STRUCT

SimROD: A Simple Adaptation Method for Robust Object Detection

ICCV 2021 (Poster)

Rindra Ramamonjison, Amin Banitalebi-Dehkordi, Xinyu Kang, Xiaolong Bai, and Yong Zhang

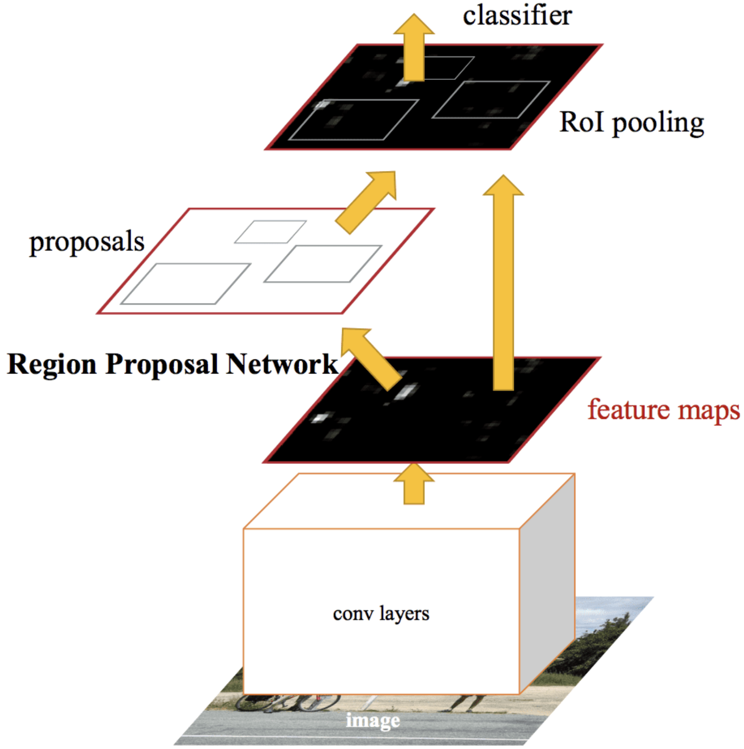
Presented by Yuzhang Hu
2021.12.19

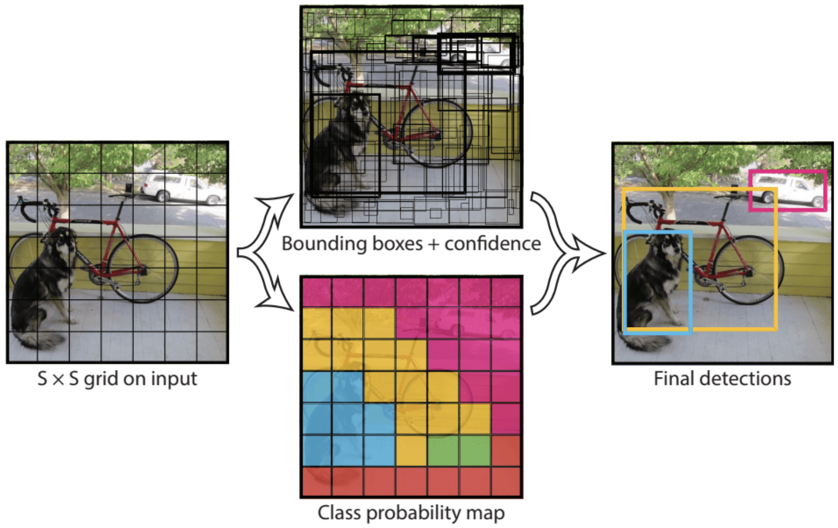
Outline

- Authorship
- Background
- Method
- Experiment
- Conclusion

Background

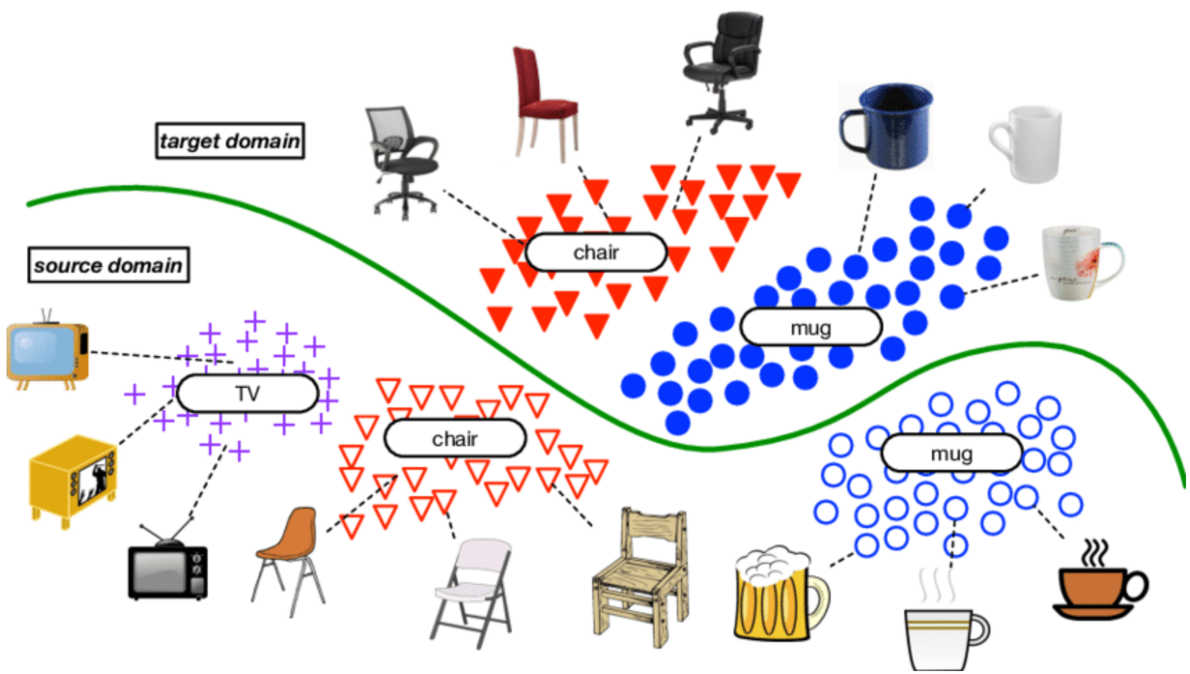
- Object detection
  - Two-stage (Region-based) vs. One-stage





Background

- Domain gap leads to performance drop in testing
- Hard to perform supervision on target domain



Background

- Popular solutions:

