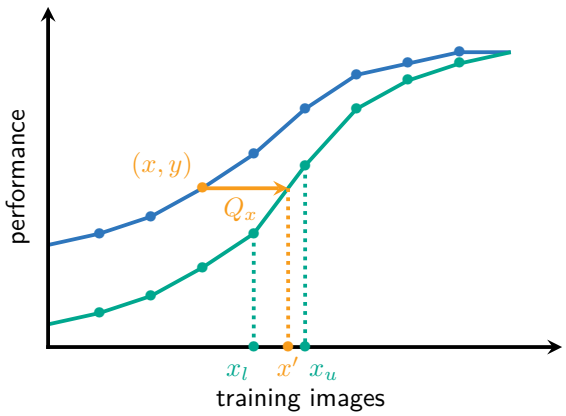

Figure 1. Computation of dataset size factors between models. The green line shows the accuracy of a baseline and the blue line shows an improved model.

B. Calculation of dataset size factor

The calculation of the dataset size factor is visualized in Figure 1. The factor indicates how much larger the training set would need to be for the baseline model to achieve the same performance as the improved model. The basis for the calculation are the accuracy curves over the sizes of the training set. For each point (x, y) of the improved model, a quotient

$$Q_x = \frac{x'}{x} \tag{1}$$

is calculated, where x' is the number of images needed or the baseline model to reach accuracy y. x' is approximated by linearly interpolating the two neighboring data points (x_l, y_l) and (x_u, y_u) on the baseline accuracy curve. The linear interpolation is performed in log-space as

$$x' = x_l \cdot \left(\frac{x_u}{x_l}\right)^f \quad \text{with} \quad f = \frac{y - y_l}{y_u - y_l} \tag{2}$$

since our results indicate a roughly linear increase of accuracy with an exponential increase of images. The final factor is calculated as the average quotient $\bar{Q}_x$ over every training set size.

C. Performance by object characteristics

In Figure 4, we compare the performance of our final model, the fine-tuning baseline, and the model purely pre-trained on COCO in terms of object size, object distance to the center, and angle of the object. The fine-tuning baseline and our final model were trained on PIROPO-5a (i.e., containing only 5 images). For each category, we report performances as AP with IoU 0.75 and IoU 0.5 (putting stronger or weaker emphasis on localization), as well as with IoU 0.1 (ignoring localization almost entirely), and IoU 0.1 with corrected false positives (ignoring localization and classification errors). The object sizes small, medium, and large correspond to the COCO sizes with $< 32^2$, $< 96^2$, and $> 96^2$ pixels, respectively.

The model purely pre-trained on COCO struggles to detect small objects and objects far from the center. Additionally, detection performance is highly dependent on the angle of objects. These results highlight the challenges our adaptation setting provides. Just as in previous experiments, fine-tuning increases performance by a large margin. However, model performance still depends on the object angle and decreases with object size and distance to the image center. Interestingly, the fine-tuned model achieves higher AP@50 on objects with a medium distance to the image center than on objects directly in the center, which are seen directly from above. While most of these top-views are detected correctly, the model still misses some of these objects, and fine-tuning does not improve performance for this case. Our domain-adaptive training boosts detection performance across all characteristics and especially improves localization performance for objects in the periphery and rotated objects. This demonstrates that even with five training images, our method can provide important regularization to the fine-tuning.

D. Discussion of ethical aspects

For the PIROPO [1] and the COCO [2] datasets, no information is provided about whether consent was obtained from the people visible on the images. For the Mirror Worlds [3] dataset, the authors state that most of the data is collected with IRB-approved studies and volunteers.

E. Full data

Performance curves in Figures 3 and 4 of the main paper as well as in Figures 2 and 3 show the mean performance over three datataset splits per size. The numeric values including the 95 % confidence interval for all curves are listed in Tables 6 to 22.

References

[1] *PIROPO database*. [Online]. Available: https://sites.google.com/site/piropodatabase/.

