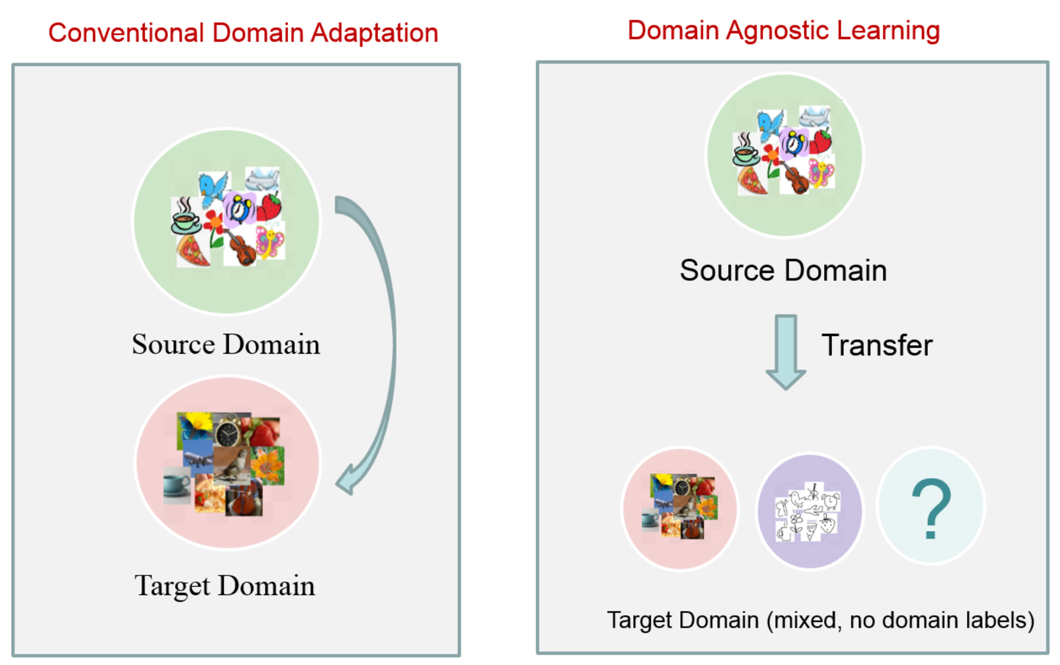
Domain Agnostic Learning with Disentangled Representations

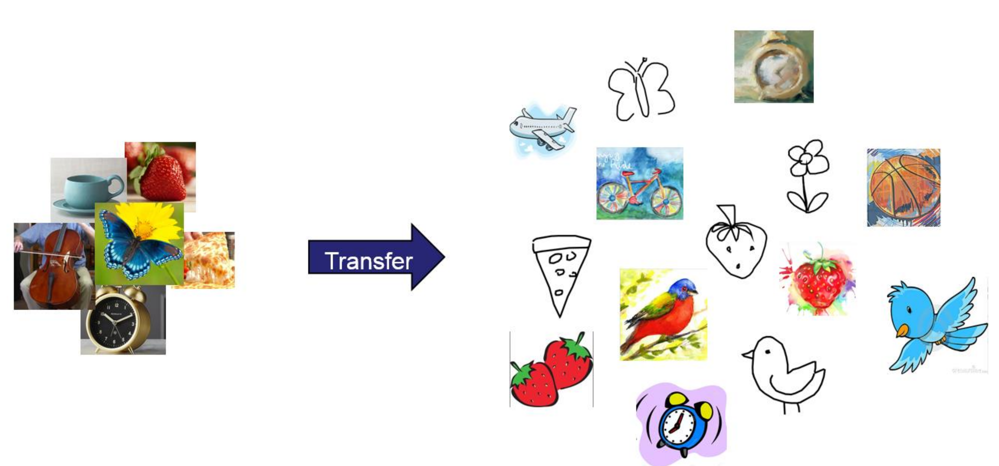
Xingchao Peng[1], Zijun Huang[2], Ximeng Sun[1], Kate Saenko[1]
[1]Boston University [2]Columbia University

Introduction

- **Conventional domain adaptation:**
  Single source domain with labels
  Single target domain without labels
- **Domain Agnostic Learning:**
  Single source domain with labels
  Mixed unlabeled target domain



