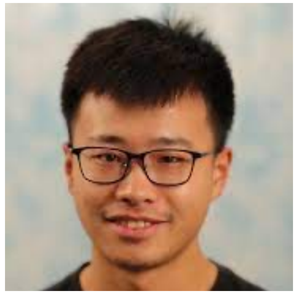
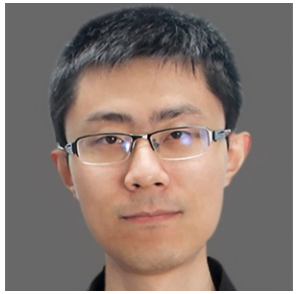
Poster Tag
THU-PM-309

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Contrastive Mean Teacher for Domain Adaptive Object Detectors



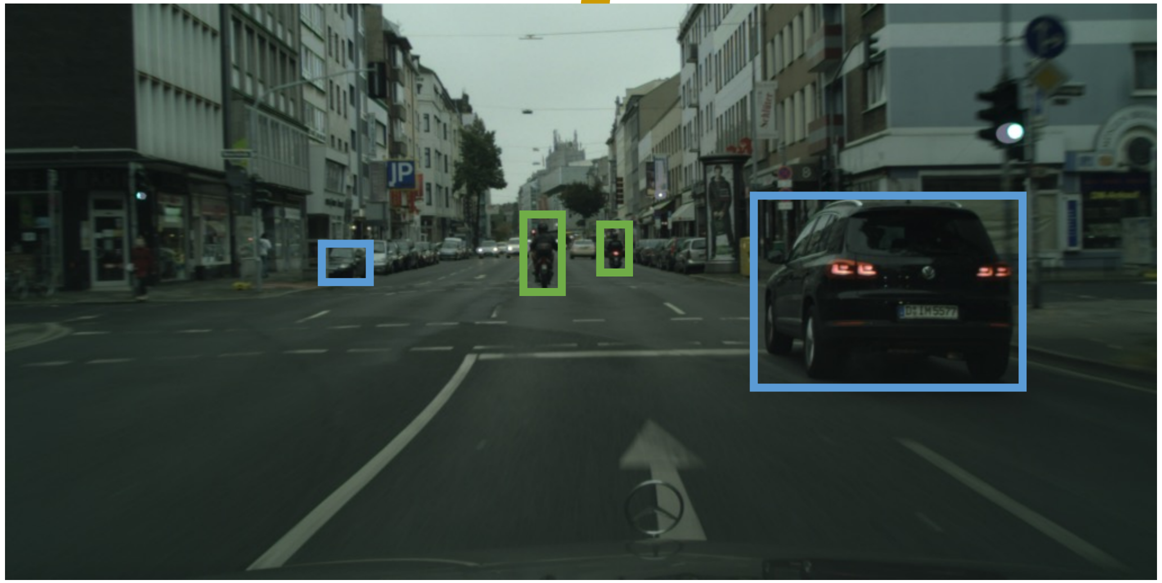
Shengcao Cao[1] Dhiraj Joshi[2] Liang-Yan Gui[1] Yu-Xiong Wang[1]