[2] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *Computer Vision – ECCV 2014*, Annotations are licensed under Creative Commons BY, Images must be used according to the Flickr Terms of Use, 2014, pp. 740–755.

[3] *Mirror Worlds Challenge*, Dataset is licensed under Creative Commons BY-NC-SA. [Online]. Available: http://www2.icat.vt.edu/mirrorworlds/challenge/index.html.

t_f	t_b	mAP↑	AP@75↑	AP@50↑	LAMR↓
0.90	0.10	0.566	0.647	0.881	0.162
0.75	0.25	**0.582**	0.650	**0.899**	**0.143**
0.70	0.30	0.578	**0.652**	0.895	0.144
0.50	0.50	0.581	0.642	0.893	0.149

Table 1. Performance for different threshold values in SPA-class on PIROPO-20a. As long as no extreme values are chosen, the exact values do not have a significant impact.

Align	GT	mAP↑	AP@75↑	AP@50↑	LAMR↓
class	✗	0.574	0.643	0.886	0.153
class	✓	0.579	0.642	**0.901**	0.138
Δx	✓	0.589	0.662	0.900	**0.137**
Δy	✓	0.589	0.665	0.894	0.141
Δw	✓	0.589	0.660	0.896	0.139
Δh	✓	0.589	0.662	0.900	0.140
$\forall \Delta d$	✓	**0.591**	**0.666**	0.899	0.139

Table 2. Performance on PIROPO-20a for different adaptation methods in the bounding box head. The combined alignment of all bounding box dimensions yields the best overall performance, especially on mAP and AP@75 which put greater emphasis on correct localization. GT denotes whether ground-truth information was used.

Method	mAP↑	AP@75↑	AP@50↑	LAMR↓
Individual heads	0.556	0.636	0.890	0.149
├ dampened gradients	0.586	0.663	0.892	0.146
├ round-robin training (RR)	0.579	**0.671**	0.883	0.155
└ RR + dampened gradients	0.582	0.662	0.893	0.143
Combined head	**0.591**	0.666	**0.899**	**0.139**

Table 3. Performance for different methods of combining bounding box adaptation heads in SPA on PIROPO-20a. Using a combined head with a single shared FC layer for all dimensions performed best.

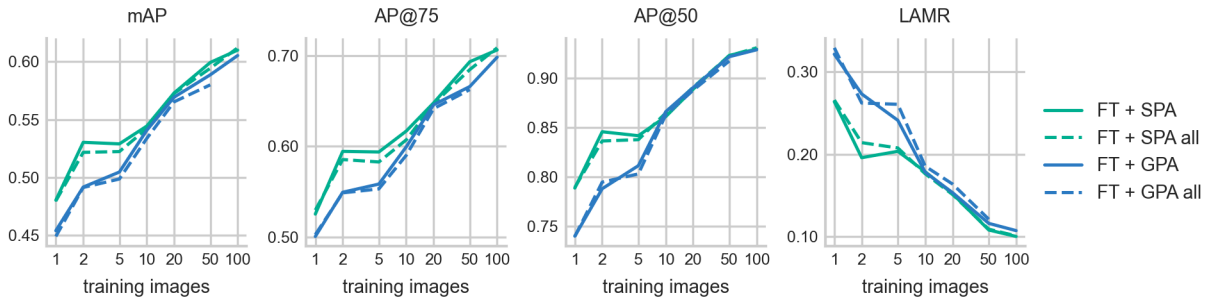

Figure 2. Performance of different versions of GPA and SPA on PIROPO. For GPA (blue), we compare aligning only the RoI features and classification features versus additionally aligning features in the bounding box regressor. For SPA (green), we compare aligning classification and bounding box regression features (final model in the main paper) versus additionally aligning the shared RoI features.

	mAP↑				AP@50↑			
n	$s=3$	$s=7$	$s=13$	$s=19$	$s=3$	$s=7$	$s=13$	$s=19$
16	0.581	0.581	0.594	0.596	0.902	0.908	0.908	0.912
20	0.581	0.588	0.591	0.594	0.900	0.904	0.911	0.908
30	0.578	0.594	0.594	0.593	0.897	0.909	0.907	0.915
50	0.584	0.589	0.596	0.596	0.906	0.902	0.910	0.912
100	0.581	0.590	0.593	0.596	0.899	0.906	0.905	0.912

Table 4. Performance of ADV on `neck 0` for different numbers and sizes of samples on PIROPO-20a. Performance clearly increases with sample size s and number n.

