Debiased Teacher for Day-to-Night Domain Adaptive Object Detection

Supplementary Material

A. The Quality Evaluation of Pseudo Labels

Due to the adverse nighttime conditions, the student model suffers from poor-quality pseudo labels generated by the teacher model. Models in the self-training framework reinforce incorrect predictions and fail to self-correct, thus causing confirmation bias. To correct the confirmation bias, we design a Pseudo Label Confirmation Calibrating module (ConCal) to generate high-quality pseudo labels. In Fig. S1, we conduct experiments on the TDND benchmark to validate the quality of generated pseudo labels by class accuracy and false positive ratio.

We observed that with the use of ConCal, the classification accuracy of pseudo labels improves by approximately 4% and continues to increase as training progresses. Meanwhile, the false positive rate of pseudo labels steadily decreases, showing a 4% reduction compared to the variant without ConCal. The improvement in classification accuracy demonstrates that the dynamic threshold based on class confidence in the Low-confidence Proposals Suppressing step effectively filters the proposals that are not correctly classified. Meanwhile, the reduction in the false negative rate indicates that this step successfully suppresses incorrect proposals, while the subsequent High-similarity Proposals Distilling step distills missed true positive proposals, further reducing the model's missed detection rate.

B. Visualization of mixed images

We do not directly mix source domain images with target domain images, as this is shown to be detrimental to cross-domain adaptation in Tab. 6. Due to the significant distribution gap between day and night, we leverage DNDT to model for the day-to-night distribution bias. This allowed us to mix night-like images, which closely resemble the target domain distribution, with target domain images. As illustrated in Fig. S3, we observe that night-like images can be naturally mixed into target domain images. This is because the camera perspectives are fixed, ensuring that the pasted objects generally align with real-world patterns (e.g., traffic lights are positioned in the sky while cars and buses remain on the ground). Meanwhile, to avoid unnecessary occlusion, the mix is not performed when the IoU between the mixed object and the original object is greater than 0.9 or occludes the original object.

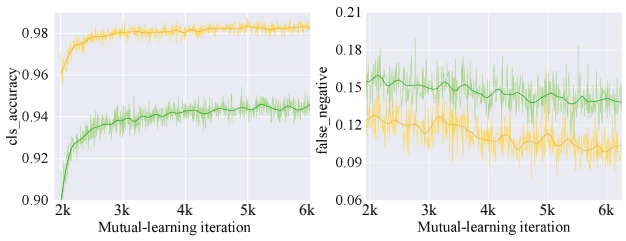

Figure S1. Pseudo label quality evaluation experiment. We employ two evaluation indicators, classification accuracy and false negative rate, to verify the quality improvement of false labels. Where yellow denotes w/ ConCal and green denotes w/o ConCal.

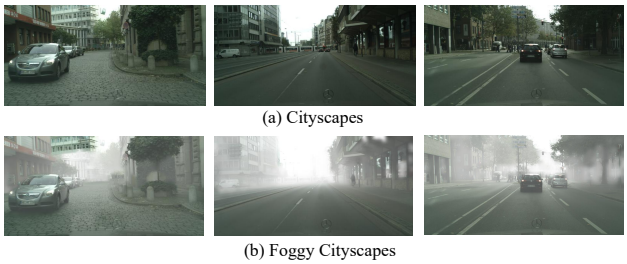

Figure S2. We present the Cityscapes dataset (a) and the Foggy Cityscapes dataset (b). Cityscapes→Foggy Cityscapes is a widely used benchmark in routine DAOD tasks, created to evaluate the detector's adaptability under adverse weather conditions.

C. Routine DAOD Benchmark

In autonomous driving scenarios, the object detector is not only influenced by day-night differences but also subjected to interference from other conditions. Adverse weather adaptation is one of the routine domain adaptive object detection tasks. Therefore, we further evaluate our DeT's domain adaptive performance in adverse weather adaptation. **Note that** our approach is mainly designed for day-to-night adaptation rather than routine cross-domain adaptation.

