Prototypical Cross-domain Self-supervised Learning for Few-shot Unsupervised Domain Adaptation

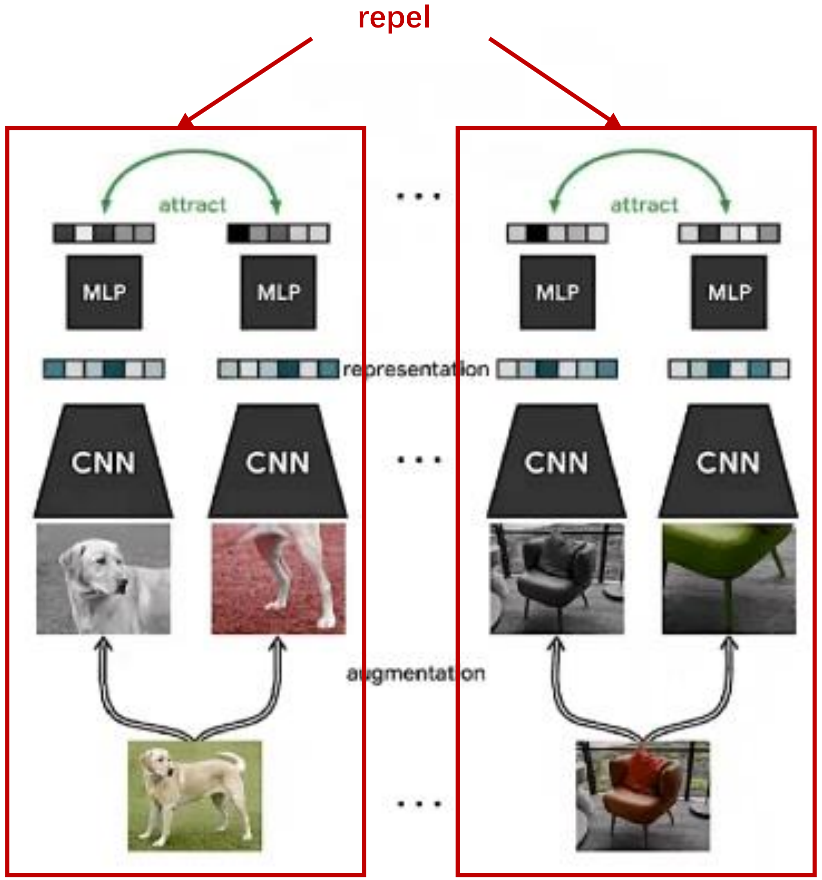
Xiangyu Yue[1,*] Zangwei Zheng[2,*] Shanghang Zhang[1] Yang Gao[3]
Trevor Darrell[1] Kurt Keutzer[1] Alberto Sangiovanni Vincentelli[1]
[1]UC Berkeley [2]Nanjing University [3]Tsinghua University

CVPR 2021

Background – Contrastive Learning

- Contrastive Learning

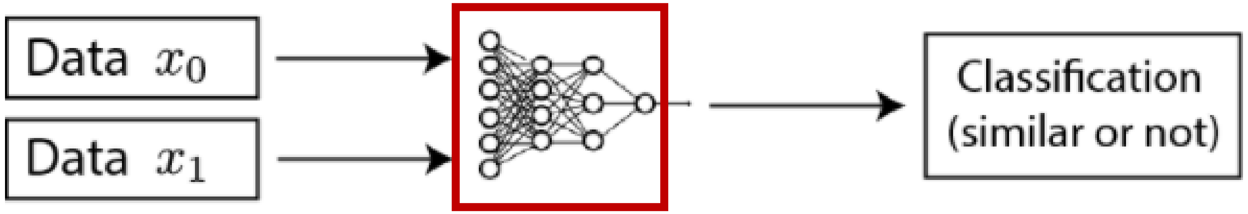
自监督学习能避免注释大型数据集带来的成本，即采用自定义pseudo-labels作为监督，并将学习的表示形式用于多个下游任务。



Background – Contrastive Learning

- Contrastive Learning

Contrastive



☐ Loss

$$L_N = -\mathbb{E}_X \left[\log \frac{\exp\left(f(x)^T f(x^+)\right)}{\exp\left(f(x)^T f(x^+)\right) + \sum_{j=1}^{\boxed{N-1}} \exp\left(f(x)^T f(x_j^-)\right)} \right]$$

✓ To learn an encoder f : $score(f(x), f(x^+)) >> score(f(x), f(x^-))$

Background – Contrastive Learning

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

- Domain Adaptation
  - 如何利用含有可利用监督信息的数据集（源域）向没有监督信息的数据集做知识迁移（目标域）
  - 源域和目标域有不同的数据分布，但是任务（label space or predict function) 相同

Target Data (X_t)

?

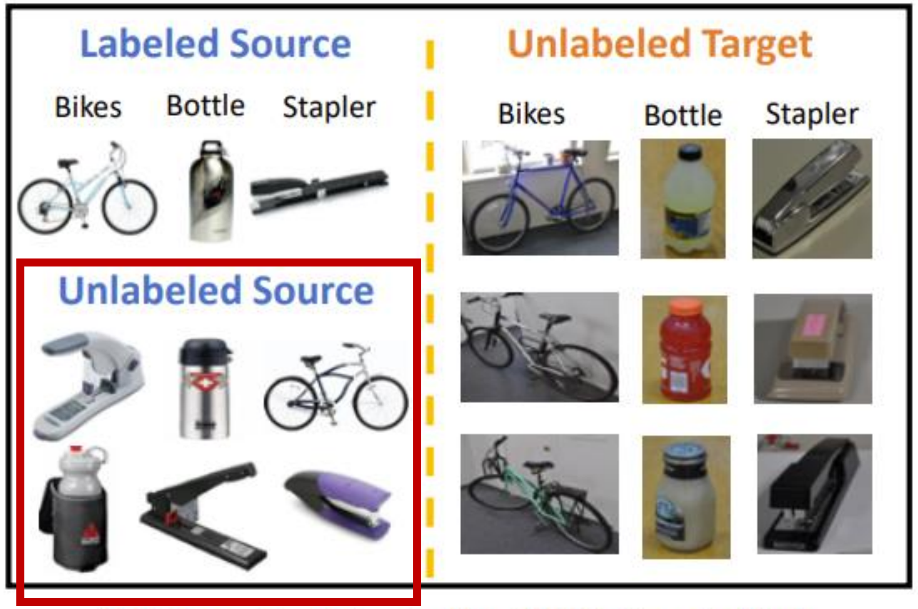
Model

Core: Distribution Alignment

Source Data (X_s , Y_s)

