Domain Adaptation with Auxiliary Target Domain-Oriented Classifier

Jian Liang [1], Dapeng Hu [1], and Jiashi Feng [2,1]

[1] National University of Singapore (NUS)

[2] Sea AI Lab (SAIL)

Outline

- Background
- Method
- Experiments

Background

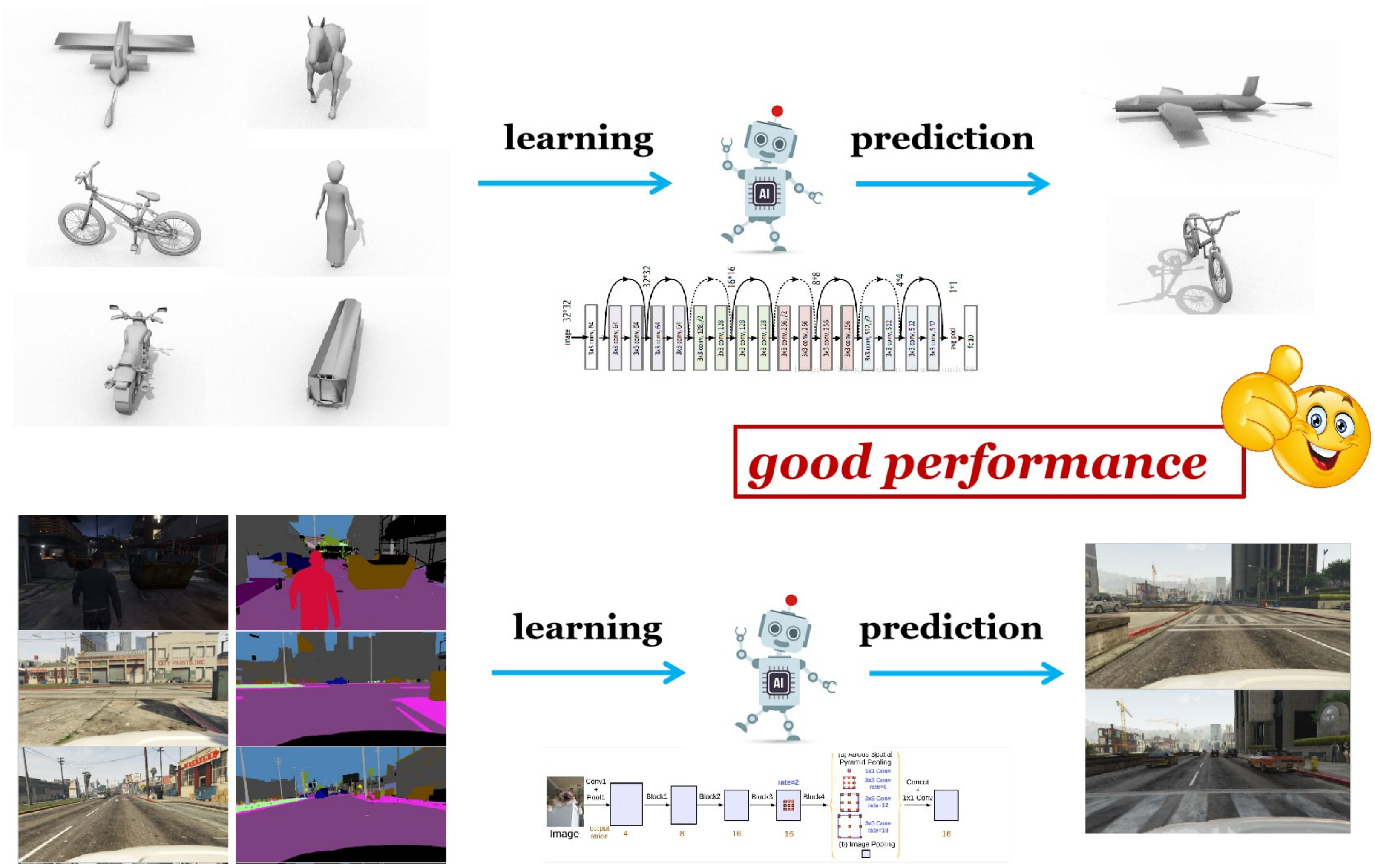


Figure 1: Test data comes from the **same distribution** as training data!

