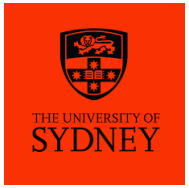
电子科技大学
University of Electronic Science and Technology of China



JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Harmonious Teacher for Cross-domain Object Detection

Jinhong Deng[1] Dongli Xu[2] Wen Li[3*] Lixin Duan[3,4]
[1]University of Electronic Science and Technology of China [2]University of Sydney
[3]Shenzhen Institute for Advanced Study, UESTC [4]Sichuan Provincial People's Hospital, UESTC
{jhdengvision, dongliixu, liwenbnu, lxduan}@gmail.com

THU-PM-308

Cross-domain Object Detection

Concepts

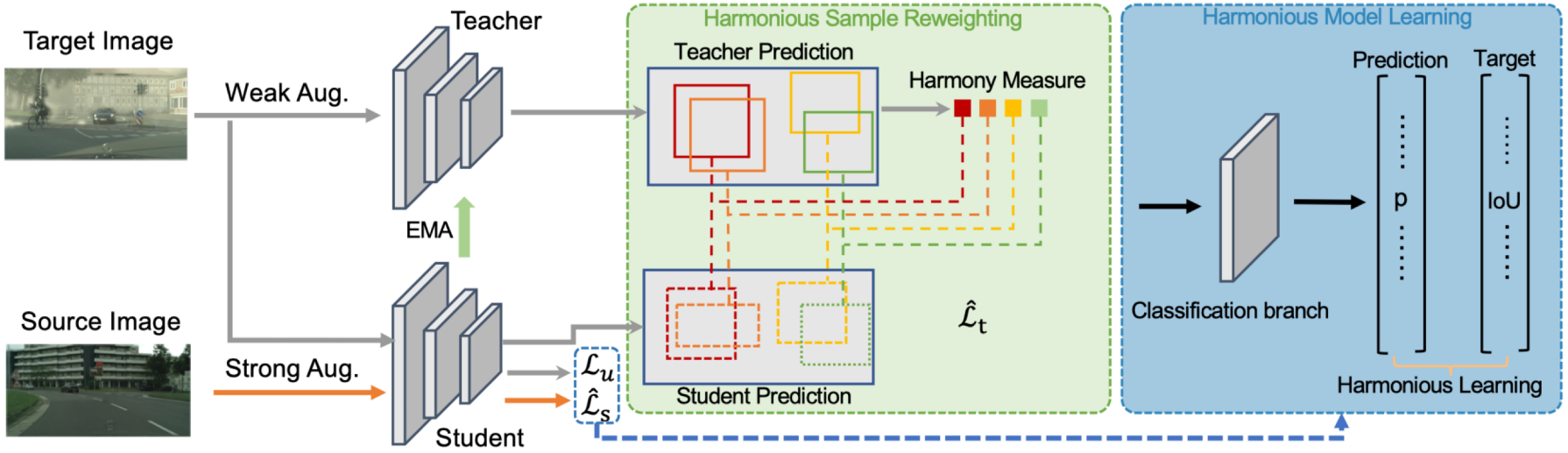
Cross-domain Object Detection (CDOD) aims to adapt a detector from label-rich source to label-scarce target domains.

Harmonious Teacher

Motivation

- ❑ Challenges of Self-training in CDOD
  - Adopt the classification score to select pseudo labels.
  - Hard thresholding for selecting confident pseudo-labeled instances.
- ❑ Harmonious Teacher
  - Harmonious Model Learning
    - Regularizing the consistency of the classification prediction and the localization score when training the detection model.
  - Harmonious Sample Reweighting
    - All pseudo-labeled samples can contribute to the model training based on prediction qualities
    - The hard threshold is not needed anymore.