Setting

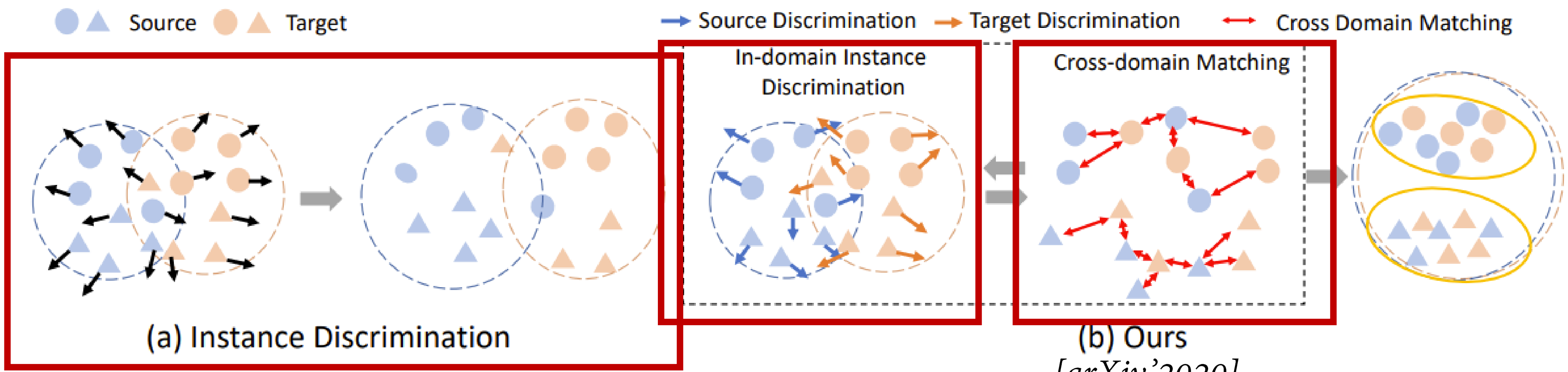
- Domain adaptation scenario with only a few source labels.



(a) Domain Adaptation With Few Labels

Method -- CDS

Pre-train
- ☐ In-domain Self-supervision
- ☐ Across-domain Self-supervision
- ☐ Domain Adaptation



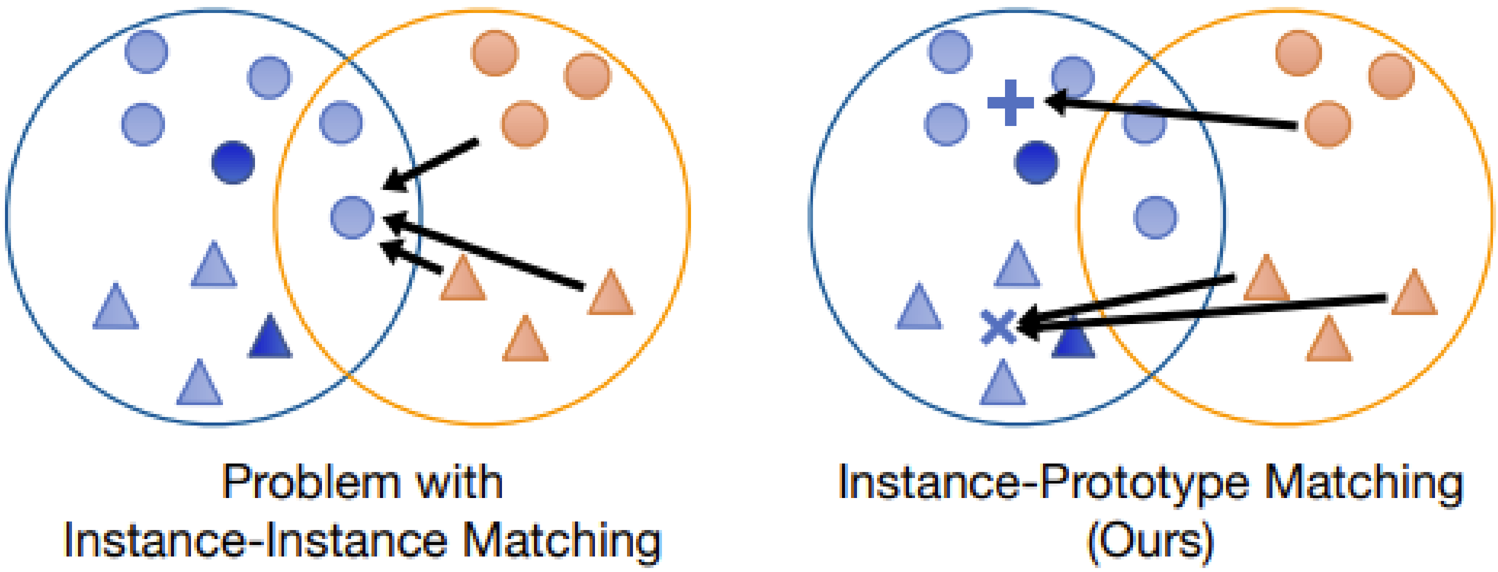
[arXiv'2020]

a) Instance discrimination distinguishes every feature from all the others without considering the domain gap, so that features of different domains are unlikely to be embedded close together.

b) In order to reduce the domain gap, CDS jointly uses in-domain instance discrimination and cross-domain alignment to learn features that are domain-invariant as well as discriminative

Kim D, Saito K, Oh T H, et al. Cross-domain self-supervised learning for domain adaptation with few source labels[J]. arXiv preprint arXiv:2003.08264, 2020.

