Categorical Domain Adaptation Theory and Algorithms

Pengcheng Xu

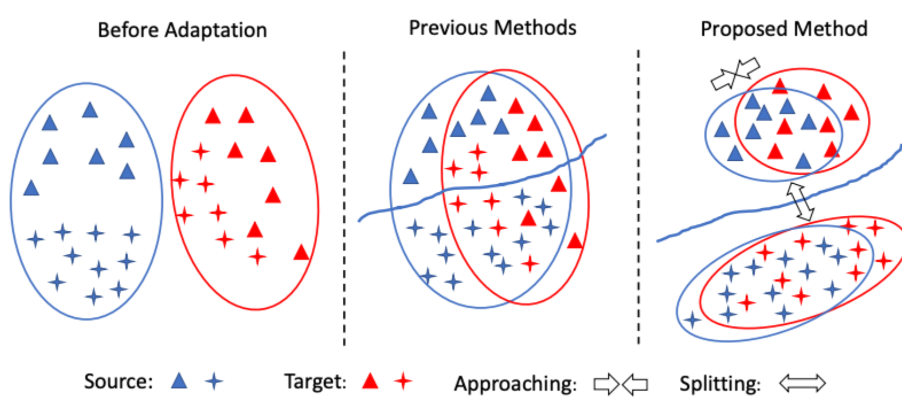
2020, SEP, 25

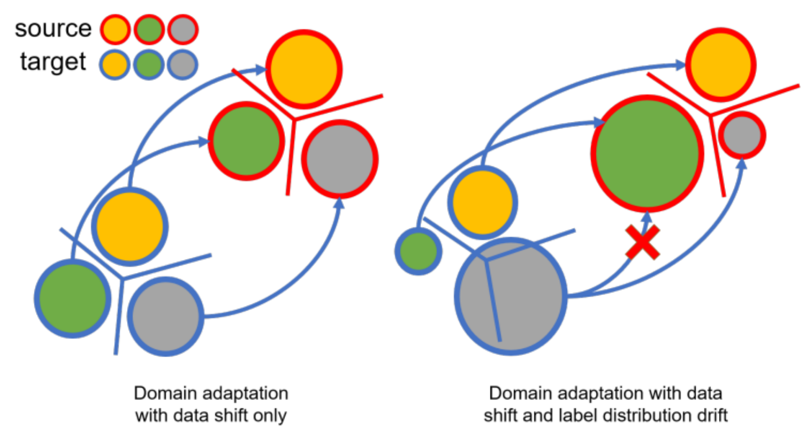
Paper list

- Contrastive Adaptation Network for Unsupervised Domain Adaptation
- Deep Subdomain Adaptation Network for Image Classification
- Opposite Structure Learning for Semi-supervised Domain Adaptation

Problems

- Issues for covariate shift assumption
  - Classification mechanism is the same; Marginal distribution is the different.
  - $P_S(y|x) = P_T(y|x)$ $\;P_S(x) \neq P_T(x)$ => Align the source & target marginal distribution
  - Negative transferring due to lack of categorical information on target.





Contrastive Adaptation Network for Unsupervised Domain Adaptation

Guoliang Kang[1], Lu Jiang[2], Yi Yang[1,3]*, Alexander G. Hauptmann[4]
[1]CAI, University of Technology Sydney, [2]Google AI, [3]Baidu Research, [4]Carnegie Mellon University
kgl.prml@gmail.com, lujiang@google.com, Yi.Yang@uts.edu.au, alex@cs.cmu.edu

Intuition

- 1. Both inter and intra domain adaptation
  - Modified MMD to CDD as distance measure
- 2. Explicitly categorical domain adaptation
  - Assign pseudo labels to unlabeled target domain
- 3. Alternative training for pseudo labeling
  - Iterate between spherical K-means and Adaptation
- 4. Class-ware sampling

