基于掩蔽图像一致性的无源领域自适应算法研究

刘二虎
安徽理工大学 计算机科学与工程学院,安徽 淮南 232001

摘 要:目的 无监督领域自适应(Unsupervised Domain Adaptation, UDA)旨在将有标签源域中的知识适应到没有任何标签的目标域中,使得目标任务同样表现良好。针对以前的 UDA 方法存在隐私泄露风险以及容易在目标域中具有相似视觉外观的类之间产生混淆的问题,提出一种基于掩蔽图像一致性的无源领域自适应算法,称为掩蔽假设迁移(Masked Hypothesis Transfer, MHT)。**方法** MHT 采用假设迁移(Hypothesis Transfer, HT)思想,冻结源模型的分类器模块(假设),通过信息最大化和自监督伪标记来学习目标特征提取器,以隐式地对齐目标域与源域的表示;此外,还提出一个掩蔽图像一致性(Masked Image Consistency, MIC)模块,显式迫使模型学习目标域的空间上下文关系来增强假设迁移(HT),MIC 强迫掩蔽目标图像的预测和伪标签之间的一致性,就必须学会从被掩蔽区域的上下文中推断其预测。**结果** 该算法在闭集 UDA、部分集 UDA 和多源 UDA 3 种适应设置下,在 4 个公共基准上进行了广泛的测试,其中在 VisDA-*C* 上达到了 87.6%的准确率,比 SHOT 高 5.3%,在 Office-Home 上达到了 72.6%的准确率,比 SHOT 高 0.9%。**结论** 实验结果表明:掩蔽图像一致性目标可以作为额外的线索增强无源领域自适应,MHT 优于其他对比方法。

关键词:无监督域适应;无源域适应;掩蔽图像建模;迁移学习
中图分类号:TP391.41; TP18 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2025.0004.004

Research on Source-free Domain Adaptation Algorithm Based on Masked Image Consistency

LIU Erhu
School of Computer Science and Engineering, Anhui University of Science and Technology, Anhui Huainan 232001, China

Abstract: Objective Unsupervised domain adaptation (UDA) aims to adapt knowledge from labeled source domains to unlabeled target domains, achieving comparable performance in target tasks. Previous UDA methods face risks of privacy leakage and confusion among classes with similar visual appearances in the target domain. This paper proposed a source-free domain adaptation method based on masked image consistency, called Masked Hypothesis Transfer (MHT). **Methods** MHT adopted the idea of hypothesis transfer (HT) by freezing the classifier module (hypothesis) of the source model and learned the target feature extractor through information maximization and self-supervised pseudo-labeling to implicitly align representations between the target and source domains. Additionally, a masked image consistency (MIC) module was introduced to explicitly enforce the model to learn spatial contextual relationships in the target domain to enhance hypothesis transfer (HT). MIC forced consistency between the predictions on masked target images and the pseudo-labels, so the model must learn to infer predictions from the context of the masked region. **Results** The algorithm was extensively tested on four public benchmarks under closed-set UDA, partial-set UDA, and multi-source UDA settings. It achieved 87.6% accuracy on VisDA-*C*, outperforming SHOT by 5.3%, and 72.6% accuracy on Office-Home, surpassing SHOT by 0.9%. **Conclusion** Experimental results demonstrate that the target of masked image consistency serves as an additional clue to enhance source-free domain adaptation and MHT outperforms other comparative methods.

Keywords: unsupervised domain adaptation; source-free domain adaptation; masked image modeling; transfer learning

收稿日期:2023-12-23 **修回日期:**2024-02-20 **文章编号:**1672-058X(2025)04-0027-09
作者简介:刘二虎(1998—),男,安徽淮北人,硕士研究生,从事迁移学习、知识蒸馏等研究.
引用格式:刘二虎. 基于掩蔽图像一致性的无源领域自适应算法研究[J]. 重庆工商大学学报(自然科学版),2025,42(4):27-35.
LIU Erhu. Research on source-free domain adaptation algorithm based on masked image consistency[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(4): 27-35.

1 引 言

卷积神经网络(Convolutional Neural Network, CNN)[1]模型得益于大规模的标注数据集,在各种视觉识别任务中取得了显著的成功。然而,收集和标注过程可能非常耗时和繁琐,例如,对单幅图像进行语义分割的注释可能需要一个多小时。因此,利用现有的已标注数据集是有益的。然而,当在这样的源数据集上训练的网络应用于实际目标数据集时,会遭受显著的性能下降,因为神经网络对域偏移[2]很敏感。为了解决域偏移问题,近年来开发了各种无监督领域自适应(Unsupervised Domain Adaptation, UDA)算法,例如对抗性训练[3-7],但是这需要同时访问源数据和目标数据,造成了隐私泄露的安全问题。这限制了它们在许多现实场景中的应用,例如源数据是私有的,或者在计算能力有限的边缘设备中进行适应。

为了解决上述问题,无源域适应(Source-Free Domain Adaptation, SFDA)获得了越来越多的关注[8-12],它是一个有趣但具有挑战性的UDA设置,在适应中只有一个训练好的源模型可用。它与标准UDA的不同之处在于,将源模型而不是源数据提供给未标记的目标域。最先进的SFDA方法(如SHOT[9]、G-SFDA[10]、FAUST[13])大多依赖于由预训练源模型生成的伪标签引导自训练机制。这些方法经常利用特征空间中每类聚类结构的知识进行标签细化。虽然这些SFDA所使用的CNN已经具备对局部和上下文关系建模的能力,但是所使用的无监督目标损失不够强大,所以不能充分发挥在目标域上使用上下文依赖的潜力,以有效地学习这些信息。

对此,本文引入一种在域适应过程中显式引导模型学习目标域上下文关系的方法,用来提升模型的鲁棒识别能力。这是通过掩蔽图像一致性(MIC)目标实现的。以图像分类为例,随机掩蔽住目标图片的补丁,迫使网络仍然能够识别出这张图片的分类结果。这样,网络就必须利用上下文来推断被掩蔽区域的语义。由于没有目标域的真实标签,本文使用伪标签,由EMA(Exponential Moving Average)教师生成,使用原始的未掩蔽的目标图像作为输入。因此,教师可以同时利用上下文和局部线索来生成可靠的伪标签。在训练过程中,目标的不同部分被掩蔽,使网络学会利用不同的上下文线索,进一步提高了鲁棒性。由于本文旨在加强假设迁移,因此教师保持了源假设,只将特征提取器实现为目标模型的EMA。实验表明:所提MHT在各个任务上有显著和一致的性能提升,充分表明了MIC在无源域适应问题上的有效性。

