

Beijing Forest Studio
北京理工大学信息系统及安全对抗实验中心



深度域适应方法

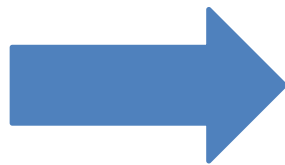
硕士研究生 巩锟

2022年11月06日

- 背景简介
- 基础概念
 - 迁移学习
 - 深度域适应
 - 深度域适应基础算法
- 算法原理
 - CDAC
- 总结
- 参考文献

- 预期收获
 - 1. 了解迁移学习和域适应的基本概念
 - 2. 理解无监督深度域适应基本算法
 - 3. 理解半监督深度域适应的基本原理
 - 4. 了解深度域适应的应用和前沿发展

- 为什么需要迁移学习

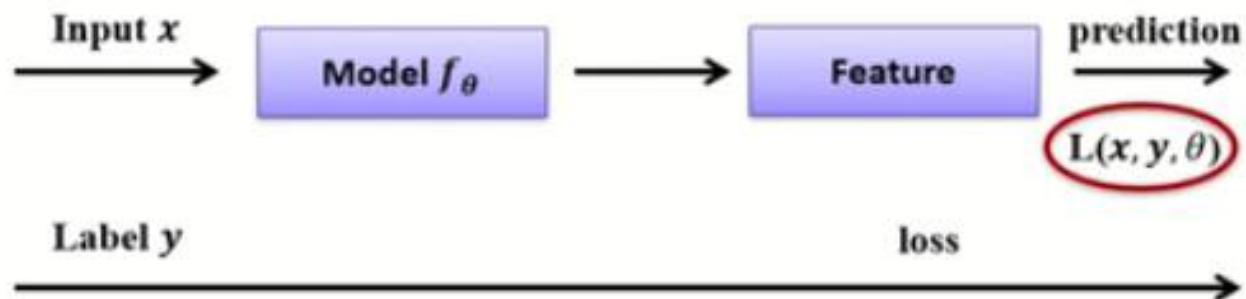


- 在一个领域学到的知识很好地迁移到相关的领域
- 让机器具备举一反三的能力
 - 现实问题：数据少或有标签数据少
 - 使之适应不同场景

- 深度迁移学习

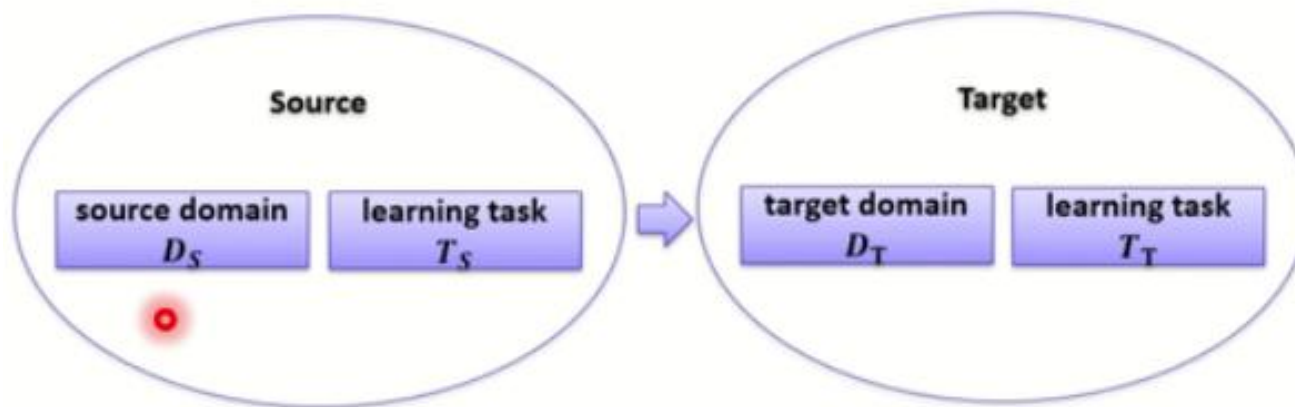
- 深度学习

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



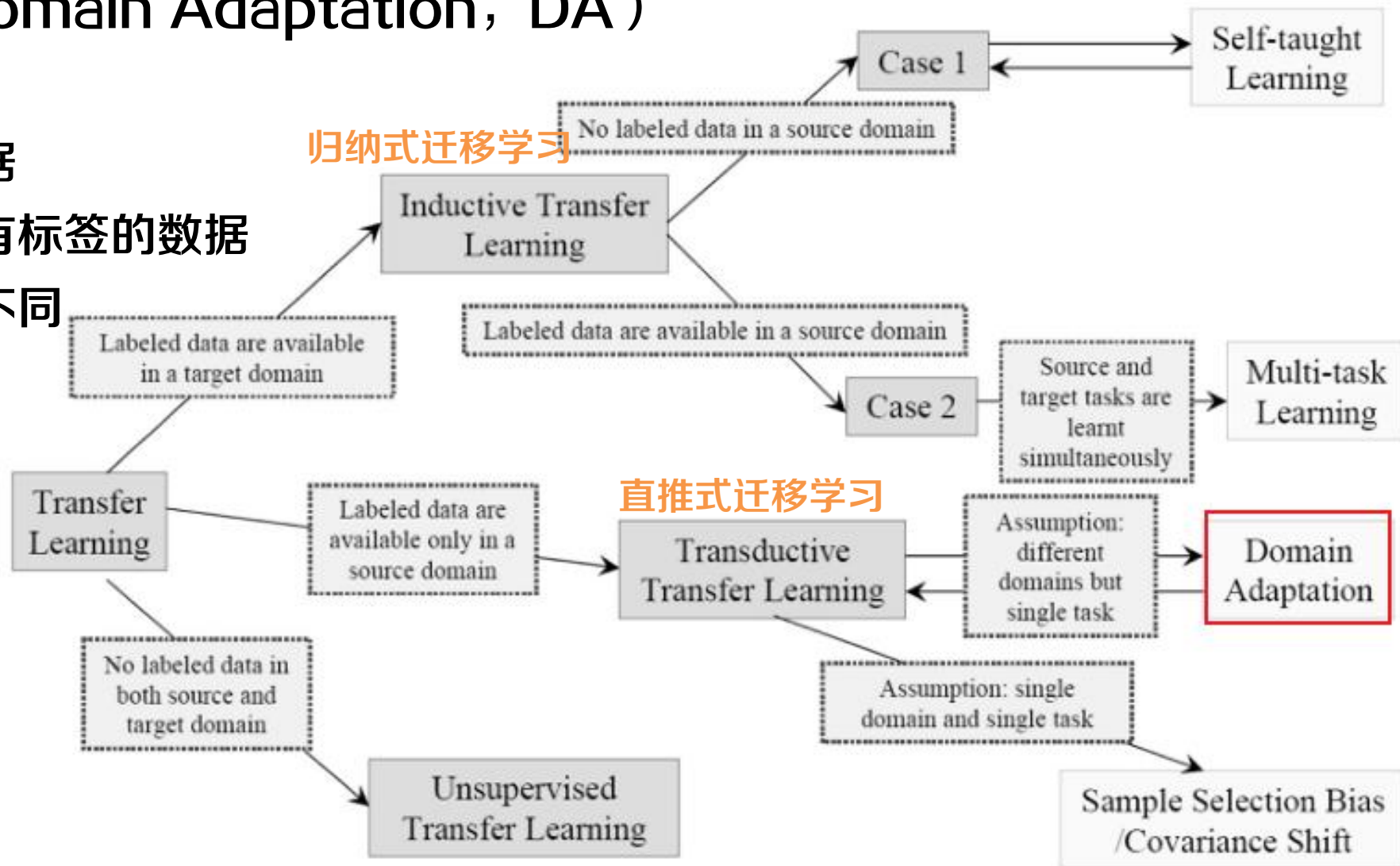
- 迁移学习

- 数据：源域和目标域不一样
 - 任务：分类或回归



- 迁移学习与域适应 (Domain Adaptation, DA)
- 域适应

- 源域：大量带标签数据
- 目标域：少量或者没有标签的数据
- 任务相同，数据分布不同

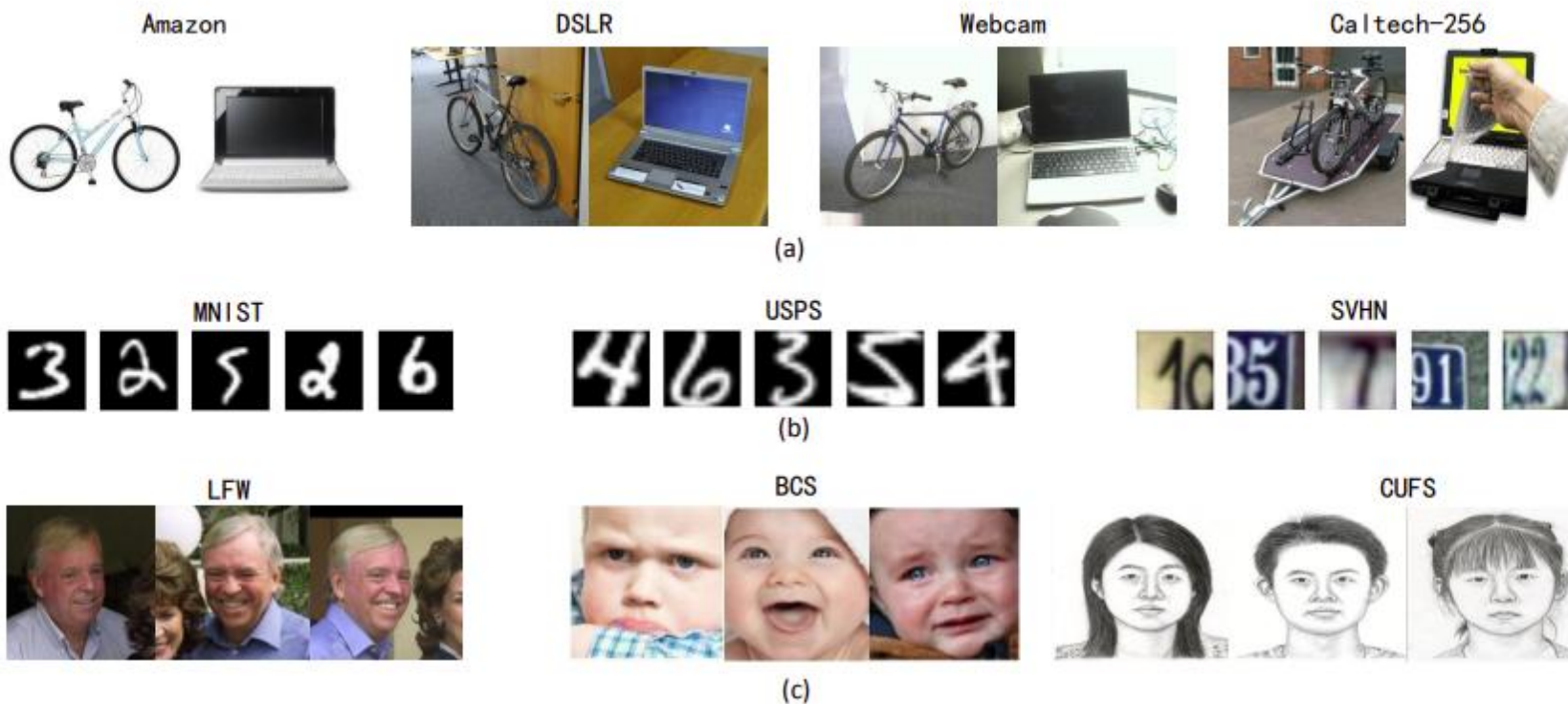


- 迁移学习与其他任务的关系
 - 迁移学习：任务不一定相同
 - 域适应：任务相同，数据分布不同

Notation	Description
$\mathbf{x}, y$	Instance/label
$\mathcal{X}, \mathcal{Y}$	Feature/label space
$\mathcal{S}$	Domain
$P(\cdot)$	Distribution
$\mathbb{E}[\cdot]$	Expectation
M	Number of source domain

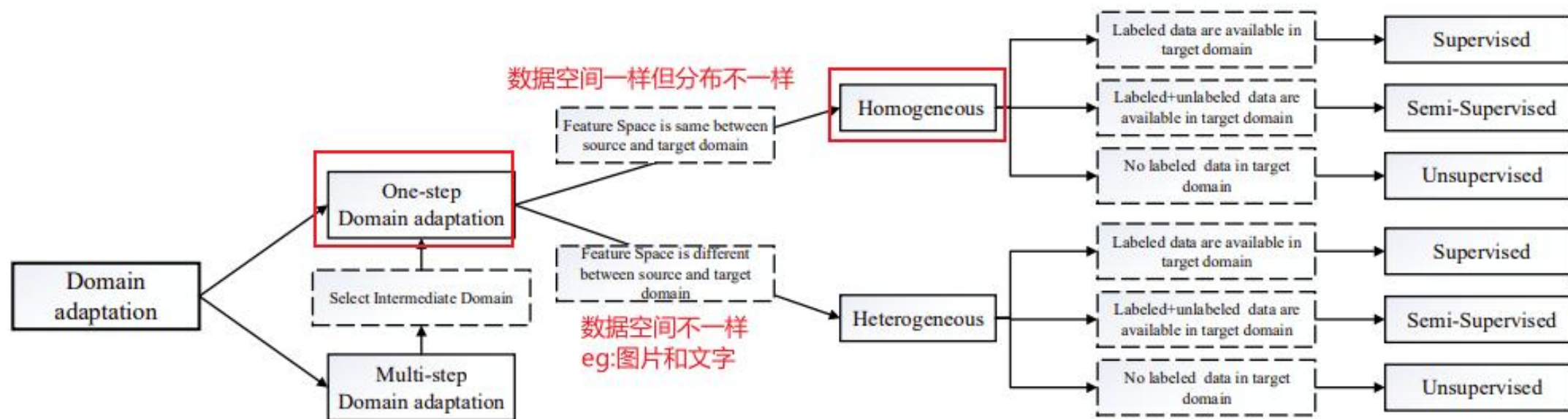