[1]University of Illinois at Urbana-Champaign [2]IBM Research

Overview

- Problem: **Unsupervised domain adaptation for object detection**

Adapt Object Detector

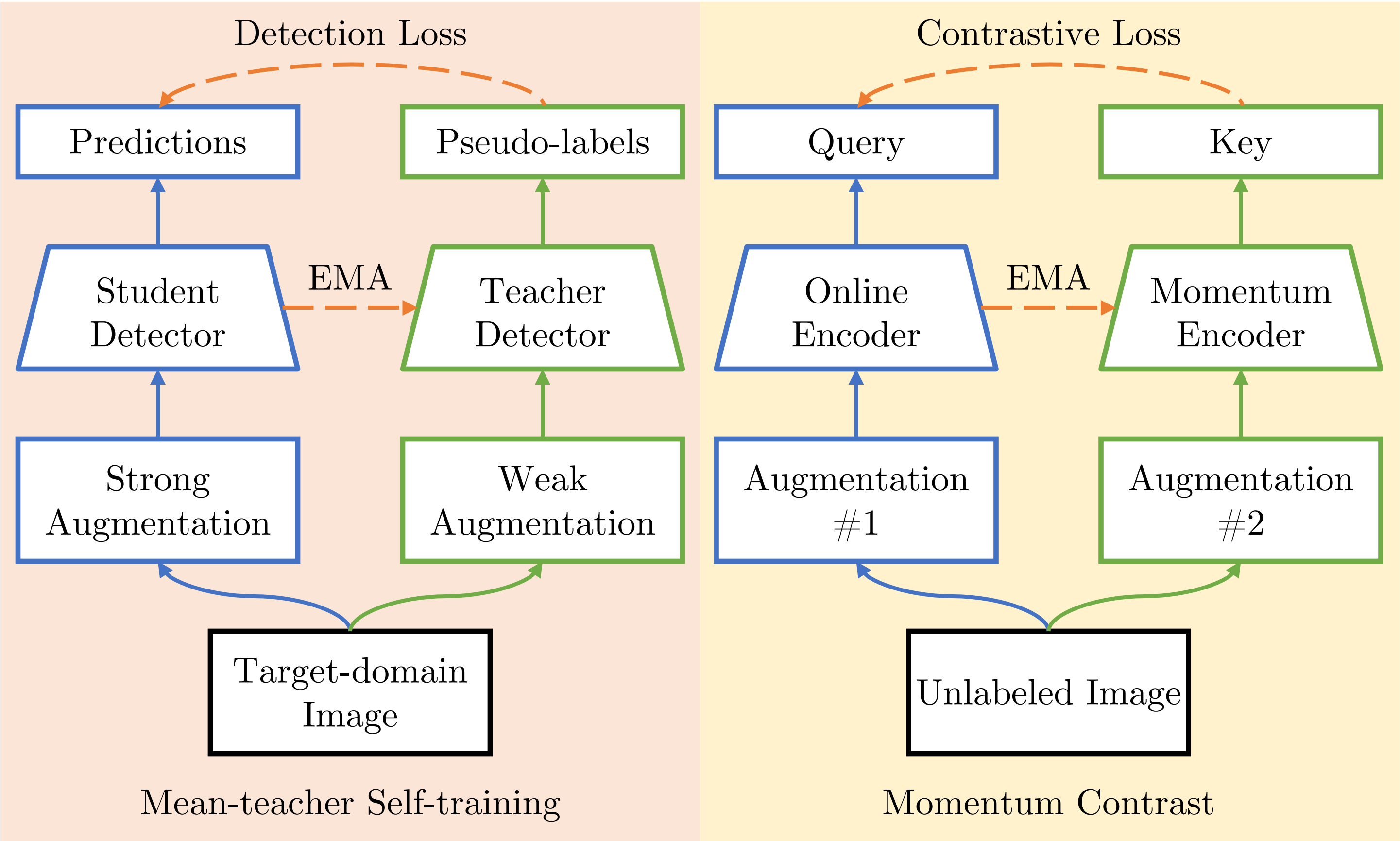


Labeled Source Domain

Unlabeled Target Domain

Overview

- Problem: Unsupervised domain adaptation for object detection
- Solution: **Contrastive Mean Teacher (CMT)**



Why do we need unsupervised domain adaptation for object detectors?

Why do we need unsupervised domain adaptation for object detectors?

Challenge 1: Object-level labels are expensive or even unavailable

Image-level annotation

"Dog"

Why do we need unsupervised domain adaptation for object detectors?