2 相关工作

2.1 无监督域适应

无监督域自适应(UDA)旨在利用不同但相关的标记数据集中的知识来帮助学习未标记数据集的判别模型。现有UDA方法主要分为两大类,即矩匹配方法[14-16]和对抗学习方法[3-4,6-7,17]。

矩匹配方法通过跨域匹配已定义好的基于矩的分布差异来学习域不变特征表示。最初,大多数方法通过最小化最大平均差异(Maximum Mean Discrepancy, MMD)来显式地对齐跨域特征分布。后来,Long等[14-15]分别引入多核最大平均差异(Multi-Kernel Maximum Mean Discrepancy, MK-MMD)和联合最大平均差异(Joint Maximum Mean Discrepancy, JMMD)进行对齐。此外,Zhang等[16]提出了边际差异(Margin Disparity Discrepancy, MDD)来减小分布差异。

受生成对抗网络(Generative Adversarial Nets, GAN)的启发,对抗学习方法通过最小最大博弈来学习域不变特征。最初,Canin等[7]引入一个额外的判别器来区分特征提取器生成的特征,成功地实现了域级自适应。后来,Saito等[6]利用两个分类器之间的差异隐式实现对抗学习来提高可迁移性。此外,Chen等[4]将类别分类器重用为判别器,通过引入(Nuclear-norm Wasserstein Discrepancy, NWD)实现了明确的域对齐和类别区分。虽然这两类方法均取得了有希望的域适应效果,但当源域数据不可访问时,它们将不再适用。

2.2 无源域适应

正是由于标准无监督域适应存在隐私泄露的安全隐患,近年来越来越多的工作关注无源域适应方法。Kim等[8]使用基于低熵样本和点到集距离的伪标签;Liang等[9]利用基于质心的自训练标签细化技术;Yang等[10]采用类似的策略,通过鼓励局部相邻样本之间的一致预测来进一步改进伪标签;Roy等[12]量化了源模型预测中的不确定性,并利用它来指导目标适应;Lee等[13]同时考虑了任意不确定性和认知不确定性来增强模型对扰动图像的鲁棒性。但是它们均没有考虑目标域中局部外观相似的类。因此,在本文中,MHT显式引入了一种促使模型充分利用上下文信息来进行鲁棒识别的方法。

3 模型建立与原理分析

本文只使用预训练的源模型并且不访问源数据来解决无监督数据处理任务。特别地,考虑K-way分类。对于一个普通的无监督数据处理任务,给定n_s个来自源域D_s的标记样本$\{x_s^i, y_s^i\}_{i=1}^{n_s}$,其中$x_s^i \in X_s$,$y_s^i \in Y_s$,和$n_t$个来自源域$D_t$的未标记样本$\{x_t^i\}_{i=1}^{n_t}$,其中$x_t^i \in X_t$。领域适应(Domain Adaptation, DA)的目标是在目标域中预测标签$\{y_t^i\}_{i=1}^{n_t}$,其中$y_t^i \in Y_t$。假设源任务$X_s \to Y_s$和目标任务$X_t \to Y_t$相同,这里MHT的目标是在仅有$\{x_t^i\}_{i=1}^{n_t}$和$f_s: X_s \to Y_s$可用的条件下,学习一个目标函数$f_t: X_t \to Y_t$来推断$\{y_t^i\}_{i=1}^{n_t}$。

3.1　源模型生成

源模型是无源域适应的先验条件,为此,设计了一个源模型网络 $f_s=G_s\circ F_s$,与图 1 中的目标模型一样,由包含骨干网络和瓶颈层的特征提取器 F_s 和分类器 G_s 组成。其中骨干网络一般为 ResNet[1],瓶颈层由一个全连接层(Fully-Connected, FC)和一个批归一化层(Batch-Normalization, BN)组成,用于降维骨干网络输出的特征表示,以适应不同数据集。分类器由一个 FC 层和一个权值归一化层(Weight-Normalization, WN)组成。通过最小化带标签平滑(Label Smoothing, LS)的交叉熵损失来获得源模型 $f_s:X_s\to Y_s$。

$$\mathscr{L}_s(f_s;X_s,Y_s)=-\mathbb{E}_{(x_s,y_s)\in X_s\times Y_s}(\boldsymbol{q}^s)^{\mathrm{T}}\log f_s(x_s) \tag{1}$$

其中,$\boldsymbol{q}^s=(1-\sigma)\boldsymbol{l}_{y_s}+\sigma/K$ 是平滑标签向量,σ 是个平滑参数,经验上设置为 0.1,$\boldsymbol{l}_j$ 表示只有第 j 个值为 1 的 K 维独热编码。

