Multi-Adversarial Domain Adaptation
(2018)

Zhongyi Pei, Zhangjie Cao, Mingsheng Long, and Jianmin Wang

Summary

Contributions

Adversarial learning has been successfully em-bedded into deep networks to learn transferable features to reduce distribution discrepancy between the source and target domains. There are, however, two technical challenges to enabling domain adaptation: (1) enhancing positive transfer by maximally matching the multi mode structures underlying data distributions across domains, and (2) alleviating negative transfer by preventing false alignment of modes in different distributions across domains. Motivated by these challenges, the authors present a multi-adversarial domain adaptation (MADA) approach, which captures multi mode structures to enable fine-grained alignment of different data distributions based on multiple domain discriminators.

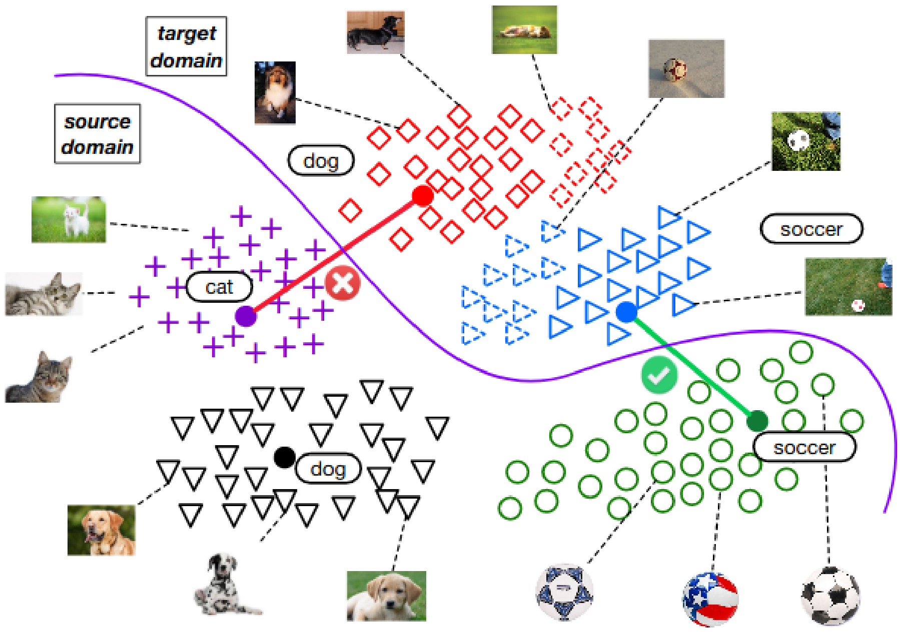


Figure 1: The difficulty of domain adaptation: discriminative structures may be mixed up or falsely aligned across domains. As an intuitive example, in this figure, the source class *cat* is falsely aligned with target class *dog*, making final classification wrong.

Method

In practical domain adaptation problems, the data distributions of the source domain and target domain usually embody complex multimode structures, reflecting either the class boundaries in supervised learning or the cluster boundaries in unsupervised learning. Thus, previous domain adversarial adaptation methods that only match the data distributions without exploiting the multimode structures maybe prone to either under transfer or negative transfer. Under transfer may happen when different modes of the distributions cannot be maximally matched. Negative transfer may happen when the corresponding modes of the distributions across domains are falsely aligned.