Categorical Distribution Alignment

MMD distance of marginal distributions

$$\mathcal{D}_{\mathcal{H}}(P,Q) \triangleq \sup_{f \sim \mathcal{H}} (\mathbb{E}_{\boldsymbol{X}^s}[f(\boldsymbol{X}^s)] - \mathbb{E}_{\boldsymbol{X}^t}[f(\boldsymbol{X}^t)])_{\mathcal{H}},$$

$$\begin{aligned}
\hat{\mathcal{D}}_l^{mmd} &= \frac{1}{n_s^2} \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} k_l(\phi_l(\boldsymbol{x}_i^s), \phi_l(\boldsymbol{x}_j^s)) \\
&+ \frac{1}{n_t^2} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} k_l(\phi_l(\boldsymbol{x}_i^t), \phi_l(\boldsymbol{x}_j^t)) \\
&- \frac{2}{n_s n_t} \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} k_l(\phi_l(\boldsymbol{x}_i^s), \phi_l(\boldsymbol{x}_j^t)),
\end{aligned}$$

Contrastive Domain Discrepancy

- Extend the marginal MMD to conditional CDD

$$\sup_{f\sim\mathcal{H}}\left(\mathbb{E}_{\boldsymbol{X}^s}[f(\phi(\boldsymbol{X}^s)|Y^s)] - \mathbb{E}_{\boldsymbol{X}^t}[f(\phi(\boldsymbol{X}^t)|Y^t)]\right)_{\mathcal{H}}.$$

- Define the divergence of any two classes

$$\hat{\mathcal{D}}^{c_1c_2}(\hat{y}_1^t, \hat{y}_2^t, \cdots, \hat{y}_{n_t}^t, \phi) = e_1 + e_2 - 2e_3$$

e1 & e2 are kernels of the same classes; e3 are kernels of different classes

- Combine the divergence of all classes

$$\hat{\mathcal{D}}^{cdd} = \underbrace{\frac{1}{M}\sum_{c=1}^{M}\hat{\mathcal{D}}^{cc}(\hat{y}_{1:n_t}^t, \phi)}_{intra} - \underbrace{\frac{1}{M(M-1)}\sum_{c=1}^{M}\sum_{\substack{c'=1 \\ c'\neq c}}^{M}\hat{\mathcal{D}}^{cc'}(\hat{y}_{1:n_t}^t, \phi)}_{inter},$$

Overall Loss:
Add all layers and CE loss

$$\hat{\mathcal{D}}_{\mathcal{L}}^{cdd} = \sum_{l=1}^{L}\hat{\mathcal{D}}_l^{cdd}.$$

$$\min_{\theta}\ell = \ell^{ce} + \beta\hat{\mathcal{D}}_{\mathcal{L}}^{cdd}$$

Alternate Optimization

- Jointly update both pseudo labels and features network

- Target cluster centers are initialized with source cluster centers

$$O^{sc} = \sum_{i=1}^{N_s} \mathbf{1}_{y_i^s=c} \frac{\phi_1(\boldsymbol{x}_i^s)}{\|\phi_1(\boldsymbol{x}_i^s)\|}, \ \mathbf{1}_{y_i^s=c} \begin{cases} 1 & \text{if } y_i^s = c \\ 0 & \text{otherwise} \end{cases}$$

- Clustering target with spherical k-means

$$\hat{y}_i^t \leftarrow \operatorname{argmin}_c dist(\phi_1(\boldsymbol{x}_i^t), O^{tc}) \qquad O^{tc} \leftarrow \sum_{i=1}^{N_t} \mathbf{1}_{\hat{y}_i^t=c} \frac{\phi_1(\boldsymbol{x}_i^t)}{\|\phi_1(\boldsymbol{x}_i^t)\|}$$

- Filtering the ambiguous samples based on distance

$$\{(\boldsymbol{x}^t, \hat{y}^t) | dist(\phi_1(\boldsymbol{x}^t), O^{t(\hat{y}^t)}) < D_0, \boldsymbol{x}^t \in \mathcal{T}\}$$

Class-Ware Sampling

- Select a subset classes with enough samples
- Sample data for each class to construct a mini-batch for intra-DA

Algorithm 1: Optimization of CAN at loop T_e.

