Supplementary Material

1 Ablation studies on other datastes

1.1 Results on Foggy Cityscapes

In Tab. 1, with careful observation of Foggy cityscapes, it becomes evident that every module introduced in both phases enhances the model's performance. Among these methods, the I2I method and IAoU loss exhibit substantial improvements, achieving 2.0% mAP and 2.8% mAP, respectively. In addition, the proposed PLF, compared to the original mean teacher model, demonstrates a growth of 3.4 % mAP. This further elucidates the efficacy of the various schemes we have proposed.

Table 1: The ablation results of Cityscapes→Foggy cityscapes. ✓:with, x:without.

GFA	I2I	IAoU	MT	PLF	mAP
x	x	x	x	x	46.5
✓	x	x	x	x	48.1
✓	✓	x	x	x	50.1
✓	✓	✓	x	x	52.9
✓	✓	✓	✓	x	55.0
✓	✓	✓	✓	✓	58.4

1.2 Results on Rain Cityscapes

The results on Rain Cityscapes are depicted in Tab. 2. We can see that, among these methods, the proposed I2I method achieves an enhancement of 3.5 % mAP and IAoU loss exhibits improvements of 2.9% mAP. What's more, the proposed PLF also demonstrates a growth of 3.4% mAP.

2 Results on BDD100K-night

We can see in Tab. 3, on BDD100K-night, Ours-YOLOv5L improves 4.1% mAP, 5.5% mAP, 5.9% mAP and 4.3% mAP over SSDA-YOLOv5L, R-YOLOv5L, Confmix, and CMT, respectively. Ours-YOLOv7 also exceeds R-YOLOv7 and SSDA-YOLOv7 by more than 4.2 % mAP and 2.9 % mAP, respectively. In addition, in Tab. 4, the IAoU loss exceeds the original one by 1.5 % mAP, highlighting the effectiveness of our proposed loss function. Moreover, we can observe that PLF outperforms the classical mean-teacher model by 2.2 % mAP, indicating the method is more suitable for cars.

Table 2: The ablation results of Cityscapes→Rain cityscapes. ✓:with, x:without.

GFA	I2I	IAoU	MT	PLF	mAP
x	x	x	x	x	43.8
✓	x	x	x	x	46.1
✓	✓	x	x	x	49.6
✓	✓	✓	x	x	52.5
✓	✓	✓	✓	x	54.1
✓	✓	✓	✓	✓	57.5

Table 3: Quantitative comparison results on the BDD100K-night.

Method	Detector	car	mAP
Baseline	YOLOv5	84.9	81.9
Baseline	YOLOv7	88.6	84.6
TDD [3]	FRCNN	76.2	79.2
CMT [1]	FRCNN	82.3	85.3
MIGADA [14]	FCOS	76.9	76.9
SIGMA++ [4]	FCOS	82.8	82.8
ConfMix [5]	YOLOv5L	83.7	83.7
R-YOLO [8]	YOLOv5L	84.1	84.1
SSDA-YOLO [13]	YOLOv5L	85.5	85.5
Ours	YOLOv5L	89.6	89.6(+7.7)
R-YOLO [8]	YOLOv7	87.1	87.1
SSDA-YOLO [13]	YOLOv7	88.4	88.4
Ours	YOLOv7	**91.3**	**91.3(+6.7)**
Oracle	YOLOv5L	91.1	91.1
Oracle	YOLOv7	93.2	93.2

3 Generation phase

3.1 Image to Image translation: Clear to adverse

We compare several GAN-based image-to-image translation approaches, including CUT and CycleGAN [6, 15] in the Fig. 1, and observe that the method employed in this paper effectively reduces noise in the clear-to-adverse process while incorporating the feature of the target domain.

3.2 Image to Image translation: Adverse to clear

Instead of using the rendering technique during the adverse-to-clear process, we employ the restoration-enabled work to generate intermediate images. The Fig. 2 below shows that the rendering technique introduces a lot of noise, but the restoration work reduces the noise while generating many source domain features.