C.1. Dataset

Cityscapes As illustrated in Fig. S2 (a), Cityscapes [6] is collected by capturing images from street scenes in daytime conditions from 50 cities and annotated across 8 classes. It contains 2,975 images for training and 500 images for validation with dense pixel-level labels. We use 2,975 images as the source domain.

Foggy Cityscapes As illustrated in Fig. S2 (b), Foggy Cityscapes [23] is synthesized from the images in the Cityscapes. Therefore, it has the same train/test split as Cityscapes. It simulates the condition of foggy weather according to depth information provided in Cityscapes and generates three levels of foggy weather. We use the densest foggy (0.02) images as the target domain and 500

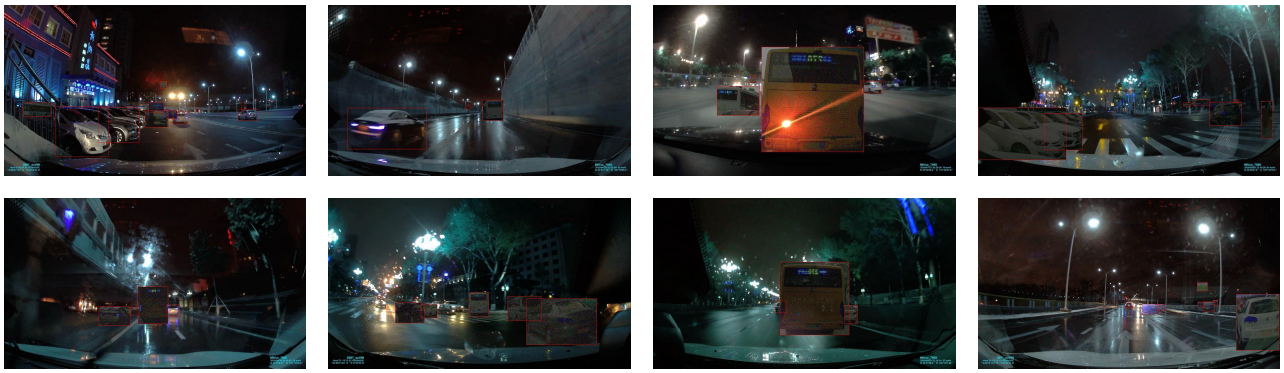

Figure S3. Visualization of the D2N Mixing results. We can observe that benefiting from the DNDT module's compensation for day-night differences, these objects from the source domain can be naturally mixed into the target domain (where red boxes denote objects).

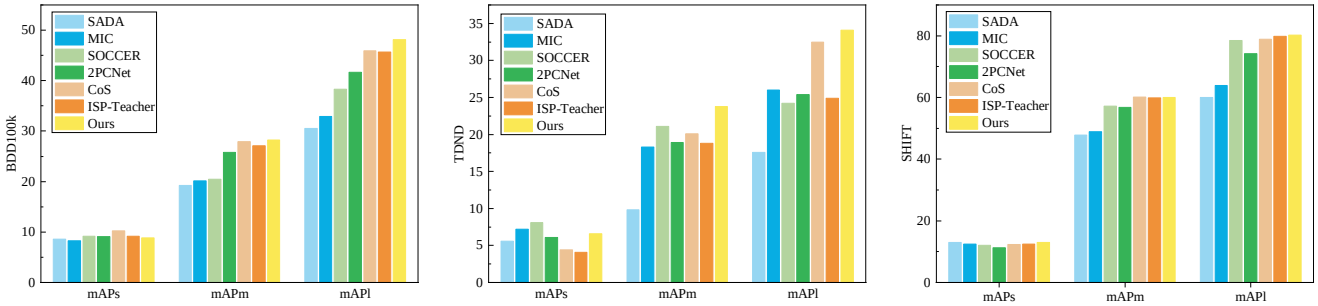

Figure S4. Quantitative results. We compare the detection results of SADA [5], MIC [13], SOCCER [7], 2PCNet [17], CoS [16], ISP-Teacher [28], and our DeT on three benchmark. Where mAPs, mAPm, mAPl denotes the mAP of small, medium, and large objects.

foggy images for validation.