Input:

source data: $\mathcal{S} = \{(\boldsymbol{x}_1^s, y_1^s), \cdots, (\boldsymbol{x}_{N_s}^s, y_{N_s}^s)\}$,

target data: $\mathcal{T} = \{\boldsymbol{x}_1^t, \cdots, \boldsymbol{x}_{N_t}^t\}$

Procedure:

Forward $\mathcal{S}$ and compute the M cluster centers O^{sc} ;

Initialize O^{tc}: $O^{tc} \leftarrow O^{sc}$;

Cluster target samples $\mathcal{T}$ using spherical K-means;

Filter the ambiguous target samples and classes;

for $(k \leftarrow 1; k \leq K; k \leftarrow k+1)$ **do**

  Class-aware sampling based on $\mathcal{C}'_{T_e}$, $\tilde{\mathcal{T}}$, and $\mathcal{S}$;

  Compute $\hat{\mathcal{D}}_{\mathcal{L}}^{cdd}$ using Eq. (6);

  Sample from $\mathcal{S}$ and compute ℓ^{ce} using Eq. (7);

  Back-propagate with the objective ℓ (Eq.(8));

  Update network parameters θ.

end

Experiment Results

Method	A → W	D → W	W → D	A → D	D → A	W → A	Average
Source-finetune	68.4 ± 0.2	96.7 ± 0.1	99.3 ± 0.1	68.9 ± 0.2	62.5 ± 0.3	60.7 ± 0.3	76.1
RevGrad [10, 11]	82.0 ± 0.4	96.9 ± 0.2	99.1 ± 0.1	79.7 ± 0.4	68.2 ± 0.4	67.4 ± 0.5	82.2
DAN [22]	80.5 ± 0.4	97.1 ± 0.2	99.6 ± 0.1	78.6 ± 0.2	63.6 ± 0.3	62.8 ± 0.2	80.4
JAN [25]	85.4 ± 0.3	97.4 ± 0.2	99.8 ± 0.2	84.7 ± 0.3	68.6 ± 0.3	70.0 ± 0.4	84.3
MADA [28]	90.0 ± 0.2	97.4 ± 0.1	99.6 ± 0.1	87.8 ± 0.2	70.3 ± 0.3	66.4 ± 0.3	85.2
Ours (intra only)	93.2 ± 0.2	98.4 ± 0.2	99.8 ± 0.2	92.9 ± 0.2	76.5 ± 0.3	76.0 ± 0.3	89.5
Ours (CAN)	**94.5 ± 0.3**	**99.1 ± 0.2**	**99.8 ± 0.2**	**95.0 ± 0.3**	**78.0 ± 0.3**	**77.0 ± 0.3**	**90.6**

Table 1. Classification accuracy (%) for all the six tasks of Office-31 dataset based on ResNet-50 [14, 15]. Our methods named "intra only" and "CAN" are trained with intra-class domain discrepancy and contrastive domain discrepancy, respectively.

Method	airplane	bicycle	bus	car	horse	knife	motorcycle	person	plant	skateboard	train	truck	Average
Source-finetune	72.3	6.1	63.4	**91.7**	52.7	7.9	80.1	5.6	90.1	18.5	78.1	25.9	49.4
RevGrad [10, 11]	81.9	77.7	82.8	44.3	81.2	29.5	65.1	28.6	51.9	54.6	82.8	7.8	57.4
DAN [22]	68.1	15.4	76.5	87.0	71.1	48.9	82.3	51.5	88.7	33.2	88.9	42.2	62.8
JAN [25]	75.7	18.7	82.3	86.3	70.2	56.9	80.5	53.8	92.5	32.2	84.5	54.5	65.7
MCD [32]	87.0	60.9	83.7	64.0	88.9	79.6	84.7	76.9	88.6	40.3	83.0	25.8	71.9
ADR [31]	87.8	79.5	83.7	65.3	92.3	61.8	88.9	73.2	87.8	60.0	85.5	32.3	74.8
SE [9]	95.9	**87.4**	**85.2**	58.6	96.2	95.7	90.6	80.0	94.8	90.8	88.4	47.9	84.3
Ours (intra only)	96.5	72.1	80.9	70.8	94.6	**98.0**	**91.7**	**84.2**	90.3	89.8	**89.4**	47.9	83.9
Ours (CAN)	**97.0**	87.2	82.5	74.3	**97.8**	96.2	90.8	80.7	**96.6**	**96.3**	87.5	**59.9**	**87.2**