3.3 Bounding box regression for feature alignment

The target domain is unlabeled, and the bounding boxes of the target domain need to be generated during the process of alignment. If the alignment of the

Table 4: The ablation results of BDD100K daytime→night. ✓:with, x:without.

GFA	I2I	IAoU	MT	PLF	mAP
x	x	x	x	x	81.9
✓	x	x	x	x	83.4
✓	✓	x	x	x	84.7
✓	✓	✓	x	x	86.2
✓	✓	✓	✓	x	87.4
✓	✓	✓	✓	✓	89.6

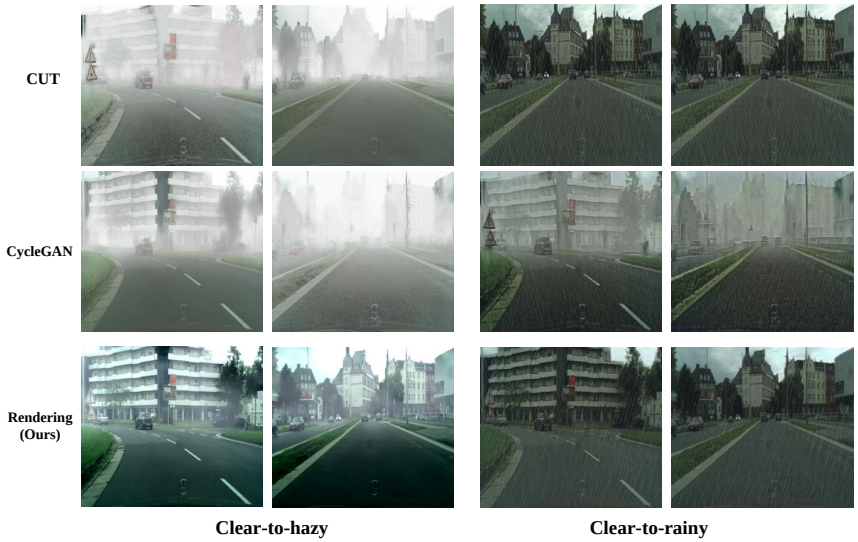

Fig. 1: Visualization of GAN-based approaches and rendering technology. The first row represents the CUT, the second row represents CycleGAN and our method is the last row. The first two columns display test results from clear-to-hazy, and the last two columns show results from clear-to-rainy.

source domain with the inaccurate regression bounding boxes of the target domain is done, it is very likely to result in negative transfer, which adversely affects the effectiveness of feature alignment. As shown in the Fig. 3, an inaccurate bounding box will result in the loss of important feature information i.e., brightly colored areas. On the other hand, an accurate bounding box will encompass more feature information.

3.4 IAoU loss: Sensitivity experiments on β values

The β values exhibit insensitivity across various datasets. To assess the sensitivity of the IAoU loss introduced in this paper, We set different values of the

Fig. 2: Visualization of rendering technology and restoration-enabled work. The first row represents the rendering technology, while the second row represents restoration-enabled work. The first two columns display test results on RTTS, the middle two columns show results on Rain Cityscapes, and the last two columns show results on BDD100K-night.

hyper-parameter β. The specific results are shown in Tab. 5. We can see that, the influence of β on model performance shows minimal variation, optimal performance is achieved in all four datasets when β is set to 0.4. This indicates that the effectiveness of β is independent of the dataset and does not require manual adjustment.

Table 5: Quantitative comparison results of the different β values. The evaluation metric is mAP.

β	Foggy	Rainy	RTTS	BDD100K
0.4	**52.9**	**52.5**	**51.8**	**86.2**
0.5	52.5	51.9	51.7	86.0
0.6	52.3	51.6	51.5	85.7
0.7	52.1	51.4	51.2	85.8
0.8	52.0	51.4	51.1	85.5

3.5 Effectiveness of the IAoU loss function

*Our proposed IAoU loss achieves optimal performance.*To assess the effectiveness of the IAoU loss introduced in this paper, we compare it with several common regression losses on RTTS. The specific results are shown in Tab. 6. We can see that, our proposed IAoU loss has an advantage of a 0.7% improvement in mAP over IoC. This demonstrates that our proposed loss function is more effective in facilitating feature alignment, thereby enhancing detection performance.

