

跨域元优化和双通道注意力结合的 少样本多源域目标检测

朱松豪, 王双丞

(南京邮电大学自动化学院, 江苏南京 210023)

摘要: 本文针对一个新的、具有挑战性的问题,即实现源域、中间域到单个目标域的知识转移,其中目标域的每个类别仅有少量标记样本. 此种情况下的知识转移过程面临以下两个困难:一是目标数据极其稀缺,从而导致没有足够的目标域特征分布;二是现有的少样本学习方法无差别地提取每部分特征,从而导致少样本目标检测性能不佳. 为解决上述问题,本文提出一种少样本多源域目标检测方法. 提出一种新的元优化机制,通过引入的混合域将源域和目标域对齐,用以缓解目标域稀缺特征分布的问题. 具体而言,首先利用图像级混合生成混合图像,和相应的标签共同构成第一个混合域;然后通过双通道注意力机制生成细粒度特征,再利用特征级混合生成特征级混合特征,和相应的标签共同构成第二个混合域;最后通过区域建议网络和感兴趣区域网络生成感兴趣区域特征,再利用感兴趣区域特征级混合生成 ROI(Region Of Interest)级混合 ROI 特征,和相应的标签共同构成第三个混合域. 生成的三个混合域共同用于计算损失函数,完成元优化过程. 提出一种包含卷积层和特征校准的双通道注意力机制,用以学习更具判别性的深度特征表征,其中卷积层用于防止关键空间信息的丢失,特征校准用于选择性地增强重要特征并削弱非重要特征. 具体而言,首先利用卷积层子模块生成粗粒度特征表示;其次,利用特征校准子模块根据特征间的相关性建立注意力权重,并将这些注意力权重与原始特征进行整合,从而有选择地强化重要区域,同时抑制不重要区域. COCO 数据集和 PASCAL-VOC 数据集的大量实验结果证明了本文提出的跨域元优化和双通道注意力结合的少样本多源域目标检测方法的有效性和鲁棒性. 在检测效果上超越了同领域内其他方法,同时在不同数据集上保持了良好的泛化性能,此外模型的参数量在同领域内相比其他方法有显著优势.

关键词: 少样本多源域目标检测;跨域元优化;双通道注意力

基金项目: 国家自然科学基金(No.52405065)

中图分类号: TP391.4

文献标识码: A

文章编号: 0372-2112(2025)10-3659-12

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20250309

Cross-Domain Meta Optimization and Dual-Channel Attention for Few-Shot Multi-Source Domain Object Detection

ZHU Song-hao, WANG Shuang-cheng

(School of Automation, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China)

Abstract: This article addresses a novel and challenging problem of knowledge transfer from the source domain and the intermediate domain to a single target domain, where each category in the target domain has few labeled samples. The knowledge transfer process in this situation faces two difficulties: the target data is extremely scarce, resulting in insufficient target domain feature distribution. Existing few-shot learning methods often extract features from each part indiscriminately, resulting in poor performance in few-shot object detection. To solve the above problems, this paper proposes a few-shot multi-source domain object detection method. A new meta optimization mechanism is proposed to align the source domain and target domain by introducing a mixed domain, alleviating the problem of scarce feature distribution in the target domain. Firstly, image-level mixing is used to generate mixed images, which together with corresponding labels form the first mixed domain. Then, fine-grained features are generated through a dual-channel attention mechanism, and feature level mixing is used to generate feature level mixed features, which together with corresponding labels form the second mixed domain. Finally, region of interest features are generated through a region recommendation network and a region of interest

network, and then ROI (Region Of Interest) level mixed ROI features are generated through feature-level mixing of the region of interest, which together with corresponding labels form the third mixed domain. The three generated mixed domains are used together to calculate the loss function and complete the meta optimization process. A dual channel attention mechanism including convolutional layers and feature calibration is proposed to learn more discriminative deep feature representations, where convolutional layers are used to prevent the loss of key spatial information, and feature calibration is used to selectively enhance important features and weaken non important features. Firstly, the convolutional layer submodules are used to generate coarse-grained feature representations. Secondly, the feature calibration submodules are used to establish attention weights based on the correlation between features, and these attention weights are integrated with the original features to selectively enhance important regions while suppressing unimportant regions. A large number of experimental results on the COCO dataset and PASCAL-VOC dataset demonstrate the effectiveness and robustness of the proposed method. It surpasses other methods in the same field in terms of detection performance, while maintaining good generalization performance on different datasets. Furthermore, the model's parameter count has significant advantages compared to other methods in the same field.

Key words: few-shot multi-source domain object detection; cross-domain meta optimization; dual-channel attention
Foundation Item(s): National Natural Science Foundation of China (No.52405065)

1 引言

本文提出了一个具有挑战性和现实性的少样本多源域目标检测问题,其现实性在于大规模数据集的收集和注释需要大量的人力、物力、财力的投入,挑战性在于罕见疾病、稀有物种、军事目标等数据样本难以获取.该问题假设每个源域包含多个标记样本,而目标域包含很少的标记样本.与现有的包括无监督域自适应、有监督域适应和多监督域自适应等在内的域自适应问题不同,本文问题中的目标领域提供了很少的标记样本用以辅助模型训练.现有的有监督域适应方法旨在将知识从单个源域转移到目标域,因此,有监督域适应方法不能用于处理涉及多个源域间与目标域间对齐的少样本多源域目标检测问题.此外,现有的多监督域自适应方法^[1,2]旨在通过将目标域与每个源域对齐,从而学习域不变特征表征.然而,这些多监督域自适应方法并不适合处理本文提出的少样本多源域目标检测问题,这是因为少样本多源域目标检测中包含少量标记样本的目标域,不适合学习域不变特征表征.最近,文献[3]提出少样本多源域自适应方法,以解决这样的一种应用场景,即源域中只有少量样本进行标记,而源域中的大多数样本和目标域中的所有样本都未进行标记.与少样本多源域自适应方法不同,少样本多源域目标检测假设所有源域样本、所有引入的混合域样本及目标域中的少数样本均进行了注释.由于使用未标记的目标域样本,从而限制了现有的无监督域自适应方法的知识转移效果;由于目标域数据稀缺,从而导致了现有的有监督域适应方法的知识转移困难;而本文提出的少样本多源域目标检测方法丰富了目标域样本,有效解决了知识转移困难问题.

本文提出一种新的元优化机制,解决所提的少样本多源域目标检测问题中的目标域数据稀缺问题.首

先,构建三个中间混合域有效丰富数据多样性,且最初将中间混合域设置靠近目标域,以减轻目标域样本稀少限制.然后,调整两张图像或两个特征的线性组合混合比引导混合域逐渐接近源域,并引导目标域进一步接近源域,实现知识从源域逐步继承至目标域.最后,在元学习机制中优化多个源域,提高少样本多源域目标检测的稳定性和鲁棒性.

现有大多数方法通过捕获复杂的通道依赖关系或引入额外的空间注意力,提高少样本目标检测性能^[4].然而,这些方法存在以下问题:一是在样本稀少情况下,复杂的网络架构可能会增加过拟合的风险;二是复杂的注意力模型可能会对模型性能产生负面影响.为解决上述问题,本文提出包含卷积子模块和特征校准子模块的双通道注意力模块.卷积子模块利用卷积层替换全连接层,提取对少样本目标检测至关重要的空间信息.特征校准子模块用于强化重要特征并抑制不重要特征,且无需提取大计算量的空间注意力,从而减少了参数规模、优化了元学习模型.

本文主要贡献可归结为如下三点:

(1)提出了一个具有挑战性的任务——少样本多源域目标检测,旨在将知识从含有多个标记样本的源域转移至含有少量标记样本的目标域.

(2)提出了一种新的元优化机制,旨在解决少样本多源域目标检测问题,该机制创建了三个中间混合域,并逐步调整混合比 λ ,以缓解目标域和源域间的域偏移.