Given that the source domain labeled information provides strong signals to reveal the multimode structures, the authors propose to split the traditional domain discriminator into K domain discriminators, each is responsible for matching the source and target domain data associated with class k, and to select the correct discriminator, we use the output of the label predictor in the form of a probability distribution $\hat{\mathbf{y}}_i = G_y\left(\mathbf{x}_i\right)$, the loss in this case can be written as follows.

$$L_d = \frac{1}{n}\sum_{k=1}^{K}\sum_{\mathbf{x}_i\in\mathcal{D}_s\cup\mathcal{D}_t} L_d^k\left(G_d^k\left(\hat{y}_i^k G_f\left(\mathbf{x}_i\right)\right), d_i\right)$$

The total loss is then computed as follows:

$$C\left(\theta_f, \theta_y, \theta_d^k\big|_{k=1}^{K}\right) = \frac{1}{n_s}\sum_{\mathbf{x}_i\in\mathcal{D}} L_y\left(G_y\left(G_f\left(\mathbf{x}_i\right)\right), y_i\right) - \frac{\lambda}{n}\sum_{k=1}^{K}\sum_{\mathbf{x}_i\in\mathcal{D}} L_d^k\left(G_d^k\left(\hat{y}_i^k G_f\left(\mathbf{x}_i\right)\right), d_i\right)$$

Results

Table 1: Accuracy (%) on *Office-31* for unsupervised domain adaptation (AlexNet and ResNet)

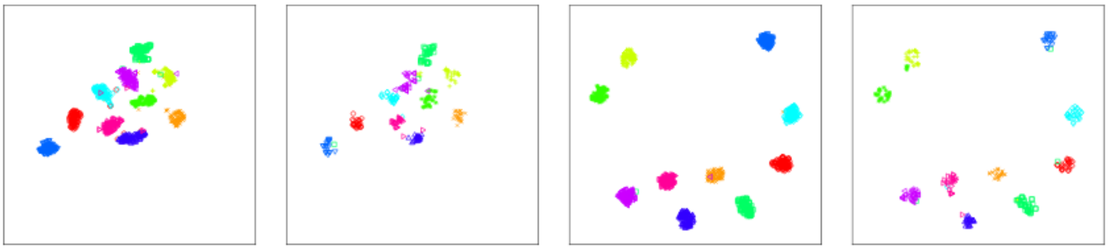
Method	A → W	D → W	W → D	A → D	D → A	W → A	Avg
AlexNet (Krizhevsky, Sutskever, and Hinton 2012)	60.6±0.4	95.4±0.2	99.0±0.1	64.2±0.3	45.5±0.5	48.3±0.5	68.8
TCA (Pan et al. 2011)	59.0±0.0	90.2±0.0	88.2±0.0	57.8±0.0	51.6±0.0	47.9±0.0	65.8
GFK (Gong et al. 2012)	58.4±0.0	93.6±0.0	91.0±0.0	58.6±0.0	52.4±0.0	46.1±0.0	66.7
DDC (Tzeng et al. 2014)	61.0±0.5	95.0±0.3	98.5±0.3	64.9±0.4	47.2±0.5	49.4±0.4	69.3
DAN (Long et al. 2015)	68.5±0.3	96.0±0.1	99.0±0.1	66.8±0.2	50.0±0.4	49.8±0.3	71.7
RTN (Long et al. 2016)	73.3±0.2	96.8±0.2	99.6±0.1	71.0±0.2	50.5±0.3	51.0±0.1	73.7
RevGrad (Ganin and Lempitsky 2015)	73.0±0.5	96.4±0.3	99.2±0.3	72.3±0.3	52.4±0.4	50.4±0.5	74.1
MADA	**78.5**±0.2	**99.8**±0.1	**100.0**±.0	**74.1**±0.1	**56.0**±0.2	**54.5**±0.3	**77.1**
ResNet (He et al. 2016)	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
TCA (Pan et al. 2011)	74.7±0.0	96.7±0.0	99.6±0.0	76.1±0.0	63.7±0.0	62.9±0.0	79.3
GFK (Gong et al. 2012)	74.8±0.0	95.0±0.0	98.2±0.0	76.5±0.0	65.4±0.0	63.0±0.0	78.8
DDC (Tzeng et al. 2014)	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 (Long et al. 2015)	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
RTN (Long et al. 2016)	84.5±0.2	96.8±0.1	99.4±0.1	77.5±0.3	66.2±0.2	64.8±0.3	81.6
RevGrad (Ganin and Lempitsky 2015)	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
MADA	**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**