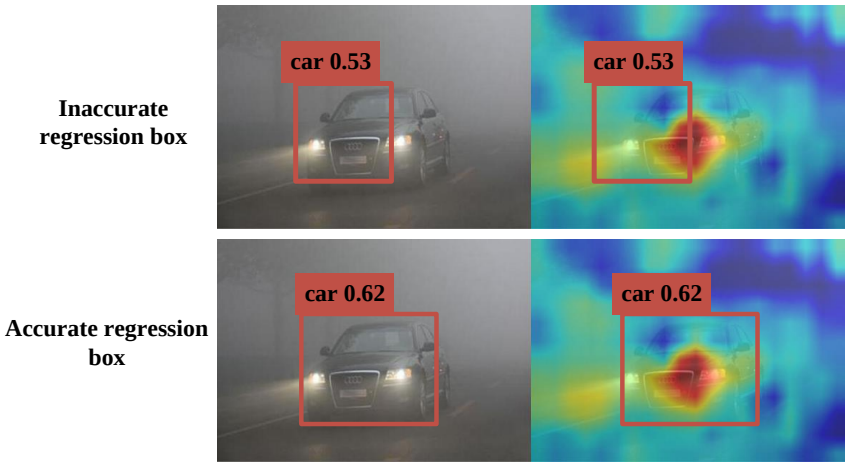

Fig. 3: The heat map of different regression boxes. The first row represents the inaccurate bounding box regression, while the second row indicates the accurate bounding box.

Table 6: The comparison of different regression loss.

loss	mAP
Ciou [12]	49.7
Eiou [11]	50.1
Siou [2]	50.4
MPDIoU [7]	50.6
IoC [10]	51.1
IAoU	51.8

3.6 Details of IoC loss and MPDIoU loss

The IoC [10] is formulated as below:

$$IoC = \frac{I - (E - U)}{E}, \tag{1}$$

where E is the minimum enclosing convex of the predicted box and the ground truth (GT), I and U is the intersection and union of the two boxes, respectively. The total loss can be the Equation 20:

$$L_{\text{IoC}} = 1 - IoC + \frac{d^2}{c^2} + \rho \times \sigma. \tag{2}$$

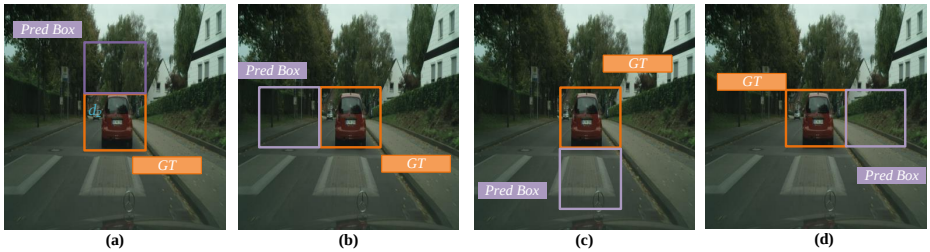

Fig. 4: The situation of prediction boxes and GT.

The penalty terms σ and ρ are as below:

$$\begin{aligned} \sigma &= \frac{(w^* - \hat{w})^2}{w^2} + \frac{\left(h^* - \hat{h}\right)^2}{h^2}, \\ \rho &= \frac{\sigma}{(1 - IoC) + \sigma}, \end{aligned} \tag{3}$$

