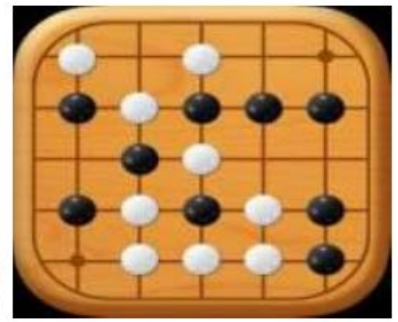
Deep Visual Domain Adaptation： A Survey

Content：

1. Introduction
    - Transfer Learning
    - Domain adaptation
2. Method
3. Experiment

Why do we need transfer learning

Machine do have its weakness, it has no ability to "transfer learning". The trained models can not be adopted in different related scenarios, for example AlphaGo can't play Chinese chess.

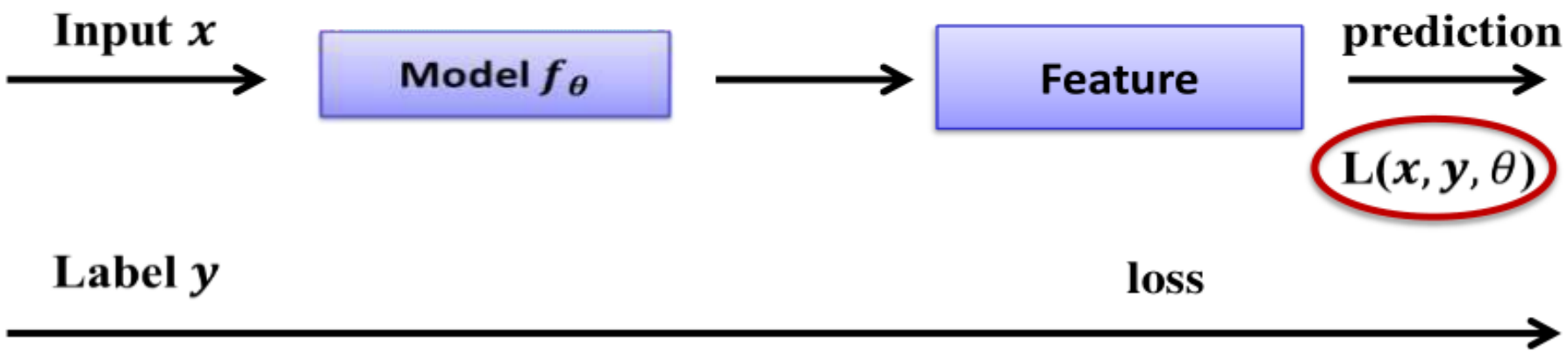


In the human evolution, the ability of transfer learning is very important. We can extend learned knowledge to other scenarios. For example, after learning riding bicycles, it is very easy to ride motorcycles.