(3)提出了一种包含特征校准子模块和卷积子模块的双通道注意力模块,旨在缓解提取全局信息时重要特征遗漏的问题,其中特征校准子模型用于强化重要特征的表征,卷积子模块用于保障空间信息的完整性.

2 相关工作

2.1 少样本学习

少样本学习方法大致可分为以下两类:(1)基于元学习的少样本学习方法^[5,6];(2)基于度量学习的少样本学习方法^[7,8].文献[9]利用抗混叠语义重建模型处理少样本学习问题,文献[10]利用变分推理进一步提高少样本学习性能,文献[11]利用基于注意力的多视图特征协作策略解决少样本学习中的数据稀疏问题.

2.2 少样本目标检测

现有的目标检测方法通常分为一阶段法和两阶段法^[12,13].文献[14]利用教师模型和元训练检测器,通过自适应伪标记引导学生模型学习.文献[15]利用几何感知特征,解决类别特征判别不足问题.然而,大数据集的泛化性能要求严重限制了这两类少样本目标检测方法的应用.

2.3 多域自适应

多源域自适应方法旨在学习所有域的域不变特征,或利用多个源域训练的子分类器学习鲁棒分类器^[16].现有的域对齐方法旨在通过对齐多源域特征学习域不变特征^[17,18],而数据增强方法旨在利用源域多样性提高模型泛化性^[19].与上述方法不同,少样本多源域目标检测中的多源域包括一个源域和通过数据增强获得的三个中间域,而非多个源域数据集.

2.4 注意力机制

文献[20]在通道注意力中引入挤压和激励机制.文献[21]提出一种局部跨通道交互策略,捕获跨通道信息并有效避免降维.文献[22]提出细粒度识别的多注意力元学习模型.与上述方法不同,本文提出的双通道注意力模块选择性地强化重要特征,并抑制不重要特征.

2.5 数据增强

文献[23]提出域增强元学习策略,通过类级和域级的混合策略增强源域,学习域不变特征.文献[24]利用元学习策略动态更新混合率,但其需要特定验证学习混合率,且不考虑多个域间的混合.相比之下,本文利用三个层级的跨域混合策略丰富目标域数据,且训练过程中自动调整混合率.

2.6 元优化

元优化旨在通过元学习,提升模型的效率和性能.文献[25]提出基于LSTM(Long Short-Term Memory)的元优化器,将梯度下降过程建模为序列学习问题,实现高效的参数更新.文献[26]通过元学习优化区域提议网络生成的候选框特征,显著提升检测性能.文献[27]通过梯度更新元模型参数,并利用先验知识加速新任务收敛.文献[28]首次将DETR(DEtection TRansformer)框架引入少样本检测,通过图像级元学习统一目标定位与分

类,消除区域建议依赖,并通过语义对齐提升泛化能力.

3 方法

3.1 问题设置

对于提出的少样本多源域目标检测问题,假设 M 个所有样本标记的源域,少量标记样本的目标域.之前领域自适应方法中的多源域表示完全独立的多个源域,而本文中所述的多源域包含源域和中间域,中间域是通过源域和目标域数据增强构成.第 i 个源域 $D_{s,i} = \{(x_{s,i}^j, y_{s,i}^j)\}_{j=1}^{N_{s,i}}$ 包含从源分布 $P_{s,i}(x,y)$ 中提取的 $N_{s,i}$ 个标记样本,目标域 $D_t = \{(x_t^j, y_t^j)\}_{j=1}^{N_t}$ 包括从目标分布 $P_t(x,y)$ 中提取的 N_t 个标记样本,这里的 $N_t \ll N_s$, 即目标域至少含有 1 个样本. $P_t(x,y) \neq P_{s,i}(x,y)$ 和 $P_{s,i}(x,y) \neq P_{s,j}(x,y) (i \neq j)$. 多个源域和目标域具有 K 个类别的相同标签空间 $Y = \{1, 2, \dots, K\}$. 本文目标是在 $\{D_{s,i}\}_{i=1}^M$ 和 D_t 上学习一个自适应模型 H , 并推广至目标域样本. 一般来说, $H = G * D$, 其中生成器 $G: x \rightarrow G$ 用于将输入样本 x 映射至嵌入空间,检测头 $D: G \rightarrow f$ 用于预测样本的类别和回归的位置.

3.2 总体介绍

为解决少样本多源域目标检测中数据稀缺和重要特征提取不足的问题,本文提出如图 1 所示的基于跨域元优化和双通道注意力的模型框架,其中混合图像批次表示源域和目标域的混合图像,特征级混合特征表示通过卷积获得的混合特征,ROI(Region Of Interest)级混合 ROI 特征表示感兴趣区域网络获得的混合特征. 首先,跨域知识转移策略包括跨域混合和跨域对齐,其中跨域混合用于丰富目标域样本并缓解目标域数据稀缺,跨域对齐用于调整跨域混合比加速知识从源域到目标域的转移. 其次,利用双通道注意块充分提取特征表征,并根据特征表征权重评估特征重要性. 最后,利用元优化过程实现从源域到目标域的知识迁移.

3.3 跨域知识转移

3.3.1 数据增强

数据增强核心思想是以凸方式随机插值构建虚拟样本,丰富样本多样性提高模型泛化性. 具体而言,给定两个样本 (x_1, y_1) 和 (x_2, y_2) , 混合样本 (\hat{x}, \hat{y}) 定义为

$$\begin{cases} \hat{x} = \lambda x_1 + (1 - \lambda)x_2 \\ \hat{y} = \lambda y_1 + (1 - \lambda)y_2 \end{cases} \quad (1)$$

其中,混合比 λ 决定了两张图像或两个特征的线性组合比例.

3.3.2 跨域混合

在元优化过程中,跨域混合包含图像级别混合、跨域特征级别混合和感兴趣区域特征级别混合.

跨域图像级混合. 受自监督学习中图像数据增强

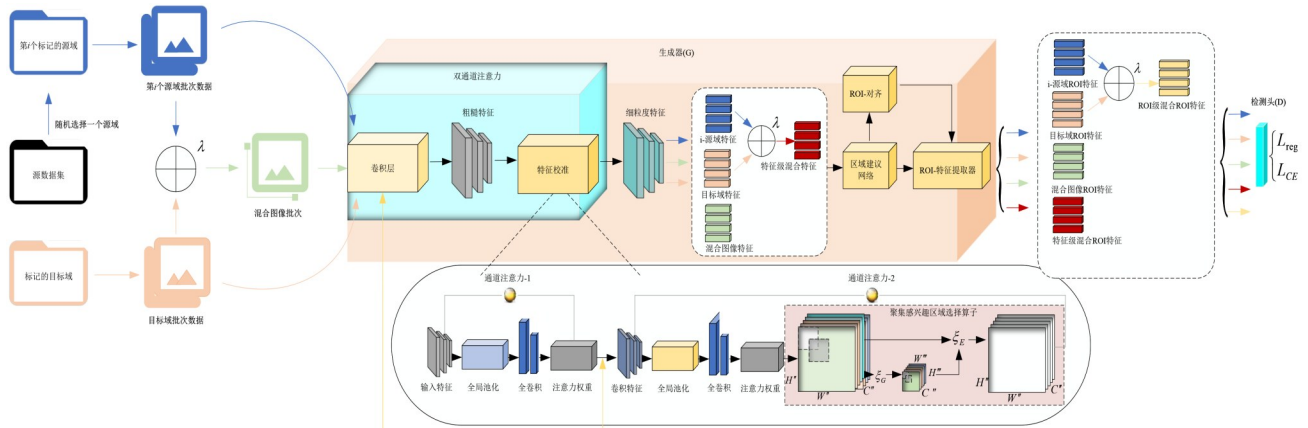


图1 所提模型的框架

策略的启发,本文将跨域图像级混合引入少样本多源域目标检测,构建具有混合标签的增强样本,缓解目标域样本稀缺问题.使用下式进行线性插值:

$$\begin{cases} \hat{x}_{img} = \lambda x_{s,i} + (1-\lambda)x_i \\ \hat{y}_{img} = \lambda y_{s,i} + (1-\lambda)y_i \end{cases} \quad (2)$$

其中, \hat{x}_{img} 和 \hat{y}_{img} 分别表示融合后的源域和目标域样本.

