2024 동계 세미나

Source-Free Domain Adaptive Object Detection

Sogang University
Vision & Display Systems Lab, Dept. of Electronic Engineering

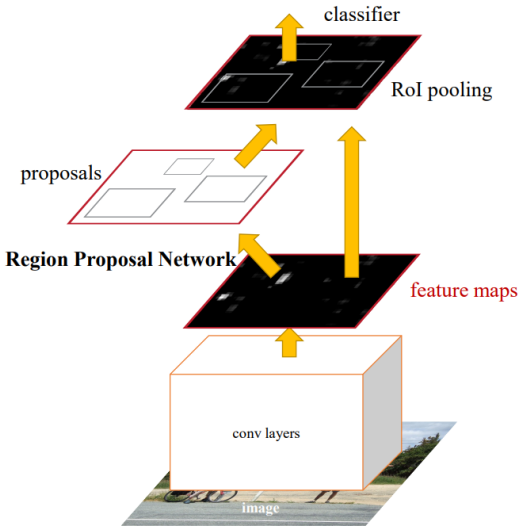
Presented By
박지원

Outline

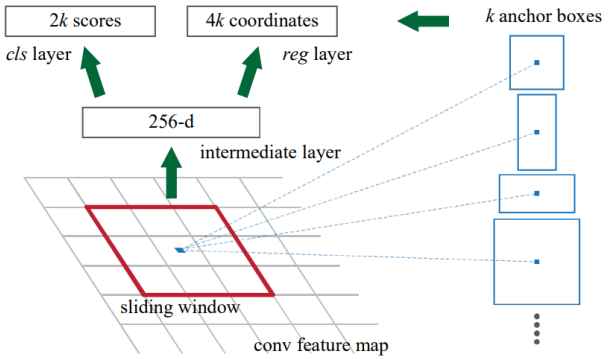
- Background
  - Object detection
  - Source-Free Domain adaptation
- Instance Relation Graph Guided Source-Free Domain Adaptive Object Detection
  - CVPR 2023
- Periodically Exchange Teacher-Student for Source-Free Object Detection
  - ICCV 2023

Background

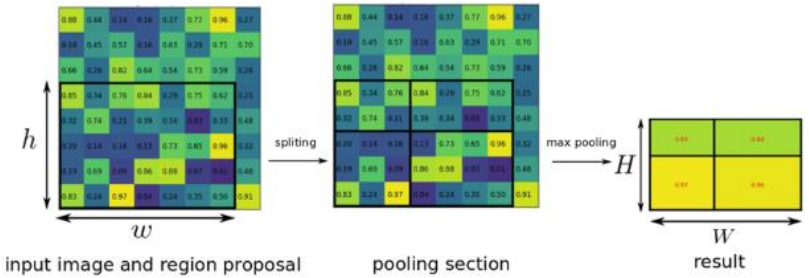
- Object detection
  - 개념
    - 이미지/비디오에서 object의 식별하고 분류하는 컴퓨터 비전 기술
  - Faster R-CNN
    - Region Proposal Network (RPN)
      - Background / foreground 구분, bounding box 결정
    - RoI(Region of Interest) pooling
      - 다양한 크기의 region proposals 로부터 고정된 크기의 feature map 얻음



