

基于改进Faster R-CNN的小样本目标检测算法

李京¹, 王子怡², 巩海鑫², 熊风光^{2,3,4*}

(1. 中北大学 国际教育学院, 山西 太原 030051; 2. 中北大学 计算机科学与技术学院, 山西 太原 030051;

3. 中北大学 机器视觉与虚拟现实山西省重点实验室, 山西 太原 030051;

4. 中北大学 山西省视觉信息处理及智能机器人工程研究中心, 山西 太原 030051)

摘要: 针对小样本场景下训练样本数量稀少导致的模型可学习的信息有限、模型的性能表现受目标的尺度变化制约以及对基类知识的灾难性遗忘等问题, 遵循迁移学习方法, 提出了一种改进Faster R-CNN的小样本目标检测算法。首先, 在特征提取阶段, 加入中心坐标注意力机制, 引导模型学习目标的核心特征, 同时引入多尺度融合与可变形卷积, 使模型能够在不同尺度特征图上提取丰富的特征信息; 然后, 引入渐进式RPN, 采用两阶段回归方法, 对生成的区域建议进行逐步优化; 最后, 引入知识补偿模块, 借助解耦机制将先验知识与新知识进行解耦, 保持对基类的良好记忆。实验结果表明, 在PASCAL VOC数据集的3种新类划分中, 所提算法与主流小样本目标检测方法FSODCR相比, 在2-Shot和5-Shot设置下的nAP分别平均提升了8.37个百分点和6.63个百分点; 在MS COCO数据集中, 所提算法较FSODCR, 在10-Shot下nAP和nAP75分别提升了1.2个百分点和0.9个百分点; 在30-Shot下nAP和nAP75分别提升了1.2个百分点和0.8个百分点, 显著提升了小样本条件下目标检测的精度与稳定性。

关键词: 目标检测; 深度学习; 小样本目标检测; 迁移学习

中图分类号: TP393

文献标识码: A

doi: 10.62756/jnuc.issn.1673-3193.2025.05.0003

引用格式: 李京, 王子怡, 巩海鑫, 等. 基于改进Faster R-CNN的小样本目标检测算法[J]. 中北大学学报(自然科学版), 2026, 47(3): 263-273.

LI Jing, WANG Ziyi, GONG Haixin, et al. A few-shot object detection algorithm based on improved faster R-CNN[J]. Journal of North University of China(Natural Science Edition), 2026, 47(3): 263-273.

A Few-Shot Object Detection Algorithm Based on Improved Faster R-CNN

LI Jing¹, WANG Ziyi², GONG Haixin², XIONG Fengguang^{2,3,4*}

(1. School of International Education, North University of China, Taiyuan 030051, China;

2. School of Computer Science and Technology, North University of China, Taiyuan 030051, China;

3. Shanxi Key Laboratory of Machine Vision & Virtual Reality, North University of China, Taiyuan 030051, China;

4. Shanxi Province's Vision Information Processing and Intelligent Robot Engineering Research Center, North University of China, Taiyuan 030051, China)

Abstract: In order to solve the problems of limited information that can be learned by the model due to the scarcity of training samples in few-shot scenarios, the performance of the model is restricted by the scale change

收稿日期: 2025-05-11

基金项目: 国家自然科学基金(62272426); 山西省科技重大专项计划“揭榜挂帅”项目(202201150401021); 山西省自然科学基金(202203021212138, 202303021211153, 202203021222027); 山西省科技成果转化引导专项(202104021301055)

作者简介: 李京(1979—), 男, 讲师, 硕士, 主要从事人工智能方面的研究。

* 通信作者: 熊风光(1979—), 男, 副教授, 博士, 主要从事人工智能、虚拟现实方面的研究。E-mail: hopenxf@nuc.edu.cn。

of the object, and the catastrophic forgetting of the base class knowledge, a few-shot object detection algorithm with improved Faster R-CNN was proposed by following the transfer learning method. Firstly, in the feature extraction stage, the central coordinate attention mechanism was added to guide the model to learn the core features of the object, and multi-scale fusion and deformable convolution were introduced to enable the model to extract rich feature information on the feature map at different scales. Then, a gradual RPN was introduced, and a two-stage regression method was used to gradually optimize the generated regional suggestions. Finally, the knowledge compensation module was introduced to decouple the prior knowledge from the new knowledge with the help of the decoupling mechanism, so as to maintain a good memory of the base class. Experimental results show that, in the three novel class splits of the PASCAL VOC, compared with the mainstream small sample object detection method FSODCR, the proposed algorithm achieves an average increased in nAP by 8.37 percentage points and 6.63 percentage points in the 2-Shot and 5-Shot settings respectively; on the MS COCO, compared with FSODCR, the proposed algorithm achieves an average increase in nAP and nAP75 of 1.2 percentage points and 0.9 percentage points respectively under the 10-Shot, and nAP and nAP75 have increased by 1.2 percentage points and 0.8 percentage points under the 30-Shot, which significantly improves the accuracy and stability of object detection under few-shot conditions.

Key words: object detection; deep learning; few-shot object detection; transfer learning

0 引 言

目标检测(Object Detection)^[1]作为计算机视觉领域的重要任务,旨在对给定图像中的目标进行分类和定位,其性能直接影响人脸识别^[2-3]、安防系统^[4-5]、目标跟踪^[6]等一系列下游任务的效果。当前主流目标检测算法(无论是单阶段的SSD、YOLO,还是两阶段的Faster R-CNN)的优异性能,均依赖于大规模标注数据的训练支撑。然而在工业缺陷检测^[7]、稀有物种检测^[8]、医学检测^[9-10]等现实场景中,标注数据的获取面临显著障碍:工业缺陷样本因生产流程稳定而稀缺,稀有物种因种群数量少而难以采集,医学病症因病例隐私与罕见性而标注困难。这种“数据匮乏”的现实矛盾,严重限制了目标检测技术在高价值领域的落地,使得小样本目标检测(Few-Shot Object Detection, FSOD)成为亟待突破的关键课题。

近年来,学者围绕基于深度学习的小样本目标检测展开了广泛且深入的研究。该领域的核心目标是让模型借鉴人类“举一反三”的学习能力,在仅依赖少量标注样本的前提下,高效完成对新类别的识别与定位任务,从而突破传统目标检测算法对大规模标注数据依赖的瓶颈。许多研究将小样本学习的思想与深度学习中的目标检测算法进行结合,如单阶段目标检测算法SSD和YOLO系列,以及两阶段目标检测算法Faster R-CNN等。Wang等^[11]提出的

