山东大学 人工智能研究中心

TIME 机器学习与数据挖掘实验室

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

MHPL: Minimum Happy Points Learning For Active Source Free Domain Adaptation

THU-AM-335

CVPR 2023

Fan Wang Zhongyi Han Zhiyan Zhang Rundong He Yilong Yin*

*corresponding author

学无止境 气有浩然

Abstract

山东大学 人工智能研究中心

- The **Topic** of this paper is Active Source Free Domain adaptation, which aims to produce remarkable performance gains when a small set of informative target samples labeled by experts.

- In this paper, we **find** the best informative samples are Minimum Happy Points that satisfy the neighbor-chaotic, individual-difference and source-dissimilar.

- We **propose a novel MHPL** framework to explore and exploit the MH points with the neighbor environment uncertainty, neighbor diversity relaxation, one-shot querying, and neighbor focal loss.

- Extensive experiments verify that MHPL significantly surpasses state-of-the-art active learning strategies and existing ASFFA methods.

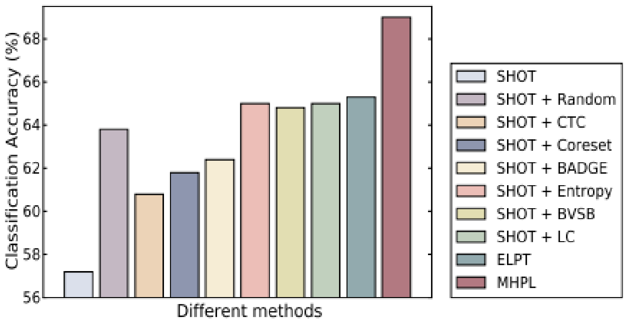
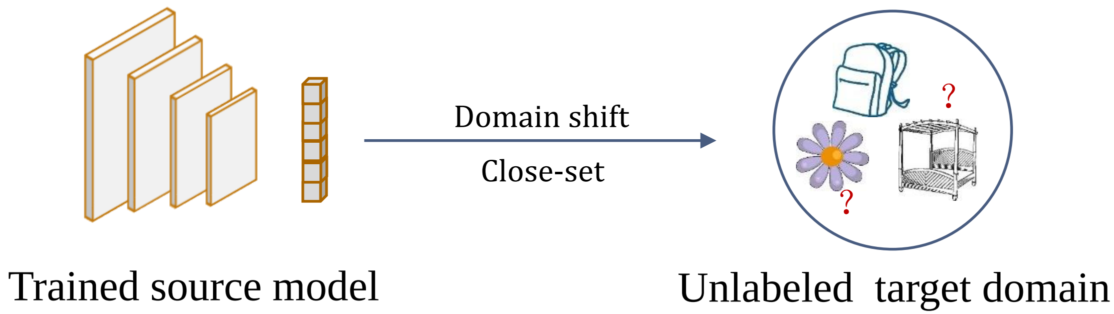


Figure 2. The comparison of ASFDA baselines (SHOT [25] + *, * denotes the active strategy), ELPT, and our MHPL with 5% active labeled target samples on Ar→Cl in the Office-Home.

Introduction

■ **Source Free Domain Adaptation**



Motivation of ASFDA

The SFDA setting faces a **performance bottleneck** with limited performance improvements due to the absence of source data and target supervised information. **Active source free domain adaptation (ASFDA)** can produce remarkable performance gains and breakthrough performance bottlenecks when a small set of informative target samples labeled by experts.

Introduction

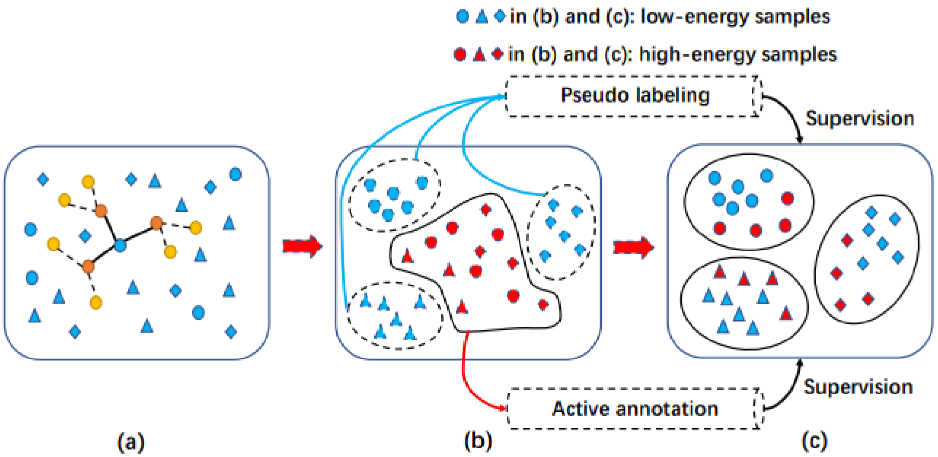
山东大学 人工智能研究中心

■ **Active Source Free Domain Adaptation (ASFDA)**

Two factors must be considered to achieve significant performance gains in ASFDA:

- ✓ Exploring samples that, once labeled, will improve accuracy significantly;
- ✓ Exploiting limited active labeled target data well in adaptation.