Learning paradigm	Training data	Test data	Condition	Test access
Multi-task learning	$\mathcal{S}^1, \dots, \mathcal{S}^n$	$\mathcal{S}^1, \dots, \mathcal{S}^n$	$\mathcal{Y}^i \neq \mathcal{Y}^j, 1 \leq i \neq j \leq n$	✓
Transfer learning	$\mathcal{S}^{src}, \mathcal{S}^{tar}$	$\mathcal{S}^{tar}$	$\mathcal{Y}^{src} \neq \mathcal{Y}^{tar}$	✓
Domain adaptation	$\mathcal{S}^{src}, \mathcal{S}^{tar}$	$\mathcal{S}^{tar}$	$P(\mathcal{X}^{src}) \neq P(\mathcal{X}^{tar})$	✓
Meta-learning	$\mathcal{S}^1, \dots, \mathcal{S}^n$	$\mathcal{S}^{n+1}$	$\mathcal{Y}^i \neq \mathcal{Y}^j, 1 \leq i \neq j \leq n + 1$	✓
Lifelong learning	$\mathcal{S}^1, \dots, \mathcal{S}^n$	$\mathcal{S}^1, \dots, \mathcal{S}^n$	$\mathcal{S}^i$ arrives sequentially	✓
Zero-shot learning	$\mathcal{S}^1, \dots, \mathcal{S}^n$	$\mathcal{S}^{n+1}$	$\mathcal{Y}^{n+1} \neq \mathcal{Y}^i, 1 \leq i \leq n$	×
Domain generalization	$\mathcal{S}^1, \dots, \mathcal{S}^n$	$\mathcal{S}^{n+1}$	$P(\mathcal{S}^i) \neq P(\mathcal{S}^j), 1 \leq i \neq j \leq n + 1$	×

