Multi-Adversarial Domain Adaption

Peng Li

10175501102@stu.ecnu.edu.cn

2019/05/21

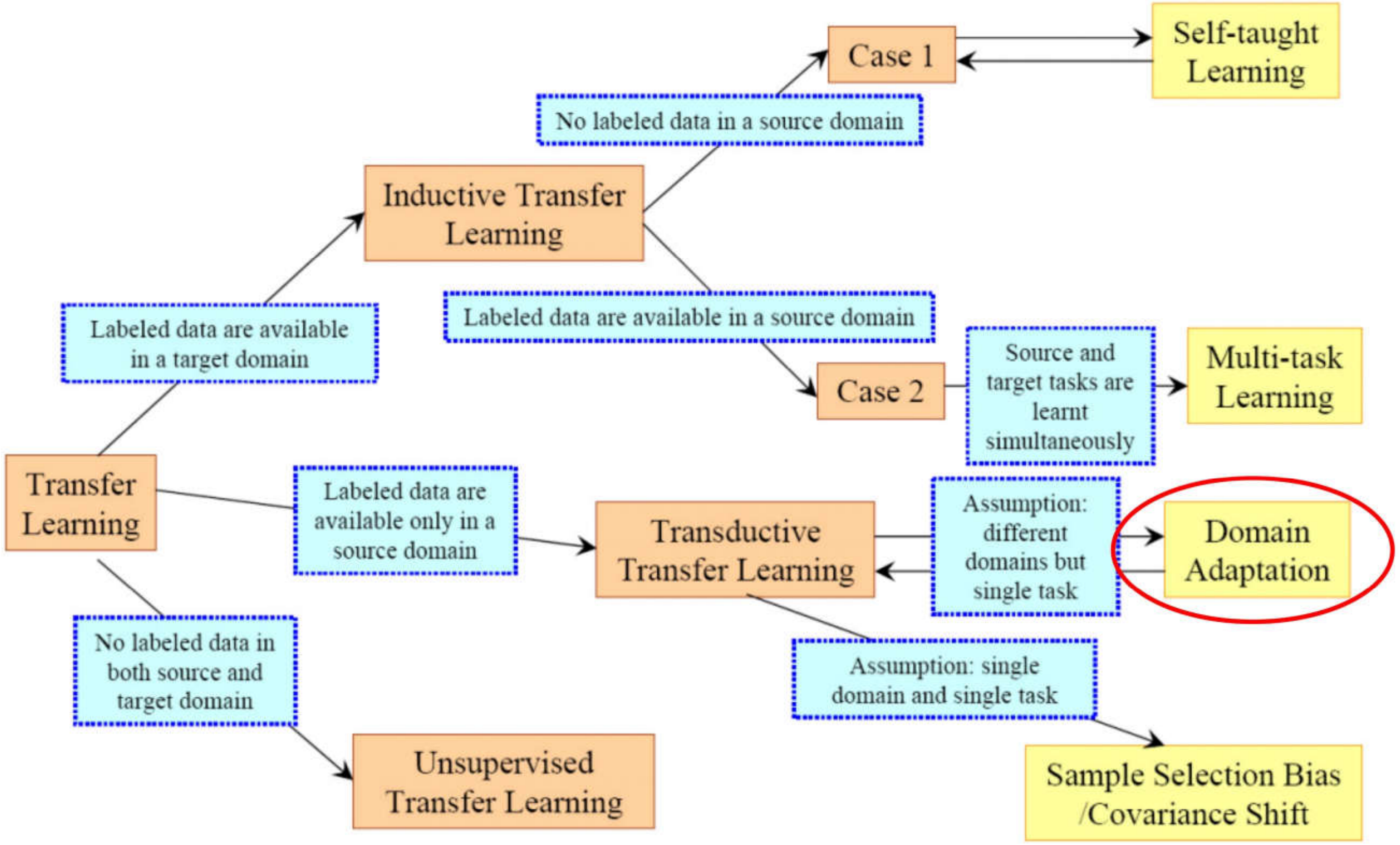
Paper & Codes

Content

- Multi-Adversarial Domain Adaption
- Multi-Adversarial Domain Adaption
- Multi-Adversarial Domain Adaption

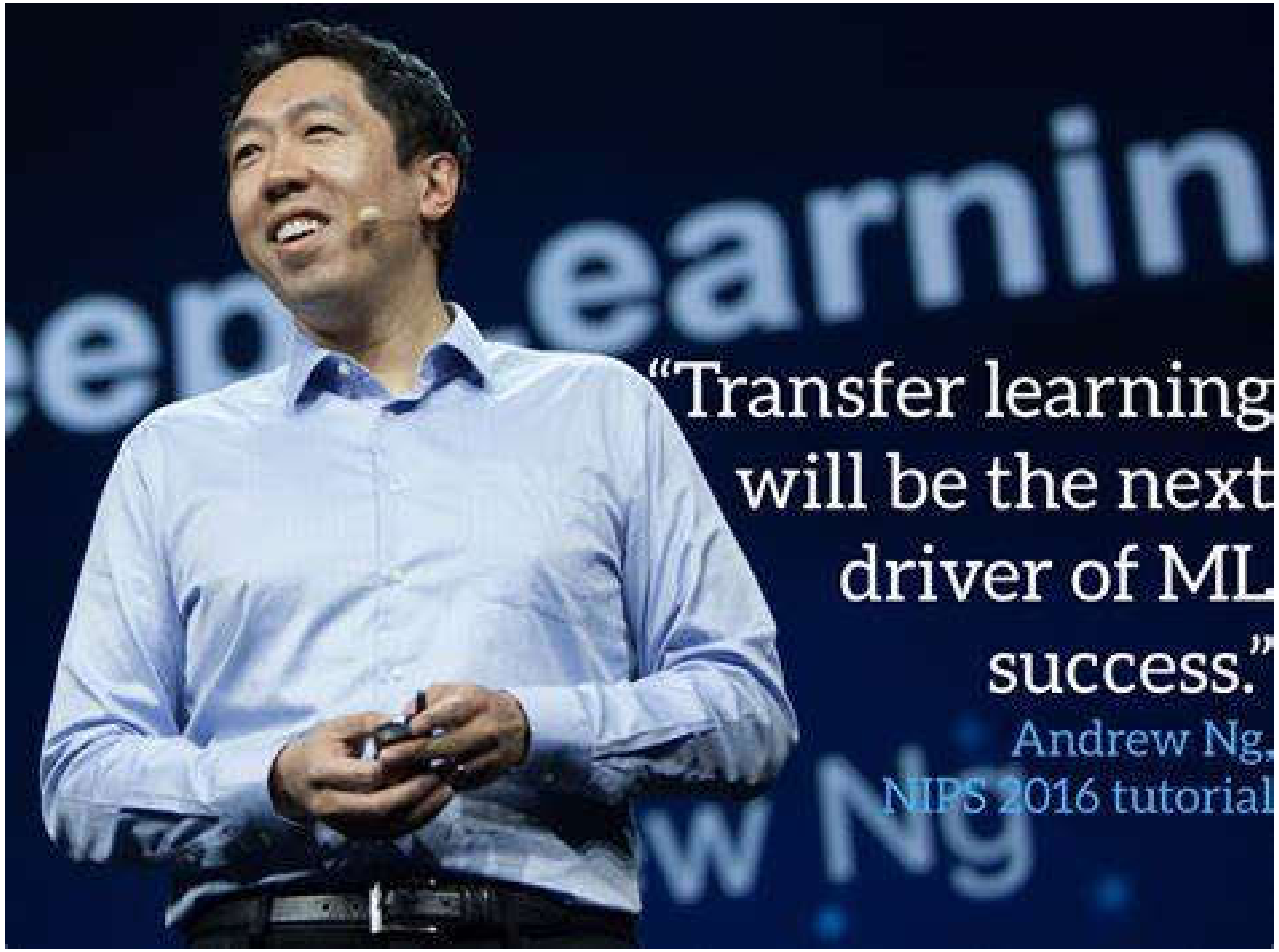
Multi-Adversarial Domain Adaption

Multi-Adversarial Domain Adaption



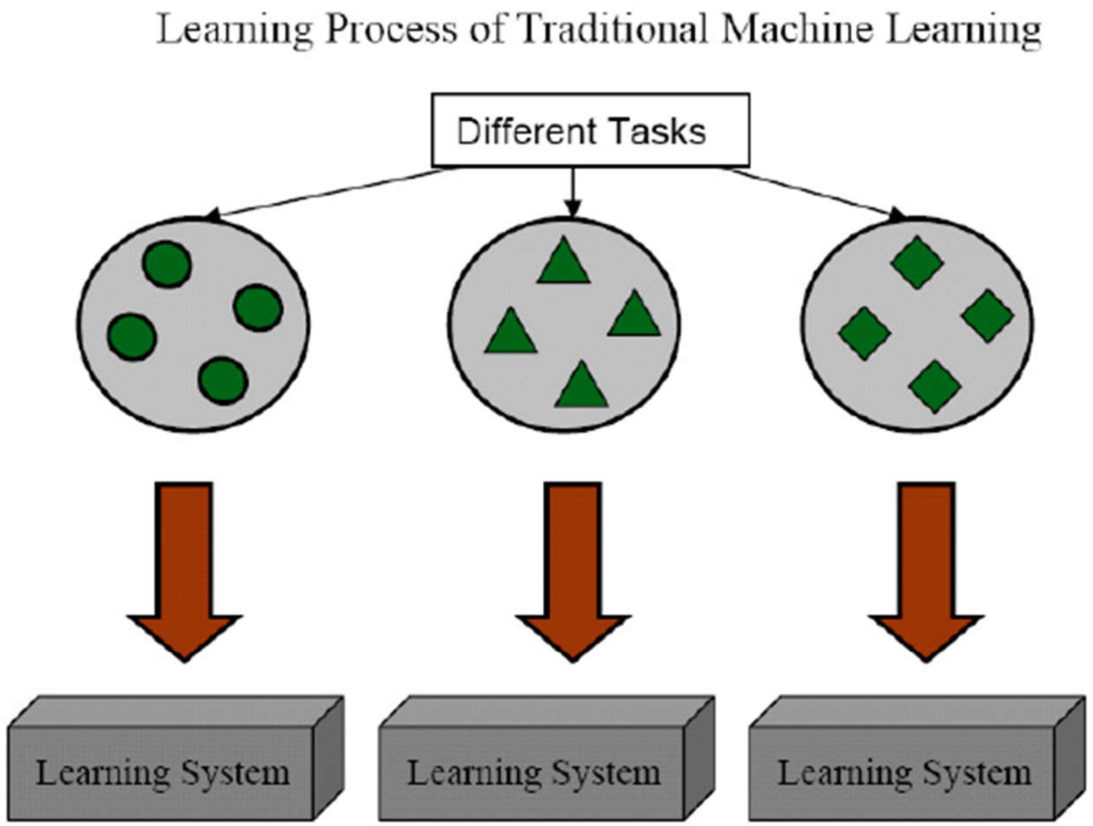
https://jamiekang.github.io/2017/06/05/domain-adversarial-training-of-neural-networks/

Transfer Learning



Transfer Learning

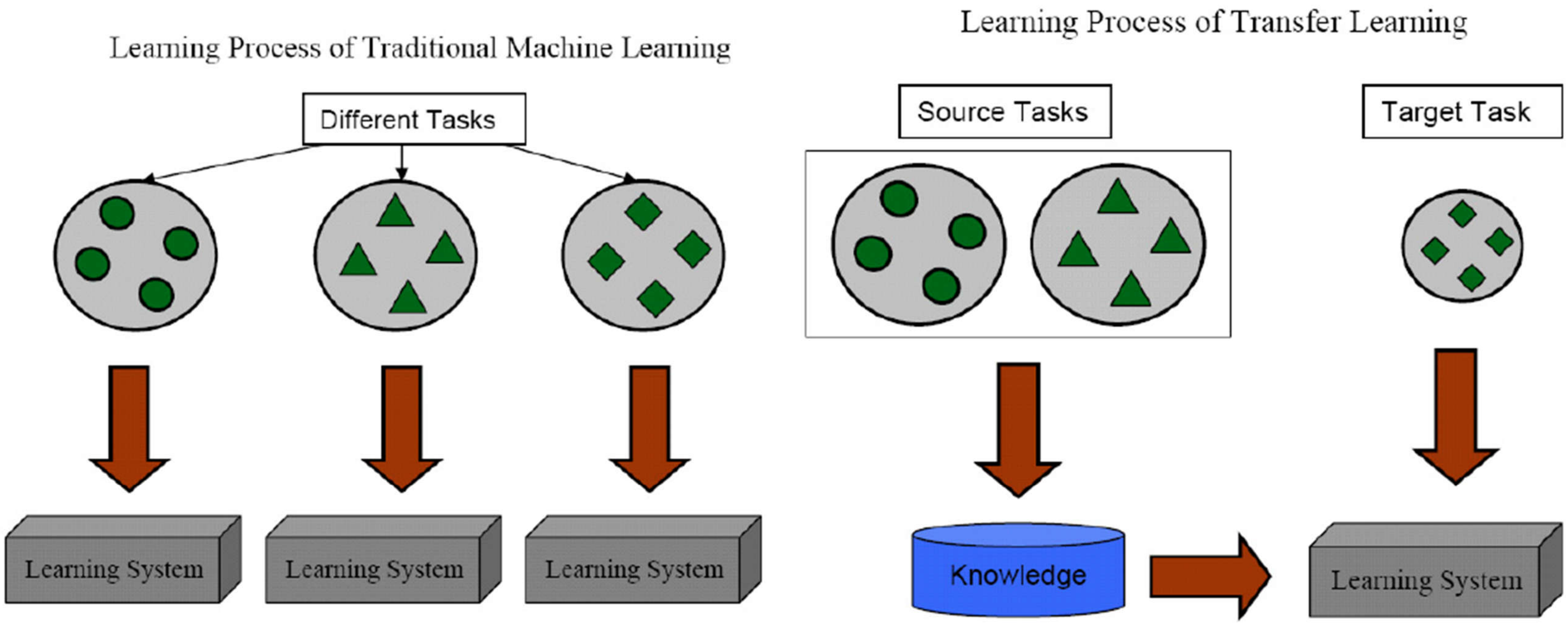
Transfer Learning



(a) Traditional Machine Learning

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5288526

Transfer Learning



(a) Traditional Machine Learning (b) Transfer Learning

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5288526

Transfer Learning

Dog/Cat
Classifier

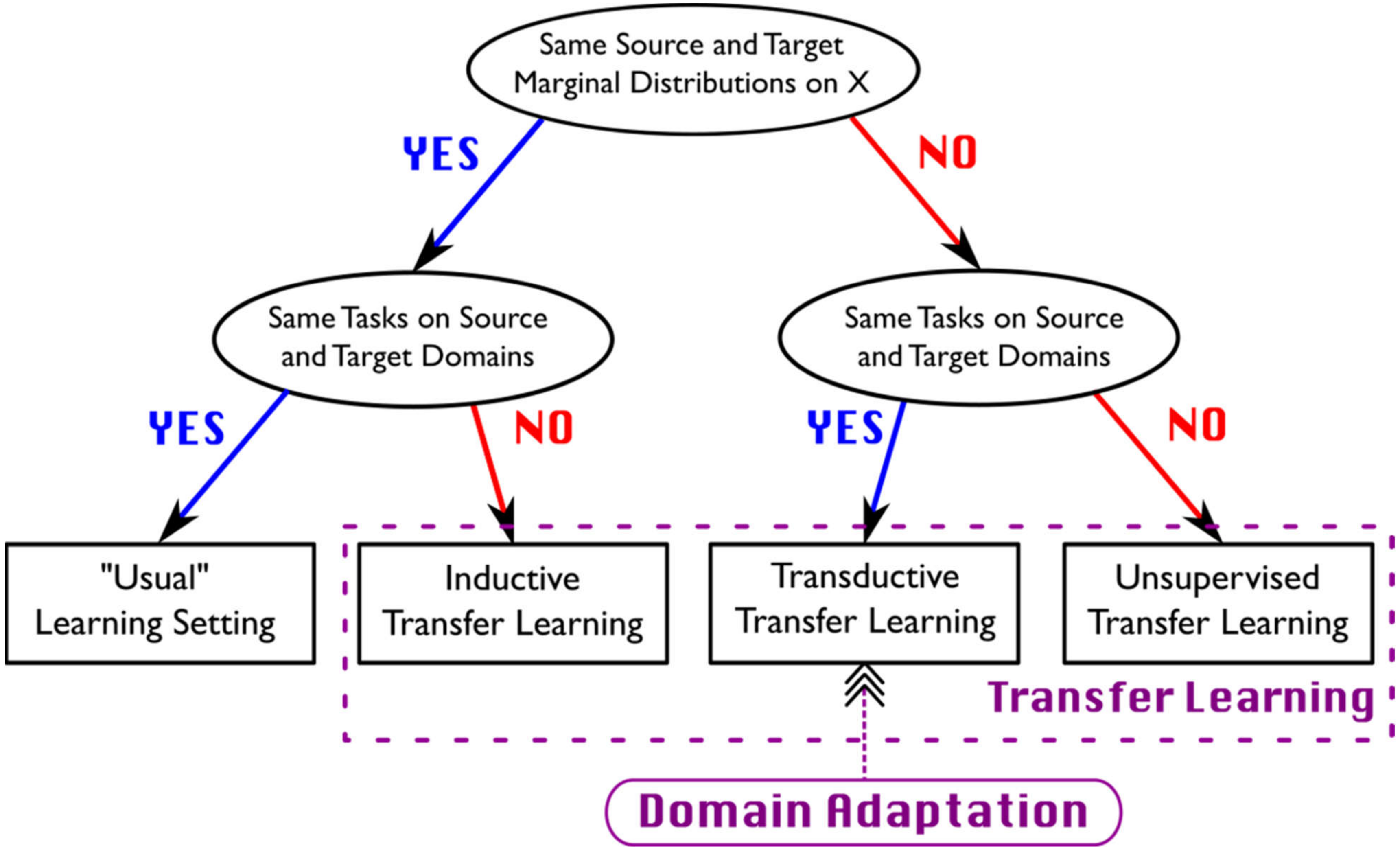
Data ***not directly related to*** the task considered