Faster R-CNN 구조



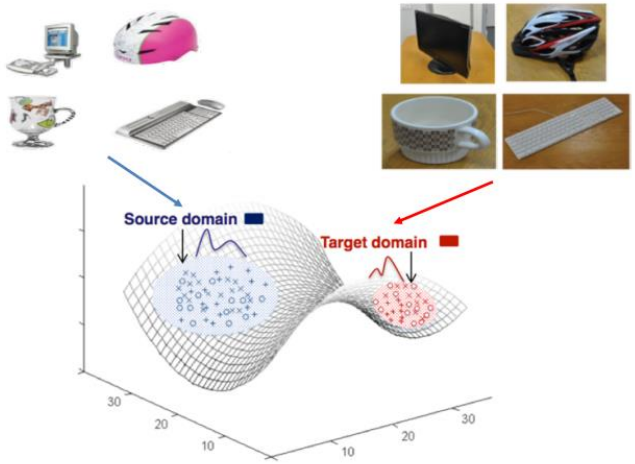
RPN



RoI pooling

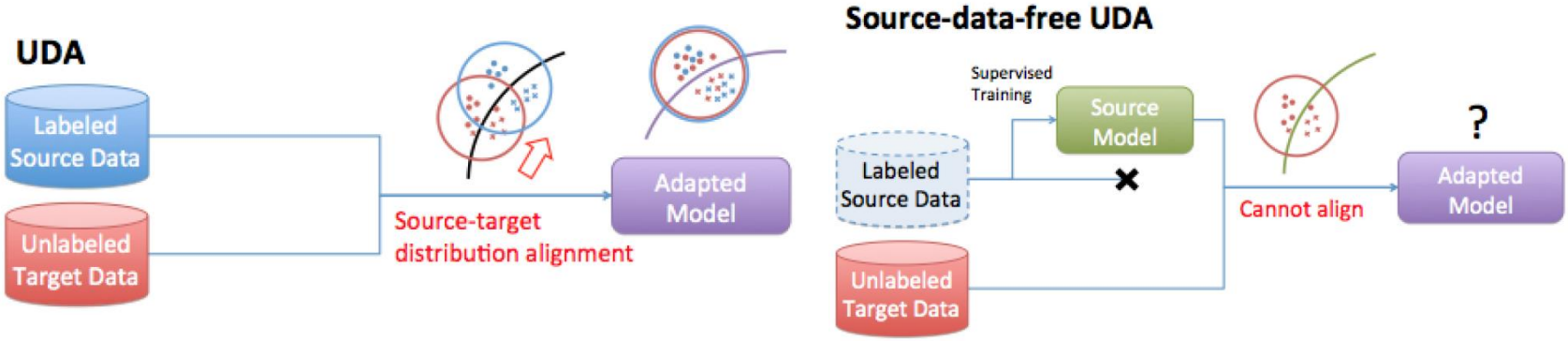
Background

- Source-free Unsupervised Domain Adaptation
  - Domain adaptation
    - 특정 domain에서 학습된 모델을 다른 domain 으로 adapt 하려는 것
      - Source domain data: 모델이 학습하는 데이터
      - Target domain data: source data 로 학습한 모델이 적응하고자 하는 데이터
    - Domain 간의 domain gap 을 극복하고 source domain 에서 학습된 모델을 target domain 에 효과적으로 적응하기 위한 방법론 연구
      - Domain gap: source domain 과 target domain 의 분포 상의 차이



Background

- Source-free Unsupervised Domain Adaptation
  - Unsupervised Domain Adaptation(UDA)
    - 타겟 도메인의 데이터가 라벨 없이도 task 를 구행할 수 있도록 학습
  - Source-free UDA
    - Source model 과 라벨이 없는 target data 를 통해 target domain 에 adapting 하는 방법론



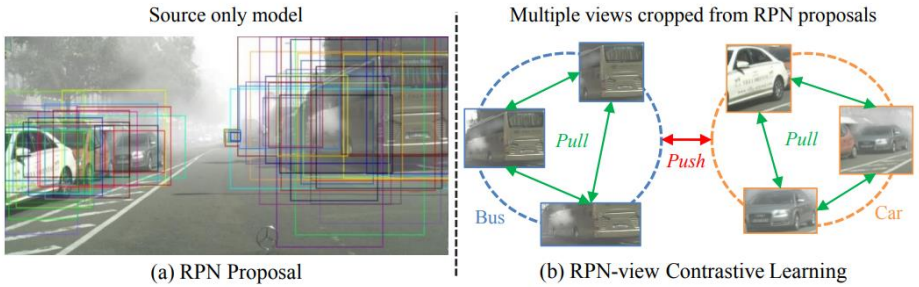
VS, Oza, et al. “Instance Relation Graph Guided Source-Free Domain Adaptive Object Detection.” CVPR, 2023.

Introduction

- Source-free domain

(a) Prediction (b) Ground truth

Figure 2. (a) Object predictions by Cityscapes-trained model on the FoggyCityscapes image. (b) Corresponding ground truth. Here, the proposals around the bus instance have inconsistent predictions, indicating that instance features are prone to large shift in the feature space, for a small shift in the proposal location.



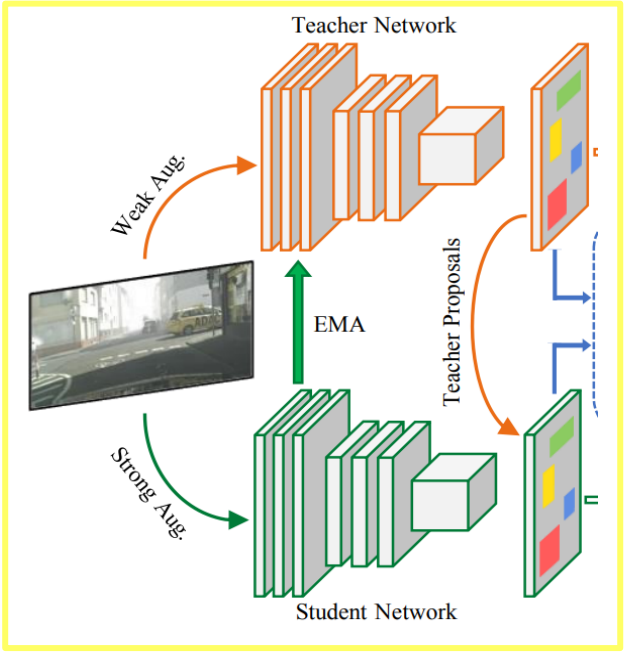
(a) RPN Proposal (b) RPN-view Contrastive Learning

Figure 3. (a) Class agnostic object proposals generated by Region Proposal Network (RPN). (b) Cropping out RPN proposals will provide multiple contrastive views of an object instance. We utilize this to improve target domain feature representations through RPN-view contrastive learning. However as RPN proposals are class agnostic, it is challenging to form positive (same class)/negative pairs (different class), which is essential for CRL.

Mean-teacher based self-training

- Teacher-student architecture
  - Teacher network
    - Weak augmentation image
    - Student network 에 region proposal 제공
    - Student network의 weight 로 모델 업데이트 (EMA)

$$\Theta_t \leftarrow \alpha\Theta_t + (1-\alpha)\Theta_s$$



Framework 일부

  - Student network
    - Strong augmentation image
    - Teacher network 로부터 제공받은 proposal 로 모델 학습

$$\Theta_s \leftarrow \Theta_s + \gamma\frac{\partial(\mathcal{L}_{SL}^{st})}{\partial\Theta_s}$$

$\mathcal{L}_{SL}^{st}$: Student loss computed using the pseudo-labels generated by the teacher network

Graph-guided contrastive learning

- Instance Relation Graph (IRG)
  - Notation

$\mathcal{G} : \mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ $\mathcal{V}$: Nodes $\mathcal{E}$: Edges