Background

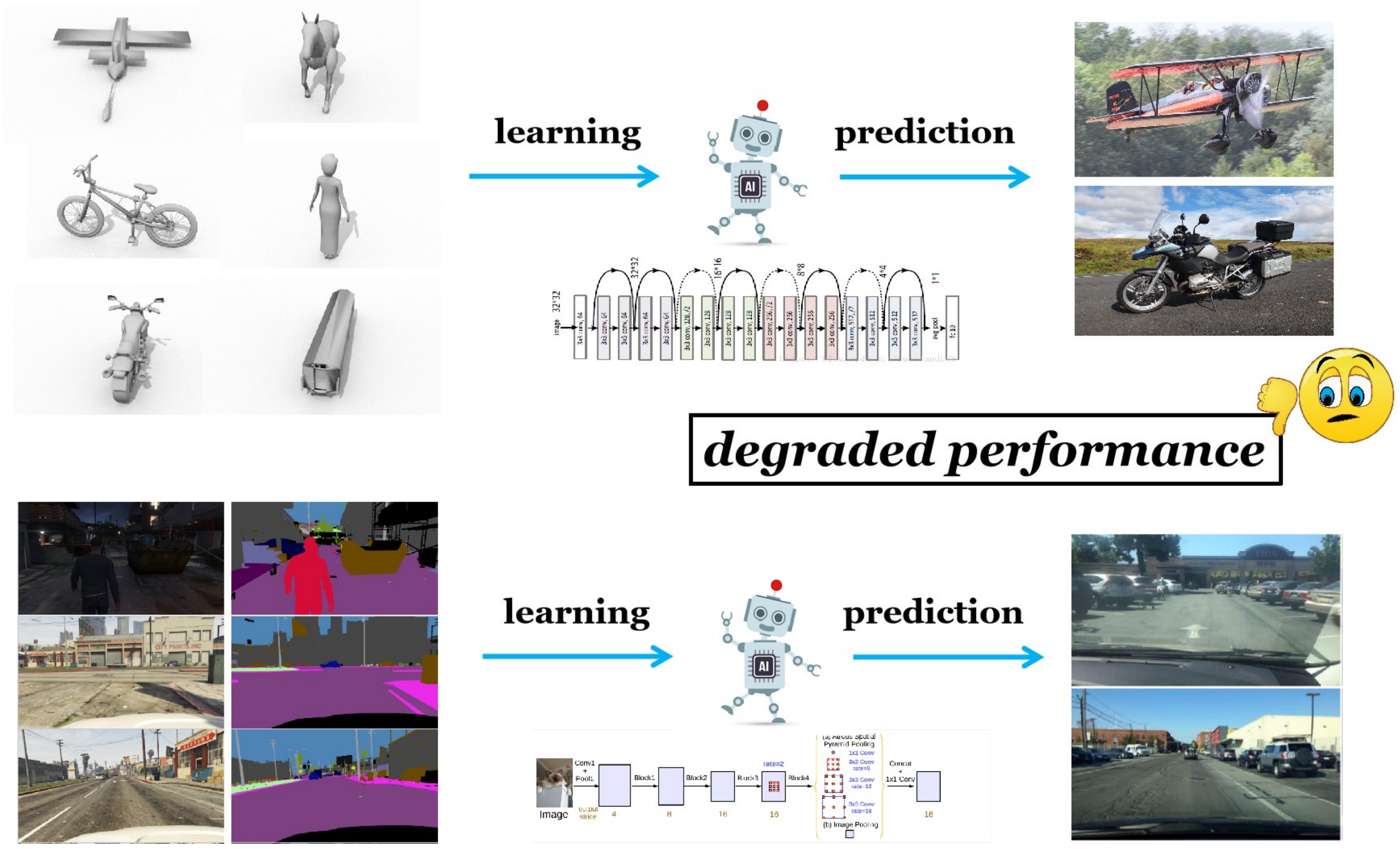
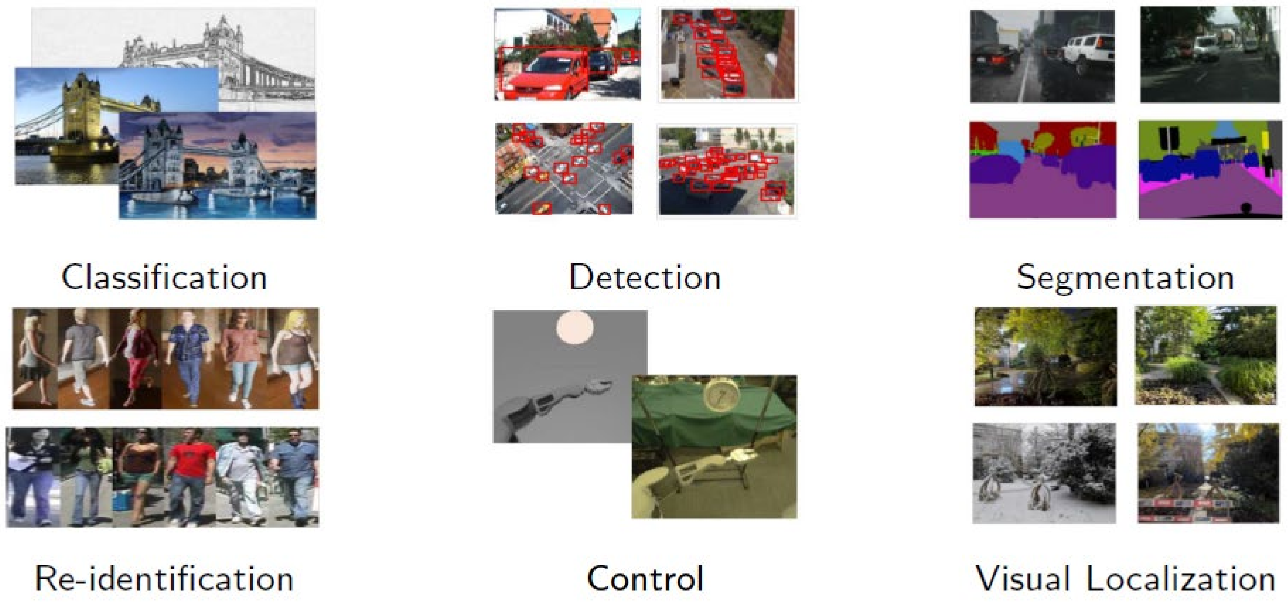


Figure 2: Test data and training data come from **different distributions**!