However, these two factors cannot be achieved by existing method, ELPT[1]:



✓The prediction uncertainty (entropy) is **error-prone** due to the miscalibrated source model.

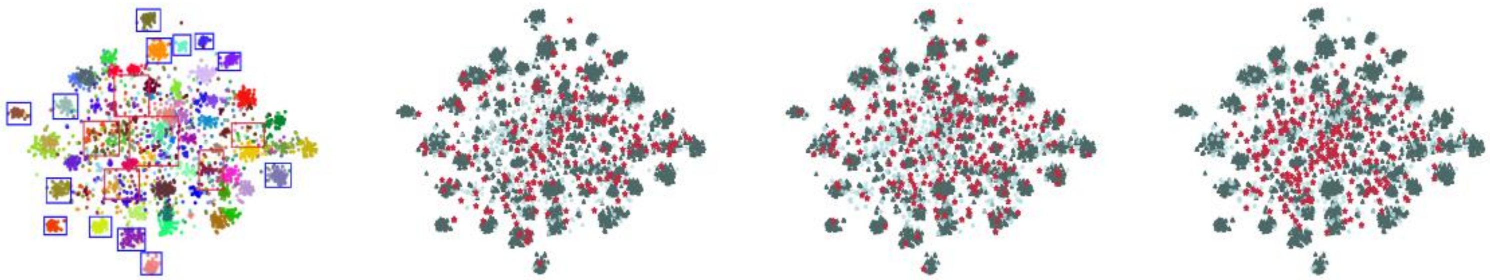
✓The **pseudo-label noise** of unlabeled samples easily influence the effect of standard cross-entropy loss on active samples.

- Xinyao Li, et al, Source-Free Active Domain Adaptation via Energy-Based Locality Preserving Transfer, ACM Multimedia 2022: 5802-5810

Minimum Happy (MH) Points

In this paper, we find the best samples for ASFDA are Minimum Happy (MH) points that satisfy the following three characteristics:

- **Neighbor-chaotic** samples refer to uncertain ones with label-chaotic neighbors. Annotating these samples can guide the learning of their confusing neighbors, bringing significant performance gains.
- **Individual-different** samples refer to diverse ones dissimilar to each other in selected uncertain samples.
- **Source-dissimilar** samples refer to those that are biased toward the target distribution.



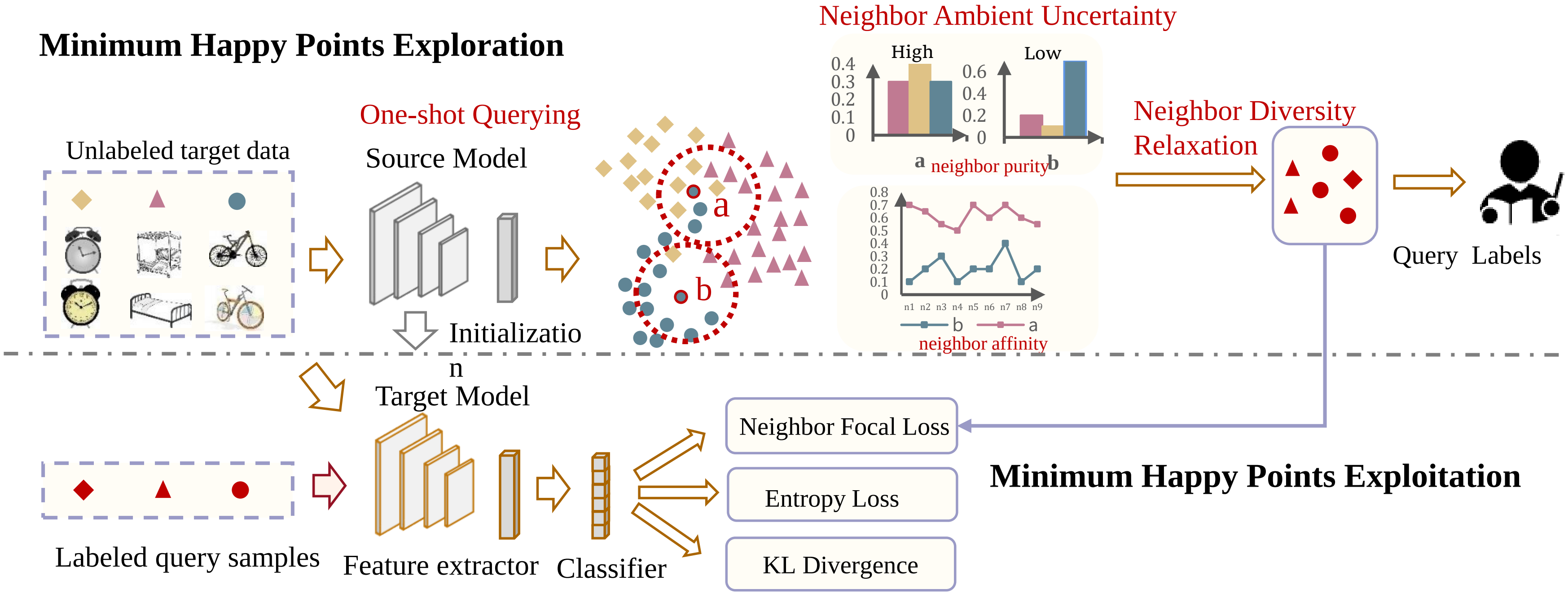
(a) Initial pseudo-labels (b) BVSB [14] (c) CoreSet [36] (d) MHPL