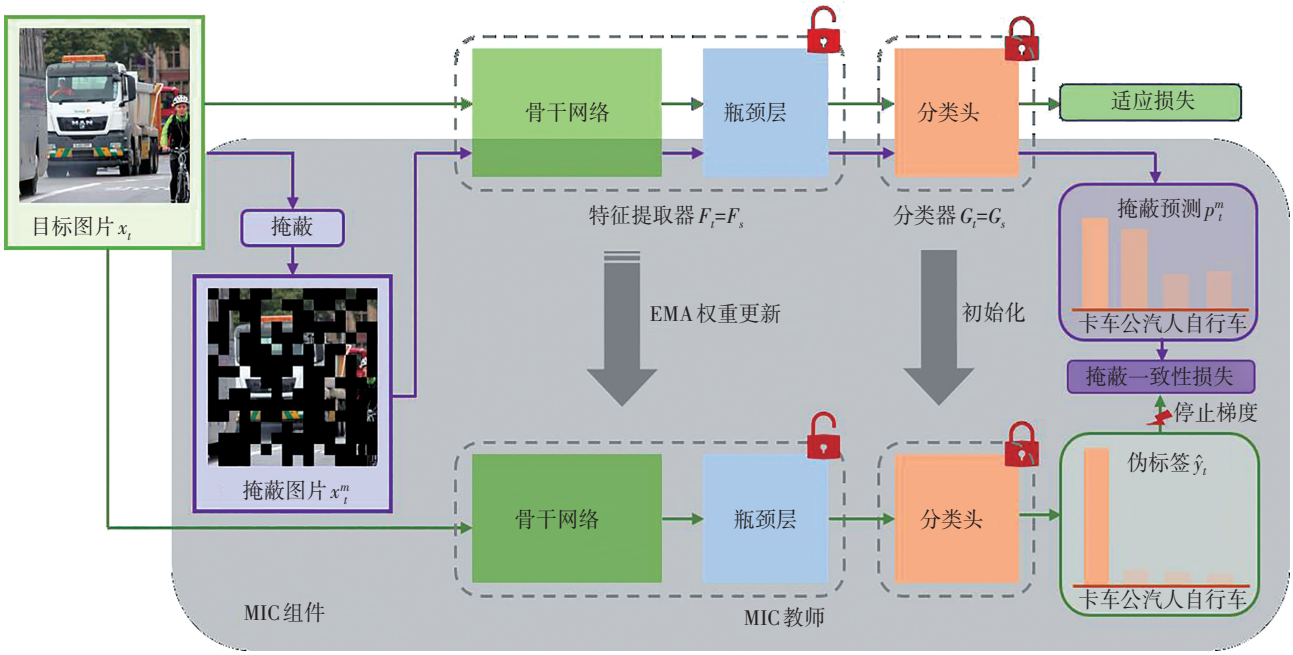


图 1　MHT 框架图

Fig. 1　Architecture of MHT

3.2　假设迁移

3.2.1　信息最大化

在无源域适应中,由于只能访问在 3.1 节中生成的源模型,不能像之前的矩匹配或域对抗学习等 UDA 方法一样在适应过程中同时使用源数据和目标数据,因此共享特征提取器是不可行的。为此,只能利用未标记目标数据 $\{x_t^i\}_{i=1}^{n_t}$ 学习特定于目标域的特征提取器。考虑到分类器(假设)编码了不可见的源数据分布信息,用固定分类器模块学习目标域特征提取器是一种隐式对齐的解决方案。具体地,如图 1 所示,目标模型在适应过程中使用与源模型相同的分类器 $G_t=G_s$,其核心思想:如果学习的目标模型能直接根据源假设进行准确分类,表明特征提取模块输出的目标特征与源特征分布很好地匹配了。

现在的问题是,什么样的目标输出才算得上准确分类。根据经验,目标输出要同时具备个体确定性和整体多样性。相应地,信息最大化(Information Maximization, IM)损失[9]满足了这一要求:

$$\mathscr{L}_{\mathrm{im}}(f_t;X_t)=H(\mathbb{E}_{x_t\in X_t}f_t(x_t))-\mathbb{E}_{x_t\in X_t}H(f_t(x_t)) \tag{2}$$

其中,H 是信息熵。第一项通过最大化平均预测的熵来促进类间多样性。第二项是最小熵,以鼓励模型对单一样本的精准预测。

3.2.2　自监督伪标记

尽管上一节中采用信息最大化的假设迁移是一种有效的对齐策略,但是信息最大化是一种无监督方式,很可能造成处于决策边界的样本划分到错误的集群中。为了缓解这种影响,遵循 SHOT[9] 提出的自监督伪标记策略,来显式监督目标特征提取器训练。首先,获得目标域中每个类的质心:

$$\boldsymbol{c}_k^{(0)}=\frac{\sum_{x_t\in X_t}\hat{f}_t(x_t)\,\hat{\boldsymbol{F}}_t(x_t)}{\sum_{x_t\in X_t}\hat{f}_t(x_t)} \tag{3}$$

其中,$\hat{f}_t=G_t\circ\hat{\boldsymbol{F}}_t$ 表示已经学习的目标模型,$\hat{f}_t(x_t)$ 表示目标模型预测,$\hat{\boldsymbol{F}}_t(x_t)$ 表示通过特征提取器 F 获得的目标特征向量。然后,通过最近邻算法获得伪标签:

$$\hat{y}_t=\underset{k}{\operatorname{argmin}}\,D(\hat{\boldsymbol{F}}_t(x_t),\boldsymbol{c}_k^{(0)}) \tag{4}$$

其中,$D(a,b)$ 表示 a 和 b 之间的余弦距离。最后,根据

新的伪标签计算目标质心:

$$\boldsymbol{c}_k^{(1)} = \frac{\sum_{x_t \in X_t} \mathbb{I}(\hat{y}_t = 1)\hat{\boldsymbol{F}}_t(x_t)}{\sum_{x_t \in X_t} \mathbb{I}(\hat{y}_t = 1)} \tag{5}$$

$$\hat{y}_t = \underset{k}{\operatorname{argmin}}\, D(\hat{\boldsymbol{F}}_t(x_t), \boldsymbol{c}_k^{(1)})$$

其中,$\mathbb{I}(\cdot)$表示指示函数。实际上,可以对质心和标签进行多轮更新。因此,自监督分类损失可以表述为

$$\mathscr{L}_{\text{cls}}(f_t; X_t, \hat{Y}_t) = -\mathbb{E}_{(x_t, \hat{y}_t) \in X_t \times \hat{Y}_t} \boldsymbol{q}_{\hat{y}_t} \log f_t(x_t) \tag{6}$$

其中,$\boldsymbol{q}_{\hat{y}_t}$表示 $\hat{y}_t$ 的 K 维独热编码向量。本文将上述两个损失统称为适应损失。

注意,所提方法 MHT 是基于 SHOT[9] 的,但与 SHOT++[18] 存在明显的不同。首先,MHT 没有采用 SHOT++中使用的旋转预测这一自监督代理任务,因为这需要额外的网络架构,显著增加了训练参数,而 MHT 并没有引进额外的训练参数。另外,SHOT++是两步训练策略,通过引入半监督学习作为假设迁移的增强手段,而 MHT 直接实现了端到端的训练,仅仅在 SHOT 这一基线上增加了 MIC 目标。

3.3 掩蔽假设迁移

为了识别目标,模型可以利用局部信息或者上下文信息,虽然卷积神经网络(Convolutional Neural Network, CNN)具有集成局部和上下文信息的能力,但是对于外观相似的目标,模型仍然容易产生分类混淆。对此,本文显式加强目标域中上下文关系的学习,引入了一个掩蔽图像一致性(MIC)模块,为具有相似局部外观类的鲁棒识别提供额外的线索,如图 1 所示。

MIC 通过随机掩蔽目标图像的补丁来保留局部信息。为此,从均匀分布中随机采样一个补丁掩膜 $\mathscr{M}$:

$$\mathscr{M}_{\substack{mb+1:(m+1)b \\ nb+1:(n+1)b}} = [v > r] \text{ with } v \sim U(0,1) \tag{7}$$