Background

- Unsupervised Domain Adaptation (DA)
  - **Source Domain** $\mathcal{D}_s$: n_s labeled samples $\{x_s^i, y_s^i\}_{i=1}^{n_s}$ from $P_S(X, Y)$;
  - **Target Domain** $\mathcal{D}_t$: n_t unlabeled samples $\{x_t^i, ?\}_{i=1}^{n_t}$ from $P_T(X, Y)$;
  - **Goal**: Use $\{x_t^i\}_{i=1}^{n_t}$ during training (transductive) and learn a good classifier to get the values of ? under **domain shift** (i.e., $P_S \neq P_T$).



Credit to Gabriela Csurka, TaskCV-2019 talk.

Background

Previous DA Methods - (I) **Input-level Pixel Transfer**

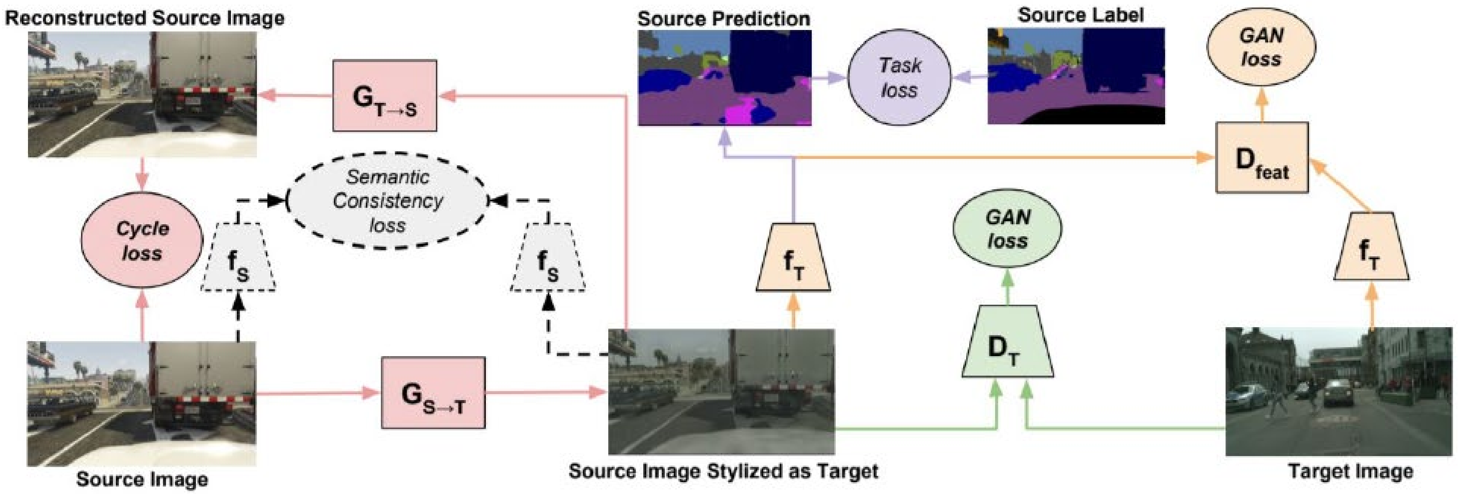


Figure 3: Cycle-consistent adversarial adaptation (CyCADA) [1] overview.

[1] Homan, Judy, et al. "CyCADA: Cycle-Consistent Adversarial Domain Adaptation." In ICML 2018.

Background

Previous DA Methods - (II) **Feature-level Alignment**

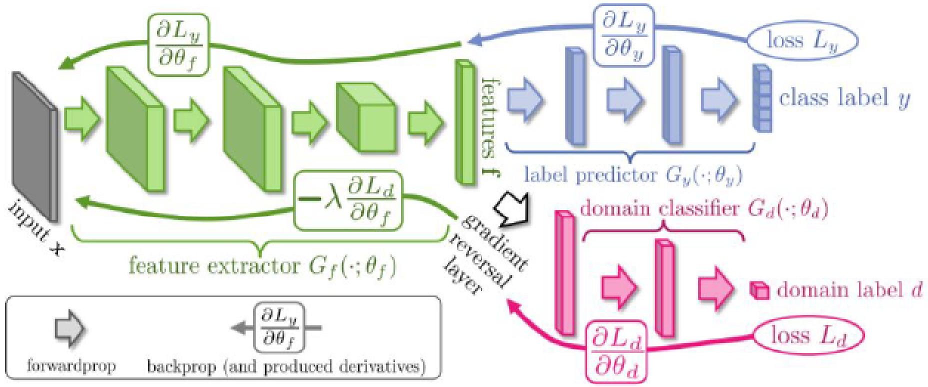


Figure 4: Unsupervised Domain Adaptation by Backpropagation (DANN) [2] overview.