Figure 1. Feature visualization for the source model with 5% actively labeled target data on the Cl→Pr task. Different colors in (a) represent different classes of pseudo-labels by clustering. Blue blocks include easily-adaptive source-similar samples with label-clean neighbors that can be learned well by SFDA methods. Red blocks include the hard-adaptive source-dissimilar samples with label-chaotic neighbors. In (b), (c), and (d), the dark green indicates that the pseudo-label is consistent with the true label, and light blue indicates the opposite. The red stars indicate the selected samples based on BVSB, CoreSet, and our MHPL.

Minimum Happy Points Learning

山东大学 人工智能研究中心

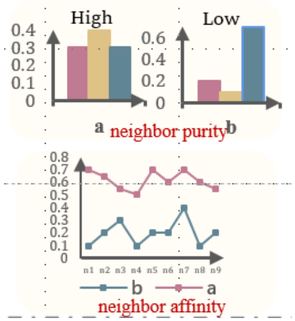
The Framework of our method MHPL is composed of two components: Minimum Happy Points Exploration and Minimum Happy Points Exploitation.



Minimum Happy Points Learning

◆ **Minimum Happy Points Exploration.**

- The neighbor environment uncertainty (NEU) is designed to explore neighbor-chaotic samples, which combines the NP (describes the chaotic degree of neighbor labels around a sample) and NA (describes how close a sample is to its neighbors)



$$\mathrm{NEU}(x) = \mathrm{NP}(x) * \mathrm{NA}(x)\,.$$

where $\mathrm{NP}(x) = -\sum_{k=1}^{K} p_k \log p_k,\ \text{s.t. } p_k \in S_N^p(x)$ $\quad \mathrm{NA}(x) = \dfrac{S_{N_1} + \ldots + S_{N_q}}{q},\ \text{s.t. } S_{N_q} \in S_N^s(x)$

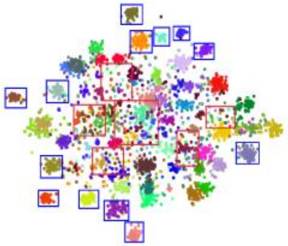
- The neighbor diversity relaxation (NDR) guarantees individual differences in neighbor-chaotic samples locally, thus redundant samples with similar features would have few chances to be chosen.

Algorithm 1 Neighbor Diversity Relaxation.

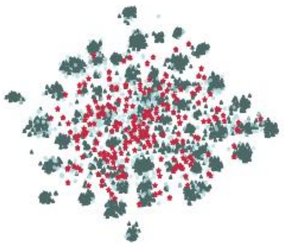
Require: D_t^L=∅, D_t, m, and o;

Sort the samples in D_t in reverse order by the value of NEU, and let $i = 0$;
while length$(D_t^L) \leq m$ **do**
  Select the candidate sample x_i and obtain its nearest neighbor set $S_N(x_i)^o$,
  if $S_N(x_i)^o \cap D_t^L = \emptyset$ **then** $D_t^L \leftarrow x_i$,
  else Skip the selection of sample x_i,
end while

- The one-shot querying (OSQ) guarantees source-dissimilar samples, selecting samples at once according to the raw source model.