  - Augmentation
    - fail to the ones not similar to the augmented samples

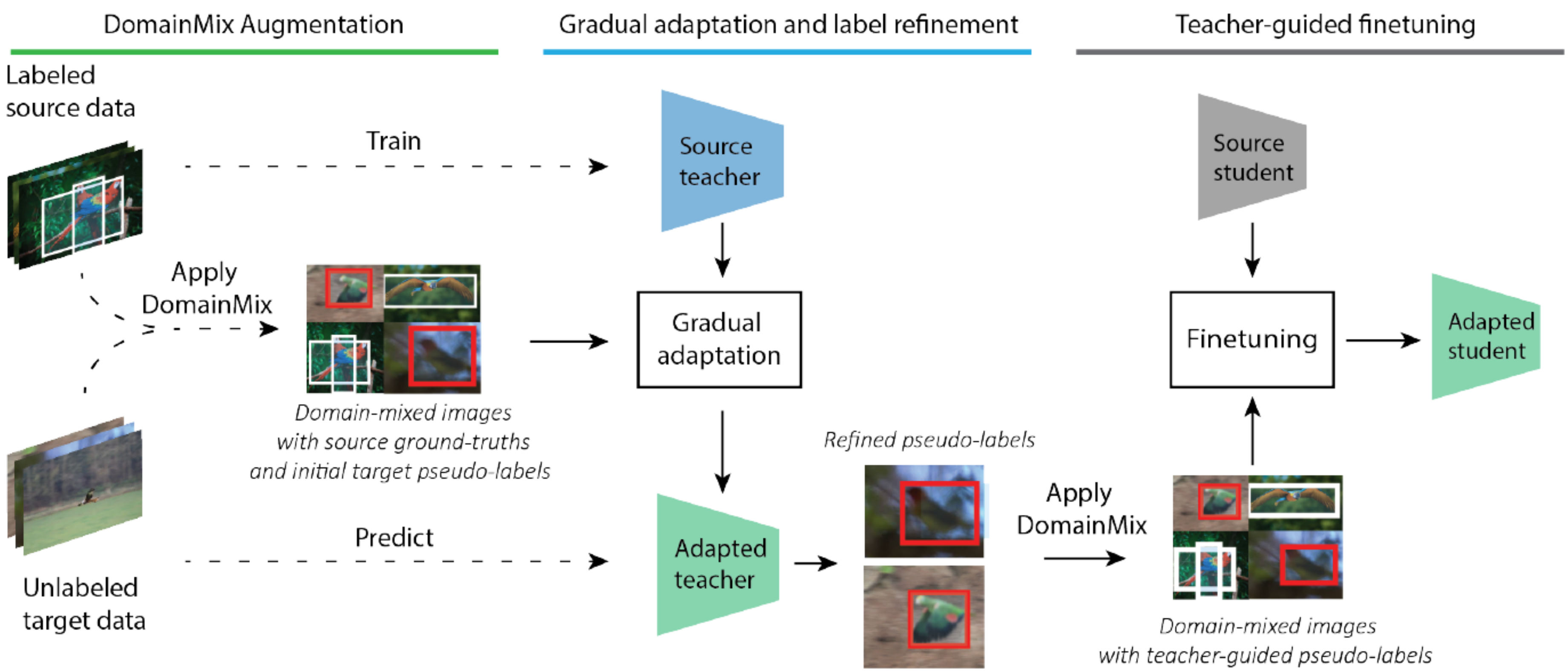
  - Domain-aligning
    - require the addition of specialized modules

Background

- Popular solutions:
  - Domain-mapping
    - generated images not of high similarity
  - Self-labeling
    - hard to generating accurate pseudo-labels

