文章编号: 1001-4543(2025)03-0298-09 DOI: 10.19570/j.cnki.jsspu.2025.03.008

基于不对称对抗和自监督增强的无监督域自适应算法

尤雨婷[a], 谭金菁[b], 高瑞玲[a], 毛文逸[a], 吕金龙[c], 侯英琦[a], 谭文安[a]

(上海第二工业大学 a. 计算机与信息工程学院; b. 经济与管理学院;
c. 智能制造与控制工程学院, 上海 201209)

摘 要: 针对现有无监督域自适应 (unsupervised domain adaptation, UDA) 算法在处理复杂领域分布差异时存在的特征对齐细粒度不足、目标域潜在特征挖掘能力有限等问题, 提出了一种基于不对称对抗和自监督增强的无监督域自适应 (UDA based on asymmetric adversarial learning and self-supervised enhancement, UDA-2A2S) 算法。该算法通过引入全局与局部域判别器协同作用机制用于捕捉细粒度的领域特征差异, 并结合渐进对齐策略动态调整特征对齐强度, 以更好地适用于复杂任务场景。此外, 设计了一种基于旋转预测的自监督学习机制, 从目标域无标注数据中挖掘潜在的图像实例结构信息, 进一步提升模型的特征学习能力和跨域迁移能力。实验结果表明, 在 Office-Home 和 DomainNet 基准数据集上, UDA-2A2S 方法平均准确率分别达到 72.4% 和 46.2%, 显著优于现有主流 UDA 方法, 验证了所提算法的优越性, 为跨域迁移学习提供了新的解决方案。

关键词: 无监督领域自适应; 不对称对抗机制; 自监督增强; 迁移学习
中图分类号: TP399 **文献标志码**: A

An Unsupervised Domain Adaptive Algorithm with Asymmetric Adversarial and Self-Supervised Enhancement

YOU Yuting[a], TAN Jinjing[a], GAO Ruiling[a], MAO Wenyi[a],
LÜ Jinlong[c], HOU Yingqi[a], TAN Wenan[a]

(a. School of Computer and Information Engineering; b. School of Economics and Management; c. School of Intelligent Manufacturing and Control Engineering, Shanghai Polytechnic University, Shanghai 201209, China)

Abstract: To address the limitations of existing unsupervised domain adaptation (UDA) algorithm in handling complex domain distribution differences, including the insufficient fine-grained feature alignment and limited ability to extract potential features in the target domain, the study proposed an UDA based on asymmetric adversarial learning and self-supervised enhancement (UDA-2A2S) algorithm. The proposed UDA-2A2S is used to capture fine-grained domain feature differences by introducing a synergistic mechanism of global and local domain discriminators, and dynamically adjusts the feature alignment strength in combination with a progressive alignment strategy to better apply to complex task scenarios. In addition, a self-supervised learning mechanism based on rotation prediction was designed to mine potential image instance structure information from unlabeled data in the target domain to further enhance the feature learning capability and cross-domain migration ability of the model. The experimental results show that the UDA-2A2S method has an average accuracy of 72.4% and 46.2% on the Office-Home and DomainNet benchmark datasets, respectively, which is significantly better than the existing mainstream UDA methods, demonstrating the superiority of the algorithm and providing a new

收稿日期: 2025-02-21
通信作者: 谭文安 (1965–), 男, 湖北公安人, 教授, 博士, 主要研究方向为协同计算与系统进化、协同学习与知识管理等。
E-mail: watan@sspu.edu.cn
基金项目: 国家自然科学基金项目 (61672022, U1904186), 上海市研究生教育学会研究课题 (ShsgeG202207) 资助

solution for cross domain transfer learning.

Keywords: unsupervised domain adaptation; asymmetric adversarial mechanisms; self-supervised enhancements; transfer learning

0　引言

深度神经网络技术在图像分类 [1]、语义分割 [2]、目标检测 [3] 等传统计算机视觉分析中取得了广泛应用和显著成就, 并逐渐扩展至机器学习、自然语言处理、具身智能等前沿技术。例如, 通过 ImageNet 数据集 [4] 的构建与应用, 深度学习模型在计算机视觉处理等图像分类中取得了显著突破, 极大推动了该领域的快速发展。然而, 深度模型的训练依赖大量标注数据, 而数据标注耗时且成本高昂。针对这一问题, 无监督域自适应 [5] (unsupervised domain adaptation, UDA) 提供了一种有效解决途径, 其核心思想是利用源域中已的标注数据及目标域中未标注的数据, 通过特征迁移 [6] 和分布对齐策略 [7], 训练一个能够在目标域上表现优异的模型。这种跨域学习方法为解决领域间分布偏移问题提供了重要思路。

传统领域适配方法, 如领域对抗神经网络 (domain adversarial neural network, DANN)[8], 通过对抗训练实现源域与目标域特征分布的对齐, 在领域自适应研究中取得了广泛应用。然而, 这些方法通常依赖单一判别器, 难以捕捉全局与局部特征的细粒度差异; 在目标域无标注或分布差异显著时, 其特征学习能力也明显不足, 导致性能下降。