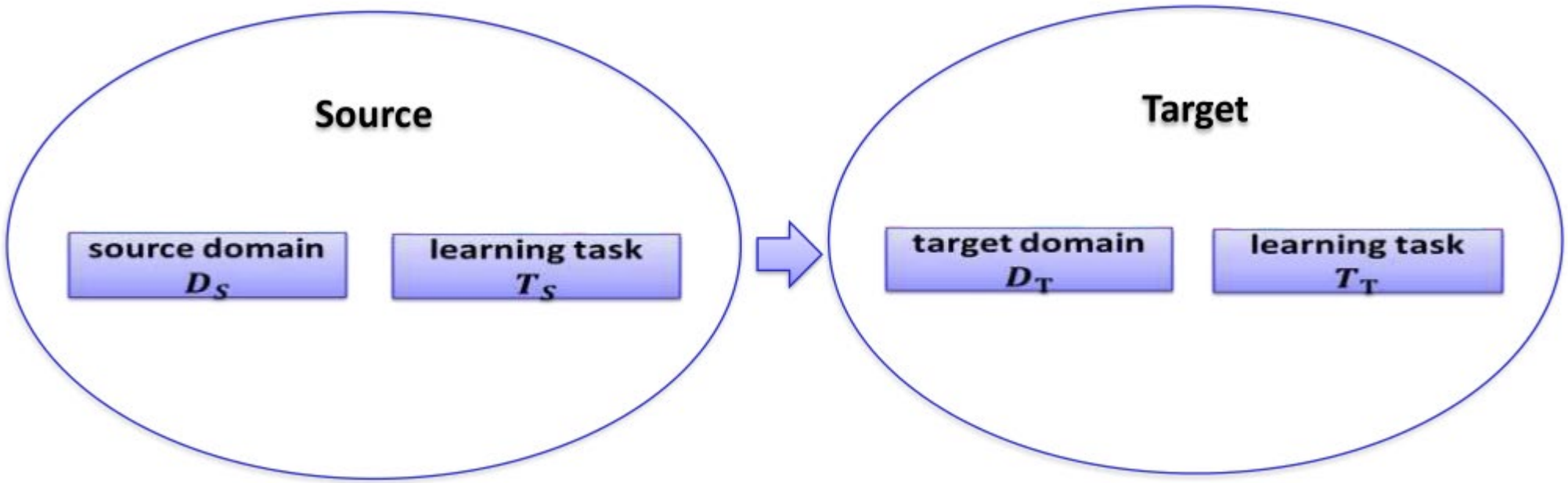
What is transfer learning?

- **Traditional deep learning when training and testing share similar distribution:**

$$\text{Empirical Risk: } \min \frac{1}{n}\sum_{i=1}^{n} L(x_i, y_i, \theta)$$

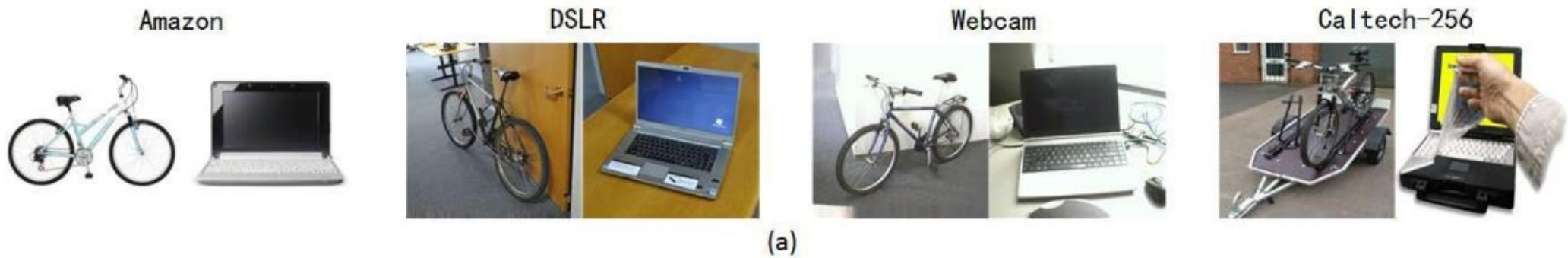


Given a source domain D_S and learning task T_S, a target domain D_T and learning task T_T, transfer learning aims to help improve the learning of the target predictive function $f_T(\cdot)$ in D_T using the knowledge in D_S and T_S, where $D_S \neq D_T$, or $T_S \neq T_T$.

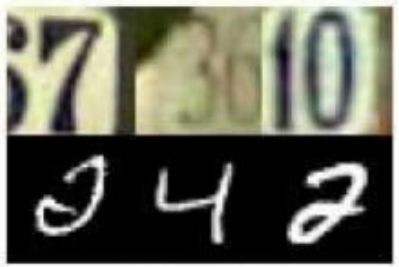


What is domain adaptation?

Due to many factors (e.g., illumination, pose, and image quality), there is always a distribution change or domain shift between two domains that can degrade the performance.