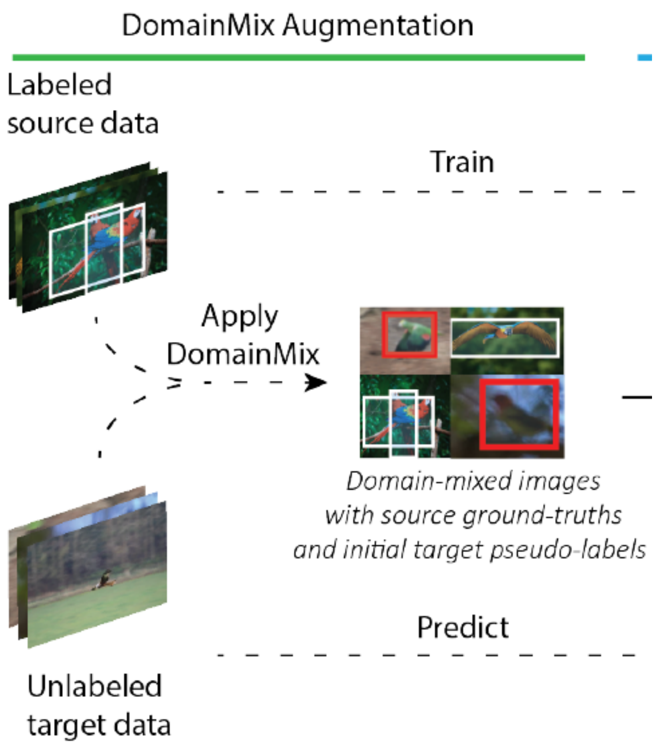
Method

- Overall Structure



Method

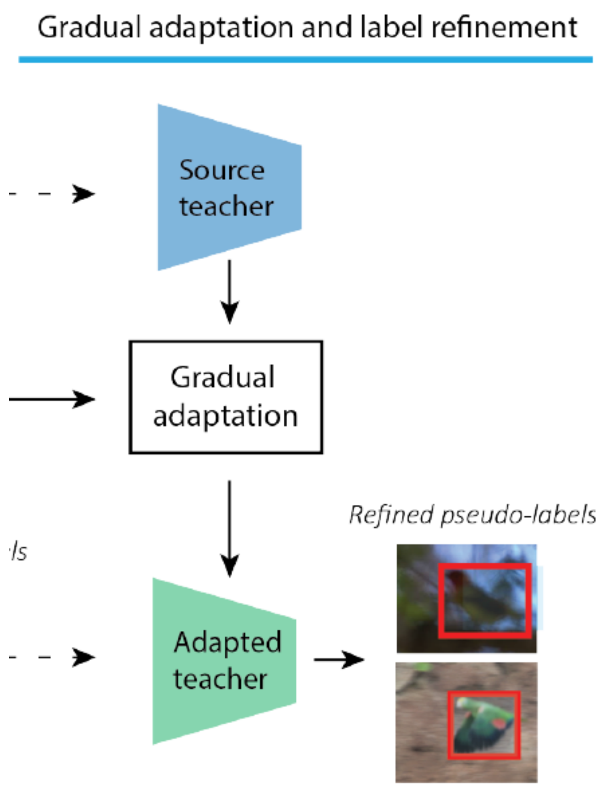
- DomainMix Augmentation



- Sample images from both the source and target domains
- Mix these images along with labels