FPN			Backbone							
0	1	4	0	1	2	3	mAP↑	AP@75↑	AP@50↑	LAMR↓
✓	✓	✓					0.598	**0.692**	0.902	0.140
✓	✓						0.586	0.673	0.894	0.147
✓		✓					**0.599**	0.679	**0.910**	0.131
	✓	✓					0.587	0.682	0.896	0.151
✓							0.595	0.675	**0.910**	**0.130**
	✓						0.566	0.651	0.875	0.171
		✓					0.579	0.666	0.899	0.144
			✓				0.565	0.632	0.870	0.172
				✓			0.567	0.637	0.875	0.167
						✓	0.571	0.640	0.874	0.161
			✓	✓	✓	✓	0.565	0.632	0.869	0.170

Table 5. Performance of ADV for alignment of different feature maps in the first stage of Faster R-CNN on PIROPO-20a. Since results were inconclusive, the best three versions were tested on the full range of training sets (see below).

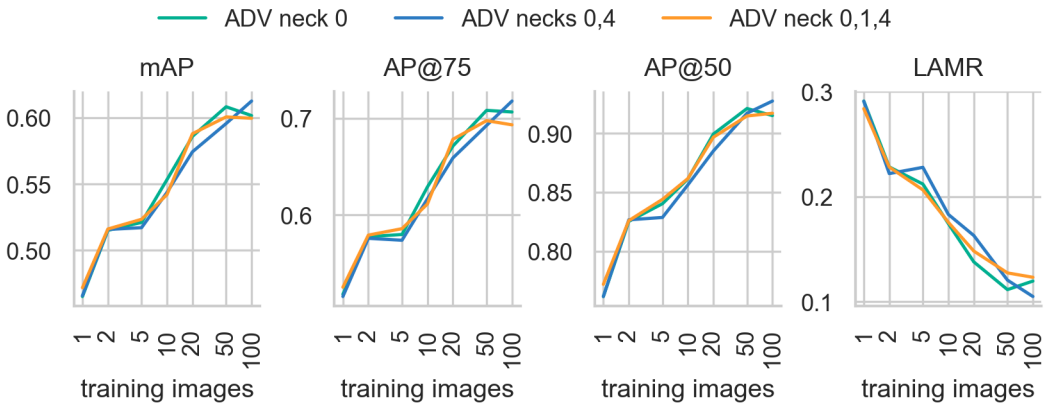

Figure 3. Performance of ADV for alignment of different combination of FPN feature maps on PIROPO-20a. Overall, adaptation only on `neck 0` (green) performed slightly better than other methods and was used in the final model, but differences in performance are very small.

Figure 4. Model performance measured as AP across object size (left), distance to image center (center), and angle of rotation (right) on PIROPO for the pre-trained model without fine-tuning (top), the fine-tuning baseline (middle), and the final method (bottom). Shown are the AP@75 (dashed/cross-hatched green), AP@50 (green), AP@10 (blue), and AP@10 with corrected classifications (orange). AP@10 corresponds to the model's performance if all bounding boxes were predicted correctly, and AP@10 with corrected classifications corresponds to performance if all detections had been correct. The difference between this last metric and $100\,\%$ is due to objects left entirely undetected.