Motivation



[arXiv'2020]

- **Left**: I-I incorrectly matches all orange samples to the same blue sample.
- **Right**: I-P robustly matches samples to the correct prototypes.

× 对于离群点比较敏感

× 没有对类别层次的语义结构进行建模，忽略了同类之间特征的语义相似性。

Kim D, Saito K, Oh T H, et al. Cross-domain self-supervised learning for domain adaptation with few source labels[J]. arXiv preprint arXiv:2003.08264, 2020.

Method -- PCS

- Structure （Instance to Prototype）

$$\mu_j^s = \frac{\mathbf{u}_j^s}{\|\mathbf{u}_j^s\|} \quad \mathbf{u}_j^s = \frac{1}{|C_j^{(s)}|}\sum_{\mathbf{v}_i^s \in C_j^{(s)}} \mathbf{v}_i^s$$

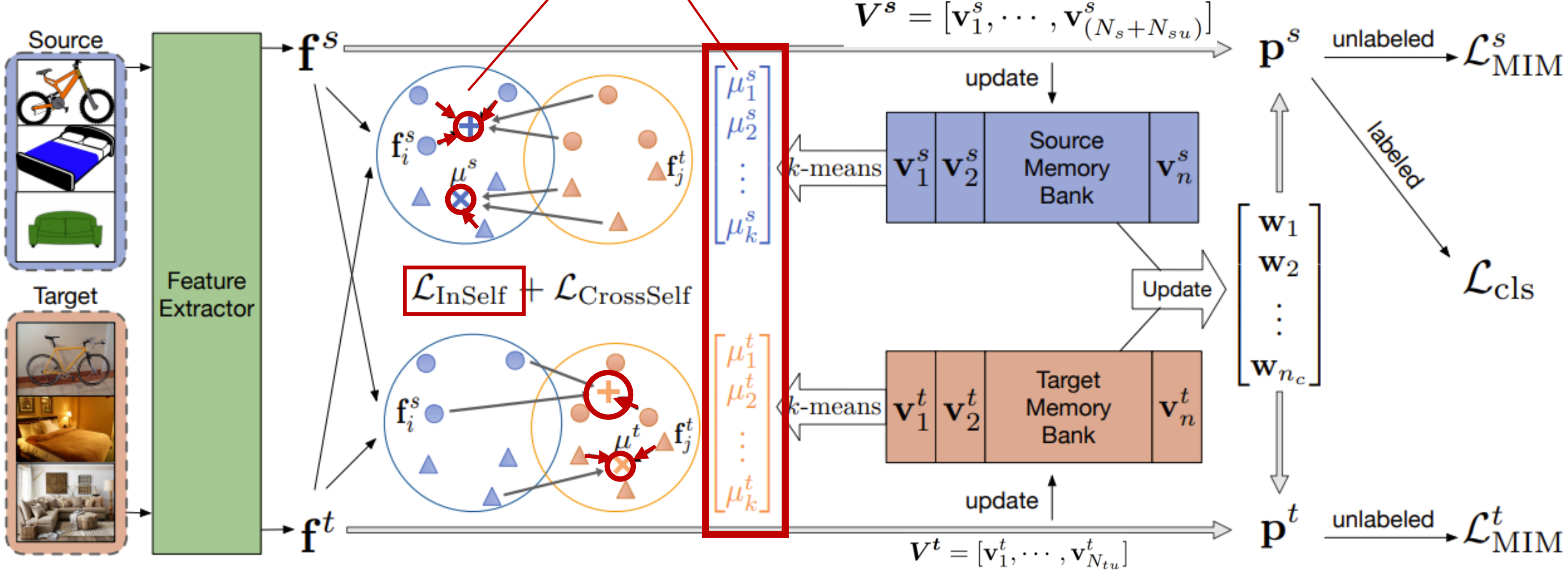


Figure 2: An overview of the PCS framework. In-domain and cross-domain self-supervision are performed between normalized feature vectors $\mathbf{f}$ and prototypes μ computed by clustering vectors $\mathbf{v}$ in memory banks. Features with confident predictions ($\mathbf{p}$) are used to adaptively update classifier vectors $\mathbf{w}$. MI maximization and classification loss are further used to extract discriminative features.

✓ In-domain Prototypical Contrastive Learning

$$P_{i,j}^s = \frac{\exp(\mu_j^s \cdot \mathbf{f}_i^s / \phi)}{\sum_{r=1}^k \exp(\mu_r^s \cdot \mathbf{f}_i^s / \phi)},$$

$$\mathcal{L}_{\text{PC}} = \sum_{i=1}^{N_s+N_{su}} \mathcal{L}_{CE}(P_i^s, c_s(i)) + \sum_{i=1}^{N_{tu}} \mathcal{L}_{CE}(P_i^t, c_t(i)) \quad (4)$$

$$\mathcal{L}_{\text{InSelf}} = \frac{1}{M}\sum_{m=1}^{M} \mathcal{L}_{\text{PC}}^{(m)}$$

$$P_i^s = [P_{i,1}^s, P_{i,2}^s, \ldots, P_{i,k}^s]$$

Method -- PCS

南京航空航天大學 NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

◻ Structure （Instance to Prototype）

$$\mu_j^s = \frac{\mathbf{u}_j^s}{\|\mathbf{u}_j^s\|} \quad \mathbf{u}_j^s = \frac{1}{|C_j^{(s)}|} \sum_{\mathbf{v}_i^s \in C_j^{(s)}} \mathbf{v}_i^s$$