为了解决上述问题, 提出了一种基于不对称对抗和自监督增强 [9] 的 UDA 算法 (UDA based on asymmetric adversarial learning and self-supervised enhancement, UDA-2A2S)。基于传统 DANN 框架引入了不对称对抗机制与自监督增强模块, 改进特征对齐方式和增强目标域特征学习能力。

本文主要贡献包括:

(1) 引入了一种不对称对抗机制: 通过全局域和局部域判别器的协同作用, 捕捉细粒度的领域特征差异, 并结合渐进对齐策略动态调整对齐力度, 以适应复杂任务场景。

(2) 引入了一种自监督增强模块: 通过旋转预测任务, 从目标域无标注数据中挖掘潜在结构信息, 增强特征表征的鲁棒性和判别力。

(3) 在两个数据集 (Office-Home 和 DomainNet) 上展示了本文提出算法的有效性, 并通过消融实验进一步评估了不同方法的贡献。

1　相关工作

1.1　基于领域对齐的 UDA 方法

领域对齐是经典的 UDA 解决方案, 通过减少领域间的分布差异来实现知识迁移。例如, DANN, Ben-David 等 [10] 提出使用对抗训练, 通过引入领域分类器与梯度反转层, 实现特征对齐, 学习领域不变表示。基于条件性对抗网络 (conditional domain adversarial networks, CDANs) 方法, Long 等 [11] 结合分类器的预测与领域对齐, 进一步提高了模型特征分布的一致性。然而, 这些方法主要聚焦于全局领域对齐, 忽略了局部特征在领域适应中的重要作用。

1.2　基于自监督学习的无监督增强

近年来, 自监督学习被广泛用于提升 UDA 的性能。Caron 等 [12] 采用深度聚类 (Deep Cluster) 提升特征表征的稳定性; Lee 等 [13] 则利用虚拟对抗扰动增强目标域特征的鲁棒性。以上方法通过设计自监督学习, 从未标注数据中挖掘潜在的图像实例结构化信息, 在目标域标注稀缺的情况下显著提升了模型泛化性。然而, 这些方法将自监督与迁移过程分离, 缺乏协同机制, 且常侧重于单一层次或模态的特征学习。

1.3　基于不对称对抗机制的优化

不对称对抗机制通过设计多层对抗策略, 提升了领域对齐的能力。例如, 基于反向传播诱导的特征加权 (backprop-induced feature weighting, BIFW) 方法, Westfechtel 等 [14] 通过回传分类误差, 强调重要特征并减弱无关特征, 从而实现更精确的领域对齐。基于边际损失的对抗学习方法 (margin-based adversarial learning for domain generalization, MADG), Dayal 等 [15] 通过最小化一个基于边际损失的差异度量来学习域不变特征。尽管这些方法针对对齐策略有了显著进步, 但其多聚焦对抗机制本身, 仍缺乏

对目标域深层特征结构的深入建模, 忽略与自监督等机制的协同。

为此, 本文提出了 UDA-2A2S 方法, 有效解决了现有方法在复杂领域分布差异下的对齐不足问题, 显著提升了目标域的特征表征能力与分类性能。

2 基于不对称对抗与自监督增强的无监督域自适应框架

2.1 UDA-2A2S 整体模型架构

UDA-2A2S 建立在 DANN 框架基础之上, 设有带标签的源域数据 $(x_s, x_y) \in D_S$ 和无标签的目标域数据 $(x_t) \in D_T$。该方法通过 1 个特征提取器网络 G_f、1 个类别预测器网络 G_y 和 1 个域判别器 G_d 构建联合优化过程。其中, G_f 接收输入样本并输出特征表示 $f = G_f(x)$, G_y 基于该特征执行分类预测, G_d 则负责判别特征来自源域还是目标域。传统 DANN 的优化目标是在特征空间中实现 2 个方面的平衡 [8]: 一方面通过最小化源域的分类损失以提升源域上分类性能; 另一方面通过对抗性的域分类器训练, 使 G_f 提取的特征在域层面上难以区分源域或目标域, 从而实现跨域特征分布对齐。

以最小化模型参数为例, DANN 的基本优化目标可表示为:

$$\min_{G_f,G_y}\max_{G_d} L(G_f,G_y,G_d) = E_{(x,y_s)\in D_S}[L_y(G_y(G_f(x_s)),y_s)] - \lambda E_{x\in\{D_S\cup D_T\}}[L_d(G_d(G_f(x)))] \tag{1}$$

式中, $\min\limits_{G_f,G_y}\max\limits_{G_d}$ 是指生成器和判别器的博弈式优化。生成器希望最小化损失, 而判别器希望最大化损失。也就是说, 生成器最小化损失以 “欺骗” 判别器, 判别器则试图准确区分数据来源, 二者博弈直至达到均衡完成训练。L_y 是源域标签监督下的分类损失 (通常为交叉熵); L_d 是域判别损失 (通常也为二元交叉熵或对数似然损失); λ 为权衡参数。UDA-2A2S 整体架构如图 1 所示。

