

## 面向单一背景的改进RetinaNet目标检测方法研究

周波, 江佩峰, 段昶, 罗月童

引用本文

周波, 江佩峰, 段昶, 罗月童. 面向单一背景的改进RetinaNet目标检测方法研究[J]. 计算机科学, 2023, 50(7): 137-142.

ZHOU Bo, JIANG Peifeng, DUAN Chang, LUO Yuetong. [Study on Single Background Object Detection Oriented Improved-RetinaNet Model and Its Application](#) [J]. Computer Science, 2023, 50(7): 137-142.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

**Similar articles recommended (Please use Firefox or IE to view the article)**

### [基于遗传算法的恶意软件对抗样本生成方法](#)

Adversarial Malware Generation Method Based on Genetic Algorithm

计算机科学, 2023, 50(7): 325-331. <https://doi.org/10.11896/jsjcx.220800176>

### [基于深度学习的活跃IPv6地址预测算法](#)

Deep Learning-based Algorithm for Active IPv6 Address Prediction

计算机科学, 2023, 50(7): 261-269. <https://doi.org/10.11896/jsjcx.220700076>

### [基于时序知识图谱嵌入的短期地铁客流量预测](#)

Short-term Subway Passenger Flow Forecasting Based on Graphical Embedding of Temporal Knowledge

计算机科学, 2023, 50(7): 213-220. <https://doi.org/10.11896/jsjcx.220600120>

### [面向自动驾驶的三维目标检测综述](#)

Review of 3D Object Detection for Autonomous Driving

计算机科学, 2023, 50(7): 107-118. <https://doi.org/10.11896/jsjcx.220700090>

### [探索站点时空移动模式:长短期交通预测框架](#)

Exploring Station Spatio-Temporal Mobility Pattern:A Short and Long-term Traffic Prediction Framework

计算机科学, 2023, 50(7): 98-106. <https://doi.org/10.11896/jsjcx.220900109>

# 面向单一背景的改进 RetinaNet 目标检测方法研究

周波 江佩峰 段昶 罗月童

合肥工业大学计算机与信息学院 合肥 230601

安全关键工业测控技术教育部工程研究中心 合肥 230009

(zhoubo810707@hfut.edu.cn)

**摘要** 基于深度学习的目标检测算法在工业缺陷检测领域得到了充分推广与应用,但少有适用于工业检测场景中单一背景算法。文中以工业检测场景中具有大量简单重复背景为出发点,对 RetinaNet 算法进行了如下改进:1)引入难负样本挖掘策略,减小了大量简单重复负样本对模型拟合正样本的影响;2)设计了自适应忽略样本选择策略,避免与正样本交并比高的样本混入负样本而混淆模型训练;3)简化了 RetinaNet 的分类子网络,降低了模型改进后的过拟合风险。在公开的 PCB 缺失孔数据集及自建的 LED 气泡数据集上,相比 RetinaNet 算法,改进后的方法在召回率上分别提升了 14.1% 和 1.8%,在精确率上分别提升了 3.6% 和 0.4%,表明改进后的方法能显著提升单一背景下的目标检测水平。

**关键词**:深度学习;RetinaNet;自适应采样;单一背景

中图分类号 TP391

## Study on Single Background Object Detection Oriented Improved-RetinaNet Model and Its Application

ZHOU Bo,JIANG Peifeng,DUAN Chang and LUO Yuetong

School of Computer Science and Information Engineering,Hefei University of Technology,Hefei 230601,China

Engineering Research Center of Safety Critical Industrial Measurement and Control Technology,Ministry of Education,Hefei 230009,China

**Abstract** Object detection algorithms based on deep learning have been fully promoted and applied in the field of industrial defect detection,but few algorithms are suitable for single background in industrial detection scenarios. This paper takes the industrial detection scene with a large number of simple repeated backgrounds as the starting point,and makes the following improvements to the RetinaNet algorithm:1)introduce the difficult negative sample mining strategy to reduce the impact of a large number of simple repeated negative samples on the model fitting positive samples;2)an adaptive ignoring sample selection strategy is designed to avoid mixing samples with high intersection ratios of positive samples into negative samples to confuse model training;3)the classification sub-network of RetinaNet is simplified,and the risk of overfitting after model improvement is reduced. Compared with the RetinaNet algorithm,the improved method improves the recall rate by 14.1% and 1.8%,and the precision rate by 3.6% and 0.4% respectively on the public PCB missing hole dataset and the self-built LED bubble dataset,indicating that the improved method can significantly improve the level of object detection in a single background.