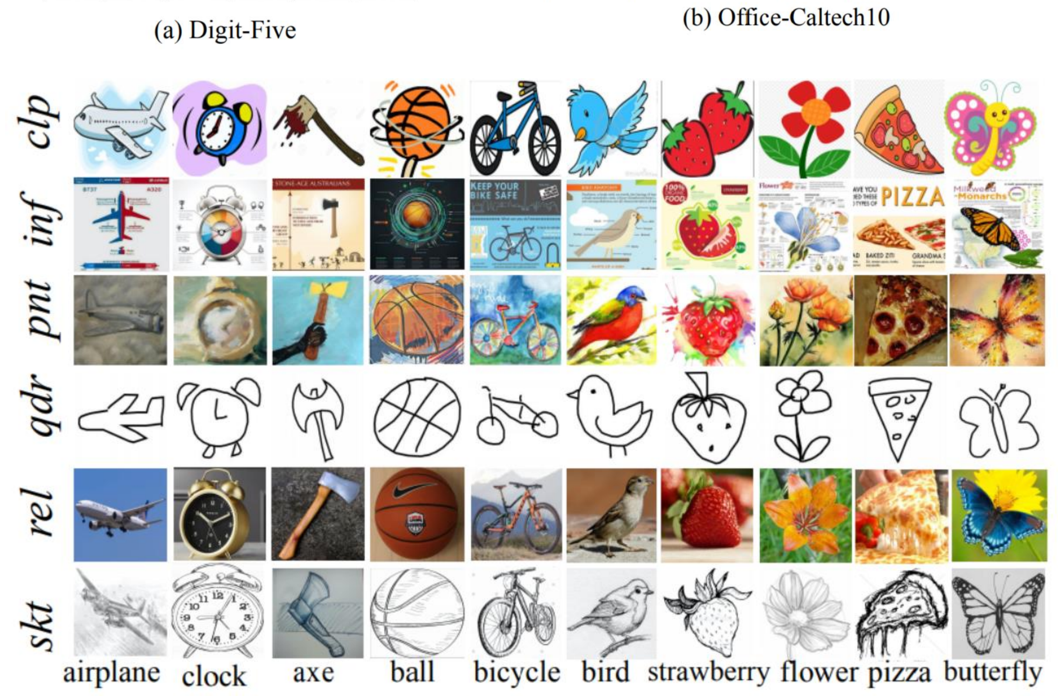
Domain Agnostic Learning



Datasets

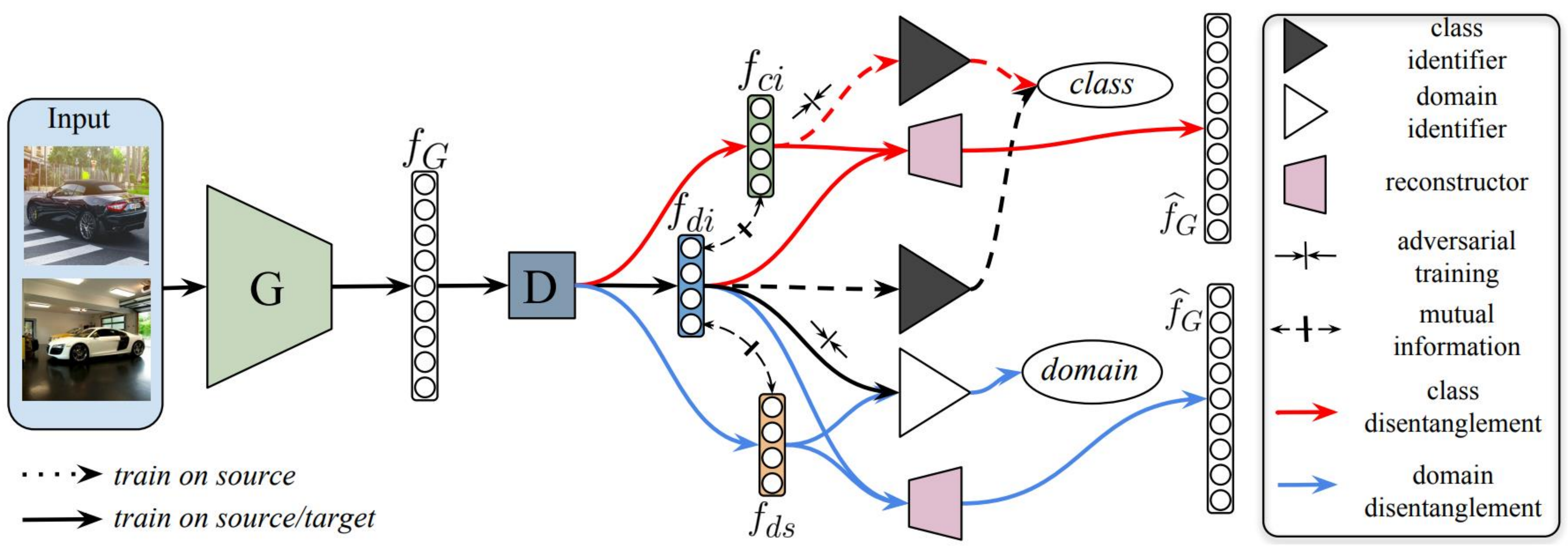
(a) Digit-Five

(b) Office-Caltech10



(c) DomainNet

Deep Adversarial Disentangled Autoencoder



Class Disentanglement:

- Train class identifier:

$$\mathcal{L}_{ce} = -\mathbb{E}_{(x_s,y_s)\sim\hat{\mathcal{D}}_s} \sum_{k=1}^{K} \mathbb{1}[k = y_s] log(C(f_D))$$

- Confuse class identifier:

$$\mathcal{L}_{ent} = -\frac{1}{n_s}\sum_{j=1}^{n_s} \log C(f_{ci}^j) - \frac{1}{n_t}\sum_{j=1}^{n_t} \log C(f_{ci}^j)$$

Domain Disentanglement:

- Adversarial loss:

$$\mathcal{L}_{DI} = -\mathbb{E}[l_f \log P(l_f)] - \mathbb{E}(1-l_f)[\log P(1-l_f)]$$

- Feature Reconstruction

$$\mathcal{L} = \|\hat{f}_G - f_G\|_F^2$$

Mutual Information Minimization:

$$I(\mathcal{D}_x; \mathcal{D}_{f_{di}}) = \int_{\mathbb{X}\times\mathbb{Z}} \log \frac{d\mathbb{P}_{XZ}}{d\mathbb{P}_X \otimes \mathbb{P}_Z} d\mathbb{P}_{XZ}$$