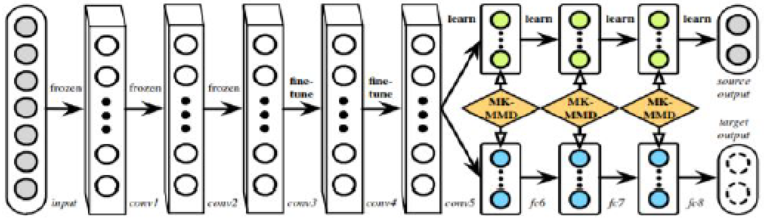


Figure 5: Deep Adaptation Networks (DAN) [3] overview.

[2] Ganin, Yaroslav, and Victor Lempitsky. "Unsupervised Domain Adaptation by Backpropagation." In ICML 2015.

[3] Long, Mingsheng, et al. "Learning Transferable Features with Deep Adaptation Networks." In ICML 2015.

Background

Previous DA Methods - (III) **Output-level Regularization**

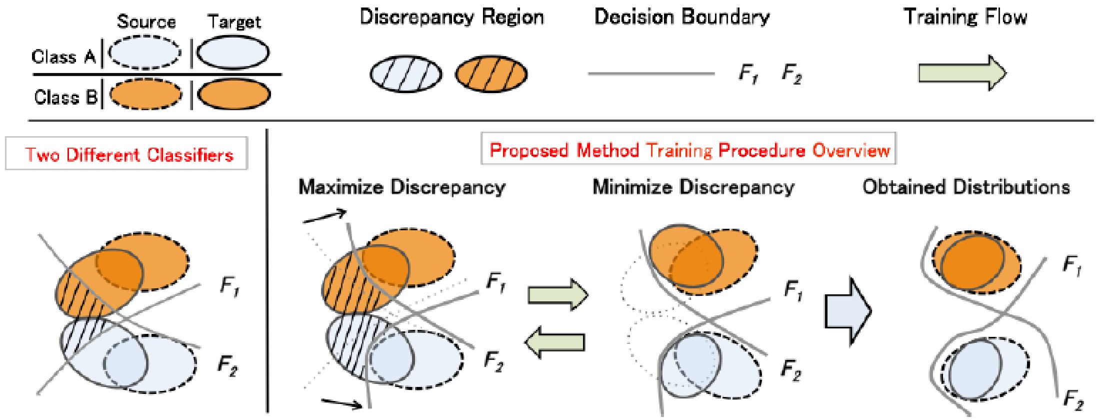


Figure 6: Maximum classifier discrepancy (MCD) [4] overview.

Or exploit the low-density separation principle:

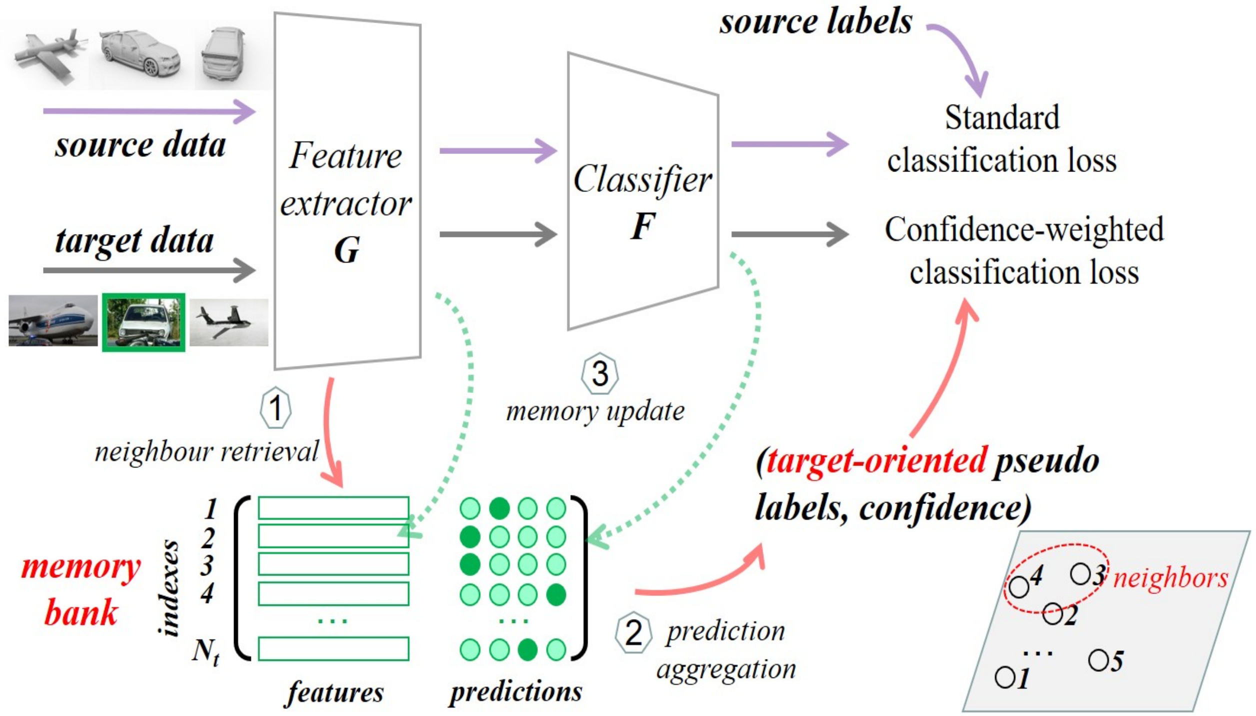
- entropy minimization
- pseudo-labeling / self-training
- virtual adversarial training
- consistency regularization