C.2. Evaluation

As illustrated in Tab. S1, our DeT achieves 51.0% mAP, surpassing all comparison DAOD methods of 0.7% mAP compared to the best-performing Faster RCNN detector MIC [13], 0.6% mAP compared to the one-stage FCOS detector HT [8], and also 1.6% mAP compared to the Deformable DETR detector BiADT [11]. Compared with SOTA self-training methods, such as PT, MIC, and MTM, DeT addresses three biases in the self-training framework, which boosts the learning of target domain knowledge to the student model. We also observe that benefiting from Inverse Class Frequency Balancing submodule our DeT achieves the smallest standard deviation of 9.02% compared to 9.30% for MTM, 10.71% for BiADT, 9.85% for HT, and 10.00% for CMT.

D. Quantitative Results

In Fig. S4, we compare more detailed performance of our DeT with other methods: SADA [5], MIC [13], SOCCER [7], 2PCNet [17], CoS [16], ISP-Teacher [28] on three benchmarks. Where mAPs, mAPm, mAPl denotes the mAP of small, medium, and large objects. In the BDD100k benchmark, our DeT outperforms other methods by 3% on mAPl, demonstrating its advantage in detecting large objects. However, it lags behind CoS and ISP-Teacher in detecting small objects. In the TDND benchmark, our DeT outperforms other methods by a large gap on mAPs, mAPm, and mAPl. This proves that DeT can solve the day-night adaptation bias problem well in the adverse autonomous driving scenario. In the SHIFT benchmark, our DeT also maintains a fundamental advantage in all three metrics, indicating that DeT is also remarkably adaptable in simulated autonomous driving scenarios.

E. Pseudo-code for Debiased Teacher

In Algorithm 1, we present a pseudo-code pipeline of our Debiased Teacher.

F. Implementation Details

Due to the varying scales of different datasets, the implementation details of some parameters may differ. Therefore, we provide detailed hyper-parameters in Tab. S2.

Table S1. Results of **Cityscapes to Foggy Cityscapes** (0.02, dense fog). The average precision (AP, %) on all classes is presented.

Method	Venues	Detector	person	rider	car	truck	bus	train	mcycle	bicycle	mAP
Source [22]	NeurIPS'15	FRCNN	39.1	22.1	42.2	20.1	30.0	6.6	28.5	35.4	30.2
SADA [5]	IJCV'21	FRCNN	50.3	45.4	62.1	32.4	48.5	52.6	31.5	29.5	44.0
PT [4]	ICML'22	FRCNN	40.2	48.8	59.7	30.7	51.8	30.6	35.4	44.5	42.7
TDD [12]	CVPR'22	FRCNN	39.6	47.5	55.7	33.8	47.6	42.1	37.0	41.4	43.1
MGA [30]	CVPR'22	FRCNN	45.7	47.5	60.6	31.0	52.9	44.5	29.0	38.0	43.6
SAD [29]	T-PAMI'23	FRCNN	38.3	47.2	58.8	34.9	57.7	48.3	35.7	42.0	45.2
MIC [13]	CVPR'23	FRCNN	52.4	47.5	67.0	40.6	50.9	55.3	33.7	33.9	47.6
CMT [3]	CVPR'23	FRCNN	45.9	55.7	63.7	39.6	66.0	38.8	41.4	51.2	50.3
EPM [14]	ECCV'20	FCOS	44.2	46.6	58.5	24.8	45.2	29.1	28.6	34.6	39.0
SCAN [19]	AAAI'22	FCOS	41.7	43.9	57.3	28.7	48.6	48.7	31.0	37.3	42.1
SIGMA [20]	CVPR'22	FCOS	44.0	43.9	60.3	31.6	50.4	51.5	31.7	40.6	44.2
OADA [25]	ECCV'22	FCOS	47.8	46.5	62.9	32.1	48.5	50.9	34.3	39.8	45.4
CIGAR [21]	CVPR'23	FCOS	46.1	47.3	62.1	27.8	56.6	44.3	33.7	41.3	44.9
CSDA [9]	ICCV'23	FCOS	46.6	46.3	63.1	28.1	56.3	53.7	33.1	39.1	45.8
IGG [18]	ACM MM'23	FCOS	44.3	44.8	62.2	35.8	54.2	50.7	38.2	38.7	46.1
HT [8]	CVPR'23	FCOS	52.1	55.8	67.5	32.7	55.9	49.1	40.1	50.3	50.4
MTTrans [26]	ECCV'22	DefDETR	47.7	49.9	65.2	25.8	45.9	33.8	32.6	46.5	43.4
O^2net [10]	ACM MM'22	DefDETR	48.7	51.5	63.6	31.1	47.6	47.8	38.0	45.9	46.8
AQT [15]	IJCAI'22	DefDETR	49.3	52.3	64.4	27.7	53.7	46.5	36.0	46.4	47.1
DA-DETR [27]	CVPR'23	DefDETR	49.9	50.0	63.1	24.0	45.8	37.5	31.6	46.3	43.5
BiADT [11]	ICCV'23	DefDETR	50.7	56.3	67.1	28.8	53.7	49.5	38.8	50.1	49.4
MTM [24]	AAAI'24	DefDETR	51.0	53.4	67.2	37.2	54.4	41.6	38.4	47.7	48.9
DeT (ours)	-	FRCNN	47.8	50.3	66.2	37.9	61.3	55.8	41.4	47.0	51.0