**Keywords** Deep learning,RetinaNet,Adaptive sampling,Single background

### 1 引言

近年来,随着深度学习技术的发展,基于深度学习的目标检测算法的检测效率迭代提升,目标检测算法在工业缺陷检测领域得到重视与推广。在部分工业质检场景中,由于拍摄环境稳定、工业产品成批次生产,拍摄得到的产品表面图片通常呈现为背景单一的特点,而一些基于深度学习的目标检测算法采用了密集采样的方案,训练模型时简单重复的背景

样本数量远超前景采样数量,使得模型对前景样本的拟合不足,导致模型的检测性能受到了较大的影响,这在要求严苛的工业检测中是难以接受的。当前主流目标检测算法一般以几类常规数据集作为测试数据,如广泛用于目标检测算法验证的 PASCAL VOC 数据集、COCO 数据集,这一类数据集中的数据背景复杂多变、相似度较低,不具备工业检测数据背景相对单一、相似度极高的特点,这就使得尽管高效的目标检测算法推出迅速,但欠缺适用于工业检测环境的算法。

到稿日期:2022-05-07 返修日期:2022-10-14

基金项目:国家自然科学基金(61602146);国家重点基础研究发展计划(2017YFB1402200);安徽省科技攻关计划(1604d0802009)

This work was supported by the National Natural Science Foundation of China(61602146),National Basic Research Program of China(2017YFB1402200)and Key Science and Technology Program of Anhui Province, China(1604d0802009).

通信作者:罗月童(ytluo@hfut.edu.cn)

当前基于深度学习的检测算法可分为以下几类:1)两阶段检测算法,如经典的 Faster R-CNN<sup>[1]</sup>, R-FCN<sup>[2]</sup>;2)单阶段检测算法,如 YOLO 系列算法中的 YOLO V4<sup>[3]</sup>、YOLO V5<sup>[4]</sup>、YOLOX<sup>[5]</sup>,以及经典的 SSD<sup>[6]</sup>算法、RetinaNET<sup>[7]</sup>算法;3)无需锚框的算法,如 CornerNet<sup>[8]</sup>, CenterNet<sup>[9]</sup>, FCOS<sup>[10]</sup>等;另外,随着 Transformer 技术的发展,新推出了 Swin Transformer<sup>[11]</sup>算法。在两阶段检测算法中,前一阶段由区域建议模块生成一组稀疏的候选目标区域,在后一阶段再对它们进行进一步分类和回归;在单阶段检测算法中,在提取图片特征后,通过密集地采样得到训练样本,然后进行分类和位置回归。一般来说,由于两阶段检测算法中存在区域建议模块,已经对负样本数量进行了有效控制,因此检测精度较单阶段检测算法高,但速度一般较单阶段算法慢,因此从样本选择方面考虑提升目标检测算法的效率是一种可行的办法<sup>[12]</sup>,如 RON<sup>[13]</sup>利用卷积特征的目标性先验约束来显著减小前景目标的搜索空间;SSD 使用难负样本挖掘策略<sup>[6]</sup>,对负样本按损失值降序排序,选择靠前的一定数量的负样本参与反向传播,并控制正负样本比例为 1:3;ATSS<sup>[14]</sup>根据锚框与目标框交并比的统计特征为每个目标设定动态交并比阈值,使目标能匹配到合适的训练正样本;RetinaNet<sup>[10]</sup>算法也通过

重塑标准交叉熵损失来解决正负样本不均衡问题,增加了困难样本在模型损失中的权重。

由于 RetinaNet 较好地均衡了检测速度和精度,具有较高的实际应用价值,已在工业检测中得到了广泛认可,本文拟在 RetinaNet 算法的基础上进行算法改进。RetinaNet 算法中的采样属于密集采样,而工业检测中的单一背景会导致大量背景样本的相似性较常规数据高,本文拟引入难负样本挖掘来减少简单负样本参与训练,制定忽略样本的选择策略,使正样本得到充分训练,调整分类子网络以降低子网络过拟合的风险,从而实现在工业检测场景中算法检测效率的提升。