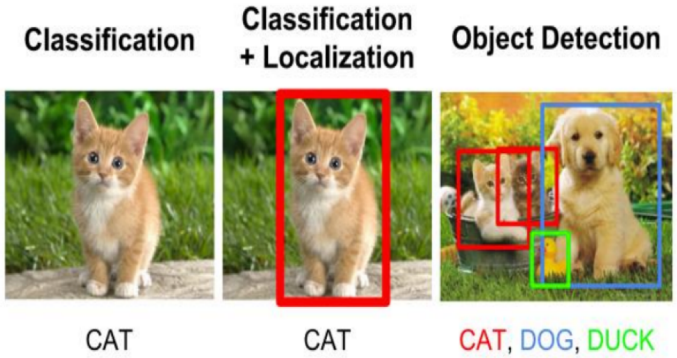
Harmonious Teacher



- Harmonious Model Learning: Regularizing the consistency of the classification prediction and the localization score when training the detection model.
- Harmonious Sample Reweighting
  - All pseudo-labeled samples can contribute to the model training based on prediction qualities
  - The hard threshold is not needed anymore.

Object Detection

Object detection aims to **recognize** and **localize** objects in images simultaneously.



Formulation

Input-target pairs $(\boldsymbol{x}, \boldsymbol{b}, \boldsymbol{c})$

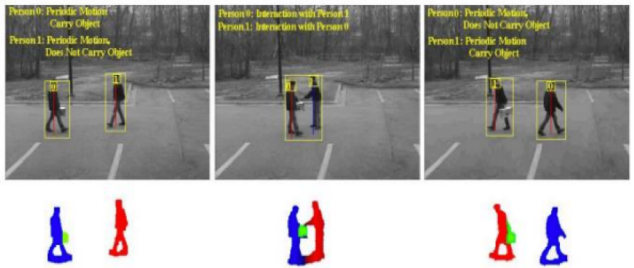
Image $x \in \mathbb{R}^{3\times H\times W}$

Bounding Boxes $\boldsymbol{b}^i = (c_x^i, c_y^i, w^i, h^i)$

Object Detection

Applications

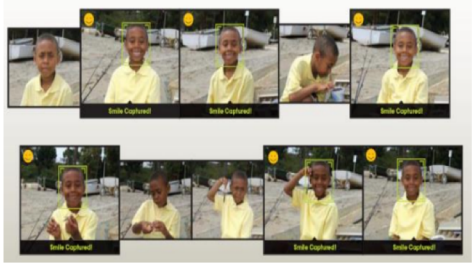
- Autonomous driving
- Image retrieval
- Video understanding
- Robotics
- Security etc.



Activity Recognition

Optical Character Recognition

Driving Vehicles



Smile Detection

Robotics

Image Search

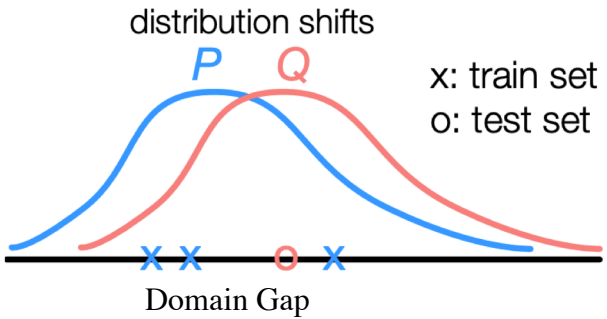
Tracking People

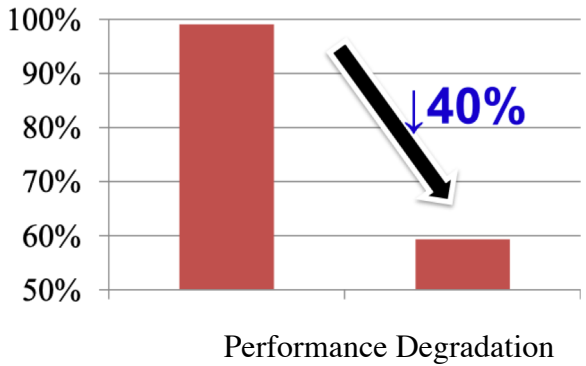
Image credit: Applications of Object Detection System

Object Detection

Challenge

- **Diverse objects** (small-scale, occlusion, pose, unknown)
- **Versatile environments** (background variation, weather, illumination, view-point)
- **Data Uncertainty** (Lack of annotation, weak annotation, distribution shift)





Deploying deep model in a novel domain leads to performance degradation due to domain gap.