Table S2. Detailed hyper-parameters of DeT for each benchmark.

Hyperparameter	Description	BDD100k	SHIFT	TDND
N_c	Number of shared cross-domain categories	9	6	6
b	Batch size	10	9	4
φ	EMA update ratio	0.9996	0.9996	0.9996
$\lambda_{D\to N}$	Weight for $D \to N$ Mixing Loss	1.0	1.0	1.0
$\lambda_{N\to D}$	Weight for $N \to D$ Mixing Loss	1.0	1.0	1.0
λ_{sup}	Weight for supervised Loss	1.0	1.0	1.0
λ_{mix}	Weight for mixing Loss	0.3	0.3	0.3
α	Balance factor of CDRC	4	4	4
β	Amplitude factor of ConCal	0.8	0.8	0.8
τ	IoU matching threshold of ConCal	0.6	0.6	0.6
δ_{base}	Base threshold of ConCal	0.8	0.8	0.8
δ_{lower}	Lower limit of δ	0.8	0.8	0.8
δ_{upper}	Upper limit of δ	0.95	0.95	0.95
δ_{min}	Minimum threshold for distill	0.5	0.5	0.5
lr	Learning rate	0.01	0.04	0.04
T_{burn_in}	Burn-in stage iterations	50000	20000	20000
$T_{max\_iterations}$	Total training iterations	100000	60000	60000

Algorithm 1 : The training pipeline of Debiased Teacher

Input: Object detectors: Student $\mathcal{S}(\cdot;\theta_s)$, Teacher $\mathcal{T}(\cdot;\theta_t)$ and θ_s/θ_t are the model parameters of Student and Teacher. Burn-in stage iterations: T_{burn_in}, Total training iterations: $T_{max\_iterations}$. Hyper-parameters: Momentum φ in EMA, adaptive threshold δ, and learning rate η.

Output: Student $\mathcal{S}(\cdot;\theta_s)$, Teacher $\mathcal{T}(\cdot;\theta_t)$ after the training of Debiased Teacher.

for $iteration \leftarrow 1$ **to** $T_{max\_iterations}$ **do**

// 1. Load data mini-batch

Sample source batch $B_s = \{(i_s^i, b_s^i, c_s^i)\}_{i=1}^{N_s} \in \mathcal{D}_s(I_s, B_s, C_s)$

Sample target batch $B_t = \{(i_t^i)\}_{i=1}^{N_t} \in \mathcal{D}_t(I_t)$

// 2. Generate night-like images from **Day-to-Night Domain Transforming** module by Eq.5∼9