Method

- Gradual self-labeling adaptation

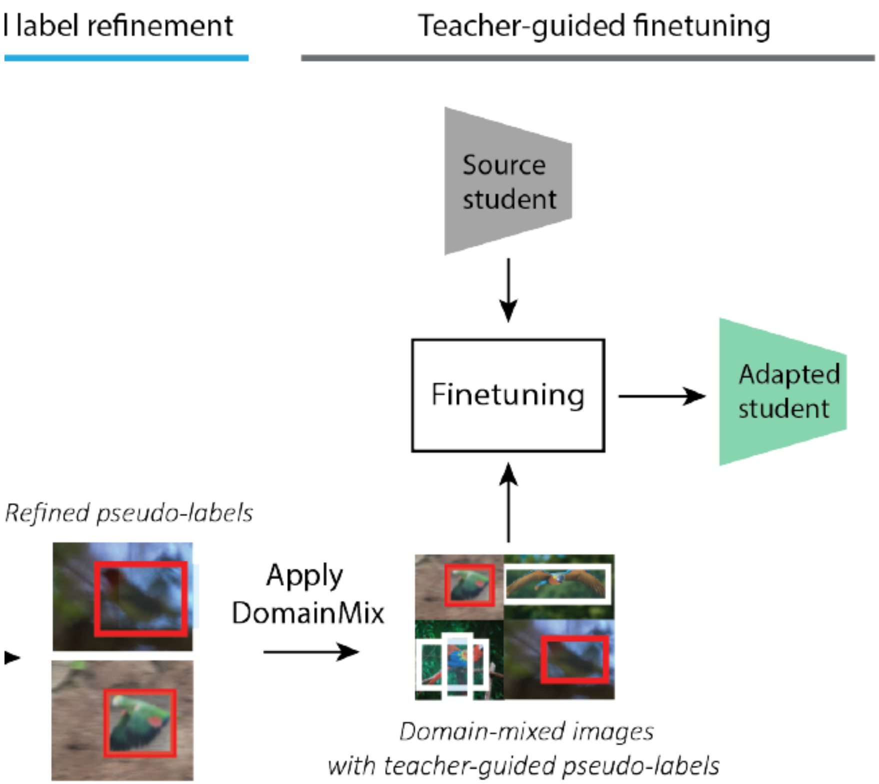


- Adaptation in two phases
- First phase:
  - Freeze all convolutional layers except for the BN layers
  - Generate more accurate pseudo-labels with the partially adapted model

- Second phase:
  - Unfreeze all layers and finetuned using the refined pseudo-labels

Method

- Teacher-guided finetuning



- Generate more accurate pseudo-labels with the final adapted teacher model
- Finetune the student model with the refined pseudo-labels

Experiment

$$\tau = \mathrm{AP}^{50}(\theta^a) - \mathrm{AP}^{50}(\theta^s),$$

$$\rho = 100 \times \frac{\mathrm{AP}^{50}(\theta^a) - \mathrm{AP}^{50}(\theta^s)}{\mathrm{AP}^{50}(\mathrm{Oracle}) - \mathrm{AP}^{50}(\theta^s)},$$

Synthetic-to-real result

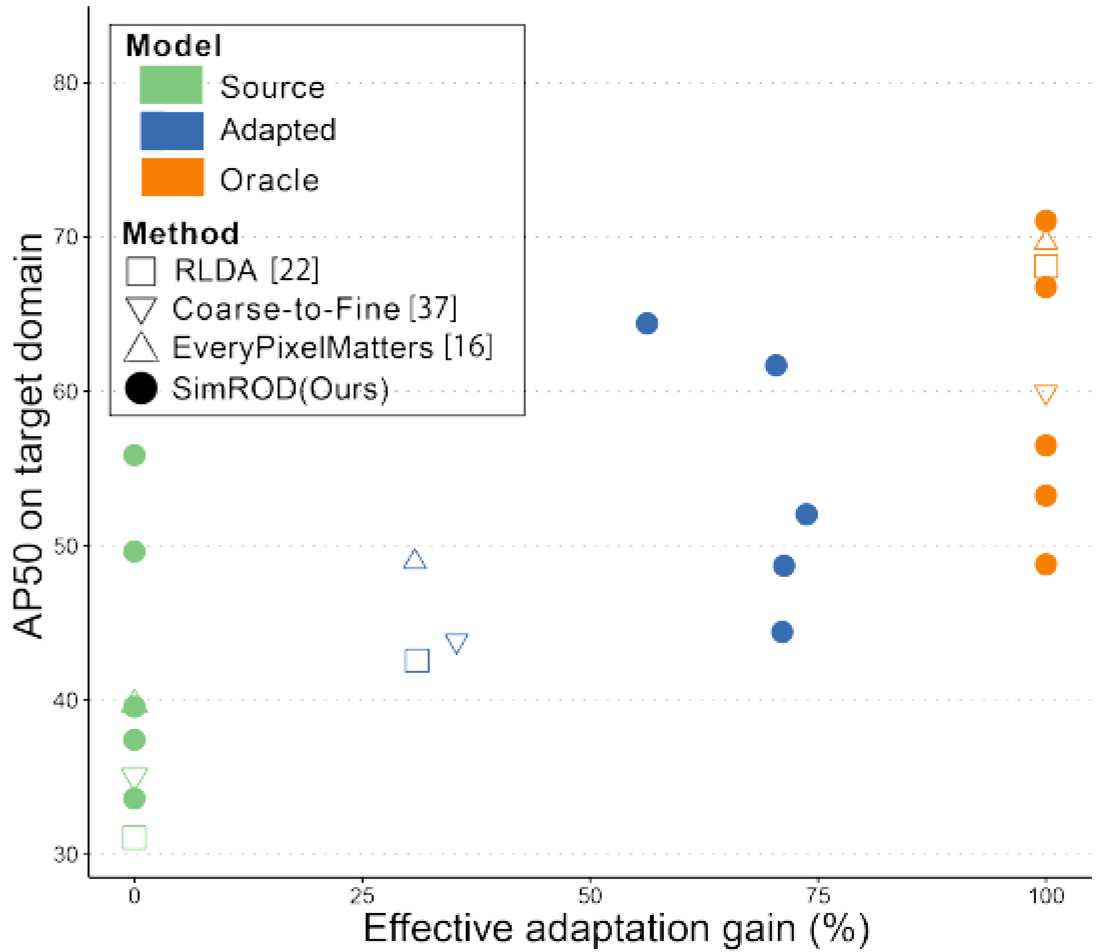