跨域特征级混合.假设一对源域和目标域特征及相应的标签分别表示为 $(g_{s,i}, y_{s,i})$ 和 (g_t, y_t) ,则可利用下式进行混合:

$$\begin{cases} \hat{g}_{feat} = \lambda g_{s,i} + (1-\lambda)g_t \\ \hat{y}_{feat} = \lambda y_{s,i} + (1-\lambda)y_t \end{cases} \quad (3)$$

其中, \hat{g}_{feat} 和 \hat{y}_{feat} 分别表示融合后的源域、目标域特征.

跨域感兴趣区域特征级混合.假设一对源域和目标域感兴趣区域特征及其相应的标签分别表示为 $(g_{s,i}^{ROI}, y_{s,i}^{ROI})$ 和 (g_t^{ROI}, y_t^{ROI}) ,则可利用下式进行混合:

$$\begin{cases} \hat{g}_{feat}^{ROI} = \lambda g_{s,i}^{ROI} + (1-\lambda)g_t^{ROI} \\ \hat{y}_{feat}^{ROI} = \lambda y_{s,i}^{ROI} + (1-\lambda)y_t^{ROI} \end{cases} \quad (4)$$

其中, \hat{g}_{feat}^{ROI} 和 \hat{y}_{feat}^{ROI} 分别表示融合后的源域、目标域感兴趣区域特征.

对特征级混合和图像级混合使用相同的混合比,

此时图像级混合样本和特征级混合样本位于同一语义空间.不同级别间的混合不存在信息冲突,这是因为三种混合样本是在前向传播过程中分别利用成对的源域样本和目标域样本生成,三种跨域混合域在元优化中的应用具体请见3.5节.

3.3.3 跨域对齐

文献[24]发现,混合比取值对域知识转移至关重要,这是因为混合比决定了源域和目标域的混合程度.也即,如果混合比为恒定值或遵循某种规律,则随后的跨域对齐要么非常困难,要么不太合适.

为缓解上述问题,引入瓦斯坦距离至“源域-混合域” $d_w(G_s, G_{mix})$ 和“目标域-混合域” $d_w(G_t, G_{mix})$,其中 G_s, G_t 和 G_{mix} 分别表示源域、目标域和中间域的嵌入.由训练过程可知,如最初混合域接近目标域,则跨域对齐相对简单,这是因为此时的 $d_w(G_s, G_{mix})$ 很小,且由于存在大量源域样本, $d_w(G_s, G_{mix})$ 可有效最小化.随混合比的逐渐增加,混合域和目标域均逐渐趋向于源域.如图2所示,这种混合比的逐步调整有助于缓解域差异,并确保整个知识转移过程的顺利进行,水平箭头表示训练过程,“对齐”表示元优化过程,混合比 λ 用于对齐源域和目标域.

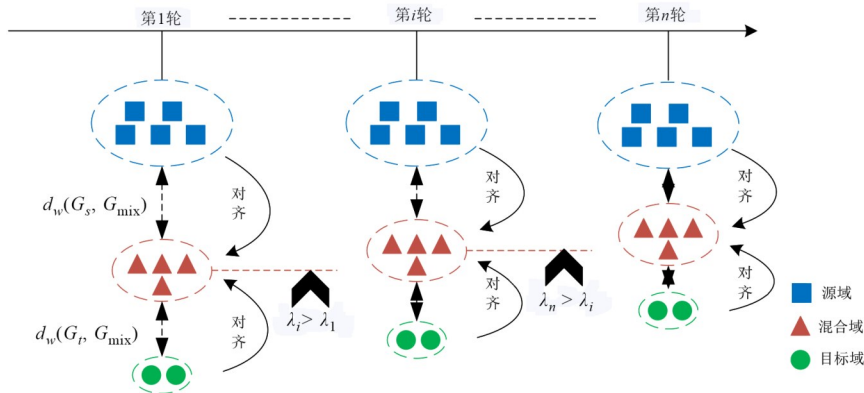


图2 跨域对齐策略示意图

跨域对齐的具体实现如下所述. 首先,利用下式所示加权因子 q 描述中间域/目标域与源域间的接近度:

$$q = \exp\left(-\frac{d_w(G_s, G_{\text{mix}})}{d_w(G_s, G_{\text{mix}}) + \beta d_w(G_t, G_{\text{mix}})}\right) \quad (5)$$

其中, β 为 0.05 的因子. G_{mix} 最初设为接近目标域. 利用式(5)渐进调整加权因子 q , 从而确保跨域对齐过程中每轮选择合适的混合比. 其次, 渐进式混合过程可用下式表示:

$$\lambda_n = \frac{n(1-q)}{N} + q\lambda_{n-1} \quad (6)$$

其中, N 表示迭代总数, $n-1$ 表示当前迭代次数, 初始混合比 λ_0 设为 0. 在以 q 作为加权因子的基础上, 利用式(6)渐进调整混合比, 从而完成跨域对齐. 最后, 为稳定训练过程, 在混合比中引入均匀分布 Λ :

$$\lambda_n = \text{Clamp}(\Lambda(\lambda_n - \sigma, \lambda_n + \sigma), \min = 0.0, \max = 1.0) \quad (7)$$

其中, σ 表示局部扰动, 这里设为 0.2. λ_n 在 $[0.0, 1.0]$ 范围内随机采样.

3.4 双通道注意力

由图 1 可知, 双通道注意力元学习模块包含卷积子模块和特征校准子模块. 首先, 利用卷积子模块生成粗粒度特征表示; 其次, 利用特征校准子模块根据特征相关性建立注意力权重, 并将这些注意力权重与原始特征进行整合, 从而有选择强化重要区域, 同时抑制不重要区域.

3.4.1 卷积层

大多数少样本学习在特征提取器后采用全连接进行特征提取, 然而, 全连接的缺点是其将输入特征图展平为一维向量, 导致重要空间信息丢失, 削弱模型泛化能力. 为克服上述问题, 这里采用特征提取器加卷积层作为卷积子模块, 提取识别目标域类别至关重要的空间信息.

卷积过程如下所述. 假设输入特征 $\mathbf{X} = [x_1, x_2, \dots, x_{C'}]$ (其中 $x \in \mathbf{R}^{H' \times W'}$), 以及卷积核 $\mathbf{K} = [k_1, k_2, \dots, k_n]$ (其中 H', W', C' 和 n' 分别表示高度、宽度、通道数和核数), 则卷积可表示为

$$\text{Conv}(X, k) = \sum_{s=1}^{C'} x_s * k'_s \quad (8)$$

其中, x_s 是 \mathbf{X} 的单通道, k'_s 是 k 的二维空间滤波器.

3.4.2 特征校准

少样本学习在卷积过程中通常会不加区分地从图像每个区域提取局部信息, 这导致以下问题: 一是过于局部化, 缺乏全局信息; 二是无法关注显著特征. 为解决上述问题, 本文提出基于人类认知机制的特征校准子模块.

特征校准子模块用于优化特征表征 (如 $U \rightarrow V$, $U = [u_1, u_2, \dots, u_{C'}] \in \mathbf{R}^{H' \times W' \times C'}$, $V = [v_1, v_2, \dots, v_{C'}] \in \mathbf{R}^{H'' \times W'' \times C''}$). 特征校准子模块包含两个递进的注意块.

(1) 第一个注意力模块. 该注意力块用于获取全局

上下文信息, 缓解过度参数化问题, 具体过程如下所述.