## 2 本文方法

本节首先介绍了 RetinaNet 算法的原理,依次从难负样本挖掘、忽略样本的选择策略、分类子网络的调整这几个改进点进行了阐述。

### 2.1 RetinaNet 简介

RetinaNet 采用骨干网络 ResNet<sup>[15]</sup>提取特征,利用特征金字塔(Feature Pyramid Networks, FPN)<sup>[16]</sup>实现多尺度特征提取,随后由两个全卷积(Fully Convolutional Networks, FCN)<sup>[17]</sup>子网络(见图 1)分别负责类别判断与位置回归。

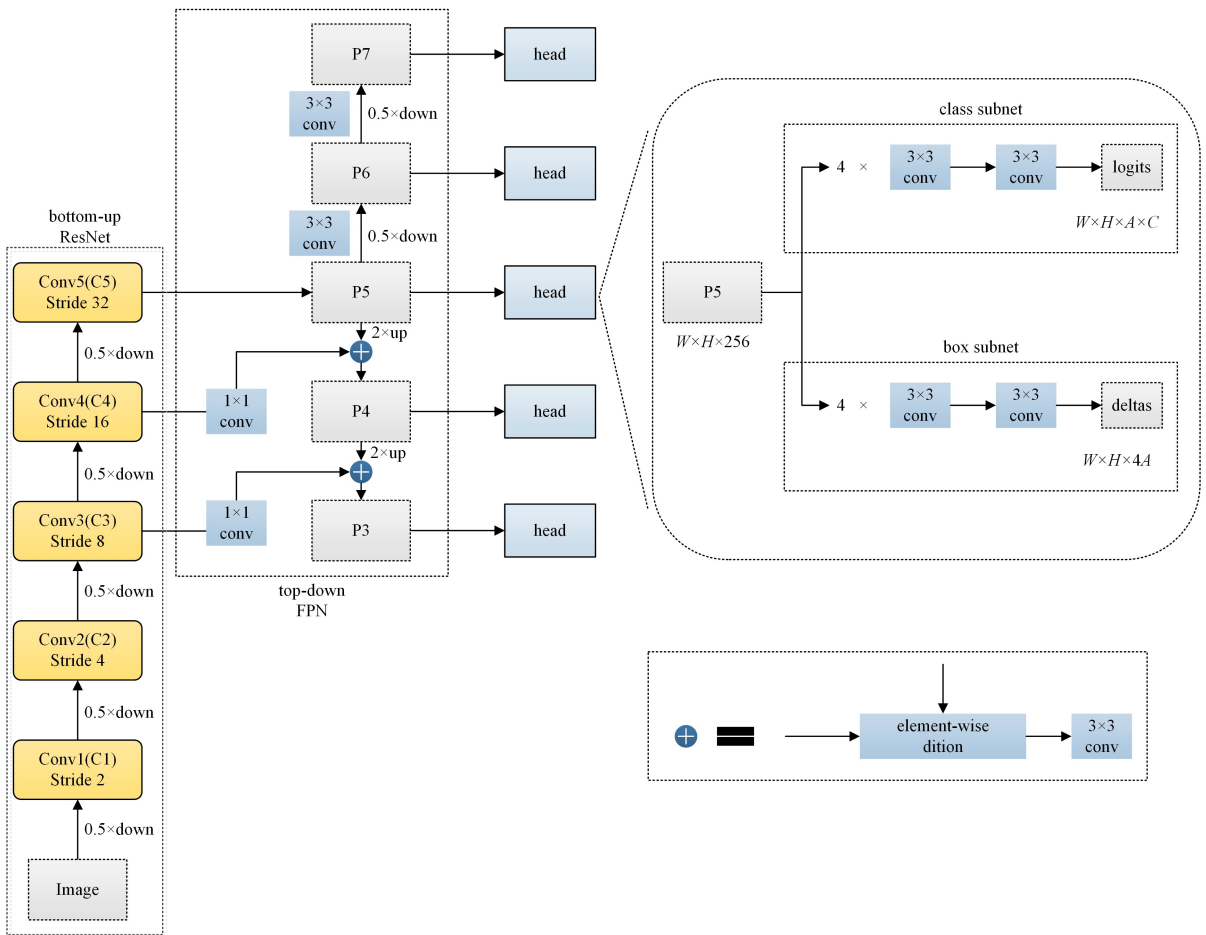


图 1 RetinaNet 的结构

Fig. 1 Architecture of RetinaNet

RetinaNet 采用了密集采样策略,具体实现如下:输入图像经 FPN 得到多尺度特征图后,在每个特征图上生成多种尺度、多种宽高比的锚框,计算锚框与目标框的交并比,与预设