Similar domain, different tasks

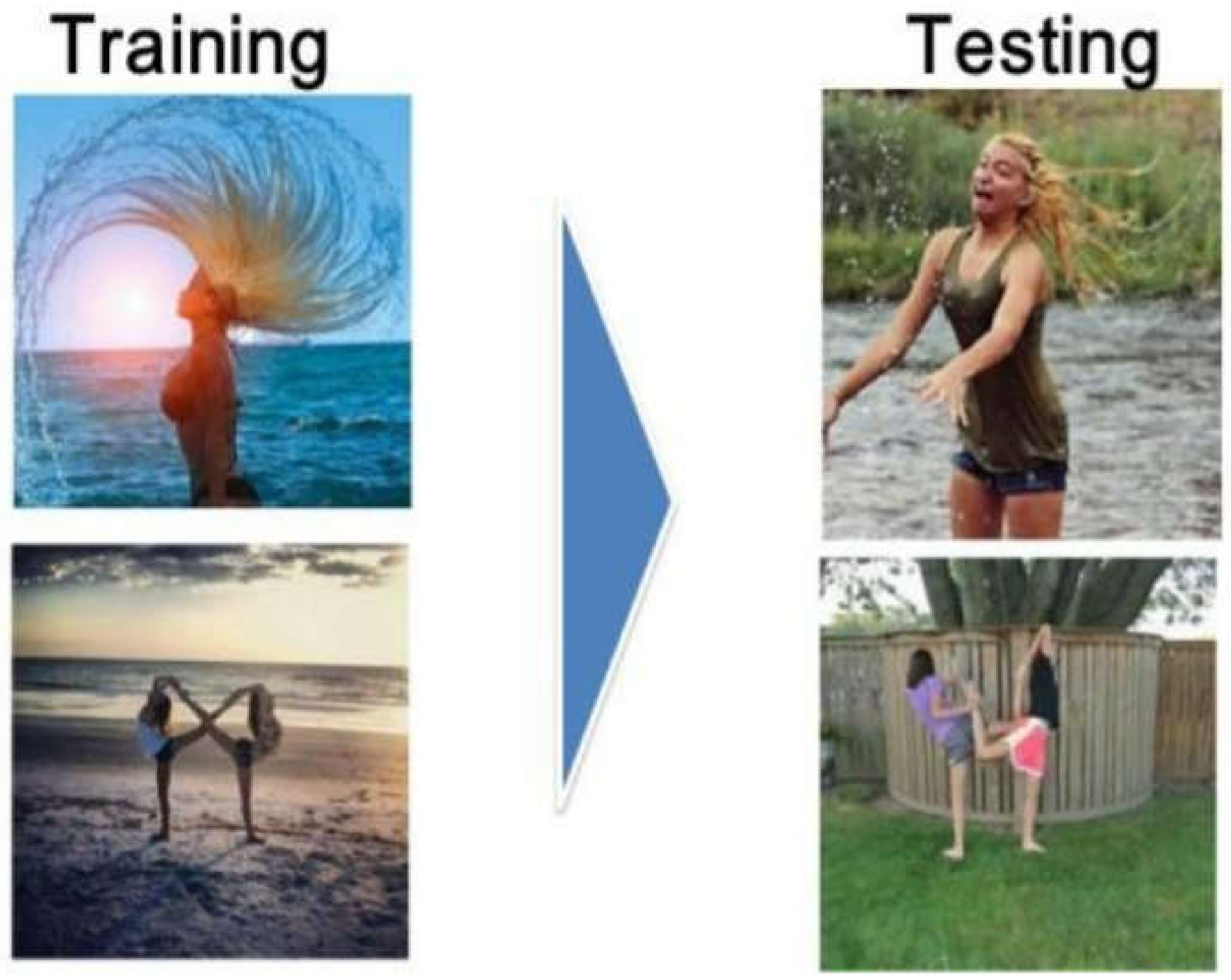
Different domains, same task

http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2016/Lecture/transfer%20(v3).pdf

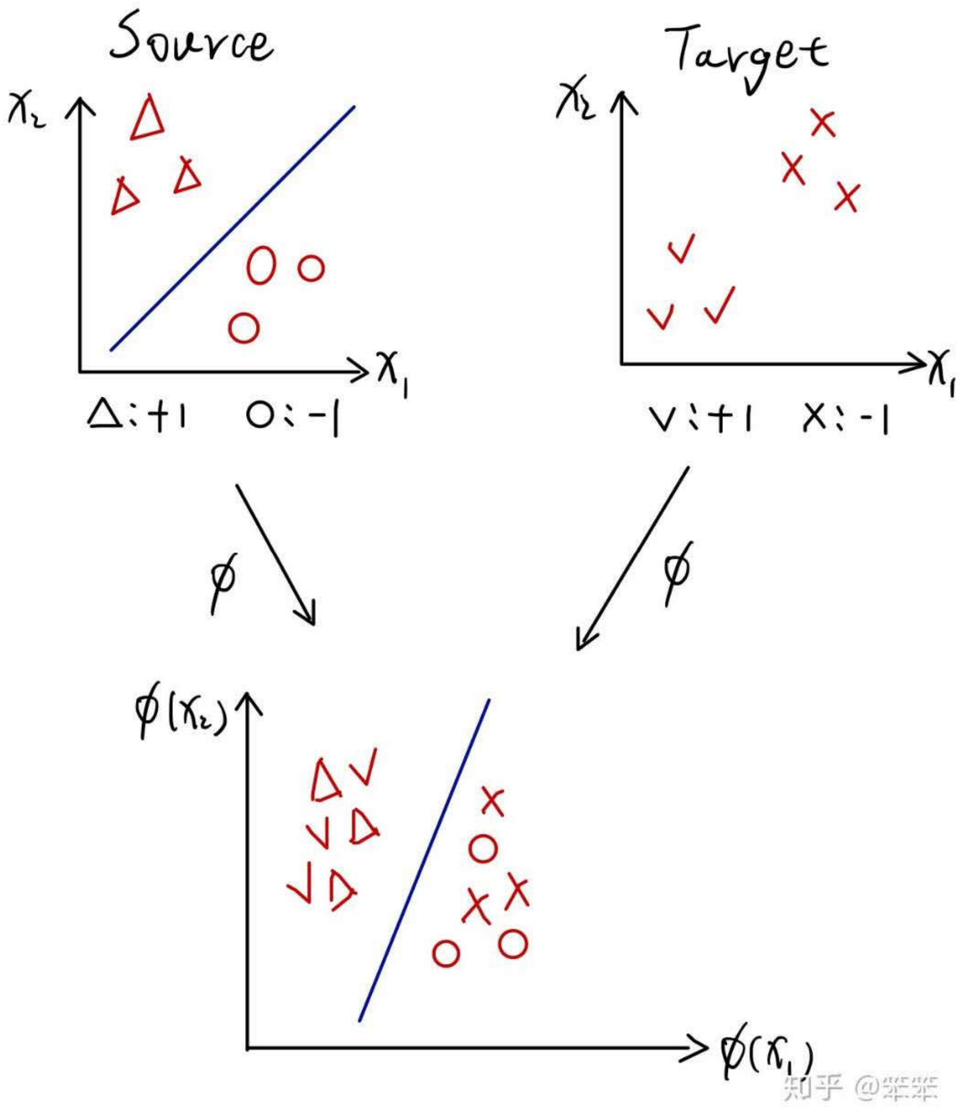
Transfer Learning



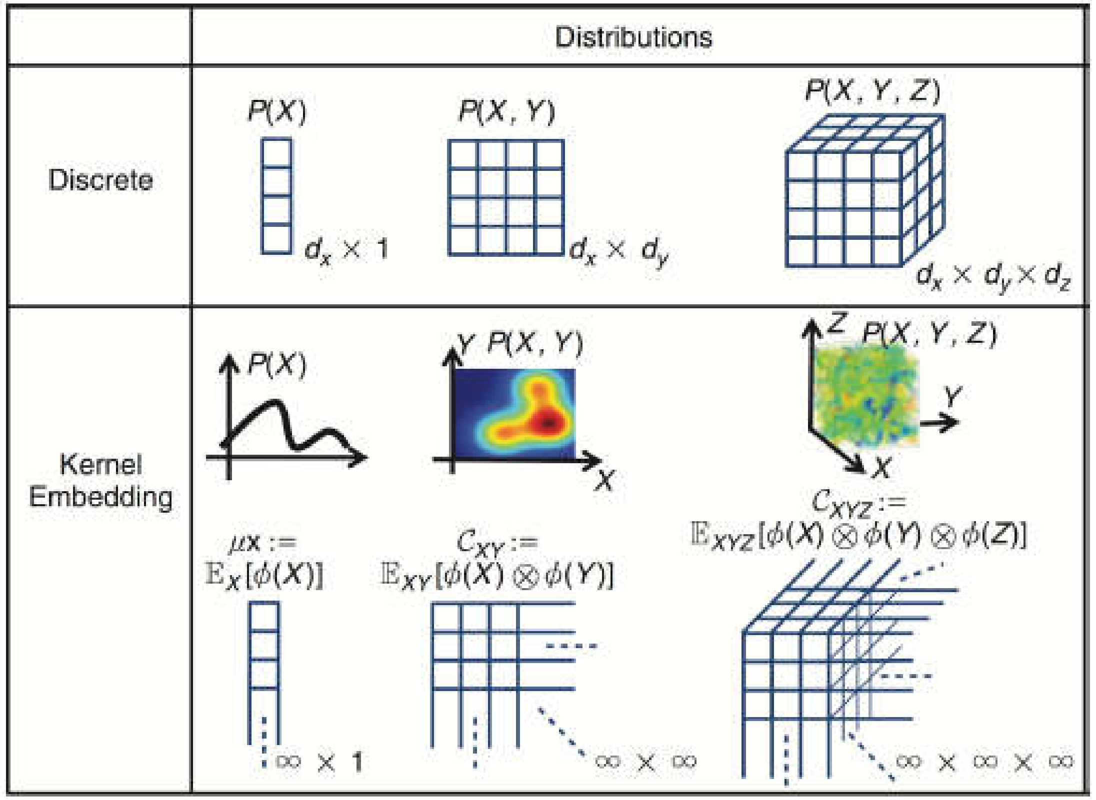
Domain Adaption



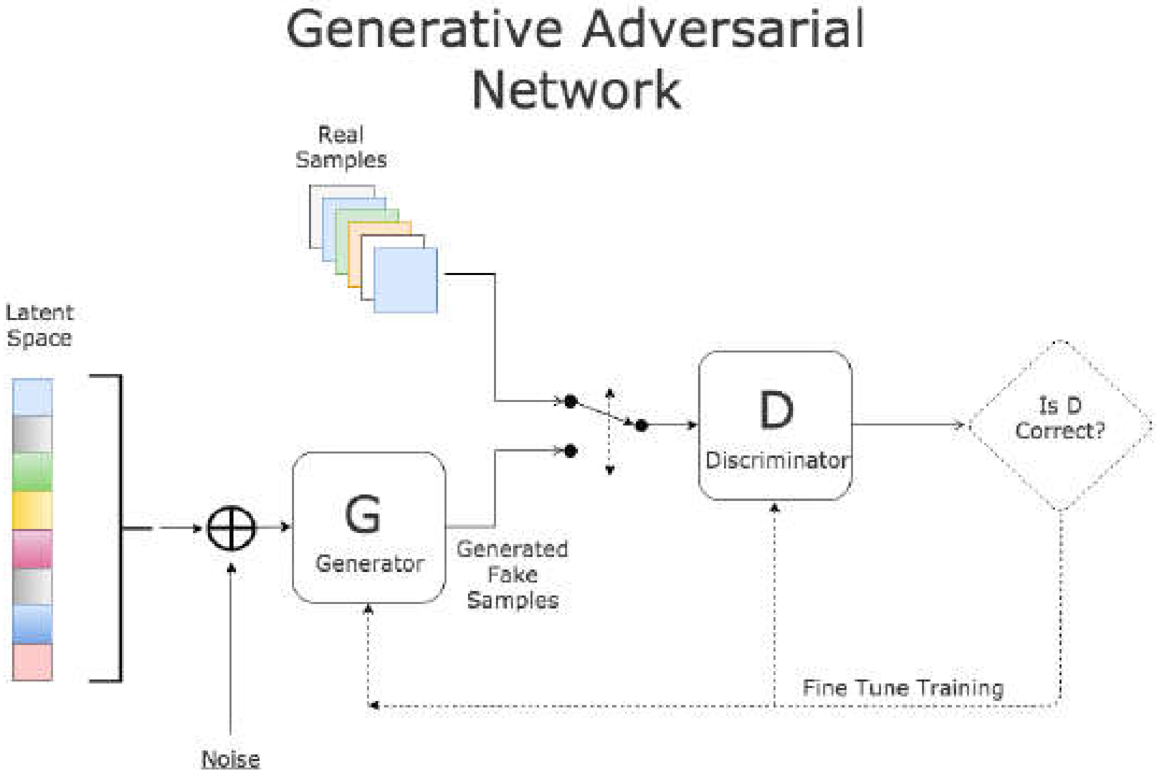
Methods of Domain Adaption



Methods of Domain Adaption



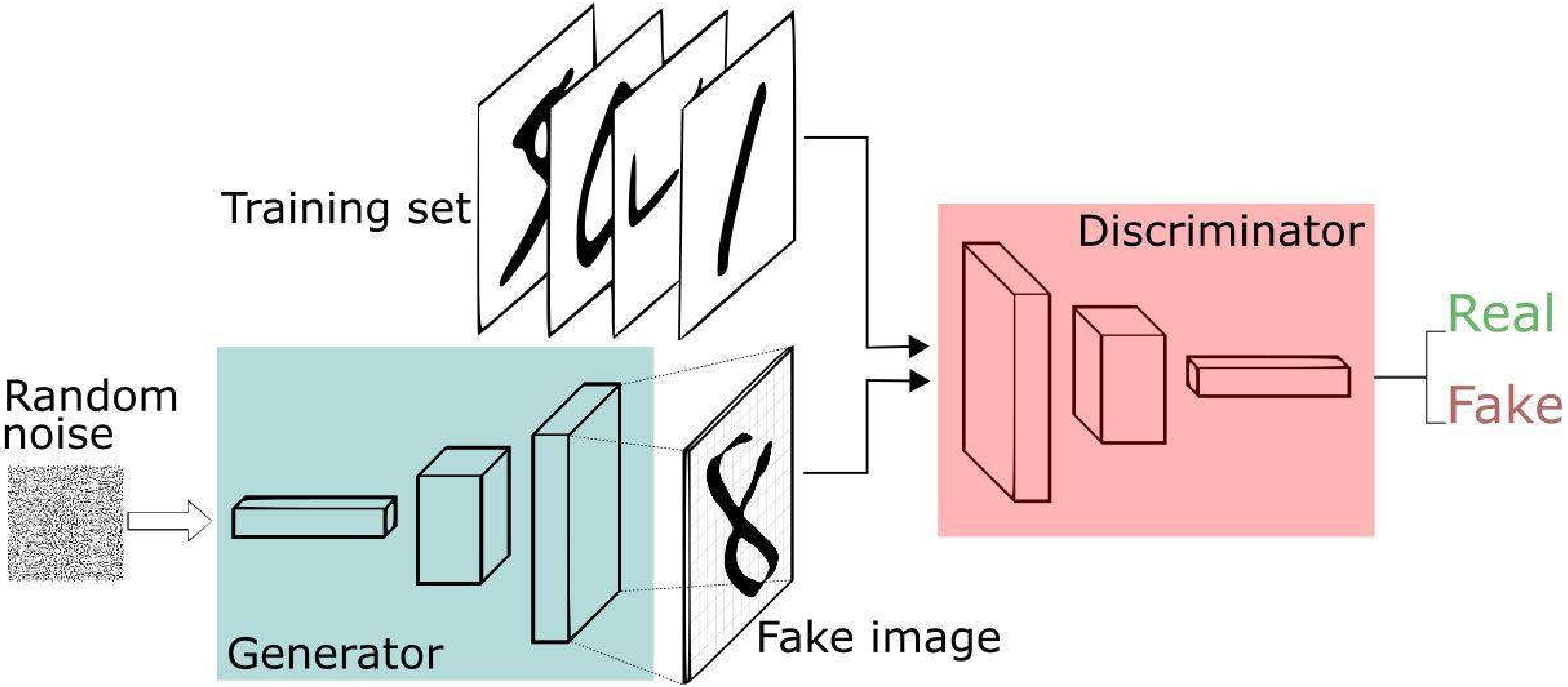
Kernel Embedding



Adversarial Learning

Multi-Adversarial Domain Adaption

Multi-Adversarial Domain Adaption



GAN, Goodfellow 2014

Multi-Adversarial Domain Adaption

		Source Data (not directly related to the task)	
		labelled	unlabeled
Target Data	labelled	Fine-tuning Multitask Learning	Self-taught learning Rajat Raina , Alexis Battle , Honglak Lee , Benjamin Packer , Andrew Y. Ng, Self-taught learning: transfer learning from unlabeled data, ICML, 2007
	unlabeled	Domain-adversarial training Zero-shot learning	Self-taught Clustering Wenyuan Dai, Qiang Yang, Gui-Rong Xue, Yong Yu, "Self-taught clustering", ICML 2008

