Adversarial Defense Teacher for Cross-Domain Object Detection under Poor Visibility Conditions

Kaiwen Wang* Yinzhe Shen† and Martin Lauer‡

Abstract: Existing object detectors encounter challenges handling domain shifts between training and deployment, particularly under poor visibility conditions like fog and night. Cutting-edge cross-domain object detection methods use teacher-student frameworks and compel teacher and student models to produce consistent predictions under weak and strong augmentations. In this paper, we reveal that manually crafted augmentations are insufficient for optimal teaching and present a simple yet effective framework named ***A****dversarial* ***D****efense* ***T****eacher* (**ADT**), leveraging adversarial defense to enhance teaching quality. Specifically, we employ adversarial attacks, encouraging the model to generalize on subtly perturbed inputs that effectively deceive the model. To address small objects under poor visibility conditions, we propose a Zoom-in Zoom-out strategy, which zooms-in images for better pseudo-labels and zooms-out images and pseudo-labels to learn refined features. Our results demonstrate that ADT achieves superior performance, reaching 54.5% mAP on Foggy Cityscapes, surpassing the previous state-of-the-art by 2.6% mAP.

Keywords: Adversarial Defense, Object Detection, Robustness, Unsupervised Domain Adaptation

1 Introduction

A noticeable decline in performance occurs when object detection algorithms are exposed to poor visibility conditions, e.g., fog, night, etc. [21]. To overcome this performance drop, supervised methods require extensive annotations, which is expensive and impractical. Cross-domain Object Detection (CDOD) [22, 6, 4, 16, 2] has thus been proposed to address this issue, where a pre-trained object detector is adapted from a labeled source domain to an unlabeled target domain. As a semi-supervised learning scheme, CDOD eliminates the need for annotating training data in the target domain, making it more practical for real-world applications.

In recent years, the self-training paradigm [6, 4, 16, 2] has shown promising results in mitigating domain shift by using teacher-student mutual learning [23]. Specifically, the entire model consists of two architecturally identical components: a student and a teacher model. The student model is trained through standard gradient updating, while the teacher model is updated using the exponential moving average (EMA) of the

*Kaiwen Wang is PhD student at the Institute of Measurement and Control Systems (MRT), KIT. (e-mail: kaiwen.wang@kit.edu).

†Yinzhe Shen is PhD student at the Institute of Measurement and Control Systems (MRT), KIT.

‡Martin Lauer is Group leader at the Institute of Measurement and Control Systems (MRT), KIT.

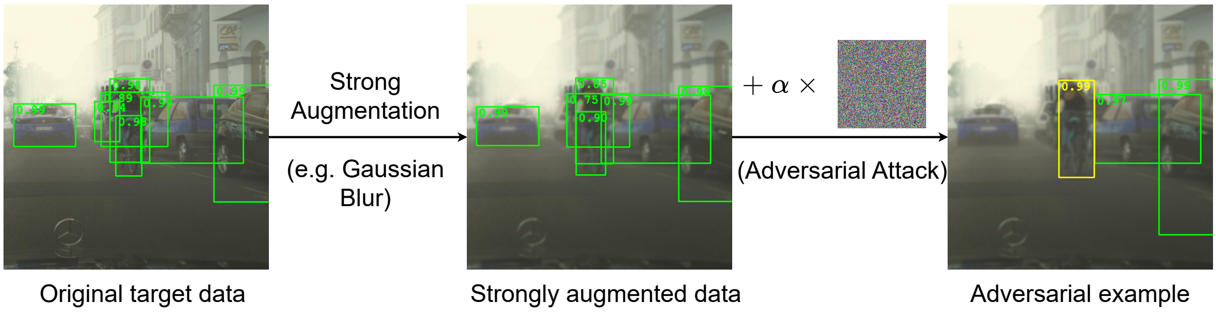

Figure 1: Current self-training methods apply manually crafted augmentations to help the student model generalize. However, these augmentations offer limited variability and suffer from overfitting. In contrast, our Adversarial Defense Teacher (ADT) framework introduces adaptive adversarial augmentations and generates tailored variations that challenge the network more effectively. Green boxes denote true positives, while yellow boxes indicate misclassifications. Best viewed in color.