Table 2. Classification accuracy (%) on the VisDA-2017 validation set based on ResNet-101 [14, 15]. Our methods named "intra only" and "CAN" are trained with intra-class domain discrepancy and contrastive domain discrepancy, respectively.

Deep Subdomain Adaptation Network for Image Classification

Yongchun Zhu[1,2], Fuzhen Zhuang[1,2], Jindong Wang[3], Guolin Ke[3], Jingwu Chen[4], Jiang Bian[3], Hui Xiong[5], and Qing He[1,2]
[1]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]Microsoft Research, [4]ByteDance, [5]the State University of New Jersey
{zhuyongchun18s,zhuangfuzhen,heqing}@ict.ac.cn, {jindong.wang,Guolin.Ke,jiang.bian}@microsoft.com
{chenjingwu}@bytedance.com, {hxiong}@rutgers.edu

Intuition

- 1 Categorical subdomain alignment
- 2. Current subdomain adversarial methods include many loss functions and converges slowly
- 3. Soft target labels for re-weighting the LMMD

Maximum Mean Discrepancy

- MMD definitions

$$d_{\mathcal{H}}(p,q) \triangleq \|\mathbf{E}_p[\phi(\mathbf{x}^s)] - \mathbf{E}_q[\phi(\mathbf{x}^t)]\|^2_{\mathcal{H}},$$

  - p=q iff the MMD is zero but we can only get an estimation of the MMD

- LMMD definitions

$$d_{\mathcal{H}}(p,q) \triangleq \mathbf{E}_c\|\mathbf{E}_{p^{(c)}}[\phi(\mathbf{x}^s)] - \mathbf{E}_{q^{(c)}}[\phi(\mathbf{x}^t)]\|^2_{\mathcal{H}},$$

$$\hat{d}_{\mathcal{H}}(p,q) = \frac{1}{C}\sum_{c=1}^{C}\left\|\sum_{\mathbf{x}_i^s \in \mathcal{D}_s} w_i^{sc}\phi(\mathbf{x}_i^s) - \sum_{\mathbf{x}_j^t \in \mathcal{D}_t} w_j^{tc}\phi(\mathbf{x}_j^t)\right\|^2_{\mathcal{H}}, \qquad w_i^c = \frac{y_{ic}}{\sum_{(\mathbf{x}_j,\mathbf{y}_j)\in\mathcal{D}} y_{jc}},$$

Deep Subdomain Network

- Align the distribution and generate more accurate pseudo labeling;
- For source, use the one-hot encoding; For target, use the soft target;

$$\min_f \frac{1}{n_s} \sum_{i=1}^{n_s} J(f(\mathbf{x}_i^s), \mathbf{y}_i^s) + \lambda \sum_{l \in L} \hat{d}_l(p, q).$$

Bound Theory

- Bound under covariate shift

$$\forall h \in \mathcal{H}, R_{\mathcal{T}}(h) \leq R_{\mathcal{S}}(h) + \frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S}, \mathcal{T}) + C, \qquad C = \min_{h \in \mathcal{H}} R_{\mathcal{S}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}}(h, f_{\mathcal{T}}),$$