Task Description

- Source data: $(x^s, y^s) \rightarrow$ Training data
- Target data: $(x^t) \rightarrow$ Testing data

mismatch

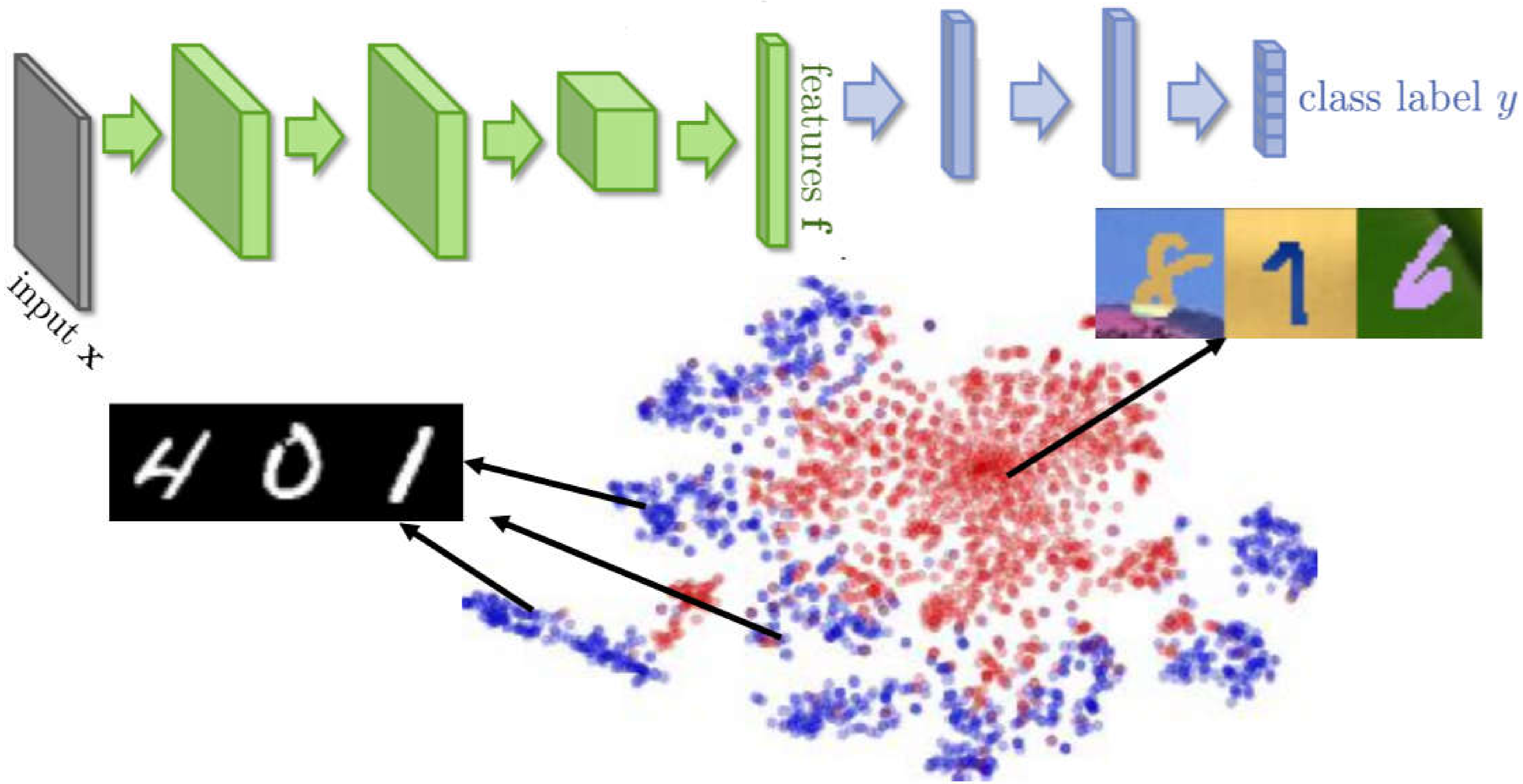
MNIST

SOURCE  with label

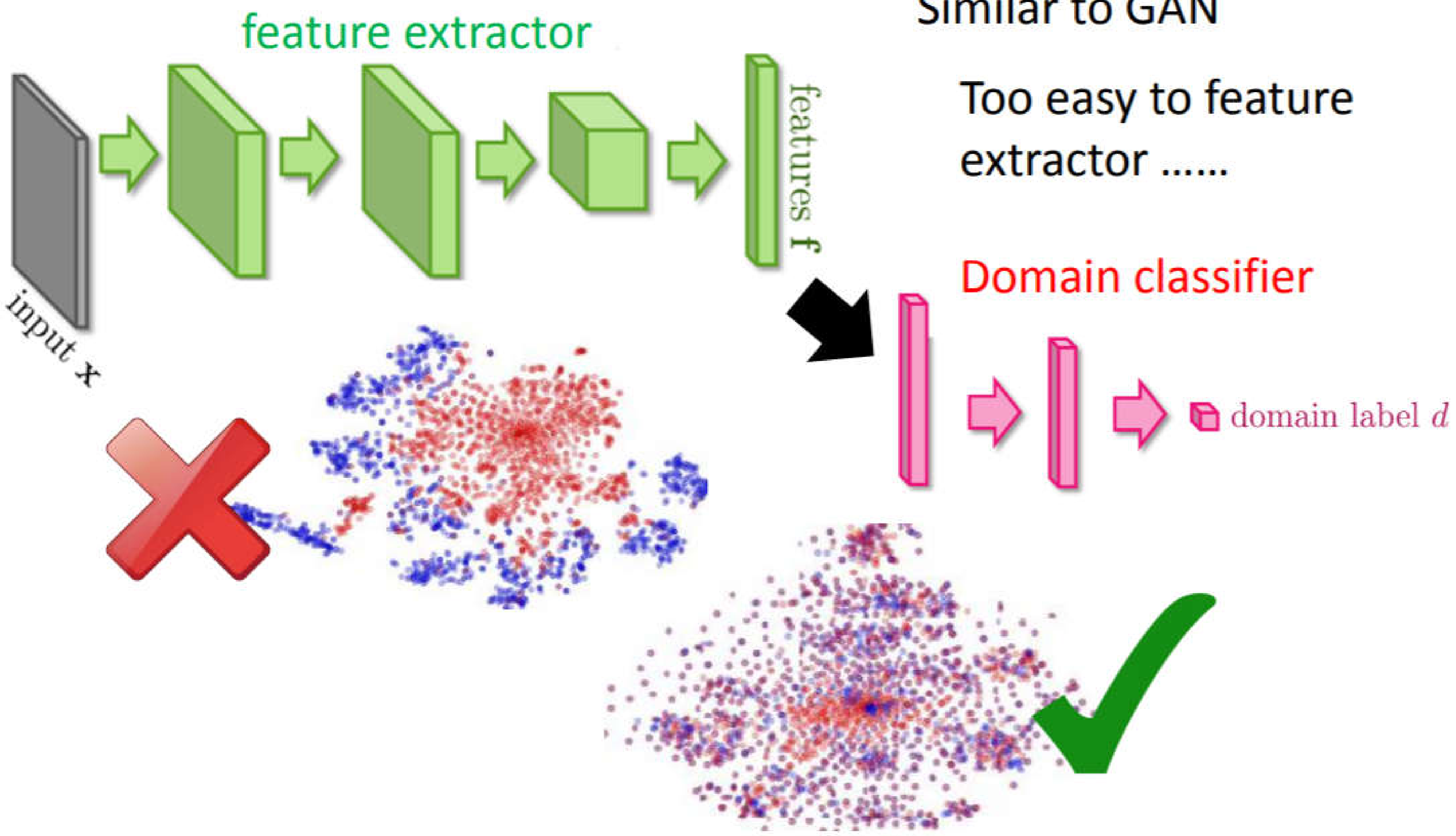
TARGET without label

MNIST-M

Domain-adversarial Training



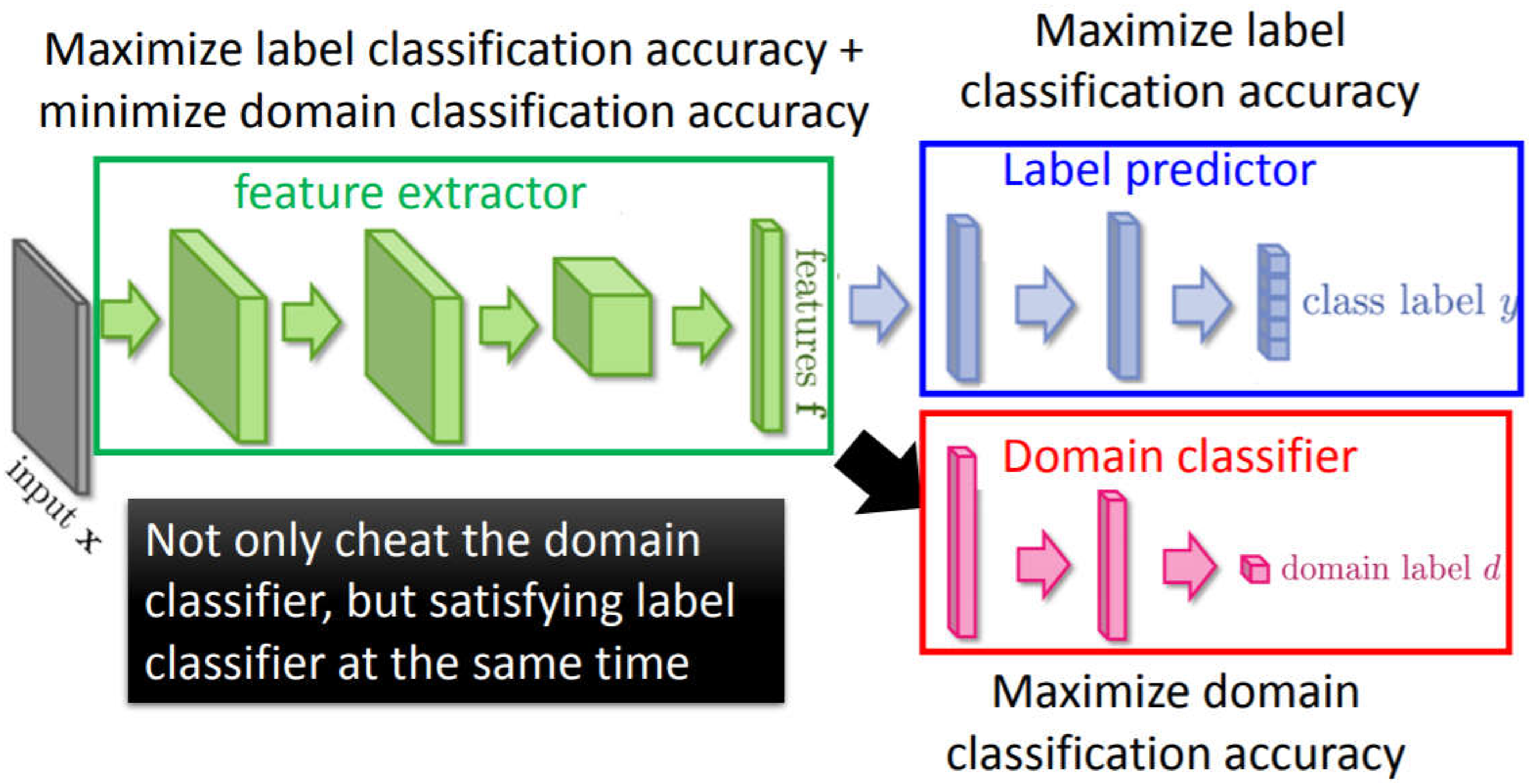
Domain-adversarial Training



Similar to GAN

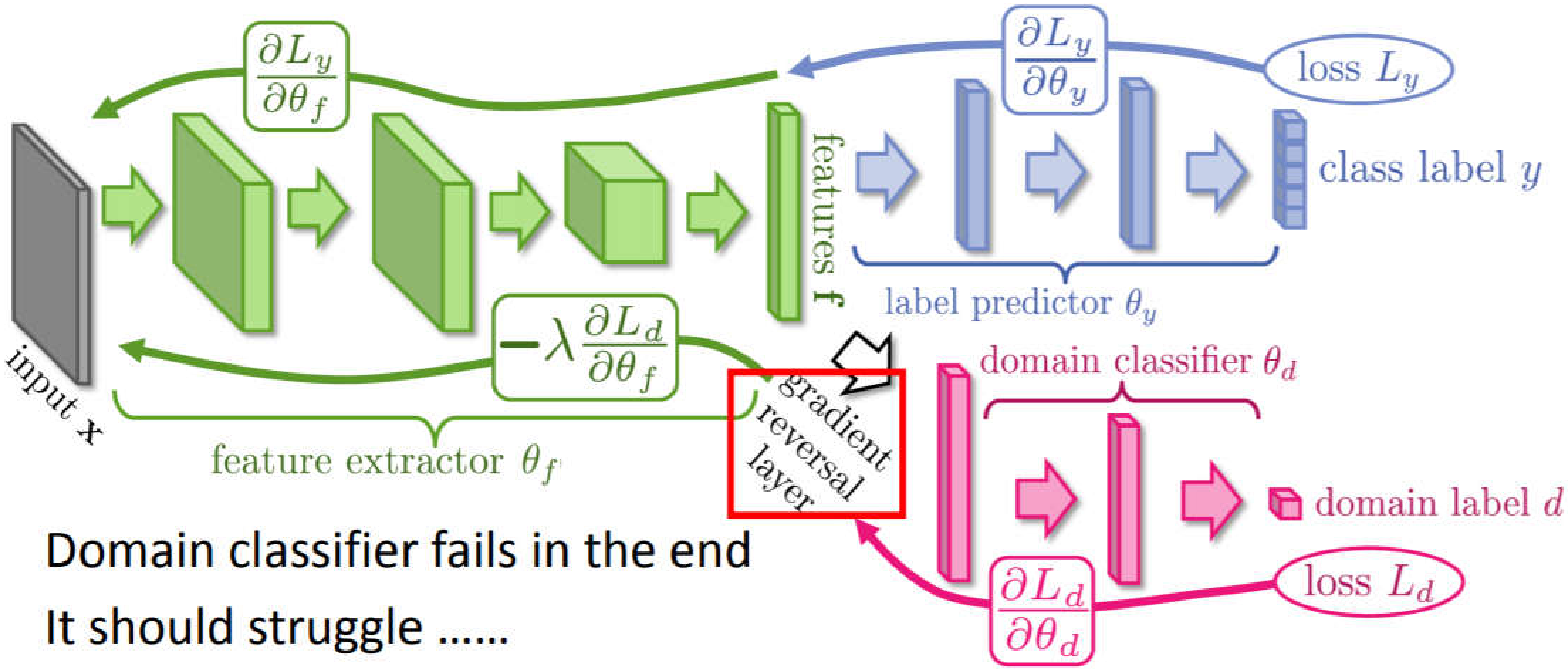
Too easy to feature extractor ……

Domain-adversarial Training



This is a big network, but different parts have different goals.