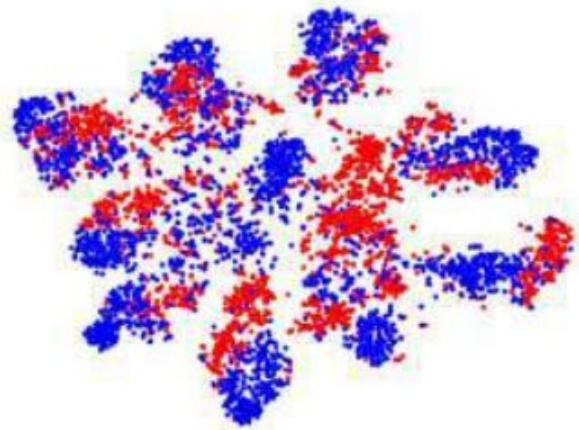


SVHN



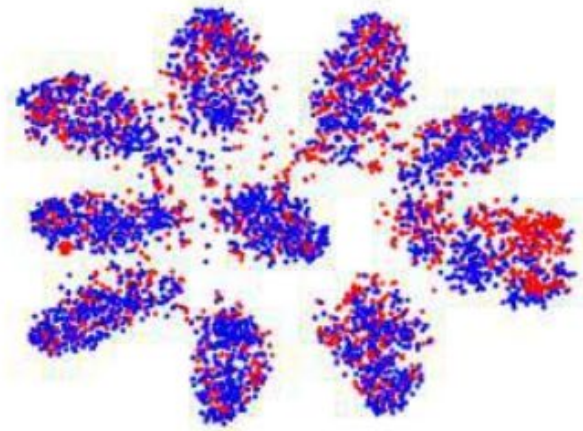
MNIST

SVHN→MNIST



(a) non-adapted

SVHN→MNIST



(a) adapted

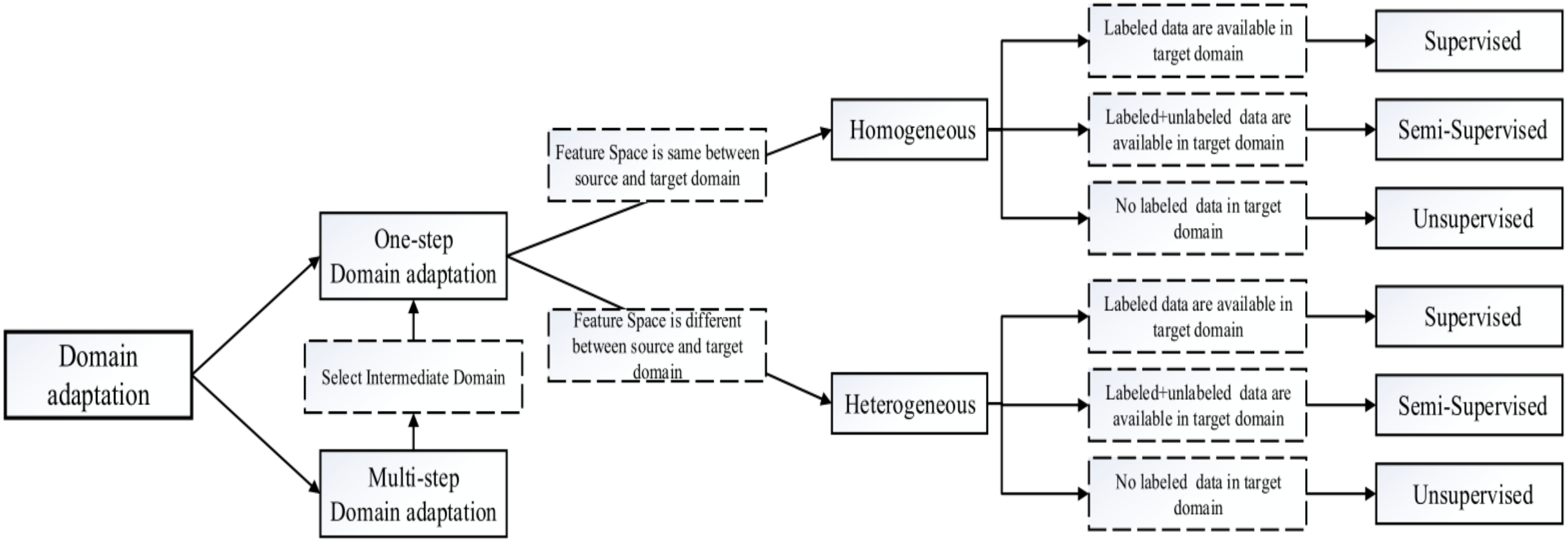


Fig. 2. An overview of different settings of domain adaptation

Method

- **Feature adaptation**

$$\min \frac{1}{n}\sum_{i=1}^{n} L(\phi(\mathbf{x}_i^s), \mathbf{y}_i^s, \theta)$$