第一步, 对输入特征 U 执行全局平均池化:

$$z_{\text{avg}} = \text{AvgPool}(u_c) = \frac{1}{H'' \times W''} \sum_{h''=1}^{H''} \sum_{w''=1}^{W''} u_c(h'', w'') \quad (9)$$

其中, u_c 表示图中 U 的第 c 个通道的特征映射, H'' 和 W'' 表示特征映射 u_c 的高度和宽度.

第二步, 利用两个卷积捕获通道相关性/注意力权重:

$$\begin{cases} a_{c'} = F_{\text{cm}}(z_{\text{avg}}, W'') = \sigma(W_2 \delta(W_1 z_{\text{avg}})) \\ W_1 \in \mathbf{R}^{\frac{C''}{r} \times C''}, W_2 \in \mathbf{R}^{C'' \times \frac{C''}{r}} \end{cases} \quad (10)$$

其中, σ 表示 sigmoid 函数, δ 表示 ReLU 函数, W_1 和 W_2 分别表示两个卷积层的权重参数, r 表示超参数.

第三步, 通过将注意力权重 $a_{c'}$ 嵌入源特征, 获得第一个注意力表征 $U' = [u'_1, u'_2, \dots, u'_{C'}]$:

$$u'_{c'} = F_{\text{scale}}(u_{c'}, a_{c'}) \quad (11)$$

其中, F_{scale} 表示源特征映射 $u_{c'}$ 和注意力权重 $a_{c'}$ 间的通道相乘.

最后, 将第一个注意力表征 U' 输入卷积模块:

$$g_{c'} = \text{Conv}(U', k) \quad (12)$$

其中, k 表示卷积核, Conv 表示为如式(8)所示的卷积.

(2) 第二个注意力模块. 此注意力块用于定位不同对象间显著差异的区域, 具体实现过程如下所述.

第一步, 对输入特征 $g_{c'}$ 执行全局最大池化:

$$\begin{aligned} z_{\text{max}} &= \text{MaxPool}(g_{c'}) = \max(g_{c'}(h'', w'')) \\ \text{where } h'' &\in (0, H''), w'' \in (0, W'') \end{aligned} \quad (13)$$

其中, H'' 和 W'' 分别表示输入特征 $g_{c'}$ 的高度和宽度.

第二步, 通过将 z_{max} 输入到 F_{scale} 函数, 获取第二个注意力表征 $v_{c'}$:

$$v_{c'} = F_{\text{scale}}(g_{c'}, m'_{c'}) = F_{\text{cm}}(z_{\text{max}}, W'') \cdot g_{c'} \quad (14)$$

第三步, 为评估空域尺寸对聚焦重要特征的影响, 引入聚集感兴趣区域选择算子:

$$\zeta_G(\Psi, e) = \left\{ e^\Psi + \delta, \delta \in \left[-|(2e^{-1})/2|, \langle (2e^{-1})/2| \right] \right\} \quad (15)$$

其中, e 表示 $(0 \sim 1)$ 间选择的范围比, Ψ 表示每个通道的输出位置.

第四步, 定义一个范围比 e 的聚集算子:

$$\zeta_G: \mathbf{R}^{H'' \times W'' \times C''} \rightarrow \mathbf{R}^{H'' \times W'' \times C''} \left(H'' = \left[\frac{H''}{e} \right], W'' = \left[\frac{W''}{e} \right] \right) \quad (16)$$

对于 V 的任何输入 v , ζ_G 满足以下方程:

$$\zeta_G(v)_{\Psi}^{c'} = \zeta_G(x \otimes I_{\zeta(\Psi, e)}) \quad (17)$$

其中, $\Psi \in \{1, 2, \dots, H''\} \times \{1, 2, \dots, W''\}$, $c' \in \{1, 2, \dots, C''\}$, $I(\cdot)$ 表示指示张量, \otimes 表示哈达玛积. 式(17)表示在通道 c' 的输出位置 Ψ , 聚集算子都有一个面积为 $(2e^{-1})^2$ 的感受野. 如感受野包围整个特征映射, 则可认为聚集算子具有全局范围.

第五步,利用激励算子获取上下文信息:

$$\xi_E(v, \hat{v}) = v \otimes f(\hat{v}) \quad (18)$$

其中, $f: \mathbf{R}^{H^m \times W^m \times C^m} \rightarrow [0, 1]^{H^m \times W^m \times C^m}$ 用于缩放聚集后的特征映射.

最后,校准模块输出精细特征 $[0, 1]^{H^m \times W^m \times C^m}$. 式(16)~(18)的示意如图1所示.

3.5 元优化过程

本文模型的总体架构如图1所示,主要由特征生成器 G 和检测头 D 组成. 由于少样本多源域目标检测中存在多个标记样本的源域和一个少量标记样本的目标域,因此这里选择域泛化框架,使用元学习训练模型.

如将模型参数 $F * G$ 表示为 θ , 则损失可定义为

$$L^T_i(\theta) = - \sum_{x,y \in T_i} \sum_{k=1}^K y_k \log(\theta(x)_k) + L_{\text{reg}} \quad (19)$$

其中, T_i 表示某个特定域(如源域或目标域), x 表示输入图像, y 表示相应的标签真值, K 表示类别数. 此外, L_{reg} 表示回归损失.

受模型不可知元学习方法^[29]的启发,利用元优化的元训练过程,伪更新模型参数:

$$\theta' = \theta - \alpha \nabla_{\theta} \sum_{T_i \in P(T)} L^T_i(\theta) \quad (20)$$

其中, α 表示更新系数,实验中设为0.001. 此外, $p(T)$ 表示域采样,实验中随机从一个源域和三个中间域采样两个.

在元测试过程中,将未被采样的源域、目标域和混合域用于更新模型参数:

$$\theta = \theta - \beta \nabla_{\theta} \sum_{T_i \notin P(T)} L^T_i(\theta') \quad (21)$$

其中, β 表示更新系数,实验中设为0.001.

4 实验

4.1 实验细节

数据集. 这里采用 MS COCO 和 ASCAL 两个公开数据集评估本文所提模型的性能. 从数据集中随机选择某些类别作为目标域,其余类别则作为源域.

性能评价协议. 为精确将目标域样本限制为少量样本,从 COCO 数据集中随机选择 10 个类别,再从这 10 个类别中随机选择少量样本构建目标域,并用这 10 个类别中的剩余样本构建测试集;此外,利用其余的 70 个类别构建元训练的源域. 类似地,从 PASCAL-VOC 数据集中随机选择 5 个类别,再从这 5 个类别中随机选择少量样本(1~10 个样本)构建目标域,并用这 5 个类别中的剩余样本构建测试集;此外,利用其余的 15 个类别构建元训练的源域. 以上的构建形式能够模拟实际的多域场景,确保源域和目标域分布异质.

对比方法. 性能比较实验选择了三种主流方法:(1)域泛化方法,例如域增强元学习方法(Domain Augmentation Meta Learning, DAML)^[23]; (2)数据增强方法,例如 Meta Mix-Up^[24]; (3)监督域自适应方法,例如分类和对比语义对齐方法(Classification and Comparison Semantic Alignment, CCSA)^[30]. 此外,以 TFA^[31]作为基线进行实验.

评估指标:平均平均精度(Mean Average Precision, MAP)表示三次实验中所有类别平均精度的平均值.