While the pre-trained model performs noticeably worse for small objects and objects towards the periphery, the fine-tuning and adaptation models almost completely close this gap. Adaptation additionally improves performance across all sizes, distances, and angles.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
0	0.383	0.441	0.594	0.458
1	0.466 ± 0.023	0.519 ± 0.022	0.760 ± 0.038	0.293 ± 0.039
2	0.512 ± 0.011	0.575 ± 0.023	0.822 ± 0.009	0.230 ± 0.019
5	0.508 ± 0.020	0.565 ± 0.025	0.814 ± 0.027	0.234 ± 0.023
10	0.535 ± 0.024	0.607 ± 0.045	0.850 ± 0.018	0.190 ± 0.022
20	0.561 ± 0.009	0.635 ± 0.008	0.865 ± 0.012	0.176 ± 0.018
50	0.592 ± 0.014	0.684 ± 0.013	0.916 ± 0.014	0.118 ± 0.017
100	0.606 ± 0.011	0.700 ± 0.015	0.920 ± 0.008	0.111 ± 0.006
200	0.635 ± 0.012	0.736 ± 0.013	0.936 ± 0.017	0.088 ± 0.018
500	0.664 ± 0.006	0.780 ± 0.007	0.953 ± 0.002	0.064 ± 0.002
1000	0.687 ± 0.001	0.825 ± 0.006	0.963 ± 0.002	0.052 ± 0.003
2357	0.709	0.854	0.973	0.037

Table 6. Mean and CI-95 for fine-tuning baseline on PIROPO.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
0	0.400	0.430	0.655	0.448
1	0.493 ± 0.028	0.536 ± 0.016	0.831 ± 0.059	0.231 ± 0.075
2	0.461 ± 0.056	0.486 ± 0.054	0.768 ± 0.098	0.321 ± 0.111
5	0.524 ± 0.018	0.572 ± 0.035	0.873 ± 0.018	0.193 ± 0.026
10	0.568 ± 0.016	0.624 ± 0.022	0.926 ± 0.009	0.126 ± 0.014
20	0.581 ± 0.001	0.646 ± 0.011	0.933 ± 0.010	0.105 ± 0.024
50	0.602 ± 0.011	0.681 ± 0.026	0.956 ± 0.005	0.072 ± 0.009
100	0.614 ± 0.005	0.699 ± 0.005	0.964 ± 0.005	0.056 ± 0.007
200	0.640 ± 0.004	0.750 ± 0.011	0.973 ± 0.002	0.042 ± 0.006
819	0.681	0.809	0.978	0.025

Table 7. Mean and CI-95 for fine-tuning baseline on Mirror Worlds.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.403 ± 0.026	0.457 ± 0.024	0.633 ± 0.053	0.424 ± 0.040
2	0.414 ± 0.012	0.470 ± 0.018	0.668 ± 0.026	0.404 ± 0.018
5	0.418 ± 0.018	0.474 ± 0.018	0.666 ± 0.039	0.399 ± 0.039
10	0.422 ± 0.007	0.477 ± 0.007	0.681 ± 0.013	0.395 ± 0.031
20	0.443 ± 0.011	0.493 ± 0.007	0.714 ± 0.018	0.351 ± 0.020
50	0.443 ± 0.008	0.490 ± 0.009	0.715 ± 0.020	0.354 ± 0.009
100	0.436 ± 0.021	0.484 ± 0.022	0.703 ± 0.038	0.370 ± 0.044

Table 8. Mean and CI-95 for unsupervised GPA on PIROPO.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.395 ± 0.008	0.452 ± 0.009	0.624 ± 0.014	0.433 ± 0.003
2	0.399 ± 0.005	0.460 ± 0.004	0.627 ± 0.008	0.426 ± 0.013
5	0.399 ± 0.013	0.456 ± 0.013	0.632 ± 0.018	0.422 ± 0.017
10	0.402 ± 0.017	0.459 ± 0.020	0.634 ± 0.027	0.423 ± 0.020
20	0.425 ± 0.017	0.480 ± 0.007	0.684 ± 0.042	0.380 ± 0.049
50	0.439 ± 0.019	0.495 ± 0.018	0.703 ± 0.030	0.353 ± 0.024
100	0.437 ± 0.021	0.490 ± 0.020	0.707 ± 0.038	0.352 ± 0.044