- **Instance adaptation**

$$\min \frac{1}{n}\sum_{i=1}^{n} w_i L(\mathbf{x}_i^s, \mathbf{y}_i^s, \theta)$$

- **Model adaptation**

$$\min \frac{1}{n}\sum_{i=1}^{n} L(\mathbf{x}_i^s, \mathbf{y}_i^s, \theta)$$

Feature adaptation

$$\min \frac{1}{n}\sum_{i=1}^{n} L(\boldsymbol{\phi}(\mathbf{x}_i^s), \mathbf{y}_i^s, \boldsymbol{\theta})$$

Maximum mean discrepancy (MMD)

Method 1.DDC

Maximum mean discrepancy (MMD) is a commonly-used statistic loss for unsupervised DA. The hidden representations of images of different domain are embedded in a reproducing kernel Hilbert space, and the mean embeddings of distributions cross domains can be explicitly matched.

1. Given two distributions s and t, the MMD between them is defined as:

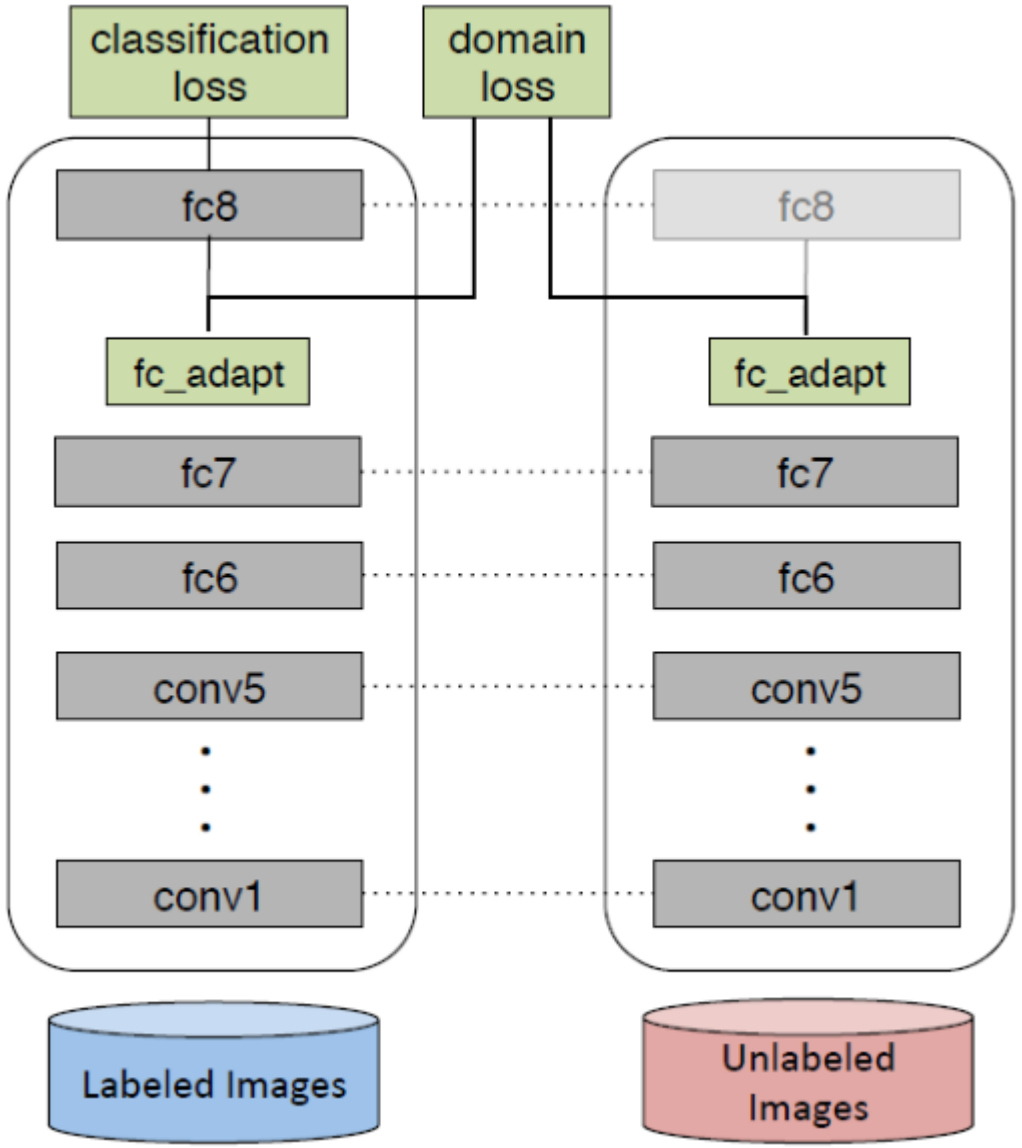
$$MMD^2(s,t) = \sup_{\|\phi\|_{\mathcal{H}} \leq 1} \left\| E_{x^s \sim s}[\phi(x^s)] - E_{x^t \sim t}[\phi(x^t)] \right\|_{\mathcal{H}}^2$$