- 什么是不同的数据分布？
 - 由于许多因素，两个域之间的分布变化或域偏移可能会降低性能
 - 源域和目标域输入边缘分布不同，条件分布相同



- 域适应（DA）分类

- 监督DA：目标域有大量带标签的数据
- 半监督DA：目标域有少量带标签的数据
- 无监督DA：目标域的数据不带标签



- 域适应方法

- 深度学习

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

- 特征适应 (共同特征空间)

$$\min \frac{1}{n} \sum_{i=1}^n L(\phi(x_i^s), y_i^s, \theta)$$

- KL divergence
- MMD
- H-divergence
- Wasserstein distance

- 实例适应 (源域和目标域像的样本)

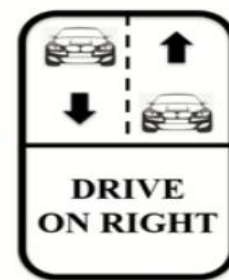
$$\min \frac{1}{n} \sum_{i=1}^n \omega_i L(x_i^s, y_i^s, \theta)$$

- 参数适应

$$\min \frac{1}{n} \sum_{i=1}^n \omega_i L(x_i^s, y_i^s, \theta)$$



VS

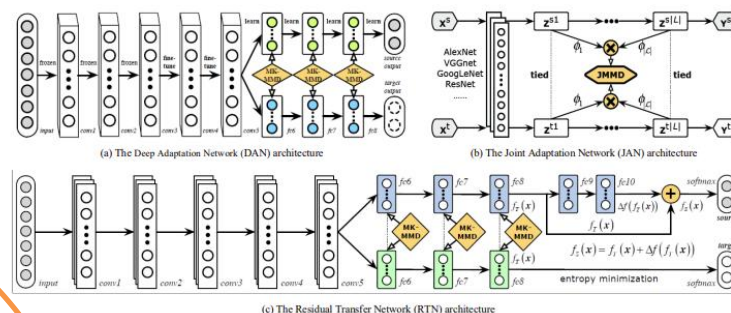
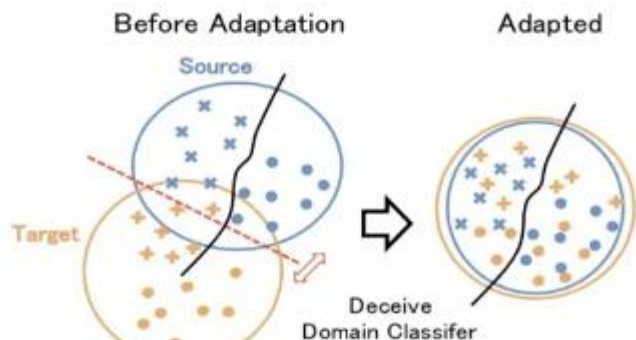


深度域适应方法（基于特征的适应）

用两个域的数据**微调**神经网络以减小域转移

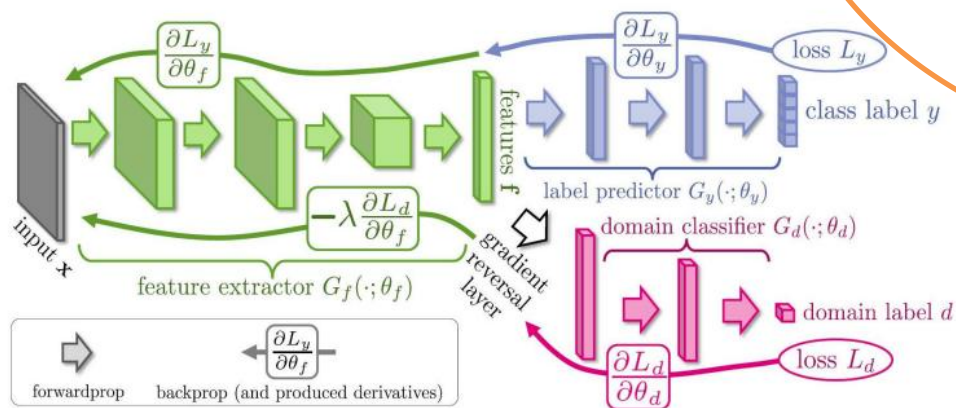
基于差异的域适应方法

Domain Adaptation



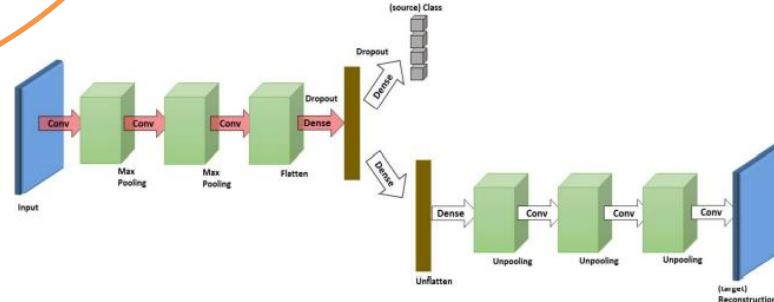
用域判别器进行**域混淆**

基于对抗的域适应方法



使用**数据重建**作为辅助任务，以确保特征不变性

基于重构的域适应方法

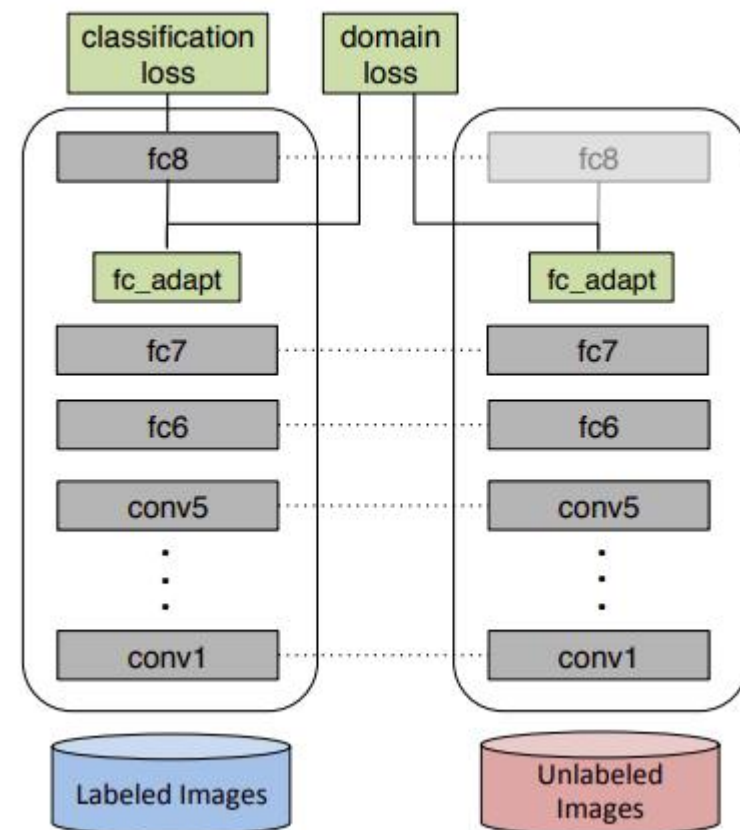


- 基于差异的域适应方法（Maximum mean discrepancy, **MMD**）
 - 无监督域自适应（Unsupervised Domain Adaptation, UDA）常用的损失函数
 - 找一个核函数，将不同域的特征表示映射到再生核的Hilbert空间，取**均值的差**并作为他们之间的距离