weights from the student model. Additionally, the consistency loss between the pseudo-labels predicted by the teacher model on weakly augmented data and the predictions of the student model on strongly augmented data guides the adaptation mutually. Prior methods [16, 4, 7, 2] suggest employing manually designed augmentations, e.g., Gaussian blur and grayscaling. One major challenge, however, is the low teaching quality resulting from the model's overfitting to manually crafted data augmentations. As shown in Fig. 1, after training for several epochs, the predictions on blurred data closely resemble those on the original target data, reducing the effectiveness of teacher-student mutual learning.

To address this issue, we propose a simple yet effective framework called Adversarial Defense Teacher (ADT), which employs adversarial defense to enhance the quality of teacher-student mutual learning. As gradient-based augmentations, adversarial attacks are more adaptive than manually crafted augmentation techniques because they are dynamically updated to exploit the model's weaknesses, generating tailored variations that challenge the network more effectively. Defending against such attacks leads to more robust training and improved model performance. Moreover, while these adversarial attacks significantly impact model predictions, they remain imperceptible to humans, ensuring that the resulting adversarial examples still belong to the same domain.

While adversarial defense significantly mitigates the domain shift, detecting small and obscure objects remains challenging under adverse visibility conditions. To tackle this issue, we introduce a Zoom-in Zoom-out strategy. Target images are zoomed in before feeding into the teacher model so that smaller objects are upscaled and thus more likely to be included in the pseudo-labels. Subsequently, we perform a zoom-out operation on both the image and pseudo-labels with the same ratio. The student model is then forced to detect downscaled objects, ensuring benefits from extracting finer features.

2 Related Works

2.1 Object Detection under Poor Visibility Conditions

Object detection aims at classifying and localizing the objects given in an input image [19, 24, 8]. A common approach is to first propose regions of interest (ROIs) using region proposal networks (RPNs) and then refine these candidates in the second stage. In this work, cross-domain object detection is explored using Faster R-CNN [8] as the baseline architecture due to its wide range of applications.

Despite the great success, object detection under poor visibility conditions, like fog, rain and night, has been proven to be vulnerable [21]. Several approaches have been proposed to address this challenge. In [1], a multimodal sensor information technique is proposed to improve object detection under adverse weather. In [20, 26, 18], input images are preprocessed to enhance the visibility conditions, such as removing haze, rain streaks and raindrops. Despite these advancements, a substantial gap persists in achieving precise object detection under challenging visibility conditions.

2.2 Cross-Domain Object Detection

Recently, considerable work has employed domain adaptation to achieve better object detection under challenging visibility conditions. Such cross-domain object detection approaches can be mainly divided into three categories: feature alignment, domain translation and self-training. Feature alignment methods [5, 22, 27, 11, 28, 3, 25] conduct adversarial learning to align the features from both domains with a gradient reverse layer (GRL). Domain translation methods [15, 12] aim at translating the source data into target-like styles and thus improve the performance of CDOD. Recently, self-training methods [6, 4, 16, 2, 7] use weak-strong augmentation and Mean Teacher (MT)[23] for teacher-student mutual learning and have demonstrated superior advantages in this field. Adaptive Teacher (AT) [16] additionally applies adversarial learning to bridge the domain gap. Probabilistic Teacher (PT) [4] and Harmonious Teacher (HT) [7] focus on improving the quality of pseudo-labels for both classification and regression. Contrastive Mean Teacher (CMT) [2] leverages contrastive learning to optimize object-level features.

2.3 Adversarial Attack and Adversarial Defense

Different from adversarial learning [16], which focuses on learning domain-invariant features, adversarial defense aims to enhance neural networks' robustness against intentional yet imperceptible perturbations, i.e., adversarial attacks. Adversarial attacks are broadly categorized into two types: white-box and black-box attacks. In a white-box attack, the adversary has complete knowledge of the network under attack, including its architecture, parameters, and training data. Fast Gradient Sign Method (FGSM) is proposed in [9] to leverage the gradients of a neural network to design adversarial examples. Building upon this, Projected Gradient Descent [17] employs multi-step perturbations for a more powerful attack. In contrast, a black-box attacker only has limited or no knowledge about the system, making the attack more challenging due to the absence of internal details.