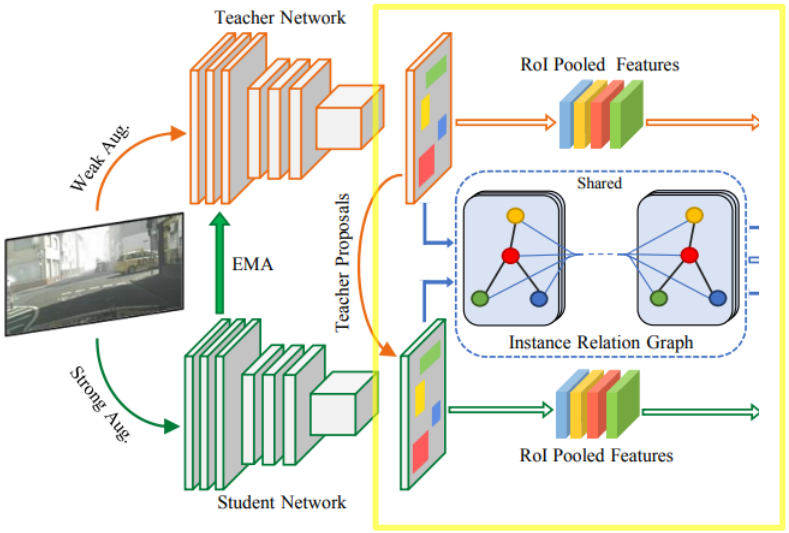
  - Nodes
    - RoI features extracted from RPN proposals
    - Teacher RPN proposals 사용
  - Edges
    - Instance relation matrix (instance들 간의 유사도)

$\mathcal{E} = [e_{ij}]_{m \times m}$ e_{ij} : Edges of the v_i^{th} and v_j^{th}

$$e_{ij} = \frac{\exp(S_{ij})}{\sum \exp(S_{ij})} \quad (softmax)$$

where $S_{ij} = f(v_i) \cdot g(v_j)^T$

(f and g : learnable function)



Framework 일부

Graph-guided contrastive learning

- Graph Distillation Loss (GDL)
  - IRG 통과 전후의 RoI features 모두 사용
    - Input feature to IRG: $F \in \mathbb{R}^{m \times d}$
    - Output features of IRG: $\tilde{F} = \mathrm{ReLU}(\mathcal{E}FW)$

$\mathcal{E}$: *instance relation matrix*

m: proposal instances 개수
d: dimention of RoI features

w: learnable weight matrix

    - 위 두 features 를 모두 RCNN classification layer 의 입력으로 사용

  - 두 features 로부터 얻은 class logits의 discrepancy minimize 학습
    - IRG 통과 전후 RoI feature 의 consistency 유지
    - Teacher 과 student 의 consistency 유지
      - Z_{st} , Z_{te} : student and teacher class logits corresponding to features F
      - $\tilde{Z}_{st}$, $\tilde{Z}_{te}$: student and teacher class logits corresponding to features $\tilde{F}$

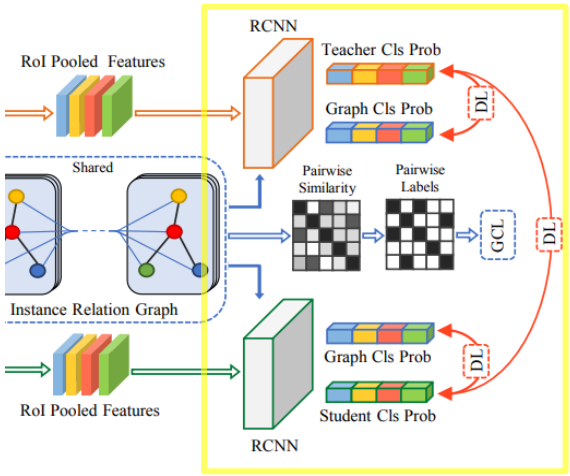
$$\mathcal{L}_{GDL} = \mathrm{KL}(\sigma(Z_{st}), \sigma(\tilde{Z}_{st})) + \mathrm{KL}(\sigma(Z_{te}), \sigma(\tilde{Z}_{te})) + \mathrm{KL}(\sigma(Z_{st}), \sigma(Z_{te}))$$

KL: Kullback–Leibler divergence
(두 확률분포의 차이를 계산)

Graph-guided contrastive learning

- Graph Contrastive Loss (GCL)
  - Contrastive learning
    - Representation space 에서 유사한 데이터는 서로 가깝게, 다른 데이터는 떨어져있도록 학습하는 기법
    - Positive/negative pair 가 필요
  - Instance pairwise labels & logits
    - Instance relation matrix 의 thresholding 을 통해 positive/negative pair 선정

$$M_{ij} = \begin{cases} 0, & e_{ij} < \epsilon \\ 1, & e_{ij} \geq \epsilon, \end{cases}$$



    - RoI feature 로부터 key, query, pairwise logits 계산

$$k_i = W_k \cdot v_i,$$
$$q_i = W_q \cdot v_i,$$
$$R_{ij} = q_i (k_j)^T,$$

Contrastive loss 계산

$$\mathcal{L}_{GCL} = \sum_{i \in I} -\log \left\{ \frac{1}{|P(i)|} \sum_{p \in P(i)} \frac{\exp\left(q_i (k_p)^T\right)}{\sum_{a \in A(i)} \exp\left(q_i (k_a)^T\right)} \right\}$$

Experiments

- Cityscapes → FoggyCityscapes 실험 결과
  - Adaptation to adverse weather

Table 1. Quantitative results (mAP) for Cityscapes → FoggyCityscapes. S: Source only, O: Oracle, UDA: Unsupervised Domain Adaptation, SFDA: Source-Free Domain Adaptation.