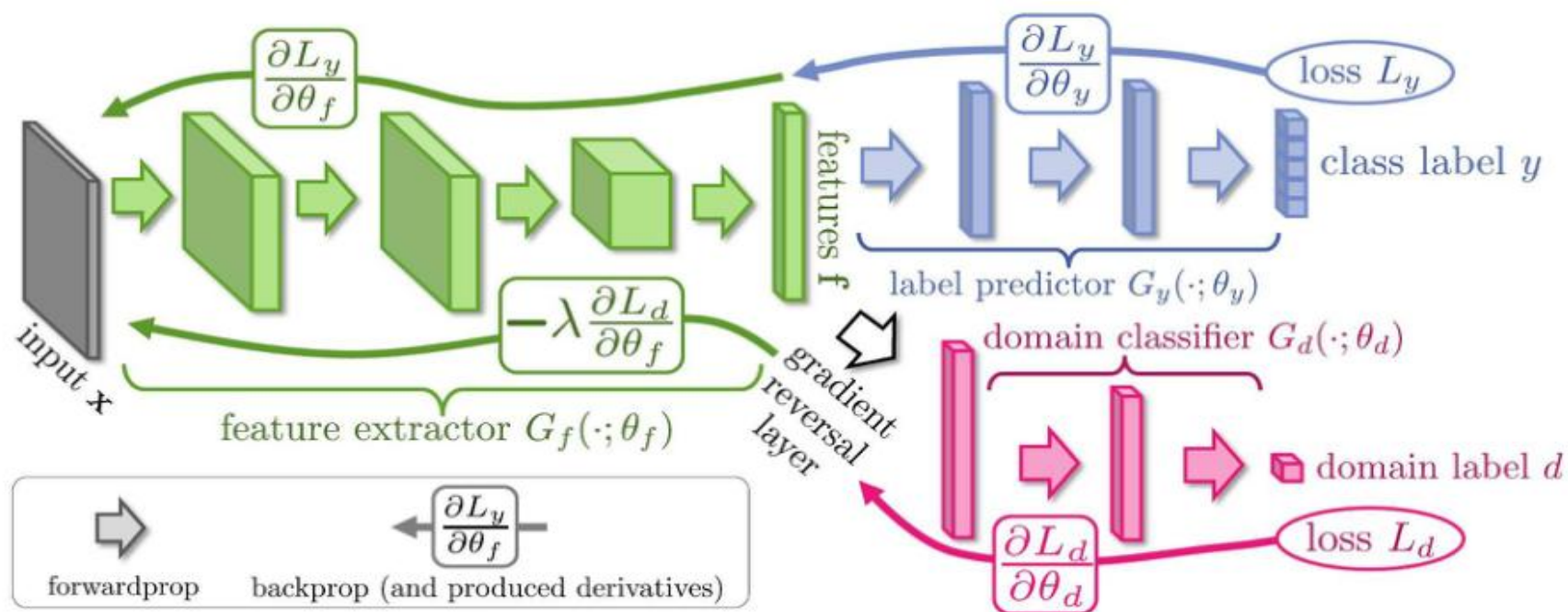
$$MMD^2(D_s, D_t) = \left\| \frac{1}{M} \sum_{i=1}^M \phi(x_i^s) - \frac{1}{N} \sum_{j=1}^N \phi(x_j^t) \right\|_H^2$$

- 损失函数计算

$$\mathcal{L} = \mathcal{L}_C(X^L, y) + \lambda MMD^2(X^s, X^t)$$



- 基于对抗的域适应方法 (DANN)
 - Max-min optimization
 - Gradient reversal layer



- 基于对抗的域适应方法（DANN）
 - Max-min optimization

$$E(\theta_f, \theta_y, \theta_d) = \sum_{\substack{i=1..N \\ d_i=0}} L_y(G_y(G_f(x_i; \theta_f); \theta_y), y_i) - \lambda \sum_{i=1..N} L_d(G_d(G_f(x_i; \theta_f); \theta_d), y_i)$$

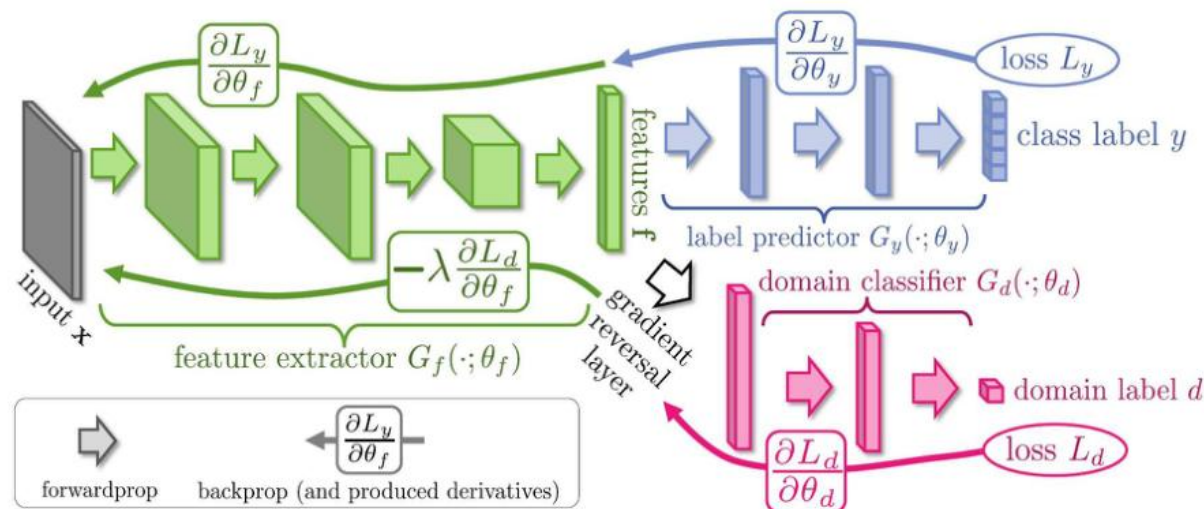
$$(\hat{\theta}_f, \hat{\theta}_y) = \arg \min_{\theta_f, \theta_y} E(\theta_f, \theta_y, \hat{\theta}_d)$$

$$\hat{\theta}_d = \arg \max_{\theta_d} E(\hat{\theta}_f, \hat{\theta}_y, \theta_d)$$

- Gradient reversal layer

$$R_\lambda(x) = x$$

$$\frac{dR_\lambda}{dx} = -\lambda I$$



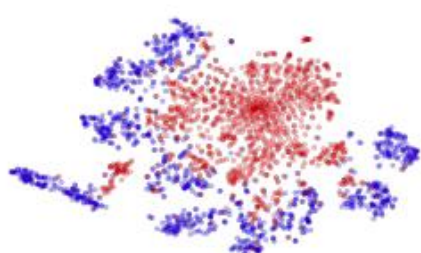
- 基于对抗的域适应方法（DANN）



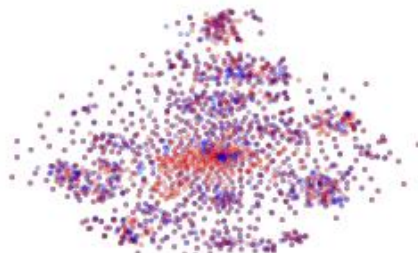
Figure 6: Examples of domain pairs used in the experiments. See Section 5.2.4 for details.

METHOD	SOURCE	MNIST	SYN NUMBERS	SVHN	SYN SIGNS
	TARGET	MNIST-M	SVHN	MNIST	GTSRB
SOURCE ONLY		.5225	.8674	.5490	.7900
SA (Fernando et al., 2013)		.5690 (4.1%)	.8644 (-5.5%)	.5932 (9.9%)	.8165 (12.7%)
DANN		.7666 (52.9%)	.9109 (79.7%)	.7385 (42.6%)	.8865 (46.4%)
TRAIN ON TARGET		.9596	.9220	.9942	.9980