$I_n = \textbf{DNDT}(I_s)$

// 3. Burn-in stage

if $iteration < T_{burn\_in}$ **then**

// 4. Compute supervised loss by Eq.1

$B'_d, C'_d = \mathcal{S}(I_n;\theta_s)$

$\mathcal{L}_{sup} = \mathcal{L}_{reg}(B'_d, B_s) + \mathcal{L}_{cls}(C'_d, C_s)$

else

// 5. Mutual-learning stage *i.e.*, **Cross-Domain Representation Compensating** stage

// 6. Update Teacher by EMA

$\theta_t \leftarrow \varphi\theta_t + (1-\varphi)\theta_s$

// 7. Obtain accurate pseudo labels by **Pseudo Label Confirmation Calibrating** module by Eq.12∼15

$\hat{Y}(\hat{B}, \hat{C}) = \textbf{ConCal}(\mathcal{S}, \mathcal{T}(I_t, I_n; \theta_t, \theta_t))$

// 8. Generate mixing images and labels from **Inverse Class Frequency Balancing** submodule by Eq.10

$I^M_{D\to N}, I^M_{N\to D}, Y^M_{D\to N}, Y^M_{N\to D} = \textbf{Bi-Mixing}(\textbf{ICFB}(I_n, I_t, Y_s, \hat{Y}))$

// 9. Compute mix training loss by Eq.11

$\mathcal{L}_{mix} = \mathcal{L}_{unsup}(I^M_{D\to N}, B^M_{D\to N}, C^M_{D\to N}) + \mathcal{L}_{unsup}(I^M_{N\to D}, B^M_{N\to D}, C^M_{N\to D})$

end if

// 10. The overall optimization objective

$\mathcal{L} = \lambda_{sup}\cdot\mathcal{L}_{sup} + \lambda_{mix}\cdot\mathcal{L}^{mix}_{consis}$

Take SGD step: $\theta_s = \theta_s - \eta \bigtriangledown_{\theta_s}\mathcal{L}$

end for

Table S3. Runtime and memory consumption during training and inference on an RTX 3090.

BDD100k Method	Training Throughput	Training GPU Memory	Inference Throughput	Inference GPU Memory
2PCNet [17]	1.39 it/s	10.08 GB	18.88 img/s	8.43 GB
CoS [16]	0.87 it/s	12.59 GB	17.05 img/s	9.58 GB
ISP-Teacher [28]	0.48 it/s	13.33 GB	17.22 img/s	10.16 GB
DeT|Ours	0.55 it/s	12.79 GB	16.01 img/s	12.64 GB

G. Runtime and Memory Analysis

To comprehensively evaluate the practical applicability of our proposed DeT framework, we conduct detailed runtime and memory consumption analysis. All runtime evaluations presented in Table S3 are conducted with identical hardware configurations and experimental settings to ensure fair comparison across different methods.

G.1. Training Efficiency Analysis

Our DeT framework demonstrates competitive training efficiency with a throughput of 0.55 it/s, which represents a 14.6% improvement over the recent state-of-the-art ISP-Teacher method (0.48 it/s). While 2PCNet achieves the highest training throughput at 1.39 it/s, this comes at the cost of significantly lower detection performance as shown in our main results. The training speed of DeT strikes an optimal balance between computational efficiency and model capability.

In terms of memory consumption during training, DeT requires 12.79 GB of GPU memory, positioning it between CoS (12.59 GB) and ISP-Teacher (13.33 GB). This moderate memory footprint is particularly noteworthy considering that DeT incorporates three distinct models.

G.2. Inference Efficiency Analysis

During inference, DeT achieves 16.01 images per second (img/s), which is competitive with existing methods. While this represents a slight decrease compared to ISP-Teacher

Figure S5. Visualization of daytime image after different generative transformation methods. The second line: CyEDA [1]. The third line: InstructPix2Pix [2], where the prompt is *"turn it into night time"*.

(17.22 img/s) and CoS (17.05 img/s), the difference is marginal (approximately 7%) and is well-justified by the substantial performance gains achieved by our method. The inference memory consumption of 12.64 GB is higher than baseline methods.