2. Denote by $\mathcal{D}_s = \{x_i^s\}_{i=1}^M$ and $\mathcal{D}_t = \{x_i^t\}_{i=1}^N$ drawn from the distributions s and t, respectively, an empirical estimate of MMD is given as:

$$MMD^2(\mathcal{D}_s, \mathcal{D}_t) = \left\| \frac{1}{M} \sum_{i=1}^{M} \phi(x_i^s) - \frac{1}{N} \sum_{i=1}^{N} \phi(x_i^t) \right\|_{\mathcal{H}}^2$$

3. The main idea of DDC and DAN is to integrate this MMD estimator:

$$\mathcal{L} = \mathcal{L}_C(X_s, y) + \lambda \sum_{l \in \mathcal{L}} L_M(D_s^l, D_t^l)$$


classification loss
domain loss
fc8
fc8
fc_adapt
fc_adapt
fc7
fc7
fc6
fc6
conv5
conv5
conv1
conv1
Labeled Images
Unlabeled Images

Experiment

A

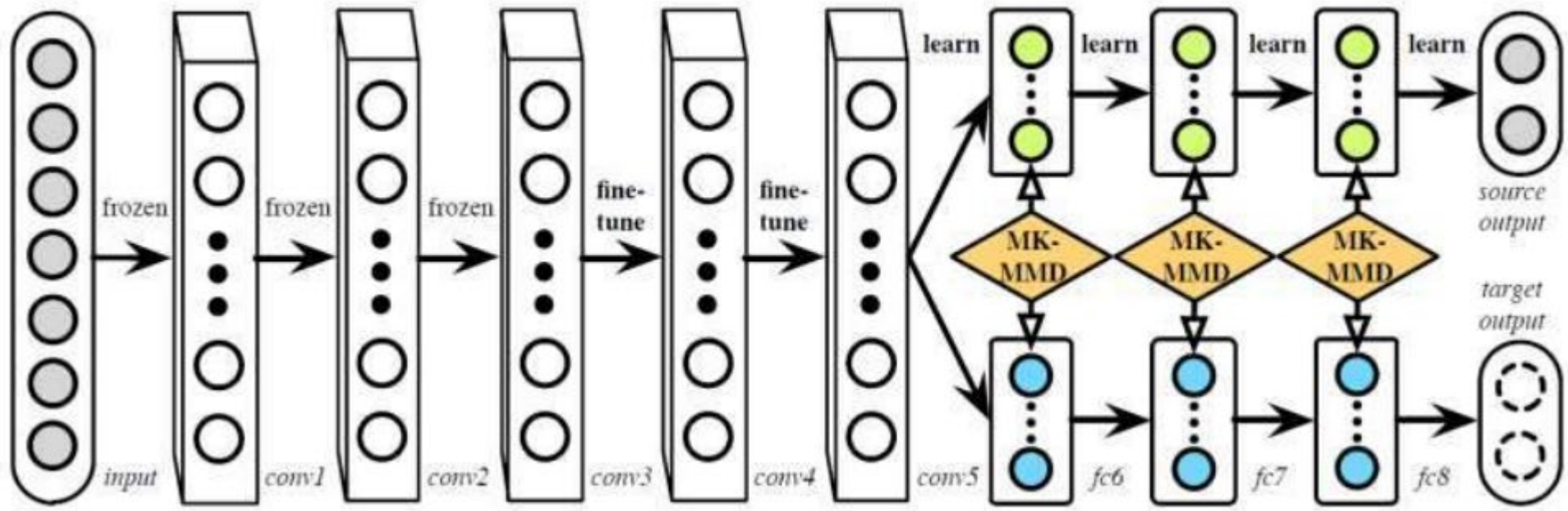