Type	Method	prsn	rider	car	truck	bus	train	mcycle	bicycle	mAP
S	Source Only	29.3	34.1	35.8	15.4	26.0	9.09	22.4	29.7	25.2
UDA	DA Faster [8]	25.0	31.0	40.5	22.1	35.3	20.2	20.0	27.1	27.6
	D&Match [33]	30.8	40.5	44.3	27.2	38.4	34.5	28.4	32.2	34.6
	MTOR [2]	30.6	41.4	44.0	21.9	38.6	40.6	28.3	35.6	35.1
	SWDA [56]	29.9	42.3	43.5	24.5	36.2	32.6	30.0	35.3	34.3
	CDN [60]	35.8	45.7	50.9	30.1	42.5	29.8	30.8	36.5	36.6
	Collaborative DA [75]	32.7	44.4	50.1	21.7	45.6	25.4	30.1	36.8	35.9
	iFAN DA [78]	32.6	48.5	22.8	40.0	33.0	45.5	31.7	27.9	35.3
	Instance DA [78]	33.1	43.4	49.6	21.9	45.7	32.0	29.5	37.0	36.5
	Progressive DA [26]	36.0	45.5	54.4	24.3	44.1	25.8	29.1	35.9	36.9
	Categorical DA [71]	32.9	43.8	49.2	27.2	45.1	36.4	30.3	34.6	37.4
	MeGA CDA [26]	37.7	49.0	52.4	25.4	49.2	46.9	34.5	39.0	41.8
	Unbiased DA [11]	33.8	47.3	49.8	30.0	48.2	42.1	33.0	37.3	40.4
SFDA	SFOD [40]	21.7	44.0	40.4	32.2	11.8	25.3	34.5	34.3	30.6
	SFOD-Mosaic [40]	25.5	44.5	40.7	**33.2**	22.2	**28.4**	34.1	39.0	33.5
	HCL [27]	26.9	**46.0**	41.3	33.0	25.0	28.1	**35.9**	40.7	34.6
	LODS [39]	34.0	45.7	48.8	27.3	39.7	19.6	33.2	37.8	35.8
	Mean-Teacher [61]	33.9	43.0	45.0	29.2	37.2	25.1	25.6	38.2	34.3
	IRG (Ours)	**37.4**	45.2	**51.9**	24.4	**39.6**	25.2	31.5	**41.6**	**37.1**
O	Oracle	38.7	46.9	56.7	35.5	49.4	44.7	35.9	38.8	43.1

Cityscapes

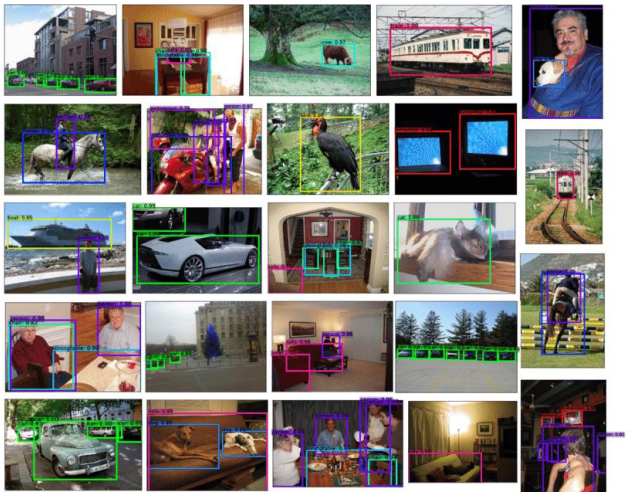
FoggyCityscapes

Experiments

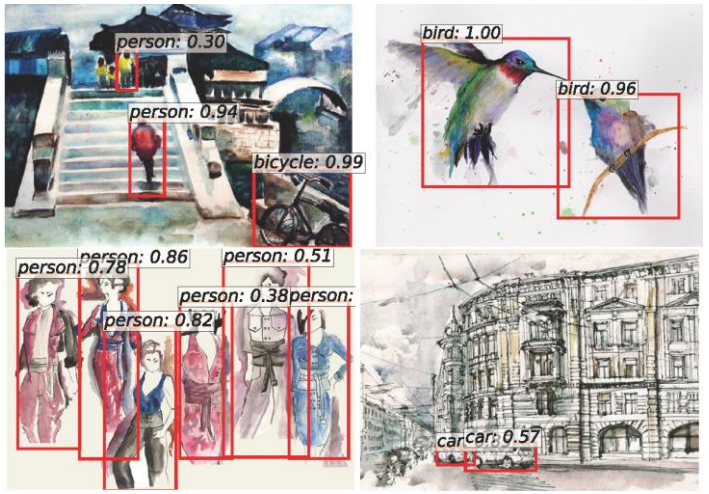
- PASCAL-VOC → Watercolor 실험 결과
  - Realistic to artistic data adaptation

Table 3. Quantitative results for PASCAL-VOC → Watercolor. S: Source only, UDA: Unsupervised Domain Adaptation, SFDA: Source-Free domain adaptation.