其中,$[\cdot]$表示艾弗森括号,b 是补丁大小,r 是掩蔽率,$m \in [0, \cdots, W/b-1]$,$n \in [0, \cdots, W/b-1]$是补丁索引,v 是从均匀分布的 $U(0,1)$ 中采样的随机值。通过掩膜和图像的逐元素乘法得到掩蔽目标图像,如式(8)所示:

$$x_t^{\mathscr{M}} = \mathscr{M} \odot x_t \tag{8}$$

由于只能利用有限的可见目标区域,使得预测很困难。为了训练网络在不访问整个图像的情况下,使用剩余的上下文线索仍然能重建正确的标签,引入 MIC 损失:

$$\mathscr{L}_{\text{mic}}(f_t; X_t^{\mathscr{M}}, \hat{Y}_t^{\mathscr{M}}) = -\mathbb{E}_{(x_t^{\mathscr{M}}, \hat{y}_t^{\mathscr{M}}) \in X_t^{\mathscr{M}} \times \hat{Y}_t^{\mathscr{M}}} \boldsymbol{q}_{\hat{y}_t^{\mathscr{M}}} \log f_t(x_t^{\mathscr{M}}) \tag{9}$$

其中,$\boldsymbol{q}_{\hat{y}_t^{\mathscr{M}}}$表示 $\hat{y}_t^{\mathscr{M}}$ 的最大 softmax 概率加权的 K 维编码向量,伪标签 $\hat{y}_t^{\mathscr{M}}$ 是对完整目标图像 x_t 的教师网络 g_t 的预测,而教师网络 g_t 通过目标模型(学生)f_t 的指数移动平均(Exponential Moving Average, EMA)逐步更新

$$g_{t+1} \leftarrow \alpha g_t + (1-\alpha) f_t \tag{10}$$

因此,这里的伪标签是稳定可靠的。注意,为了防止破坏源假设信息,本文固定教师网络的分类器模块。

综上所述,MHT 训练的总目标可以表述为

$$\mathscr{L}(F_t) = \mathscr{L}_{\text{im}}(f_t; X_t) + \beta \mathscr{L}_{\text{cls}}(f_t; X_t, \hat{Y}_t) + \gamma \mathscr{L}_{\text{mic}}(f_t; X_t^M, \hat{Y}_t^M) \tag{11}$$

β 和 γ 为超参数,遵循 SHOT[9],经验上设置为 0.3,而 γ 设置为 1。

4 仿真实验与结果分析

4.1 数据集

为了验证 MHT 的有效性,使用 4 种流行的公共视觉基准进行实验。Office-31 是一个标准的 UDA 基准,包含 3 个域(Amazon (A), DSLR (D) 和 Webcam (W)),每个域由办公环境下的 31 个对象类组成。Office-Caltech 涵盖了 Office-31(A、D、W)与 Caltech-256 (C)之间的 10 个共有类别,并形成一个新的基准。Office-Home 是一个具有显著域偏移的中型基准,它由 4 个不同的领域(艺术图片(Artistic images, Ar),剪贴画(Clip Art, Cl),产品图片(Product images, Pr)和现实世界图片(Real-World images, Re)组成。VisDA-C 是一个具有挑战性的大规模基准,主要关注 12 类 synthesis-to-real 目标识别任务。

4.2 实现细节

通过 PyTorch 实现了提出的 MHT,并报告了 3 次运行之间的平均准确率。遵循 SHOT[9],对于 Office-31 和 Office-Home,采用 ResNet-50[1] 作为主干;对于 Office-Caltech 和 VisDA-C,采用 ResNet-101[1] 作为主干。利用 ImageNet 预训练权重初始化,采用小批量 SGD 对模型进行优化,除了 VisDA-C 的主干网络学习率为 1e-4,其他数据集为 1e-3,瓶颈层和分类头的学习率是其 10 倍,动量为 0.9,权重衰减为 1e-3。训练时的批大小设置为 64。Office-31、Office- home、VisDA-C 和 Office-Caltech 源模型训练的最大 epoch 数分别设置为 100、50、10 和 100。而目标模型的训练 epoch 统一设置为 15。表 1、表 2 和表 3 中的“SF”表示 Source-Free,即适应中不使用源数据。Source Model Only 表示使用预训练源模型预测的目标精度。SHOT * 表示根据公开代码的复现结果。

4.3 实验结果及分析

在 4 个广泛使用的基准数据集上展示了 MHT 的适应性能,涵盖 closed-set、partial-set 和 multi-source 3 种设置,结果如表 1、表 2、表 3、表 4、表 5 所示,每个任务上 SFDA 最好的结果以粗体显示。为了公平性,本文

仅对比设置相同的基线方法。

Office-31、Office-Home 和 VisDA-C(closed-set)。在表 1 中 Office-31 数据集上,MHT 模型获得了 89.2% 的精度,在 4 个任务上获得了最先进的性能。相比 SOTA (State of the Art)方法仍然获得了具有竞争力的结果。在表 2 的 Office-Home 上,MHT 在一半的任务上取得了最先进的表现,将平均准确率从 SHOT 的 71.7%提高到 72.6%。尤其地,如表 3 所示,在最具挑战的 VisDA-C 上,MHT 几乎达到了每类上最佳的分类精度,甚至大幅度超越之前接触的源数据标准 UDA 算法。这些结果均表明掩蔽图像一致性的有效性。

Office-Home(partial-set)和 Office-Caltech (multi-source),如表 4 所示,MHT 在 Office-Home 上验证了其部分集域适应的效果,相比于 SHOT,实现了整体的提升,从 79.0%→79.8%。同样地,在表 5 中,MHT 在 Office-Caltech 上实现了类似的增益,其中 $\mathscr{R}$ 表示除源域外的剩余 3 个域。这表明 MHT 在处理非对称 UDA 任务和多源条件下的适应能力。

表 1　在 Office-31 数据集上的 closed-set UDA 分类精度

Table1　Classification accuracies on Office-31 dataset for closed-set UDA (Resnet-50)　/%

方法(源域→目标域)	SF	A→D	A→W	D→A	D→W	W→A	W→D	Avg.
ResNet-50[1]	-	68.9	68.4	62.5	96.7	60.7	99.3	76.1
DANN[7]	×	79.7	82.0	68.2	96.9	67.4	99.1	82.2
DAN[14]	×	78.6	80.5	63.6	97.1	62.8	99.6	80.4
MCD[6]	×	92.2	88.6	69.5	98.5	69.7	100.0	86.5
BNM[19]	×	90.3	91.5	70.9	98.5	71.6	100.0	87.1
SAFN+ENT[20]	×	90.7	90.1	73.0	98.6	70.2	99.8	87.1
SWD[21]	×	94.7	90.4	70.3	98.7	70.5	100.0	87.4
CDAN+BSP[22]	×	93.0	93.3	73.6	98.2	72.6	100.0	88.5
MetaAlign[23]	×	94.5	93.0	75.0	98.6	73.6	100.0	89.2
Source model only	-	80.1	76.1	58.7	95.5	61.5	99.2	78.5
SFDA[8]	✓	91.1	98.2	99.5	92.2	71.0	71.2	87.2
SHOT*[9]	✓	94.4	90.5	73.5	98.1	74.3	99.9	88.5
U-SFAN+[12]	✓	94.2	**92.8**	**74.6**	98.0	74.4	99.0	88.8
MHT(Ours)	✓	**95.8**	91.6	73.9	**99.0**	**74.8**	**100.0**	**89.2**