(a) Initial pseudo-labels



(d) MHPL

Minimum Happy Points Learning

山东大学 人工智能研究中心

◆ **Minimum Happy Points Exploitation.**

- **Neighbor Focal** (**NF**) Loss is designed to make the model focus more on informative MH points to ensure the generalization of the target domain. In particular, NF loss assigns the neighbor purity and a larger weight α to the MH points:

$$L_{NF}(h_t;\mathcal{D}_t^L) = -\mathbb{E}_{x\in\mathcal{D}_t^L}\sum_{k=1}^{K}\alpha \mathrm{NP}(x) l_k \log(\delta_k(h_t(x)))\,,$$

  Meanwhile, the NF loss assigns a smaller weight β to the rest target samples with pseudo-labels:

$$L_{NF}(h_t;\mathcal{D}_t^U) = -\mathbb{E}_{x\in\mathcal{D}_t^U}\sum_{k=1}^{K}\beta l_k \log(\delta_k(h_t(x)))\,.$$

  In summary, NF loss is defined as the combination of the above two parts:

$$L_{NF}(h_t;\mathcal{D}_t) = L_{NF}(h_t;\mathcal{D}_t^L) + L_{NF}(h_t;\mathcal{D}_t^U)$$

- **Entropy loss** and **KL divergence** are introduced to guarantee the unambiguous and balanced classes, which has been widely used in clustering and several DA works:

$$L_{ent}(h_t;\mathcal{D}_t) = -\mathbb{E}_{x\in\mathcal{D}_t}\sum_{k=1}^{K}\delta_k(h_t(x))\log(\delta_k(h_t(x)))\,,$$

$$L_{div}(h_t;\mathcal{D}_t) = -\mathbb{E}_{x\in\mathcal{D}_t}\sum_{k=1}^{K}\mathrm{KL}(\hat{p}_k||q_k),$$

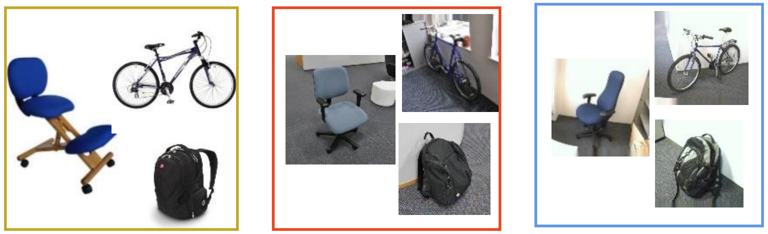
Final optimization goal: $L = L_{NF}(h_t;\mathcal{D}_t) + L_{ent}(h_t;\mathcal{D}_t) + L_{div}(h_t;\mathcal{D}_t)\,.$

Experiments

山东大学 人工智能研究中心

Datasets

- Office-31



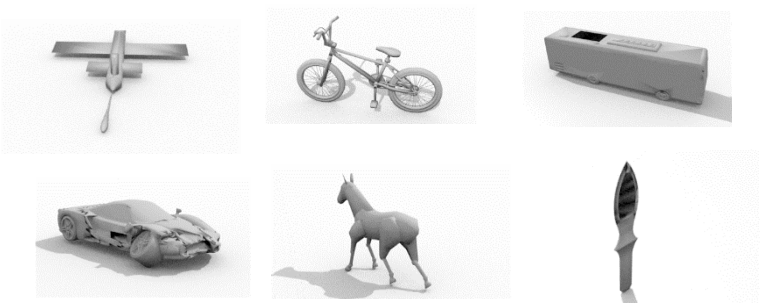
Amazon   Dslr   Webcam

- Office-Home

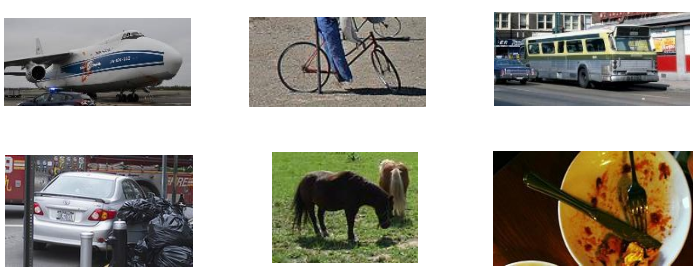
Art   Clipart   Product   RealWorld

- VisDA-2017

Synthetic Source



Real target



Experiments

山东大学 人工智能研究中心

Main Results

Table 1. Accuracy (%) on Office-Home (ResNet50) under different settings with 5% labeled target samples ("SF" in tables denotes source data free, *i.e.*, adaptation without source data).