Table 9. Mean and CI-95 for unsupervised ADV on PIROPO.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.454 ± 0.033	0.501 ± 0.030	0.740 ± 0.073	0.321 ± 0.067
2	0.492 ± 0.016	0.549 ± 0.028	0.788 ± 0.016	0.273 ± 0.018
5	0.505 ± 0.019	0.558 ± 0.037	0.811 ± 0.031	0.241 ± 0.031
10	0.541 ± 0.028	0.599 ± 0.044	0.866 ± 0.020	0.178 ± 0.023
20	0.569 ± 0.012	0.646 ± 0.014	0.890 ± 0.009	0.153 ± 0.013
50	0.589 ± 0.008	0.666 ± 0.003	0.922 ± 0.010	0.116 ± 0.012
100	0.605 ± 0.007	0.698 ± 0.001	0.929 ± 0.005	0.107 ± 0.005

Table 10. Mean and CI-95 for FT + GPA on PIROPO.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.449 ± 0.037	0.503 ± 0.032	0.740 ± 0.059	0.329 ± 0.071
2	0.491 ± 0.014	0.549 ± 0.032	0.795 ± 0.026	0.262 ± 0.033
5	0.499 ± 0.024	0.553 ± 0.036	0.803 ± 0.036	0.261 ± 0.047
10	0.534 ± 0.027	0.590 ± 0.043	0.863 ± 0.025	0.185 ± 0.031
20	0.565 ± 0.009	0.641 ± 0.023	0.888 ± 0.011	0.164 ± 0.015
50	0.580 ± 0.010	0.662 ± 0.010	0.917 ± 0.008	0.121 ± 0.009
2357	0.702	0.840	0.971	0.039

Table 11. Mean and CI-95 for FT + GPA all on PIROPO.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.480 ± 0.013	0.525 ± 0.019	0.789 ± 0.024	0.264 ± 0.026
2	0.530 ± 0.018	0.594 ± 0.032	0.846 ± 0.004	0.196 ± 0.003
5	0.529 ± 0.018	0.594 ± 0.025	0.842 ± 0.017	0.204 ± 0.002
10	0.544 ± 0.022	0.617 ± 0.033	0.861 ± 0.018	0.179 ± 0.014
20	0.573 ± 0.005	0.648 ± 0.015	0.889 ± 0.005	0.152 ± 0.001
50	0.599 ± 0.014	0.693 ± 0.019	0.923 ± 0.008	0.108 ± 0.001
100	0.610 ± 0.011	0.706 ± 0.011	0.929 ± 0.004	0.100 ± 0.008
2357	0.711	0.846	0.971	0.042

Table 12. Mean and CI-95 for FT + SPA on PIROPO.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.480 ± 0.008	0.530 ± 0.009	0.789 ± 0.020	0.266 ± 0.021
2	0.522 ± 0.015	0.585 ± 0.023	0.836 ± 0.005	0.214 ± 0.010
5	0.522 ± 0.021	0.583 ± 0.030	0.838 ± 0.024	0.208 ± 0.025
10	0.542 ± 0.026	0.607 ± 0.038	0.865 ± 0.018	0.176 ± 0.001
20	0.573 ± 0.007	0.647 ± 0.010	0.891 ± 0.005	0.151 ± 0.003
50	0.594 ± 0.009	0.684 ± 0.011	0.923 ± 0.011	0.109 ± 0.015
100	0.612 ± 0.006	0.708 ± 0.011	0.931 ± 0.003	0.101 ± 0.007
2357	0.707	0.850	0.971	0.040