MNIST → MNIST-M: top feature extractor layer

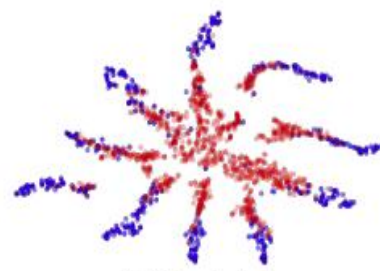


(a) Non-adapted

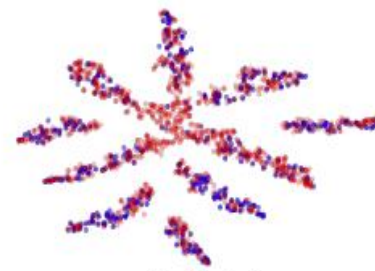


(b) Adapted

SYN NUMBERS → SVHN: last hidden layer of the label predictor

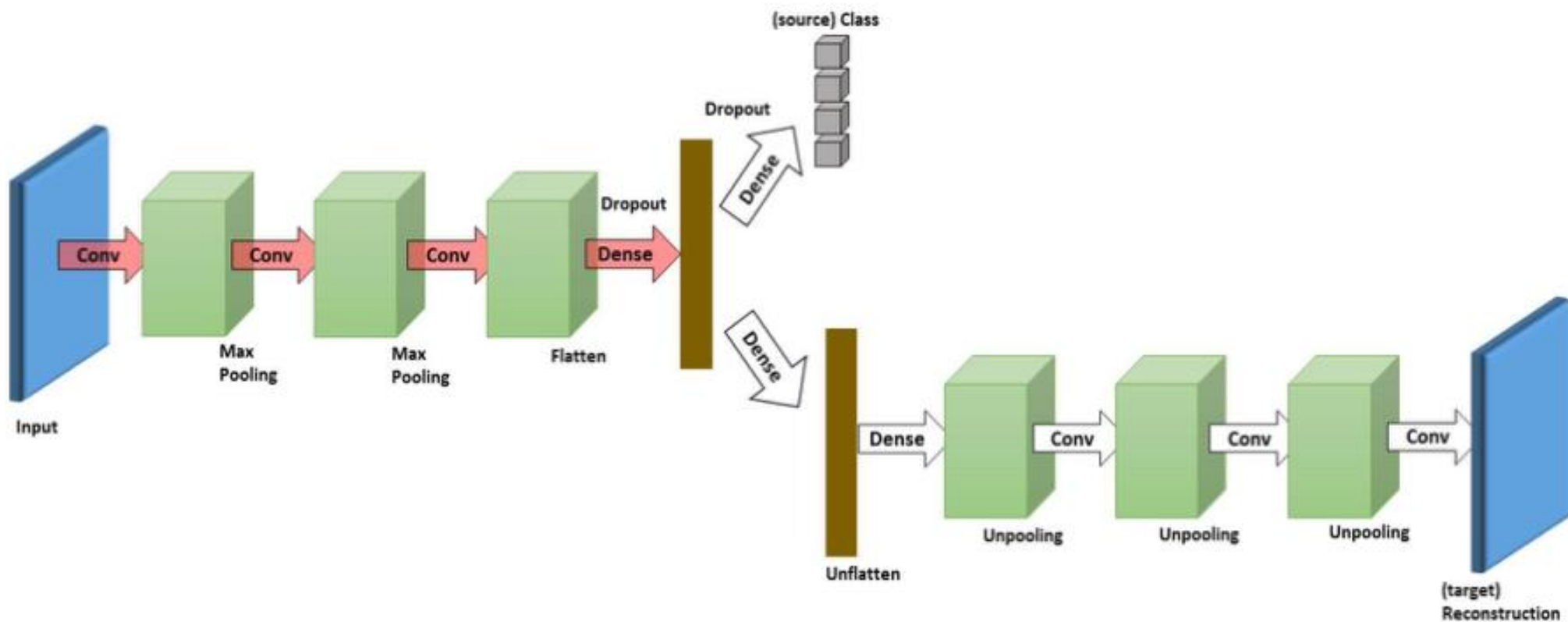


(a) Non-adapted



(b) Adapted

- 基于重构的域适应方法（DRCN）
 - 源标签监督分类
 - 目标数据无监督重建





CDAC

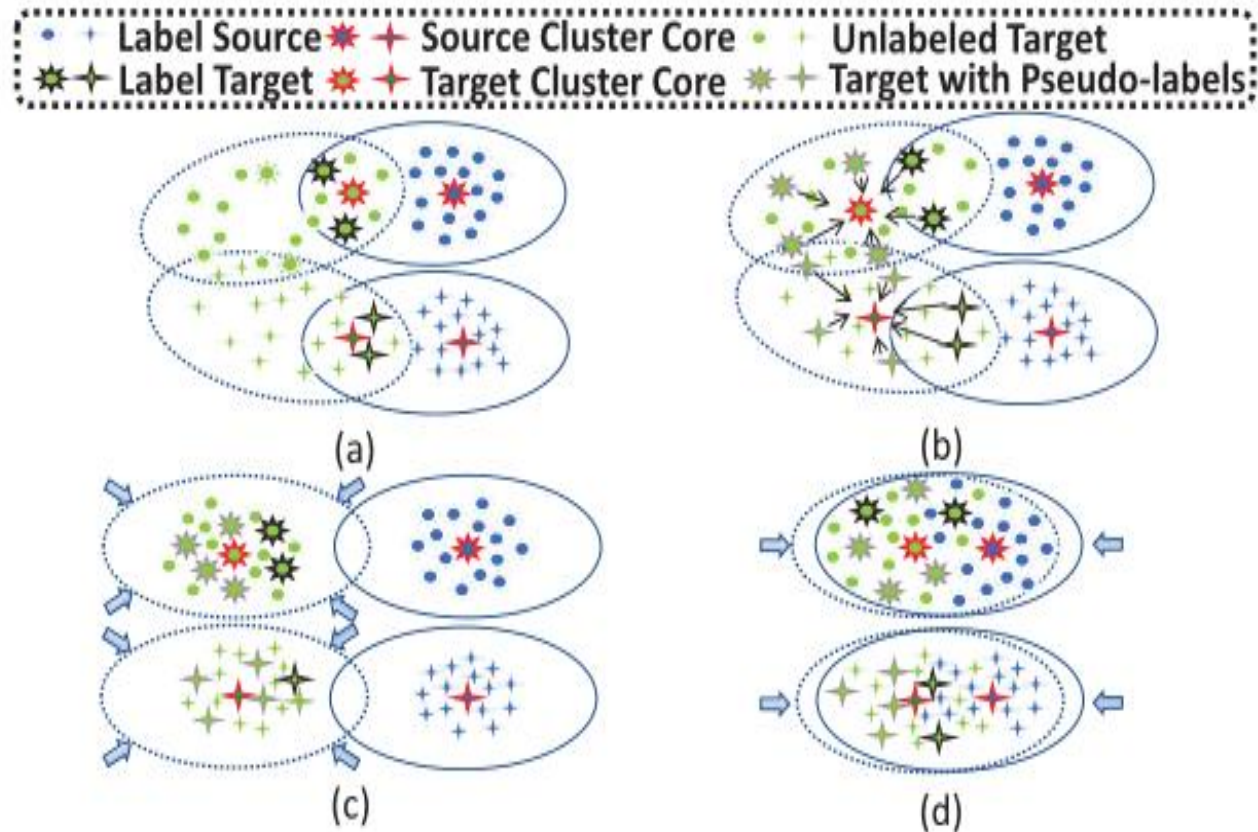
**Cross-Domain Adaptive Clustering
for Semi-Supervised Domain**

T	降低跨域特征失配可能性，提高自适应模型泛化性能
I	源域数据、目标域数据
P	1.带标签的源和目标域数据训练网络 2.打伪标签，增强对抗性自适应聚类损失； 3.计算一致性损失 4.模型优化
O	目标域分类结果

P	解决跨域特征难以匹配的问题，提高模型泛化性
C	源域数据带标签，目标域数据带少量标签
D	如何实现源域和目标域逐簇特征对齐
L	CVPR 2021(CCF A)