[4] Saito, Kuniaki, et al. "Maximum classifier discrepancy for unsupervised domain adaptation." In CVPR 2018.

Outline

- Background
- Method
- Experiments

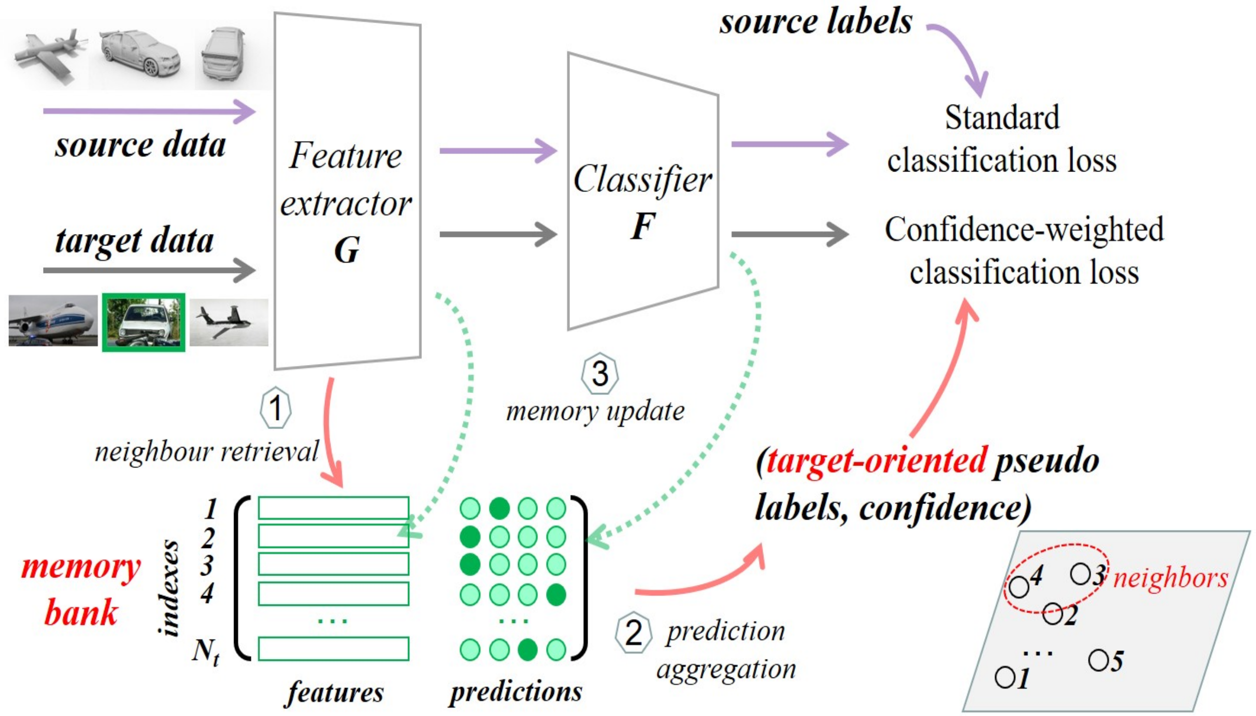
Method - Framework



✓ We argue that the degraded performance on the target domain mainly results from the bias learned in the classifier.

✓ To address this, we propose a new pseudo-labeling framework termed **Auxiliary Target Domain-Oriented Classifier (ATDOC)** for DA problems.

Method – Details of ATDOC-NC



✓ Pseudo labels are obtained via nearest centroid classifier (NC).