表 2　在 Office-Home 数据集上的 closed-set UDA 分类精度

Table 2　Classification accuracies on Office-Home dataset for closed-set UDA (Resnet-50)　/%

方法(源域:目标域)	SF	Ar:Cl	Ar:Pr	Ar:Re	Cl:Ar	Cl:Pr	Cl:Re	Pr:Ar	Pr:Cl	Pr:Re	Re:Ar	Re:Cl	Re:Pr	Avg.
ResNet-50[1]	-	34.9	50.0	58.0	37.4	41.9	46.2	38.5	31.2	60.4	53.9	41.2	59.9	46.1
DAN[14]	×	43.6	57.0	67.9	45.8	56.5	60.4	44	43.6	67.7	63.1	51.5	74.3	56.3
DANN[7]	×	45.6	59.3	70.1	47.0	58.5	60.9	46.1	43.7	68.5	63.2	51.8	76.8	57.6
MCD[6]	×	48.9	68.3	74.6	61.3	67.6	68.8	57.0	47.1	75.1	69.1	52.2	79.6	64.1
CDAN+BSP[22]	×	52.0	68.6	76.1	58.0	70.3	70.2	58.6	50.2	77.6	72.2	59.3	81.9	66.3
SAFN[20]	×	52.0	71.7	76.3	64.2	69.9	71.9	63.7	51.4	77.1	70.9	57.1	81.5	67.3
BNM[19]	×	52.3	73.9	80.0	63.3	72.9	74.9	61.7	49.5	79.7	70.5	53.6	82.2	67.9
SCDA[17]	×	57.5	76.9	80.3	65.7	74.9	74.5	65.5	53.6	79.8	74.5	59.6	83.7	70.5
MetaAlign[23]	×	59.3	76.0	80.2	65.7	74.7	75.1	65.7	56.5	81.6	74.1	61.1	85.2	71.3
DALN[4]	×	57.8	79.9	82.0	66.3	76.2	77.2	66.7	55.5	81.3	73.5	60.4	85.3	71.8
Source model only	-	44.1	66.4	74.3	51.6	62.2	65.0	52.4	41.2	73.1	65.0	45.8	76.9	59.8
SFDA[8]	✓	48.4	73.4	76.9	64.3	69.8	71.7	62.7	45.3	76.6	69.8	50.5	79.0	65.7
G-SFDA[10]	✓	57.9	78.6	81.0	66.7	77.2	77.2	65.6	56.0	**82.2**	72.0	57.8	83.4	71.3
FAUST[13]	✓	**59.4**	78.5	79.4	62.7	77.6	75.0	64.5	**61.0**	78.3	72.7	**64.8**	**85.9**	71.6
SHOT*[9]	✓	56.8	78.3	81.7	67.8	77.2	78.1	**68.2**	54.7	82.1	72.9	58.4	84.2	71.7
U-SFAN+[12]	✓	57.8	77.8	81.6	67.9	77.3	79.2	67.2	54.7	81.2	73.3	60.3	83.9	71.9
MHT(Ours)	✓	57.2	**80.3**	**82.7**	**69.4**	**79.0**	**79.3**	67.8	56.1	81.7	73.7	59.2	85.0	**72.6**

表 3 在 VisDA-C 数据集上的 closed-set UDA 分类精度

Table 3 Classification accuracies on VisDA-C dataset for closed-set UDA (Resnet-101) /%

方法(Synthesis→ Real)	SF	plane	bcycl	bus	car	horse	knife	mcycle	person	plant	sktbrd	train	truck	Per-class
ResNet-101[1]	-	55.1	53.3	61.9	59.1	80.6	17.9	79.7	31.2	81.0	26.5	73.5	8.5	52.4
DANN[7]	×	81.9	77.7	82.8	44.3	81.2	29.5	65.1	28.6	51.9	54.6	82.8	7.8	57.4
DAN[14]	×	87.1	63.0	76.5	42	90.3	42.9	85.9	53.1	49.7	36.3	85.8	20.7	61.1
CDAN+BSP[22]	×	92.4	61.0	81.0	57.5	89.0	80.6	90.1	77.0	84.2	77.9	82.1	38.4	75.9
SAFN[20]	×	93.6	61.3	84.1	70.6	94.1	79.0	91.8	79.6	89.9	55.6	89.0	24.4	76.1
SWD[21]	×	90.8	82.5	81.7	70.5	91.7	69.5	86.3	77.5	87.4	63.6	85.6	29.2	76.4
DTA[24]	×	93.7	82.2	85.6	83.8	93.0	81.0	90.7	82.1	95.1	78.1	86.4	32.1	81.5
BCDM[25]	×	95.1	87.6	81.2	73.2	92.7	95.4	86.9	82.5	95.1	84.8	88.1	39.5	83.4
SDAT[3]	×	95.8	85.5	76.9	69.0	93.5	97.4	88.5	78.2	93.1	91.6	86.3	55.3	84.3
Source model only	-	58.6	16.4	52.8	65.1	71.7	6.3	81.9	26.7	79.6	33.3	84.4	7.1	48.7
SFDA[8]	✓	86.9	81.7	84.6	63.9	93.1	91.4	86.6	71.9	84.5	58.2	74.5	42.7	76.7
SHOT*[9]	✓	94.8	88.2	79.4	56.3	93.0	93.7	79.9	80.7	88.1	89.0	86.0	58.3	82.3
A^2NET[11]	✓	94.0	87.8	85.6	66.8	93.7	95.1	85.8	81.2	91.6	88.2	86.5	56.0	84.3
FAUST[13]	✓	96.7	77.6	87.6	73.3	95.5	95.4	**92.9**	83.6	95.3	89.5	87.7	46.9	85.2
G-SFDA[10]	✓	96.1	88.3	85.5	74.1	**97.1**	95.4	89.5	79.4	**95.4**	92.9	89.1	42.6	85.4
MHT(Ours)	✓	**96.9**	**91.7**	**88.0**	**75.4**	96.6	**95.8**	89.3	**84.2**	**92.2**	**93.4**	**89.9**	**57.4**	**87.6**

表 4 在 Office-Home 数据集上的 partial-set UDA 分类精度

Table 4 Classification accuracies on Office-Home dataset for partial-set UDA (Resnet-50) /%