2.2 不对称对抗模块

基于 DANN 架构, 引入不对称对抗 [16] 机制和渐进对齐策略。使用两个域判别器分别对齐全局特征与局部特征, 以避免传统 DANN 中对齐粗糙的问题。其改造后的模型架构如图 2 所示。

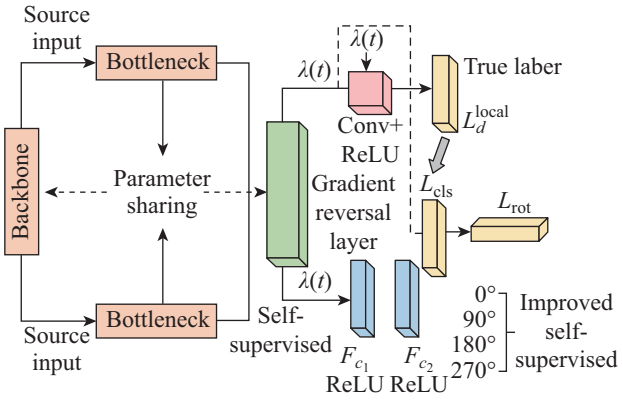

图 1 UDA-2A2S 整体网络架构

Fig. 1 The overall architecture of UDA-2A2S framework

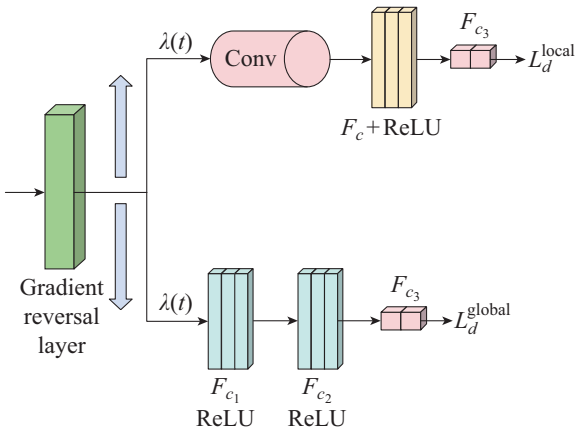

图 2 改进后的不对称对抗模块架构

Fig. 2 The improved architecture of asymmetric adversarial module

设 $G_{d,g}$ 为全局域判别器, $G_{d,l}$ 为局部域判别器。全局域判别器对整体特征分布进行域区分, 而局部域判别器则针对特征中的子空间或局部特征片段进行区分, 以捕捉细粒度的差异。此时, 全局和局部域判别损失为:

$$L_d^{\text{global}}(G_{d,g},G_f) = E_{x\in\{D_S\cup D_T\}}[-d\log(G_{d,g}(G_f(x))) - (1-d)\log(1-G_{d,g}(G_f(x)))] \tag{2}$$

$$L_d^{\text{local}}(G_{d,l},G_f) = E_{x\in\{D_S\cup D_T\}}[-d\log(G_{d,l}(G_f^{\text{local}}(x))) - (1-d)\log(1-G_{d,l}(G_f^{\text{local}}(x)))] \tag{3}$$

式中, G_f^{local} 表示从特征中抽取局部特征表示或子空间特征表示的操作。整体损失为两者加权求和:

$$L_d^{\text{all}}(G_{d,g},G_{d,l},G_f) = \alpha L_d^{\text{global}}(G_{d,g},G_f) + (1-\alpha)L_d^{\text{local}}(G_{d,l},G_f) \tag{4}$$

式中, α 为权重参数。

与此同时, 为实现渐进对齐策略, 引入时间或迭代步数相关的衰减系数 $\lambda(t)$, 使域对抗力度在初期增强、后期减弱, 促使模型先充分对齐域分布, 再逐步聚焦目标域精细特征学习。$\lambda(t)$ 设计为一个随训练轮数 t 增加逐步变化的函数, 如 $\lambda(t)=2/(1+\mathrm{e}^{-\gamma t})-1$, 其中 γ 为调节参数。

最终整体目标函数计算如下:

$$\begin{aligned}\min_{G_f,G_y}\max_{G_{d,g},G_{d,l}} L(G_f,G_y,G_{d,g},G_{d,l})=\\ E_{(x_s,y_s)\in D_S}[L_y(G_y(G_f(x_s)),y_s)]-\\ \lambda(t)E_{x\in\{D_S\cup D_T\}}[L_d^{\mathrm{all}}(G_{d,g},G_{d,l},G_f)]\end{aligned}\tag{5}$$

2.3　自监督增强模块

在无监督域自适应场景中, 目标域缺乏标注使得对抗对齐难以保证类判别能力。为此引入伪标签策略与自监督增强, 以提升目标域特征的鲁棒性与区分能力。这里的自监督增强采用传统的旋转预测 [17], 让模型从无标签数据中挖掘有助于判别的通用特征。其改造后的模型主要架构如图 3 所示。