Challenge 1: Object-level labels are expensive or even unavailable

Object-level annotation

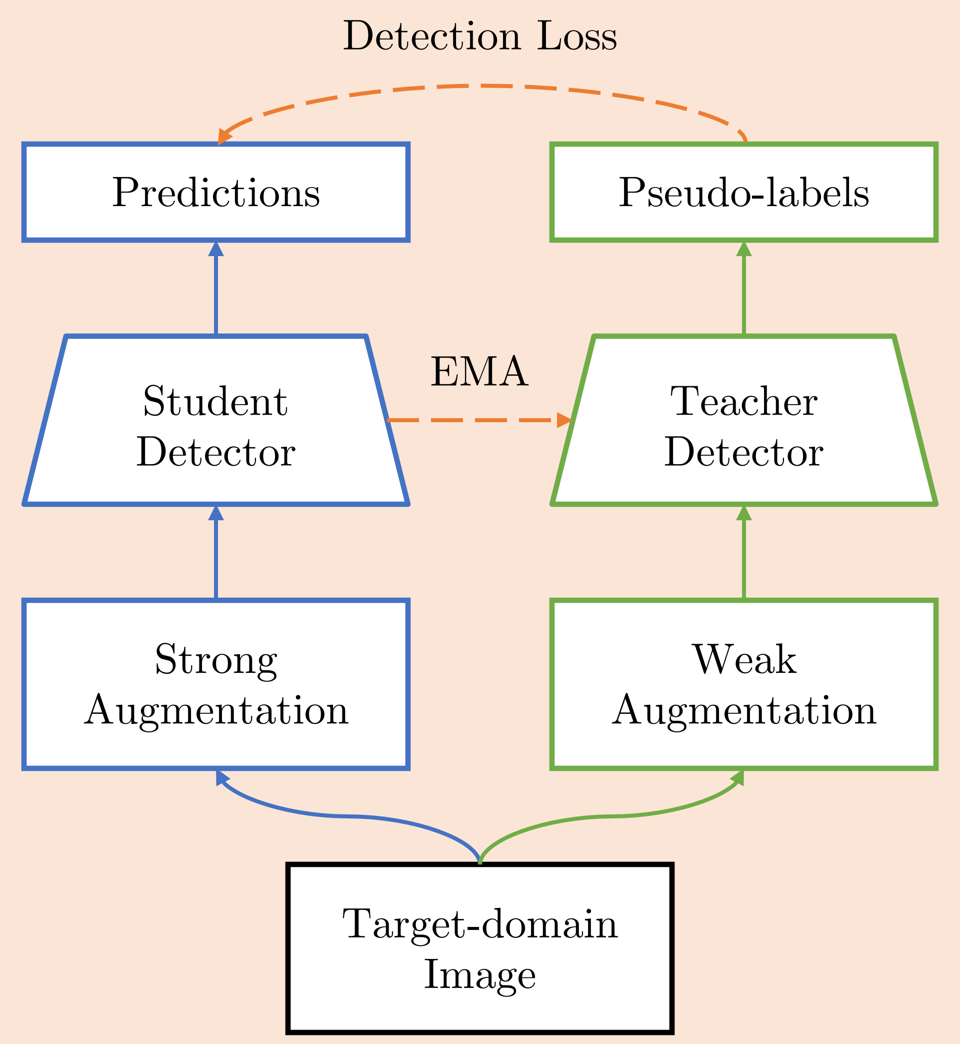
- Category of each object
  - Bounding box
  - Pixel-level mask

Why do we need unsupervised domain adaptation for object detectors?

Challenge 2: Real-world applications may face a huge domain gap or different data distribution