3 Preliminary

CDOD aims at mitigating the impact of domain shift between the source domain $\mathcal{D}_s = \{X_s, B_s, C_s\}$ and the target domain $\mathcal{D}_t = \{X_t\}$ on object detectors. Source images X_s are labeled with corresponding bounding box annotations B_s and class labels C_s, while target images X_t are not annotated. In the context of poor visibility conditions, the source domain denotes clear weather in the daytime, while the target domain encompasses various poor visibility conditions, such as fog and night.

State-of-the-art methods [4, 16, 2, 7] employ the MT framework [23] and weak-strong augmentation for CDOD. A source model is first pre-trained on labeled source data, serving as the initial model for two architecturally identical models: the teacher and the student model. The teacher model generates pseudo-labels with high confidence B'_t and C'_t on weakly augmented (e.g., random horizontal flip, etc.) target samples X_t^w, while the student model is trained on both the labeled source data $\{X_s, B_s, C_s\}$ and strongly augmented (e.g., Gaussian blur, grayscaling, etc.) target data $\{X_t^s, B'_t, C'_t\}$. The consistency loss between the pseudo-labels generated by the teacher model on weakly augmented target samples X_t^w and the predictions of the student model on strongly augmented target samples X_t^s improves the generalization capability of the student model through gradient back-propagation. Concurrently, the teacher model is updated through the EMA of the weights of the student model as in Eq. 1, performed without any gradient involvement.

$$\theta_{\text{teacher}} \leftarrow \beta\theta_{\text{teacher}} + (1-\beta)\theta_{\text{student}} \tag{1}$$

where θ_{teacher} and θ_{student} denote the weights of the teacher and student models respectively.

However, one drawback emerges. The randomly selected strong augmentation may not guide the student model in the most informative direction, i.e., the direction where the student model makes the most inconsistent predictions.

4 Adversarial Defense Teacher

In this section, we introduce the proposed Adversarial Defense Teacher (ADT) framework, which not only addresses the identified issue adaptively by applying adversarial defense for CDOD (Sec. 4.1) but also incorporates a reasonable inductive bias for poor visibility conditions (Sec. 4.2).

4.1 Adversarial Defense for CDOD

Given an input image, adversarial attack methods [9, 17] leverage the gradient of the loss concerning the input to generate a new image that maximizes the loss. This generated adversarial example can be expressed mathematically as in Eq. 2.

$$x_{\text{adv}} = x + \alpha \cdot \text{sgn}(\nabla_x \mathcal{L}(\theta, x, y)) \tag{2}$$

where the parameter α governs the magnitude of the adversarial perturbation, and $\text{sgn}(\cdot)$ refers to the sign function. Furthermore, $\mathcal{L}(\theta, x, y)$ indicates the loss between the prediction and the ground truth y.