Cross-domain Object Detection

Concepts

Cross-domain Object Detection (CDOD) aims to adapt a detector from label-rich source to label-scarce target domains.

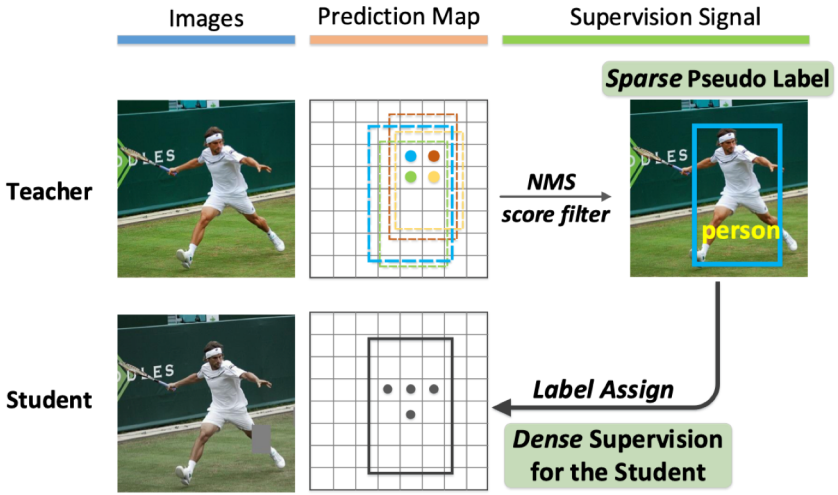
Cross-domain Object Detection

Related Works

- Domain Alignment
  - Adversarial Training (AT)
  - Foreground-aware AT
  - Conditional AT
  - Prototype Alignment
  - Disentanglement
  - Style Transfer
- Self-training
  - Mean Teacher
  - Uncertainty Estimation