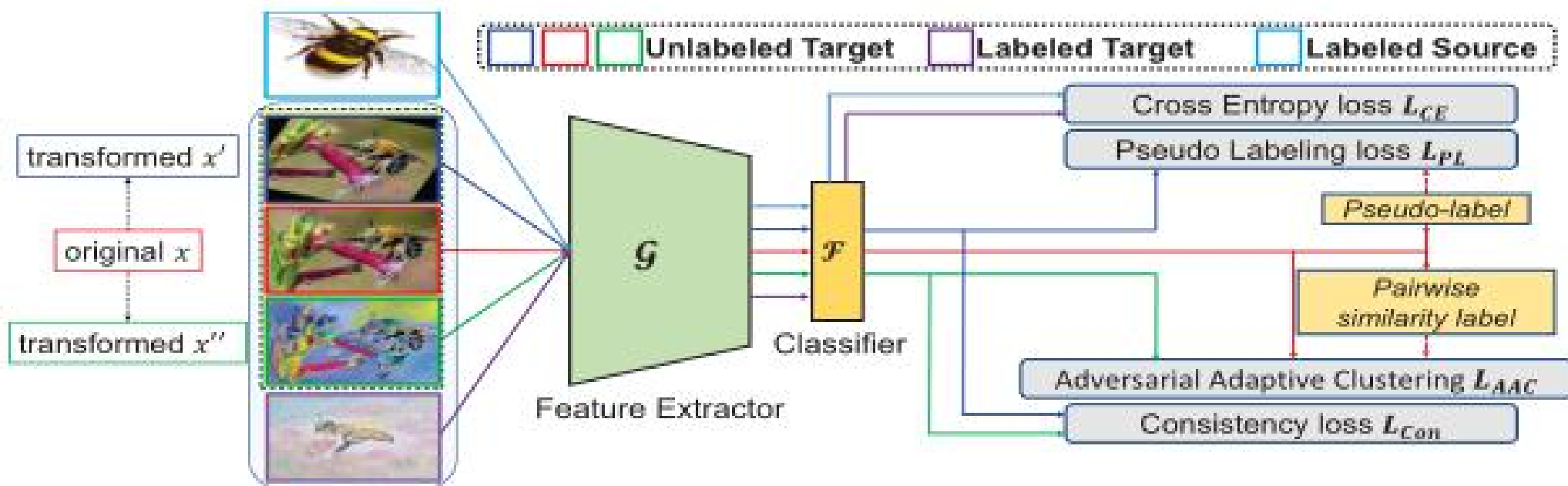
• Overview

- a) 有监督确保部分跨域特征对齐
- b) 伪标记，形成增强目标簇
- c) 最小化目标域特征簇的对抗性自适应聚类损失
- d) 最大化对抗性自适应聚类损失以促进跨域聚类特征对齐



- 模型框架

- 带标签数据训练网络
- 在未标记的目标数据上添加损失，以进一步优化模型
 - L_{AAC} 、 L_{pL} 、 L_{Con}



- 对抗自适应聚类（特征对齐）

- 将目标域特征分组到簇中，执行跨域聚类特征对齐，同时实现域间自适应和域内自适应

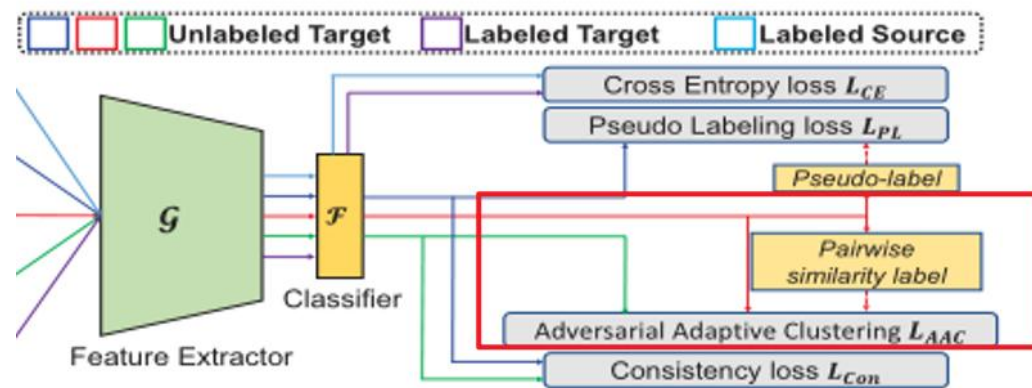
- 计算目标域未标记样本特征之间的成对相似性
 - 训练分类器，使之具有成对特征相似性的样本预测标签保持一致

- Pairwise similar label

- 特征相似度计算
 - 标记索引排序
 - 成对样本伪标记

- 配对样本 $s_{ij} = 1$;否则 $s_{ij} = 0$

$$s_{ij} = \mathbb{1}\{\text{top}k(\mathcal{G}(x_i^u)) = \text{top}k(\mathcal{G}(x_j^u))\}$$

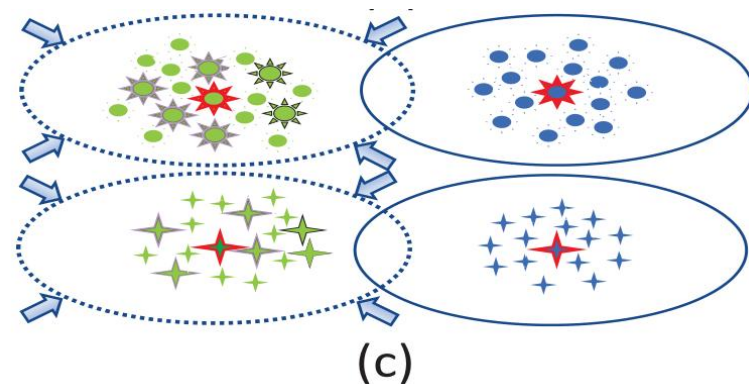


对抗自适应聚类

– 对抗自适应聚类损失

- **最小化**使目标域相似样本特征接近，不同样本特征远离，从而使特征在目标域形成簇。

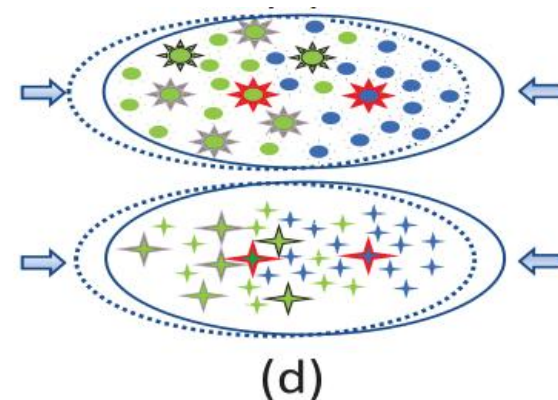
$$L_{AAC} = - \sum_{i=1}^M \sum_{j=1}^M s_{ij} \log(\mathbf{p}_i^T \mathbf{p}'_j) + (1 - s_{ij}) \log(1 - \mathbf{p}_i^T \mathbf{p}'_j)$$



– 梯度反转层

$$\theta_{\mathcal{G}}^* = \arg \min_{\theta_{\mathcal{G}}} \mathbf{L}_{CE} + \lambda \mathbf{L}_{AAC},$$

$$\theta_{\mathcal{F}}^* = \arg \min_{\theta_{\mathcal{F}}} \mathbf{L}_{CE} - \lambda \mathbf{L}_{AAC},$$



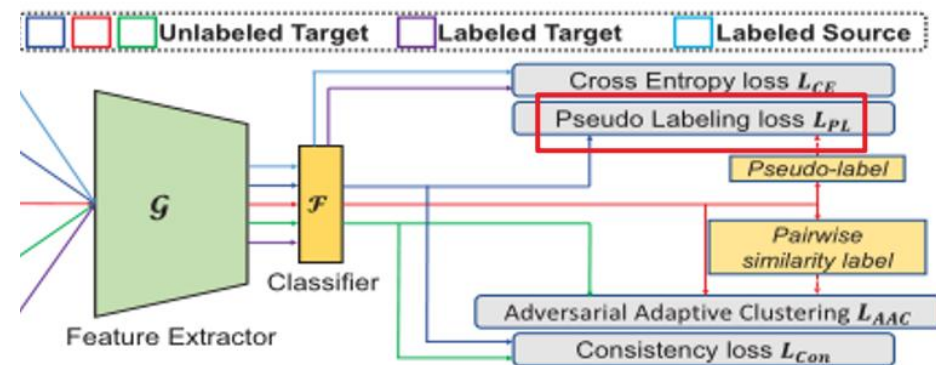
• 类伪标签

- L_{AAC} 很难在模型训练期间在目标域中形成稳定和准确的聚类核
- 渐进伪标记技术 (FixMatch)

$$\mathbf{p}_j = p(x_j^u) = \sigma(\mathcal{F}(\mathcal{G}(x_j^u)))$$

$$\hat{y}_j^u = \arg \max(\mathbf{p}_j)$$

$$L_{PL} = - \sum_{j=1}^M \mathbb{1}\{\max(\mathbf{p}_j) \geq \tau\} \cdot \hat{y}_j^u \log(p(x_j'')),$$