Example: Weather change

Recent Paradigm: Mean-teacher Self-training

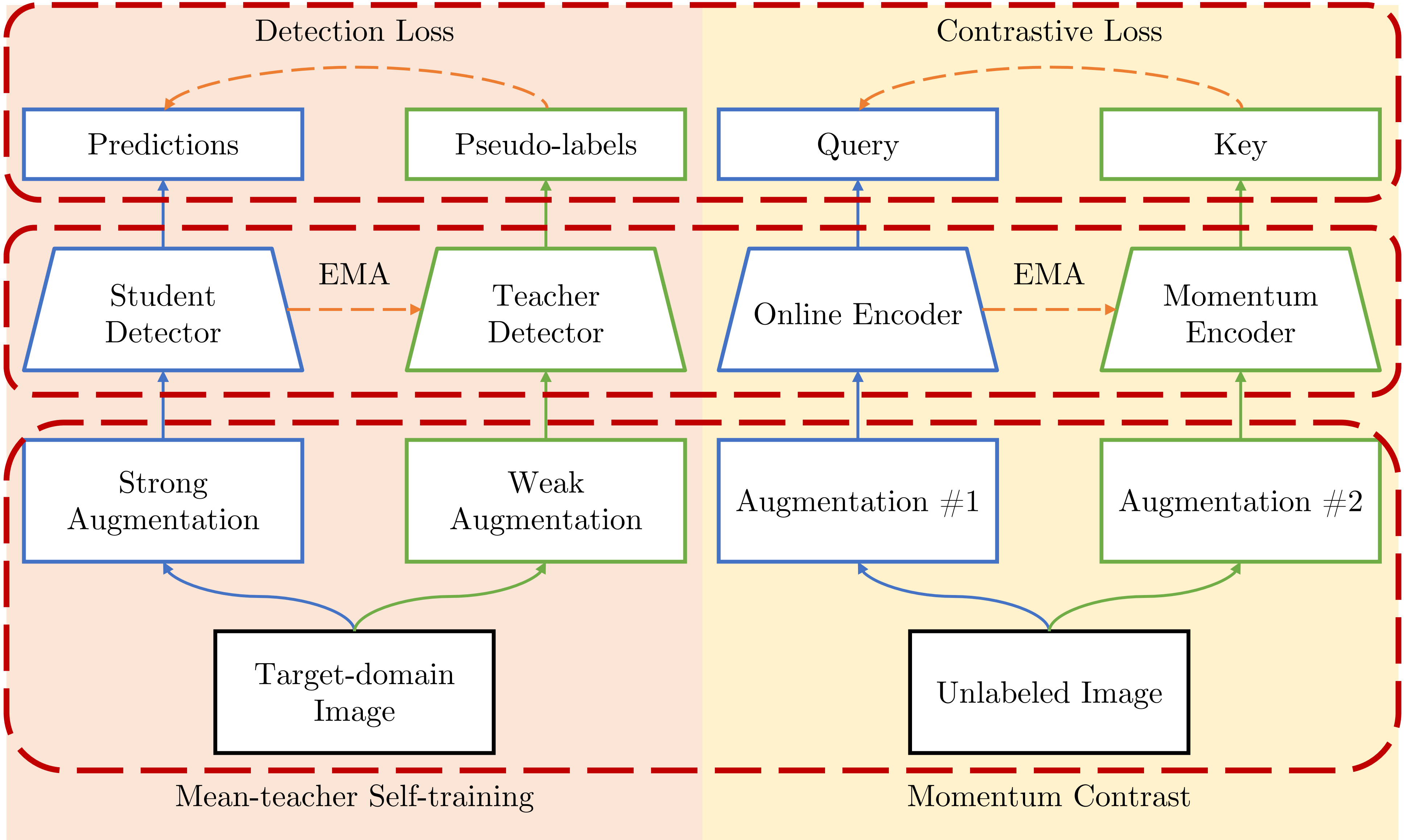


Mean-teacher Self-training

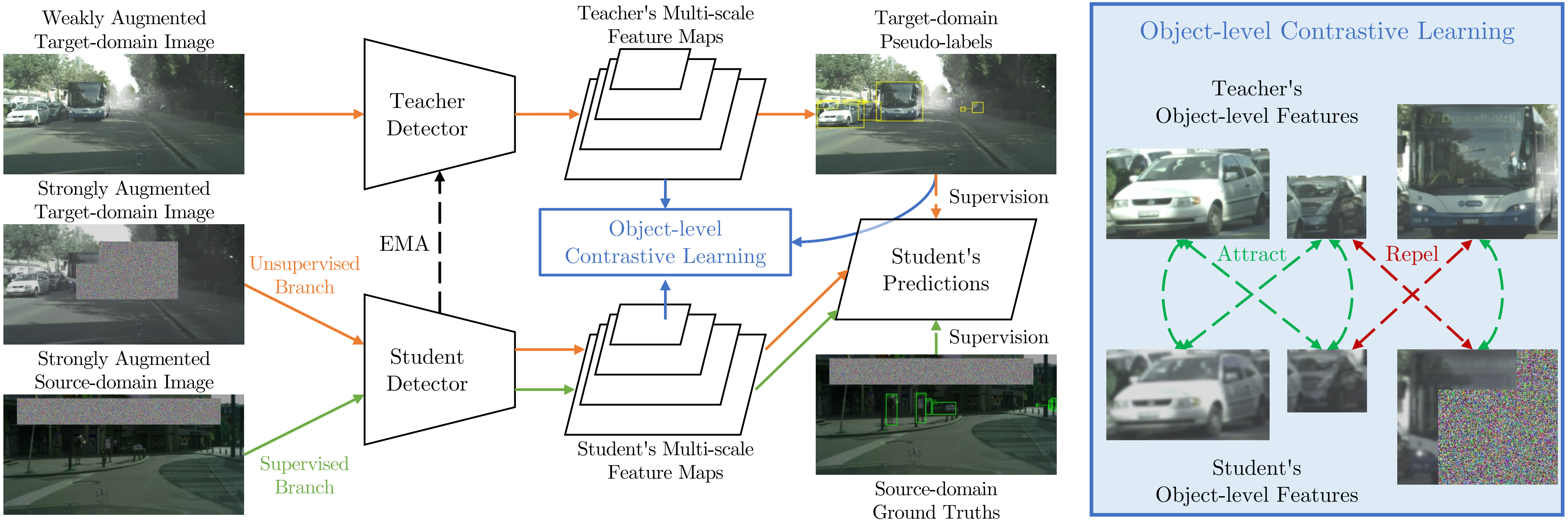
Leads to state-of-the-art domain adaptive object detectors