Self-training

Pseudo Labeling in Object Detection



1. Predictions of teacher network.
2. Non-Maximum Suppression
3. Score-based thresholding
4. Label Assignment

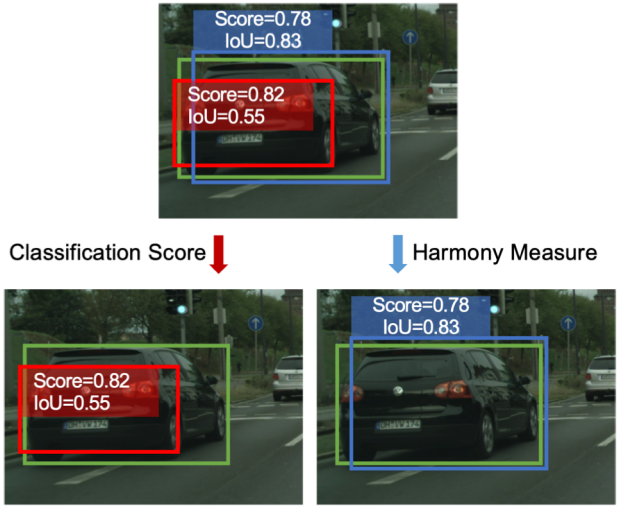
Harmonious Teacher

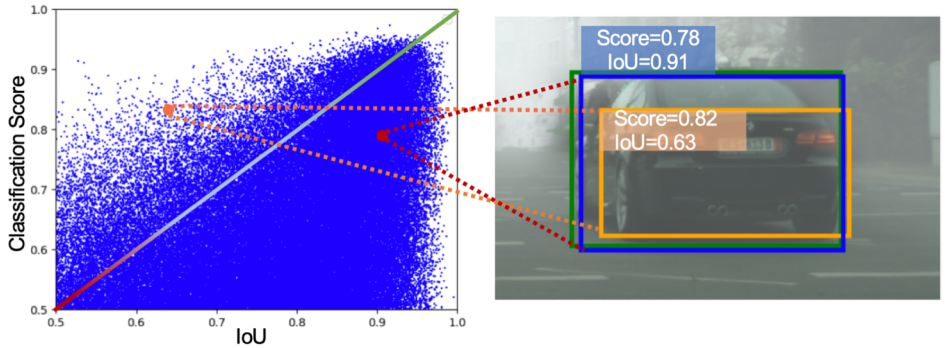
Motivation; Method; Experiments

Harmonious Teacher

Motivation

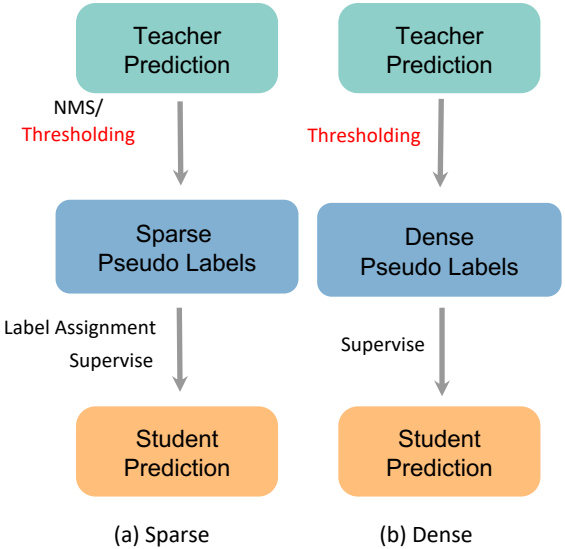
- Challenges
  - Adopt the classification score to select pseudo labels.
    - Inconsistency between classification and localization scores.





Motivation

- Challenges
  - Hard thresholding for selecting confident pseudo-labeled instances.
    - Ignore the valuable hard examples to the model training.