- L_{pL} 损失用于增强对抗性自适应聚类损失
 - 识别出具有高置信度的伪标签并将其用于模型训练
 - 使目标域特征簇与源域中的更好对齐

• 总体损失

– 一致性损失

- 确保模型对两个数据增强后的图片预测保持一致

$$L_{Con} = w(t) \sum_{j=1}^M \|p'_j - p''_j\|^2$$

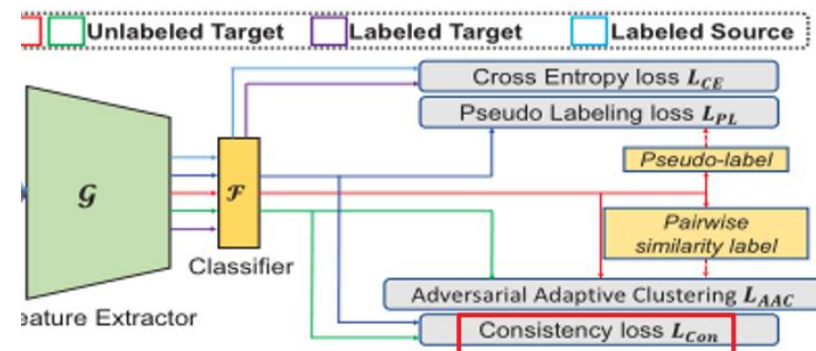
$$w(t) = \nu e^{-5(1-\frac{t}{T})^2}$$

- $w(t)$ 是斜升函数 (ICLR 2017)。对于有监督损失和无监督损失，利用与时间有关的权重函数来加权， $w(t)$ 的分布类似于高斯函数的上升区间，随着训练次数的增加，无监督损失的权重占比越高。

– 总体损失

$$\theta_G^* = \arg \min_{\theta_G} L_{CE} + \lambda L_{AAC} + L_{PL} + L_{Con}$$

$$\theta_F^* = \arg \min_{\theta_F} L_{CE} - \lambda L_{AAC} + L_{PL} + L_{Con}$$



- 实验设计

- 数据集

- DomainNet

- 多元域适应数据集

- 只选择4个域，分别是Real、Clipart、Painting和Sketch（RCPS），每个域包含126类别的图像。

- Office Home

- UDA数据集

- 4个域，Real、Clipart、Art和Product（RCaP），共有65类

- Office

- 3个域，单反、网络摄像头和亚马逊（DWA），共有31个类别

- 基准数据集设置

- 参考现有SSDA方法，标记目标数据的数量（每类1-shot或3-shot）

- 实验细节

- 主干网络G

- Alexnet和Resnet网络

- 初始化

- 在ImageNet4上预训练模型初始化特征提取器，并随机初始化线性分类层

- 训练方式

- 在源域和目标域的标记数据上使用标准交叉熵损失训练模型
 - 在未标记的目标数据上添加提出的损失，以进一步优化模型

$$\theta_G^* = \arg \min_{\theta_G} L_{CE} + \lambda L_{AAC} + L_{PL} + L_{Con}$$

$$\theta_F^* = \arg \min_{\theta_F} L_{CE} - \lambda L_{AAC} + L_{PL} + L_{Con}$$

实验分析(On DomianNet)

Table 1. Accuracy(%) on *DomainNet* under the settings of 1-shot and 3-shot using Alexnet and Resnet34 as backbone networks.

Net	Method	R→C		R→P		P→C		C→S		S→P		R→S		P→R		Mean	
		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
Alexnet	S+T	43.3	47.1	42.4	45.0	40.1	44.9	33.6	36.4	35.7	38.4	29.1	33.3	55.8	58.7	40.0	43.4
	DANN	43.3	46.1	41.6	43.8	39.1	41.0	35.9	36.5	36.9	38.9	32.5	33.4	53.5	57.3	40.4	42.4
	ENT	37.0	45.5	35.6	42.6	26.8	40.4	18.9	31.1	15.1	29.6	18.0	29.6	52.2	60.0	29.1	39.8
	MME	48.9	55.6	48.0	49.0	46.7	51.7	36.3	39.4	39.4	43.0	33.3	37.9	56.8	60.7	44.2	48.2
	Meta-MME	-	56.4	-	50.2	-	51.9	-	39.6	-	43.7	-	38.7	-	60.7	-	48.8
	BiAT	54.2	58.6	49.2	50.6	44.0	52.0	37.7	41.9	39.6	42.1	37.2	42.0	56.9	58.8	45.5	49.4
	APE	47.7	54.6	49.0	50.5	46.9	52.1	38.5	42.6	38.5	42.2	33.8	38.7	57.5	61.4	44.6	48.9
	CDAC	56.9	61.4	55.9	57.5	51.6	58.9	44.8	50.7	48.1	51.7	44.1	46.7	63.8	66.8	52.1	56.2
Resnet34	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	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
	ENT	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	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
	UODA	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
	Meta-MME	-	73.5	-	70.3	-	72.8	-	62.8	-	68.0	-	63.8	-	79.2	-	70.1
	BiAT	73.0	74.9	68.0	68.8	71.6	74.6	57.9	61.5	63.9	67.5	58.5	62.1	77.0	78.6	67.1	69.7
	APE	70.4	76.6	70.8	72.1	72.9	76.7	56.7	63.1	64.5	66.1	63.0	67.8	76.6	79.4	67.6	71.7
	CDAC	77.4	79.6	74.2	75.1	75.5	79.3	67.6	69.9	71.0	73.4	69.2	72.5	80.4	81.9	73.6	76.0

消融实验

- 验证每个损失项的有效性
- 数据集: DomainNet
- 网络: Resnet34
- 标记目标数据的数量: 3-shot
- 基线实验: 仅在两个域的标记样本使用交叉熵损失

Table 4. Accuracy(%) of CDAC using Resnet34 as the backbone on *DomainNet* under the setting of 3-shot. In the UDA setting, the supervised cross-entropy loss L_{CE} refers to the model trained only with labeled source samples.

Net	Setting	L_{CE}	L_{AAC}	L_{PL}	L_{Con}	R→C	R→P	P→C	C→S	S→P	R→S	P→R	Mean
Resnet34	UDA	✓				57.8	61.4	58.1	53.4	58.2	49.6	72.3	58.6
	UDA	✓	✓			64.6	64.8	64.8	59.9	62.2	58.8	73.0	64.0
	UDA	✓		✓		68.0	73.2	68.3	61.8	67.0	63.1	76.5	68.2
	UDA	✓	✓	✓		76.9	73.9	73.9	66.2	70.2	69.0	79.3	72.8
	UDA	✓	✓	✓	✓	77.1	74.4	73.2	67.0	70.4	69.6	79.6	73.0
	SSDA	✓				60.0	62.2	59.4	55.0	59.5	50.1	73.9	60.0
	SSDA	✓	✓			69.4	68.1	68.3	62.8	65.6	62.0	76.9	67.6
	SSDA	✓		✓		76.7	73.6	76.3	66.9	70.3	69.3	80.4	73.4
	SSDA	✓	✓	✓		78.7	74.9	78.5	69.7	73.2	71.1	81.6	75.3
	SSDA	✓	✓	✓	✓	79.6	75.1	79.3	69.9	73.4	72.5	81.9	76.0