两阶段微调方法(Two-Stage Fine-tuning Approach, TFA)采用两阶段训练策略,简洁高效,性能超过大部分基于元学习的方法,成为当前最有效的迁移学习基线模型。Sun等^[12]提出的FSCE在TFA的基础上引入对比学习思想,通过增加对比分支和区域建议的数量来改善模型的性能,还将位置相近的区域建议作为正样本,较远的区域作为负样本,使类内距离更近,类间距离更远。Qiao等^[13]提出了DeFCRN,旨在解决区域建议网络在前景识别方面表现不佳的问题。在微调过程中,没有直接冻结参数,而是通过调整反向传播过程中的梯度系数,达到解耦RPN和RCNN的效果,从而减轻了两个子网络在训练中的冲突。Kristan等^[14]提出了一种全新的、整合检测与分割的架构用于小样本计数。与传统将检测和分割独立处理的方式不同,这种统一架构打破了常规,让检测和分割在小样本计数任务中协同工作,可能提升小样本条件下目标识别和数量统计的效率与精度。Zhao等^[15]提出了一种创新的合成孔径雷达(SAR)图像小样本目标检测算法,通过多种模块设计解决了SAR图像检测中的难题。尽管上述方法在小样本目标检测任务中已取得显著进展,但仍面临以下挑战:训练样本的稀缺性导致模型难以充分学习目标特征,严重影响检测性能;传统的特征提取方法对多尺度目标的适应性较差,在小样本场景下泛化能力受限;基于迁移学习的方法在学习新类时普遍存在灾难性遗忘问题,导致基类知识丢失,这些局限严重制约着小样本目标检测模型的

性能表现。为解决上述关键挑战,本文在 Faster R-CNN 的基础上进行改进,主要工作如下:

- 1) 在特征提取阶段加入中心坐标注意力机制(Central Coordinate Attention Module, CCAM),通过像素级注意力特征融合策略以及中心重要性加权的方法,将模型的注意力引导至中心及周围区域,使模型能更好地聚焦于关键特征。同时,加入可变形卷积特征金字塔(DFPN),将可变形卷积与特征金字塔进行结合,使模型能够灵活地适应目标的形态变化,从不同尺度的特征图上提取丰富的特征信息。
- 2) 对 RPN 进行改进,采用两阶段回归策略,实现对区域建议的进一步优化。
- 3) 在 ROI 特征提取器中引入知识补偿模块,采用解耦机制,将学习到的先验知识与新知识进行解耦,缓解新类训练过程中对基类知识的灾难性遗忘问题,使模型能够在学习新类知识的同

时,保留已有的基类知识。

1 相关研究

1.1 Faster R-CNN 结构

Faster R-CNN 是两阶段目标检测框架的代表性算法,主要由 4 个核心模块构成:特征提取网络 Backbone、区域建议网络 RPN、ROI 特征提取器以及检测器,如图 1 所示。首先,主干网络(如 VGG-16 或 ResNet)从输入图像提取特征,得到相应的特征图;随后,RPN 通过预先定义的锚框在特征图中识别可能包含目标对象的候选框,并对候选框进行前景和背景分类与位置回归;为适应不同大小的候选框,ROI 池化将其转换为固定大小的特征图,之后送入全连接层进行分类和边界框回归;最终,通过非极大值抑制去除重叠度较高的冗余框,得到检测结果。

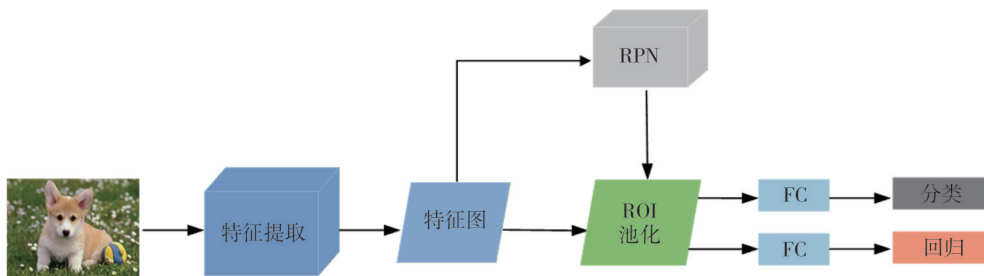


图 1 Faster R-CNN 结构

Fig. 1 Structure of Faster R-CNN

1.2 两阶段微调策略

在小样本目标检测任务中,训练集由基类数

据集和新类数据集组成。基类数据集中的类别为基类(C_{base}),新类数据集中的类别为新类(C_{novel}),且基类与新类之间无交集。

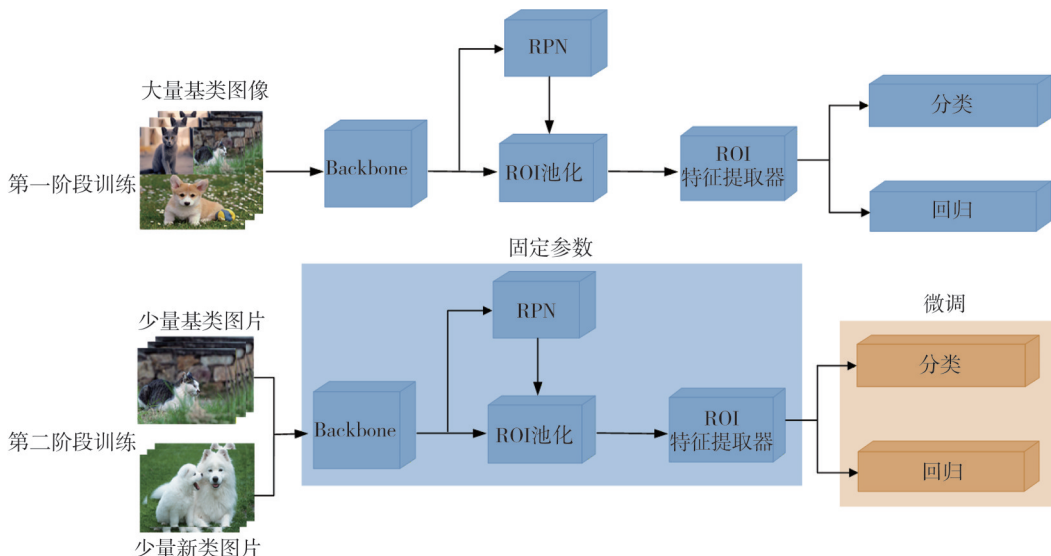


图 2 两阶段微调训练过程

Fig. 2 Two-stage fine-tuning training process

基类数据集中每个类别都含有大量的训练样本,新类数据集中每个类别只有少量数据。通用的迁移学习范式首先使用基类数据集在初始模型上进行训练,得到基础模型。接着,在由少量样本构成的数据集上对基础模型进行微调,得到最终模型,该过程表示为

$$M_{\text{init}} \xrightarrow{D_{\text{base}}} M_{\text{base}} \xrightarrow{D_{\text{fine-tune}}} M_{\text{final}} \quad (1)$$