H. Discussion of the Day-Night Transformation Methods

H.1. Physics-based Modeling vs. Data-driven Generative Approaches

In light of the reviewers' questions, we replaced the DNDT module with readily available generative methods (based on diffusion models and CycleGAN), attempting to use generative models to mitigate the distribution bias between the source and target domains. As shown in Fig. S5, both the CycleGAN-based CyEDA [1] and the diffusion model-based InstructPix2Pix [2] can generate visually realistic nighttime images. However, when we replaced DNDT and conducted experiments, the results were not satisfactory, see Tab. S4. However, when replacing our DNDT with the above generative methods, we observe consistent performance degradation:

- InstructPix2Pix (Diffusion-based): 3.6% mAP decrease
- CyEDA (CycleGAN-based): 6.0% mAP decrease

While generative models such as diffusion models and CycleGAN can produce visually appealing nighttime effects, they often fail to capture the true statistical distribution of real nighttime data, which is critical for domain adaptation tasks. In contrast, our DNDT module, by incorporating physical priors such as illumination changes, noise characteristics, and flare effects, produces transformed images that closely align with the actual nighttime data distribution. This alignment is crucial for reducing distribution bias, which is one of the three core biases our method addresses.

Table S4. Compare more day-to-night (D2N) modules on **B**.

row	D2N Module	Type	**B** (mAP)
1	InstructPix2Pix|Brooks et al.	Diffusion	47.3 (-3.6)
2	CyEDA|Beh, J.C. et al.	CycleGAN	44.9 (-6.0)
3	DNDT|Ours	Physical Prior	**50.9**

H.2. Limitations and Future Directions

We acknowledge that our physics-based approach, while effective for the day-to-night domain adaptation task, may have limitations in scenarios where the physical modeling assumptions do not hold or where more complex visual transformations are required.

Future work could explore hybrid approaches that combine the distributional advantages of physics-based modeling with the expressive power of modern generative models. Such combinations could potentially leverage the strengths of both paradigms while mitigating their respective limitations.

References

[1] Jing Chong Beh, Kam Woh Ng, Jie Long Kew, Che-Tsung Lin, Chee Seng Chan, Shang-Hong Lai, and Christopher Zach. Cyeda: Cycle-object edge consistency domain adaptation. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 2986–2990. IEEE, 2022. 5

[2] Tim Brooks, Aleksander Holynski, and Alexei A Efros. Instructpix2pix: Learning to follow image editing instructions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18392–18402, 2023. 5

[3] Shengcao Cao, Dhiraj Joshi, Liang-Yan Gui, and Yu-Xiong Wang. Contrastive mean teacher for domain adaptive object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23839–23848, 2023. 3

[4] Meilin Chen, Weijie Chen, Shicai Yang, Jie Song, Xinchao Wang, Lei Zhang, Yunfeng Yan, Donglian Qi, Yueting Zhuang, Di Xie, et al. Learning domain adaptive object detection with probabilistic teacher. *arXiv preprint arXiv:2206.06293*, 2022. 3

[5] Yuhua Chen, Haoran Wang, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Scale-aware domain adaptive faster r-cnn. *International Journal of Computer Vision*, 129(7):2223–2243, 2021. 2, 3

[6] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 1

[7] Yiming Cui, Liang Li, Jiehua Zhang, Chenggang Yan, Hongkui Wang, Shuai Wang, Jin Heng, and Wu Li. Stochastic context consistency reasoning for domain adaptive object detection. In *ACM Multimedia 2024*, 2024. 2

[8] Jinhong Deng, Dongli Xu, Wen Li, and Lixin Duan. Harmonious teacher for cross-domain object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23829–23838, 2023. 2, 3

[9] Changlong Gao, Chengxu Liu, Yujie Dun, and Xueming Qian. Csda: Learning category-scale joint feature for domain adaptive object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11421–11430, 2023. 3

[10] Kaixiong Gong, Shuang Li, Shugang Li, Rui Zhang, Chi Harold Liu, and Qiang Chen. Improving transferability for domain adaptive detection transformers. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1543–1551, 2022. 3

[11] Liqiang He, Wei Wang, Albert Chen, Min Sun, Cheng-Hao Kuo, and Sinisa Todorovic. Bidirectional alignment for domain adaptive detection with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18775–18785, 2023. 2, 3