Updating centroids:

$$c_j = \sum\nolimits_{i \in B_t} \mathbb{1}_{[j=\hat{y}_i]} G(x_i^t) / \sum\nolimits_{i \in B_t} \mathbb{1}_{[j=\hat{y}_i]},$$

$$c_j^m = \gamma c_j + (1-\gamma) c_j^m, \; m = 1, 2, \cdots, K,$$

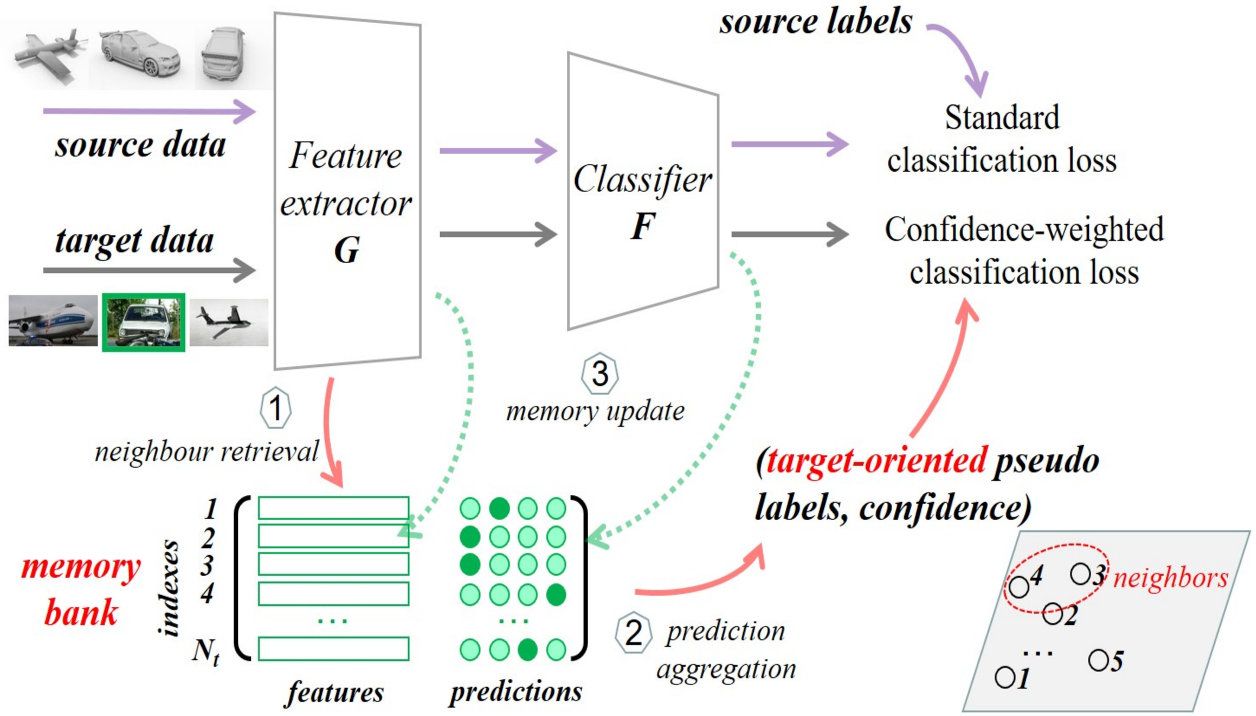
Obtaining pseudo labels:

$$\hat{y}_i = \arg\min\nolimits_{j=1}^{K} d\left(G(x_i^t), c_j^m\right), \; i = 1, 2, \cdots, N_t,$$

Loss function:

$$\mathcal{L}_{nc} = -\frac{\lambda}{N_{tu}} \sum\nolimits_{i=1}^{N_{tu}} \log p_{i,\hat{y}_i}.$$

Method – Details of ATDOC-NA



✓ Pseudo labels are obtained via neighborhood aggregation (NA).

Updating memory bank:

$$\breve{p}^m_{i,k} = p^2_{i,k} / \sum_i p^2_{i,k}.$$

Class balancing

Obtaining pseudo labels:

$$\hat{q}_i = \frac{1}{m} \sum_{j \neq i, j \in \mathcal{N}_i} \breve{p}_j$$

Loss function:

$$\mathcal{L}_{na} = -\frac{\lambda}{N_{tu}} \sum_{i=1}^{N_{tu}} \hat{q}_{i,\hat{y}_i} \log p_{i,\hat{y}_i}.$$

Outline

- Background
- Method
- Experiments

Experiments

- Results for Closed-set Unsupervised DA

Table 3. Accuracy (%) on Office-Home for closed-set UDA (ResNet-50).