Partial-set DA (源域:目标域)	Ar:Cl	Ar:Pr	Ar:Re	Cl:Ar	Cl:Pr	Cl:Re	Pr:Ar	Pr:Cl	Pr:Re	Re:Ar	Re:Cl	Re:Pr	Avg.
ResNet-50[1]	46.3	67.5	75.9	59.1	59.9	62.7	58.2	41.8	74.9	67.4	48.2	74.2	61.3
IWAN[26]	53.9	54.5	78.1	61.3	48.0	63.3	54.2	52.0	81.3	76.5	56.8	82.9	63.6
SAN[5]	44.4	68.7	74.6	67.5	65.0	77.8	59.8	44.7	80.1	72.2	50.2	78.7	65.3
ETN[27]	59.2	77.0	79.5	62.9	65.7	75	68.3	55.4	84.4	75.7	57.7	84.5	70.5
SAFN[20]	58.9	76.3	81.4	70.4	73.0	77.8	72.4	55.3	80.4	75.8	60.4	79.9	71.8
Source model only	43.3	68.6	79.8	53.2	58.9	64.0	59.7	41.8	76.3	70.3	47.7	77.7	61.8
SHOT*[9]	59.8	**86.6**	**92.9**	73.7	75.5	**85.4**	80.1	64.4	**90.0**	84.4	63.5	91.8	79.0
MHT(Ours)	**60.5**	84.7	91.6	**76.0**	**78.2**	85.2	**81.5**	**67.4**	**90.0**	**84.6**	**65.8**	**92.0**	**79.8**

表 5 在 Office-Caltech 数据集上的 multi-source UDA 分类精度

Table 5 Classification accuracies on Office-Caltech dataset for multi-source UDA (Resnet-101) /%

Multi-source (ℜ→)	ℜ→A	ℜ→C	ℜ→D	ℜ→*W*	Avg.
ResNet-101[1]	88.7	85.4	98.2	99.1	92.9
DAN[14]	91.6	89.2	99.1	99.5	94.8
DCTN[28]	92.7	90.2	99.0	99.4	95.3
MCD[6]	92.1	91.5	99.1	99.5	95.6
M^3SDA-β[2]	94.5	92.2	99.2	99.5	96.4
Source model only	95.6	92.8	99.4	98.3	96.5
SHOT*[9]	95.9	**96.4**	98.7	99.3	97.6
MHT(Ours)	**96.4**	96.0	**99.4**	**99.7**	**97.9**

4.4 消融实验

为了验证所提 3 个优化目标的独立有效性,在 3 个数据集上进行了详细的消融实验,如表 6 所示。可以看出:自监督伪标记和信息最大化都是有益的,两者的结合相互促进,进一步增强了目标域的性能(即 SHOT)。对于本文提出的 MIC,在 3 个数据集上均显著提升了 SHOT,达到了 83.1%的平均精度。表明了 MIC 的有效性和泛化性。

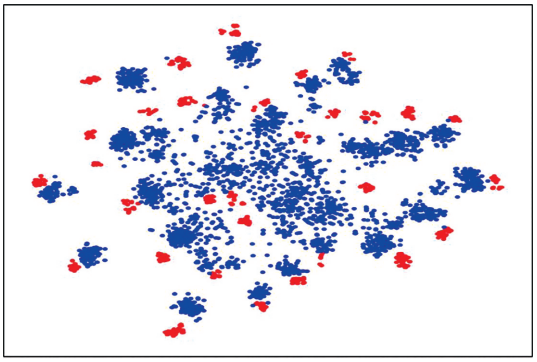
表 6 在 3 个 closed-set UDA 数据集上平均精度的消融研究

Table 6 Ablation study on average accuracy on three closed-set UDA datasets /%

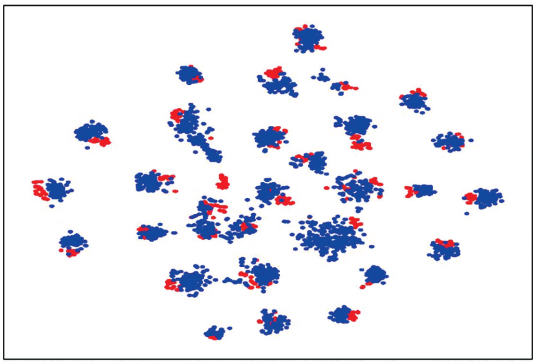
组件/数据集	Office-31	Office-Home	VisDA-*C*	Avg.
Source model only	78.5	59.8	48.7	62.3
Self-supervised PL	87.6	68.9	80.7	79.1
$\mathscr{L}_{im}$	87.3	70.5	80.4	79.4
$\mathscr{L}_{im}$+Self-supervised PL	88.6	71.8	82.9	81.1
$\mathscr{L}_{im}$+Self-supervised PL+MIC(Ours)	89.2	72.6	87.6	83.1

4.5　特征可视化

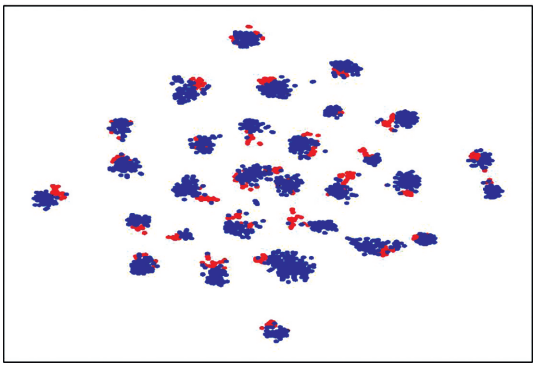
图 2 展示了在 Office-31 数据集的 W→A 迁移任务上使用 t-SNE 绘制的特征提取器的输出特征。红色代表源域,蓝色代表目标域。Source model only 混成一团,在目标域上仅有基本的判别能力,SHOT 虽然取得了显著的改进,但仍然存在类内松散、类间混淆的问题,而 MHT 形成了更紧凑的对齐和更清晰的决策边界。



(a) Source model only



(b) SHOT

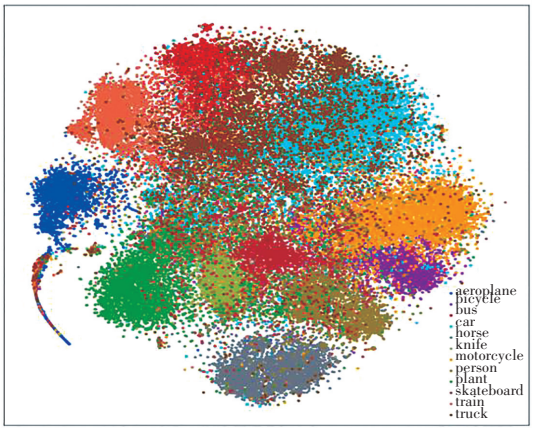


(c) MHT(Ours)