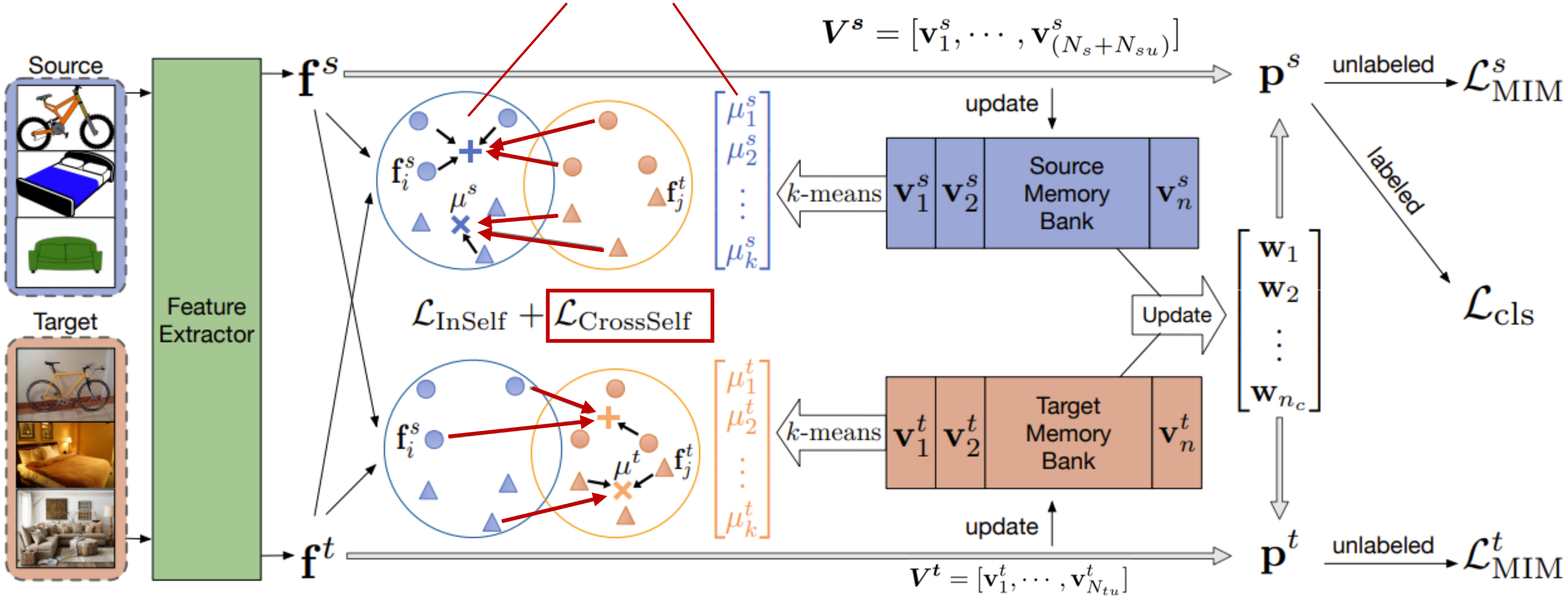


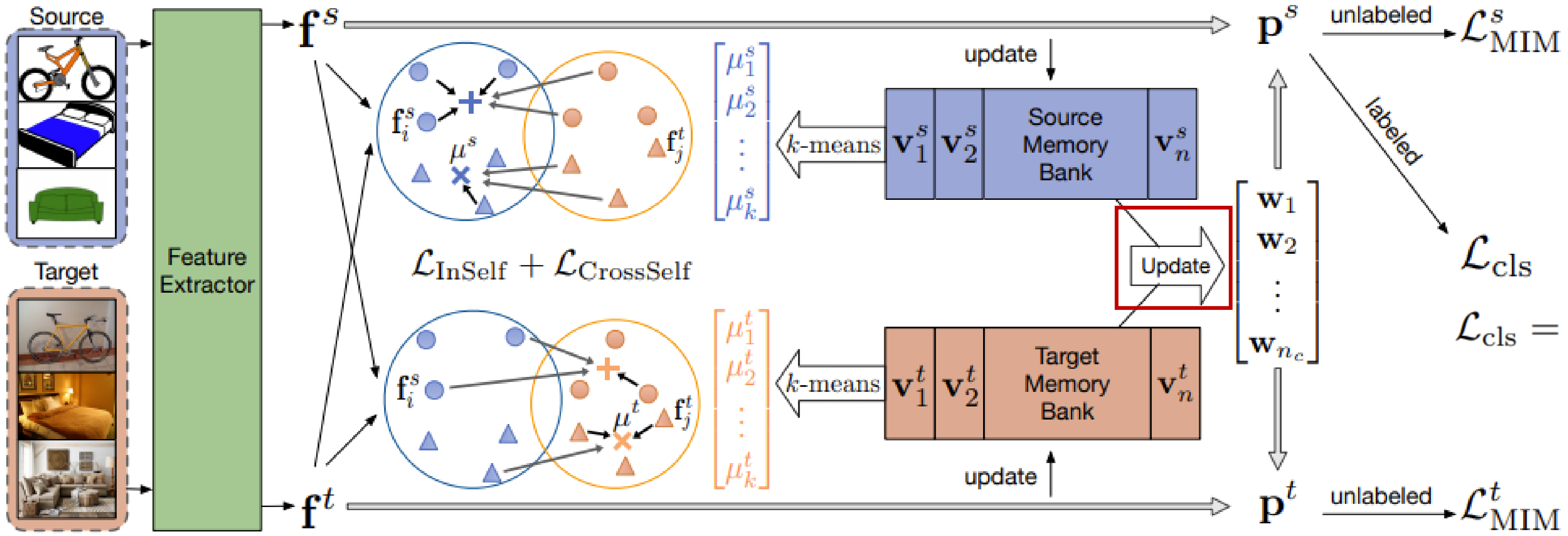
Figure 2: An overview of the PCS framework. In-domain and cross-domain self-supervision are performed between normalized feature vectors $\mathbf{f}$ and prototypes μ computed by clustering vectors $\mathbf{v}$ in memory banks. Features with confident predictions ($\mathbf{p}$) are used to adaptively update classifier vectors $\mathbf{w}$. MI maximization and classification loss are further used to extract discriminative features.