Table 2: Accuracy (%) on *ImageCLEF-DA* for unsupervised domain adaptation (AlexNet and ResNet)

Method	I → P	P → I	I → C	C → I	C → P	P → C	Avg
AlexNet (Krizhevsky, Sutskever, and Hinton 2012)	66.2±0.2	70.0±0.2	84.3±0.2	71.3±0.4	59.3±0.5	84.5±0.3	73.9
DAN (Long et al. 2015)	67.3±0.2	80.5±0.3	87.7±0.3	76.0±0.3	61.6±0.3	88.4±0.2	76.9
RTN (Long et al. 2016)	67.4±0.3	82.3±0.3	89.5±0.4	78.0±0.2	63.0±0.2	90.1±0.1	78.4
RevGrad (Ganin and Lempitsky 2015)	66.5±0.5	81.8±0.4	89.0±0.5	79.8±0.5	63.5±0.4	88.7±0.4	78.2
MADA	**68.3**±0.3	**83.0**±0.1	**91.0**±0.2	**80.7**±0.2	**63.8**±0.2	**92.2**±0.3	**79.8**
ResNet (He et al. 2016)	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
DAN (Long et al. 2015)	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
RTN (Long et al. 2016)	**75.6**±0.3	86.8±0.1	95.3±0.1	86.9±0.3	72.7±0.3	92.2±0.4	84.9
RevGrad (Ganin and Lempitsky 2015)	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
MADA	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**

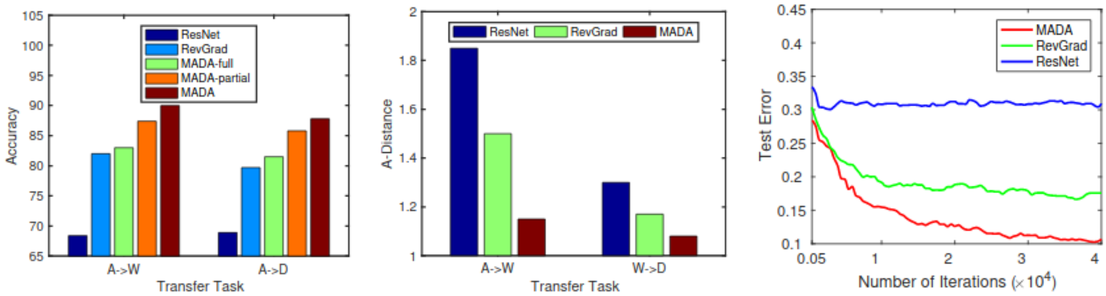
Table 3: Accuracy (%) on *Office-31* for domain adaptation from 31 classes to 25 classes (AlexNet)

Method	A → W	D → W	W → D	A → D	D → A	W → A	Avg
AlexNet (Krizhevsky, Sutskever, and Hinton 2012)	58.2±0.4	95.9±0.2	99.0±0.1	60.4±0.3	49.8±0.5	47.3±0.5	68.4
RevGrad (Ganin and Lempitsky 2015)	65.1±0.5	91.7±0.3	97.1±0.3	60.6±0.3	42.1±0.4	42.9±0.5	66.6
MADA	**70.8**±0.2	**96.6**±0.1	**99.5**±.0	**69.6**±0.1	**51.4**±0.2	**54.2**±0.3	**73.7**



(a) RevGrad: source=**A** (b) RevGrad: target=**W** (c) MADA: source=**A** (d) MADA: target=**W**

Figure 3: The t-SNE visualization of deep features extracted by RevGrad (a)(b) and MADA (c)(d).



(a) Sharing Strategies (b) $\mathcal{A}$-distance (c) Convergence

Figure 4: Empirical analysis: (a) Sharing strategies, (b) $\mathcal{A}$-distance, and (c) Convergence performance.