- Relaxation of the bound

$$R(f_1, f_2) \leq R(f_1, f_3) + R(f_2, f_3).$$

Then, we have:

$$\begin{aligned} C &= \min_{h \in \mathcal{H}} R_{\mathcal{S}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}}(h, f_{\mathcal{T}}) \\ &\leq \min_{h \in \mathcal{H}} R_{\mathcal{S}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}}(f_{\mathcal{S}}, f_{\mathcal{T}}) \\ &\leq \min_{h \in \mathcal{H}} R_{\mathcal{S}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}}(f_{\mathcal{S}}, f_{\hat{\mathcal{T}}}) \\ &\quad + R_{\mathcal{T}}(f_{\mathcal{T}}, f_{\hat{\mathcal{T}}}), \end{aligned}$$

3rd item is the subdomain discrepancy

4th item is expected to be reduced with training

Experiments

TABLE I
ACCURACY (%) ON IMAGECLEF-DA FOR UNSUPERVISED DOMAIN ADAPTATION (RESNET50).

Method	I → P	P → I	I → C	C → I	C → P	P → C	Avg
ResNet [1]	74.8±0.3	83.9±0.1	91.5±0.3	78.0±0.2	65.5±0.3	91.2±0.3	80.7
DDC [39]	74.6±0.3	85.7±0.8	91.1±0.3	82.3±0.7	68.3±0.4	88.8±0.2	81.8
DAN [13]	75.0±0.4	86.2±0.2	93.3±0.2	84.1±0.4	69.8±0.4	91.3±0.4	83.3
DANN [17]	75.0±0.6	86.0±0.3	96.2±0.4	87.0±0.5	74.3±0.5	91.5±0.6	85.0
D-CORAL [16]	76.9±0.2	88.5±0.3	93.6±0.3	86.8±0.6	74.0±0.3	91.6±0.3	85.2
JAN [26]	76.8±0.4	88.0±0.2	94.7±0.2	89.5±0.3	74.2±0.3	91.7±0.3	85.8
MADA [15]	75.0±0.3	87.9±0.2	96.0±0.3	88.8±0.3	75.2±0.2	92.2±0.3	85.8
CAN [31]	78.2	87.5	94.2	89.5	75.8	89.2	85.7
iCAN [31]	79.5	89.7	94.7	89.9	78.5	92.0	87.4
CDAN [14]	76.7±0.3	90.6±0.3	97.0±0.4	90.5±0.4	74.5±0.3	93.5±0.4	87.1
CDAN+E [14]	77.7±0.3	90.7±0.2	**97.7**±0.3	91.3±0.3	74.2±0.2	94.3±0.3	87.7
DSAN	**80.2**±0.2	**93.3**±0.4	97.2±0.2	**93.8**±0.2	**80.8**±0.4	**95.9**±0.4	**90.2**

TABLE II
ACCURACY (%) ON OFFICE-31 FOR UNSUPERVISED DOMAIN ADAPTATION (RESNET50).

Method	A → W	D → W	W → D	A → D	D → A	W → A	Avg
ResNet [1]	68.4±0.5	96.7±0.5	99.3±0.1	68.9±0.2	62.5±0.3	60.7±0.3	76.1
DDC [39]	75.8±0.2	95.0±0.2	98.2±0.1	77.5±0.3	67.4±0.4	64.0±0.5	79.7
DAN [13]	83.8±0.4	96.8±0.2	99.5±0.1	78.4±0.2	66.7±0.3	62.7±0.2	81.3
D-CORAL [16]	77.7±0.3	97.6±0.2	99.7±0.1	81.1±0.4	64.6±0.3	64.0±0.4	80.8
DANN [17]	82.0±0.4	96.9±0.2	99.1±0.1	79.7±0.4	68.2±0.4	67.4±0.5	82.2
ADDA [30]	86.2±0.5	96.2±0.3	98.4±0.3	77.8±0.3	69.5±0.4	68.9±0.5	82.9
JAN [26]	85.4±0.3	97.4±0.2	99.8±0.2	84.7±0.3	68.6±0.3	70.0±0.4	84.3
MADA [15]	90.0±0.1	97.4±0.1	99.6±0.1	87.8±0.2	70.3±0.3	66.4±0.3	85.2
GTA [43]	89.5±0.5	97.9±0.3	99.8±0.4	87.7±0.5	72.8±0.3	71.4±0.4	86.6
CAN [31]	81.5	98.2	99.7	85.5	65.9	63.4	82.4
iCAN [31]	92.5	98.8	**100.0**	90.1	72.1	69.9	87.2
CDAN [14]	93.1±0.2	98.2±0.2	**100.0**±.0	89.8±0.3	70.1±0.4	68.0±0.4	86.6
CDAN+E [14]	**94.1**±0.1	**98.6**±0.1	**100.0**±.0	**92.9**±0.2	71.0±0.3	69.3±0.3	87.7
DSAN	93.6±0.2	98.3±0.1	**100.0**±0.0	90.2±0.7	**73.5**±0.5	**74.8**±0.4	**88.4**