Method	Ar→Cl	Ar→Pr	Ar→Re	Cl→Ar	Cl→Pr	Cl→Re	Pr→Ar	Pr→Cl	Pr→Re	Re→Ar	Re→Cl	Re→Pr	Avg.
ResNet-50 [25]	44.9	66.3	74.3	51.8	61.9	63.6	52.4	39.1	71.2	63.8	45.9	77.2	59.4
MinEnt [22]	51.0	71.9	77.1	61.2	69.1	70.1	59.3	48.7	77.0	70.4	53.0	81.0	65.8
BNM [16]	56.7	77.5	81.0	67.3	76.3	77.1	65.3	55.1	82.0	73.6	57.0	84.3	71.1
MCC [30]	56.3	77.3	80.3	67.0	77.1	77.0	66.2	55.1	81.2	73.5	57.4	84.1	71.0
Pseudo-labeling	54.1	74.1	78.4	63.3	72.8	74.0	61.7	51.0	78.9	71.9	56.6	81.9	68.2
ATDOC-NC	54.4	77.6	80.8	66.5	75.6	75.8	65.9	51.9	81.1	72.7	57.0	83.5	70.2
ATDOC-NA	58.3	**78.8**	**82.3**	**69.4**	*78.2*	**78.2**	67.1	56.0	*82.7*	72.0	58.2	*85.5*	*72.2*
CDAN+E [48]	54.6	74.1	78.1	63.0	72.2	74.1	61.6	52.3	79.1	72.3	57.3	82.8	68.5
+ BSP [9]	57.1	73.4	77.5	64.2	71.8	74.3	64.0	56.7	81.0	73.4	59.1	83.3	69.6
+ BNM [16]	58.1	77.2	81.1	67.5	75.3	77.2	65.5	*56.8*	82.6	74.1	*59.9*	84.6	71.7
+ MCC [30]	*58.9*	77.6	80.7	67.0	75.1	77.1	65.8	*56.8*	82.2	73.9	59.8	84.5	71.6
+ Pseudo-labeling	57.3	76.6	79.2	66.6	74.0	76.6	66.1	53.6	81.0	*74.3*	58.9	84.2	70.7
+ ATDOC-NC	55.9	76.3	80.3	63.8	75.7	76.4	63.9	53.7	81.7	71.6	57.7	83.3	70.0
+ ATDOC-NA	**60.2**	77.8	*82.2*	*68.5*	**78.6**	77.9	**68.4**	**58.4**	**83.1**	**74.8**	**61.5**	**87.2**	**73.2**
SAFN [76]	52.0	71.7	76.3	64.2	69.9	71.9	63.7	51.4	77.1	70.9	57.1	81.5	67.3
CADA-P [35]	56.9	76.4	80.7	61.3	75.2	75.2	63.2	54.5	80.7	73.9	**61.5**	84.1	70.2
DCAN [42]	54.5	75.7	81.2	67.4	74.0	76.3	*67.4*	52.7	80.6	74.1	59.1	83.5	70.5
SHOT [45]	57.1	*78.1*	81.5	68.0	*78.2*	*78.1*	*67.4*	54.9	82.2	73.3	58.8	84.3	71.8

Experiments

- Results for Semi-supervised DA

Table 4. Accuracy (%) on DomainNet-126 for Semi-supervised DA (SSDA) using a ResNet-34 backbone.

Method	C → S		P → C		P → R		R → C		R → P		R → S		S → P		Average	
	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot
ResNet-34 [25]	54.8	57.9	59.2	63.0	73.7	75.6	61.2	63.9	64.5	66.3	52.0	56.0	60.4	62.2	60.8	63.6
MinEnt [22]	56.3	61.5	67.7	71.2	76.0	78.1	66.1	71.6	68.9	70.4	60.0	63.5	62.9	66.0	65.4	68.9
MCC [30]	56.8	60.5	62.8	66.5	75.3	76.5	65.5	67.2	66.9	68.1	57.6	59.8	63.4	65.0	64.0	66.2
BNM [16]	58.4	62.6	69.4	72.7	77.0	79.5	69.8	73.7	69.8	71.2	61.4	65.1	64.1	67.6	67.1	70.3
Pseudo-labeling	62.5	64.5	67.6	70.7	78.3	79.3	70.9	72.9	69.2	70.7	62.0	64.8	67.0	68.6	68.2	70.2
ATDOC-NC	58.1	62.2	65.8	70.2	76.9	78.7	69.2	72.3	69.8	70.6	60.4	65.0	65.5	68.1	66.5	69.6
ATDOC-NA	**65.6**	**66.7**	*72.8*	74.2	**81.2**	**81.2**	**74.9**	**76.9**	**71.3**	**72.5**	*65.2*	64.6	**68.7**	**70.8**	**71.4**	**72.4**
MixMatch [3]	59.3	62.7	66.7	68.7	74.8	78.8	69.4	72.6	67.8	68.8	62.5	65.6	66.3	67.1	66.7	69.2
w/ Pseudo-labeling	59.6	62.6	67.5	69.6	74.8	78.6	70.0	73.0	68.6	69.3	63.2	65.9	66.6	67.3	67.2	69.5
w/ ATDOC-NC	60.2	63.4	65.2	69.5	75.0	78.9	68.4	73.0	68.7	70.1	60.9	64.5	65.3	67.1	66.2	69.5
w/ ATDOC-NA	*64.6*	*65.9*	70.7	72.2	*80.3*	*80.8*	*74.0*	75.2	70.2	71.2	**65.7**	*67.7*	*68.5*	*69.4*	*70.6*	*71.8*
MME [59]	56.3	61.8	69.0	71.7	76.1	78.5	70.0	72.2	67.7	69.7	61.0	61.9	64.8	66.8	66.4	68.9
BiAT [29]	57.9	61.5	71.6	*74.6*	77.0	78.6	73.0	74.9	68.0	68.8	58.5	62.1	63.9	67.5	67.1	69.7
Meta-MME [40]	-	62.8	-	72.8	-	79.2	-	73.5	-	70.3	-	63.8	-	68.0	-	70.1
APE [32]	56.7	63.1	**72.9**	**76.7**	76.6	79.4	70.4	*76.6*	*70.8*	*72.1*	63.0	*67.8*	64.5	66.1	67.6	71.7

Experiments

- Results for Partial-set Unsupervised DA

Table 5. Accuracy (%) on Office-Home for Partial-set UDA (PDA) using a ResNet-50 backbone.