Where $\hat{x}$ and x^* represent the prediction bounding box and the ground truth, and w and h correspond to the width and height of the minimum enclosing convex. Compared to our proposed IAoU loss, the main term of IoC loss lacks balancing coefficients and degrades to IOU when the two boxes are merely touching, as shown in Fig. 4. In addition, its numerator is the minimum enclosing convex of the two boxes, which is not friendly to small targets compared to the union and slows down the convergence speed. Moreover, the penalty term for IoC loss σ will fail to converge when the two boxes are in a surrounding situation. The equation for MPDIoU [7] is shown below:

$$MPDIoU = IoU - \frac{d_1^2}{h^2 + w^2} - \frac{d_2^2}{h^2 + w^2}, \tag{4}$$

$$\begin{aligned} \mathcal{L}_{MPDIoU} &= 1 - MPDIoU, \\ d_1^2 &= \left(x_1^* - \hat{x_1}\right)^2 + \left(\mathrm{y}_1^* - \hat{y_1}\right)^2, \\ d_2^2 &= \left(x_2^* - \hat{x_2}\right)^2 + \left(\mathrm{y}_2^* - \hat{y_2}\right)^2, \end{aligned} \tag{5}$$

where (x_1, y_1) and (x_2, y_2) denote the coordinates of their respective top-right and bottom-left corners, and w and h correspond to the width and height of this image. MPDIoU relies solely on the penalty term to generate the gradient when the two boxes do not intersect, resulting in slow convergence. Additionally, the penalty term denominator in the equation is based on the width and height of the images, which is not sensitive to small target regression and can also slow down the model's convergence.

4 Composition phase

4.1 PLF: Sensitivity experiments on θ_1 , θ_2, θ_3, and θ_3^1 values

The θ_2 value impacts the model performance, but the optimal mAP is achieved at 0.45, which does not need to be manually adjusted due to changes in the dataset. We set the low threshold θ_1 at 0.1 and the high thresholds θ_2 at 0.45, 0.55, 0.65, 0.75, and 0.85 for sensitivity experiments, respectively. As shown in Tab. 7, the four datasets are not very sensitive to the value of θ_2, and all of them achieve the optimal value when it is 0.45. The way we manually set it is more relevant than the dynamic threshold selecting methods in the literature [9]. Further illustrating the effectiveness of the proposed high and low thresholds. The Tab. 8 and Tab. 9 demonstrate that the results are not significantly impacted by the values of θ_3 and θ_3^1, the highest mAP is attained when θ_3 is 0.5 and θ_3^1 is 0.2.

Table 7: Quantitative comparison results of the θ_1 and θ_2 values. The evaluation metric is mAP. Dynamic indicates the dynamic threshold strategy in [9].

$\theta_1 = 0.1$	θ_2	Foggy	Rainy	RTTS	BDD100K
	0.45	**58.4**	**57.5**	**58.9**	**89.6**
	0.55	58.1	57.2	58.5	89.2
	0.65	57.9	56.7	58.3	89.1
	0.75	57.5	56.5	58.2	88.9
	0.85	57.1	56.2	57.9	88.5
	Dynamic	57.6	56.9	58.1	89.1

Table 8: Quantitative comparison results of the $\theta_3$values. The evaluation metric is mAP.

$\theta_3^1 = 0.2$	θ_3	Foggy	Rainy	RTTS
	0.40	57.7	56.6	57.8
	0.50	**58.4**	**57.5**	**58.9**
	0.60	57.4	57.4	56.6

4.2 PLF: When the student model outperforms the teacher model

In this paper, we consider the prediction from both the teacher and student model, we only select predictions where the teacher model outperforms the student model as candidate pseudo-labels. When the student model exceeds the teacher model, we simply discard it instead of adopting the student model's predictions as valid pseudo-labels. We can see in Tab. 10, that our algorithm will

Table 9: Quantitative comparison results of the θ_3^1values. The evaluation metric is mAP.

$\theta_3 = 0.5$	θ_3^1	Foggy	Rainy	RTTS
	0.10	57.7	56.2	58.0
	0.20	**58.4**	**57.5**	**58.9**
	0.30	58.1	56.6	57.3

enhance the model's performance by reducing some misdirection compared to solely considering the teacher model. And the teacher-student decision is slightly inferior to our methods. We posit that this phenomenon could stem from the ease with which misinformation can be conveyed once the student model attains an advanced level, potentially leading to self-misguidance.