Experiments

TABLE IV
ACCURACY (%) ON VISDA-2017 FOR UNSUPERVISED DOMAIN ADAPTATION (RESNET101).

Method	airplane	bicycle	bus	car	horse	knife	motorcycle	person	plant	skateboard	train	truck	Avg
ResNet [1]	72.3	6.1	63.4	**91.7**	52.7	7.9	80.1	5.6	90.1	18.5	78.1	25.9	49.4
DANN [17]	81.9	**77.7**	82.8	44.3	81.2	29.5	65.1	28.6	51.9	54.6	82.8	7.8	57.4
DAN [13]	68.1	15.4	76.5	87.0	71.1	48.9	82.3	51.5	88.7	33.2	88.9	42.2	62.8
JAN [26]	75.7	18.7	82.3	86.3	70.2	56.9	80.5	53.8	92.5	32.2	84.5	**54.5**	65.7
MCD [49]	87.0	60.9	**83.7**	64.0	**88.9**	**79.6**	84.7	**76.9**	88.6	40.3	83.0	25.8	71.9
DSAN	**90.9**	66.9	75.7	62.4	**88.9**	77.0	**93.7**	75.1	**92.8**	**67.6**	**89.1**	39.4	**75.1**

Opposite Structure Learning for Semi-supervised Domain Adaptation

[1]Can Qin, [1]Lichen Wang, [2]Qianqian Ma,[1]Yu Yin,[1]Huan Wang [1,3]Yun Fu
[1] Department of Electrical & Computer Engineering, Northeastern University
[2]Department of Electrical & Computer Engineering, Boston University
[3]Khoury College of Computer Science, Northeastern University
qin.ca@husky.neu.edu, wanglichenxj@gmail.com, maqq@bu.edu,
yin.yu1@husky.neu.edu, wang.huan@husky.neu.edu, yunfu@ece.neu.edu

Intuition

- Problems:
- 1. Conditional distribution mismatch between source and targets.
- 2. Biased decision boundary towards source domain.

- Methods:
- Well **clustered** target domain for conditional mismatch; Well **scattered** source domain to regulate model.
- 1. Source **scattering and expansion**; Target classifier improves **intra-class** density and enlarge **inter-class** divergence.
- 2. Mode collapse: adversarial training between feature extractor and classifier.

Entropy Minimization

- Clustering by minimizing the conditional entropy

$$H_{tar} = -\mathbb{E}_{\boldsymbol{x}^u \sim \mathcal{U}} \sum_{k=1}^{K} [p_2(y=k|\boldsymbol{x}^u) \log p_2(y=k|\boldsymbol{x}^u)],$$

- Scattering by maximizing the conditional entropy

$$H_{src} = -\mathbb{E}_{\boldsymbol{x}^s \sim \mathcal{S}} \sum_{k=1}^{K} [p_1(y=k|\boldsymbol{x}^s) \log p_1(y=k|\boldsymbol{x}^s)],$$

- End-to-end trained with gradient reversal layer.