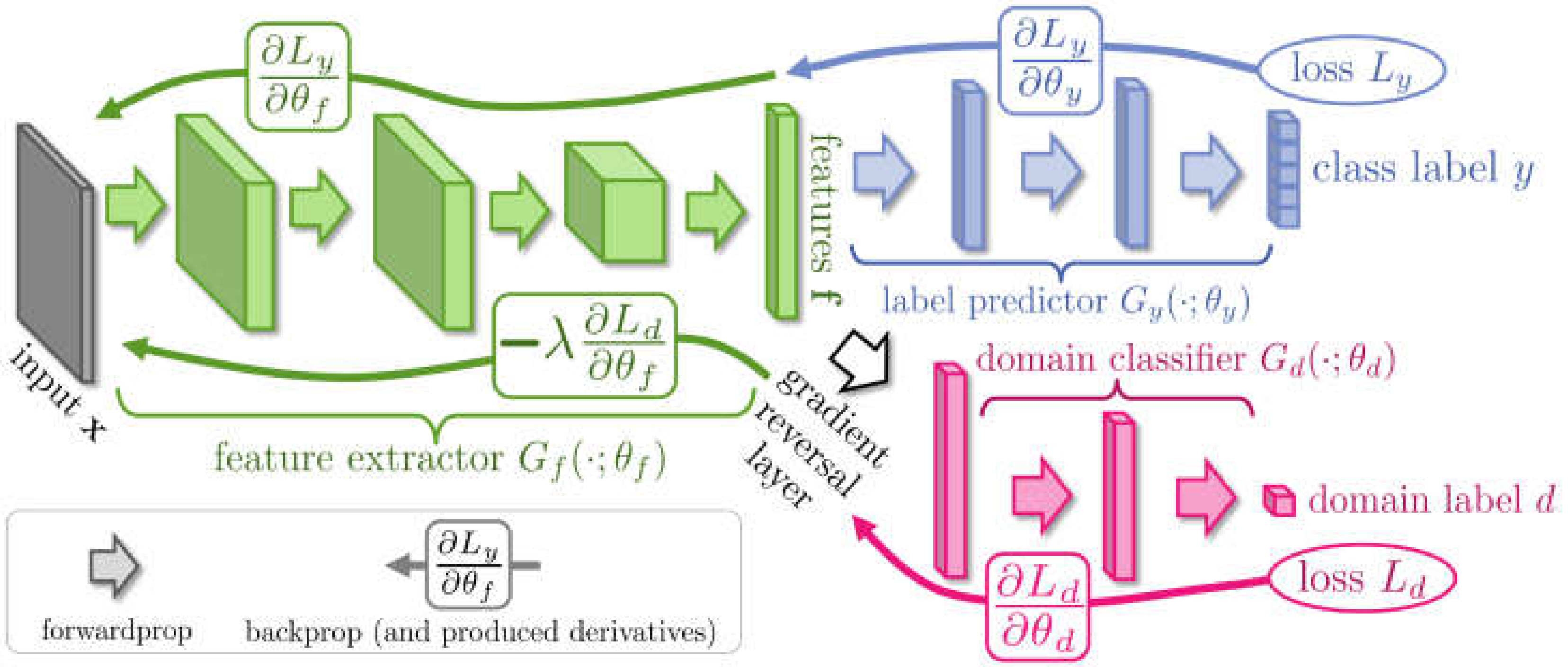
Domain-adversarial Training



Domain classifier fails in the end

It should struggle ……

Domain-adversarial Training

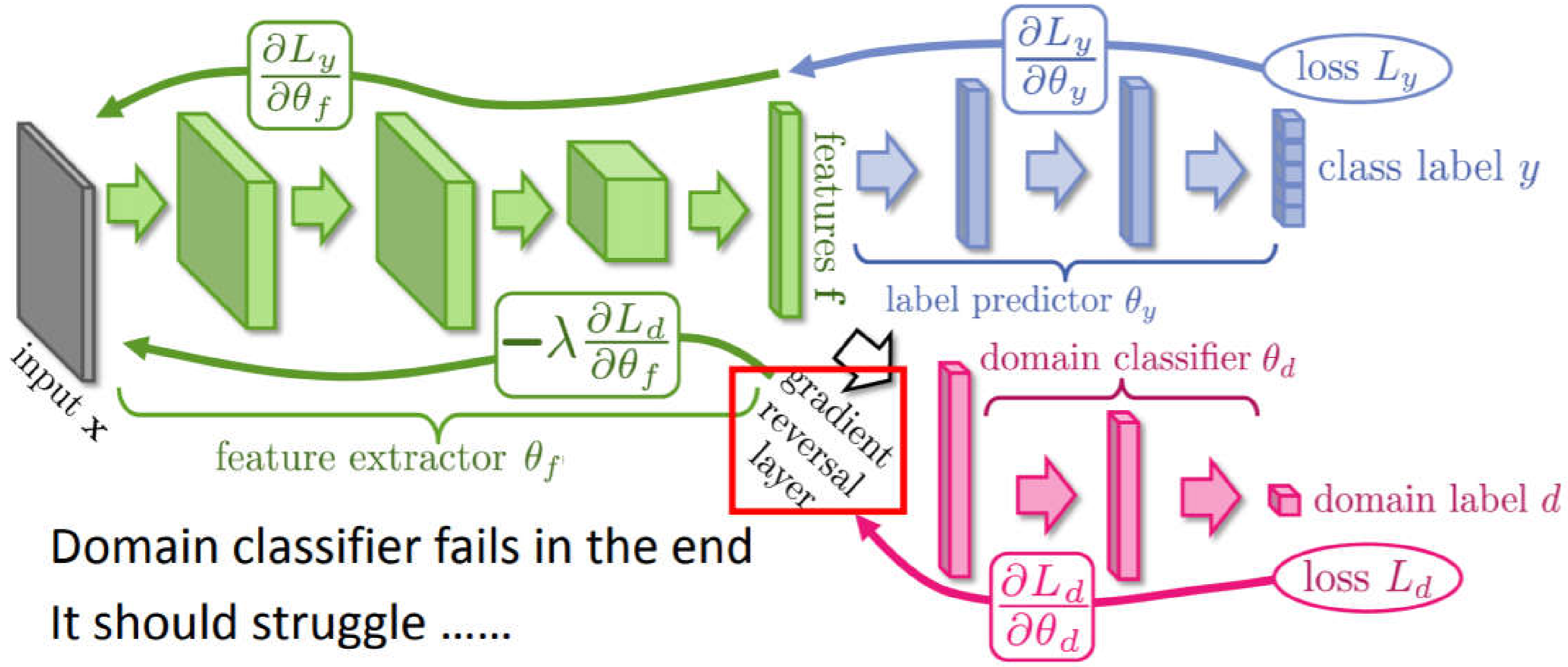


- Adversarial adaptation: learning features indistinguishable across domains

$$E(\theta_f,\theta_y,\theta_d)=\sum_{\mathbf{x}_i\in\mathcal{D}_s} L_y\big(G_y(G_f(\mathbf{x}_i)),y_i\big)-\lambda\sum_{\mathbf{x}_i\in\mathcal{D}_s\cup\mathcal{D}_t} L_d\big(G_d(G_f(\mathbf{x}_i)),d_i\big)$$

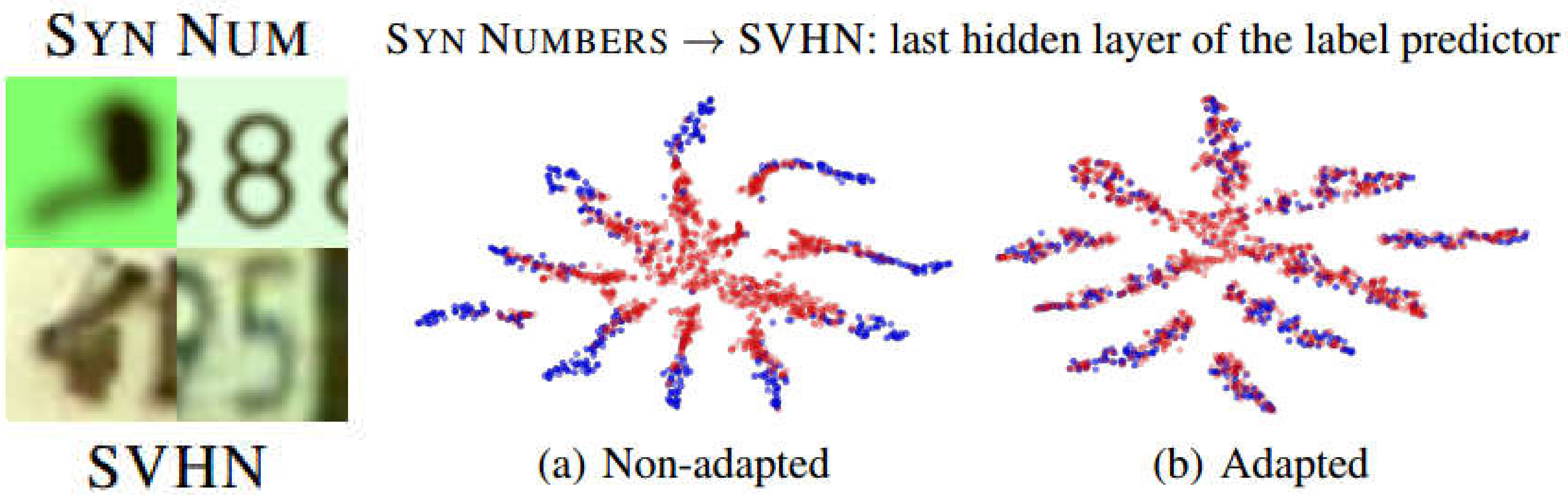
$$(\hat{\theta}_f,\hat{\theta}_y)=\underset{\theta_f,\theta_y}{\arg\min}\,E(\theta_f,\theta_y,\theta_d)\quad(\hat{\theta}_d)=\underset{\theta_d}{\arg\max}\,E(\theta_f,\theta_y,\theta_d)$$