基于两阶段微调的方法 TFA 是最早采用迁移学习范式的方法,该方法使用 Faster R-CNN 为基础模型,分为两个阶段,如图 2 所示。在第一阶段,利用大量的基类样本对 Faster R-CNN 进行训练,得到基础模型。在第二阶段,冻结除检测头

以外的其他网络参数,仅利用少量样本对分类器和回归器进行微调。该方法简单高效,能够有效地将从基类中学到的知识迁移到新类中,从而在小样本的情况下显著提升模型的检测性能。

2 改进 Faster R-CNN 的算法研究

2.1 模型总体结构

本文模型结构如图 3 所示,在 Faster R-CNN 的基础上加入中心坐标注意力机制、可变形卷积特征金字塔、渐进式 RPN、知识补偿模块。首先,输入图像进入 Faster R-CNN 的主干网络,ResNet101 对图像进行初步的特征提取,生成一系列特征图。

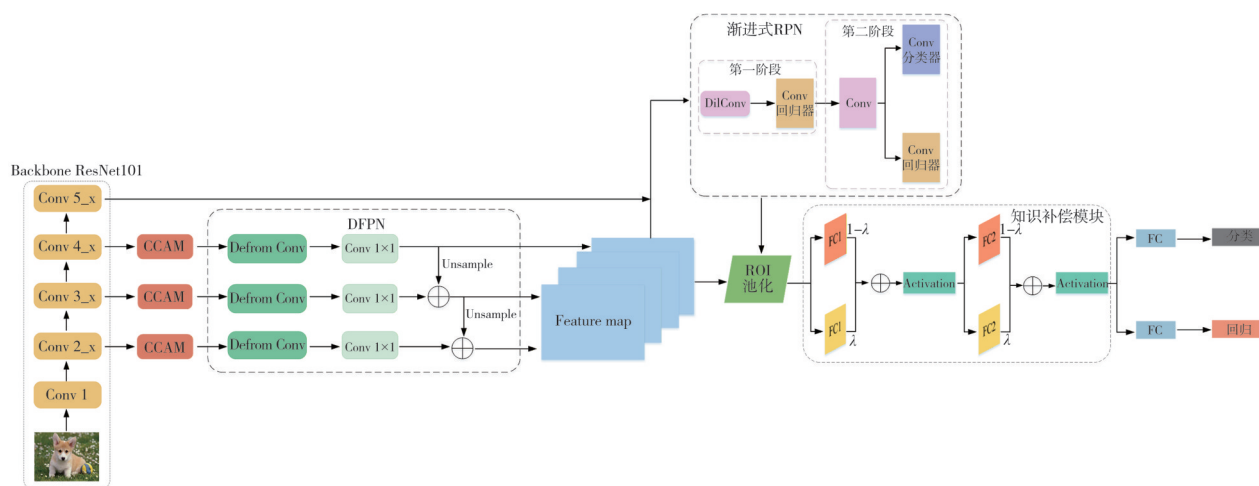


图3 模型总体结构图

Fig. 3 Overall structure diagram of the model

这些特征图随后被送入中心坐标注意力机制 CCAM 中,通过加权增强中心区域的重要信息。接着,进入可变形卷积特征金字塔 DFPN,通过多尺度特征提取和可变形卷积操作,进一步捕捉不同尺寸和形状的特征。得到的多尺度特征图进入渐进式 RPN,通过两阶段回归生成候选区域。在完成区域建议的生成后,经过 ROI 池化操作,将生成的区域建议尺寸进行统一。随后,通过引入的知识补偿模块,利用解耦机制减少模型对基类知识的遗忘。最后,送入检测器进行分类和边界框回归,得到最终的检测结果。

2.2 中心坐标注意力机制

在小样本目标检测中,由于训练样本有限,模型容易过拟合且难以提取到具有代表性的特征。因此,通过引入注意力机制,能够帮助模型聚焦于关键区域,避免无关区域的干扰,使模型

可以更有效地从有限的样本中学习重要特征,提高特征表达的质量和准确性。现有的注意力机制,如 SE、CBAM 以及 CA 均强调不同通道、不同行或不同列的特征重要性,而忽略了局部区域像素级特征的重要性。CCAM 中心坐标注意力通过对 CA 进行改进,引入了像素级注意力特征融合策略以及中心重要性加权机制,其结构如图 4 所示。具体而言,像素级注意力特征融合策略将输入特征图与加权系数矩阵进行融合,突出特征图中的重要像素,使模型能够更有效地识别核心区域,解决了不同像素特征对目标分类结果存在影响的问题。中心重要性加权机制则通过给予中心位置更高的权重,突出中心特征,有助于区分中心与边界区域的特征信息,减少边界处的噪声干扰。