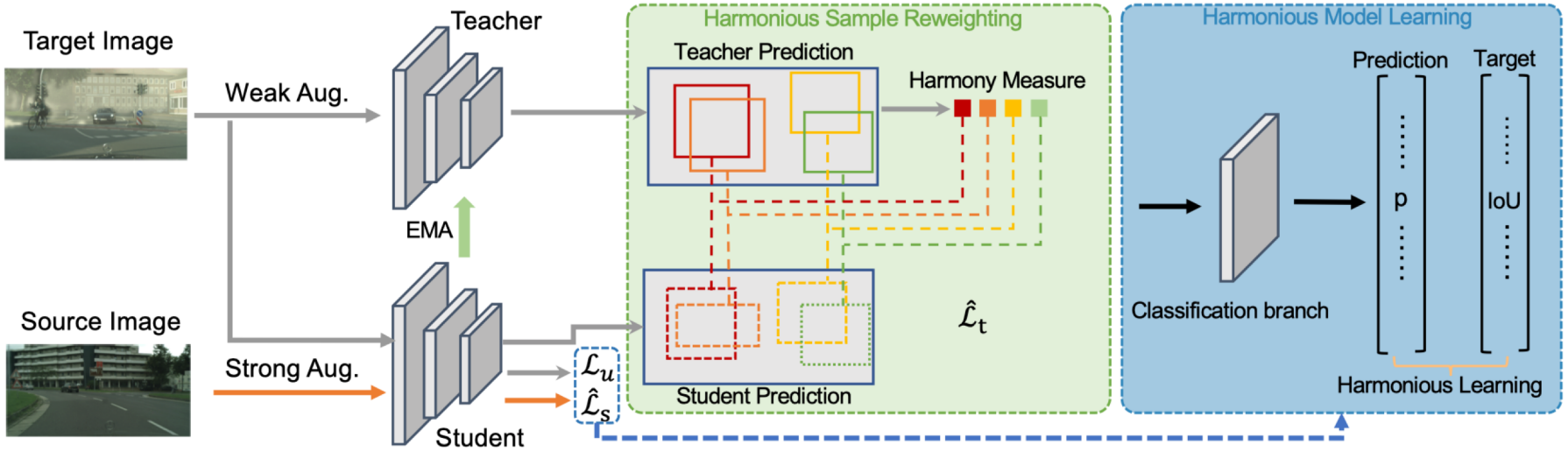


Harmonious Teacher

Motivation

- ❑ Challenges
  - Adopt the classification score to select pseudo labels.
  - Hard thresholding for selecting confident pseudo-labeled instances.
- ❑ Harmonious Teacher
  - Harmonious Model Learning
    - Regularizing the consistency of the classification prediction and the localization score when training the detection model.
  - Harmonious Sample Reweighting
    - All pseudo-labeled samples can contribute to the model training based on prediction qualities
    - The hard threshold is not needed anymore.

Harmonious Teacher



- Harmonious Model Learning: Regularizing the consistency of the classification prediction and the localization score when training the detection model.
- Harmonious Sample Reweighting
  - All pseudo-labeled samples can contribute to the model training based on prediction qualities
  - The hard threshold is not needed anymore.

Harmonious Teacher

Harmonious Model Learning

- Improve the consistency of the classification and the localization score

$$\mathcal{L}_h(y,p) = \begin{cases} -y\left(y\log(p) + (1-y)\log(1-p)\right) & y > 0 \\ -\alpha p^{\gamma}\log(1-p) & y = 0 \end{cases}$$

- Supervised Harmonious Learning in the Source Domain

$$\hat{\mathcal{L}}_{cls} = \sum_i \sum_c \mathcal{L}_h\left(y_{i,c}, p_{i,c}\right)$$

- Unsupervised Harmonious Learning in the Target Domain

$$\mathcal{L}_u = \sum_i \sum_c \mathcal{L}_h\left(\hat{u}_{i,c}, q_{i,c}\right)$$

  - pick the maximum IoU as a substitute of GT-IoU.

Harmonious Sample Reweighting

- Harmonious Measure

$$h = p^{\beta} u^{(1-\beta)}$$

  - It considers the joint quality from classification and localization branches.
  - It stands for the harmony between scores from classification and localization branches.

- Harmonious Weighting

$$\hat{\mathcal{L}}_t = \sum_i e^{(1-h_i)} \left(\sum_c \hat{\mathcal{L}}^t_{cls} \left(\hat{y}_{i,c}, p_{i,c} \right) + \mathcal{L}^i_{reg} \right)$$