Method	Ar→Cl	Ar→Pr	Ar→Re	Cl→Ar	Cl→Pr	Cl→Re	Pr→Ar	Pr→Cl	Pr→Re	Re→Ar	Re→Cl	Re→Pr	Avg.
ResNet-50 [25]	43.5	67.8	78.9	57.5	56.2	62.2	58.1	40.7	74.9	68.1	46.1	76.3	60.9
MinEnt [22]	45.7	73.3	81.6	64.6	66.2	73.0	66.0	52.4	78.7	74.8	56.7	80.8	67.8
MCC [30]	54.1	75.3	79.5	63.9	66.3	71.8	63.3	55.1	78.0	70.4	55.7	76.7	67.5
BNM [16]	54.6	77.2	81.1	64.9	67.9	72.8	62.6	55.7	79.4	70.5	54.7	77.6	68.2
Pseudo-labeling	51.9	70.7	77.5	61.7	62.4	67.8	62.9	54.1	73.8	70.4	56.7	75.0	65.4
ATDOC-NC	59.5	**80.3**	*83.8*	*71.8*	*71.6*	79.7	70.6	*59.4*	82.2	**78.4**	**61.1**	81.5	*73.3*
ATDOC-NA	*60.1*	76.9	**84.5**	**72.8**	71.2	**80.9**	**73.9**	**61.8**	83.8	77.3	*60.4*	80.4	**73.7**
ETN [6]	59.2	77.0	79.5	62.9	65.7	75.0	68.3	55.4	*84.4*	75.7	57.7	*84.5*	70.5
SAFN [76]	58.9	76.3	81.4	70.4	*73.0*	77.8	*72.4*	55.3	80.4	75.8	*60.4*	79.9	71.8
RTNet_{adv} [12]	**63.2**	*80.1*	80.7	66.7	69.3	77.2	71.6	53.9	**84.6**	*77.4*	57.9	**85.5**	72.3

Experiments

- Results for Semi-supervised Learning (SSL)

Table 6. Accuracy (%) on Office-Home and DomainNet-126 for scarce-labeled SSL (ResNet-50).

Dataset	Office-Home					DomainNet-126				
Method	Ar	Cl	Pr	Re	Avg.	C	P	R	S	Avg.
ResNet-50 [25]	48.7	42.1	68.9	66.6	56.6	41.6	46.2	66.4	33.3	46.9
MinEnt [22]	51.7	44.5	72.4	68.9	59.4	43.8	48.6	68.8	35.4	49.2
MCC [30]	58.9	*47.7*	*77.4*	74.3	64.6	45.7	49.5	70.9	38.5	51.2
BNM [16]	*59.0*	46.0	76.5	71.5	63.2	44.8	47.2	69.9	35.5	49.4
Pseudo-labeling	47.3	41.4	71.4	66.1	56.6	41.0	46.3	72.5	33.1	48.2
ATDOC-NC	56.0	43.4	76.6	72.9	62.2	45.6	51.5	72.2	35.2	51.1
ATDOC-NA	**59.1**	46.6	**78.4**	**75.9**	**65.0**	**54.7**	**60.0**	**75.5**	*38.6*	**57.2**
MixMatch [3]	52.2	41.9	73.1	69.1	59.1	41.2	38.7	64.3	34.2	44.6
w/ Pseudo-labeling	53.4	42.5	72.6	69.5	59.5	40.4	39.1	64.7	34.1	44.6
w/ ATDOC-NC	54.6	44.4	72.7	71.0	60.7	41.2	38.3	63.4	34.4	44.3
w/ ATDOC-NA	56.4	**48.3**	74.7	*75.2*	*63.6*	*49.9*	*51.4*	*72.8*	**40.8**	*53.7*

Summary

- ✓ we propose ATDOC, a new framework to combat classifier bias that **provides a new perspective of addressing domain shift**.
- ✓ we exploit the memory bank and develop two types of **non-parametric classifiers**, not involving complicated network architectures with extra parameters.
- ✓ despite its simplicity, ATDOC achieves **competitive or better results than prior state-of-the-arts under a variety of DA settings**, e.g., PDA and SSDA.
- ✓ we study an SSL setting with only a few annotated data points available and find ATDOC **performs better than other SSL techniques even without domain shift**.

Thanks for listening!

- If you require any further information, feel free to contact me.

  Email: liangjian92@gmail.com

- Code is available at https://github.com/tim-learn/ATDOC/.

paper

code