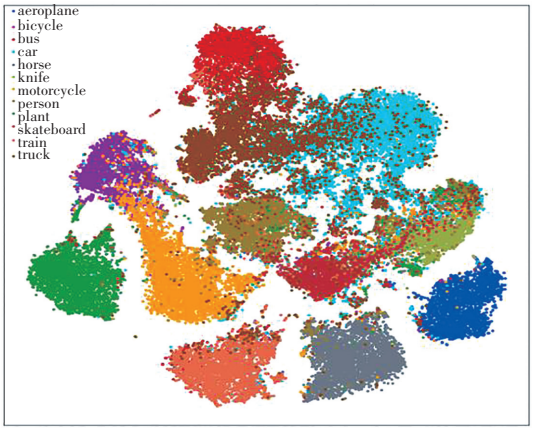
图 2　不同算法适应后的 t-SNE 可视化图

Fig. 2　t-SNE visualization graphs after adaptation of different algorithms

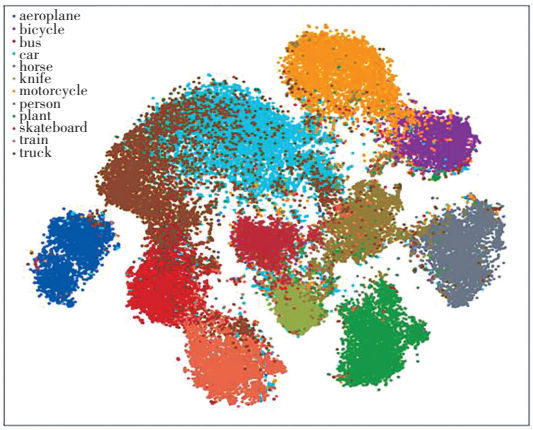
为了清晰地观察目标模型的可判别性。图 3 可视化了在 VisDA-C 数据集上的按类 t-SNE 图,不同颜色表示不同的类。因为 VisDA-C 类别较少,而每个类中样本多,所以比较容易看出适应效果的差异性。从图 3 可以看出:MHT 和 SHOT 均实现了比 Source model only 好得多的区分能力。仔细观察,所提 MHT 仍然进一步改进了 SHOT,例如,蓝绿色集群(汽车)和灰色集群(卡车)变得内部更加紧凑且彼此间更加分离。这表明 MHT 对区分相似类的优势和潜力。相应地,提供了它们之间的定量分析结果,通过图 4 相似性矩阵显示,也验证了 MHT 学习不同类的关键特征,显著缓解了类混淆问题。



(a) Source model only



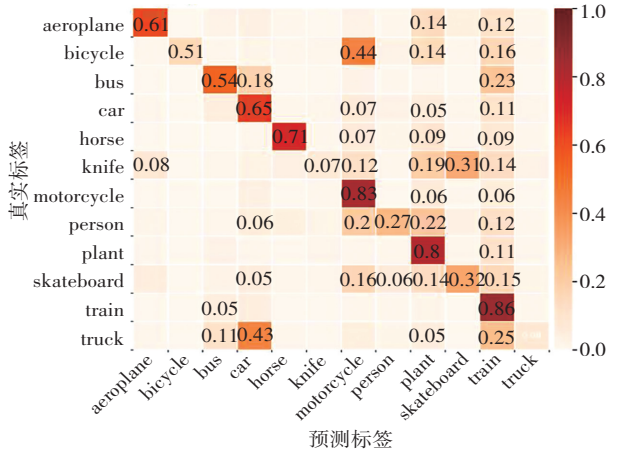
(b) SHOT



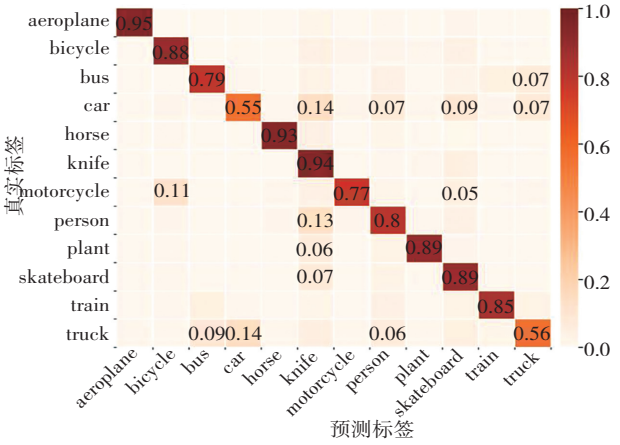
(c) MHT(Ours)

图 3　不同算法在 VisDA-C 上学习的按类 t-SNE 图

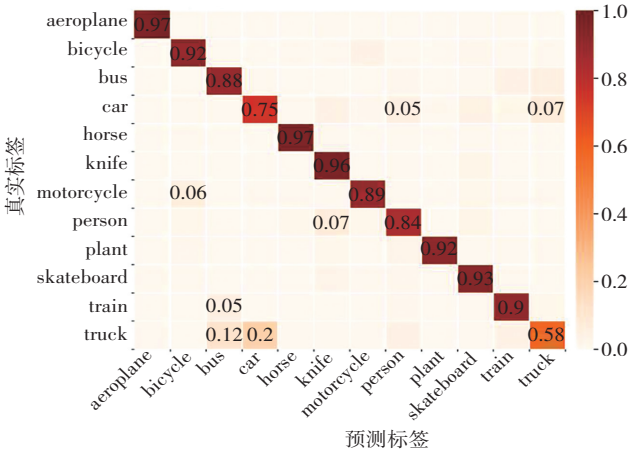
Fig. 3　Class-wise t-SNE graphs learned by different algorithms on VisDA-C



(a) Source model only



(b) SHOT



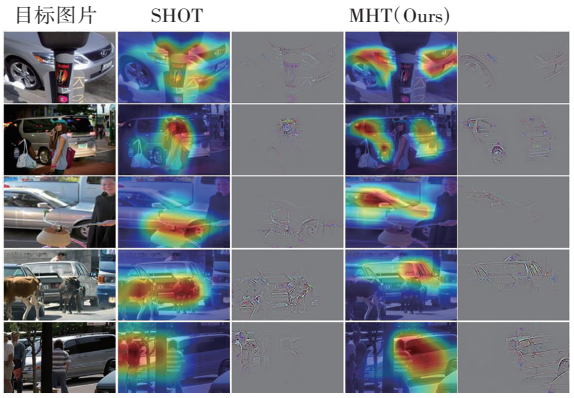
(c) MHT(Ours)