4.2 实验结果

PASCAL VOC 实验结果. 由表1所示实验结果可得以下结论. 与基线^[31]相比,本文方法的检测性能得到

表1 不同方法在 PASCAL VOC 数据集的实验结果

方法	组 1					组 2					组 3				
	1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
DAML ^[23]	49.1	49.5	50.4	50.9	52.2	27.2	28.1	42.4	43.6	44.9	51.1	39.9	45.2	51.3	51.5
CCSA ^[30]	51.1	52.3	53.2	53.7	54.7	28.5	28.9	43.2	45.5	46.8	52.3	50.6	56.7	57.2	57.9
TFA ^[31]	36.8	29.1	43.6	55.7	57.0	18.2	29.0	33.4	35.5	39.0	27.7	33.6	42.5	48.7	50.2
TFA+CORPNs ^[31]	39.8	36.1	44.7	55.7	56.0	23.5	26.9	34.1	35.1	39.1	30.8	34.8	42.8	49.5	49.8
IKD ^[32]	41.0	51.7	55.7	61.8	65.4	30.7	39.0	42.5	46.6	51.7	37.9	47.1	51.7	56.8	59.5
Meta Faster-RCNN ^[33]	43.0	54.5	60.6	66.1	65.4	27.7	35.5	46.1	47.8	51.4	40.6	46.4	53.4	59.9	58.6
IRP ^[34]	47.0	44.8	53.4	62.9	65.2	29.8	32.6	46.3	47.7	53.0	40.1	45.9	49.6	57.0	59.7
UNP ^[35]	43.7	58.3	59.8	63.7	64.2	28.1	42.8	47.7	49.5	50.3	38.4	49.3	53.8	57.7	58.7
FPD ^[36]	37.7	51.2	59.0	64.7	67.8	28.7	40.0	44.3	50.2	55.5	29.2	48.3	52.1	58.4	62.1
VFA ^[37]	57.7	64.6	64.7	67.2	67.4	41.4	46.2	51.1	51.8	51.6	48.9	54.8	56.6	59.0	58.9
CRENet ^[38]	49.3	55.2	55.0	64.7	64.3	30.4	34.6	48.3	49.7	53.4	42.6	44.1	50.9	58.4	62.1
FSMT ^[39]	44.6	53.1	51.5	60.7	63.5	28.4	33.5	44.4	47.0	51.3	41.7	45.7	46.9	54.5	57.1
OPN ^[40]	53.7	58.8	60.2	65.0	66.9	41.8	46.4	50.1	53.6	47.1	53.3	56.1	60.0	61.6	54.2
IFRCNN ^[41]	35.7	38.9	46.9	60.1	62.2	22.1	28.4	38.6	39.4	43.2	31.0	36.8	42.3	51.6	55.0
DRC-FSOD ^[42]	47.2	53.4	56.6	63.5	65.6	28.5	32.9	46.3	46.9	53.3	42.4	44.9	49.5	56.6	59.6
Ours	58.3	64.7	65.3	70.9	68.7	42.4	46.7	51.3	51.5	55.7	53.6	57.4	60.9	64.4	62.9

显著提高,其中当目标域样本数量为1时,三个组别的MAP50分别增加了21.5、24.2和25.9.与文献[32]相比,当目标域样本数量为5时,本文方法在三个组别上的MAP50分别提高了9.1、4.9和7.6.与其他少样本目标检测方法相比,本文方法的检测性能也有一定程度提高.

MS COCO2014实验结果.由表2所示实验结果可

得以下结论.与基线^[31]相比,当目标域别样本数量为1时,本文方法分别将MAP、MAP50和MAP75提高了2.8、6.3和5.5.与Meta Faster RCNN^[33]相比,对于不同数量的目标域样本,本文方法的MAP分别提高了1.1、0.8和1.3.与最新的少样本目标检测方法相比,本文方法的检测性能显著提高,表明该方法具有很强的竞争力.

表2 不同方法在MS COCO2014数据集的实验结果

方法	1 样本			2 样本			3 样本		
	MAP	MAP50	MAP75	MAP	MAP50	MAP75	MAP	MAP50	MAP75
DAML ^[23]	4.5	11.6	5.2	5.9	11.2	6.4	7.6	13.8	8.1
CCSA ^[30]	4.1	11.1	4.5	5.3	10.3	5.8	7.1	13.3	7.5
TFA+CORPNs ^[31]	3.4	5.8	3.8	4.6	8.3	4.8	6.6	12.1	6.5
IKD ^[32]	5.3	11.2	6.1	7.8	9.5	7.0	8.9	11.5	9.0
Meta Faster-RCNN ^[33]	5.1	10.7	4.3	7.6	16.3	6.2	8.3	15.2	8.0
FSMT ^[39]	5.3	10.7	7.2	7.6	12.9	7.3	9.2	16.7	9.5
OPN ^[40]	5.2	9.9	6.6	6.4	12.4	6.7	7.6	13.7	8.1
FSDetView ^[43]	3.2	8.9	1.4	4.9	13.3	2.3	6.7	18.6	2.9
MPF-Net ^[44]	4.4	11.4	4.9	5.7	10.9	6.2	7.4	13.5	7.8
DRL ^[45]	5.6	11.7	8.2	7.9	15.2	7.6	8.7	17.8	9.2
Ours	6.2	12.1	9.3	8.4	16.6	8.2	9.6	19.3	10.7

4.3 消融实验

4.3.1 元优化的消融实验

从表3所示实验结果可以得出以下结论:(1)索引2/3策略的性能分别优于索引1/2的策略,这突显了本文所提渐进混合策略的优越性;(2)跨域特征级混合策

略比跨域图像级混合策略表现出更好的性能,平均改善率超过4.0;(3)如果只能选择一个策略,跨域特征级混合策略将是更好选择,如果策略数量没有限制,图像级混合策略和特征级混合策略的组合可进一步提高模型准确性.

表3 PASCAL VOC数据集上不同数据增强策略的消融实验结果

索引	图像级混合 ^[46]	特征级混合 ^[46]	跨域混合	组1					组2					组3				
				1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
1	√	—	—	36.8	39.1	43.6	48.7	50.0	20.2	25.0	29.4	32.5	34.0	27.7	31.6	38.5	44.7	48.2
2	—	√	—	39.5	45.2	47.6	50.2	52.1	22.6	27.7	31.5	35.1	36.6	28.9	32.9	39.1	46.3	49.5
3	—	—	√	42.7	49.2	50.3	53.2	54.9	25.0	29.9	33.3	37.3	39.5	31.1	34.7	39.9	48.6	51.1

注:加粗数据为最优结果.

在PASCAL VOC数据集上进行对比实验,研究不同域自适应策略对模型性能的影响.文献[47]采用 $\lambda \sim \text{Beta}(0, 1)$ 形式的域混合比采样策略,本文采用如式(6)所示的自动更新的域混合比采样策略.从表4所示实验结果可得以下结论:(1)相较于 $\lambda \sim \text{Beta}(0, 1)$ 形式的域混合比采样策略,本文自动更新的域混合比采样策略表现出更好的性能;(2)自动更新的域混合比采样策略和跨域混合策略可协同工作,展示最佳性能.

从图3所示损失曲线可得以下结论:(1)固定采样策略的训练收敛速度较慢;(2)相较于固定采样策略,Beta采样策略下的模型稳定性较差;(3)相较于Beta采样策略和固定采样策略,自动更新策略收敛速度较快,并保持损失下降平缓直至模型收敛.

图4给出域混合比 λ 和域接近比 $(1-q)$ 对模型性能的影响,其中 q 表示与源域的接近程度, $(1-q)$ 表示与目标域的接近程度.由图4(a)所示实验结果可知,由于混合域逐渐向源域移动,混合比在训练过程中不断增加.由图4(b)所示实验结果,可以得出以下结论:(1)训练早期阶段,混合域更接近目标域;(2)随着域混合比的逐渐增加,混合域逐渐靠近源域,表明跨域混合策略确实有效缓解了域差距.图4(c)给出训练过程中不同域自适应策略的检测性能.图4(d)给出不同域自适应策略的检测性能.