Table 10: Different prediction selection from the teacher-student model. The evaluation metric is mAP. Only the teacher indicates to consider the prediction from the teacher model, teacher-student awareness (ours) represents our strategy, and the teacher-student decision means selecting the optimal output from the teacher-student model.

Prediction Selection	Foggy	Rainy	RTTS	BDD100K
Only teacher	57.5	56.2	57.6	88.1
teacher-student aware(ours)	**58.4**	**57.5**	**58.9**	**89.6**
teacher-student decision	57.7	57.1	58.1	88.9

4.3 Visualization of pseudo labels

In Fig. 5, the baseline pseudo-labels encounter difficulties with noise and small targets. In phase one, our restoration-enabled method reduces noise in the translation to narrow the image-level domain shift. Feature alignment based on the proposed IAoU loss reduces the domain gap from the instance level. In phase two, we use image restoration and super-resolution as data augmentation to improve the texture information of noise targets and details of small targets in the target domain. The two phases above significantly reduce the **miss detection** of noise and small targets in the candidate pseudo-labels. Alternatively, our filtering strategy incorporates regression thresholds to eliminate **false detection** and leverages the student-aware method to prevent **misguidance from the teacher**.

4.4 Analysis of Complexity

As shown in Tab. 11, our model shows a marginal increase in computation and training time compared to the baseline. This demonstrates that our model

Table 11: Comparison of model complexity.

Phase	Model	Gflops	Training Time/h	Fps
Generation	Baseline	108.3	0.915	95.23
	Ours	111.4	1.674	93.45
Phase	Model	Gflops	Training Time/h	Fps
Composition	Baseline	110.9	1.980	84.03
	Ours	113.4	2.639	76.92

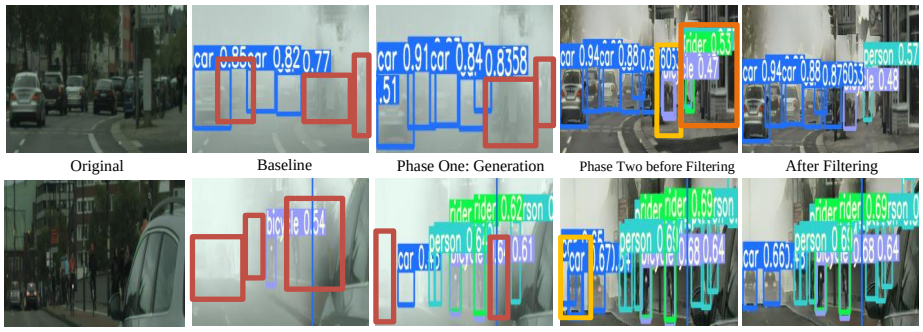

Fig. 5: Visualization of pseudo labels during different phases. Red boxes: missed targets, yellow: inaccurate regression boxes, orange: incorrect detection from the teacher model.

achieves significant performance improvements with very limited extra complexity.

5 Visualization of detection results

As depicted in Fig. 6 and Fig. 7, we compare our model with YOLOv5L, SSDA-YOLOv5L, and R-YOLOv5L by visualizing the detection results. We can see that the other two methods have relatively significant missing ones and some false detections, but our model can detect small targets at long distances without any false predictions, significantly enhancing the model's detection performance in adverse weather conditions.

Fig. 6: Visualization of detection results. The first row represents the original YOLOv5L model, while the second row represents Ours-YOLOv5L. The first two columns display test results on RTTS, the middle two columns show results on Rain Cityscapes, and the last two columns show results on Foggy Cityscapes, the blue boxes indicate missing targets, which demonstrates that our model greatly improves detection accuracy.