Method	Arch.	Backbone	Source	AP50	Oracle	τ	ρ	Reference
DAF [6]	F-RCNN	V	30.10	39.00	-	8.90	-	CVPR 2018
MAF [11]	F-RCNN	V	30.10	41.10	-	11.00	-	ICCV 2019
RLDA [22]	F-RCNN	I	**31.08**	**42.56**	68.10	**11.48**	**31.01**	ICCV 2019
SCDA [38]	F-RCNN	V	34.00	43.00	-	9.00	-	CVPR 2019
MDA [36]	F-RCNN	V	34.30	42.80	-	8.50	-	ICCV 2019
SWDA [27]	F-RCNN	V	34.60	42.30	-	7.70	-	CVPR 2019
Coarse-to-Fine [37]	F-RCNN	V	**35.00**	43.80	59.90	8.80	35.34	CVPR 2020
SimROD (self-adapt)	YOLOv5	S320	33.62	38.73	48.81	5.11	33.66	Ours
SimROD (w. teacher X640)	YOLOv5	S320	33.62	**44.70**	48.81	**11.08**	**72.93**	Ours
MTOR [4]	F-RCNN	R	39.40	46.60	-	7.20	-	CVPR 2019
EveryPixelMatters [16]	FCOS	V	**39.80**	49.00	69.70	9.20	30.77	ECCV 2020
SimROD (self adapt)	YOLOv5	S416	39.57	44.21	56.49	4.63	27.37	Ours
SimROD (w. teacher X1280)	YOLOv5	S416	39.57	**52.05**	56.49	**12.47**	**73.73**	Ours