为获取输入特征图在不同位置的远程依赖关系,CCAM 首先对输入特征图分别在水平和垂直

方向上进行全局平均池化操作,得到在宽度和高度两个方向上的特征表示。具体为

$$f_i^h = \frac{1}{W} \sum_{0 \leq j \leq W} x_{i,j}, \quad (2)$$

$$f_j^v = \frac{1}{H} \sum_{0 \leq i \leq H} x_{i,j}. \quad (3)$$

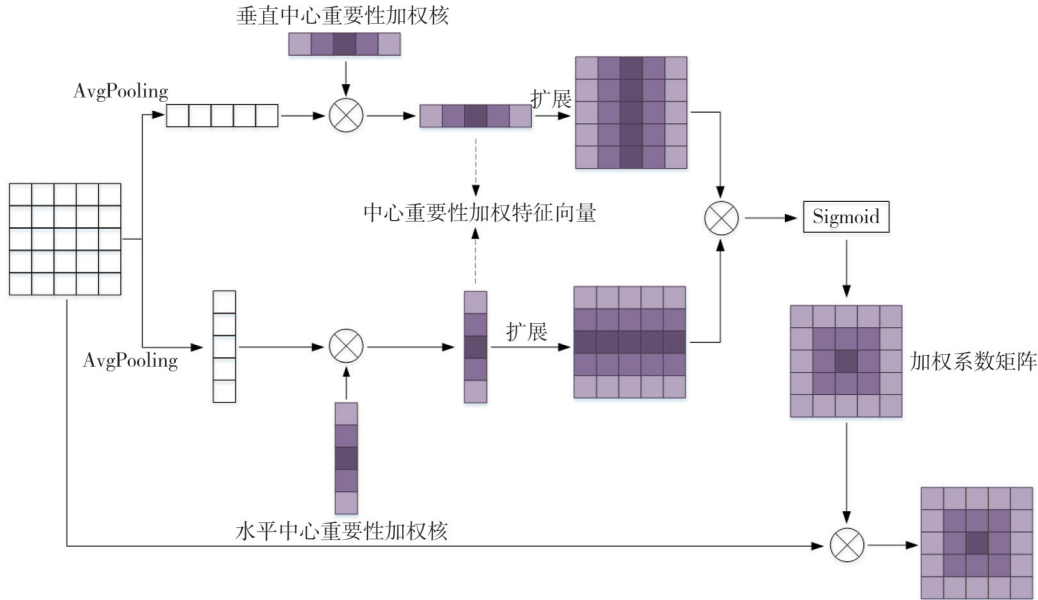


图4 CCAM的网络结构

Fig. 4 Network architecture of CCAM

接着,将加权后的特征向量与原始特征向量进行融合,得到中心重要性加权特征向量。

$$g_v(i,j) = f_j^v \odot x_v(i,j), \quad (5)$$

$$g_h(i,j) = f_i^h \odot x_h(i,j), \quad (6)$$

式中: $x_v(i,j)$ 和 $x_h(i,j)$ 分别为垂直中心重要性加权值和水平中心重要性加权值; $g_v(i,j)$ 和 $g_h(i,j)$ 分别为融合后的垂直中心重要性加权特征向量和水平中心重要性加权特征向量; $\odot$ 为元素相乘。

此时,由于上述步骤得到的特征向量大小与原始输入特征图的大小不同,需要对这些特征向量进行扩展,以便与输入特征图的大小一致。如图5所示,将垂直方向的特征向量进行行扩展,水平方向的特征向量则进行列扩展,从而得到与输入特征图大小相同的扩展特征图。

再通过元素加法操作将垂直和水平方向的扩展特征向量进一步融合。

$$G(i,j) = \text{Add}_{0 \leq i \leq H, 0 \leq j \leq W} (G_v(i,j), G_h(i,j)). \quad (7)$$

接着,使用sigmoid函数对融合后的结果进行处理,计算得到加权系数矩阵,该矩阵的值介于0~1之间,用于表示每个位置的权重。

随后,对池化后的水平和垂直方向的向量进行中心重要性加权,突出输入特征图的中心特征信息。

$$x(i,j) = \exp\left(-\frac{(i-i^*)^2 + (j-j^*)^2}{2\sigma^2}\right), \quad (4)$$

式中: $x(i,j)$ 为中心重要性加权值; $x(i^*,j^*)$ 为中心位置坐标; σ 为权值参数,设为5。

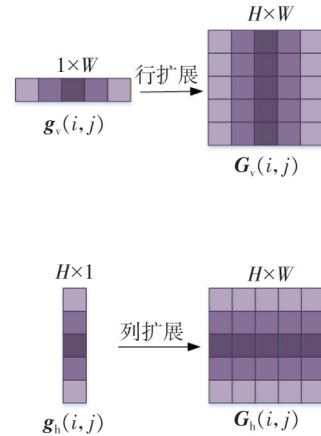


图5 扩展过程

Fig. 5 Expansion process

$$W(i,j) = \text{sigmoid}(G(i,j)). \quad (8)$$

最后,将原始输入特征图与加权系数矩阵进行融合,得到最终的输出特征图。

$$Y(i,j) = X(i,j) \odot W(i,j). \quad (9)$$

2.3 可变形卷积特征金字塔

由于目标尺度具有多样性,模型在处理多尺度目标时面临较大困难,尤其是在小样本目标检

测中。特征金字塔(Feature Pyramid Networks, FPN)作为一种经典的多尺度特征融合方法,通过对网络中不同层次特征的融合,能够增强模型对不同尺度目标的感知能力,提升其在多尺度场景中的表现。然而,传统的FPN在面对具有明显尺度变化的目标时,存在一定的局限性。特别是在目标形态发生变化或尺寸较小的情况下,FPN的固定卷积难以适应目标的形态和尺度差异,导致模型可能无法充分挖掘目标的细节特征,从而影响检测精度。为克服这一问题,本文提出一种改进的多尺度特征融合方法—可变形卷积特征金字塔(DFPN),如图6所示。

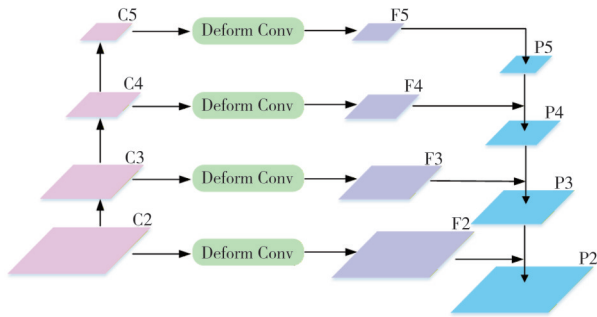


图6 DFPN的结构

Fig. 6 Architecture of DFPN

首先,将经过主干网络输出的不同尺度的特征图C2、C3、C4以及C5送入可变形卷积进行处理。然后,与传统的FPN类似,对通过可变形卷积得到的新特征图进行 1×1 卷积操作,调整通道数,生成F2、F3、F4、F5特征图。接着,对高层特征图进行上采样,再将上采样后的高层特征图与对应的低层特征图进行逐层相加。最后,得到多尺度融合后的特征图P2、P3、P4以及P5。

2.4 渐进式RPN

Meta-DETR^[16]中表明,RPN提取的区域建议的准确性,在很大程度上影响着小样本目标检测的性能表现。由于后续检测头中的分类和回归任务高度依赖RPN提出的区域建议,当RPN生成的区域建议质量较低时,低质量的区域建议可能会给分类和回归任务带来困难,最终导致模型难以准确检测出新类。

渐进式RPN采用从粗到细的两步回归策略,具体流程如图7所示。在第一个回归阶段,渐进式RPN首先将输入的特征图送入空洞卷积层提取粗略特征。经过空洞卷积处理后,特征图被进一步输入到回归器中。通过这一阶段,渐进式RPN

生成了初步的候选框。在第二个回归阶段,网络将重点关注第一阶段生成的粗略锚框。首先,对这些粗略锚框进行进一步的卷积操作。随后,把经过增强处理的特征图输入到第二阶段的分类器和回归器中。通过这两个阶段的工作,渐进式RPN能生成更加准确可靠的区域建议,为后续的检测任务提供更高质量的输入。

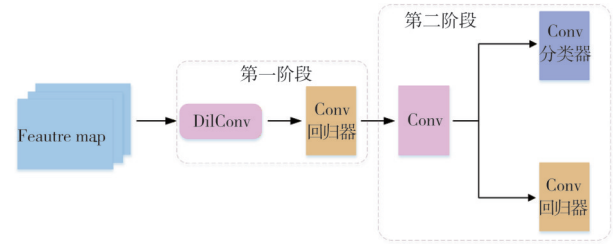


图7 渐进式RPN结构

Fig. 7 Network architecture of progressive RPN

渐进式RPN的损失函数为

$$L_{\text{rpn}} = L_{\text{reg}}^{s_1} + L_{\text{reg}}^{s_2} + L_{\text{cls}}^{s_2}, \quad (10)$$