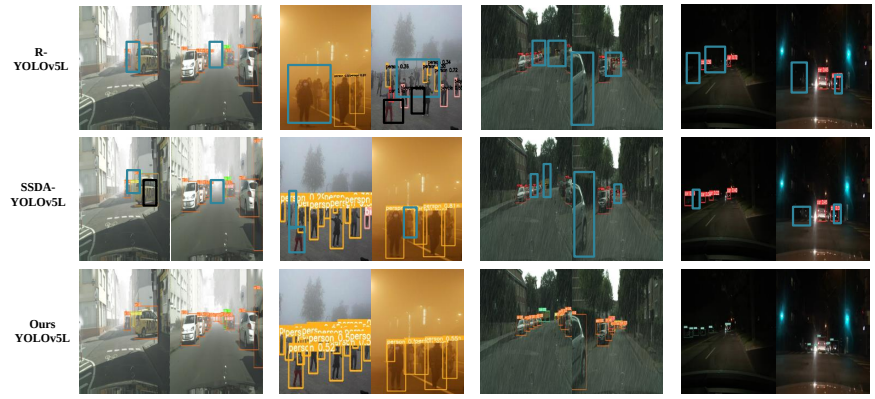

Fig. 7: Visualization of detection results by different models. The first row represents the SSDA-YOLOv5L model, while the second row represents the R-YOLOv5L model and our model is the last row. The first two columns display test results on Foggy cityscapes, the second two columns display test results on RTTS, the third two columns show results on Rain Cityscapes, and the last two columns show results on BDD100K-night. The blue boxes indicate missing targets and the black boxes are wrong detection, which demonstrates that our model greatly improves detection accuracy.

References

1. Cao, S., Joshi, D., Gui, L., Wang, Y.: Contrastive mean teacher for domain adaptive object detectors. In: CVPR. pp. 23839–23848 (2023)

2. Gevorgyan, Z.: Siou loss: More powerful learning for bounding box regression. arXiv preprint arXiv:2205.12740 (2022)
3. 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: CVPR. pp. 9560–9570 (2022)
4. Li, W., Liu, X., Yuan, Y.: SIGMA++: improved semantic-complete graph matching for domain adaptive object detection. TPAMI **45**(7), 9022–9040 (2023)
5. Mattolin, G., Zanella, L., Ricci, E., Wang, Y.: Confmix: Unsupervised domain adaptation for object detection via confidence-based mixing. In: WACV. pp. 423–433 (2023)
6. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: ECCV. pp. 319–345 (2020)
7. Silang, M., Yong, X.: Mpdiou: A loss for efficient and accurate bounding box regression. arXiv preprint arXiv:2307.07662 (2023)
8. Wang, L., Qin, H., Zhou, X., Lu, X., Zhang, F.: R-yolo: A robust object detector in adverse weather. IEEE Transactions on Instrumentation and Measurement **72**, 1–11 (2022)
9. Xu, B., Chen, M., Guan, W., Hu, L.: Efficient teacher: Semi-supervised object detection for yolov5. arXiv preprint arXiv **abs/2302.07577** (2023)
10. Zhang, Y., Shi, Z., Zhang, Y.: Adioc loss: An auxiliary descent ioc loss function. Engineering Applications of Artificial Intelligence **116**, 105453 (2022)
11. Zhang, Y., Ren, W., Zhang, Z., Jia, Z., Wang, L., Tan, T.: Focal and efficient IOU loss for accurate bounding box regression. Neurocomputing **506**, 146–157 (2022)
12. Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D.: Distance-iou loss: Faster and better learning for bounding box regression. In: AAAI. vol. 34, pp. 12993–13000 (2020)
13. Zhou, H., Jiang, F., Lu, H.: SSDA-YOLO: semi-supervised domain adaptive YOLO for cross-domain object detection. Comput. Vis. Image Underst. **229**, 103649 (2023)
14. Zhou, W., Du, D., Zhang, L., Luo, T., Wu, Y.: Multi-granularity alignment domain adaptation for object detection. In: CVPR. pp. 9571–9580 (2022)
15. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV. pp. 2223–2232 (2017)