$$\Theta^*_{\mathcal{F}_1} = \arg\min_{\Theta_{\mathcal{F}_1}} \alpha \mathcal{L}_{src} + (1-\alpha)\mathcal{L}_{tar} + \beta H_{src},$$

$$\Theta^*_{\mathcal{F}_2} = \arg\min_{\Theta_{\mathcal{F}_2}} (1-\alpha)\mathcal{L}_{src} + \alpha \mathcal{L}_{tar} - \lambda H_{tar}.$$

$$\Theta^*_{\mathcal{G}} = \arg\min_{\Theta_{\mathcal{G}}} \mathcal{L}_{src} + \mathcal{L}_{tar} - \beta H_{src} + \lambda H_{tar}.$$

Experiments

Table 1. Quantitative results (%) on the benchmark of DomainNet.

Methods	R→C		R→P		P→C		C→S		S→P		R→S		P→R		Avg	
	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}
S+T	55.6	60.0	60.6	62.2	56.8	59.4	50.8	55.0	56.0	59.5	46.3	50.1	71.8	73.9	56.9	60.0
DANN [7]	58.2	59.8	61.4	62.8	56.3	59.6	52.8	55.4	57.4	59.9	52.2	54.9	70.3	72.2	58.4	60.7
ADR [26]	57.1	60.7	61.3	61.9	57.0	60.7	51.0	54.4	56.0	59.9	49.0	51.1	72.0	74.2	57.6	60.4
CDAN [15]	65.0	69.0	64.9	67.3	63.7	68.4	53.1	57.8	63.4	65.3	54.5	59.0	73.2	78.5	62.5	66.5
ENT [10]	65.2	71.0	65.9	69.2	65.4	71.1	54.6	60.0	59.7	62.1	52.1	61.1	75.0	78.6	62.6	67.6
MME [24]	70.0	72.2	67.7	69.7	69.0	71.7	56.3	61.8	64.8	66.8	61.0	61.9	76.1	78.5	66.4	68.9
Ours	**72.7**	**75.4**	**70.3**	**71.5**	**69.8**	**73.2**	**60.5**	**64.1**	**66.4**	**69.4**	**62.7**	**64.2**	**77.3**	**80.8**	**68.5**	**71.2**

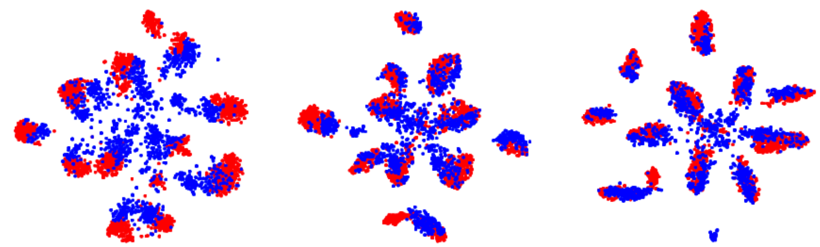
The applied backbone is Resnet34 [12] and Avg means the average results on previous adaptation scenarios.

Experiments

Table 3. Quantitative results (%) of ablation study.

COMPONENTS					R→S	R→C
1-C	2-C	$\mathcal{H}_{tar}$	$\mathcal{H}_{tar}$+$\mathcal{H}_{src}$	ST	$1_{shot}/3_{shot}$	$1_{shot}/3_{shot}$
√		√			61.0/61.9	70.0/72.2
√			√		60.4/61.2	69.2/71.5
	√	√			61.1/62.8	70.5/72.4
	√		√		62.2/63.9	71.6/74.0
	√	√		√	61.2/62.6	70.7/72.8
	√		√	√	**62.7/64.2**	**72.7/75.4**

The applied backbone is Resnet34 [12].



(a) (b) (c)

Figure 4. The visualization results of t-SNE [17] on the shared ten-class features on the adaptation scenario *Real* to *Sketch*, *i.e.,* R→S, obtained by (a) S+T, (b) MME [24] and (c) Ours. The figures are captured under 3-shot SSDA setting. The feature points of source and target domains are indicated by red and blue spots.

Q&A

Thank You Very Much