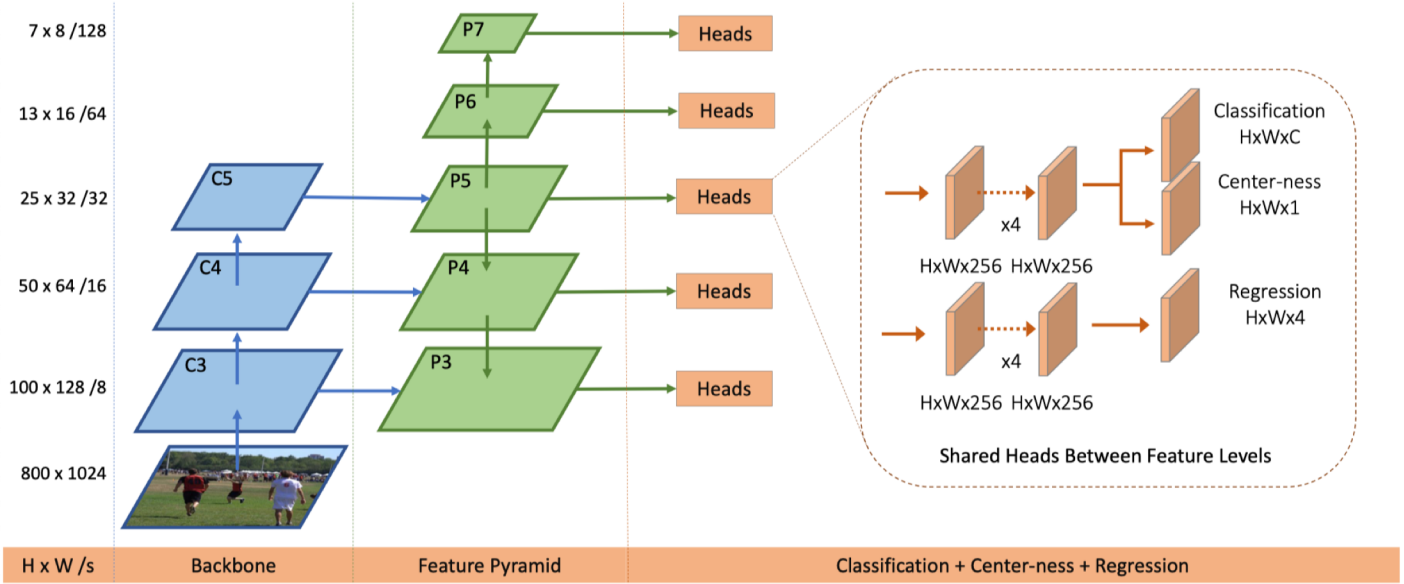
- Overall Objective

$$\mathcal{L} = \hat{\mathcal{L}}_s + \lambda \mathcal{L}_u + \lambda_1 \hat{\mathcal{L}}_t$$

Harmonious Teacher

Experiments

- Implementation Details: FCOS, VGG16, PyTorch



Harmonious Teacher

Results

- Cityscapes to FoggyCityscapes

Table 1. Quantitative results on adaptation from Cityscapes to Foggy Cityscapes with VGG16 backbone network.

Method	Reference	Detector	person	rider	car	truck	bus	train	mcycle	bicycle	mAP
SWDA [30]	CVPR'19	Faster RCNN	29.9	42.3	43.5	24.5	36.2	32.6	30.0	35.3	34.3
CFDA [47]	CVPR'20	Faster RCNN	43.2	37.4	52.1	34.7	34.0	46.9	29.9	30.8	38.6
HTCN [2]	CVPR'20	Faster RCNN	33.2	47.5	47.9	31.6	47.4	40.9	32.3	37.1	39.8
UMT [6]	CVPR'21	Faster RCNN	33.0	46.7	48.6	34.1	56.5	46.8	30.4	37.4	41.7
MeGA [37]	CVPR'21	Faster RCNN	37.7	49.0	52.4	25.4	49.2	46.9	34.5	39.0	41.8
ICCR-VDD [40]	ICCV'21	Faster RCNN	33.4	44.0	51.7	**33.9**	52.0	34.7	34.2	36.8	40.0
TIA [46]	CVPR'22	Faster RCNN	34.8	46.3	49.7	31.1	52.1	48.6	37.7	38.1	42.3
TDD [13]	CVPR'22	Faster RCNN	39.6	47.5	55.7	33.8	47.6	42.1	37.0	41.4	43.1
MGA [49]	CVPR'22	Faster RCNN	45.7	47.5	60.6	31.0	52.9	44.5	29.0	38.0	43.6
PT [3]	ICML'22	Faster RCNN	40.2	48.8	59.7	30.7	51.8	30.6	35.4	44.5	42.7
EPM [14]	ECCV'20	FCOS	41.9	38.7	56.7	22.6	41.5	26.8	24.6	35.5	36.0
SCAN [18]	AAAI'22	FCOS	41.7	43.9	57.3	28.7	48.6	48.7	31.0	37.3	42.1
KTNet [35]	ICCV'21	FCOS	46.4	43.2	60.6	25.8	41.2	40.4	30.7	38.8	40.9
SSAL [25]	NeurIPS'21	FCOS	45.1	47.4	59.4	24.5	50.0	25.7	26.0	38.7	39.6
SIGMA [19]	CVPR'22	FCOS	44.0	43.9	60.3	31.6	50.4	51.5	31.7	40.6	44.2
OADA [43]	ECCV'22	FCOS	47.8	46.5	62.9	32.1	48.5	**50.9**	34.3	39.8	45.4
HT	-	FCOS	**52.1**	**55.8**	**67.5**	32.7	**55.9**	49.1	**40.1**	**50.3**	**50.4**

Harmonious Teacher

Results

- Ablation Studies

Table 5. Ablation studies of HT on Cityscapes → FoggyCityscapes. SHL and UHL denote supervised harmonious and unsupervised harmonious loss. HM-Rank indicates that we use the HM to select pseudo labels. HW is the harmonious weighting.