✓ **Cross-domain** Instance-Prototype SSL

$$P_{i,j}^{s\to t} = \frac{\exp(\mu_j^t \cdot \mathbf{f}_i^s/\tau)}{\sum_{r=1}^k \exp(\mu_r^t \cdot \mathbf{f}_i^s/\tau)} \quad \Big| \quad H(P_i^{s\to t}) = -\sum_{j=1}^{k} P_{i,j}^{s\to t} \log P_{i,j}^{s\to t}. \quad \Big| \quad \mathcal{L}_{\text{CrossSelf}} = \sum_{i=1}^{N_s+N_{su}} H(P_i^{s\to t}) + \sum_{i=1}^{N_{tu}} H(P_i^{t\to s})$$

Method -- PCS

Structure （Instance to Prototype）



$$\mathcal{L}_{\text{cls}} = \mathbb{E}_{(\mathbf{x},y)\in\mathcal{D}_s}\mathcal{L}_{CE}(\mathbf{p}(\mathbf{x}), y)$$

$$\mathcal{D}^{(i)}_{su} = \{\mathbf{x} | \mathbf{x} \in \mathcal{D}_{su}, \mathbf{p}(\mathbf{x})_i > t\}$$

- W(类别权重向量)应该对于某个类别的特征具有代表性。
- 使用few-shot labeled源域样本和**置信度高**的样本来估计每个类别的类别原型。

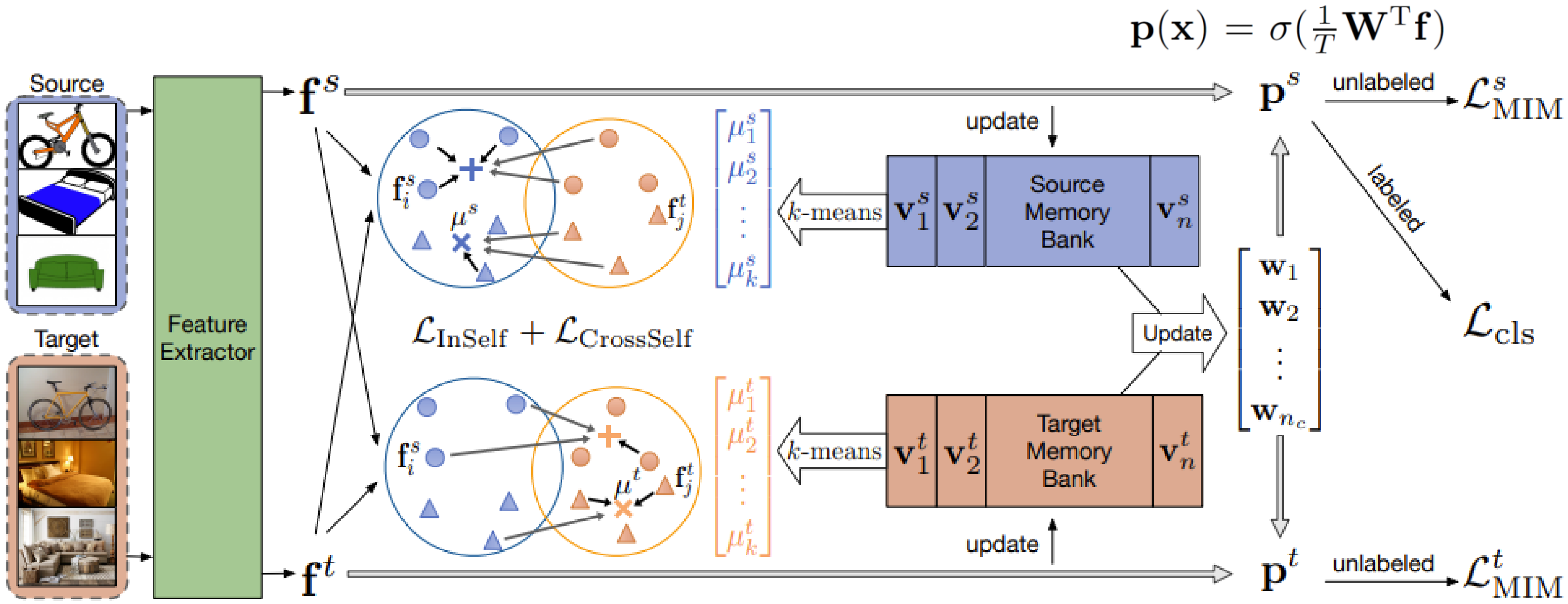
$$\begin{cases} \hat{\mathbf{w}}^s_i = \dfrac{1}{|\mathcal{D}^{(i)}_{s+}|} \displaystyle\sum_{\mathbf{x}\in\mathcal{D}^{(i)}_{s+}} \boldsymbol{V}^s(\mathbf{x}) \\ \hat{\mathbf{w}}^t_i = \dfrac{1}{|\mathcal{D}^{(i)}_{tu}|} \displaystyle\sum_{\mathbf{x}\in\mathcal{D}^{(i)}_{tu}} \boldsymbol{V}^t(\mathbf{x}) \end{cases}$$

✓ Adaptive Prototypical Classifier Update(APCU)

$$\mathbf{w}_i = \begin{cases} unit(\hat{\mathbf{w}}^s_i) & \text{if } |\mathcal{D}^{(i)}_{tu}| < t_w \\ unit(\hat{\mathbf{w}}^t_i) & \text{otherwise} \end{cases} \quad (11)$$

Method -- PCS

Structure (Instance to Prototype)



✓ Mutual Information Maximization

$$\mathcal{I}(y;\mathbf{x}) = \mathcal{H}(\mathbf{p}_0) - \mathbb{E}_{\mathbf{x}}[\mathcal{H}(p(y|\mathbf{x};\theta))], \quad (12)$$

$$\max \mathcal{H}(\mathbb{E}_{\mathbf{x}}[p(y|\mathbf{x};\theta)]) - \mathbb{E}_{\mathbf{x}}[\mathcal{H}(p(y|\mathbf{x};\theta))]$$

$$\mathcal{L}_{\mathrm{MIM}} = -\mathcal{I}(y;\mathbf{x}) \quad (13)$$

$$\mathcal{L}_{PCS} = \mathcal{L}_{\mathrm{cls}} + \lambda_{\mathrm{in}} \cdot \mathcal{L}_{\mathrm{InSelf}} + \lambda_{\mathrm{cross}} \cdot \mathcal{L}_{\mathrm{CrossSelf}} + \lambda_{\mathrm{mim}} \cdot \mathcal{L}_{\mathrm{MIM}} \quad (14)$$

Experiments

Table 1: Adaptation accuracy (%) comparison on 1-shot and 3-shots per class on the Office dataset.