式中: $L_{\text{reg}}^{s_1}$ 和 $L_{\text{reg}}^{s_2}$ 分别为第一阶段和第二阶段的边界框回归损失; $L_{\text{cls}}^{s_2}$ 为第二阶段的分类损失。

第一阶段和第二阶段的回归损失均采用IOU损失,表示为

$$L_{\text{reg}}^{s_i} = \frac{\lambda_{\text{reg}}}{N_{\text{pos}}^{s_i}} \sum_{j=0}^{N_{\text{pos}}^{s_i}} -\ln(\text{IOU}(B_j^s, B_j^{gt})), i \in \{1, 2\}, \quad (11)$$

式中: i 代表第一、二个回归阶段; $N_{\text{pos}}^{s_i}$ 指在第 i 个回归阶段中被预测为正样本框的预测框数量; B_j^s 是第 i 个回归阶段的第 j 个预测框; B_j^{gt} 指与第 i 个回归阶段的第 j 个正样本预测框对应的真实目标框; λ_{reg} 为平衡分类损失和回归损失之间的平衡因子,设置为10。

第二个回归阶段的分类损失为交叉熵损失

$$L_{\text{cls}}^{s_2} = \frac{1}{N} \sum_{i=0}^N -(y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)), \quad (12)$$

式中: N 为采样框的数量; p_i 为第 i 个采样框,属于前景的预测概率; y_i 为第 i 个采样框的真实标签。

2.5 知识补偿模块

在小样本迁移学习的目标检测方法中,普遍存在模型对先验知识即基类知识的遗忘问题。当模型首先利用大量的基类数据进行训练,学习到丰富的知识后,在新类数据上进行微调以学习新类知识时,模型会迅速地遗忘之前在基类学习中

所获取的知识。这种现象严重影响了模型的性能,因此需要在网络中引入知识补偿模块。

知识补偿模块通过将先验知识与新知识进行解耦,使模型在学习新类知识的同时,依然能够保持对基类知识的良好记忆。具体实现过程如图 8 所示,通过在 ROI 池化后引入解耦机制,使用并行分支和融合分支替代原始的单一分支结构,实现对先验知识和新知识的解耦。具体而言,针对 ROI 特征提取器中的两个全连接层,为每个全连接层(FC_i)额外添加一个带有预训练参数的并行全连接层,这些预训练参数来自基类训练阶段,包含模型在基类学习中积累的先验知识。在微调阶段,将新添加的分支进行冻结,以此保护已学习到的先验知识,使其在学习新类知识的过程中不会被轻易修改或覆盖。另一分支则保持可训练状态,用于学习新类知识。此外,将两个并行分支的全连接层进行融合,ROI 特征提取器可以同时兼顾来自冻结层的先验知识和来自可训练层的新知识。

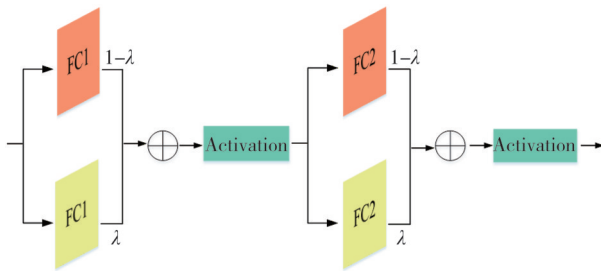


图 8 知识补偿模块网络结构

Fig. 8 Network architecture of the knowledge augmentation module

知识补偿模块的计算公式为

$$F_{out} = \delta(\lambda(W_{FC}^{frozen}FC_i + b_{FC}^{frozen}) + (1-\lambda) \cdot (W_{FC}FC_i + b_{FC})), \quad (13)$$

式中: δ 为非线性激活函数; F_{out} 和 FC_i 分别为知识补偿模块的输出和输入; λ (本文设为 0.5) 为平衡先验知识和新类知识的系数; W_{FC}^{frozen} 和 b_{FC}^{frozen} 分别为冻结分支的权重参数和偏差; W_{FC} 和 b_{FC} 分别为训练分支的权重参数和偏差。

3 实验

3.1 数据集与实验设置

PASCAL VOC 包含 20 个类别,分为 15 个基础类别和 5 个新类别。对于新类别的 K-Shot 设定, K 值取 1, 2, 3, 5, 10。本文遵循现有文献的划分方法,采用 3 种不同的新类别划分方式,分别

为 Split1: {bird, bus, cow, motorbike, sofa}、Split2: {aeroplane, bottle, cow, horse, sofa} 和 Split3: {boat, cat, motorbike, sheep, sofa}。

MS COCO 涵盖 80 个不同的常见类别,选取 MS COCO 数据集与 PASCAL VOC 数据集共有的 20 个类别为新类别,其余 60 个类别则作为基础类别。由于 MS COCO 数据集具有丰富的类别和复杂的场景信息,目标检测的难度较大,尤其是在数据量较少的情况下,因此 K 值取 10 和 30。

本文以 Faster R-CNN 为基础模型,使用在 ImageNet 上预训练的 ResNet101 为主干网络。在微调阶段中,使用 FSOD 设定。实验采用 Pytorch1.7.1 深度学习框架,Python 版本为 3.9。在模型训练阶段,使用随机梯度下降(SGD)优化器调整网络参数, batch size 的大小为 4, 动量大小为 0.9。

3.2 对比实验

将本文提出的算法与现有的小样本目标检测方法 TFA w/fc^[11]、TFA w/cos^[11]、FSCE^[12]、LSC^[14]、FSODCR^[15]、FSRW^[17]、Meta RCNN^[18]、FSDetView^[19]、DCNet^[20]、LSTD^[21]进行对比。

对于 PASCAL VOC 数据集,分别测试了不同方法在新类 Split1、Split2、Split3 上的新类平均精度 (novel Average Precision, nAP), 实验结果如表 1~表 3 所示。

表 1 在 PASCAL VOC 数据集 Split1 新类上的 nAP 结果对比

Tab. 1 Comparison of nAP results on novel classes of Split1 on PASCAL VOC

方法	nAP/%				
	1-Shot	2-Shot	3-Shot	5-Shot	10-Shot
TFA w/fc ^[11]	36.8	29.1	43.6	55.7	57.0
TFA w/cos ^[11]	39.8	36.1	44.7	55.7	56.0
FSCE ^[12]	44.2	43.8	51.4	61.9	63.4
LSC ^[14]	31.5	35.2	42.8	50.3	58.7
FSODCR ^[15]	44.7	38.5	45.1	52.6	60.3
FSRW ^[17]	14.2	23.6	29.8	36.5	35.6
Meta RCNN ^[18]	19.9	25.5	35.0	45.7	51.5
FSDetView ^[19]	24.2	35.3	42.2	49.1	57.4
DCNet ^[20]	33.9	37.4	43.7	51.1	59.6
LSTD ^[21]	8.2	11.0	12.4	29.1	38.5
本文方法	41.5	46.7	51.7	57.2	64.1