[12] Mengzhe He, Yali Wang, Jiaxi Wu, Yiru Wang, Hanqing Li, Bo Li, Weihao Gan, Wei Wu, and Yu Qiao. Cross domain object detection by target-perceived dual branch distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9570–9580, 2022. 3

[13] Lukas Hoyer, Dengxin Dai, Haoran Wang, and Luc Van Gool. Mic: Masked image consistency for context-enhanced domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11721–11732, 2023. 2, 3

[14] 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 *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, pages 733–748. Springer, 2020. 3

[15] Wei-Jie Huang, Yu-Lin Lu, Shih-Yao Lin, Yusheng Xie, and Yen-Yu Lin. Aqt: Adversarial query transformers for domain adaptive object detection. In *31st International Joint Conference on Artificial Intelligence, IJCAI 2022*, pages 972–979. International Joint Conferences on Artificial Intelligence, 2022. 3

[16] Yuan Jicheng, Le-Tuan Anh, Hauswirth Manfred, and Le-Phuoc Danh. Cooperative students: Navigating unsupervised domain adaptation in nighttime object detection. *2024 IEEE International Conference on Multimedia and Expo (ICME)*, 2024. 2, 4

[17] Mikhail Kennerley, Jian-Gang Wang, Bharadwaj Veeravalli, and Robby T Tan. 2pcnet: Two-phase consistency training for day-to-night unsupervised domain adaptive object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11484–11493, 2023. 2, 4

[18] Pengteng Li, Ying He, F. Richard Yu, Pinhao Song, Dongfu Yin, and Guang Zhou. Igg: Improved graph generation for domain adaptive object detection. In *Proceedings of the 31st ACM International Conference on Multimedia*, page 1314–1324, 2023. 3

[19] Wuyang Li, Xinyu Liu, Xiwen Yao, and Yixuan Yuan. Scan: Cross domain object detection with semantic conditioned adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1421–1428, 2022. 3

[20] Wuyang Li, Xinyu Liu, and Yixuan Yuan. Sigma: Semantic-complete graph matching for domain adaptive object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5291–5300, 2022. 3

[21] Yabo Liu, Jinghua Wang, Chao Huang, Yaowei Wang, and Yong Xu. Cigar: Cross-modality graph reasoning for domain adaptive object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23776–23786, 2023. 3

[22] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015. 3

[23] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126:973–992, 2018. 1

[24] Weixi Weng and Chun Yuan. Mean teacher detr with masked feature alignment: A robust domain adaptive detection trans-

former framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5912–5920, 2024. 3

[25] Jayeon Yoo, Inseop Chung, and Nojun Kwak. Unsupervised domain adaptation for one-stage object detector using offsets to bounding box. In *European Conference on Computer Vision*, pages 691–708. Springer, 2022. 3

[26] Jinze Yu, Jiaming Liu, Xiaobao Wei, Haoyi Zhou, Yohei Nakata, Denis Gudovskiy, Tomoyuki Okuno, Jianxin Li, Kurt Keutzer, and Shanghang Zhang. Mttrans: Cross-domain object detection with mean teacher transformer. In *European Conference on Computer Vision*, pages 629–645. Springer, 2022. 3

[27] Jingyi Zhang, Jiaxing Huang, Zhipeng Luo, Gongjie Zhang, Xiaoqin Zhang, and Shijian Lu. Da-detr: Domain adaptive detection transformer with information fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23787–23798, 2023. 3

[28] Yin Zhang, Yongqiang Zhang, Zian Zhang, Man Zhang, Rui Tian, and Mingli Ding. Isp-teacher: Image signal process with disentanglement regularization for unsupervised domain adaptive dark object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7387–7395, 2024. 2, 4

[29] Qianyu Zhou, Qiqi Gu, Jiangmiao Pang, Xuequan Lu, and Lizhuang Ma. Self-adversarial disentangling for specific domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 3

[30] Wenzhang Zhou, Dawei Du, Libo Zhang, Tiejian Luo, and Yanjun Wu. Multi-granularity alignment domain adaptation for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9581–9590, 2022. 3