Method	Office: Target Acc. on 1-shot / 3-shots						
	A→D	A→W	D→A	D→W	W→A	W→D	Avg
SO	27.5 / 49.2	28.7 / 46.3	40.9 / 55.3	65.2 / 85.5	41.1 / 53.8	62.0 / 86.1	44.2 / 62.7
MME [59]	21.5 / 51.0	12.2 / 54.6	23.1 / 60.2	60.9 / 89.7	14.0 / 52.3	62.4 / 91.4	32.3 / 66.5
CDAN [45]	11.2 / 43.7	6.2 / 50.1	9.1 / 65.1	54.8 / 91.6	10.4 / 57.0	41.6 / 89.8	22.2 / 66.2
SPL [71]	12.0 / 77.1	7.7 / 80.3	7.3 / 74.2	7.2 / 93.5	7.2 / 64.4	10.2 / 91.6	8.6 / 80.1
CAN [38]	25.3 / 48.6	26.4 / 45.3	23.9 / 41.2	69.4 / 78.2	21.2 / 39.3	67.3 / 82.3	38.9 / 55.8
MDDIA [35]	45.0 / 62.9	54.5 / 65.4	55.6 / 67.9	84.4 / 93.3	53.4 / 70.3	79.5 / 93.2	62.1 / 75.5
CDS [39]	33.3 / 57.0	35.2 / 58.6	52.0 / 67.6	59.0 / 86.0	46.5 / 65.7	57.4 / 81.3	47.2 / 69.3
DANN + ENT [18]	32.5 / 57.6	37.2 / 54.1	36.9 / 54.1	70.1 / 87.4	43.0 / 51.4	58.8 / 89.4	46.4 / 65.7
MME + ENT	37.6 / 69.5	42.5 / 68.3	48.6 / 66.7	73.5 / 89.8	47.2 / 63.2	62.4 / 95.4	52.0 / 74.1
CDAN + ENT	31.5 / 68.3	26.4 / 71.8	39.1 / 57.3	70.4 / 88.2	37.5 / 61.5	61.9 / 93.8	44.5 / 73.5
CDS + ENT	40.4 / 61.2	44.7 / 66.7	66.4 / 73.1	71.6 / 90.6	58.6 / 71.8	69.3 / 86.1	58.5 / 74.9
CDS + MME + ENT	39.4 / 61.6	43.6 / 66.3	66.0 / 74.5	75.7 / 92.1	53.1 / 73.0	70.9 / 90.6	58.5 / 76.3
CDS + CDAN + ENT	52.6 / 65.1	55.2 / 68.8	65.7 / 71.2	76.6 / 88.1	59.7 / 71.0	73.3 / 87.3	63.9 / 75.3
CDS / MME + ENT†	55.4 / 75.7	57.2 / 77.2	62.8 / 69.7	84.9 / 92.1	62.6 / 69.9	77.7 / 95.4	65.3 / 80.0
CDS / CDAN + ENT†	53.8 / 78.1	65.6 / 79.8	59.5 / 70.7	83.0 / 93.2	57.4 / 64.5	77.1 / 97.4	66.1 / 80.6
PCS (Ours)	**60.2 / 78.2**	**69.8 / 82.9**	**76.1 / 76.4**	**90.6 / 94.1**	**71.2 / 76.3**	**91.8** / 96.0	**76.6 / 84.0**
Improvement	+4.8 / +0.1	+4.2 / +3.1	+9.7 / +1.9	+5.7 / +0.9	+8.6 / +3.3	+14.1 / -1.4	+10.5 / +3.4

† Two-stage training results reported in [39].

ENT: Entropy Minimization

Experiments

Table 3: Adaptation accuracy (%) comparison on 3% and 6% labeled samples per class on the Office-Home dataset.