由表中实验结果可知,本文方法在 3 种不同的划分方式下均取得了较好的效果。在 Split1 新类上的结果对比中,本文算法在 2-Shot、3-Shot 和 10-Shot 下的检测精度达到较高水平。然而,在 1-Shot 的检测精度低于 FSCE 和 FSODCR,这主要是由于在

Split1的划分中,模型对于motorbike、cow、bus这类较大尺寸物体的检测效果表现良好,但对于bird这类目标较小的物体,以及sofa这类受遮挡较为严重的物体,模型在仅有一张训练样本的情况下难以提取到足够的有效特征,从而导致1-Shot的检测精度较低。在Split2新类上的结果对比中,本文算法在全样本量下表现较优,且在不同设置下均有一定优势,仅1-Shot的检测精度低于FSODCR;在Split3的2-Shot~5-Shot均高于对比方法,虽然在1-Shot略低于FSODCR、在10-Shot以略低于LSC,但综合所有Shot设置下的表现,本文方法具有较好的稳定性和整体精度,这也证明了本文算法在极端数据稀缺条件下的有效性。

表 2 在 PASCAL VOC 数据集 Split2 新类上的 nAP 结果对比
Tab. 2 Comparison of nAP results on novel classes of Split2 on PASCAL VOC

方法	nAP/%				
	1-Shot	2-Shot	3-Shot	5-Shot	10-Shot
TFA w/fc ^[11]	18.2	29.0	33.4	35.5	39.0
TFA w/cos ^[11]	23.5	26.9	34.1	35.1	39.1
FSCE ^[12]	27.3	29.5	43.5	44.2	50.2
LSC ^[14]	20.8	25.3	32.7	38.6	46.1
FSODCR ^[15]	32.6	27.4	35.9	41.8	49.5
FSRW ^[17]	12.3	19.6	25.1	31.4	29.8
Meta RCNN ^[18]	10.4	19.4	29.6	34.8	45.4
FSDetView ^[19]	21.6	24.6	31.9	37.0	45.7
DCNet ^[20]	23.2	24.8	30.6	36.7	46.6
LSTD ^[21]	11.4	3.8	5.0	15.7	31.0
本文方法	32.4	38.2	44.6	50.3	55.9

表 3 在 PASCAL VOC 数据集 Split3 新类上的 nAP 结果对比
Tab. 3 Comparison of nAP results on novel classes of Split3 on PASCAL VOC

方法	nAP/%				
	1-Shot	2-Shot	3-Shot	5-Shot	10-Shot
TFA w/fc ^[11]	27.7	33.6	42.5	48.7	50.2
TFA w/cos ^[11]	30.8	34.8	42.8	49.5	49.8
FSCE ^[12]	37.2	41.9	47.5	54.6	58.5
LSC ^[14]	39.5	33.7	40.2	47.3	58.8
FSODCR ^[15]	31.6	36.4	42.1	49.1	54.3
FSRW ^[17]	12.5	21.3	26.8	33.8	31.0
Meta RCNN ^[18]	14.3	18.2	27.5	41.2	48.1
FSDetView ^[19]	21.2	30.0	37.2	43.8	49.6
DCNet ^[20]	32.3	34.9	39.7	42.6	50.7
LSTD ^[21]	12.6	8.5	15.0	27.3	36.3
本文方法	35.6	42.5	48.1	55.9	57.8

同时,所提算法在PASCAL VOC数据集的3种新类划分中,与主流小样本目标检测方法FSODCR相比:2-Shot设置下,在新类上的nAP平均提升8.37百分点,其中,Split1、Split2和Split3在2-Shot设置下的nAP分别提升8.2百分点,10.8百分点和

6.1百分点;5-Shot设置下在新类上的nAP平均提升6.63百分点,其中,Split1、Split2和Split3在5-Shot的nAP分别提升4.6百分点,8.5百分点和6.8百分点。在具体场景,在Split1新类中,本文算法在3-Shot和10-Shot设置下的nAP优于FSODCR,Split2全Shot设置下性能在各项指标上均高于对比方法,Split3的2-Shot至5-Shot设置下nAP高于其他对比方法。

对于MS COCO数据集,在10-Shot和30-Shot上评估算法的检测性能,表4展示了在新类上的nAP和nAP75(IoU为0.75下的nAP),由于MS COCO数据集复杂度较高,类别众多,且场景也更为复杂多样,这使模型性能提升面临着相当大的挑战。从表中结果可以看出,在综合指标nAP上,本文算法在10-Shot和30-Shot设置下仅低于LSC;但是,在精度nAP75上,本文算法在10-Shot和30-Shot设置下均高于其他对比方法。

表 4 在 MS COCO 数据集新类上的 nAP 结果对比
Tab. 4 Comparison of nAP results on novel classes of the MS COCO %

方法	10-Shot		30-Shot	
	nAP	nAP75	nAP	nAP75
TFA ^[11]	10.0	9.3	13.7	13.4
FSCE ^[12]	11.9	10.5	16.4	16.2
LSC ^[14]	12.5	10.2	17.3	15.8
FSODCR ^[15]	11.2	9.9	15.7	15.5
MetaDet ^[16]	7.1	6.1	11.3	8.1
FSRW ^[17]	5.6	4.6	9.1	7.6
Meta RCNN ^[18]	8.7	6.6	12.4	10.8
本文方法	12.4	10.8	16.9	16.3

同时,在MS COCO数据集上,本文方法较FSODCR,在10-Shot设置下的nAP提升1.2百分点,nAP75提升0.9百分点;30-Shot设置下nAP提升1.2百分点,nAP75提升0.9百分点,且10-Shot、30-Shot的nAP75均为最优。上述结果验证了所提算法在小样本条件下对检测精度的显著提升,性能优于FSODCR。

本文所提方法能取得较好的检测精度,主要得益于一系列改进措施:1)中心坐标注意力机制引导模型聚焦关键特征,使模型在小样本下能更好地提取有效信息;2)可变形卷积特征金字塔结合可变形卷积与特征金字塔,增强了模型对不同尺度和形态目标的特征捕捉能力;3)渐进式RPN通过两步回归优化区域建议,为后续检测提供高质量输入;4)知识补偿模块解耦先验知识与新知

识,防止模型在学习新类时遗忘基类知识,从而综合提升了模型在 MS COCO 数据集新类检测中的性能,优于多数对比方法。

3.3 消融实验

为验证本文所提中心坐标注意力机制(CCAM)、可变形卷积特征金字塔(DFPN)、渐进

式 RPN 及知识补偿模块的有效性,在 PASCAL VOC 数据集的 Split1 新类上设计了消融实验。实验以原始 Faster R-CNN 为基础模型,通过逐步叠加 CCAM、CCAM+DFPN、CCAM+DFPN+渐进式 RPN、全部模块,对比不同组合在 1-Shot、2-Shot、3-Shot、5-Shot 和 10-Shot 设置下的 nAP,结果如表 5 所示。