超参数(设超参数  $k_1, k_2$ , 满足条件  $0 < k_1 \leq k_2 < 1$ )进行对比,将交并比值位于  $[0, k_1)$  的锚框赋值为负样本,位于  $[k_2, 1]$  的锚框赋值为正样本,对于某个标记目标,若不存在与其交并

比高于  $k_2$  的锚框,则取与其交并比最大的锚框赋值为该目标的正样本,而其他未赋值样本则被作为忽略样本,随后将采样样本输入子网络中,分别计算目标分类损失和边界框回归损失,两者相加得到总损失。

在子网络的分类损失计算中,RetinaNet 提出了 Focal loss<sup>[7]</sup>,其通过改进交叉熵损失函数,降低了训练过程中“简单”样本的损失权重,使算法能够聚焦于“困难”样本的学习,从而提高识别率。Focal loss 的计算式如式(1)所示:

$$FL(p_t) = -\alpha_t (1 - p_t)^\gamma \log(p_t) \quad (1)$$

其中,  $p_t$  指训练样本类别预测的置信度,  $\alpha_t$  和  $\gamma$  为超参数,本文分别采用经验值 0.25 和 2.0。

## 2.2 难负样本挖掘

RetinaNet 采用的 Focal loss 减少了大量简单负样本的损失贡献,有效降低了样本分布不均衡带来的影响,并在 MS COCO 数据集上进行了对比实验,采用 Focal loss 优于使用难负样本挖掘(Hard Negative Mining, HNM)。但在工业缺陷检测场景中,由于各训练数据的背景基本相同,采样后负样本特征的分布基本一致,尽管 Focal loss 已经降低“简单”负样本的损失贡献,但由于采用了密集采样策略,“雷同”特征样本在损失中的比重依然偏大。因此,本文使用难负样本挖掘进行了尝试,仅允许少量困难负样本和所有正样本参与损失计算,预期能够缓解在单一背景下的采样样本不均衡问题。

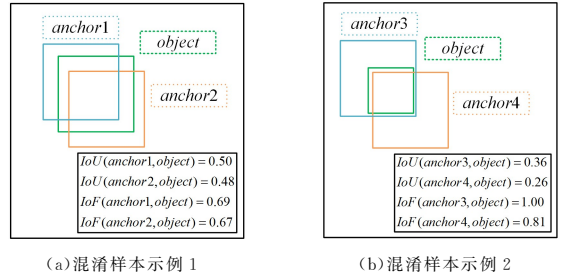
本文难负样本挖掘的具体做法如下:在训练阶段,按损失值大小对所有负样本进行降序排序,选择损失值高的负样本参与损失计算,确保困难负样本得到充分训练,并控制正负样本的比例为 1:3。与 RetinaNet 算法相比,本文所采用方法仅允许少量的困难负样本参与损失计算,这有可能打破分类损失和回归损失的平衡。为了维持分类损失和回归损失的平衡,本文对分类损失乘以一个动态计算的加权值  $\lambda$ ,以使分类损失和回归损失处于同一数量级上。 $\lambda$  的计算式如式(2)所示:

$$\lambda = Pow(10, E(N(reg\_loss)) - E(N(cls\_loss))) \quad (2)$$

其中,  $Pow$  为幂函数,  $N$  为科学计数法转换,  $E$  指取指数,  $reg\_loss$  为回归损失,  $cls\_loss$  为分类损失。

## 2.3 自适应忽略样本选择

如 2.1 节所述,RetinaNet 算法中采用静态超参数  $k_1, k_2$  来区分正负样本、忽略样本。在实际检测任务中,检测效果对超参数的变动十分敏感,例如当  $k_1 = k_2 = 0.5$  时,可能会存在与目标交并比接近 0.5 的正样本和负样本,这些正负样本实际上十分相似,如图 2(a)所示,anchor2 作为负样本会混淆模型的训练;当  $k_1 = 0.4, k_2 = 0.5$ ,且与目标重叠的锚框的交并比都低于  $k_2$  时,会选择与其交并比最大的锚框赋值为该目标的正样本,如图 2(b)所示,anchor3 会被赋为正样本,anchor4 则被赋为负样本,但实际上它们与目标的 IoF(Intersection over Foreground,指边界框与前景框的重合区域与前景区域的面积比值,用于衡量边界框对前景区域的覆盖程度)都高于 0.8,这意味着它们对应的特征都可能拥有完全覆盖目标区域的感受野,因此 anchor4 作为负样本也可能会混淆模型的训练。