Domain-adversarial Training

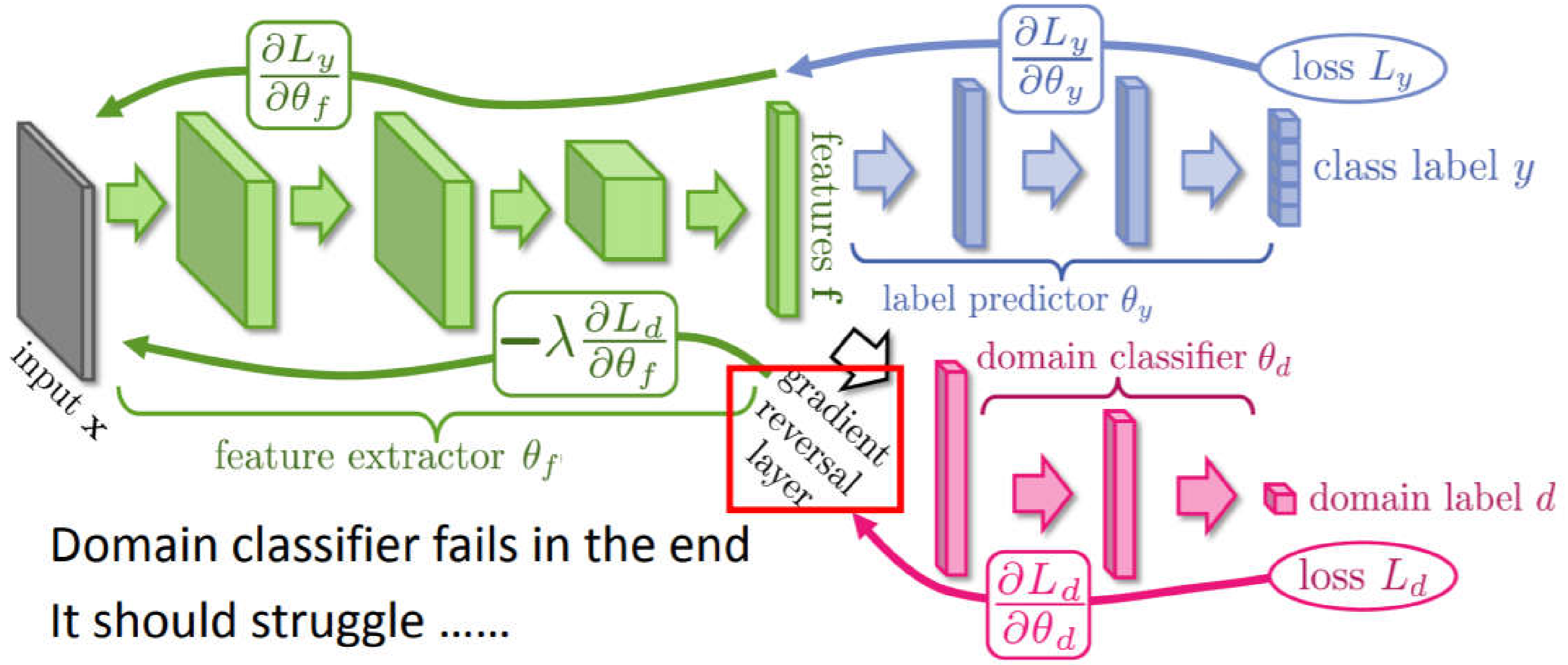


Domain classifier fails in the end

It should struggle ……

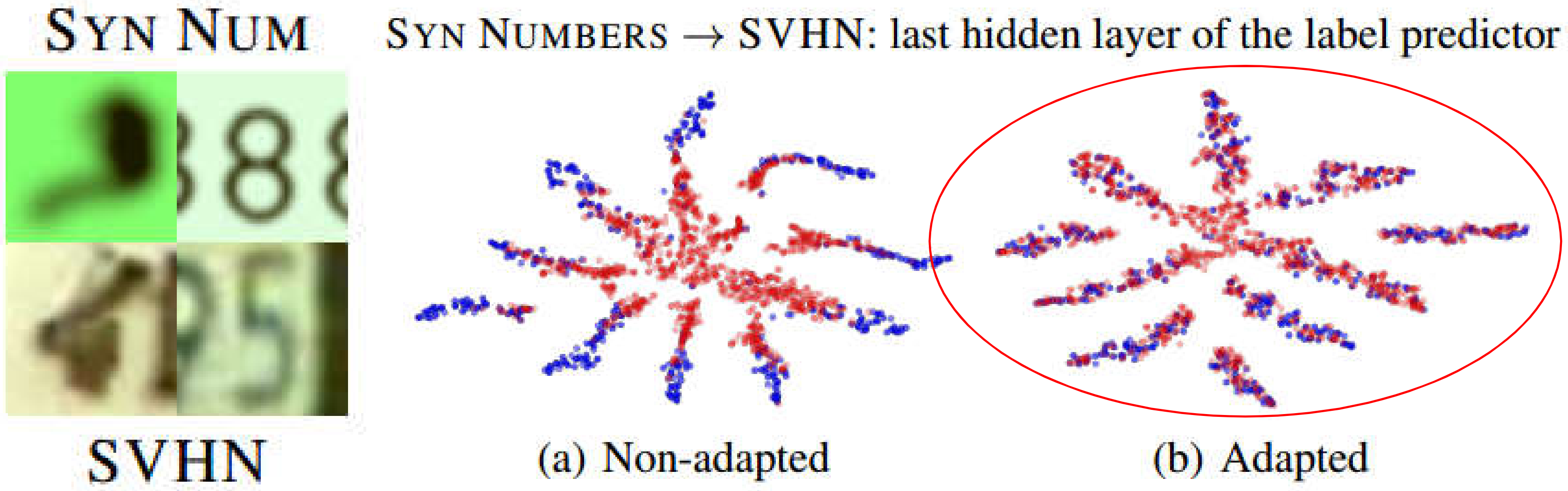


Domain-adversarial Training



Domain classifier fails in the end

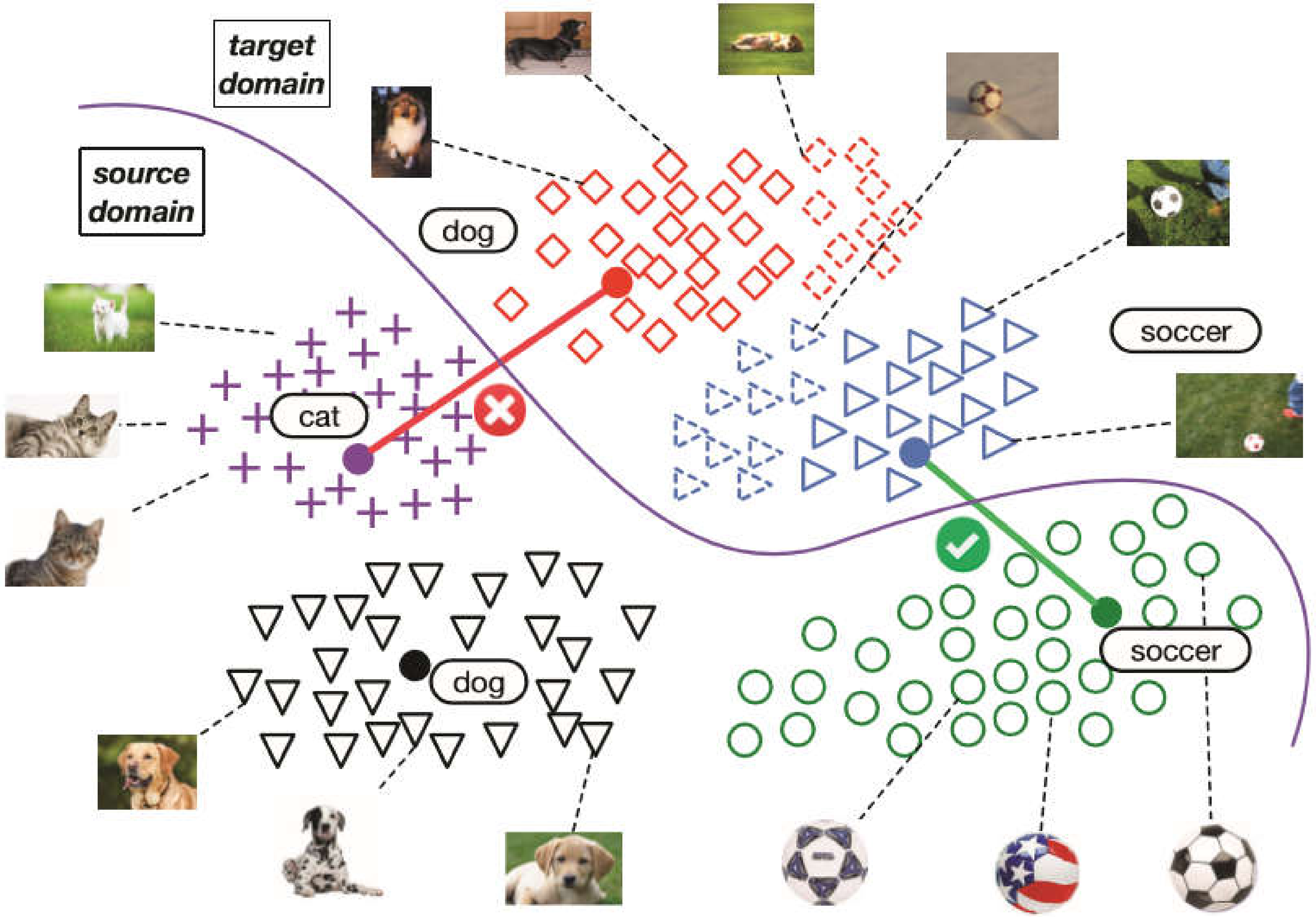
It should struggle ……



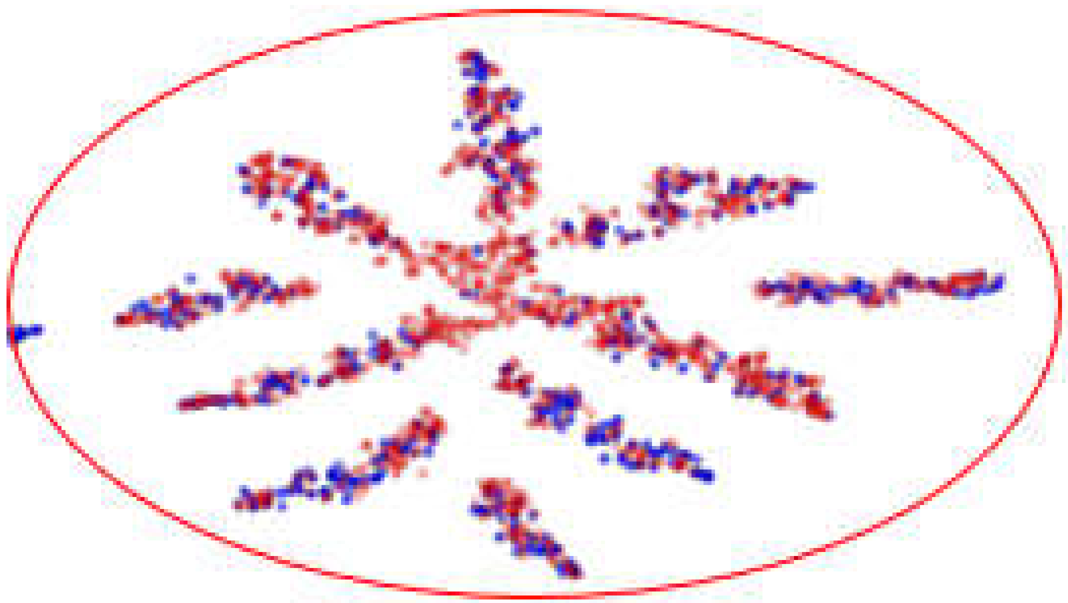
(a) Non-adapted (b) Adapted

Multi-Adversarial Domain Adaption

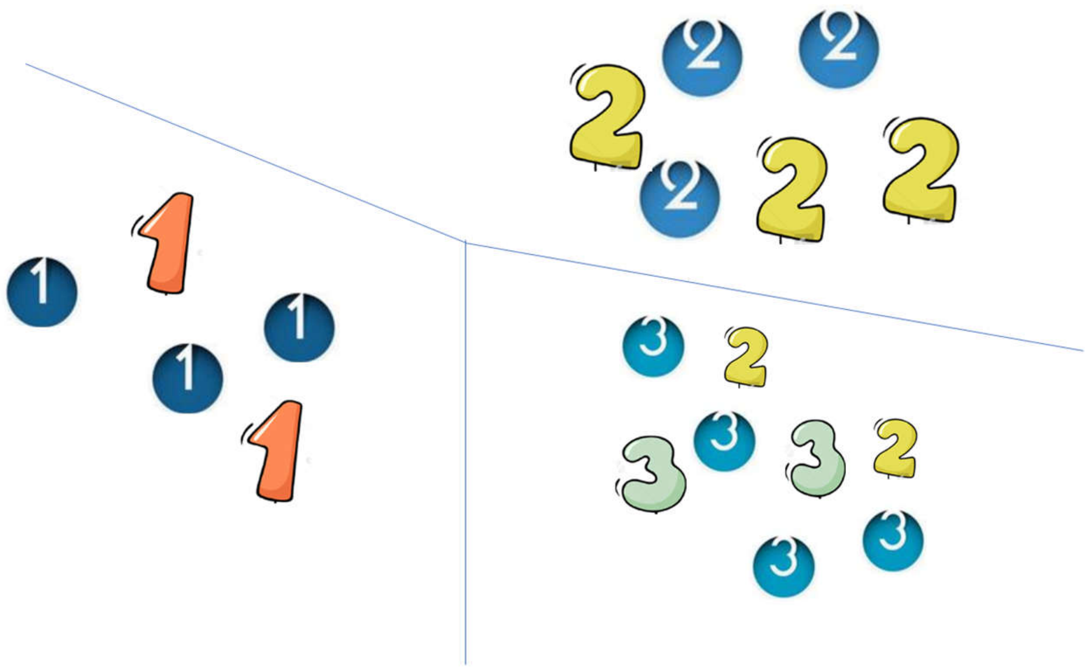
Existing Problem



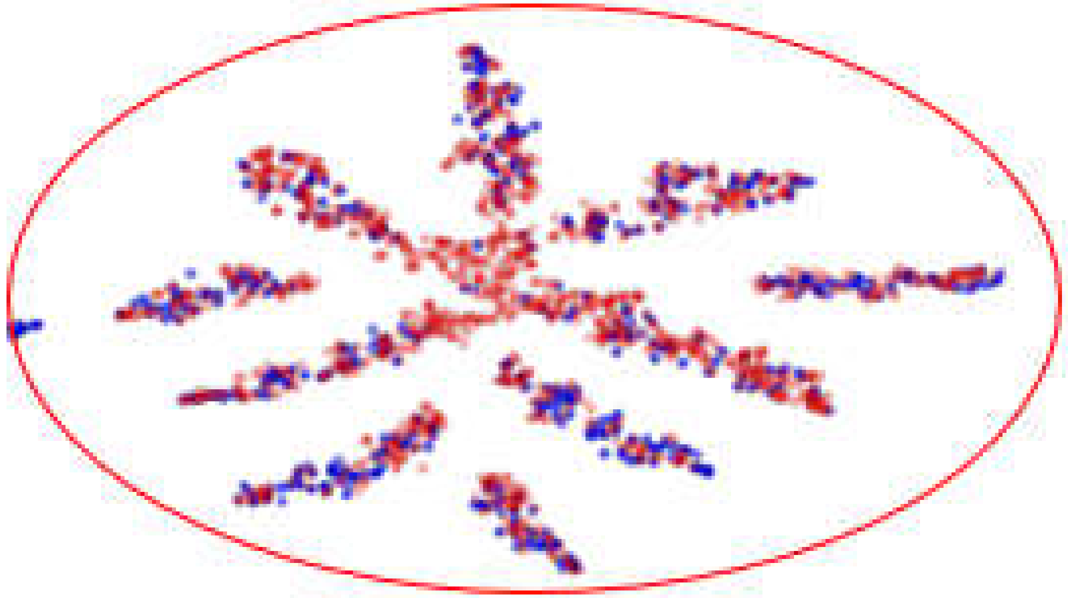
Existing Problem : Negative Transfer



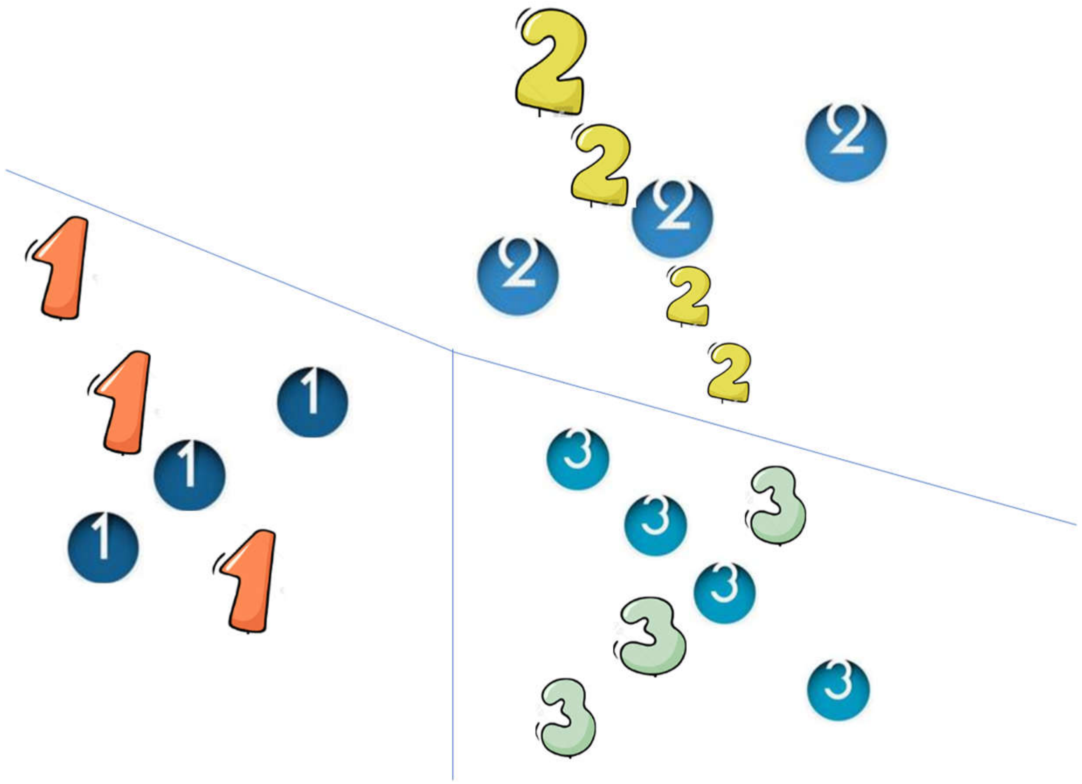
(b) Adapted



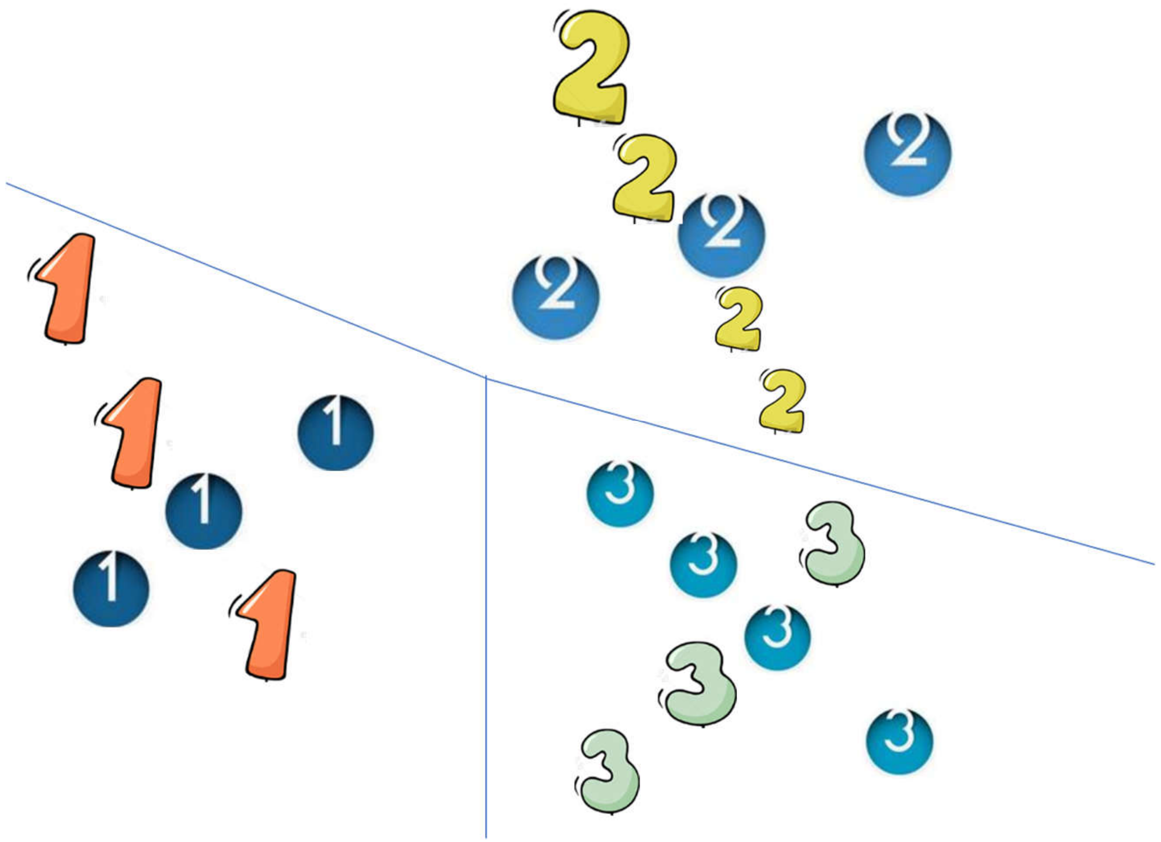
Existing Problem : Under Transfer

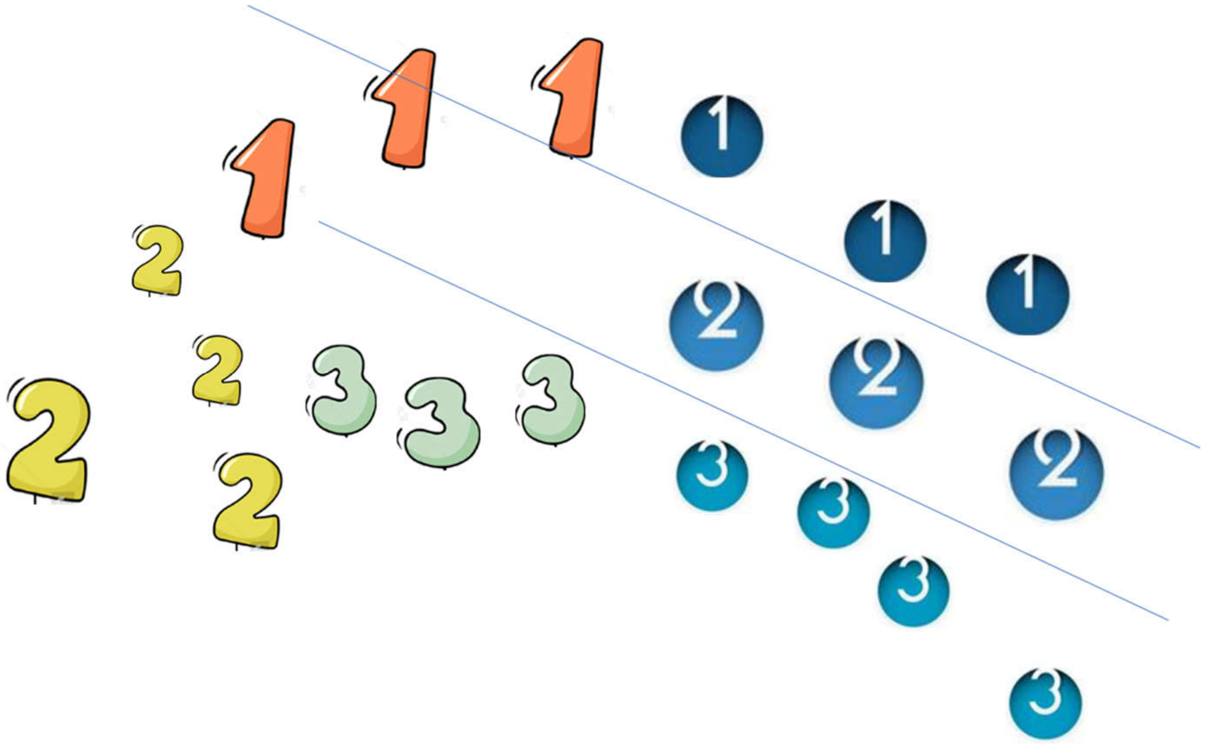


(b) Adapted

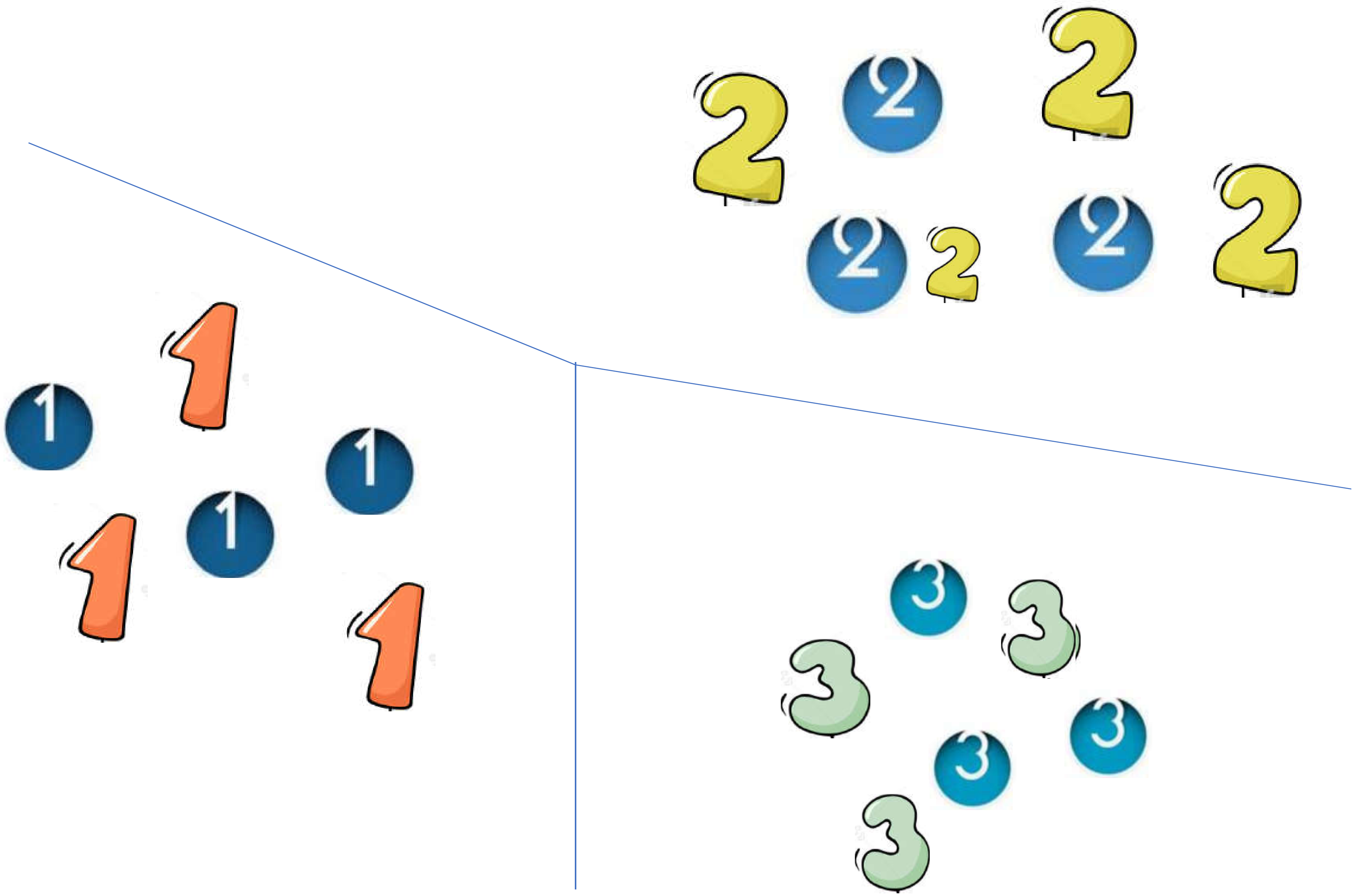


Existing Problem : Under Transfer

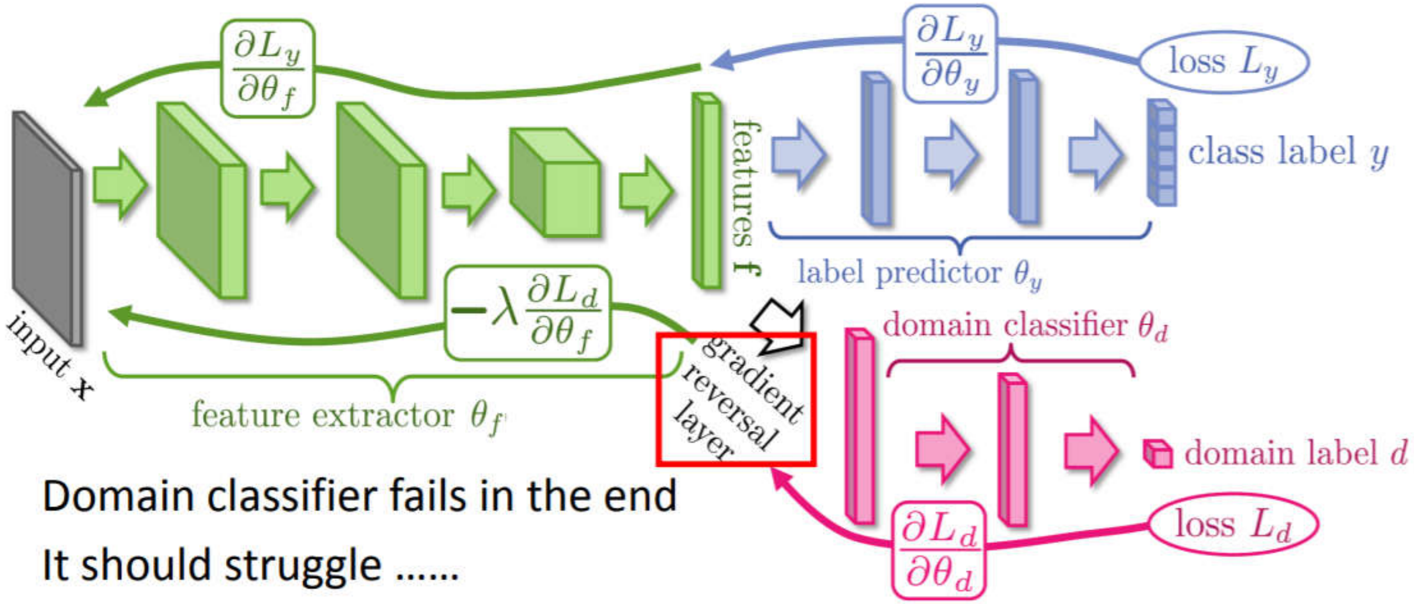




Existing Problem : Target

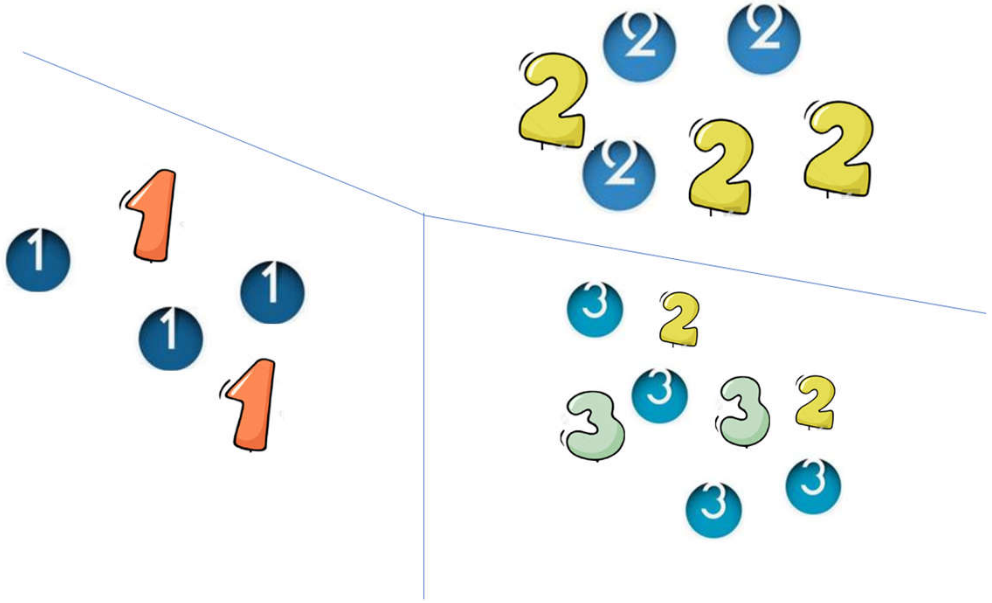