Table 13. Mean and CI-95 for FT + SPA all on PIROPO.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.465 ± 0.024	0.518 ± 0.029	0.762 ± 0.045	0.291 ± 0.044
2	0.515 ± 0.013	0.577 ± 0.019	0.825 ± 0.015	0.228 ± 0.025
5	0.521 ± 0.033	0.580 ± 0.031	0.840 ± 0.042	0.212 ± 0.033
10	0.553 ± 0.015	0.629 ± 0.017	0.861 ± 0.017	0.174 ± 0.018
20	0.586 ± 0.013	0.671 ± 0.006	0.899 ± 0.014	0.138 ± 0.013
50	0.608 ± 0.005	0.708 ± 0.011	0.921 ± 0.006	0.112 ± 0.007
100	0.602 ± 0.010	0.707 ± 0.016	0.915 ± 0.009	0.120 ± 0.009

Table 14. Mean and CI-95 for FT + ADV on PIROPO.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.466 ± 0.024	0.515 ± 0.025	0.762 ± 0.045	0.291 ± 0.039
2	0.516 ± 0.009	0.576 ± 0.018	0.827 ± 0.005	0.222 ± 0.006
5	0.517 ± 0.024	0.574 ± 0.031	0.829 ± 0.016	0.228 ± 0.016
10	0.544 ± 0.019	0.617 ± 0.034	0.856 ± 0.018	0.183 ± 0.023
20	0.574 ± 0.034	0.659 ± 0.047	0.885 ± 0.028	0.163 ± 0.027
50	0.596 ± 0.010	0.693 ± 0.021	0.917 ± 0.003	0.121 ± 0.005
100	0.613 ± 0.009	0.718 ± 0.020	0.927 ± 0.002	0.105 ± 0.011

Table 15. Mean and CI-95 for FT + ADV on necks 0 and 4 on PIROPO.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.472 ± 0.026	0.525 ± 0.028	0.772 ± 0.043	0.284 ± 0.045
2	0.516 ± 0.017	0.579 ± 0.027	0.826 ± 0.020	0.228 ± 0.030
5	0.523 ± 0.036	0.586 ± 0.050	0.844 ± 0.052	0.206 ± 0.048
10	0.542 ± 0.033	0.611 ± 0.052	0.862 ± 0.034	0.176 ± 0.037
20	0.588 ± 0.012	0.678 ± 0.015	0.897 ± 0.011	0.148 ± 0.012
50	0.601 ± 0.009	0.698 ± 0.024	0.915 ± 0.007	0.127 ± 0.016
100	0.600 ± 0.003	0.693 ± 0.012	0.917 ± 0.003	0.123 ± 0.003

Table 16. Mean and CI-95 for FT + ADV on necks 0, 1, and 4 on PIROPO.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.475 ± 0.008	0.516 ± 0.008	0.785 ± 0.002	0.270 ± 0.004
2	0.530 ± 0.022	0.596 ± 0.034	0.845 ± 0.009	0.198 ± 0.008
5	0.534 ± 0.024	0.603 ± 0.031	0.851 ± 0.027	0.193 ± 0.028
10	0.552 ± 0.039	0.628 ± 0.053	0.871 ± 0.031	0.165 ± 0.034
20	0.586 ± 0.003	0.677 ± 0.010	0.901 ± 0.005	0.139 ± 0.010
50	0.597 ± 0.007	0.691 ± 0.015	0.916 ± 0.009	0.117 ± 0.011
100	0.603 ± 0.010	0.703 ± 0.009	0.923 ± 0.001	0.114 ± 0.005

Table 17. Mean and CI-95 for FT + ADV + SPA on PIROPO.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.497 ± 0.024	0.536 ± 0.022	0.854 ± 0.037	0.214 ± 0.051
2	0.440 ± 0.051	0.459 ± 0.043	0.753 ± 0.083	0.341 ± 0.089
5	0.545 ± 0.022	0.595 ± 0.029	0.899 ± 0.023	0.158 ± 0.035
10	0.590 ± 0.008	0.650 ± 0.012	0.947 ± 0.004	0.089 ± 0.009
20	0.600 ± 0.001	0.677 ± 0.007	0.947 ± 0.007	0.089 ± 0.016
50	0.612 ± 0.009	0.698 ± 0.015	0.953 ± 0.005	0.075 ± 0.009
100	0.634 ± 0.002	0.730 ± 0.009	0.963 ± 0.004	0.056 ± 0.008