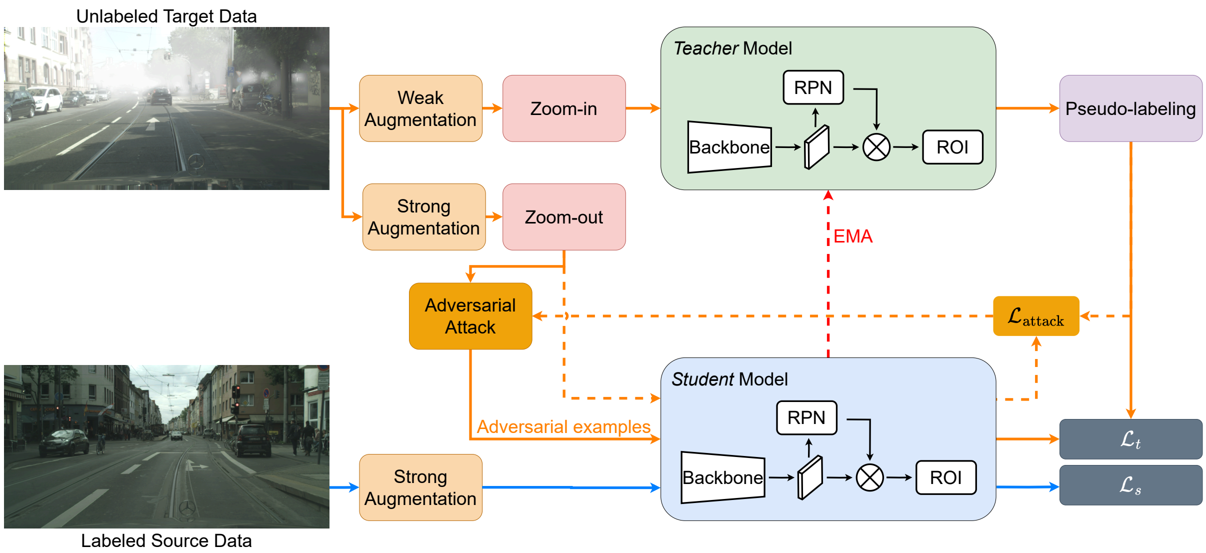

Figure 2: **Overview of the proposed Adversarial Defense Teacher.** Our model includes two branches: 1) supervised branch (blue lines): strongly augmented source data is fed into the student model. 2) unsupervised branch (orange lines): the teacher model processes weakly augmented and zoomed-in data to generate pseudo-labels with high confidence. Adversarial attacks (dashed lines) are conducted on the student model based on the inconsistency loss $\mathcal{L}_{\text{attack}}$ between pseudo-labels and predictions on strongly augmented and zoomed-out data. The resulting adversarial examples are reintroduced to the student model. Best viewed in color.

However, ground truth is not available for the unlabeled target data under the setting of CDOD. To achieve an adaptive data augmentation, we take the pseudo-labels generated by the teacher model as ground truth and calculate an attack loss $\mathcal{L}_{\text{attack}}$ for the student model. Based on adversarial examples generated to increase $\mathcal{L}_{\text{attack}}$, the student model is trained to defend itself from such attacks. With each student model being more robust, the teacher model is expected to benefit as well.

Consistent with common CDOD methods [6, 16], we also employ a confidence threshold to filter pseudo-labels. The pseudo-labels typically exhibit a higher precision than recall. This implies that positive objects identified in pseudo-labels are likely correct, yet there is a notable risk of overlooking a portion of actual positive instances, i.e., false negatives. Consequently, in the process of conducting adversarial attacks, our objective is to deceive the model in a way that disrupts its accurate detection of positive objects within pseudo-labels.

As shown in Fig. 2, we conduct adversarial attacks on the strongly augmented data, making it more challenging for the model to produce consistent predictions. The overall loss used for adversarial attacks is defined as:

$$\mathcal{L}_{\text{attack}} = \mathcal{L}^{\text{rpn}}_{\text{attack}}(X^s_t, B'_t, C'_t) + \mathcal{L}^{\text{roi}}_{\text{attack}}(X^s_t, B'_t, C'_t) \tag{3}$$

After applying PGD on the student model using $\mathcal{L}_{\text{attack}}$, adversarial examples are generated and then fed into the student model again. Due to the adversarial attack, the student model is deceived effectively to output highly inconsistent predictions as the pseudo-labels predicted by the teacher model, which in turn enhances the effectiveness of teacher-student mutual learning.

4.2 Zoom-in Zoom-out Strategy

After conducting adversarial defense, the model is robust enough to correctly detect objects despite various augmentations or perturbations. However, particularly under challenging visibility conditions, obscure objects of smaller sizes are still hard to capture. This challenge arises from the low recall of pseudo-labels, hindering the model's ability to effectively identify small objects. When encountering obscure objects, humans often exhibit a natural tendency to zoom in on an image for closer inspection. This zoom-in behavior facilitates the recognition of obscured details, allowing for improved comprehension of the object. Remarkably, even after zooming out again, humans can retain the ability to recognize previously obscure objects. Inspired by this, we propose a Zoom-in Zoom-out strategy, intending to increase the recall of pseudo-labels predicted by the teacher model and encourage the student model to extract detailed features.