- Adaptive Teacher (*CVPR*'22)
- Probabilistic Teacher (*ICML*'22)

Aligning Mean Teacher and Momentum Contrast



Contrastive Mean Teacher (CMT)

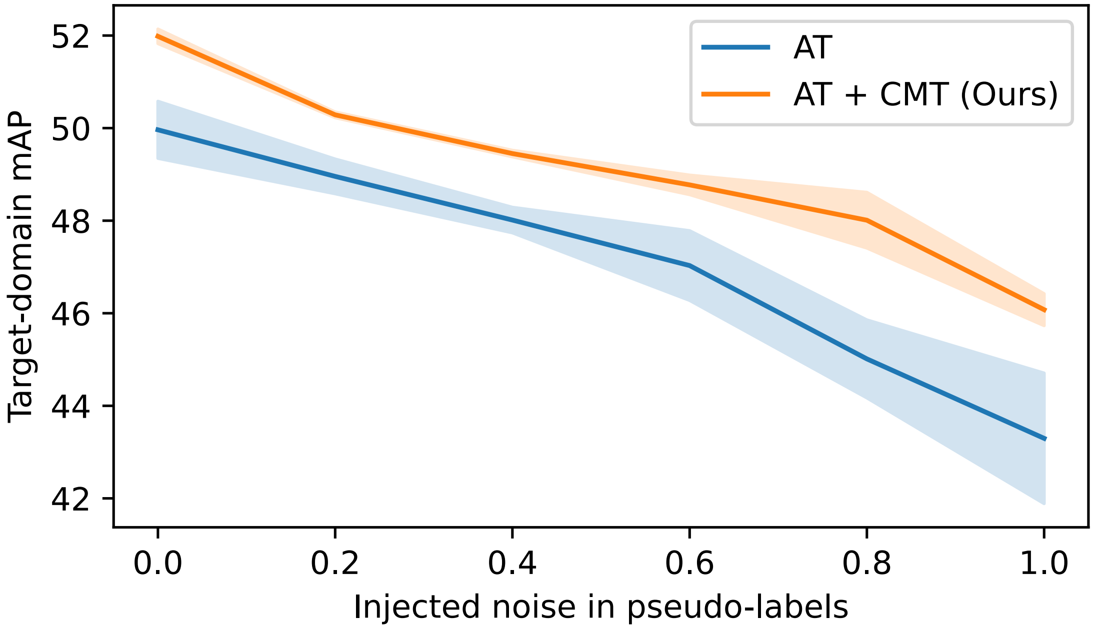


New State of the Art on Cityscapes → Foggy Cityscapes

Method	person	rider	car	truck	bus	train	motor	bike	mAP
Source	27.9	33.4	40.4	12.1	23.2	10.1	20.7	30.9	24.8
Oracle	41.2	49.1	61.6	32.6	56.6	49.0	37.9	42.4	46.3
PDA (*WACV*'20)	36.0	45.5	54.4	24.3	44.1	25.8	29.1	35.9	36.9
ICR-CCR (*CVPR*'20)	32.9	43.8	49.2	27.2	36.4	36.4	30.3	34.6	37.4
PT (*ICML*'22)	43.2	52.4	63.4	33.4	56.6	37.8	41.3	48.7	47.1
PT (*ICML*'22) + CMT (Ours)	45.6	55.1	66.5	34.0	59.4	42.4	43.9	47.4	49.3 (+2.2)
AT (*CVPR*'22)	46.3	55.9	64.3	38.5	61.1	39.3	40.8	52.3	49.8
AT (*CVPR*'22) + CMT (Ours)	47.0	55.7	64.5	39.4	63.2	51.9	40.3	53.1	**51.9 (+2.1)**