D

W

	$A \rightarrow W$	$D \rightarrow W$	$W \rightarrow D$	Average
GFK(PLS,PCA) [16]	15.0 ± 0.4	44.6 ± 0.3	49.7 ± 0.5	36.4
SA [13]	15.3	50.1	56.9	40.8
DA-NBNN [31]	23.3 ± 2.7	67.2 ± 1.9	67.4 ± 3.0	52.6
DLID [8]	26.1	68.9	84.9	60.0
DeCAF$_6$ S [11]	52.2 ± 1.7	91.5 ± 1.5	–	–
DaNN [14]	35.0 ± 0.2	70.5 ± 0.0	74.3 ± 0.0	59.9
Ours	$\mathbf{59.4 \pm 0.8}$	$\mathbf{92.5 \pm 0.3}$	$\mathbf{91.7 \pm 0.8}$	**81.2**

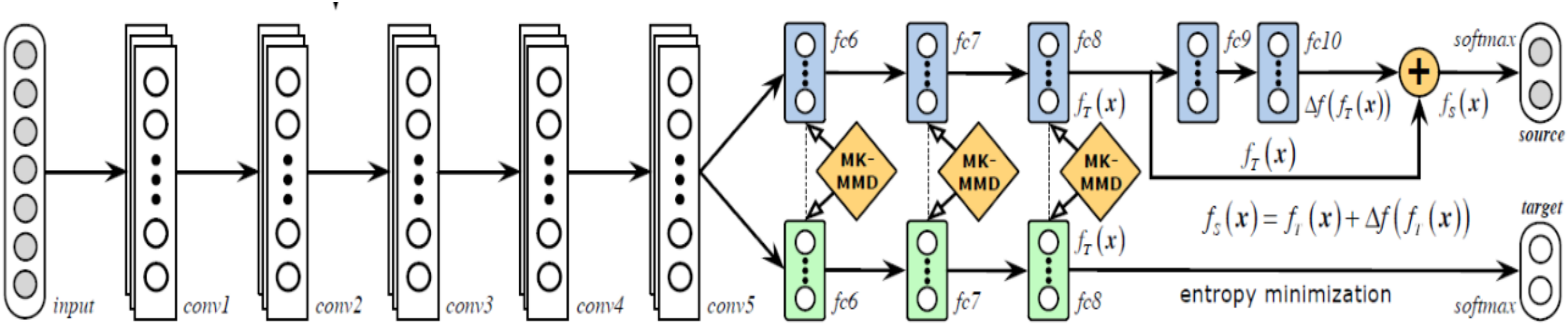
Method 2.DAN

DAN that matches the shift in marginal distributions across domains by adding multiple adaptation layers and exploring multiple kernels, assuming that the conditional distributions remain unchanged.