Method	SHL	UHL	HM-Rank	HW	mAP (%)
Baseline	-	-	-	-	37.3
Proposed	✓				39.1
	✓	✓			40.5
	✓	✓	✓		45.2
	✓	✓		✓	50.4

Results

- Qualitative Results

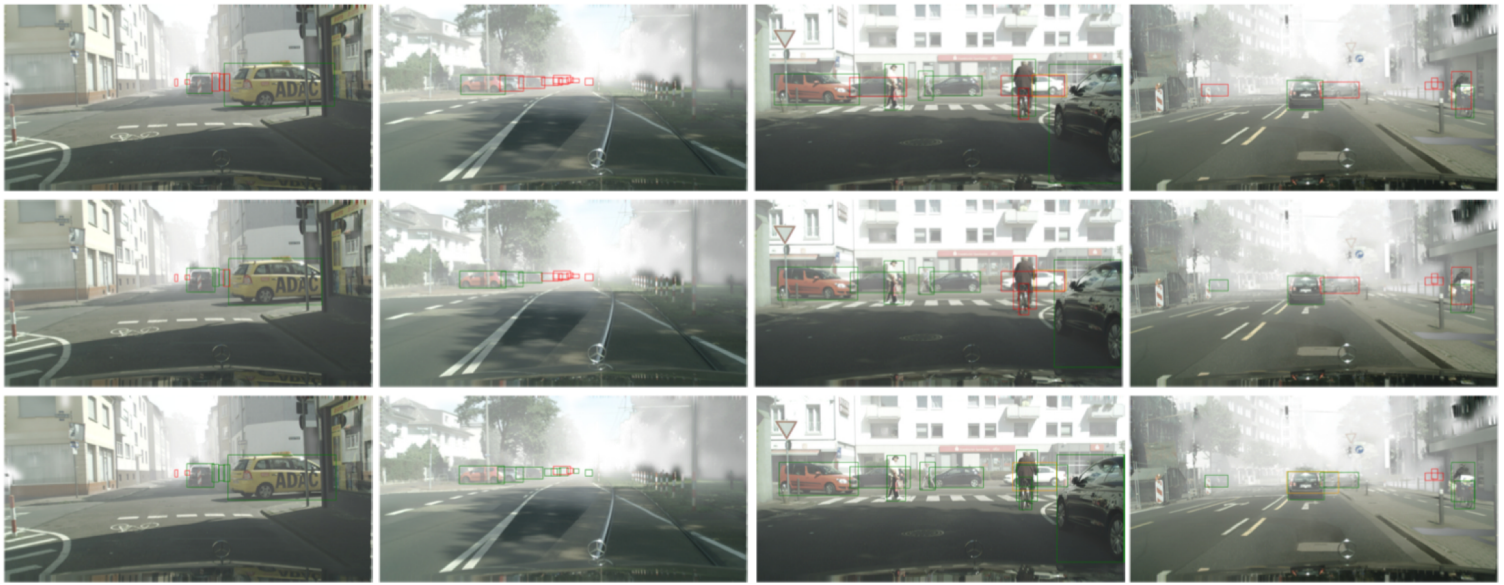


Figure 5. Qualitative results on the target domain of Cityscapes to Foggy Cityscapes for Source Only [36] (top row), SIGMA [19] (middle row) and Ours (bottom row). Green, red and orange boxes indicate true positive (TP), false negative (FN) and false positive (FP), respectively. We set the score threshold to 0.7 for better visualization. Best appreciated when viewed in color and zoomed up.

Thank you!

Email: jhdengvision@gmail.com