Multi-Adversarial Domain Adaption

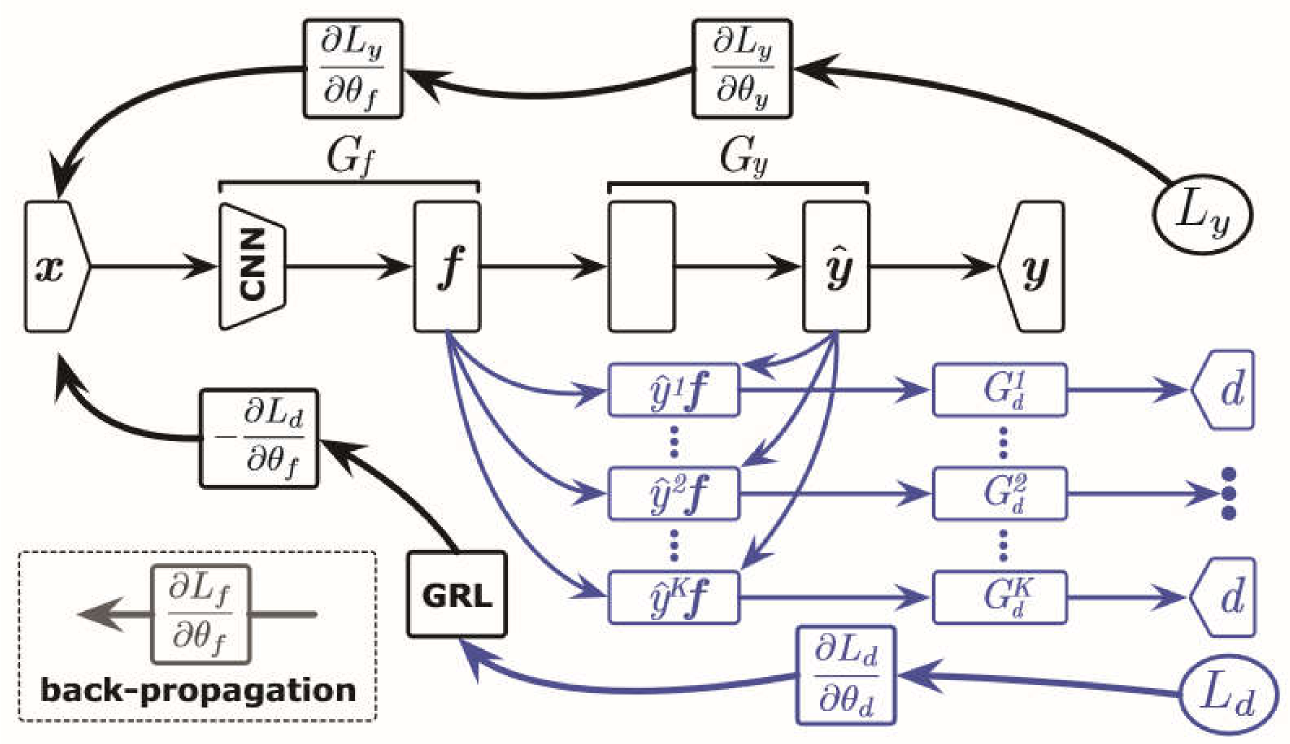


Domain classifier fails in the end

It should struggle ……



Multi-Adversarial Domain Adaption



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

$$(\hat{\theta}_f, \hat{\theta}_y) = \arg\min_{\theta_f, \theta_y} C\left(\theta_f, \theta_y, \theta_d^k|_{k=1}^K\right),$$

$$(\hat{\theta}_d^1, \ldots, \hat{\theta}_d^K) = \arg\max_{\theta_d^1, \ldots, \theta_d^K} C\left(\theta_f, \theta_y, \theta_d^k|_{k=1}^K\right).$$

Multi-Adversarial Domain Adaption

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**

Multi-Adversarial Domain Adaption

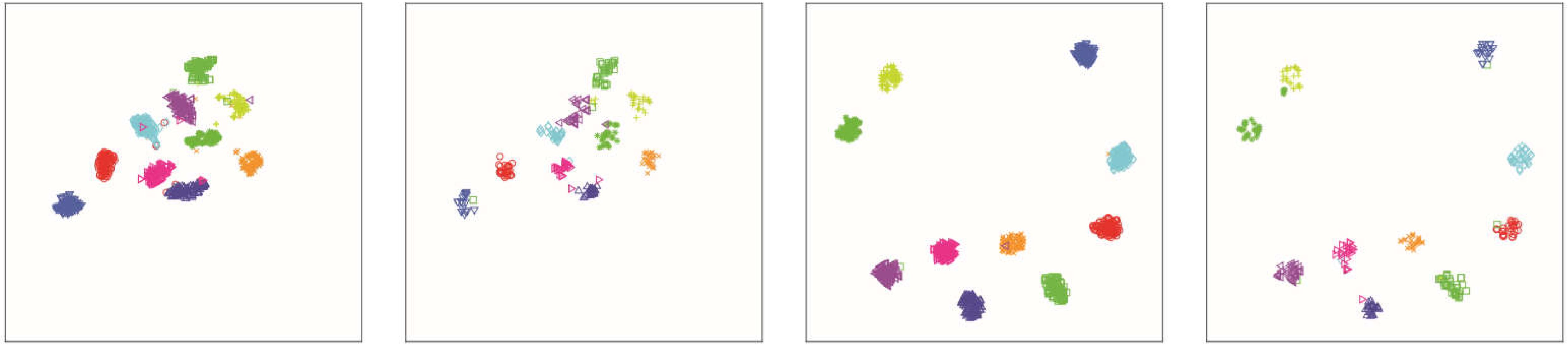
Table 2: Accuracy (%) on *ImageCLEF-DA* for unsupervised domain adaptation (AlexNet and ResNet)

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**

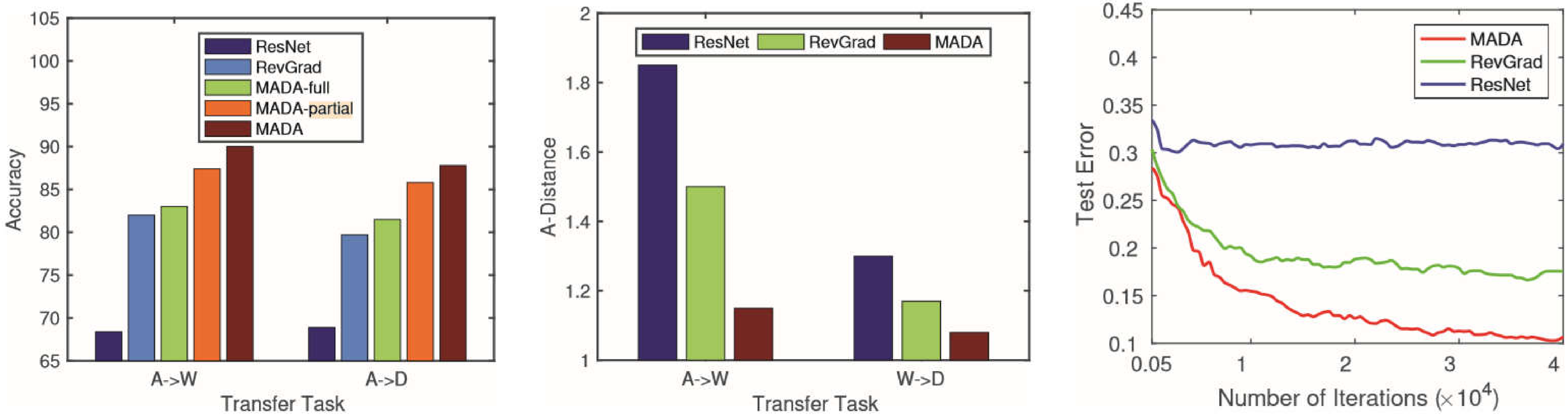
Multi-Adversarial Domain Adaption



(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).