Method	Office-Home: Target Acc. (%)												
	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg
3% labeled source													
SO	24.4	38.3	43.1	26.4	34.7	33.7	27.5	26.5	42.6	41.2	29.0	52.3	35.0
MME [59]	4.5	15.4	25.0	28.7	34.1	37.0	25.6	25.4	44.9	39.3	29.0	52.0	30.1
CDAN [45]	5.0	8.4	11.8	20.6	26.1	27.5	26.6	27.0	40.3	38.7	25.5	44.9	25.2
MDDIA [35]	21.7	37.3	42.8	29.4	43.9	44.2	37.7	29.5	51.0	47.1	29.2	56.4	39.1
CAN [38]	17.1	30.5	33.2	22.5	34.5	36.0	18.5	19.4	41.3	28.7	18.6	43.2	28.6
CDS [39]	33.5	41.1	41.9	45.9	46.0	49.3	44.7	37.8	51.0	51.6	35.7	53.8	44.4
DANN + ENT [18]	19.5	30.2	38.1	18.1	21.8	24.2	31.6	23.5	48.1	40.7	28.1	50.2	31.2
MME + ENT	31.2	35.2	40.2	37.3	39.5	37.4	48.7	42.9	60.9	59.3	46.4	58.6	44.8
CDAN + ENT	20.6	31.4	41.2	20.6	24.9	30.6	33.5	26.5	56.7	46.9	29.5	48.4	34.2
CDS + ENT	39.2	46.1	47.8	49.9	50.7	54.1	48.0	43.5	59.3	58.6	44.3	59.3	50.1
CDS + MME + ENT	39.4	48.0	52.1	50.0	52.3	56.4	47.8	44.2	60.6	61.1	45.3	62.1	51.6
CDS + CDAN + ENT	43.8	55.5	60.2	51.5	56.4	59.6	51.3	46.4	64.5	62.2	52.4	70.2	56.2
CDS / MME + ENT†	41.7	49.4	57.8	51.8	52.3	55.9	54.3	46.2	69.0	65.6	52.2	68.2	55.4
CDS / CDAN + ENT†	37.7	49.2	56.5	49.8	51.9	55.9	50.0	42.3	68.1	63.1	48.7	67.5	53.4
PCS (Ours)	**42.1**	**61.5**	**63.9**	**52.3**	**61.5**	**61.4**	**58.0**	**47.6**	**73.9**	**66.0**	**52.5**	**75.6**	**59.7**
Improvement	-1.7	+6.0	+6.1	+3.7	+5.1	+1.8	+3.7	+1.2	+4.9	+0.4	+0.1	+5.4	+3.5
6% labeled source													
SO	28.7	45.7	51.2	31.9	39.8	44.1	37.6	30.8	54.6	49.9	36.0	61.8	42.7
MME [59]	27.6	43.2	49.5	41.1	46.6	49.5	43.7	30.5	61.3	54.9	37.3	66.8	46.0
CDAN [45]	26.2	33.7	44.5	34.8	42.9	44.7	42.9	36.0	59.3	54.9	40.1	63.6	43.6
MDDIA [35]	25.1	44.5	51.9	35.6	46.7	50.3	48.3	37.1	64.5	58.2	36.9	68.4	50.3
CAN [38]	20.4	34.7	44.7	29.0	40.4	38.6	33.3	21.1	53.4	36.8	19.1	58.0	35.8
CDS [39]	38.8	51.7	54.8	53.2	53.3	57.0	53.4	44.2	65.2	63.7	45.3	68.6	54.1
DANN + ENT [18]	22.4	32.9	43.5	23.2	30.9	33.3	33.2	26.9	54.6	46.8	32.7	55.1	36.3
MME + ENT	37.2	42.4	50.9	46.1	46.6	49.1	53.5	45.6	67.2	63.4	48.1	71.2	51.8
CDAN + ENT	23.1	35.5	49.2	26.1	39.2	43.8	44.7	33.8	61.7	55.1	34.7	67.9	42.9
CDS + ENT	42.9	55.5	59.5	55.2	55.1	59.1	54.3	46.9	68.1	65.7	50.6	71.5	57.0
CDS + MME + ENT	41.7	58.1	61.7	55.7	56.2	61.3	54.6	47.3	68.6	66.4	50.3	72.1	57.8
CDS + CDAN + ENT	45.4	60.4	65.5	54.9	59.2	63.8	55.4	49.0	71.6	66.6	54.1	75.4	60.1
CDS / MME + ENT†	44.1	51.6	63.3	53.9	55.2	62.0	56.5	46.6	70.9	67.7	54.7	74.7	58.4
CDS / CDAN + ENT†	39.0	51.3	63.1	51.0	55.0	63.6	57.8	45.9	72.8	65.8	50.4	73.5	57.4
PCS (Ours)	**46.1**	**65.7**	**69.2**	**57.1**	**64.7**	**66.2**	**61.4**	**47.9**	**75.2**	67.0	53.9	**76.6**	**62.6**
Improvement	+0.7	+5.3	+3.7	+2.2	+5.5	+2.4	+3.6	-1.1	+2.4	-0.7	-0.8	+1.2	+2.5

† Two-stage training results reported in [39].

Experiments

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Table 4: Adaptation accuracy (%) comparison on 0.1% and 1% labeled samples per class on the VisDA-2017 dataset.

Method	VisDA: Target Acc. (%)	
	0.1% Labeled	1% Labeled
SO	47.9	51.4
MME [59]	55.6	69.4
CDAN [45]	58.0	61.5
MDDIA [35]	68.9	71.3
CAN [38]	51.3	57.2
CDS [39]	34.2	67.5
DANN + ENT [18]	44.5	50.2
MME + ENT	54.0	66.1
CDAN + ENT	57.7	58.1
CDS + ENT	49.8	75.3
CDS + ENT + MME	60.0	<u>78.3</u>
CDS / MME + ENT†	62.5	69.4
CDS / CDAN + ENT†	<u>69.0</u>	69.1
PCS (Ours)	**78.0**	**79.0**
Improvement	+9.0	+0.7

† Two-stage training results reported in [39].