Type	Method	bike	bird	car	cat	dog	prsn	mAP
S	Source only	68.8	46.8	37.2	32.7	21.3	60.7	44.6
UDA	DA Faster [8]	75.2	40.6	48.0	31.5	20.6	60.0	46.0
	BDC Faster [56]	68.6	48.3	47.2	26.5	21.7	60.5	45.5
	BSR [32]	82.8	43.2	49.8	29.6	27.6	58.4	48.6
	WST [32]	77.8	48.0	45.2	30.4	29.5	64.2	49.2
	SWDA [56]	71.3	52.0	46.6	36.2	29.2	67.3	50.4
	HTCN [3]	78.6	47.5	45.6	35.4	31.0	62.2	50.1
	I^3Net [4]	81.1	49.3	46.2	35.0	31.9	65.7	51.5
	Unbiased DA [11]	88.2	55.3	51.7	39.8	43.6	69.9	55.6
SFDA	PL [30]	74.6	46.5	45.1	27.3	25.9	54.4	46.1
	SFOD [40]	**76.2**	44.9	49.3	**31.6**	30.6	55.2	47.9
	Mean-teacher [61]	73.6	47.6	46.6	28.5	29.4	56.6	47.1
	IRG (Ours)	75.9	**52.5**	**50.8**	30.8	**38.7**	**69.2**	**53.0**



PASCAL-VOC



Watercolor

Experiments

- Ablation study on FoggyCityscapes
  - Input of student and teacher network
    - Weak-Weak (WW)
    - Strong-Strong (SS)
    - Strong-Weak (SW)
  - Graph Distillation Loss (GDL)
  - Graph Contrastive Loss (GCL)

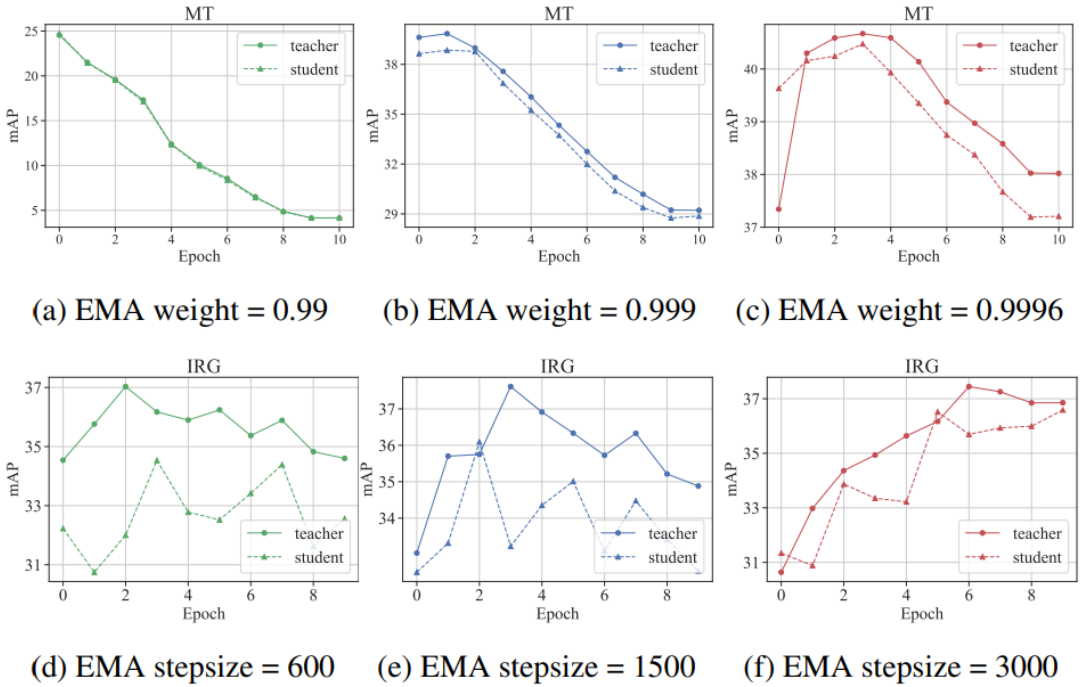
Table 5. Ablation study on FoggyCityscapes.

Method	PL	GDL	GCL	prsn	rider	car	truc	bus	train	mcycle	bcycle	mAP
Source Only	✗	✗	✗	25.8	33.7	35.2	13.0	28.2	9.1	18.7	31.4	24.4
MT + WW	✓	✗	✗	35.8	42.6	43.9	23.1	32.7	11.0	29.9	38.7	32.2
MT + SS	✓	✗	✗	32.8	41.4	43.8	18.2	28.6	11.2	24.6	38.3	29.9
MT + SW	✓	✗	✗	33.9	43.0	45.0	**29.1**	37.2	25.1	25.5	38.2	34.3
Ours	✓	✓	✗	37.2	43.1	51.0	28.6	**40.1**	21.2	28.2	37.1	35.9
Ours	✓	✓	✓	**37.4**	**45.2**	**51.9**	24.4	39.6	**25.2**	**31.5**	**41.6**	**37.1**

Liu, Lin, et al. “Periodically Exchange Teacher-Student for Source-Free Object Detection.” ICCV, 2023.