Analytical Experiment and Ablation Study

Benefit of CMT is more pronounced when pseudo-labels are noisier

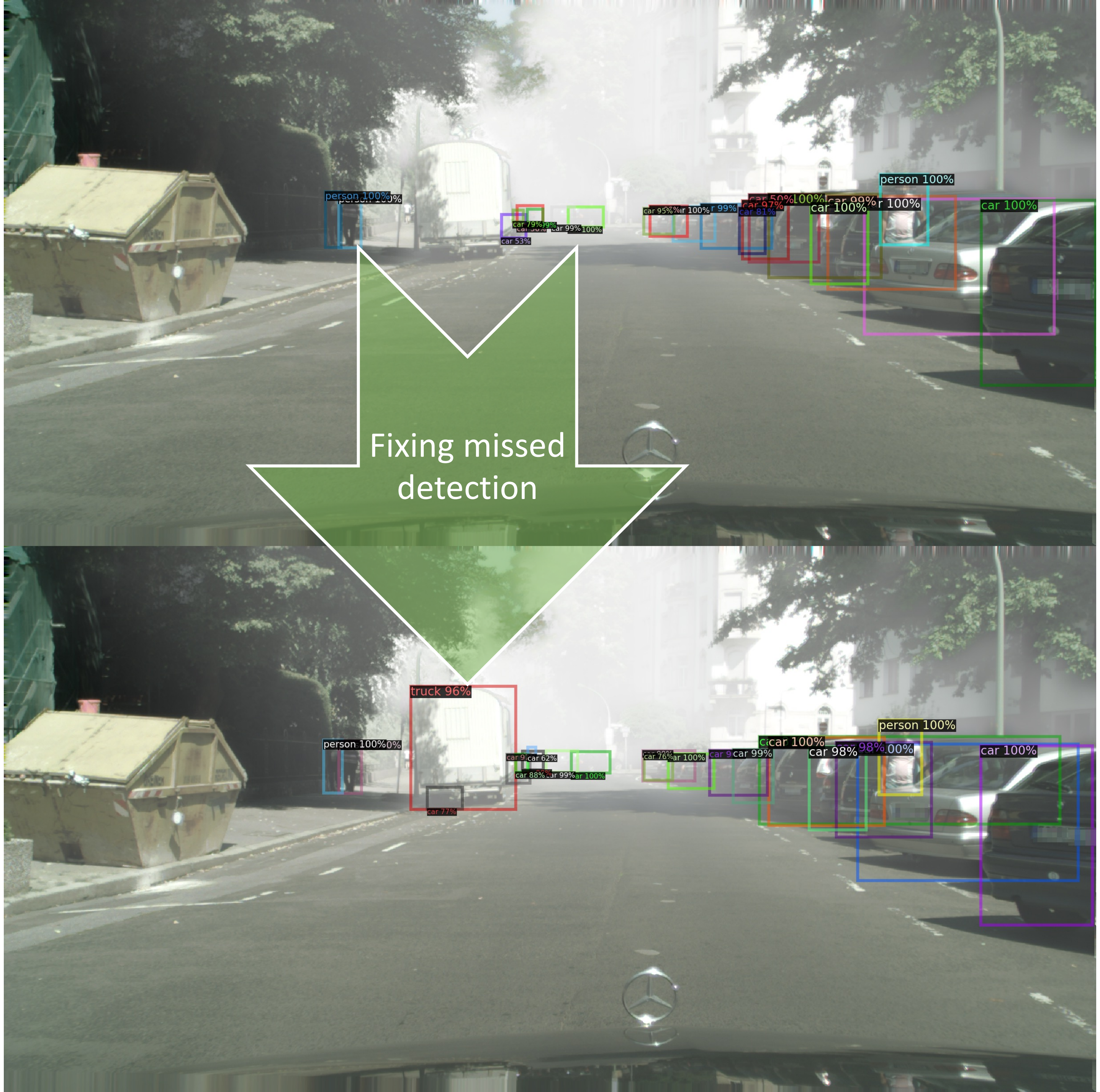


Both our techniques improves object-level contrastive learning

1) Building contrastive pairs using predicted classes
2) Learning from features of multiple scales

Method	Class-based Contrast	Multi-scale Features	mAP	Gain w.r.t. PT
PT	-	-	47.1	-
PT + CMT (Ours)	✗	✗	47.8	+0.7
	✗	✓	48.2	+1.1
	✓	✗	48.7	+1.6
	✓	✓	**49.3**	**+2.2**


bus 84%
car 100%
Fixing mis-classification
train 98%
car 100%


Fixing missed detection
truck 96%

Take-away

Code available here:

- Real-world challenges call for unsupervised domain adaption for object detection
- We align Mean Teacher and Momentum Contrast into one unified framework Contrastive Mean Teacher (CMT)
- Our CMT achieves new state-of-the-art performance on various benchmarks including Cityscapes → Foggy Cityscapes