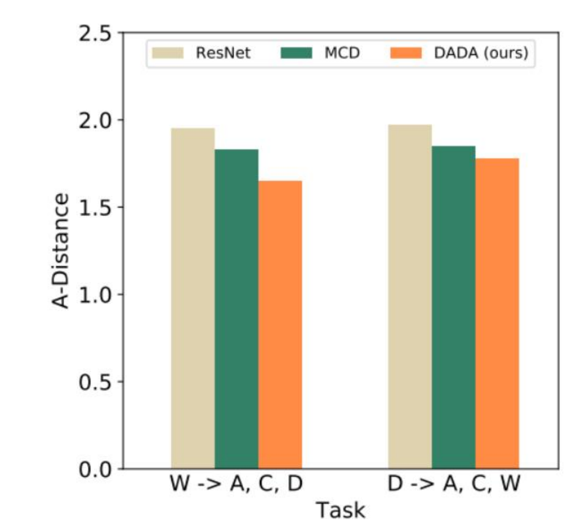
Ring-style Normalization

$$\mathcal{L}_{ring} = \frac{1}{2n}\sum_{i=1}^{n}(\|T(x_i)\|_2 - R)^2$$

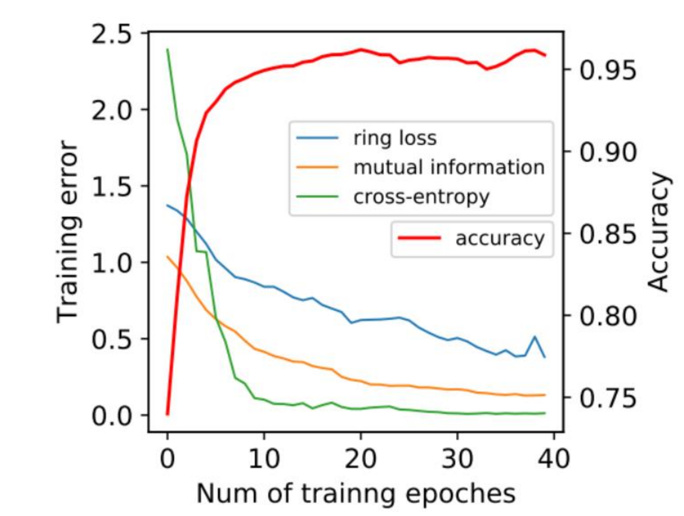
Experiments on Office-Caltech10

Table 2. Accuracy on *Office-Caltech10* dataset with DAL protocal. The methods in the above table are based on "AlexNet" backbone and the methods below are based on the "ResNet" backbone. For both backbones, our model outperforms other baselines.

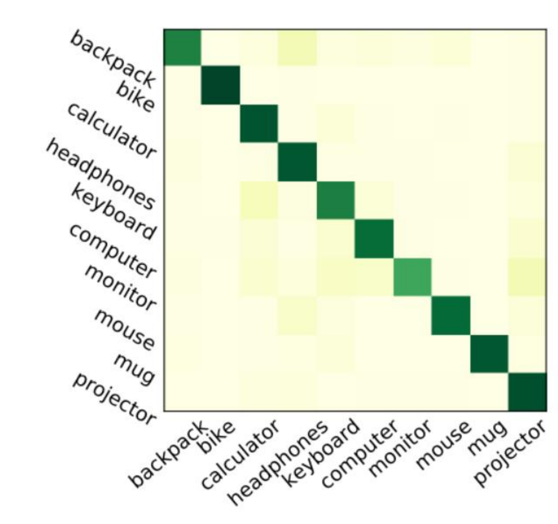
Method	A → C,D,W	C → A,D,W	D → A,C,W	W → A,C,D	Average
AlexNet (Krizhevsky et al., 2012)	83.1±0.2	88.9±0.4	86.7±0.4	82.2±0.3	85.2
DAN (Long et al., 2015)	82.5±0.3	86.2±0.4	75.7±0.5	80.4±0.2	81.2
RTN (Long et al., 2016)	85.2±0.4	89.8±0.3	81.7±0.3	83.7±0.4	85.1
JAN (Long et al., 2017)	83.5±0.3	88.5±0.2	80.1±0.3	85.9±0.4	84.5
DANN (Ganin & Lempitsky, 2015)	85.9±0.4	90.5±0.3	88.6±0.4	90.4±0.2	88.9
DADA (Ours)	86.3±0.3	91.7±0.4	89.9±0.3	91.3±0.3	**89.8**
ResNet (He et al., 2016)	90.5±0.3	94.3±0.2	88.7±0.4	82.5±0.3	89.0
SE (French et al., 2018)	90.3±0.4	94.7±0.4	88.5±0.3	85.3±0.4	89.7
MCD (Saito et al., 2018)	91.7±0.4	**95.3**±0.3	89.5±0.2	84.3±0.2	90.2
DANN (Ganin & Lempitsky, 2015)	91.5±0.4	94.3±0.4	90.5±0.3	86.3±0.3	90.6
DADA (Ours)	**92.0**±0.4	95.1±0.3	**91.3**±0.4	**93.1**±0.3	**92.9**