Introduction

- Mean-teacher framework 의 문제점 지적
  - Teacher 모델은 student 모델의 오류를 누적하는 문제점 존재
    - Source로 pretrain된 모델이 target domain에 적용될 때 내재된 bias 존재
      - Teacher 모델은 source로 pretrain 된 모델에서 파생된 student 모델의 EMA 업데이트이기 때문

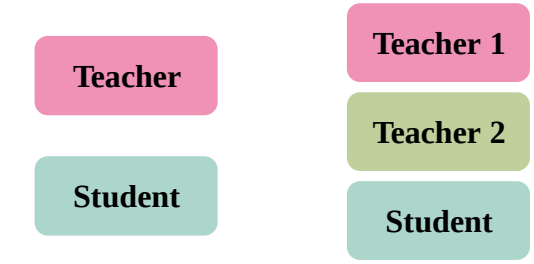


(a) EMA weight = 0.99 (b) EMA weight = 0.999 (c) EMA weight = 0.9996

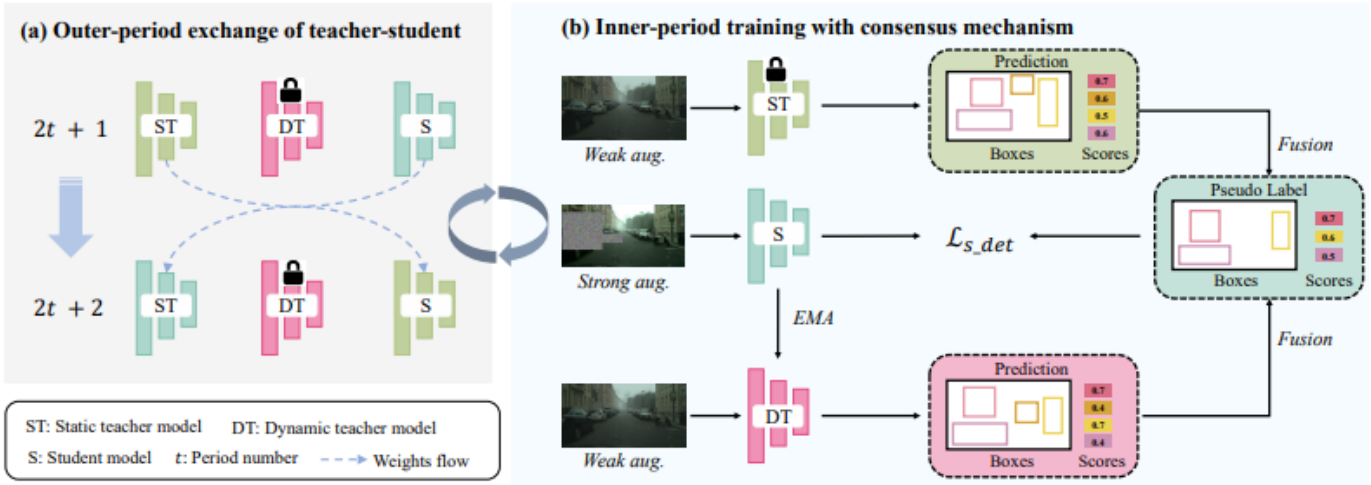
(d) EMA stepsize = 600 (e) EMA stepsize = 1500 (f) EMA stepsize = 3000

Introduction

- Framework
  - Outer-period exchange of teacher-student
    - 매 epoch 마다 static teacher 모델과 student 모델의 가중치를 교환
  - Inner-period training with consensus mechanism
    - 매 iteration 마다 dynamic teacher 은 student 모델의 EMA 로 업데이트
      - 이때 static teacher 모델의 가중치는 고정



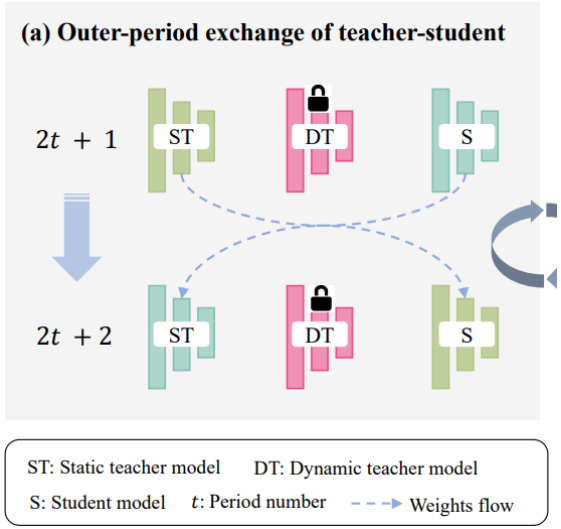
General framework Proposed framework



Framework

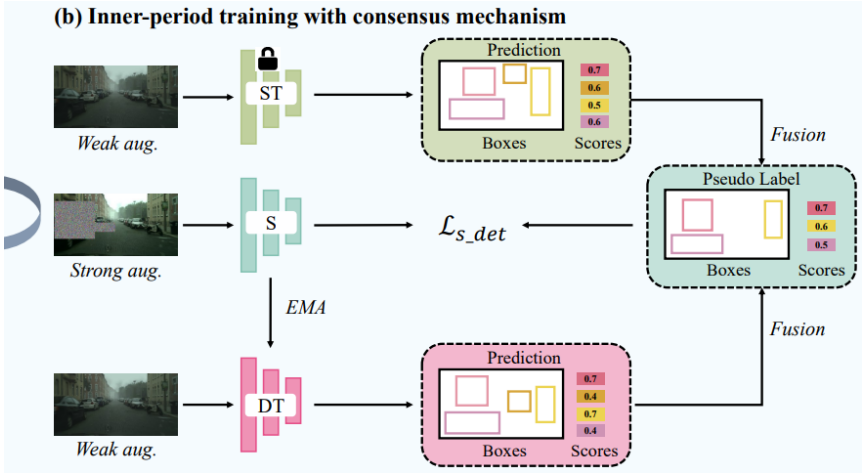
Outer-period exchange of teacher-student