Categories	Method	SF	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
None	Source-only	✓	45.5	68.4	75.2	53.4	63.7	65.6	52.4	41.0	73.6	65.9	46.3	78.2	60.8
SFDA	CPGA [33]	✓	59.3	78.1	79.8	65.4	75.5	76.4	65.7	58.0	81.0	72.0	64.4	83.3	71.6
	SHOT [25]	✓	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
	NRC [53]	✓	57.7	80.3	82.0	68.1	79.8	78.6	65.3	56.4	83.0	71.0	58.6	85.6	72.2
	A^2Net [50]	✓	58.4	79.0	82.4	67.5	79.3	78.9	68.0	56.2	82.9	74.1	60.5	85.0	72.8
Active DA	AADA [40]	×	56.6	78.1	79.0	58.5	72.7	71.0	60.1	53.1	77.0	70.6	57.0	84.5	68.3
	TQS [6]	×	58.6	81.1	81.5	61.1	76.1	73.3	61.2	54.7	79.7	73.4	58.9	86.3	72.5
	Clue [32]	×	58.0	79.3	80.9	68.8	77.5	76.7	66.3	57.9	81.4	75.6	60.8	86.3	72.5
	EADA [51]	×	63.6	84.4	83.5	70.7	83.7	80.5	73.0	63.5	85.2	79.4	65.4	88.6	76.7
ASFDA	Base	✓	57.2	78.5	81.5	68.5	79.1	78.6	67.5	56.3	82.2	73.7	58.5	83.6	72.1
	Random	✓	63.8	81.4	83.9	71.3	82.2	81.4	68.8	62.4	83.3	76.1	63.8	85.8	75.2
	CTC	✓	60.8	78.7	82.2	69.3	79.2	79.8	68.6	59.4	82.2	74.6	61.7	84.4	73.4
	CoreSet [36]	✓	61.8	81.8	83.3	71.1	82.9	81.6	70.7	60.5	84.7	76.1	61.7	86.1	75.2
	BADGE [2]	✓	62.4	82.7	83.9	71.5	83.0	81.8	71.2	62.7	84.6	76.2	62.9	87.8	75.9
	Entropy [46]	✓	65.0	84.0	85.9	71.8	83.8	82.6	70.7	63.8	85.1	77.8	64.1	88.1	76.9
	BVSB [14]	✓	64.8	84.4	85.5	72.0	83.2	83.4	70.4	63.9	85.0	77.5	65.0	88.1	76.9
	LC [10]	✓	65.0	84.0	85.4	72.1	83.0	82.8	71.0	64.9	85.1	78.0	64.8	87.9	77.0
	ELPT [24]	✓	65.3	84.1	84.9	**72.9**	84.4	82.8	69.8	63.3	86.1	76.2	65.6	89.1	77.0
	MHPL	✓	**69.0**	**85.7**	**86.4**	72.6	**87.4**	**84.2**	**73.3**	**67.4**	**86.4**	**80.1**	**69.6**	**89.8**	**79.3**

Table 3. Accuracy (%) on VisDA-2017 (ResNet-50) with 5% labeled target samples.

Method	AADA [40]	TQS [6]	Clue [32]	EADA [51]	MHPL
Acc (%)	80.8 ± 0.4	83.1 ± 0.4	85.2 ± 0.4	88.3 ± 0.1	$\mathbf{89.6 \pm 0.1}$

Table 2. Accuracy (%) on VisDA-2017 (ResNet-101) and Office-31 (ResNet-50) with 5% labeled target samples ("SF" in tables denotes source data free, *i.e.*, adaptation without source data).

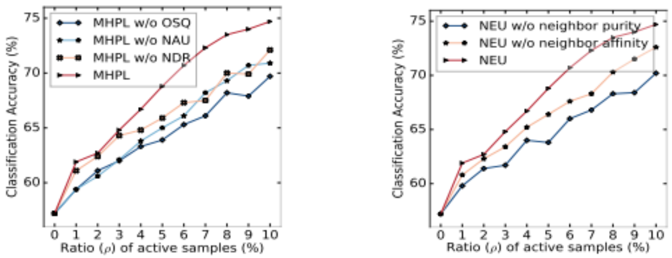
Categories	Method	SF	VisD A-2017	A→D	A→W	D→A	D→W	W→A	W→D	Avg
None	Source-only	✓	50.0	79.3	76.6	60.5	96.6	63.8	98.8	79.3
SFDA	SHOT [25]	✓	82.4	94.0	90.1	74.7	98.4	74.3	99.9	88.6
	CPGA [33]	✓	86.0	94.4	94.1	76.0	98.4	76.6	99.8	89.9
	HCL [12]	✓	83.5	90.8	91.3	72.7	98.2	72.7	**100.0**	87.6
	NRC [53]	✓	85.9	96.0	90.8	75.3	99.0	75.0	**100.0**	89.4
	A^2Net [50]	✓	84.3	94.5	94.0	76.7	99.2	76.1	**100.0**	90.1
Active DA	AADA [40]	×	-	89.2	87.3	78.2	99.5	78.7	**100.0**	88.8
	TQS [6]	×	-	92.8	92.2	80.6	**100.0**	80.4	**100.0**	91.1
	Clue [32]	×	-	92.0	87.3	79.0	99.2	79.6	99.8	89.5
	EADA [51]	×	-	97.7	96.6	82.1	**100.0**	**82.8**	**100.0**	**93.2**
ASFDA	Base	✓	83.3	93.8	91.5	76.0	99.0	74.7	**100.0**	89.2
	Random	✓	85.1	94.0	94.7	77.7	98.9	77.6	**100.0**	90.5
	CTC	✓	84.0	93.8	90.8	77.3	99.0	76.2	**100.0**	89.5
	CoreSet [36]	✓	85.9	93.4	92.5	78.4	99.1	78.2	**100.0**	90.3
	BADGE [2]	✓	86.0	94.2	93.5	79.2	99.1	79.2	**100.0**	90.9
	Entropy [46]	✓	86.7	95.6	95.4	80.3	99.1	80.1	**100.0**	91.8
	BVSB [14]	✓	86.5	96.4	95.7	79.2	99.1	80.5	**100.0**	91.9
	LC [10]	✓	86.7	95.6	95.4	80.0	99.1	80.6	**100.0**	91.8
	ELPT [24]	✓	89.2	**98.0**	**97.2**	81.2	99.4	80.7	**100.0**	92.8
	MHPL	✓	**91.3**	97.8	96.7	**82.5**	99.3	82.6	**100.0**	**93.2**

