ConfMix: Unsupervised Domain Adaptation for Object Detection via Confidence-based Mixing Supplementary Material

Giulio Mattolin[1], Luca Zanella[2], Elisa Ricci[1,2], Yiming Wang[2]
[1]University of Trento, Trento, Italy [2]Fondazione Bruno Kessler, Trento, Italy
lzanella@fbk.eu

In this Supplementary Material, we provide additional experiments to demonstrate the effect of pretrained detector backbone, i.e. with/without COCO-pretrained weights, on the adaptation performance using our proposed **ConfMix**. Moreover, we also justify key hyperparameter choices in the implementation, including the confidence threshold C_{th} to filter the detections on top of non-maximum suppression, and α that is used to scale the impact of the current iteration to the total number of iterations for the calculation of the shifting weight δ. Consistently with the main manuscript, we conduct the ablation study on the Sim10K $\rightarrow$ Cityscapes setup.

Does backbone initialisation negate the effect of domain adaptation? As there are works using ImageNet-pretrained networks [5, 2, 9, 10, 3, 12, 13, 4, 1] and works that do not specify whether they are pretrained or not [11, 8, 7, 6], we are motivated to examine how the initialisation of our backbone affects the proposed adaptation strategy. Therefore, we experiment with the initialisation of the backbone with random weights. Compared to the setting described in the main manuscript, we only vary the confidence threshold C_{th} used to filter the detections on top of the non-maximum suppression from 0.25 to 0.3 for the random weight setting, in order to account for less reliable predictions at the initial training phase. As shown in Table 1, with random weights initialisation, Source only, **ConfMix** and Oracle achieve lower performance than their corresponding ones with the COCO pretrained weights. Moreover, we notice that **ConfMix** with random weights obtains a mAP gain of +12.3% and +28.4% compared to its Source-only counterpart, on Sim10K$\rightarrow$Cityscapes and KITTI$\rightarrow$Cityscapes, respectively. While with COCO-pretrained weights, **ConfMix** achieves a mAP gain of +6.8% and +12.3% compared to its Source-only counterpart, on Sim10K$\rightarrow$Cityscapes and KITTI$\rightarrow$Cityscapes, respectively. This shows that the adaptation of **ConfMix** is more effective when the backbone is not pretrained, although its general detection performance on the target domain is bounded by the Oracle's performance.

				Sim10K$\rightarrow$ Cityscapes	KITTI$\rightarrow$ Cityscapes
Method	Detector	Backbone	Pretrained	mAP	mAP
Source only	YOLOv5	CSP-Darknet53	No	33.9	21.7
ConfMix (Ours)	YOLOv5	CSP-Darknet53	No	46.2	50.1
Oracle	YOLOv5	CSP-Darknet53	No	64.1	64.1
Source only	YOLOv5	CSP-Darknet53	COCO	49.5	39.9
ConfMix (Ours)	YOLOv5	CSP-Darknet53	COCO	56.3	52.2
Oracle	YOLOv5	CSP-Darknet53	COCO	70.3	70.3

Table 1. Quantitative results (mAP) for Sim10K/KITTI $\rightarrow$ Cityscapes benchmark.

C_{th}	0.1	0.25	0.5	0.7
mAP	42.1	47.7	**52.7**	41.8

Table 2. Target detection accuracy with different confidence thresholds C_{th}.

How does C_{th} for filtering out detections affect adaptation? We experiment with varying confidence threshold C_{th}, i.e. 0.1, 0.25 (our setting), 0.5, and 0.7, to filter the detections on top of non-maximum suppression. The weight used to balance the consistency loss $\gamma = 1$ is fixed throughout this experiment. As shown in Table 2, using $C_{th} = 0.5$ leads to the best adaptation performance, with a mAP improvement of +10.6% and +10.9% compared to $C_{th} = 0.1$ and $C_{th} = 0.7$, respectively. While the use of low values for C_{th} leads the model to keep erroneous pseudo detections, the use of high values filters out those pseudo detections that are useful for the model to learn the target features.

However, the use of C_{th} is closely related to C^{γ}_{th}, which is the confidence threshold used to calculate the reliability of pseudo detections γ, defined as the ratio of valid detections after non-maximum suppression with confidence greater than C^{γ}_{th}. We therefore vary C^{γ}_{th} from 0.4 to 0.9 under $C_{th} = 0.25$, and from 0.6 to 0.9 under $C_{th} = 0.5$ (note that the value of C^{γ}_{th} should be larger than C_{th} to function meaningfully). From the results reported in Table 3, we can see that the best performance is given by the combination of $C_{th} = 0.25$ and $C^{\gamma}_{th} = 0.5$, which is our experimental setting reported in the main manuscript.

How does α for the confidence transition affect adaptation? We analyse the effect of the hyperparameter α used to calculate the shifting weight δ for the smooth transition from C_{det} to C_{comb} during training. As illustrated in Fig-

C_{th}	0.25	0.25	0.25	0.25	0.25	0.25	0.5	0.5	0.5	0.5
C_{th}^{γ}	0.4	0.5	0.6	0.7	0.8	0.9	0.6	0.7	0.8	0.9
mAP	55.1	**56.3**	55.3	56.0	55.6	53.3	55.0	54.4	53.8	51.5

Table 3. Target detection accuracy with different confidence thresholds C_{th}^{γ} under $C_{th} = 0.25$ and $C_{th} = 0.5$.