- Epoch 기준 outer-period exchange of teacher-student
  - 각 모델에 제공하는 이점
    - Student: static teacher 모델은 student 모델에 대한 성능 lower bound 역할
      - 만약 성능 저하된 teacher로 인해 student가 무너진다면 이 교환을 통해 student 가 이전 epoch으로 돌아가게 함으로써 감소 추세를 완화시킬 수 있음
    - Dynamic teacher: 과거 student 로부터 교환된 현재의 student의 일시적 앙상블
      - 실제로 dynamic teacher 의 업데이트 속도는 감소되면서 noise 에 대한 저항력 향상



Inner-period training with consensus mechanism

- Iteration 기준 inner-period training
  - Dynamic teacher model update
    - Student 모델의 temporal ensemble 로 EMA 업데이트
      - Student 모델과 static teacher 모델의 weight 교환으로 인한 시간적 ensemble 해석 가능
    - 이때 static teacher 모델은 weight 고정
    - Consensus mechanism 을 통해 pseudo-label 의 품질 향상



EMA: Exponential Moving Average

$$\Theta_{DT} \leftarrow \alpha\Theta'_{DT} + (1-\alpha)\Theta_S$$

Inner-period training with consensus mechanism

- Consensus mechanism
  - Filering
    - Teacher model 의 noise 를 극복하기 위해 threshold 0.5 를 통해 low confidence 예측을 pre-filter
  - Fusion
    - Weakly-augmentation target image x_t 에 대한 static / dynamic teacher 의 prediction을 아래와 같이 정의

$$Y_{ST} = \{(b^i_{ST}, c^i_{ST}, y^i_{ST})\}^n_{i=0}$$

$$Y_{DT} = \{(b^j_{DT}, c^j_{DT}, y^j_{DT})\}^m_{j=0}$$

b: bounding box coordinates
c: classification confidence
y: category label of each predicted object

n, m: # of predicted objects of static/dynamic teacher

    - 동일 class 에 속하고 static/dynamic teacher 의 prediction 간에 높은 IOU 를 갖는 object 선택

$$IOU(b^i_{ST}, b^j_{DT}) \geq \eta \quad \& \quad y^i_{ST} = y^j_{DT}$$

Inner-period training with consensus mechanism

- Consensus mechanism

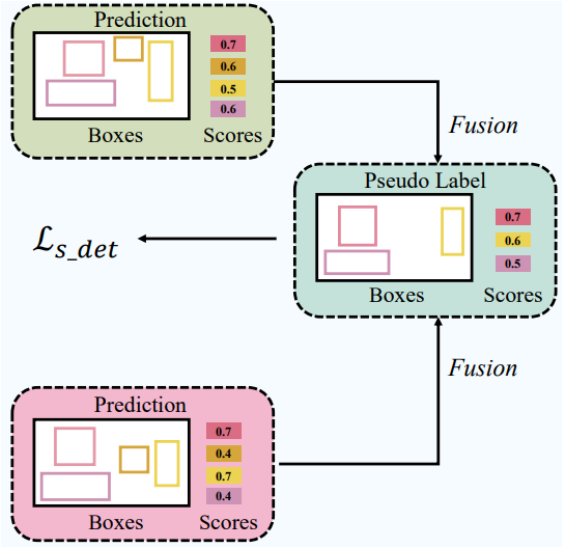
  - Fusion

    - Bounding box 결정

$$\widetilde{b} = \frac{1}{C}\left(\sum_{i=1}^{N} c_{ST}^{i} * b_{ST}^{i} + \sum_{j=1}^{M} c_{DT}^{j} * b_{DT}^{j}\right),$$

    - Classification confidence 결정

$$\widetilde{c} = \frac{\beta}{N}\sum_{i=1}^{N} c_{ST}^{i} + \frac{1-\beta}{M}\sum_{j=1}^{M} c_{DT}^{j},$$



    - 위의 과정을 통해 refine한 pseudo label 을 통해 student model supervision

$$\widetilde{Y} = \{(\widetilde{b}, \widetilde{c}, \widetilde{y})\}$$

$$\mathcal{L}_{s\_det} = \sum_{\bar{x}_t \in \mathcal{X}_T} \mathcal{L}_{cls}^{RPN}(\Theta_S(\bar{x}_t), \widetilde{y}) + \mathcal{L}_{reg}^{RPN}(\Theta_S(\bar{x}_t), \widetilde{b}) + \quad \leftarrow \text{RPN loss}$$

$$\mathcal{L}_{cls}^{ROI}(\Theta_S(\bar{x}_t), \widetilde{y}) + \mathcal{L}_{reg}^{ROI}(\Theta_S(\bar{x}_t), \widetilde{b}), \quad \leftarrow \text{ROI loss}$$