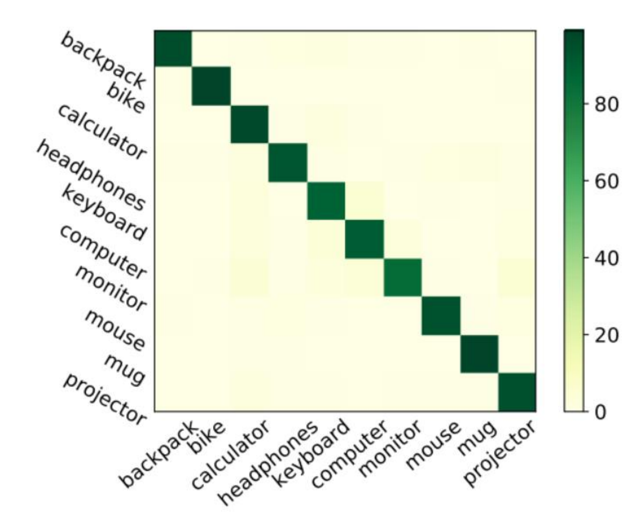
(a) $\mathcal{A}$-Distance



(b) Training loss for C→A,D,W



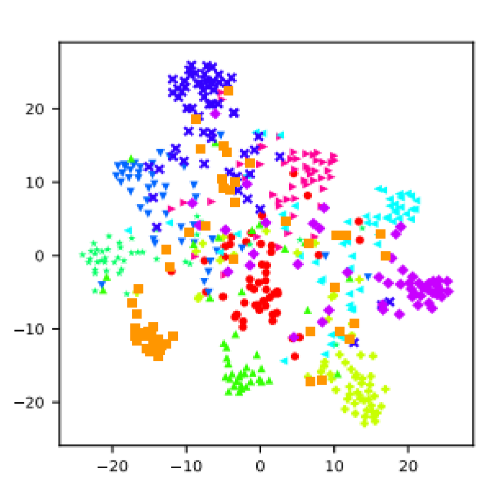
(c) MCD confusion matrix



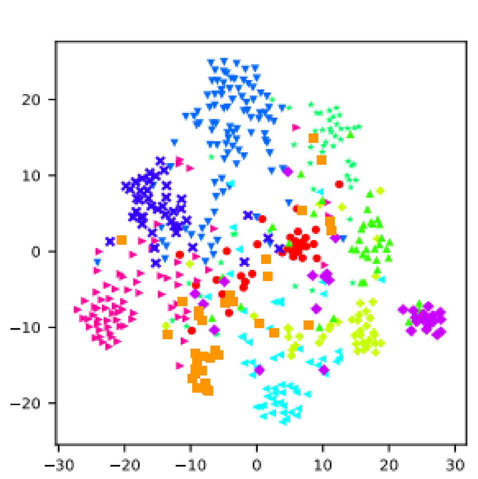
(d) DADA confusion matrix

Experiments on Digit-Five dataset

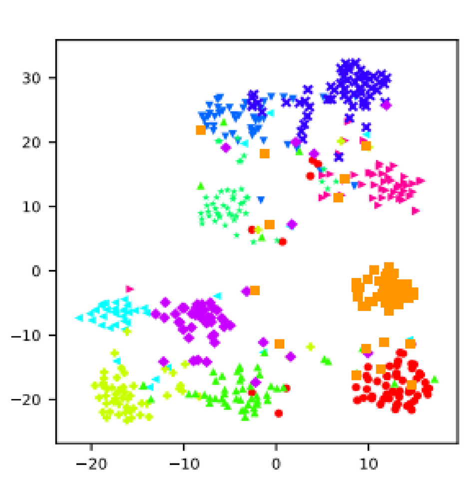
Models	mt→mm,sv,sy,up	mm→mt,sv,sy,up	sv→mt,mm,sy,up	sy→mt,mm,sv,up	up→mt,mm,sv,sy	Avg
Source Only	20.5±1.2	53.5±0.9	62.9±0.3	77.9±0.4	22.6±0.4	47.5
DAN (Long et al., 2015)	21.7±1.0	55.3±0.7	63.2±0.5	79.3±0.2	40.2±0.4	51.9
DANN (Ganin & Lempitsky, 2015)	22.8±1.1	45.2±0.6	61.8±0.2	79.3±0.3	38.7±0.6	49.6
ADDA (Tzeng et al., 2017)	23.4±1.3	54.8±0.8	63.5±0.4	79.6±0.3	43.5±0.5	52.9
UFDN (Liu et al., 2018a)	20.2±1.5	41.6±0.7	64.5±0.4	60.7±0.3	44.6±0.2	46.3
MCD (Saito et al., 2018)	28.7±1.3	43.8±0.8	75.1±0.3	78.9±0.3	**55.3**±0.4	56.4
DADA+*class* (I)	28.9±1.2	50.1±0.9	65.4±0.2	79.8±0.1	50.4±0.3	54.9
DADA+*domain* (II)	34.1±1.7	57.1±0.4	71.3±0.4	82.5±0.3	45.4±0.4	57.5
DADA+*ring* (III)	35.3±1.5	57.5±0.6	80.1±0.3	82.9±0.2	46.2±0.3	60.4
DADA+*rec* (IV)	**39.4**±1.4	**61.1**±0.7	**80.1**±0.4	**83.7**±0.2	47.2±0.4	**62.3**