Table 18. Mean and CI-95 for FT + ADV + SPA on Mirror Worlds.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.424 ± 0.023	0.487 ± 0.028	0.685 ± 0.051	0.376 ± 0.054
2	0.424 ± 0.042	0.481 ± 0.047	0.671 ± 0.075	0.389 ± 0.062
5	0.467 ± 0.045	0.533 ± 0.045	0.731 ± 0.055	0.332 ± 0.077
10	0.507 ± 0.034	0.590 ± 0.052	0.762 ± 0.030	0.304 ± 0.033
20	0.492 ± 0.010	0.564 ± 0.018	0.738 ± 0.019	0.319 ± 0.020
50	0.512 ± 0.005	0.587 ± 0.008	0.770 ± 0.006	0.285 ± 0.007
100	0.516 ± 0.026	0.606 ± 0.027	0.758 ± 0.044	0.327 ± 0.054

Table 19. Mean and CI-95 for fine-tuning trained on Mirror Worlds and tested on PIROPO.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.425 ± 0.010	0.455 ± 0.011	0.695 ± 0.015	0.392 ± 0.013
2	0.492 ± 0.027	0.526 ± 0.027	0.813 ± 0.057	0.261 ± 0.070
5	0.479 ± 0.029	0.516 ± 0.044	0.785 ± 0.051	0.296 ± 0.052
10	0.487 ± 0.007	0.535 ± 0.016	0.795 ± 0.016	0.283 ± 0.018
20	0.495 ± 0.012	0.540 ± 0.008	0.811 ± 0.023	0.268 ± 0.026
50	0.519 ± 0.014	0.568 ± 0.022	0.851 ± 0.012	0.223 ± 0.016
100	0.532 ± 0.006	0.587 ± 0.014	0.861 ± 0.007	0.212 ± 0.008

Table 20. Mean and CI-95 for fine-tuning trained on PIROPO and tested on Mirror Worlds.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.402 ± 0.047	0.459 ± 0.056	0.650 ± 0.095	0.407 ± 0.098
2	0.398 ± 0.047	0.446 ± 0.062	0.642 ± 0.071	0.427 ± 0.070
5	0.476 ± 0.019	0.541 ± 0.024	0.744 ± 0.051	0.322 ± 0.058
10	0.501 ± 0.012	0.573 ± 0.002	0.765 ± 0.037	0.292 ± 0.041
20	0.514 ± 0.012	0.593 ± 0.018	0.774 ± 0.015	0.279 ± 0.007
50	0.547 ± 0.014	0.638 ± 0.013	0.818 ± 0.015	0.230 ± 0.020
100	0.502 ± 0.031	0.593 ± 0.035	0.732 ± 0.052	0.344 ± 0.084

Table 21. Mean and CI-95 for FT + ADV + SPA trained on Mirror Worlds and tested on PIROPO.

Images	mAP↑	AP@75↑	AP@50↑	LAMR↓
1	0.444 ± 0.003	0.471 ± 0.008	0.729 ± 0.002	0.361 ± 0.002
2	0.490 ± 0.042	0.528 ± 0.047	0.807 ± 0.071	0.268 ± 0.082
5	0.486 ± 0.018	0.519 ± 0.015	0.809 ± 0.053	0.273 ± 0.055
10	0.497 ± 0.022	0.539 ± 0.030	0.821 ± 0.048	0.253 ± 0.055
20	0.484 ± 0.035	0.505 ± 0.034	0.831 ± 0.077	0.256 ± 0.103
50	0.468 ± 0.035	0.501 ± 0.044	0.805 ± 0.036	0.288 ± 0.041
100	0.491 ± 0.018	0.520 ± 0.021	0.823 ± 0.011	0.270 ± 0.015

Table 22. Mean and CI-95 for FT + ADV + SPA trained on PIROPO and tested on Mirror Worlds.