During the zoom-in phase, the teacher model takes the zoomed-in image as input so that smaller objects can be better detected, thereby increasing the recall of the generated pseudo-labels. On the other hand, we zoom out the image and corresponding pseudo-labels before feeding them into the student model. As all objects are downscaled, the student model is enforced to extract features from smaller details. To prevent the size distribution from being disturbed by this strategy, zoomed-out pseudo-labels that are smaller than a specific threshold are removed.

5 Experiments

5.1 Fog: Cityscapes → Foggy Cityscapes

This experiment evaluates ADT on the commonly used benchmark Cityscapes → Foggy Cityscapes. Results are shown in Table 5.1. We present the CDOD training and evaluation results, comparing them with the performance of source (fully supervised on the source domain) and oracle models (fully supervised on the target domain). We consider both images with the highest fog severity ("0.02" split) and all images ("All" split) in Foggy Cityscapes. Similar to many other Mean Teacher-based approaches [16, 2, 7], our method surpasses the performance of the oracle models on both splits, which are directly trained on the labeled target domain. This indicates the effectiveness of the teacher-student mutual learning framework in transferring cross-domain knowledge by leveraging images from both domains.

Notably, ADT surpasses the previous best performance (from CMT) by 0.8% mAP on the "0.02" split and 2.6% mAP on the "All" split. The relatively larger gain on the "All" split underscores ADT's robustness in learning from a more diverse set of unlabeled data. This is particularly crucial in real-world applications where acquiring abundant unlabeled data is feasible, but labeling them is resource-intensive. ADT's capacity to consistently improve target-domain performance aligns well with the evolving landscape of growing unlabeled data, making it well-suited for such real-world scenarios.

Method	Reference	Split	person	rider	car	truck	bus	train	mcycle	bicycle	mAP
Source	-	0.02	25.9	29.4	35.4	6.9	19.8	4.3	16.1	22.7	20.1
Oracle	-	0.02	41.9	48.1	64.3	29.1	52.0	38.7	35.7	42.5	44.0
DA-Faster[5]	CVPR'18	0.02	25.0	31.0	40.5	22.1	35.3	20.2	20.0	27.1	27.6
SW[22]	CVPR'19	0.02	29.9	42.3	43.5	24.5	36.2	32.6	30.0	35.3	34.3
UMT[6]	CVPR'21	0.02	33.0	46.7	48.6	34.1	56.5	46.8	30.4	37.4	41.7
PT[4]	ICML'22	0.02	40.2	48.8	59.7	30.7	51.8	30.6	35.4	44.5	42.7
TDD[10]	CVPR'22	0.02	39.6	47.5	55.7	33.8	47.6	42.1	37.0	41.4	43.1
AT†[16]	CVPR'22	0.02	45.3	55.7	63.6	36.8	64.9	34.9	42.1	**51.3**	49.3
CMT[2]	CVPR'23	0.02	45.9	55.7	63.7	**39.6**	**66.0**	38.8	41.4	51.2	50.3
HT[7]	CVPR'23	0.02	**52.1**	55.8	67.5	32.7	55.9	49.1	40.1	50.3	50.4
ADT	Ours	0.02	49.4	**57.9**	**67.6**	35.8	55.4	**51.9**	**42.2**	48.6	**51.1**
Source	-	All	35.1	37.8	51.4	16.6	22.8	12.6	26.4	36.5	29.9
Oracle	-	All	46.8	51.6	68.7	33.6	56.1	45.7	42.1	48.9	49.2
SW‡[22]	CVPR'19	All	34.2	46.3	51.0	28.7	44.9	24.0	33.8	37.1	37.5
PDA[13]	WACV'20	All	36.0	45.5	54.4	24.3	44.1	25.8	29.1	35.9	36.9
ICR-CCR[25]	CVPR'20	All	32.9	43.8	49.2	27.2	36.4	36.4	30.3	34.6	37.4
PT[4]	ICML'22	All	43.2	52.4	63.4	33.4	56.6	37.8	41.3	48.7	47.1
AT[16]	CVPR'22	All	45.5	55.1	64.2	35.0	56.3	54.3	38.5	51.9	50.9
CMT[2]	CVPR'23	All	47.0	55.7	64.5	**39.4**	63.2	51.9	40.3	**53.1**	51.9
ADT	Ours	All	**51.1**	**58.4**	**71.3**	37.6	**63.5**	**56.5**	**46.7**	51.3	**54.5**

Table 1: Results of Cityscapes → Foggy Cityscapes.