Multi-Adversarial Domain Adaption

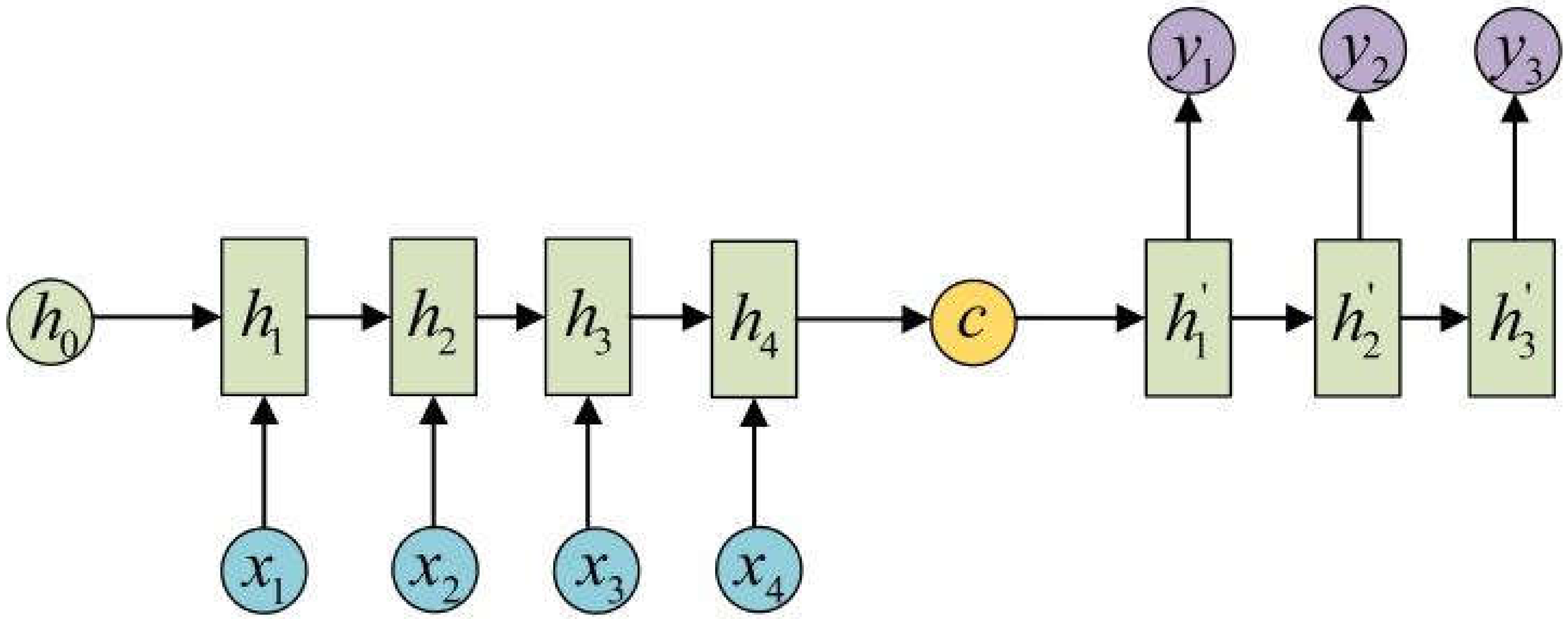


(a) Sharing Strategies (b) $\mathcal{A}$-distance (c) Convergence

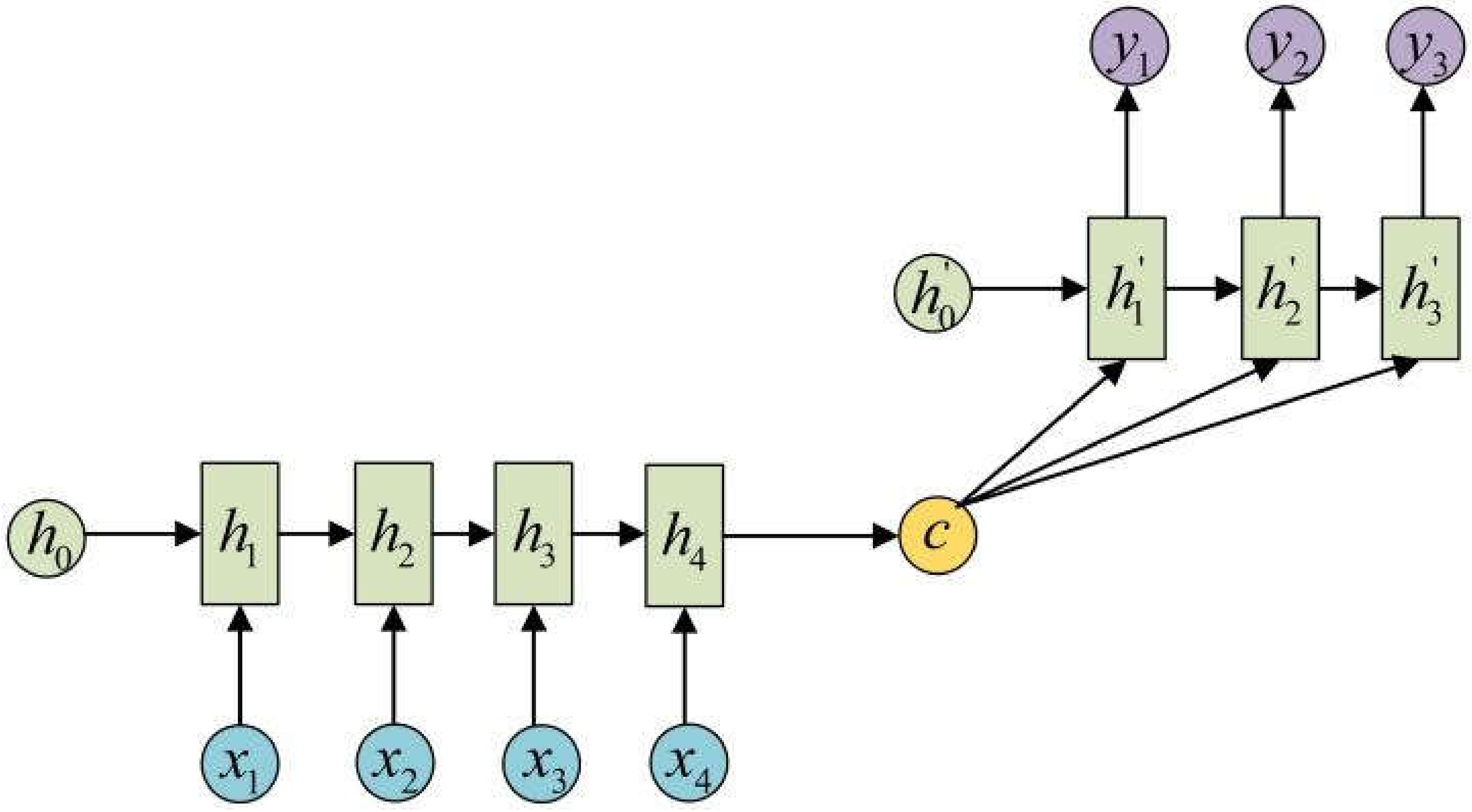
Figure 4: Empirical analysis: (a) Sharing strategies, (b) $\mathcal{A}$-distance, and (c) Convergence performance.