由表5所示实验结果可知采样混合域+源域时的模型性能优于采样混合域.这是因为采样混合域数据会导致数据不足问题,而采样混合域+源域可最大程

表 4 PASCAL VOC 数据集上不同域自适应策略的消融实验结果

混合比 λ	方法	组 1			组 2			组 3		
		1	2	3	1	2	3	1	2	3
固定采样策略	图像级混合 ^[46]	37.7	39.4	43.9	20.9	25.5	29.2	27.8	31.6	38.8
	特征级混合 ^[46]	39.2	45.5	47.7	22.8	27.6	31.3	28.6	32.8	39.6
Beta 采样策略 ^[47]	跨域混合	42.6	49.1	50.8	24.9	29.4	33.1	31.1	34.5	39.2
	图像级混合 ^[46]	38.1	40.2	44.5	21.5	26.3	30.1	28.2	32.2	39.4
	特征级混合 ^[46]	39.9	46.3	48.2	23.4	28.5	32.6	29.3	33.5	40.0
自动更新策略	跨域混合	43.1	49.9	51.2	25.8	30.6	33.8	31.9	35.7	40.6
	图像级混合 ^[46]	38.8	41.1	45.2	22.3	27.1	30.9	29.5	33.1	40.2
	特征级混合 ^[46]	40.6	47.5	48.9	24.5	29.2	33.7	30.4	34.1	41.6
	跨域混合	44.6	51.7	52.7	27.6	33.7	38.1	35.6	39.3	44.3

注:加粗数据为最优结果.

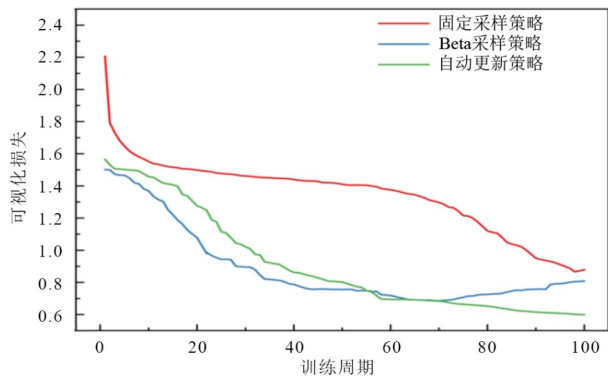


图 3 自动更新策略训练过程收敛性、稳定性与固定/随机采样策略的对比实验

度丰富目标域数据,从而提高模型的检测效果,完成源域到目标域的知识迁移.

4.3.2 双通道注意力模块的消融实验

表 6 给出 PASCAL VOC 数据集的消融实验结果,验证了卷积子模块对模型性能的影响.由表 6 所示实验

结果可得以下结论:(1)在三个目标域样本的条件下,卷积子模块分别使模型性能提高了 4.1%、4.6% 和 2.2%;(2)卷积子模块方法是有效的,且与 C-Mix-Up 策略兼容;(3)与 ResNet50 和 ResNet101 相比,使用卷积子模块的 ResNet50 能够获得更好的性能,同时参数量更少,这证明了本文的假设,卷积子模块将特征提取器后的全连接层换为卷积层,能够保护对识别对象类别至关重要的空间信息.

表 7 给出双通道注意力的消融实验结果,一方面探索双通道注意力对检测性能的影响,另一方面证明双通道注意力的优越性,其中“+”表示不同模块的组合顺序:“+”前模块的特征表征作为“+”后模块的输入,“&”表示不同模块的层级顺序:“&”前后的两个模块输入相同特征表征,再对两个模块的输出特征进行元素级相乘运算,“C-Mix+CN”表示跨域混合和卷积层.由表 7 所示实验结果可以得到以下结论:

(1)当仅引入全局最大池化(Global Max Pooling,

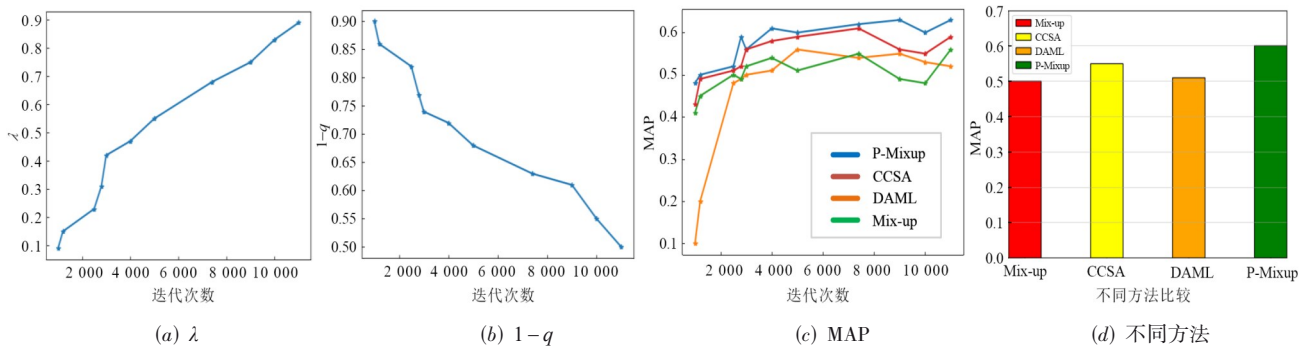


图 4 PASCAL VOC 数据集训练过程中模型超参数变化趋势及不同方法性能比较

表 5 PASCAL VOC 数据集上元训练采样域的消融实验结果

混合比 λ	方法	组 1			组 2			组 3		
		1	2	3	1	2	3	1	2	3
Beta 采样策略	采样混合域	41.3	47.8	50.1	23.3	28.8	31.1	29.8	33.5	38.2
	采样混合域+源域	43.1	49.9	51.2	25.8	30.6	33.8	31.9	35.7	40.6

表 6 PASCAL VOC 数据集上不同骨干网络的消融实验结果

C-Mix	方法	参数量/M	组 1			组 2			组 3		
			1	2	3	1	2	3	1	2	3
√	Resnet50	35.8	44.6	50.7	52.7	27.6	32.7	37.1	35.6	40.3	45.3
√	Resnet101	42.3	45.5	50.9	54.6	30.4	34.8	39.2	36.6	41.5	46.0
√	卷积子模块(Resnet50+CN)	39.9	46.2	51.2	56.8	32.6	37.1	41.7	38.3	42.8	47.5

注:加粗数据为最优结果.

表 7 PASCAL VOC 数据集上双通道注意力的消融实验结果

C-Mix+CN	方法	参数量/M	推理时间/fps	组 1			组 2			组 3		
				1	2	3	1	2	3	1	2	3
√	GMP+GIR	32.2	4.33	48.2	54.9	60.6	34.6	39.8	45.8	41.8	48.5	54.0
√	GAP	30.2	4.67	48.8	55.4	61.6	34.8	42.0	48.1	42.2	49.4	54.9
√	GAP+(GMP+GIR)	33.4	5.12	50.1	57.2	63.4	35.2	43.1	49.0	43.3	51.2	55.8
√	(GMP+GIR)&GAP	33.4	3.98	49.3	55.8	61.3	34.4	39.7	46.9	41.6	48.2	53.9
√	(GMP+GIR)+GAP	33.4	3.68	48.0	54.9	60.7	34.0	39.3	46.2	41.2	48.0	53.3
√	SENet ^[21]	30.8	4.83	47.7	54.4	60.2	33.5	38.4	45.7	40.4	47.4	52.1
√	CBAM ^[7]	31.2	4.56	46.6	53.9	59.5	31.0	36.7	43.4	39.9	45.8	50.6
√	SENet ^[21] +CBAM ^[7]	31.4	4.95	47.9	54.8	60.5	33.9	38.6	45.9	40.7	47.8	52.5
√	CBAM ^[7] +SENet ^[21]	31.0	4.83	47.6	54.5	60.2	32.2	37.7	44.6	40.1	46.8	51.6

注:加粗数据为最优结果.

GMP)+聚集感兴趣区域(General Index Register, GIR)时,模型准确性较低,这是因为单通道注意力会导致全局信息不完整.