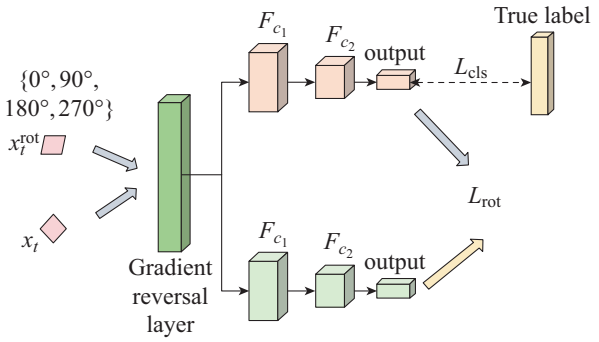

图 3　改进后的自监督模型架构

Fig. 3　Improved self-supervised model architecture

将 $y_t'=\arg\max_c p(y=c|x_t)$ 作为伪标签, 并将 (x_t,y_t') 加入伪标注集合 D_t。对这些样本计算伪标签损失:

$$L_y^T=E_{(x_t,y_t')\in D_T'}L_y(G_y(G_f(x_t)),y_t')\tag{6}$$

式中, L_y 通常为交叉熵损失。参数 τ 控制了伪标签的置信度门槛, 在训练早期应适当设置较高 τ 值, 以确保伪标签的可靠性。

在生成伪标签的同时, 引入旋转预测作为自监督任务, 通过对目标域图像施加随机旋转并训练模型判断旋转角度, 引导特征提取器学习更丰富、稳定的特征表示。此时的旋转预测损失可写为:

$$\begin{aligned}L_{\mathrm{rot}}(G_{\mathrm{rot}},G_f)=\\ E_{x_t\in D_T}\Bigg[-\sum_{r'\in\{0,90,180,270\}}q(r'=r)\times\\ \log p(r'=r|G_{\mathrm{rot}}(G_f(x_t^{\mathrm{rot}})))\Bigg]\end{aligned}\tag{7}$$

式中: $q(r'=r)$ 是指示实际旋转角度 r 的分布 (即对应的 one-hot 标签); $p(r'=r)G_{\mathrm{rot}}(G_f(x_t^{\mathrm{rot}}))$ 是模型对旋转角度的预测概率。

通过最小化旋转预测损失 L_{rot}, 促使模型学习对几何变换不敏感的通用特征。结合伪标签监督, 有助于在无监督条件下获得更稳健且更具可迁移能力的特征表示。

总体而言, 模型的最终目标可表示为多重损失项的加权求和形式:

$$\begin{aligned}L_{\mathrm{total}}=L_{\mathrm{cls}}^{\mathrm{src}}+\lambda_D((1-\alpha)L_{\mathrm{adv}}^{\mathrm{global}}+\alpha L_{\mathrm{adv}}^{\mathrm{local}})+\\ \lambda_{\mathrm{pseudo}}L_{\mathrm{pseudo}}+\lambda_{\mathrm{ssl}}L_{\mathrm{rot}}\end{aligned}\tag{8}$$

式中: $L_{\mathrm{cls}}^{\mathrm{src}}$ 为源域分类损失; $L_{\mathrm{adv}}^{\mathrm{global}}$ 与 $L_{\mathrm{adv}}^{\mathrm{local}}$ 分别为全局和局部域判别损失; L_{pseudo} 为目标域伪标签监督损失; L_{rot} 为自监督旋转预测损失; λ_D, α, $\lambda_{\mathrm{pseudo}}$, λ_{ssl} 为相应权重参数与渐进衰减因子, 旨在平衡各项损失在不同训练阶段的影响力度。通过上述综合的优化目标与策略, 模型得以在无监督域自适应场景中渐进且稳健地学习到跨域可转移、目标域可辨别的特征表示。

3　实验结果

3.1　数据集简介与实验设置

3.1.1　数据集简介

选用的数据集为 Office-Home 和 DomainNet 数据集。

Office-Home 数据集是一个常用的无监督域自适应基准数据集, 包含来自 4 个领域的图像类别: 艺术品 (art, A)、剪贴画 (clipart, C)、产品图像 (product, P) 和真实世界图像 (real-world, R)。数据集中共有 65 个类别, 涵盖办公与家居环境中的常见物品, 具有显著的领域间差异。Office-Home 数据集在评估域适应模型的跨域迁移 [18] 能力和鲁棒性方面具有重要价值。

DomainNet 数据集是一个大规模、多领域的无监督域适应基准数据集, 涵盖了 6 个不同风格的

视觉域: 剪贴画 (clipart, C)、素描 (sketch, S)、绘画 (painting, P)、真实图像 (real, R)、图标 (infograph, I) 和速写 (quickdraw, Q)。数据集中共有 345 个类别, 包含从日常物品到动物和自然场景等多样化的视觉对象。由于其规模庞大、类别多样、域间差异显著, DomainNet 数据集广泛用于评估域自适应模型在多源域和多目标域场景下的性能与泛化能力。两个数据集的示例如图 4 所示。