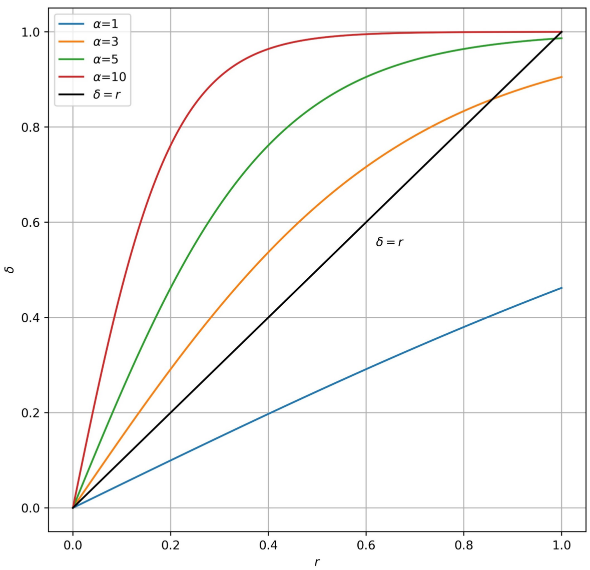

Figure 1. Visualisation of the evolution of δ throughout the training iterations with different values of α. Note that $\delta = r$ is the linear function already ablated in the main manuscript.

α	1	3	5	10
mAP	55.5	55.7	**56.3**	55.4

Table 4. Target detection accuracy with different confidence transition magnitudes.

ure 1, a lower value of α gives a greater importance to the less strict confidence C_{det}, while a higher value of α gives a greater importance to the stricter confidence C_{comb}. As shown in Table 4, using $\alpha = 5$ leads to the best mAP performance, with an improvement of +0.8% compared to $\alpha = 1$, +0.6% compared to $\alpha = 3$, and +0.9% compared to $\alpha = 10$.

References

[1] Cheng-Chun Hsu, Yi-Hsuan Tsai, Yen-Yu Lin, and Ming-Hsuan Yang. Every pixel matters: Center-aware feature alignment for domain adaptive object detector. In *Proceedings of European Conference on Computer Vision*, pages 733–748. Springer, 2020.

[2] Congcong Li, Dawei Du, Libo Zhang, Longyin Wen, Tiejian Luo, Yanjun Wu, and Pengfei Zhu. Spatial attention pyramid network for unsupervised domain adaptation. In *Proceedings of European Conference on Computer Vision*, pages 481–497. Springer, 2020.

[3] Wuyang Li, Xinyu Liu, Xiwen Yao, and Yixuan Yuan. Scan: Cross domain object detection with semantic conditioned adaptation. In *AAAI*, volume 6, page 7, 2022.

[4] Wuyang Li, Xinyu Liu, and Yixuan Yuan. Sigma: Semantic-complete graph matching for domain adaptive object detection. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5291–5300, 2022.

[5] Xianfeng Li, Weijie Chen, Di Xie, Shicai Yang, Peng Yuan, Shiliang Pu, and Yueting Zhuang. A free lunch for unsupervised domain adaptive object detection without source data. In *Proceedings of Conference on Artificial Intelligence*, volume 35, pages 8474–8481, 2021.

[6] Farzaneh Rezaeianaran, Rakshith Shetty, Rahaf Aljundi, Daniel Olmeda Reino, Shanshan Zhang, and Bernt Schiele. Seeking similarities over differences: Similarity-based domain alignment for adaptive object detection. In *Proceedings of IEEE/CVF International Conference on Computer Vision*, pages 9204–9213, 2021.

[7] Peng Su, Kun Wang, Xingyu Zeng, Shixiang Tang, Dapeng Chen, Di Qiu, and Xiaogang Wang. Adapting object detectors with conditional domain normalization. In *Proceedings of European Conference on Computer Vision*, pages 403–419. Springer, 2020.

[8] Vibashan Vs, Vikram Gupta, Poojan Oza, Vishwanath A Sindagi, and Vishal M Patel. Mega-cda: Memory guided attention for category-aware unsupervised domain adaptive object detection. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4516–4526, 2021.

[9] Vibashan VS, Poojan Oza, and Vishal M Patel. Instance relation graph guided source-free domain adaptive object detection. *arXiv preprint arXiv:2203.15793*, 2022.

[10] Minghao Xu, Hang Wang, Bingbing Ni, Qi Tian, and Wenjun Zhang. Cross-domain detection via graph-induced prototype alignment. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12355–12364, 2020.

[11] Fuxun Yu, Di Wang, Yinpeng Chen, Nikolaos Karianakis, Tong Shen, Pei Yu, Dimitrios Lymberopoulos, Sidi Lu, Weisong Shi, and Xiang Chen. Sc-uda: Style and content gaps aware unsupervised domain adaptation for object detection. In *Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1061–1070, 2022.

[12] Ganlong Zhao, Guanbin Li, Ruijia Xu, and Liang Lin. Collaborative training between region proposal localization and classification for domain adaptive object detection. In *Proceedings of European Conference on Computer Vision*, pages 86–102. Springer, 2020.

[13] Wenzhang Zhou, Dawei Du, Libo Zhang, Tiejian Luo, and Yanjun Wu. Multi-granularity alignment domain adaptation for object detection. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9581–9590, 2022.