Table 4. Accuracy (%) on challenging tasks under different networks with 5% labeled target samples on Office-home.

Networks	active samples	Ar→Cl	Cl→Ar	Cl→Pr	Pr→Cl	Avg
VGG16	Source-only	35.2	48.0	60.5	35.4	44.8
	LC points	46.2	62.1	75.7	48.4	58.1
	MH Points	49.6	63.0	78.0	50.4	60.3
ResNet50	Source-only	45.5	53.4	63.7	46.3	52.2
	LC points	58.7	69.0	80.7	60.9	67.3
	MH Points	60.3	69.9	83.8	61.9	69.0

Experiments

- **Ablation analysis**

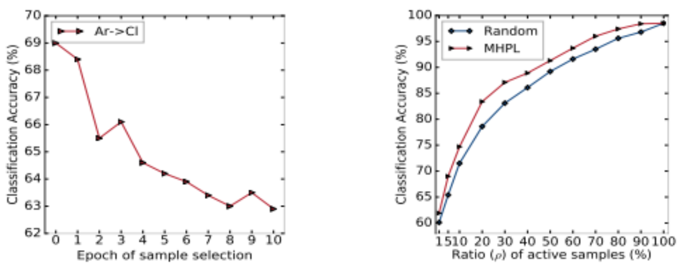


(a) Ar→Cl (b) Ar→Cl

Figure 4. Ablation study on key components of MHPL and NEU at different selection ratios.

(a) Ar→Cl without NDR (b) Ar→Cl

Figure 5. T-SNE visualization of representations with and without NDR in Ar→Cl on Office-Home.



(a) Ar→Cl (b) Ar→Cl

Figure 6. Ablation study on OSQ and random selection.

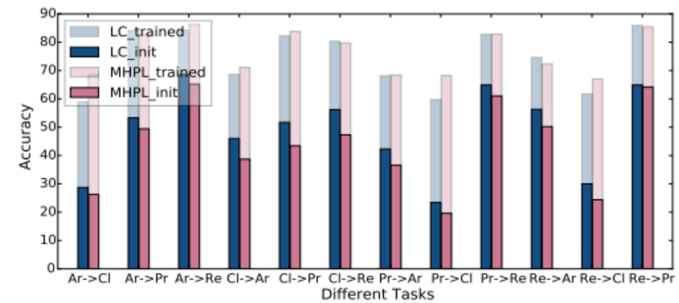
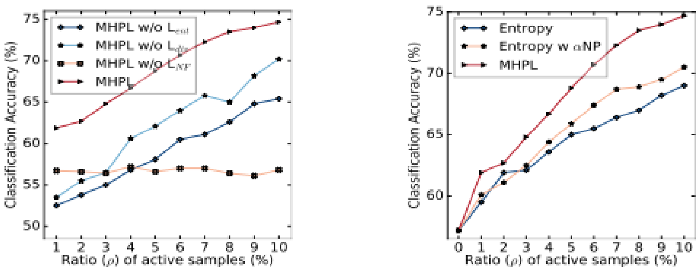


Figure 7. Mean accuracy of initial neighbors before and after training on Office-Home.