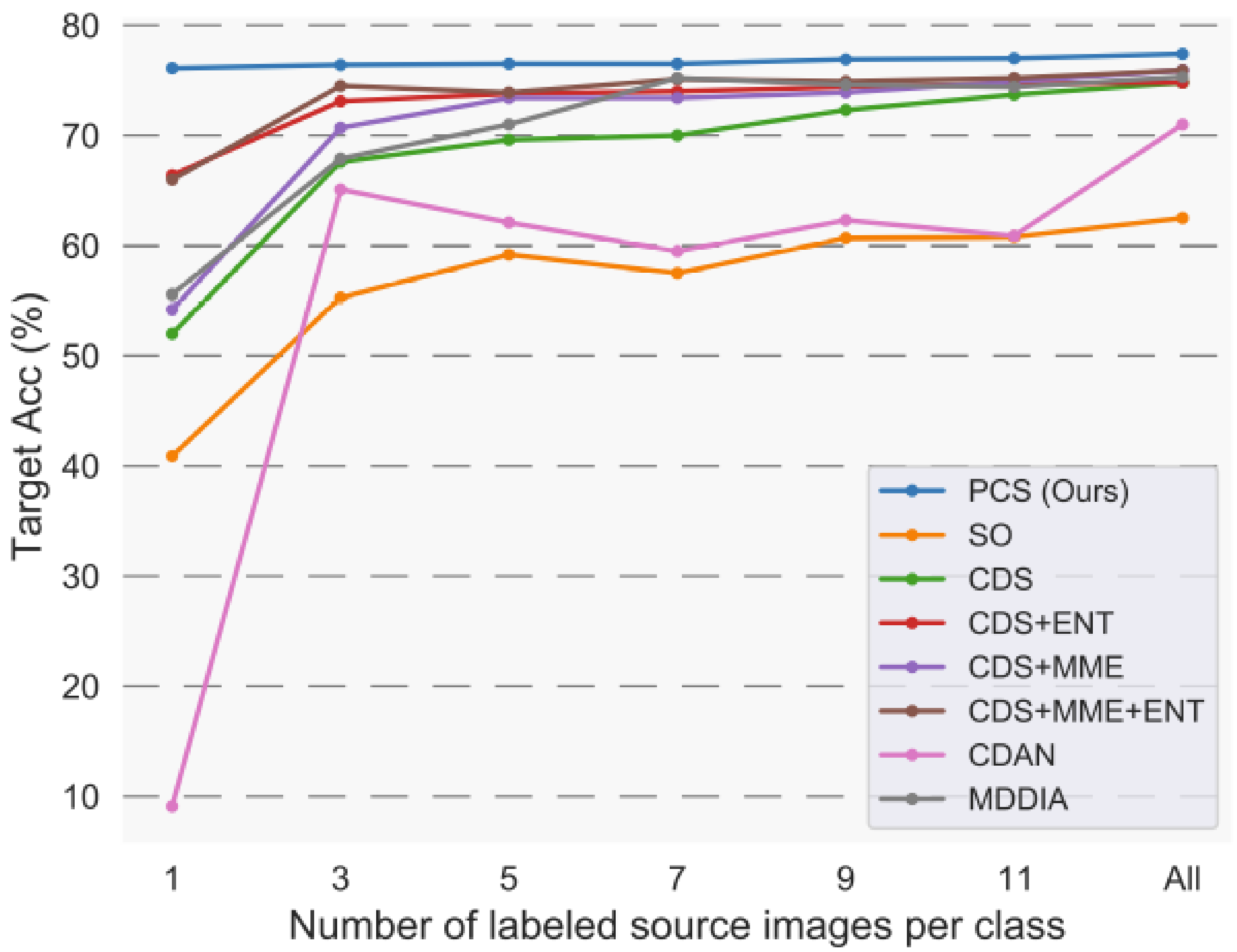
Experiments

Table 5: Adaptation accuracy (%) comparison on 1-shot and 3-shots per class on the DomainNet dataset.

Method	DomainNet: Target Acc. (%)							
	R→C	R→P	R→S	P→C	P→R	C→S	S→P	Avg
1-shot labeled source								
SO	18.4	30.6	16.7	16.2	28.9	12.7	10.5	19.1
MME [59]	13.8	29.2	9.7	16.0	26.0	13.4	14.4	17.5
CDAN [45]	16.0	25.7	12.9	12.6	19.5	7.2	8.0	14.6
MDDIA [35]	18.0	30.6	15.9	15.4	27.4	9.3	10.2	18.1
CAN [38]	18.3	22.1	16.7	13.2	23.9	11.1	12.1	16.8
CDS [39]	16.7	24.4	11.1	14.1	15.9	13.4	19.0	16.4
CDS + ENT	21.7	30.1	18.2	17.4	20.5	18.6	22.7	21.5
CDS + MME + ENT	21.2	28.8	15.5	15.8	17.6	19.0	20.7	19.8
PCS (Ours)	**39.0**	**51.7**	**39.8**	**26.4**	**38.8**	**23.7**	**23.6**	**34.7**
Improvement	**+17.3**	**+21.1**	**+21.6**	**+9.0**	**+9.9**	**+4.7**	**+0.9**	**+13.2**
3-shots labeled source								
SO	30.2	44.2	25.7	24.6	49.8	24.2	23.2	31.7
MME [59]	22.8	46.5	14.5	25.1	50.0	20.1	24.9	29.1
CDAN [45]	30.0	40.1	21.7	21.4	40.8	17.1	19.7	27.3
MDDIA [35]	41.4	50.7	37.4	31.4	52.9	23.1	24.1	37.3
CAN [38]	28.1	33.5	25	24.7	46.9	23.3	20.1	28.8
CDS [39]	35.0	43.8	36.7	34.1	36.8	31.1	34.5	36.0
CDS + ENT	44.5	52.2	40.9	40.0	47.2	33.0	40.1	42.5
CDS + MME + ENT	43.8	54.9	41.1	38.9	45.9	32.8	38.7	42.3
PCS (Ours)	**45.2**	**59.1**	**41.9**	**41.0**	**66.6**	31.9	37.4	**46.1**
Improvement	**+0.7**	**+6.9**	**+0.8**	**+1.0**	**+13.7**	-0.9	-2.7	**+3.6**

Experiments

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS



Experiments

Ablation Study

Table 2: Performance contribution of each part in PCS framework on Office.

Method	Office: Target Acc. on 1-shot / 3-shots						
	A→D	A→W	D→A	D→W	W→A	W→D	Avg
$\mathcal{L}_{\text{cls}}$	27.5 / 49.2	28.7 / 46.3	40.9 / 55.3	65.2 / 85.5	41.1 / 53.8	62.0 / 86.1	44.2 / 62.7
$+\mathcal{L}_{\text{InSelf}}$	39.0 / 55.6	38.6 / 55.1	47.2 / 68.5	71.7 / 89.4	50.9 / 68.4	65.1 / 90.6	52.1 / 71.3
$+\mathcal{L}_{\text{CrossSelf}}$	47.2 / 71.1	52.7 / 70.6	59.0 / 75.5	76.4 / 90.3	58.5 / 74.1	66.9 / 91.8	60.1 / 78.9
$+\mathcal{L}_{\text{MIM}}$	52.8 / 73.5	57.5 / 71.2	67.2 / 76.3	78.9 / 91.4	64.2 / 74.3	68.7 / 92.2	64.9 / 79.8
+APCU (PCS)	**60.2 / 78.2**	**69.8 / 82.9**	<u>76.1</u> / **76.4**	**90.6 / 94.1**	**71.2 / 76.3**	**91.8 / 96.0**	**76.6 / 84.0**
PCS w/o MIM	<u>59.0</u> / <u>75.9</u>	<u>58.6</u> / <u>76.5</u>	**76.2 / 76.4**	<u>87.8</u> / <u>93.2</u>	<u>68.7</u> / <u>74.7</u>	<u>89.8</u> / <u>95.0</u>	<u>73.5</u> / <u>82.0</u>

Thanks