表 5 在 PASCAL VOC 数据集 Split1 新类上的消融实验结果

Tab. 5 Ablation experiment results on novel classes of Split1 on PASCAL VOC

CCAM	DFPN	渐进式 RPN	知识补偿	nAP/%				
				1-Shot	2-Shot	3-Shot	5-Shot	10-Shot
×	×	×	×	30.1	35.8	37.6	45.5	52.0
✓	×	×	×	32.1	37.6	41.3	47.4	54.1
✓	✓	×	×	36.2	41.8	46.3	52.0	59.8
✓	✓	✓	×	37.5	43.6	47.2	54.7	60.3
✓	✓	✓	✓	41.5	46.7	51.7	57.2	64.1

分析表 5 中数据可知,各模块对模型性能的提升呈现显著的累积效应:

1) 基线模型:在 1-Shot 至 10-Shot 设置下的 nAP 值最低,性能表现为所有组合中最差,验证了小样本场景下原始 Faster R-CNN 的局限性。

2) CCAM:相较于基线模型,1-Shot~10-Shot 的 nAP 分别提升 2.0 百分点,1.8 百分点,3.7 百分点,1.9 百分点和 2.1 百分点,表明中心坐标注意力机制通过聚焦目标核心区域特征,有效增强了模型对关键信息的捕捉能力,尤其在 3-Shot 设置下提升最为明显。

3) CCAM+DFPN:在 CCAM 基础上进一步引入可变形卷积特征金字塔后,性能实现第二次跃升。1-Shot 至 10-Shot 的 nAP 较上一阶段分别提升 4.1 百分点,4.2 百分点,5.0 百分点,4.6 百分点和 5.7 百分点,其中 10-Shot 设置下提升幅度最大,证明多尺度特征融合与可变形卷积的结合,显著改善了模型对不同尺度、形态目标的适应性,这对小样本场景下的特征提取至关重要。

4) CCAM+DFPN+渐进式 RPN:加入两阶段回归优化的区域建议网络后,nAP 在各 Shot 设置下继续提升 1.3 百分点~2.7 百分点,其中 5-Shot 设置下提升 2.7 百分点,说明渐进式 RPN 生成的高质量区域建议为后续检测任务提供了更优输入,减少了低质量候选框对分类与回归的干扰。

5) 全部模块(CCAM+DFPN+渐进式 RPN+知识补偿):最终组合在所有 Shot 设置下均达到最优性能,1-Shot 至 10-Shot 的 nAP 分别

为 41.5%,46.7%,51.7%,57.2% 和 64.1%,较上一阶段提升 2.5 百分点~4.5 百分点,且较基线模型的最大提升达 14.1 百分点(3-Shot 设置)。这表明知识补偿模块通过解耦先验知识与新知识,有效缓解了基类知识遗忘问题,使模型在学习新类特征时仍能保持对已有知识的记忆,从而实现性能的最终突破。

综上,消融实验结果充分证明:本文所提的 4 个模块并非冗余设计,而是从特征提取、区域建议优化、知识保留 3 个关键环节形成互补,共同提升了小样本目标检测的精度与稳定性。其中,DFPN 在多尺度特征融合方面贡献最为突出(平均提升 4.72 百分点),知识补偿模块在缓解灾难性遗忘方面效果显著(平均提升 3.58 百分点),而 CCAM 与渐进式 RPN 则分别在核心特征聚焦与候选框质量优化上发挥了不可替代的作用。

3.4 可视化分析

图 9 和图 10 分别为 Faster R-CNN 与本文提出算法在 PASCAL VOC 数据集上的可视化检测结果对比。

1) 图 9 第一行中,Faster R-CNN 存在明显漏检,未能识别出“cow”和“sofa”;而图 10 第一行显示,本文算法不仅成功检测出这些目标,还精准识别出第三张图像中仅部分显露的“sofa”;

2) 图 9 第二行中,Faster R-CNN 既漏检了“sofa”,又将“cow”错误分类;图 10 第二行则表明,本文算法可有效检出遮挡严重的“sofa”,并正确识别“cow”;