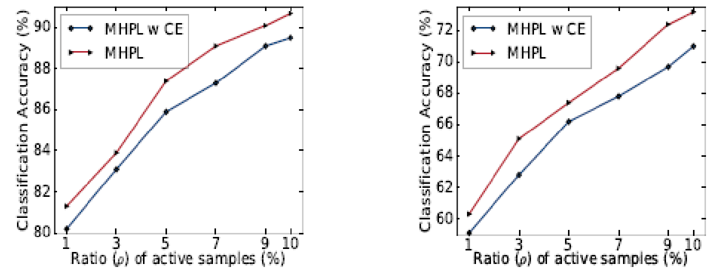
Experiments

■ Ablation analysis



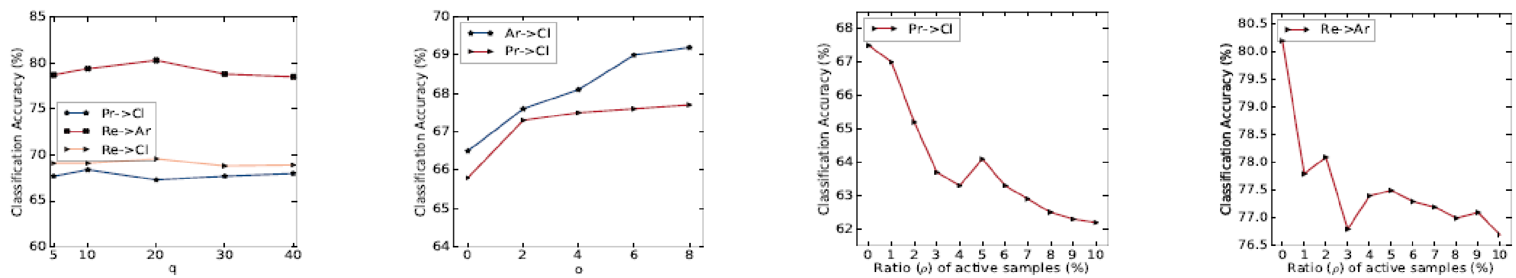
(a) Ar→Cl (b) Ar→Cl

Figure 8. Ablation studies on loss functions.



(a) Cl→Pr (b) Pr→Cl

Figure 3. Ablation study on focal loss. 'MHPL w CE' represents replacing the focal loss with normal CE loss in MHPL.



(a) Neighbor number of NAU (b) Neighbor number of NDR (c) Pr→Cl (d) Re→Ar

Figure 1. Ablation studies on hyperparameters q, o and one-shot querying.

Experiments

山东大学
人工智能研究中心

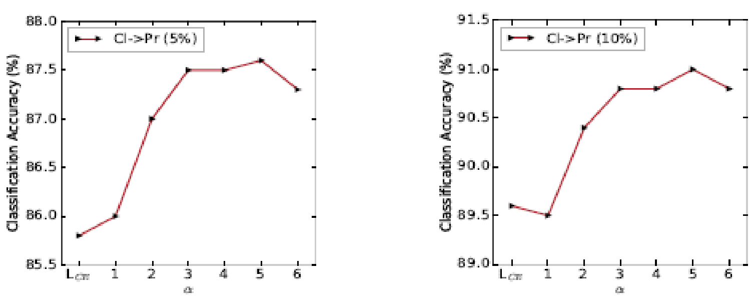
■ Ablation analysis

Table 1. Accuracy (%) on Office-Home of different networks with 5% labeled target samples.