图 4 数据集简介

Fig. 4 Introduction to the datasets

3.1.2 实验设置

实验环境和主要超参数设置见表 1。

表 1 实验环境和超参数设置

Tab. 1 Experimental environment and hyperparameter settings

超参数	值
GPU 型号	NVIDIA RTX 3090
CPU 型号	Intel Core i9-10900K
内存	64 GB
框架版本	PyTorch 2.0
优化器	Adam
初始学习率	0.001
域判别器学习率	0.000 1
批量大小	64
训练轮数	100
伪标签置信度阈值	初始 0.9, 逐渐降低
衰减参数	0.05
自监督旋转角度	0°、90°、180°、270°

3.2 对比试验

对比试验的训练指标评估遵循无监督域自适应任务。对于 Office-Home 数据集, 测试剪贴画 (C)、产品图像 (P) 和真实世界图像 (R) 和艺术品 (A)4 个域以及其他 3 个域进行域自适应测试, 每个结果用不同的随机种子测试 3 次, 并取平均值。采用 12 次实验的平均准确率来表达数据集的整体评价。

训练 DomainNet 数据集包含 6 个数据域, 鉴于其中 A 和 B 两个域与其他数据差异太大, 并且在使用 Resnet50 作为主干的情况下很难获得更好的特征表达, 因此, 只选择数据集中的 4 个域 (真实图像 (real, R)、绘画 (painting, P)、素描 (sketch, S)、剪贴画 (clipart, C)) 进行与 Office-Home 相同的测试。其总体评价指标仍为 12 对任务的平均准确率。其实验结果如表 2 所示。

实验结果表明, 本文方法 (Ours) 在 Office-Home 12 个迁移任务中均优于现有主流模型, 平均精度达 72.4%, 高于 Meta-Align(71.3%) 和 ToAlign(71.8%)。在 A2P、C2A、R2A 等任务中优势明显, 验证了其在特征提取、域适应与分类任务中的综合优化效果。

从具体迁移任务的表现来看, 在 C2A、P2C 等高差异任务中, 模型仍保持较高准确率, 展现出良好鲁棒性; 在 P2R、R2P 等简单任务中准确率实现了 83% 及以上, 体现出更优的特征对齐与判别能力。整体结果验证了不对称对抗与自监督增强模块的有效性。表 3 所示为 domainNet 数据集域自适应的结果。

实验结果表明, 本文方法 (Ours) 在 DomainNet 数据集的跨域迁移任务上整体表现优于现有的主流域自适应方法, 平均精度达到 46.2%, 超越了 ToAlign 的 45.4%。在 P2R 和 S2C 等高差异任务中分别取得了 58.7% 和 56.2% 的准确率, 并在 R2S 和 C2S 任务中取得了 40.2% 和 42.3% 的准确率, 显著优于 DANN、PAN[30] 等基线模型, 说明其对不同风格和图像表现形式的适应能力更强。

实验结果表明, 不对称对抗与自监督增强的综合策略不仅有效缩小了源、目标域分布差异, 也挖掘了更多目标域判别信息, 在多种跨域任务中展现出更高的鲁棒性与稳定性, 验证了该方法在无监督域自适应中的有效性与广泛适用性。

表 2　Office-Home 对比实验结果表
Tab. 2　Office-Home comparison experiment results table

Model	Office-Home 数据集												
	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg
ResNet-50[19]	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
MCD[20]	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
CDANs[21]	50.7	70.6	76.0	57.6	70.0	70.0	57.4	50.9	77.3	70.9	56.7	81.6	65.8
ALDA	53.7	70.1	76.4	60.2	72.6	71.5	56.8	51.9	77.1	70.2	56.3	82.1	66.6
TADA	53.1	72.3	77.2	59.1	71.2	72.1	59.7	53.1	78.4	72.4	60.0	82.9	67.6
MDD[22]	54.9	73.7	77.8	60.0	71.4	71.8	61.2	53.6	78.1	72.5	60.2	82.3	68.1
BNM	56.2	73.7	79.0	63.1	73.6	74.0	62.4	54.8	80.7	72.4	58.9	83.5	69.4
GSDA[23]	61.3	76.1	79.4	65.4	73.3	74.3	65.0	53.2	80.0	72.2	60.6	83.1	70.3
GVB[24]	57.0	74.7	79.8	64.6	74.1	74.6	65.2	55.1	81.0	74.6	59.7	84.3	70.4
E-Mix	57.7	76.0	79.8	63.6	74.1	75.0	63.4	56.4	79.7	72.8	62.4	85.5	70.6
HDA[25]	56.8	75.2	79.8	65.1	73.9	75.2	66.3	56.7	81.8	75.4	59.7	84.7	70.9
Meta-Align[26]	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
ToAlign	57.9	76.9	80.8	66.7	75.6	77.0	67.8	57.0	82.5	75.1	60.0	84.9	71.8
Ours	60.2	77.5	81.3	67.1	75.9	77.4	68.1	57.8	83.0	76.2	62.0	85.8	72.4