- 对抗自适应聚类的有效性（右图）

- 使用**聚类核心距离**(CCD)测量同一类源和目标域特征聚类之间的距离

- 跨域特征簇越对齐，CCD越小

- 对比实验：S+T、CDAC w/o L_{PL}

- 结论

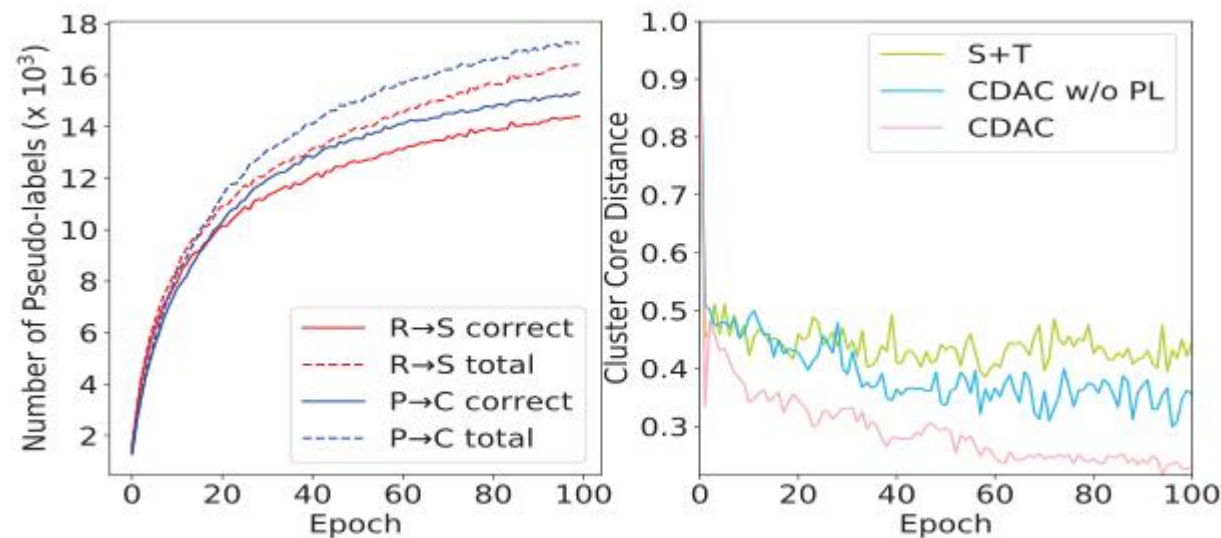
- 所有方法的CCD训练期间逐渐减小

- CDAC w/o L_{PL} 和CDAC可以更好进行特征对齐

- CDAC表现出最佳分类性能

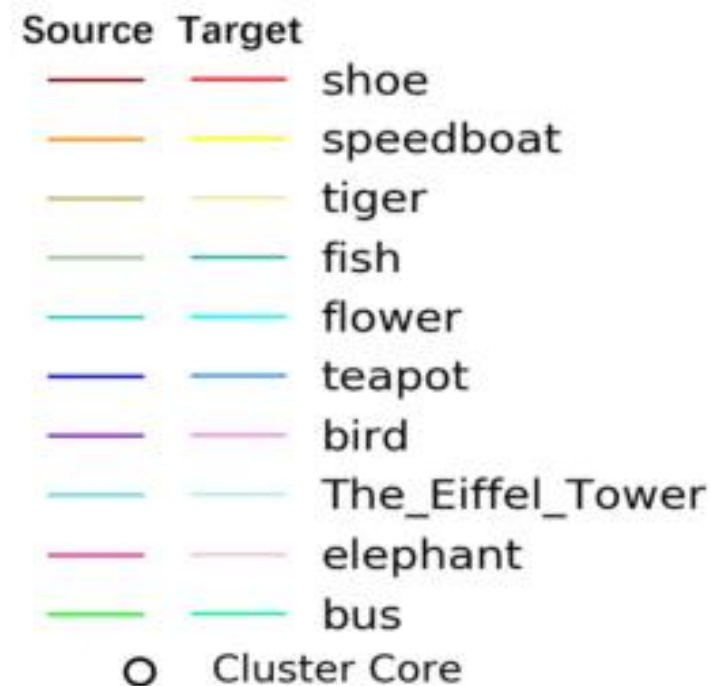
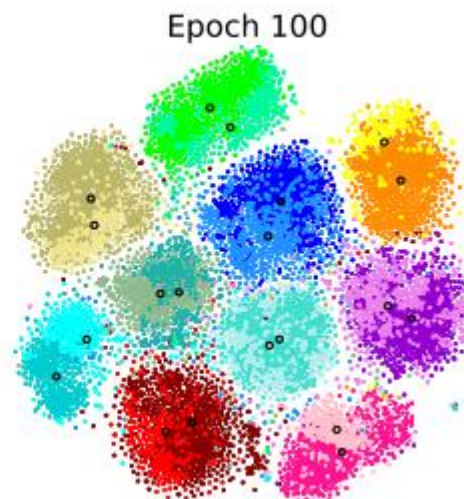
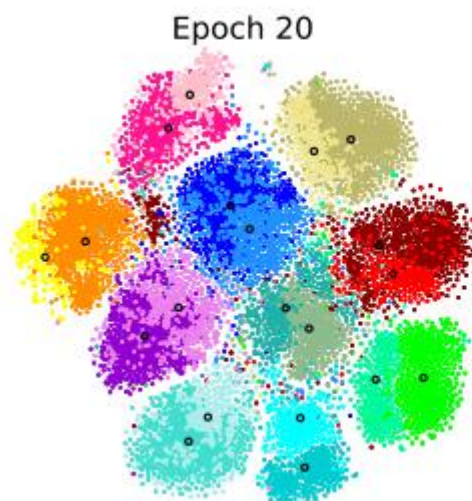
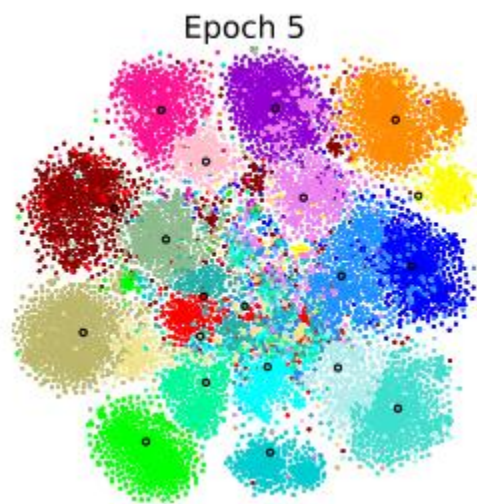
- 伪标记的有效性（左图）

- 两种自适应情况下每个Epoch总训练样本的59.9%和63.8%



- 特征可视化

- 可视化集群的变化及每个类对应的**集群核心**
- 结论
 - 目标特征逐渐向目标簇核收敛
 - 目标域每个簇逐渐向其对应的源簇核靠近



- 优势
 - 关注域内级别的类对齐，缓解域偏移问题
 - 增强目标域中集群核心的鲁棒性和能力，促进对抗性学习
- 劣势
 - 关注类对齐任务，对域内知识的利用不足



总结

- 应用

- 图像

- 目标检测
 - 人脸识别

- 自然语言处理

- 文本分类
 - 假新闻检测

- 发展前沿

- 多源/多目标域适应（多源迁移学习）
 - 域泛化

1 Image classification



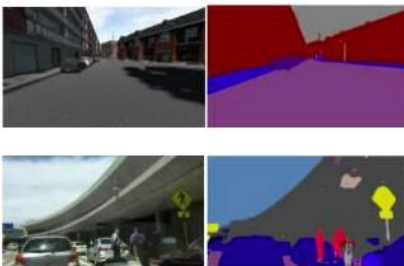
2 Face recognition



3 Object detection



4 Semantic segmentation



5 Person re-identification



6 Image-to-image translation



- [1] Wang M, Deng W. Deep visual domain adaptation: A survey[J]. Neurocomputing, 2018, 312: 135-153.**
- [2] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on knowledge and data engineering, 2009, 22(10): 1345-1359.**
- [3] Li J, Li G, Shi Y, et al. Cross-domain adaptive clustering for semi-supervised domain adaptation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 2505-2514.**
- [4]Wang J, Lan C, Liu C, et al. Generalizing to unseen domains: A survey on domain generalization[J]. IEEE Transactions on Knowledge and Data Engineering, 2022.**
- [5] <https://www.bilibili.com/video/BV1ct41167kV?p=1>**

谢谢！

大成若缺，其用不弊。大盈
若冲，其用不穷。大直若屈。
大巧若拙。大辩若讷。静胜
躁，寒胜热。清静为天下正。