Model	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
Alexnet	Source-only	26.1	37.4	49.5	24.5	39.6	40.0	25.3	24.0	49.8	40.0	31.3	59.8	37.3
	Base	30.7	50.3	56.2	34.5	51.4	53.4	33.0	26.8	58.9	44.1	36.2	64.9	45.0
	Random	37.8	58.0	61.0	39.7	58.1	59.4	36.2	34.2	62.8	48.7	45.0	69.8	50.9
	CTC	37.3	55.3	58.5	38.2	56.6	56.9	35.4	33.3	60.4	45.5	43.0	65.6	48.8
	CoreSet [11]	36.1	55.1	59.9	38.9	57.9	58.8	38.0	31.3	62.4	47.7	42.0	69.1	49.8
	BADGE [2]	37.3	54.7	59.7	39.3	58.3	58.1	38.8	33.3	63.6	49.1	43.1	69.1	50.4
	Entropy [14]	38.3	56.1	62.7	40.5	59.0	60.7	37.3	35.9	64.6	48.8	45.3	71.0	51.7
	BVSB [6]	39.2	57.7	62.8	39.7	60.1	61.5	36.7	36.7	64.9	48.9	45.8	71.6	52.1
	LC [5]	39.0	56.8	63.4	39.8	59.7	60.6	37.4	36.7	64.9	48.5	45.8	71.4	52.0
	MHPL	**45.5**	**64.4**	**64.9**	**42.8**	**67.3**	**61.7**	**40.8**	**44.4**	**66.4**	**50.5**	**50.9**	**75.7**	**56.3**
VGG	Source-only	35.2	59.8	69.4	48.0	60.5	62.6	46.7	30.8	70.5	58.8	35.4	74.3	54.3
	Base	44.3	73.1	74.0	59.3	72.5	71.5	55.9	42.0	76.6	63.5	46.6	80.4	63.3
	Random	52.6	77.0	76.9	61.8	76.8	74.9	58.7	49.4	78.8	66.3	53.7	83.3	67.5
	CTC	50.8	74.3	76.0	60.3	75.2	73.5	58.2	47.9	77.3	64.9	52.1	81.0	66.0
	CoreSet [11]	49.7	77.2	76.6	63.3	77.2	74.9	60.1	46.4	79.0	66.5	50.7	83.4	67.1
	BADGE [2]	51.0	77.4	77.0	63.2	77.0	75.3	60.0	47.2	79.3	66.9	51.4	83.4	67.4
	Entropy [14]	52.6	78.1	78.0	63.1	77.9	72.2	60.4	49.0	80.6	68.5	54.1	84.1	68.6
	BVSB [6]	54.7	77.1	78.6	63.5	78.6	77.1	59.7	50.7	80.2	68.3	54.6	85.3	69.0
	LC [5]	53.0	78.7	78.4	63.0	78.0	76.8	60.4	49.2	80.5	68.5	54.3	85.1	68.8
	MHPL	**60.4**	**82.4**	**80.5**	**66.7**	**82.3**	**78.4**	**63.5**	**56.9**	**81.9**	**70.8**	**59.3**	**87.2**	**72.5**

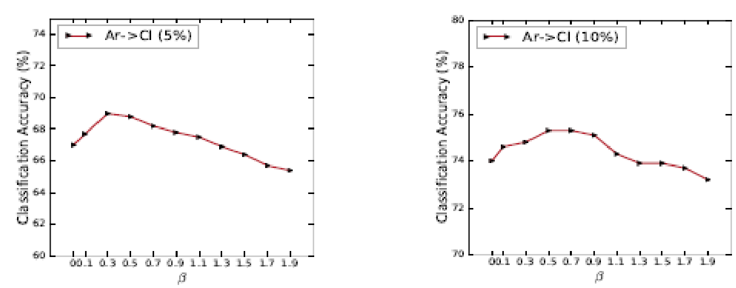
Table 2. Accuracy (%) on Office-31 of different networks with 5% labeled target samples.

Model	Method	A→D	A→W	D→A	D→W	W→A	W→D	Avg
Alexnet	Source-only	53.6	47.9	37.9	94.2	36.4	97.8	61.3
	Base	69.2	63.1	69.6	95.7	48.6	98.6	70.8
	Random	73.1	68.2	58.2	95.9	55.4	99.6	75.0
	CTC	71.5	65.5	56.3	95.6	52.7	99.0	73.4
	CoreSet [11]	70.3	69.9	59.2	96.0	54.2	98.4	74.7
	BADGE [2]	71.3	69.9	57.2	96.4	53.9	99.4	74.7
	Entropy [14]	74.2	71.2	57.1	98.9	54.8	99.6	76.0
	BVSB [6]	73.1	73.8	61.7	98.7	58.6	99.6	77.6
	LC [5]	74.1	74.1	58.6	99.1	55.4	99.6	76.8
	MHPL	**81.1**	**78.6**	**68.9**	**98.5**	**67.0**	**100.0**	**82.4**
VGG	Source-only	75.2	72.7	63.8	95.4	63.7	99.8	78.5
	Base	88.0	87.2	72.9	97.2	71.6	99.8	86.1
	Random	87.4	89.1	74.4	97.5	73.8	99.8	87.0
	CTC	88.0	86.9	73.4	97.1	71.9	99.8	86.2
	CoreSet [11]	89.0	88.7	75.1	97.5	74.1	99.8	87.4
	BADGE [2]	89.2	88.1	75.5	97.5	73.3	99.8	87.2
	Entropy [14]	87.8	89.1	77.7	98.2	74.7	100.0	87.9
	BVSB [6]	91.6	88.4	76.3	98.5	75.2	100.0	88.3
	LC [5]	90.8	88.9	77.2	98.5	75.3	100.0	88.5
	MHPL	**93.0**	**93.0**	**79.9**	**98.9**	**80.3**	**100.0**	**90.9**



(a) Cl→Pr (b) Cl→Pr

Figure 5. Ablation study on α.



(a) Ar→Cl (b) Ar→Cl

Figure 6. Ablation study on β.

山东大学
人工智能研究中心

TIME 机器学习与数据挖掘实验室

Q & A

Thanks!

学无止境　气有浩然