表 3　DomainNet 数据集结果
Tab. 3　DomainNet dataset results

Model	DomainNet 数据集												
	R2C	R2P	R2S	C2R	C2P	C2S	P2R	P2C	P2S	S2R	S2C	S2P	Avg
ResNet-50[19]	41.6	42.7	29.6	42.4	27.2	32.1	49.5	32.5	26.7	38.7	40.8	27.5	35.9
MSTN	27.2	32.9	24.3	28.1	21.1	24.1	30.7	19.8	22.5	24.3	26.2	23.5	25.4
RSDA	27.2	35.8	24.3	36.9	24.9	31.1	41.3	26.1	24.7	29.4	26.2	27.7	29.6
MCD[20]	36.3	36.5	24.9	40.3	25.8	32.1	43.6	29.6	25.7	34.1	39.1	26.8	32.9
DANN[8]	45.9	44.5	35.4	46.8	30.5	36.7	48.0	34.7	32.1	47.1	46.4	38.4	40.5
CAN[27]	40.7	37.7	33.7	54.9	31.4	37.3	51.0	33.6	30.9	52.1	42.1	32.0	39.8
PAN	49.2	48.1	36.4	49.6	33.2	38.7	51.8	36.0	32.9	49.1	50.9	39.8	43.0
CDANs[21]	50.1	48.3	39.0	50.0	33.3	39.3	52.2	36.4	33.6	48.4	49.2	38.6	43.2
HDA[25]	46.3	47.5	34.3	49.9	33.9	37.9	55.2	40.8	32.7	49.0	49.7	40.0	43.1
FixBi[28]	51.1	49.1	39.6	50.0	34.5	41.1	52.2	36.4	33.6	50.8	53.5	41.6	44.5
ToAlign[29]	50.8	50.7	35.1	49.5	33.8	41.4	57.9	43.5	36.2	47.9	55.5	41.6	45.4
Ours	52.3	51.5	40.2	51.8	35.2	42.3	58.7	44.1	37.0	51.9	56.2	42.8	46.2

3.3　超参数敏感性实验

除了对比实验外, 本文还进行了超参数敏感性实验, 系统评估了模型在不同学习率、批量大小、域判别器权重和自监督损失系数等关键参数变化下的稳定性与鲁棒性。该实验不仅揭示了各超参数对跨域迁移性能的影响, 也为实际应用中的参数选择提供了有效指导。

首先, 本文利用 Office-Home 数据集给出了最基础的学习率超参数实验, 如表 4 所示。

实验表明, 模型对学习率和对抗损失系数具有良好的鲁棒性, 在学习率为 0.001、参数 α 为 0.5 时可获得最优性能 (72.4%), 为模型训练提供了稳定的超参数选择依据。

其次, 本文探究了 α 值在不对称对抗损失中对模型的影响, 如表 5 所示。

实验结果显示, 不对称对抗损失中的参数 α 对

表 4 学习率超参数实验结果表

Tab. 4 Table of results of hyperparameter experiments on learning rate

Lr	Office-Home 数据集												
	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg
0.005	54.7	72.3	76.2	61.4	69.8	71.1	61.0	52.3	77.4	69.2	56.4	79.9	66.9
0.004	56.8	74.2	78.1	63.7	71.3	73.2	63.2	54.1	79.2	71.5	58.7	81.5	68.7
0.003	58.9	75.6	79.8	65.8	73.2	75.0	65.4	55.9	81.0	73.1	60.1	83.2	70.3
0.002	59.5	76.7	80.5	66.3	75.0	76.1	66.2	57.0	82.1	74.5	61.2	84.3	81.2
0.001	60.2	77.5	81.3	67.1	75.9	77.4	68.1	57.8	83.0	76.2	62.0	85.8	72.4

表 5 Office-Home 实验结果表

Tab. 5 Office-Home experiment results table

α 值	Office-Home 数据集												
	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg
0.3	58.9	76.2	79.8	65.7	74.1	75.3	66.5	56.9	81.4	74.8	61.0	84.0	71.0
0.4	59.4	76.8	80.2	66.0	75.0	76.0	67.2	57.2	82.1	75.2	61.4	84.6	71.5
0.6	59.8	77.1	80.7	66.6	75.4	76.5	67.7	57.5	82.5	75.6	61.7	85.1	71.8
0.7	60.0	77.3	81.0	66.9	75.7	77.0	67.9	57.6	82.7	75.9	61.9	85.4	72.1
0.5	60.2	77.5	81.3	67.1	75.9	77.4	68.1	57.8	83.0	76.2	62.0	85.8	72.4

模型性能有显著影响。当 α 取 0.5 时, 模型在 Office-Home 数据集上取得最高准确率 72.4%, 实现了全局与局部特征对齐与分类性能的良好平衡。

3.4 消融实验

为了进一步评估模型中关键模块的有效性, 本文在 DomainNet 数据集进行消融实验, 分别移除不对称对抗损失和自监督增强, 观察模型性能变化。对比完整模型与不同消融配置下的结果, 可直观展示这两模块在特征对齐与目标域分类中的贡献, 并加深对模型内部机制和性能改进来源的理解。其详细结果如表 6 所示。