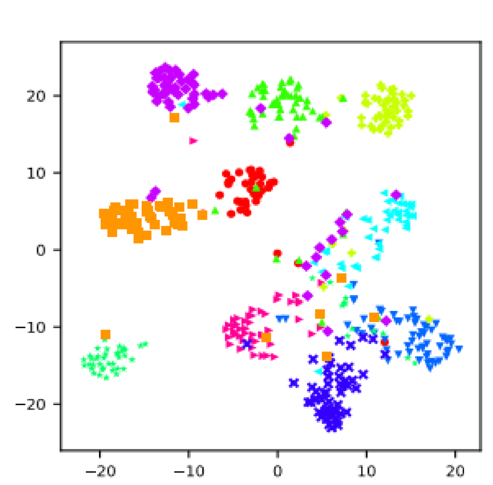
(a) Source Features



(b) UFDN Features



(c) MCD Features



(d) DADA Features

Figure 3. Feature visualization: t-SNE plot of source features, UFDN (Liu et al., 2018a) features, MCD (Saito et al., 2018) features and DADA features on agnostic target domain in *sv →mm,mt,up,sy* setting. We use different markers and different colors to denote different categories. (Best viewed in color.)

Experiment on DomainNet

Table 3. Accuracy on the DomainNet dataset (Peng et al., 2018) dataset with DAL protocol. The table below shows the results based on AlexNet (Krizhevsky et al., 2012) backbone and the below are the results of ResNet (He et al., 2016) backbone. For both setting, our model outperforms other baselines.

Models	clp→inf,pnt qdr,rel,skt	inf→clp,pnt, qdr,rel,skt	pnt→clp,inf, qdr,rel,skt	qdr→clp,inf, pnt,rel,skt	rel→clp,inf, pnt,qdr,skt	skt→clp,inf, pnt,qdr,rel	Avg
AlexNet (Krizhevsky et al., 2012)	22.5±0.4	15.3±0.2	21.2±0.3	6.0±0.2	17.2±0.3	21.8±0.3	17.3
DAN (Long et al., 2015)	23.7±0.3	14.9±0.4	22.7±0.2	7.6±0.3	19.4±0.4	23.4±0.5	18.6
RTN (Long et al., 2016)	21.4±0.3	14.2±0.3	21.0±0.4	7.7±0.2	17.8±0.3	20.8±0.4	17.2
JAN (Long et al., 2017)	21.1±0.4	16.5±0.2	21.6±0.3	9.9±0.1	15.4±0.2	22.5±0.3	17.8
DANN (Ganin & Lempitsky, 2015)	24.1±0.2	15.2±0.4	24.5±0.3	8.2±0.4	18.0±0.3	24.1±0.4	19.1
DADA (Ours)	23.9±0.4	17.9±0.4	25.4±0.5	9.4±0.2	20.5±0.3	**25.2**±0.4	20.4
ResNet101 (He et al., 2016)	25.6±0.2	16.8±0.3	25.8±0.4	9.2±0.2	20.6±0.5	22.3±0.1	20.1
SE (French et al., 2018)	21.3±0.2	8.5±0.1	14.5±0.2	**13.8**±0.4	16.0±0.4	19.7±0.2	15.6
MCD (Saito et al., 2018)	25.1±0.3	19.1±0.4	**27.0**±0.3	10.4±0.3	20.2±0.2	22.5±0.4	20.7
DADA (Ours)	**26.1**±0.4	**20.0**±0.3	26.5±0.4	12.9±0.4	**20.7**±0.4	22.8±0.2	**21.5**

Conclusion

- Deep Features are high entangled
- Disentangle features to class-irrelevant and domain-specific features
- Disentangle features to domain-specific and domain-invariant features
- Mutual Information Minimization
- Ring Loss Normalization