5.2 Night: BDD100K Daytime → BDD100K Night

In this experiment, we evaluate ADT on the widely used BDD100K benchmark, focusing on the daytime-to-night domain shift scenario. Results are shown in Table 5.2.

Method	Reference	pedestrian	rider	car	truck	bus	mcycle	bicycle	traffic light	traffic sign	mAP
Source	-	50.0	28.9	66.6	47.8	47.5	32.8	39.5	41.0	56.5	41.1
Oracle	-	52.1	35.0	73.6	53.5	54.8	36.0	41.8	52.2	63.3	46.2
DA-Faster †[5]	CVPR'18	50.4	30.3	66.3	46.8	48.3	32.6	41.4	41.0	56.2	41.3
UMT [6]	CVPR'21	46.5	26.1	46.8	44.0	46.3	28.2	40.2	31.6	52.7	36.2
TDD [10]	CVPR'22	43.1	20.7	68.4	33.3	35.6	16.5	25.9	43.1	59.5	34.6
AT [16]	CVPR'22	42.3	30.4	60.8	48.9	52.1	34.5	42.7	29.1	43.9	38.5
2PCNet[14]	CVPR'23	54.4	30.8	73.1	**53.8**	55.2	37.5	44.5	**49.4**	65.2	46.4
ADT	Ours	**55.2**	**35.5**	**73.2**	53.3	**55.5**	**40.5**	**47.3**	46.8	**67.0**	**47.5**

Table 2: Results of BDD100K daytime → BDD100K night.

5.3 Ablation Studies

We further conduct ablation studies on important components in Table 5.3 and observe the performance gain from each component. The addition of Adversarial Defense demonstrates huge performance improvements compared to the Mean Teacher baseline on both

Methods		Cityscapes → Foggy Cityscapes			BDD100K daytime → BDD100K night		
AD	ZZ	mAP@[.5:.95]	mAP@0.5	mAP@0.75	mAP@[.5:.95]	mAP@0.5	mAP@0.75
✓	✓	**29.6**	**54.5**	**27.9**	**23.5**	**47.5**	**20.2**
✓		28.3	52.0	26.5	23.2	47.2	19.5
		26.4	49.3	24.5	22.7	46.0	19.5

Table 3: **Ablation studies.** The last row indicates the Mean Teacher baseline. AD represents Adversarial Defense and ZZ refers to Zoom-in Zoom-out strategy.

fog (2.7% AP) and night (1.2%) adaptation. Meanwhile, the Zoom-in Zoom-out strategy leads to performance improvements of 2.5% in foggy adaptation and 0.3% in night adaptation. The effectiveness of the Zoom-in Zoom-out strategy is observed to be slightly diminished in night adaptation due to the distinct domain shifts encountered in foggy weather and at night. In nighttime conditions, streetlights and vehicle indicators exhibit visual similarities to traffic lights but on a smaller scale, resulting in less performance improvement.

6 Conclusions

In this work, we tackle the challenge of Cross-Domain Object Detection under poor visibility conditions and propose a novel framework *Adversarial Defense Teacher*. We reveal that manually crafted augmentations only offer limited variations for mutual learning. The integration of adversarial defense into ADT strategically guides the student model to update itself in the most informative direction. Additionally, we present a Zoom-in Zoom-out strategy to address small object detection under adverse visibility conditions. This strategy involves zooming in on target images for pseudo-label generation by the teacher model and subsequently zooming out, along with pseudo-labels, for input into the student model. Our experiments validate the effectiveness of the proposed approach, particularly in the context of fog and night adaptations.