(2)当先引入全局平均池化(Global Average Pooling, GAP),再引入全局最大池化+聚集感兴趣区域(GMP+GIR)时,模型性能达到最佳,而当调换这两者顺序后,模型性能有所降低,这是因为全局平均池化(GAP)关注公共信息,而全局最大池化+聚集感兴趣区域(GMP+GIR)容易受到极值影响,因此,在增强显著特征的基础上,计算特征映射每个通道的平均值则会产生偏差,进而影响模型性能.因此,本次实验过程中,首先使用全局平均池化(GAP)聚合全局信息,然后使用全局最大池化+聚集感兴趣区域(GMP+GIR)关注显著区域,从而提高元学习模型在少样本多源域目标检测任务上的精度.

(3)SENet 仅使用全局平均池化(GAP)生成的权重调整原始特征,从而导致模型无法关注目标域样本的显著信息.

(4)CBAM 注意力模块引入过多参数,从而导致少样本学习模型容易过拟合.

(5)本文的方法优于SENet+CBAM 和CBAM+SENet 结构. SENet 本质上类似于全局平均池化, CBAM 本质上类似于全局最大池化,本文的方法兼顾二者优势,在此基础上利用聚集感兴趣区域(GIR)进一步提取感兴趣区域特征,因此本文的方法优于上述两种结构.

从图 5 所示的热图对比结果可知,本文模型有效提取重要信息,同时抑制无关信息.这与本文假设一致:双通道注意力模块根据特征映射的信息权重评估特征映射的重要性,从而区分重要信息和无关信息.

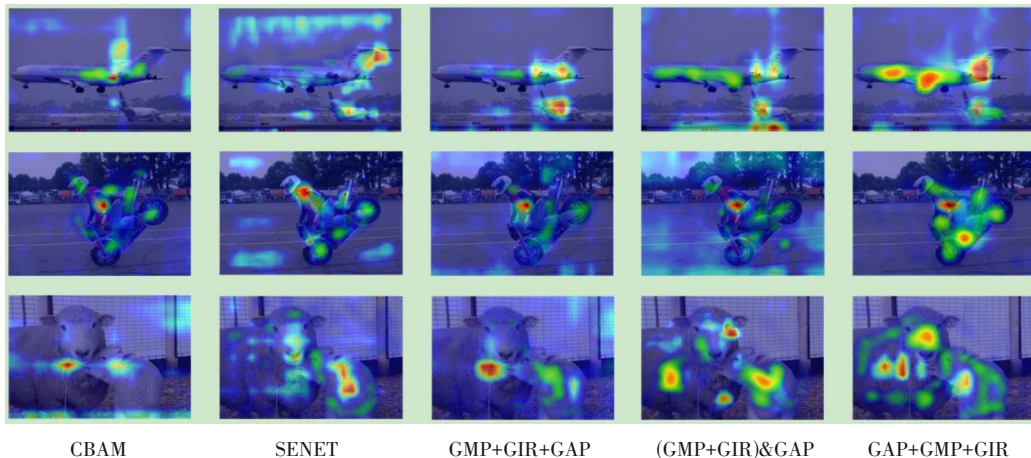


图 5 不同模型的热图

4.3.3 可视化分析

图6给出了五种不同少样本目标检测方法在MS COCO数据集上的检测结果,其中源域类别包括行人、餐桌、杯子,目标域类别包括摩托车、头盔、勺子、蛋糕、马.由图6所示结果可知,相较于其他方法,本文方法在新域样本检测结果的错检率、漏检率明显减少,同时预测的置信度显著提高;此外,即使源域类别和目标

域类别的类间差异较大时,源域类别和目标域类别的预测准确度和定位准确度基本接近,这说明不同域间知识转移不存在类别间偏差;另外,对于图像中存在遮挡的摩托车、行人等难检类别和相似度较高的勺子、牙刷等难检类别,本文方法仍然能准确识别.这说明本文方法有效缓解了域差异过大引起的知识转移困难的问题.

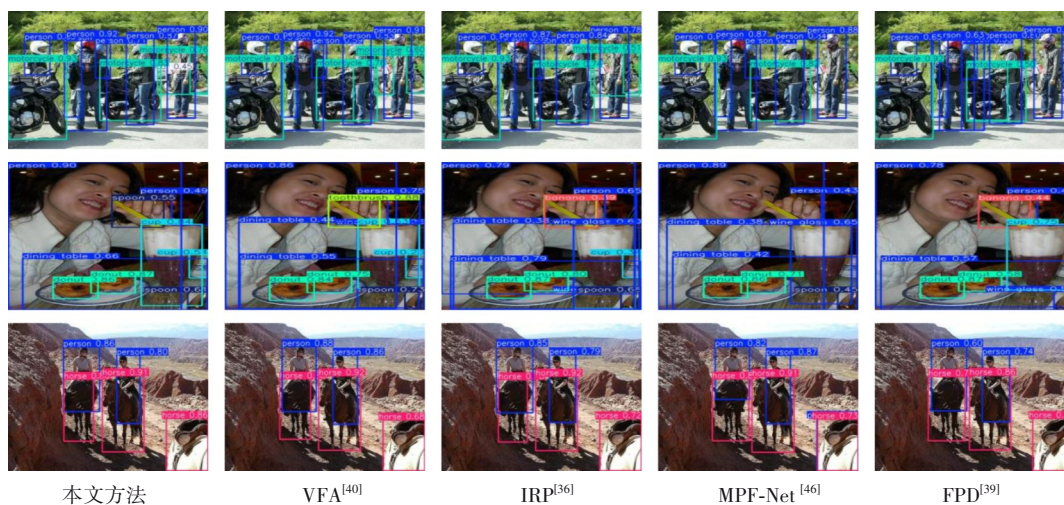


图6 不同少样本目标检测方法的可视化对比

4.3.4 泛化性分析

如表8所示,为了验证模型的泛化性能,在元训练过程中采用MS COCO2014数据集作为训练数据集,采用MS COCO2017数据集进行模型性能验证.与表2的模型检测效果对比,本文方法的泛化性能良好且优于领域内其他方法.例如,当目标域样本数量为1时,本文方法的MAP、MAP50和MAP75仅降低了0.1、0.2和0.2,DAML^[23]的MAP、MAP50和MAP75降低了0.7、0.6和0.7;当目标域样本数量为3时,本文方法的MAP、MAP50和MAP75仅降低了0.1、0.2和0.2,CCSA^[30]的MAP、MAP50和MAP75降低了0.6、0.8和0.7.

表8 不同方法在MS COCO2017数据集的实验结果

方法	1 样本			3 样本		
	MAP	MAP50	MAP75	MAP	MAP50	MAP75
FSDetView ^[43]	2.5	8.1	1.1	6.1	18.1	2.4
MPF-Net ^[44]	3.6	10.8	4.1	6.6	12.6	7.1
DRL ^[45]	4.5	10.9	7.4	8.0	16.8	8.3
DAML ^[23]	3.8	11.0	4.5	6.9	12.6	7.4
CCSA ^[30]	3.5	10.5	3.9	6.5	12.5	6.8
Ours	6.1	11.9	9.1	9.5	19.1	10.5

5 结论

本文提出了一个具有挑战性和现实性的问题,即少样本多源域目标检测问题.本文认为,目标域稀缺的

特征分布是少样本多源域目标检测的挑战,并通过元优化实现多源域到目标域的知识迁移.首先,提出了一种跨域混合策略,利用图像和特征三级混合丰富目标域数据.其次,提出了一种跨域对齐策略,通过调整跨域混合比 λ 缓解目标域和源域间的域偏差.最后,提出了一种双通道注意力机制,利用特征权重强化重要特征,进而捕获类别间差异.实验结果表明,该方法进一步提高了少样本目标检测的准确性和鲁棒性.

本文方法专注于源域数据和有限目标域数据间的知识转移,从而限制了增强数据的多样性,进而随着目标类别的增加,导致检测任务变得更加困难.未来研究方向着眼于使用多个源域数据集和目标域数据集进行增强,进一步扩大增强数据的多样性.