图 4 不同算法在 VisDA-*C* 上学习的相似性矩阵

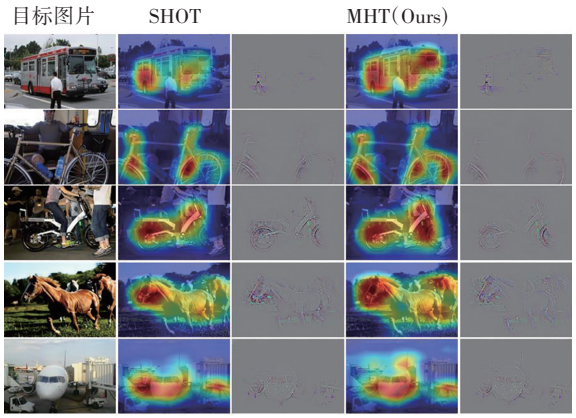
Fig. 4 Similarity matrices learned by different algorithms on VisDA-*C*

4.6 视觉解释性

为了将所提 MHT 与基线 SHOT 进行直观比较,本文通过 GradCAM 和 GuidedBackprop 进一步提供了视觉解释。如图 5 所示,分别从异类预测和同类预测两个角度比较,图 5(a)为异类预测,图 5(b)为同类预测。为了对比,选择具有代表性的图片。对于异类预测,SHOT 以高置信度预测了错误的前景对象,如在以下 4 个均为汽车的遮挡图中预测为滑板、人、马。相反,得益于 MIC 的上下文探索,成功识别出了汽车目标。而对于同类预测,MHT 仍然表现出比 SHOT 更全面的语义理解能力,更加清楚地显示出目标的轮廓形状、目标个数等。这意味着 MHT 的上下文理解有利于复杂环境中的鲁棒识别。



(a) 异类预测



(b) 同类预测

图 5 SHOT 和 MHT 的可解释性对比

Fig. 5 Comparison of explainability between SHOT and MHT

5 结 论

提出一种新的不需要访问源图像的 UDA 方法——MHT。该方法利用假设迁移的思想训练特征提取器。具体地,采用信息最大化和自监督伪标记两种无监督学习目标。为了增强特征提取器的鲁棒识别能力,提出一种显式引导模型探索上下文关系的掩蔽图像一致性目标。通过广泛的实证结果表明,MHT 在 4 个 UDA 基准上取得了最先进或具有竞争力的结果,并且利用各种可视化方法定性和定量地解释了所提算法的优越性。本文并没有在目标检测、语义分割、行人重识别领域上验证所提 MHT 的有效性,这将在未来的工作中进行彻底研究,因为无源域适应主要面向实际场景应用,所以 MHT 具有广泛的应用前景。另外,未来计划在目标模型学习的过程中加入监督性更强的代理任务,来帮助目标域的精准识别。

参考文献(References):

[1] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 770-778.

[2] PENG X, BAI Q, XIA X, et al. Moment matching for multi-source domain adaptation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2019: 1406-1415.

[3] RANGWANI H, AITHAL S K, MISHRA M, et al. A closer look at smoothness in domain adversarial training[C]//International Conference on Machine Learning. IEEE, 2022: 18378-18399.

[4] CHEN L, CHEN H, WEI Z, et al. Reusing the task-specific classifier as a discriminator: Discriminator-free adversarial domain adaptation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2022: 7181-7190.

[5] CAO Z, LONG M, WANG J, et al. Partial transfer learning with selective adversarial networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 2724-2732.

[6] SAITO K, WATANABE K, USHIKU Y, et al. Maximum classifier discrepancy for unsupervised domain adaptation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 3723-3732.

[7] GANIN Y, USTINOVA E, AIAKAN H, et al. Domain-adversarial training of neural networks[J]. Journal of Machine Learning Research, 2016, 17(59): 1-35.

[8] KIM Y, CHO D, HAN K, et al. Domain adaptation without source data[J]. IEEE Transactions on Artificial Intelligence, 2021, 2(6): 508-518.

[9] LIANG J, HU D, FENG J. Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation[C]//International Conference on Machine Learning, 2020: 6028-6039.

[10] YANG S, WANG Y, VAN DE WEIJER J, et al. Generalized source-free domain adaptation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2021: 8978-8987.

[11] XIA H, ZHAO H, DING Z. Adaptive adversarial network for source-free domain adaptation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2021: 9010-9019.

[12] ROY S, TRAPP M, PILZER A, et al. Uncertainty-guided source-free domain adaptation[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 537-555.

[13] LEE J, LEE G. Feature alignment by uncertainty and self-training for source-free unsupervised domain adaptation[J]. Neural Networks, 2023, 161: 682-692.

[14] LONG M, CAO Y, WANG J, et al. Learning transferable features with deep adaptation networks[C]//International Conference on MachineLearning. IEEE, 2015: 97-105.

[15] LONG M, ZHU H, WANG J, et al. Deep transfer learning with joint adaptation networks[C]//International Conference on Machine Learning. IEEE, 2017: 2208-2217.

[16] ZHANG Y, LIU T, LONG M, et al. Bridging theory and algorithm for domain adaptation[C]//International Conference on Machine Learning. IEEE, 2019: 7404-7413.

[17] LI S, XIE M, LV F, et al. Semantic concentration for domain adaptation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2021: 9102-9111.

[18] LIANG J, HU D, WANG Y, et al. Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(11): 8602-8617.

[19] CUI S, WANG S, ZHUO J, et al. Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2020: 3941-3950.

[20] XU R, LI G, YANG J, et al. Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2019: 1426-1435.

[21] LEE C Y, BATRA T, BAIG M H, et al. Sliced Wasserstein discrepancy for unsupervised domain adaptation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 10285-10295.

[22] CHEN X, WANG S, LONG M, et al. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation[C]//International Conference on Machine Learning. IEEE, 2019: 1081-1090.

[23] WEI G, LAN C, ZENG W, et al. MetaAlign: Coordinating domain alignment and classification for unsupervised domain adaptation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2021: 16643-16653.

[24] LEE S, KIM D, KIM N, et al. Drop to adapt: Learning discriminative features for unsupervised domain adaptation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2019: 91-100.

[25] LI S, LV F, XIE B, et al. Bi-classifier determinacy maximization for unsupervised domain adaptation[J]. Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(10): 8455-8464.

[26] ZHANG J, DING Z, Li W, et al. Importance weighted adversarial nets for partial domain adaptation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 8156-8164.

[27] CAO Z, YOU K, LONG M, et al. Learning to transfer examples for partial domain adaptation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 2985-2994.

[28] XU R, CHEN Z, ZUO W, et al. Deep cocktail network: Multi-source unsupervised domain adaptation with category shift[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 3964-3973.

责任编辑:李翠薇