表 6 消融实验结果

Tab. 6 Ablation test results

消融设置	DomainNet 数据集												
	R2C	R2P	R2S	C2R	C2P	C2S	P2R	P2C	P2S	S2R	S2C	S2P	Avg
DANN	45.9	44.5	35.4	46.8	30.5	36.7	48.0	34.7	32.1	47.1	46.4	38.4	40.5
+不对称	49.6	48.3	38.1	49.2	33.4	40.1	53.2	39.0	34.8	50.3	51.0	40.7	44.0
+自监督增强	50.2	49.1	38.9	50.1	34.2	41.0	54.1	40.1	35.7	50.8	52.0	41.2	44.8
+ALL	52.3	51.5	40.2	51.8	35.2	42.3	58.7	44.1	37.0	51.9	56.2	42.8	46.2

实验结果表明, 不对称对抗损失和自监督增强模块对模型性能的提升具有显著贡献。相比 DANN, 单引入不对称对抗损失便使平均准确率由 40.5% 升至 44.0%, 通过全局与局部特征对齐有效缓解域间分布差异, 尤其在 P2R、S2C 等高难度任务中分别提升 5.2% 和 4.6%, 展现了更强的鲁棒性。

此外, 单独加入自监督增强后, 准确率达 44.8%, 较 DANN 提升 4.3%, 说明该模块能挖掘目标域潜在结构, 增强辨识能力。完整配置 (+ALL) 时准确率达 46.2%, 显著优于基础模型和单模块设置, 表明两模

块协同作用在复杂跨域任务中表现更佳。

3.5　Grad-Cam 可视化实验

本文在 Office-Home 数据集上进行了可视化实验, 旨在直观展示模型在不同迁移任务中的注意力分布和判别依据。热力图结果表明, 模型能聚焦目标域图像中的关键区域, 验证了其对特征空间的适应性与分类表现, 进一步支持了模型在无监督域自适应任务中的有效性与可靠性。实验结果如图 5 所示。

从 Grad-Cam 可视化结果表明, 模型能在图像中准确关注于具有判别力的关键区域, 如目标对象的结构性特征, 说明所提模型具备良好的特征提取能力与可解释性。整体而言, Grad-Cam 可视化结果直观地展示了模型的决策依据, 验证了其在跨域任务中对关键区域的有效捕捉和判别能力。

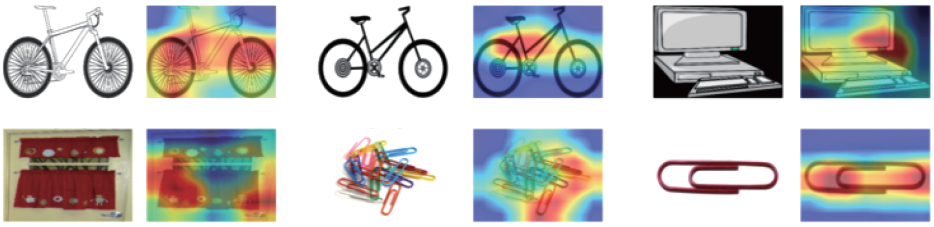

图 5　Grad-Cam 实验结果

Fig. 5　Grad-Cam experimental results

3.6　模型复杂度分析

从实验结果可以看出, 不同模型在参数量和帧率 (frame per second, FPS) 上的表现存在明显差异。CAN 模型的参数量最少 (12.5 MB), 因此 FPS 最高 (35.2), 而 ToAlign 模型参数量最大 (30.8 MB), FPS 最低 (18.1), 表明模型参数量的增加通常会导致推理速度下降。Fixbi 和 HDA 在参数量和 FPS 之间呈现一定的权衡, HDA 的参数量较 Fixbi 大 (25.3 MB vs 18.7 MB), 但 FPS 相对较低 (22.4 vs 28.9)。相比之下, 本文方法 (Ours) 在参数量 (22.1 MB) 和 FPS(31.6) 之间达到了较好的平衡, 参数量适中但推理速度较快, 表明其在保持一定复杂度的同时, 能够提供更优的实时性能。

表 7　模型复杂度分析

Tab. 7　Model complexity analysis

Model	Parames/MB	FPS
CAN	12.5	35.2
Fixbi	18.7	28.9
HDA	25.3	22.4
ToAlign	30.8	18.1
Ours	22.1	31.6

4　结　论

本文提出了一种基于不对称对抗和自监督增强的无监督域自适应算法。其创新性贡献是引入了全局与局部域判别器协同作用的机制, 实现细粒度的特征对齐。此外, 通过渐进对齐策略动态调整对齐力度, 并结合旋转预测等自监督增强模块, 从目标域中挖掘潜在结构信息, 显著提升了跨域任务的表现力。实验结果表明, 所提出方法在 Office-Home 和 DomainNet 数据集的多个迁移任务上均取得了优于现有方法的性能, 充分验证了其在处理复杂领域分布差异与目标域标注稀缺场景中的有效性。未来工作中, 可以进一步探索更复杂的多源域适配问题, 以及将不对称对抗机制与其他自监督任务相结合, 以应对更多元化的实际应用场景。

参考文献:

[1] HASAN M D A, BHARGAV T, SANDEEP V, et al. Image classification using convolutional neural networks [J]. Automatic Documentation and Mathematical Linguistics, 2024, 58(S3): S143-S149.