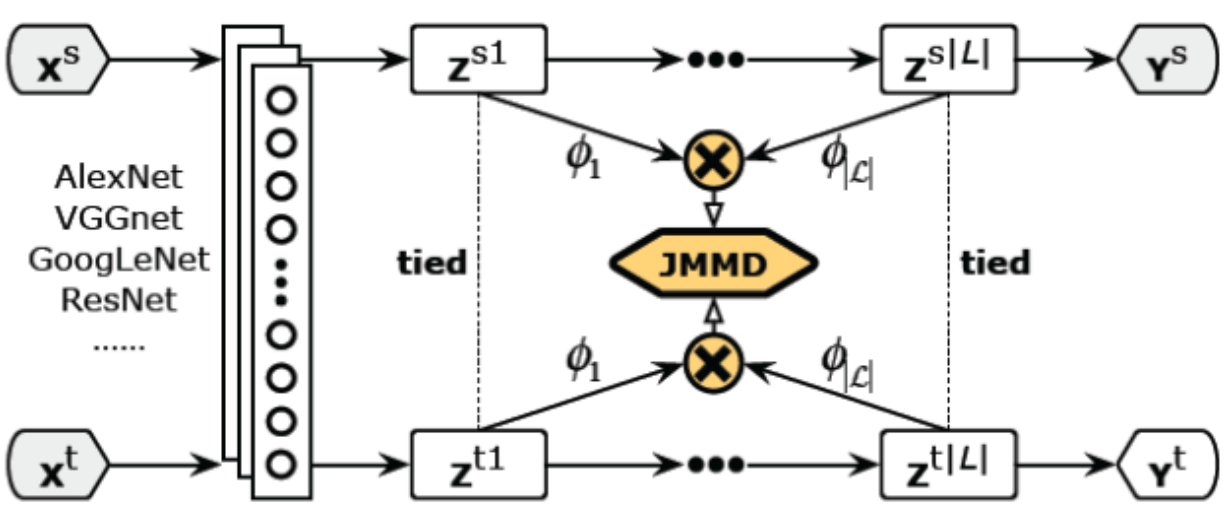
Method 3.RTN

Since deep features eventually transition from general to specific along the network:
(1) fully connected layers fc6-fc8 are tailored to model task-specific structures, hence they are not safely transferable and should be adapted with MK-MMD minimization;
(2) supervised classifiers are not safely transferable, hence they are bridged by the residual layers fc9-fc10 such that $f_S(x) = f_T(x) + \Delta f(f_T(x))$