- Source domain: Sim10K
- Target domain: Cityscapes

Experiment

Synthetic-to-real result

- Source domain: Sim10K
- Target domain: Cityscapes



Experiment

$$\tau = \mathrm{AP}^{50}(\theta^a) - \mathrm{AP}^{50}(\theta^s),$$

$$\rho = 100 \times \frac{\mathrm{AP}^{50}(\theta^a) - \mathrm{AP}^{50}(\theta^s)}{\mathrm{AP}^{50}(\mathrm{Oracle}) - \mathrm{AP}^{50}(\theta^s)},$$

Cross-camera result

Method	Arch.	Backbone	Source	AP50	Oracle	τ	ρ	Reference
DAF [6]	F-RCNN	V	30.20	38.50	-	8.30	-	CVPR 2018
MAF [11]	F-RCNN	V	30.20	41.00	-	10.80	-	ICCV 2019
RLDA [22]	F-RCNN	I	31.10	42.98	68.10	11.88	32.11	ICCV 2019
PDA [17]	F-RCNN	V	30.20	43.90	55.80	13.70	53.52	WACV 2020
SimROD (self-adapt)	YOLOv5	S416	31.61	35.94	56.15	4.33	17.65	Ours
SimROD (w. teacher X1280)	YOLOv5	S416	**31.61**	**45.66**	56.15	**14.05**	**57.27**	Ours
SCDA [38]	F-RCNN	V	**37.40**	42.60	-	5.20	-	CVPR 2019
EveryPixelMatters [16]	FCOS	R	35.30	45.00	70.40	9.70	27.64	ECCV 2020
SimROD (self adapt)	YOLOv5	M416	36.09	42.94	59.29	6.85	29.51	Ours
SimROD (w. teacher X1280)	YOLOv5	M416	36.09	**47.52**	59.29	**11.43**	**49.26**	Ours