[2] 田萱, 王亮, 丁琪. 基于深度学习的图像语义分割方法综述 [J]. 软件学报, 2019, 30(2): 440-468.

[3] 罗会兰, 陈鸿坤. 基于深度学习的目标检测研究综述 [J]. 电子学报, 2020, 48(6): 1230-1239.

[4] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database [C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009: 248-255.

[5] LI K Q Y, LU J, ZUO H, et al. Source-free multidomain adaptation with fuzzy rule-based deep neural networks [J]. IEEE Transactions on Fuzzy Systems, 2023, 31(12): 4180-4194.

[6] 邹建. 基于卷积注意力特征迁移学习的滚动轴承故障

诊断 [J]. 计算机测量与控制, 2024, 32(1): 23-29.

[7] 李灿, 王广斌, 赵树标, 等. 采用集约化开放集差异分布对齐策略的轴承故障诊断方法 [J]. 中国机械工程, 2024, 35(9): 1622-1633.

[8] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks [J]. Journal of machine learning research, 2016, 17(1): 2096.

[9] 耿传兴, 谭正豪, 陈松灿. 自监督增广的监督分类学习增强 [J]. 软件学报, 2022, 34(4): 1870-1878.

[10] BERNHARD S, JOHN P, THOMAS H. Analysis of representations for domain adaptation[C]//Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference. Cambridge: MIT Press, 2007: 137-144.

[11] LONG M S, ZHU H, WANG J M, et al. Deep transfer learning with joint adaptation networks [C]//International Conference on Machine Learning. Sydney: PMLR, 2017: 3470-3479.

[12] CARON M, BOJANOWSKI P, JOULIN A, et al. Deep clustering for unsupervised learning of visual features [C]//Proceedings of the European Conference on Computer Vision. Swiss: Springer, 2018: 139-156.

[13] 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. Seoul: IEEE, 2019: 91-100.

[14] WESTFECHTEL T, YEH H W, MENG Q E, et al. Backprop induced feature weighting for adversarial domain adaptation with iterative label distribution alignment [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Hawaii: IEEE, 2023: 392-401.

[15] DAYAL A, CENKERAMADDI L R, VIMAL K B, et al. MADG: margin-based adversarial learning for domain generalization [EB/OL]. [2024-12-28]. https://doi.org/10.48550/arXiv.2311.08503.

[16] 林佳伟, 王士同. 用于无监督域适应的深度对抗重构分类网络 [J]. 计算机科学与探索, 2022, 16(5): 1107-1116.

[17] GIDARIS S, SINGH P, KOMODAKIS N, et al. Unsupervised representation learning by predicting image rotations [EB/OL]. [2024-12-28]. https://arxiv.org/abs/1803.07728.

[18] 王熙源, 张战成, 徐少康, 等. 面向手术导航 3D/2D 配准的无监督跨域迁移网络 [J]. 计算机应用, 2024, 44(9): 2911-2918.

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

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

[21] LONG M S, CAO Z J, WANG J M, et al. Conditional adversarial domain adaptation [C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2018: 1647-1657.

[22] ZHANG Y C, LIU T L, LONG M S, et al. Bridging theory and algorithm for domain adaptation [C]//International Conference on Machine Learning. Long Beach: ICML, 2019: 12805-12823.

[23] LIU Z S, LV Q J, LEE C H, et al. GSDA: generative adversarial network-based semi-supervised data augmentation for ultrasound image classification [J]. Heliyon, 2023, 9(9): e19585.

[24] CUI S H, WANG S H, ZHUO J B, et al. Gradually vanishing bridge for adversarial domain adaptation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 12455-12464.

[25] CUI S H, JIN X, WANG S H, et al. Heuristic domain adaptation [J]. Advances in Neural Information Processing Systems, 2020, 33: 7571-7583.

[26] WEI G Q, LAN C L, ZENG W J, et al. Metaalign: coordinating domain alignment and classification for unsupervised domain adaptation [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 16638-16648.

[27] KANG G L, JIANG L, YANG Y, et al. Contrastive adaptation network for unsupervised domain adaptation [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 4888-4897.

[28] NA J, JUNG H, CHANG H J, et al. Fixbi: bridging domain spaces for unsupervised domain adaptation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Jersey City: IEEE, 2021: 1094-1103.

[29] WEI G Q, LAN C L, ZENG W J, et al. Toalign: task-oriented alignment for unsupervised domain adaptation [J]. Advances in Neural Information Processing Systems, 2021, 34: 13834-13846.

[30] YANG L, XU S, FAN J F, et al. A pixel-level deep segmentation network for automatic defect detection [J]. Expert Systems with Applications, 2023, 215: 119388.