参考文献

- [1] YAO X X, ZHAO S C, XU P F, et al. Multi-source domain adaptation for object detection[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2022: 3253-3262.
- [2] WU C W, CAO G T, LI Y, et al. Chaos to order: A label propagation perspective on source-free domain adaptation[EB/OL]. (2023-08-14)[2025-09-30]. <https://arxiv.org/abs/2301.08413>.
- [3] YUE X Y, ZHENG Z W, REED C, et al. Multi-source few-

- shot domain adaptation[EB/OL]. (2021-09-25) [2025-09-30]. <https://arXiv.org/abs/2109.12391>.
- [4] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[M]//Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [5] LIU F, ZHANG X S, PENG Z L, et al. Integrally migrating pre-trained transformer encoder-decoders for visual object detection[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2024: 6802-6811.
- [6] 刘鑫磊, 冯林, 廖凌湘, 等. 基于元学习的图卷积网络少样本学习模型. 电子学报, 2024, 52(3): 885 - 897.
LIU X L, FENG L, LIAO L X, et al. Few-shot learning on graph convolutional network based on meta learning[J]. Acta Electronica Sinica, 2024, 52(3): 885-897. (in Chinese)
- [7] ZHANG C, CAI Y J, LIN G S, et al. DeepEMD: Few-shot image classification with differentiable earth mover's distance and structured classifiers[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 12200-12210.
- [8] YANG B Y, LIU C, LI B H, et al. Prototype mixture models for few-shot semantic segmentation[M]//Computer Vision - ECCV 2020. Cham: Springer, 2020: 763-778.
- [9] LIU B H, DING Y, JIAO J B, et al. Anti-aliasing semantic reconstruction for few-shot semantic segmentation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 9742-9751.
- [10] XU J, LIU B, XIAO Y S. A variational inference method for few-shot learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(1): 269-282.
- [11] SHAO S, XING L, WANG Y J, et al. Attention-based multi-view feature collaboration for decoupled few-shot learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(5): 2357-2369.
- [12] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[M]//Computer Vision - ECCV 2020. Cham: Springer, 2020: 213-229.
- [13] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 779-788.
- [14] TANG Y B, CAO Z Q, YANG Y Q, et al. Semi-supervised few-shot object detection via adaptive pseudo labeling[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(4): 2151-2165.
- [15] MA J W, NIU Y L, XU J C, et al. DiGeo: Discriminative geometry-aware learning for generalized few-shot object detection[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 3208-3218.
- [16] SUN Q, CHATTOPADHYAY R, PANCHANATHAN S, et al. A two-stage weighting framework for multi-source domain adaptation[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems. New York: ACM, 2011: 505-513.
- [17] MUANDET K, BALDUZZI D, SCHOLKOPF B. Domain generalization via invariant feature representation[C]//Proceedings of the 30th International Conference on Machine Learning. Cambridge: PMLR, 2013, 28(1): 10-18.
- [18] LI H L, PAN S J, WANG S Q, et al. Domain generalization with adversarial feature learning[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 5400-5409.
- [19] ZHOU K Y, YANG Y X, QIAO Y, et al. Domain generalization with mixstyle[EB/OL]. (2021-04-05) [2025-09-30]. <https://arxiv.org/abs/2104.02008>.
- [20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [21] WANG Q L, WU B G, ZHU P F, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 11531-11539.
- [22] ZHU Y H, LIU C L, JIANG S Q. Multi-attention meta learning for few-shot fine-grained image recognition[C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. New York: ACM, 2021: 1090-1096.
- [23] SHU Y, CAO Z J, WANG C Y, et al. Open domain generalization with domain-augmented meta-learning[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 9619-9628.
- [24] MAI Z J, HU G S, CHEN D X, et al. Metamixup: Learning adaptive interpolation policy of mixup with metalearning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(7): 3050-3064.
- [25] RAVI S, LAROCHELLE H. Optimization as a model for few-shot learning[C]//Proceedings of the 5th International Conference on Learning Representations. Appleton: ICLR, 2017: 67413369.
- [26] YAN X P, CHEN Z L, XU A N, et al. Meta R-CNN: Towards general solver for instance-level low-shot learning[C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2020: 9576-9585.

- [27] DOMOGUEN J K L, NAVAL P C. Dynamic model-agnostic meta-learning for incremental few-shot learning[C]//2022 26th International Conference on Pattern Recognition. Piscataway: IEEE, 2022: 4927-4933.
- [28] ZHANG G J, LUO Z P, CUI K W, et al. Meta-DETR: Image-level few-shot detection with inter-class correlation exploitation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(11): 12832-12843.
- [29] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//Proceedings of the 34th International Conference on Machine Learning. New York: ACM, 2017, 70: 1126-1135.
- [30] MOTIAN S, PICCIRILLI M, ADJEROH D A, et al. Unified deep supervised domain adaptation and generalization[C]//2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 5716-5726.
- [31] WANG X, HUANG T E, DARRELL T, et al. Frustratingly simple few-shot object detection[C]//Proceedings of the 37th International Conference on Machine Learning. New York: ACM, 2020: 9919-9928.
- [32] LI J M, ZHANG Y N, QIANG W W, et al. Disentangle and remerge: Interventional knowledge distillation for few-shot object detection from a conditional causal perspective[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(1): 1323-1333.
- [33] HAN G X, HUANG S Y, MA J W, et al. Meta faster R-CNN: Towards accurate few-shot object detection with attentive feature alignment[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(1): 780-789.
- [34] SHANGGUAN Z Y, ROSTAMI M. Improved region proposal network for enhanced few-shot object detection[J]. Neural Networks, 2024, 180: 106699.
- [35] YAN B W, LANG C B, CHENG G, et al. Understanding negative proposals in generic few-shot object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(7): 5818-5829.
- [36] WANG Z C, YANG B, YUE H N, et al. Fine-grained prototypes distillation for few-shot object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(6): 5859-5866.
- [37] HAN J M, REN Y Q, DING J, et al. Few-shot object detection via variational feature aggregation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(1): 755-763.
- [38] YIN A T, WANG Y N, MAO J X, et al. Category-contextual relation encoding network for few-shot object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(9): 8355-8367.
- [39] QIN J H, XU Y, FU Y F, et al. FSMT: Few-shot object detection via Multi-Task Decoupled[J]. Pattern Recognition Letters, 2025, 192: 8-14.
- [40] WANG B X, YU D H. Orthogonal progressive network for few-shot object detection[J]. Expert Systems with Applications, 2025, 264: 125905.
- [41] WEI Y J, LONG S W, WANG Y T. Improved few-shot object detection method based on faster R-CNN[J]. IET Image Processing, 2025, 19: e70038.
- [42] WU J Q, LEI J, TIAN H, et al. Dynamic routing and calibration for few-shot object detection[C]//ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2025: 1-5.
- [43] XIAO Y, MARLET R. Few-shot object detection and viewpoint estimation for objects in the wild[M]//Computer Vision - ECCV 2020. Cham: Springer, 2020: 192-210.
- [44] CHEN H, WANG Q, XIE K L, et al. MPF-Net: Multi-projection filtering network for few-shot object detection[J]. Applied Intelligence, 2024, 54(17): 7777-7792.
- [45] LIU W J, CAI X J, WANG C, et al. Dynamic relevance learning for few-shot object detection[J]. Signal, Image and Video Processing, 2025, 19(4): 297.
- [46] WU Y, INKPEN D, EL-ROBY A. Dual mixup regularized learning for adversarial domain adaptation[M]//Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 540-555.
- [47] ZHANG H Y, CISSE M, DAUPHIN Y N, et al. Mixup: Beyond empirical risk minimization[EB/OL]. (2018-04-27)[2025-09-20]. <https://arXiv.org/abs/1710.09412>.

作者简介



朱松豪 男,1973年生.博士.主要研究方向为图像处理、人工智能.
E-mail: zhush@nupt.edu.cn



王双丞 男,2002年生.硕士研究生.主要研究方向为模式识别与机器视觉.
E-mail: 2686404323@qq.com