Bonus

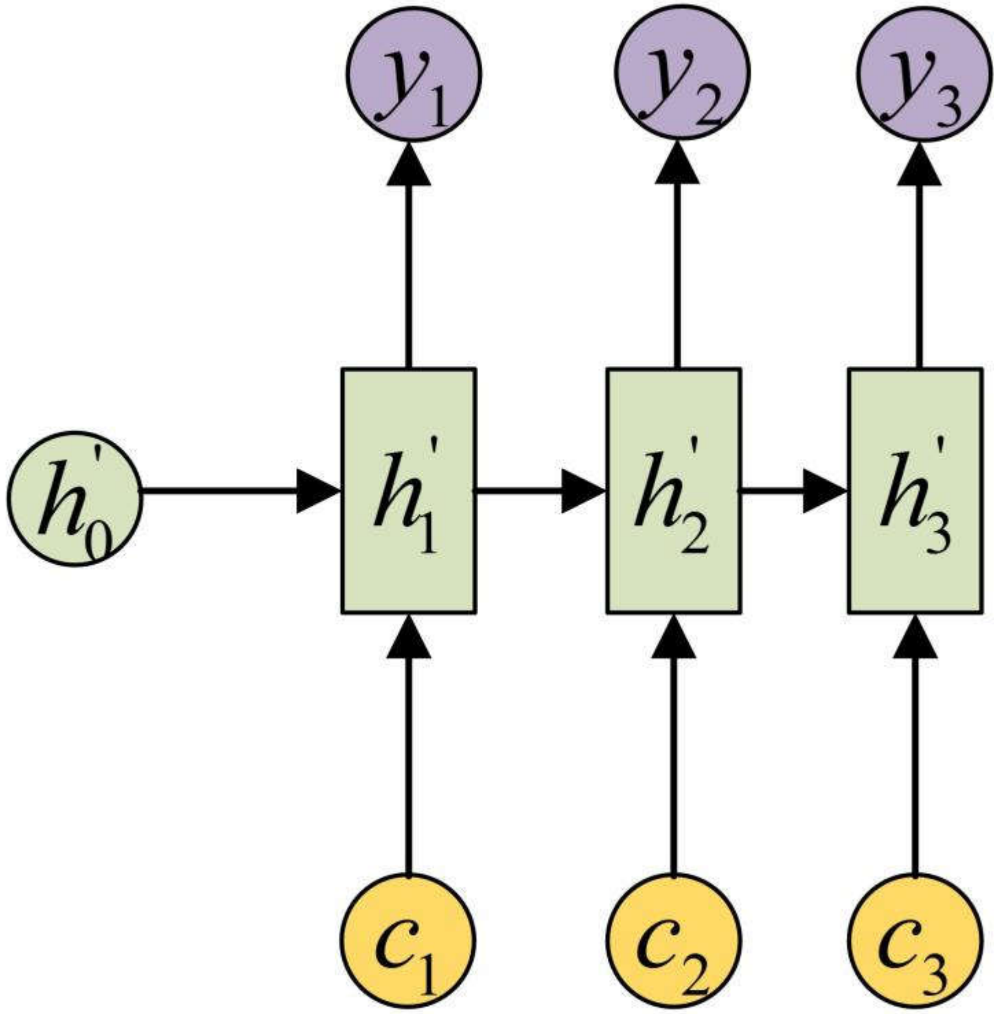
Attention Based Model



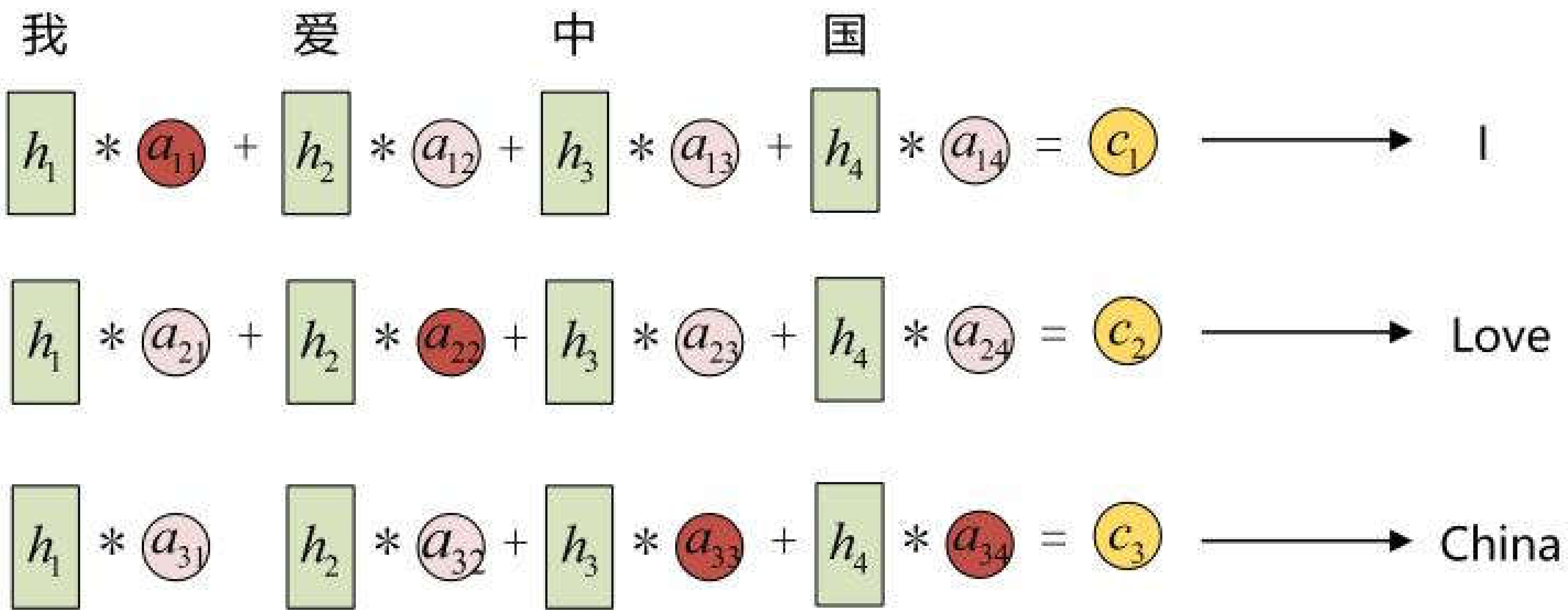
Attention Based Model



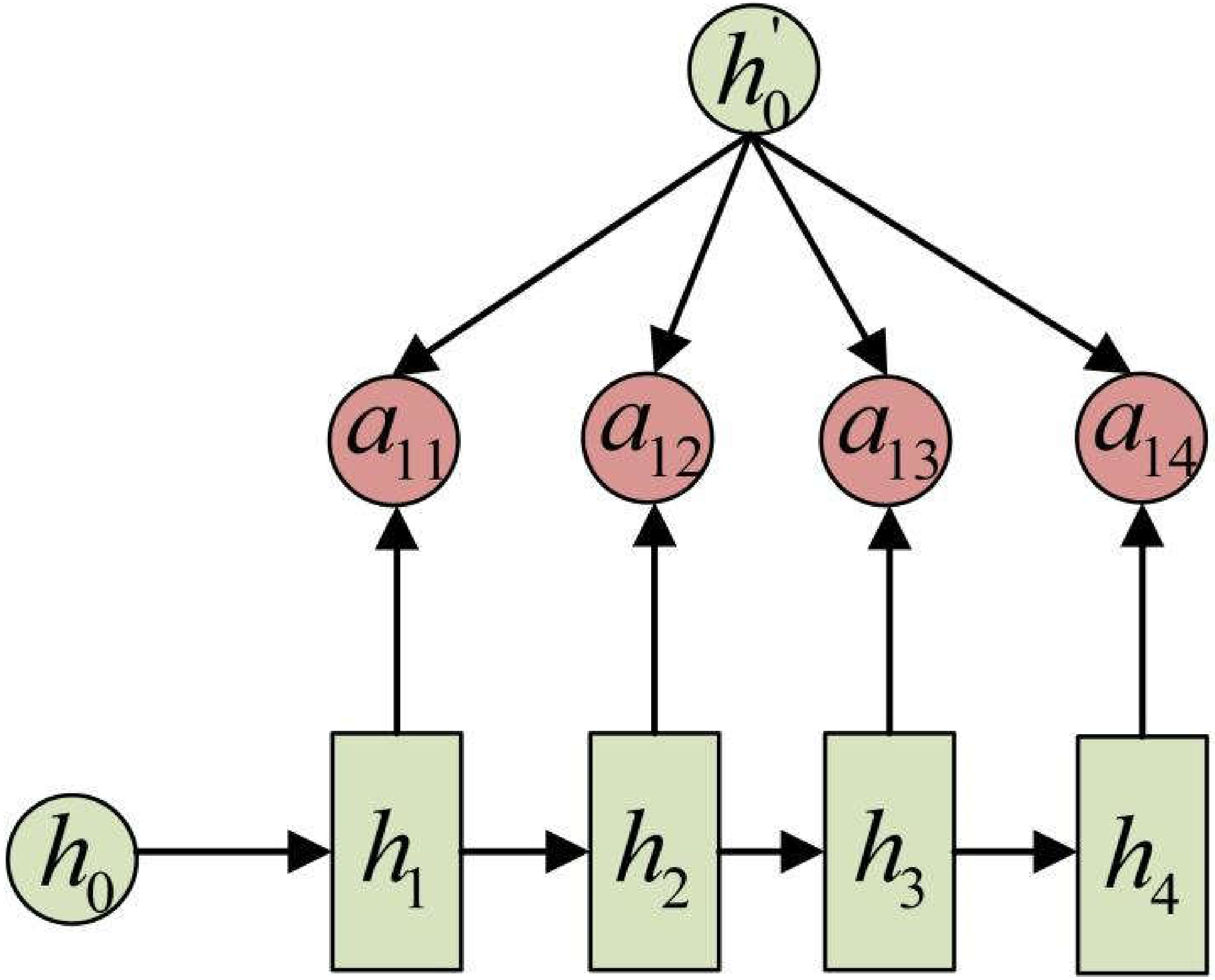
Attention Based Model



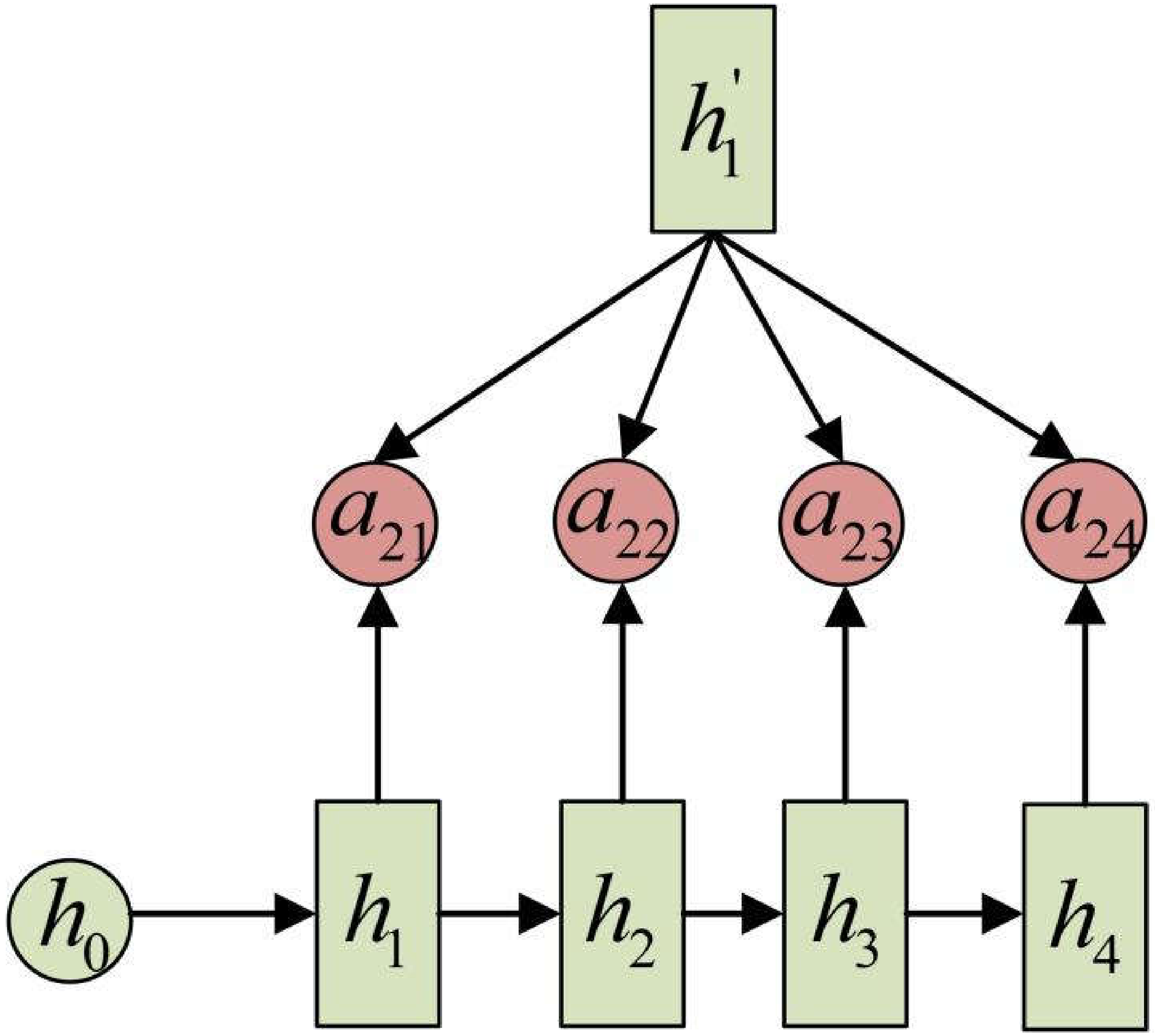
Attention Based Model



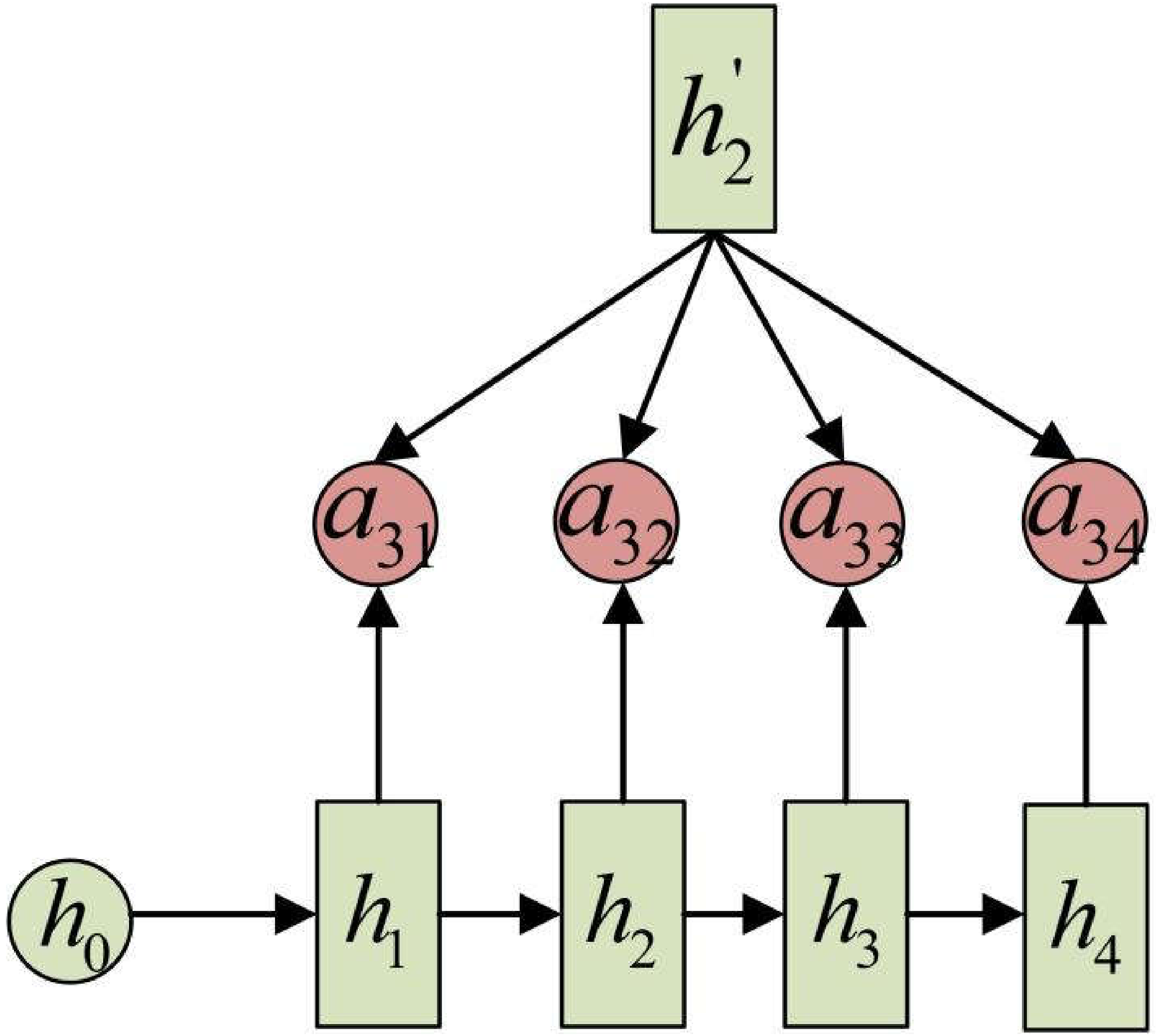
Attention Based Model



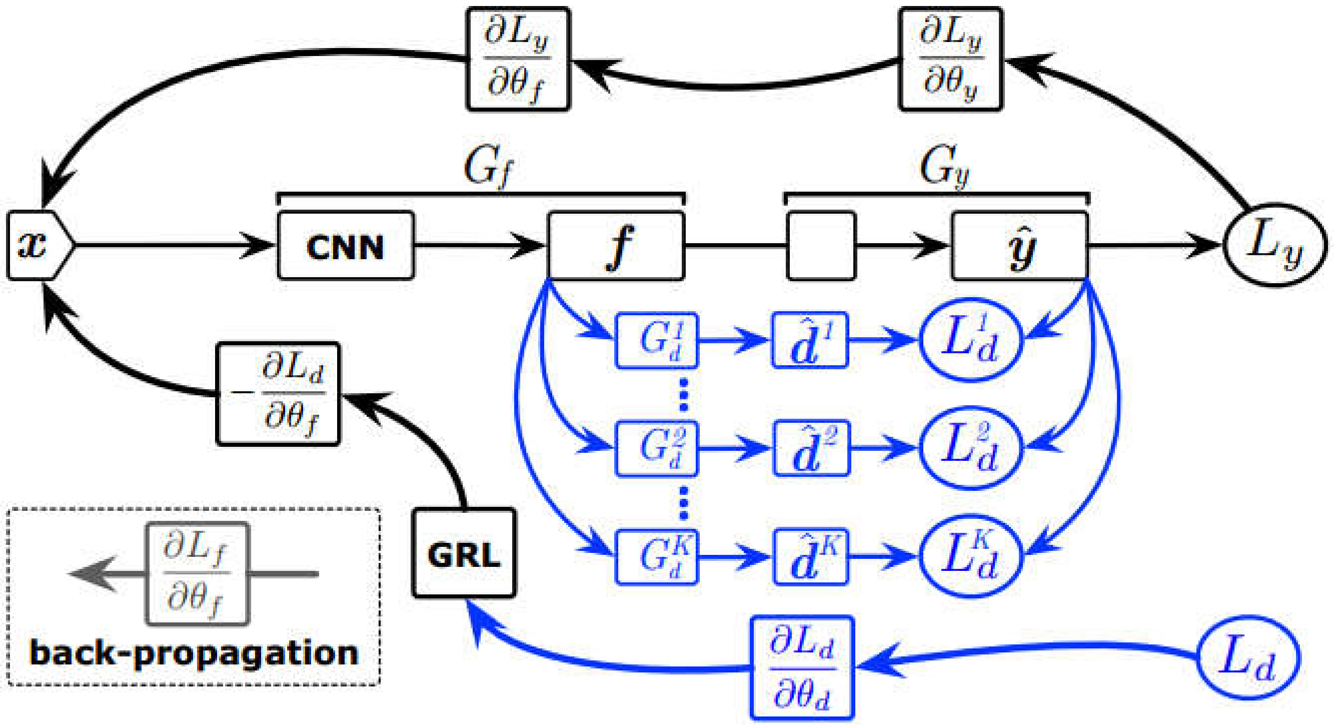
Attention Based Model



Attention Based Model



Selective Adversarial Network (SAN)

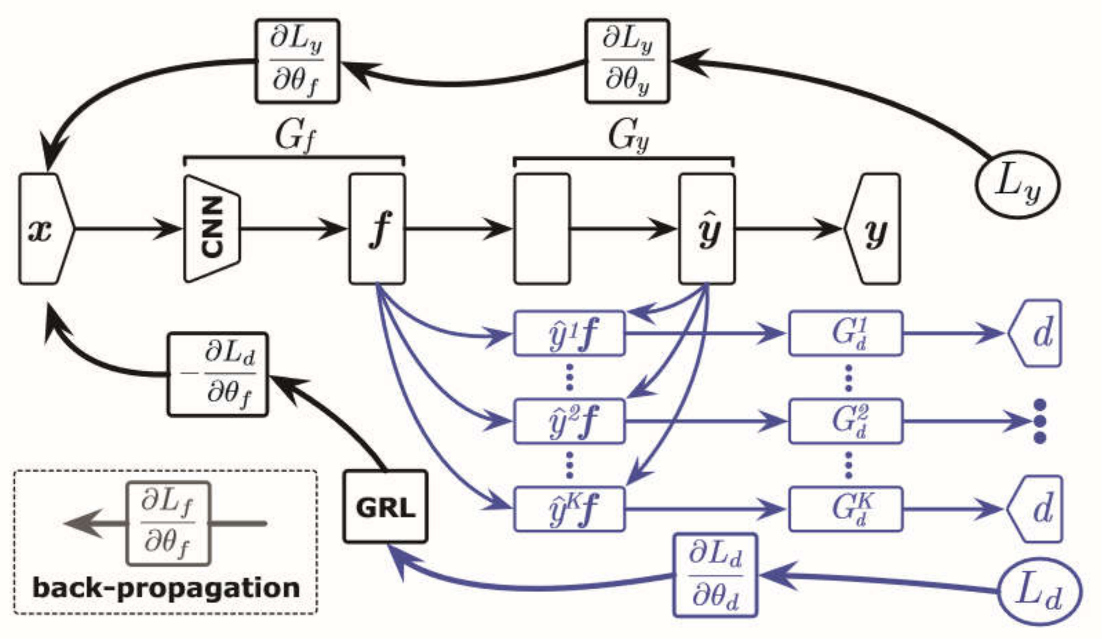


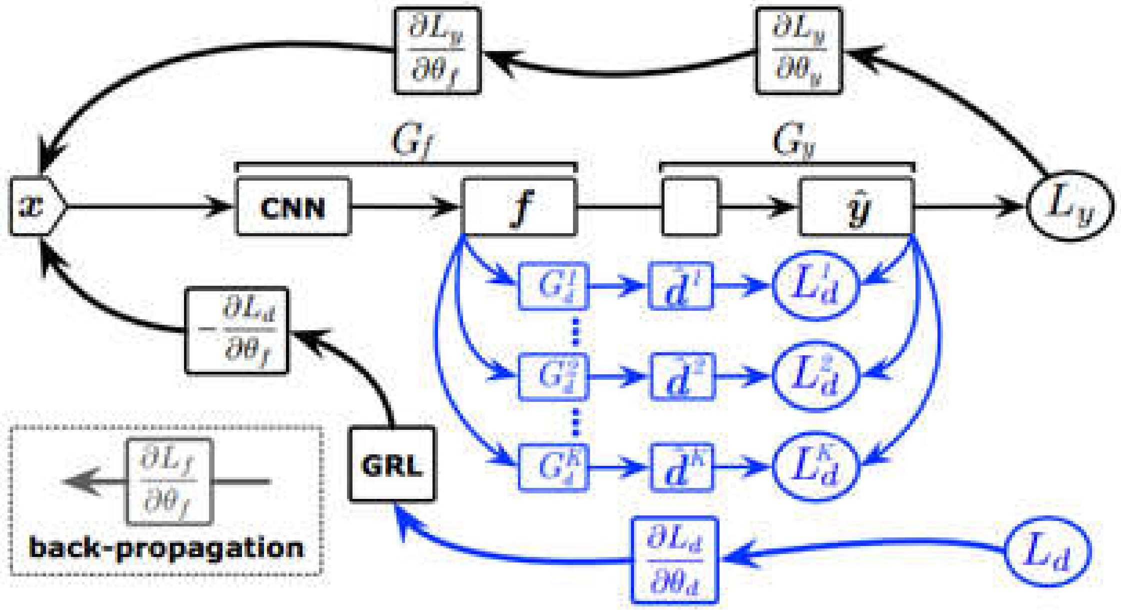
$$C\left(\theta_f,\theta_y,\theta_d^k\big|_{k=1}^{|\mathcal{C}_s|}\right)=\frac{1}{n_s}\sum_{\mathbf{x}_i\in\mathcal{D}_s}L_y\left(G_y\left(G_f\left(\mathbf{x}_i\right)\right),y_i\right)+\frac{1}{n_t}\sum_{\mathbf{x}_i\in\mathcal{D}_t}H\left(G_y\left(G_f\left(\mathbf{x}_i\right)\right)\right)$$

$$-\frac{\lambda}{n_s+n_t}\sum_{k=1}^{|\mathcal{C}_s|}\left[\frac{1}{n_t}\sum_{\mathbf{x}_i\in\mathcal{D}_t}\hat{y}_i^k\right]\sum_{\mathbf{x}_i\in\mathcal{D}_s\cup\mathcal{D}_t}\hat{y}_i^k L_d^k\left(G_d^k\left(G_f\left(\mathbf{x}_i\right)\right),d_i\right)$$

http://ise.thss.tsinghua.edu.cn/~mlong/doc/transfer-learning-talk.pdf

MADA vs SAN





AAAI2018: 2017-09-11 2017-11-28

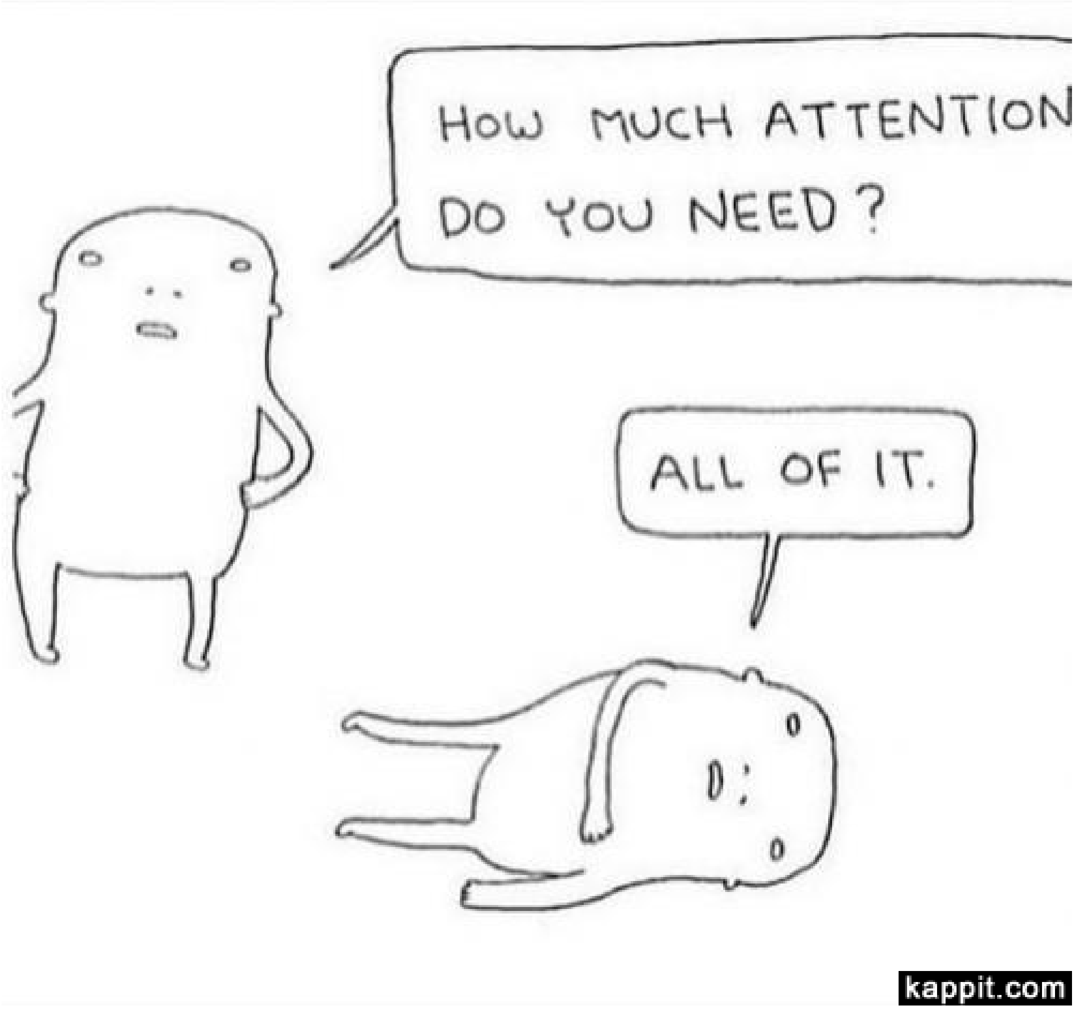
CVPR2018: 2017-11-15 2018-02-19

References

References

- Everything about Transfer Learning and Domain Adaptation
- Mingsheng Long’s Homepage
- 什么是多模态机器学习？
- 迁移学习--综述
- 完全图解RNN、RNN变体、Seq2Seq、Attention机制
- 一文读懂生成对抗网络（GANs）
- Partial Transfer Learning with Selective Adversarial Networks

http://jd92.wang/



Thanks !