Experiments

- Cityscapes → FoggyCityscapes 실험 결과

Cityscapes

FoggyCityscapes

	Methods	Person	Rider	Car	Truck	Bus	Train	Motor	Bicycle	mAP
	Source only (Single level)	23.4	23.8	29.7	8.1	12.9	5.0	18.3	24.5	18.2
	Source only (All levels)	35.1	39.4	47.0	10.7	32.5	10.1	30.0	36.9	30.7
UDAOD	MAF [13]	28.2	39.5	43.9	23.8	39.9	33.3	29.2	33.9	34.0
	SW-Faster [32]	32.3	42.2	47.3	23.7	41.3	27.8	28.3	35.4	34.8
	iFAN [52]	32.6	40.0	48.5	27.9	45.5	31.7	22.8	33.0	35.3
	CR-DA-DET [44]	32.9	43.8	49.2	27.2	45.1	36.4	30.3	34.6	37.4
	AT-Faster [14]	34.6	47.0	50.0	23.7	43.3	**38.7**	33.4	38.8	38.7
SFOD	SED(Mosaic) [25]	33.2	40.7	44.5	25.5	39.0	22.2	28.4	34.1	33.5
	HCL [17]	26.9	46.0	41.3	**33.0**	25.0	28.1	35.9	40.7	34.6
	A^2SFOD [8]	32.3	44.1	44.6	28.1	34.3	29.0	31.8	38.9	35.4
	SOAP [43]	35.9	45.0	48.4	23.9	37.2	24.3	31.8	37.9	35.5
	LODS [24]	34.0	45.7	48.8	27.3	39.7	19.6	33.2	37.8	35.8
	IRG [39]	37.4	45.2	51.9	24.4	39.6	25.2	31.5	41.6	37.1
	Ours (Single level)	42.0	48.7	56.3	19.3	39.3	5.5	34.2	41.6	35.9
	Ours (All levels)	**46.1**	**52.8**	**63.4**	21.8	**46.7**	5.5	**37.4**	**48.4**	**40.3**
	Oracle	51.3	57.5	70.2	30.9	60.5	26.9	40.0	50.4	48.5

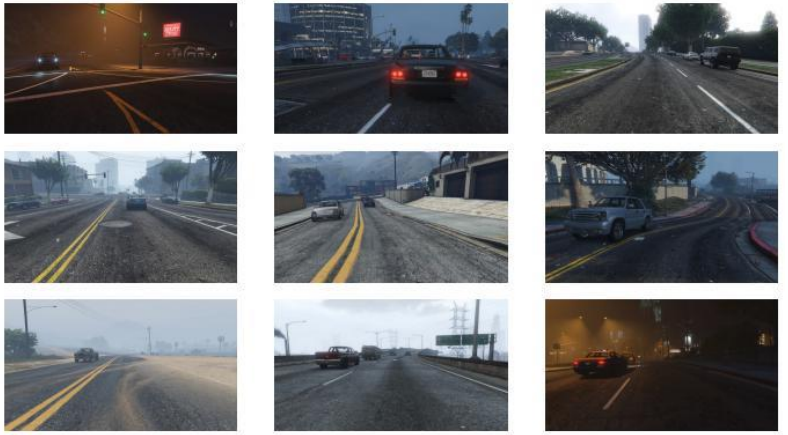
Table 1: Results of adaptation from normal to foggy weather (C2F). "Source only" and "Oracle" refer to the models trained by only using labeled source domain data and labeled target domain data, respectively.

Experiments

- Synthetic → Real 실험 결과

Methods	mAP	Methods	mAP
Source only	40.5	NL [19]	43.0
MAF [13]	41.1	UMT [10]	43.1
AT-Faster [14]	42.1	MeGA-CDA [38]	44.8
HTCN [2]	42.5	CR-DA-DET [44]	46.1
SED [25]	42.3	A^2SFOD [8]	44.0
SED(Mosaic) [25]	43.1	Ours	**57.8**
IRG [39]	43.2	Oracle	68.9

Table 4: Results of adaptation from synthetic to real scenes (S2C).



Synthetic (Sim10K)
video game GTA5

Real (Cityscapes)

Experiments

- Ablation study & training stability test

Foggy level	Method	DT	ST	mAP
All levels	Source only	-	-	30.7
	Single-teacher	-	✓	36.6
	Single-teacher	✓	-	38.0
	Ours	✓	✓	**40.3**
Single level	Source only	-	-	18.2
	Single-teacher	-	✓	27.2
	Single-teacher	✓	-	32.9
	Ours	✓	✓	**35.9**

Table 5: Results of single-teacher and multi-teacher methods on C2F benchmark. **DT** and **ST** represent the dynamic teacher and static teacher, respectively.

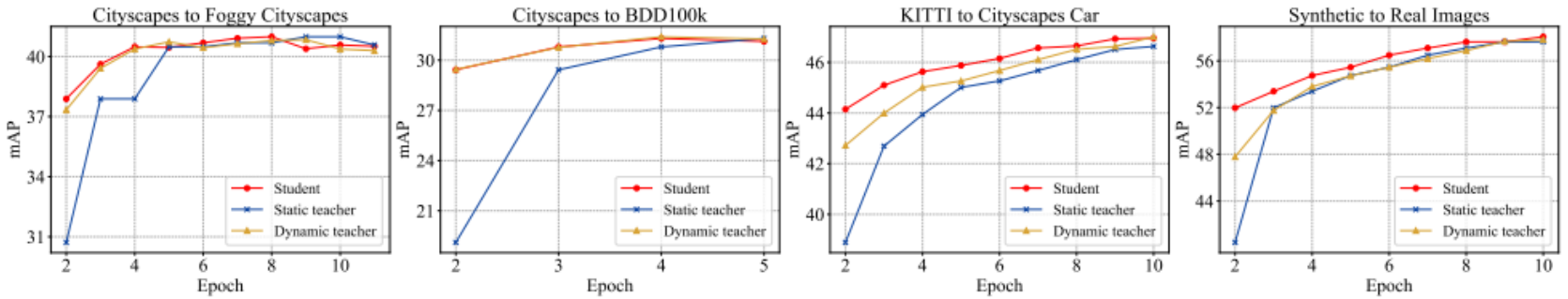


Figure 4: The training curves of each model within the multi-teacher framework during the whole training process.

감사합니다