References

[1] Bijelic, M., Gruber, T., Mannan, F., Kraus, F., Ritter, W., Dietmayer, K., Heide, F.: Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11682–11692 (2020)

[2] Cao, S., Joshi, D., Gui, L.Y., Wang, Y.X.: Contrastive mean teacher for domain adaptive object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23839–23848 (2023)

[3] Chen, C., Zheng, Z., Ding, X., Huang, Y., Dou, Q.: Harmonizing transferability and discriminability for adapting object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8869–8878 (2020)

[4] Chen, M., Chen, W., Yang, S., Song, J., Wang, X., Zhang, L., Yan, Y., Qi, D., Zhuang, Y., Xie, D., et al.: Learning domain adaptive object detection with probabilistic teacher. In: International Conference on Machine Learning. pp. 3040–3055. PMLR (2022)

[5] Chen, Y., Li, W., Sakaridis, C., Dai, D., Van Gool, L.: Domain adaptive faster r-cnn for object detection in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3339–3348 (2018)

[6] Deng, J., Li, W., Chen, Y., Duan, L.: Unbiased mean teacher for cross-domain object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4091–4101 (2021)

[7] Deng, J., Xu, D., Li, W., Duan, L.: Harmonious teacher for cross-domain object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23829–23838 (2023)

[8] Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448 (2015)

[9] Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572 (2014)

[10] He, M., Wang, Y., Wu, J., Wang, Y., Li, H., Li, B., Gan, W., Wu, W., Qiao, Y.: Cross domain object detection by target-perceived dual branch distillation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9570–9580 (2022)

[11] He, Z., Zhang, L.: Multi-adversarial faster-rcnn for unrestricted object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6668–6677 (2019)

[12] Hsu, H.K., Yao, C.H., Tsai, Y.H., Hung, W.C., Tseng, H.Y., Singh, M., Yang, M.H.: Progressive domain adaptation for object detection. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 749–757 (2020)

[13] Hsu, H.K., Yao, C.H., Tsai, Y.H., Hung, W.C., Tseng, H.Y., Singh, M., Yang, M.H.: Progressive domain adaptation for object detection. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 749–757 (2020)

[14] Kennerley, M., Wang, J.G., Veeravalli, B., Tan, R.T.: 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. pp. 11484–11493 (2023)

[15] Kim, T., Jeong, M., Kim, S., Choi, S., Kim, C.: Diversify and match: A domain adaptive representation learning paradigm for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12456–12465 (2019)

[16] Li, Y.J., Dai, X., Ma, C.Y., Liu, Y.C., Chen, K., Wu, B., He, Z., Kitani, K., Vajda, P.: Cross-domain adaptive teacher for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7581–7590 (2022)

[17] Madry, A., Makelov, A., Schmidt, L., Tsipras, D., Vladu, A.: Towards deep learning models resistant to adversarial attacks. arXiv preprint arXiv:1706.06083 (2017)

[18] Qian, R., Tan, R.T., Yang, W., Su, J., Liu, J.: Attentive generative adversarial network for raindrop removal from a single image. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2482–2491 (2018)

[19] Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)

[20] Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. pp. 154–169. Springer (2016)

[21] Rothmeier, T., Wachtel, D., von Dem Bussche-Hünnefeld, T., Huber, W.: I had a bad day: Challenges of object detection in bad visibility conditions. In: 2023 IEEE Intelligent Vehicles Symposium (IV). pp. 1–6. IEEE (2023)

[22] Saito, K., Ushiku, Y., Harada, T., Saenko, K.: Strong-weak distribution alignment for adaptive object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6956–6965 (2019)

[23] Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems **30** (2017)

[24] Tian, Z., Shen, C., Chen, H., He, T.: Fcos: Fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9627–9636 (2019)

[25] Xu, C.D., Zhao, X.R., Jin, X., Wei, X.S.: Exploring categorical regularization for domain adaptive object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11724–11733 (2020)

[26] Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detection and removal from a single image. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1357–1366 (2017)

[27] Zhou, W., Du, D., Zhang, L., Luo, T., Wu, Y.: Multi-granularity alignment domain adaptation for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9581–9590 (2022)

[28] Zhu, X., Pang, J., Yang, C., Shi, J., Lin, D.: Adapting object detectors via selective cross-domain alignment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 687–696 (2019)