- Source domain: KITTI
- Target domain: Cityscapes

Experiment

$$\tau = \mathrm{AP}^{50}(\theta^a) - \mathrm{AP}^{50}(\theta^s),$$

$$\rho = 100 \times \frac{\mathrm{AP}^{50}(\theta^a) - \mathrm{AP}^{50}(\theta^s)}{\mathrm{AP}^{50}(\mathrm{Oracle}) - \mathrm{AP}^{50}(\theta^s)},$$

Ablation Study

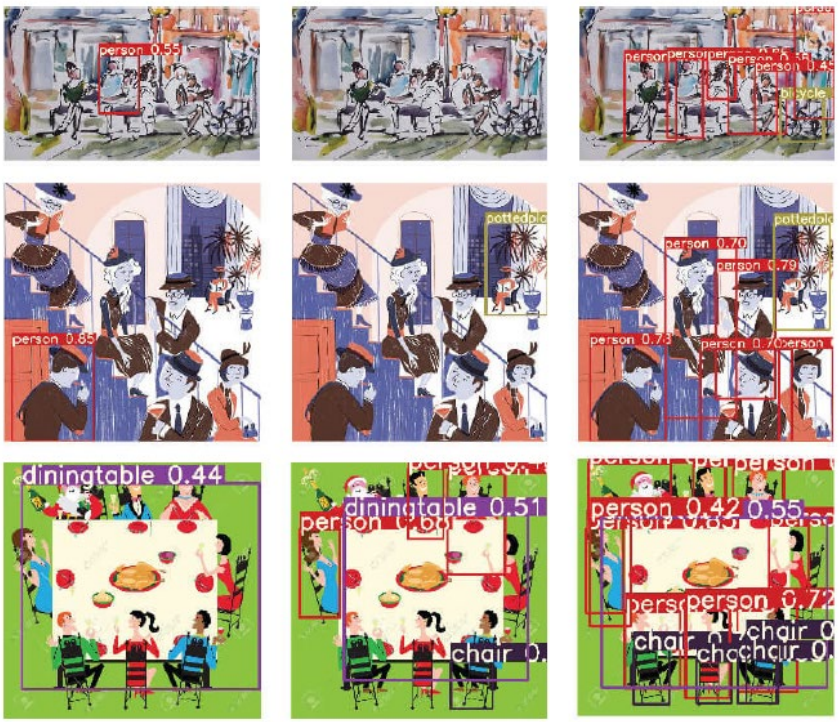
Method	TG	DMX	GA	FT	mPC^{50}	τ_c
Source					53.78	0.0
BN-Adapt			✓		64.60	10.8
BN-A + DMX		✓	✓		66.78	13.0
SimROD w/o TG		✓	✓	✓	71.81	18.0
SimROD w/o GA	✓	✓		✓	73.45	19.7
SimROD	✓	✓	✓	✓	75.40	21.7

Table 7. Ablation study on Pascal-C with yolov5m. See [2] for ablations with other models. TG, GA, DMX, and FT denote Teacher Guidance, Gradual Adaption, DomainMix, and Fine-Tuning.

- Self adaptation is not enough (Line 2-4)
- The gradual adaptation is important in refining pseudo-labels

Experiment

Qualitative analysis



STAC DT+PL Ours

STAC DT+PL Ours

Conclusion

- Disentangle the pseudo-label generation to a stronger teacher net for easier adaptation
- Gradual adaptation with DomainMix to improve the accuracy of the pseudo-label

Thanks!