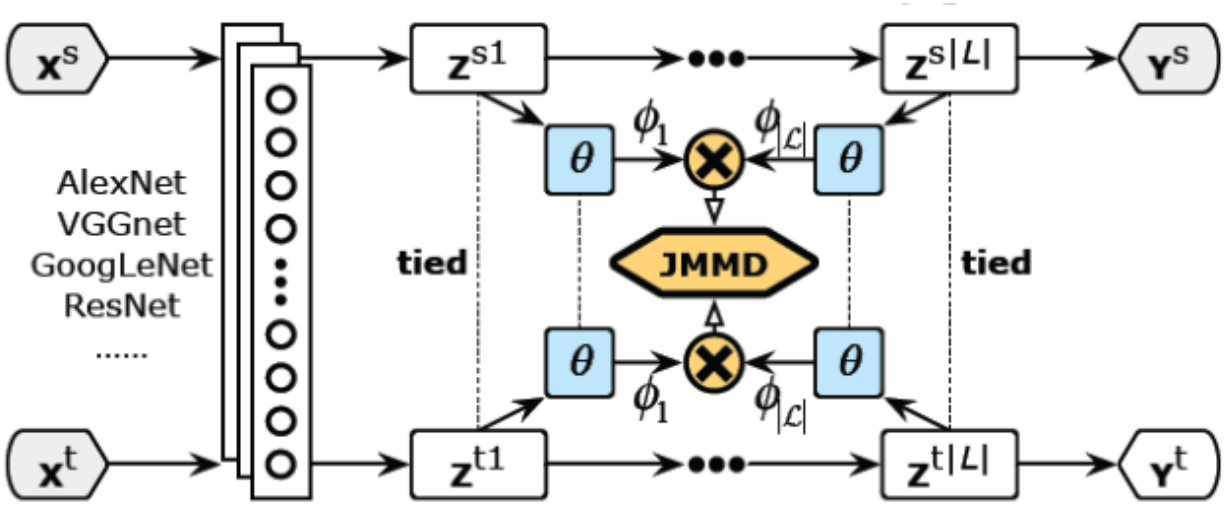


Method 4.JAN

Transfer learning will become more challenging as domains may change by the joint distributions P(X,Y) of input features X and output labels Y. The distribution shifts may stem from the marginal distributions P(X), the conditional distributions P(Y|X), or both.



(a) Joint Adaptation Network (JAN)



(b) Adversarial Joint Adaptation Network (JAN-A)

Experiment

Table 1. Classification accuracy (%) on *Office-31* dataset for unsupervised domain adaptation (AlexNet and ResNet)

Method	A → W	D → W	W → D	A → D	D → A	W → A	Avg
AlexNet (Krizhevsky et al., 2012)	61.6±0.5	95.4±0.3	99.0±0.2	63.8±0.5	51.1±0.6	49.8±0.4	70.1
TCA (Pan et al., 2011)	61.0±0.0	93.2±0.0	95.2±0.0	60.8±0.0	51.6±0.0	50.9±0.0	68.8
GFK (Gong et al., 2012)	60.4±0.0	95.6±0.0	95.0±0.0	60.6±0.0	52.4±0.0	48.1±0.0	68.7
DDC (Tzeng et al., 2014)	61.8±0.4	95.0±0.5	98.5±0.4	64.4±0.3	52.1±0.6	52.2±0.4	70.6
DAN (Long et al., 2015)	68.5±0.5	96.0±0.3	99.0±0.3	67.0±0.4	54.0±0.5	53.1±0.5	72.9
RTN (Long et al., 2016)	73.3±0.3	**96.8**±0.2	**99.6**±0.1	71.0±0.2	50.5±0.3	51.0±0.1	73.7
RevGrad (Ganin & Lempitsky, 2015)	73.0±0.5	96.4±0.3	99.2±0.3	72.3±0.3	53.4±0.4	51.2±0.5	74.3
JAN (ours)	74.9±0.3	96.6±0.2	99.5±0.2	71.8±0.2	**58.3**±0.3	55.0±0.4	76.0
JAN-A (ours)	**75.2**±0.4	96.6±0.2	**99.6**±0.1	**72.8**±0.3	57.5±0.2	**56.3**±0.2	**76.3**
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)	72.7±0.0	96.7±0.0	99.6±0.0	74.1±0.0	61.7±0.0	60.9±0.0	77.6
GFK (Gong et al., 2012)	72.8±0.0	95.0±0.0	98.2±0.0	74.5±0.0	63.4±0.0	61.0±0.0	77.5
DDC (Tzeng et al., 2014)	75.6±0.2	96.0±0.2	98.2±0.1	76.5±0.3	62.2±0.4	61.5±0.5	78.3
DAN (Long et al., 2015)	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
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 & 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
JAN (ours)	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
JAN-A (ours)	**86.0**±0.4	96.7±0.3	99.7±0.1	**85.1**±0.4	**69.2**±0.4	**70.7**±0.5	**84.6**

REFERENCES

DDC： Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., and Darrell, T. Deep domain confusion: Maximizing for domain invariance. CoRR, abs/1412.3474, 2014.

DAN： M. Long , Y. Cao , J. Wang , M. Jordan , Learning transferable features with deep adaptation networks, in ICML, 2015,pp. 97–105 .

RTN： M. Long , H. Zhu , J. Wang , M.I. Jordan , Unsupervised domain adaptation with residual transfer networks, in NIPS,
2016, pp. 136–144 .

JAN： M. Long, J. Wang, and M. I. Jordan. Deep transfer learning with joint adaptation networks. arXiv preprint arXiv:1605.06636, 2016.

Thanks