图9第三行中, Faster R-CNN将“sofa”误判为“bus”;图10第三行中, 本文算法对该类目标实现了精准检测。

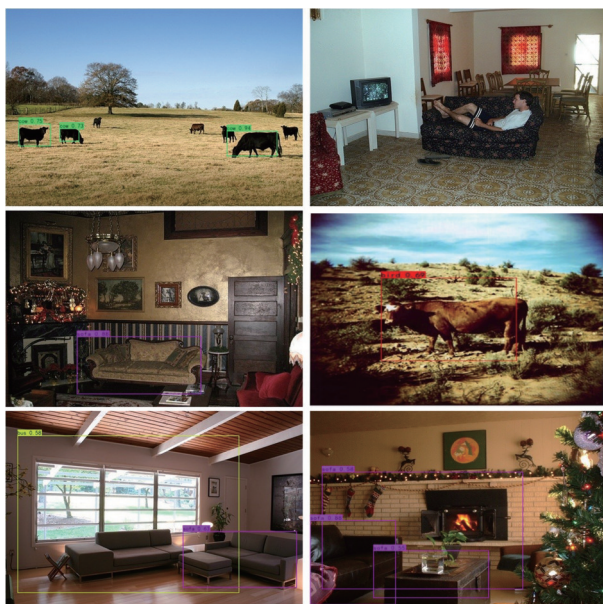


图9 Faster R-CNN检测结果可视化

Fig. 9 Visualization of detection results based on Faster R-CNN

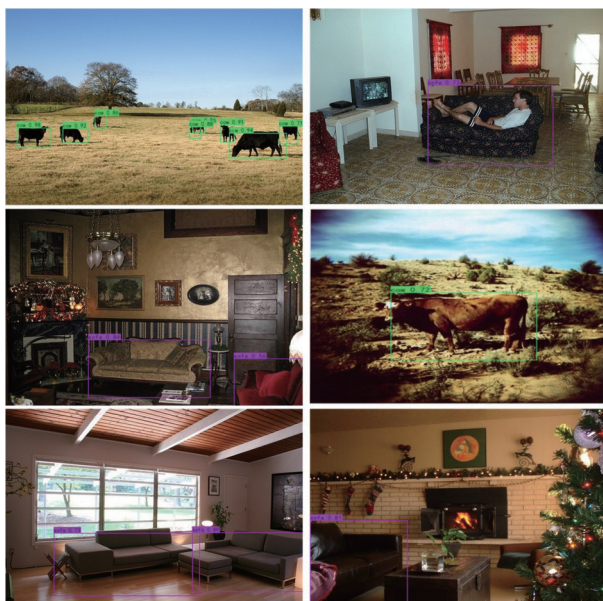


图10 本文算法检测结果的可视化

Fig. 10 Visualization of detection results of the algorithm in this paper

综上, 通过图9与图10的对比可见, 本文算法能有效减少漏检与误检现象, 验证了改进策略对检测鲁棒性的增强作用。

4 结论

针对小样本目标检测场景中存在的训练样本稀少导致特征学习不充分、目标尺度变化制约模型性

能及对基类知识存在灾难性遗忘等关键问题, 本文遵循迁移学习微调范式, 提出了一种基于改进Faster R-CNN的小样本目标检测算法。该算法通过多模块协同优化, 构建了“特征聚焦+多尺度适应+区域优化+知识补偿”的全链路解决方案。在PASCAL VOC和MS COCO数据集上的对比实验结果表明, 本文所提算法与当前主流的小样本目标检测方法相比具有更优越的检测性能, 能够进一步提升了小样本条件下目标检测的精度; 同时, 消融实验也证明了本文所提“CCAM+DFPN+渐进式RPN+知识补偿”框架的有效性。未来, 本文将探索更高效的注意力机制或语义引导策略, 提升模型对单个样本的特征挖掘能力, 适应更严苛的小样本条件, 以提升本文所提算法在1-Shot下的性能。

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors declare no relevant conflict of interests.

参考文献:

- [1] MUDAVATH T, MAMIDI A. Object detection challenges: Navigating through varied weather conditions-A comprehensive survey[J]. Journal of Ambient Intelligence and Humanized Computing, 2025, 16(2): 443-457.
- [2] 冀乃庚, 王伟鹏, 窦逸辛. 端云人脸识别系统计算卸载策略设计[J]. 计算机科学, 2025, 52(S1): 848-854.
JI Naigeng, WANG Weipeng, DOU Yixin. Design of computation offloading strategy for device-cloud face recognition system[J]. Computer Science, 2025, 52(S1): 848-854. (in Chinese)
- [3] 常星花, 王建荣. 基于优化YOLOv7-Tiny的表情识别算法[J]. 测试技术学报, 2025, 39(2): 113-120.
CHANG Xinghua, WANG Jianrong. Expression recognition algorithm based on optimized YOLOv7-tiny[J]. Journal of Test and Measurement Technology, 2025, 39(2): 113-120. (in Chinese)
- [4] FAN W, DING J, PENG B, et al. Multi-scale fire target detection algorithm using YOLO-fire[J]. Journal of Measurement Science and Instrumentation, 2025, 16(4): 625-636.
- [5] 施玮, 吴坚. 基于DEMATEL和TOPSIS的低空安防系统方案多属性评价方法[J]. 科学技术与工程, 2025, 25(14): 6109-6117.
SHI Wei, WU Jian. A multi-attribute evaluation

- method for low-altitude security system solutions based on DEMATEL and TOPSIS[J]. Science Technology and Engineering, 2025, 25(14): 6109-6117. (in Chinese)
- [6] 陈怡娜. 基于计算机视觉技术的目标跟踪方法研究[J]. 自动化仪表, 2025, 46(6): 48-52.
CHEN Yina. Research on target tracking method based on computer vision technology [J]. Process Automation Instrumentation, 2025, 46(6): 48-52. (in Chinese)
- [7] 杨杰, 胡文军, 臧影. 特征扰动池融合机制的多类工业缺陷检测[J]. 中国图象图形学报, 2025, 30(5): 1404-1418.
YANG Jie, HU Wenjun, ZANG Ying. Feature disturbance pool-based fusion mechanism for multi-class industrial defect detection [J]. Journal of Image and Graphics, 2025, 30(5): 1404-1418. (in Chinese)
- [8] 李洪波, 何启学, 鲍胜利. 基于改进 YOLOv8n 的轻量级入侵物种检测算法[J]. 计算机应用, 2025, 45(S2): 278-286.
LI Hongbo, HE Qixue, BAO Shengli. Lightweight invasive species detection algorithm based on improved YOLOv8n [J]. Journal of Computer Applications, 2025, 45(S2): 278-286. (in Chinese)
- [9] KATZMANN A, TAUBMANN O, AHMAD S, et al. Explaining clinical decision support systems in medical imaging using cycle-consistent activation maximization[J]. Neurocomputing, 2021, 458: 141-156.
- [10] 柴锐, 罗钰冰, 秦品乐, 等. 基于边缘融合和多尺度特征增强的结肠癌腺体分割网络[J]. 中北大学学报(自然科学版), 2025, 46(4): 411-421.
CHAI Rui, LUO Yubing, QIN Pinle, et al. Colon cancer gland segmentation network based on edge fusion and multi scale feature enhancement[J]. Journal of North University of China (Natural Science Edition), 2025, 46(4): 411-421. (in Chinese)
- [11] WANG X, HUANG T E, DARRELL T, et al. Frustratingly simple few-shot object detection [DB/OL]. (2020-05-16) [2025-05-11] <https://arxiv.org/abs/2003.06957>.
- [12] SUN B, LI B, CAI S, et al. FSCE: Few-shot object detection *via* contrastive proposal encoding [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 7348-7358.
- [13] QIAO L, ZHAO Y, LI Z, et al. DeFRCN: Decoupled faster R-CNN for few-shot object detection [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2022: 8661-8670.
- [14] KRISTAN M, LUKEŽIČ A, PELHAN J, et al. A novel unified architecture for low-shot counting by detection and segmentation [C]//Advances in Neural Information Processing Systems 37, 2024: 66260-66282.
- [15] ZHAO P, CHEN J, WAN H, et al. Few-shot object detection for SAR images *via* context-aware and robust Gaussian flow representation [J]. Remote Sensing, 2025, 17(3): 391.
- [16] ZHANG G, LUO Z, CUI K, et al. Meta-DETR: Few-shot object detection *via* unified image-level meta-learning [DB/OL]. (2021-09-20) [2025-05-11]. <https://arxiv.org/abs/2103.11731v2>.
- [17] KANG B, LIU Z, WANG X, et al. Few-shot object detection *via* feature reweighting [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 8419-8428.
- [18] YAN X, CHEN Z, XU A, et al. Meta R-CNN: Towards general solver for instance-level low-shot learning [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2020: 9576-9585.
- [19] XIAO Y, LEPETIT V, MARLET R. Few-shot object detection and viewpoint estimation for objects in the wild [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(3): 3090-3106.
- [20] WANG Y X, RAMANAN D, HEBERT M. Meta-learning to detect rare objects [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2020: 9924-9933.
- [21] CHEN H, WANG Y, WANG G, et al. LSTD: A low-shot transfer detector for object detection [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 2836-2843.