(a) 混淆样本示例 1

(b) 混淆样本示例 2

图 2 混淆样本示例

Fig. 2 Obfuscated sample examples

针对以上问题,文献[14]提出了更灵活鲁棒的自适应训练样本选择策略,但是该策略只定义了负责将锚框赋值为正样本或负样本的单个动态阈值,而没有定义与正样本或目标接近的容易混淆模型训练的忽略样本;文献[18]提出了忽略区域和忽略区域不采负样本的策略,该算法首先在行人检测任务中进行了实验,具体的做法是将存在密集人群或其他非行人的人的区域标注为忽略区域,若锚框覆盖的忽略区域占锚框区域的比例大于设定阈值,则将该锚框定义为忽略样本。

本文受文献[14]中的自适应训练样本的定义和文献[18]中的忽略样本的定义的启发,设计了一个自适应忽略样本选择策略(Adaptive Ignore Sample Selection, AISS)。首先根据与目标匹配的正样本的交并比信息,为每个目标设置选择候选忽略样本的动态 IoU 阈值,然后根据预设的静态 IoF 阈值,从候选忽略样本中选择确定最终的忽略样本,避免与正样本或目标接近的样本被作为负样本而混淆模型训练,进而提升模型训练的效果和稳定性。

AISS 的具体过程如图 3 所示,统计目标  $i$  的所有正样本中最小交并比  $IoU_i$ ,选择与目标  $i$  的交并比大于  $IoU_i \times w$  的非正样本作为候选忽略样本,然后计算候选忽略样本的目标  $i$  和锚框的  $IoF$ ,选择  $IoF$  大于  $u$  的候选忽略样本作为忽略样本,其中实验中使用网格搜索的方法寻找最优的  $w$  和  $u$  的取值,经验取值分别为 0.8 和 0.7;另外,RetinaNet 为使不同大小、不同形状的目标都可以匹配到合适的锚框,需要在所有特征图上设置不同尺度、不同宽高的锚框,而在同一位置上的多个锚框都具有相同的感受野。为此,本文方法将与正样本同一位置上的其他非正样本都赋值为忽略样本。

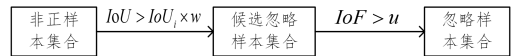


图 3 自适应忽略样本选择过程

Fig. 3 Adaptive ignore sample selection process

## 2.4 分类子网络的结构调整

如前文所述,RetinaNet 算法中,使用两个全卷积子网络分别负责分类和位置回归,两个子网络分别是由 5 层  $3 \times 3$  的步长为 1 的卷积组成,其中两个子网络的预期输出不同,它们最后一层卷积的通道数分别是同一位置所设锚框个数  $\times$  目标类别数和同一位置所设锚框个数  $\times$  位置回归系数个数。对于负责分类的子网络,影响权重学习的样本包括允许贡献分类损失的所有正样本和所有负样本,对于负责位置回归的子网络,影响权重学习的样本包括允许贡献回归损失的所有正样本。在应用难负样本挖掘和自适应忽略样本选择后,负样本

中仅存在与正样本交并比较大的少量困难样本,样本数量的大幅减少导致分类网络过拟合风险增大。

针对深度神经网络的过拟合问题的常见解决方法可分为3类:1)添加更多的训练数据;2)添加约束项;3)降低模型复杂度<sup>[19]</sup>。方法1)、方法2)的设定均与具体的应用场景密切相关,并不适用于本文,鉴于 RetinaNet 中涉及的子网络具有一定的复杂度调整空间,本文采用方法3)。分类子网络中,最后一层卷积需要负责将特征表示空间映射到目标类别空间,其他4层卷积则负责针对分类任务对特征进行进一步整合。根据分类子网络的特点,可以通过减少分类子网络中用于特征整合的卷积层来降低模型的复杂度。为了验证过拟合问题和探究具体的结构调整策略,分别采用 RetinaNet 和添加 HNM 与 AISS 策略的 RetinaNet 在 PCB 数据集上开展了网络结构调整的实验,实验数据详情以及参数设置可见第3节,实验的模型评估结果如表1所列。根据表1中的结果可以得出以下结论:1)随着分类子网的层数减少,改进前后的两个模型的召回率都在不断地提升,说明分类子网层数少的模型对数量少的正样本具有更强的泛化能力;2)对于未应用 HNM 与 AISS 的 RetinaNet 模型,其允许数量多的负样本参与训练过程,当减少分类子网层数后,模型的精确率降低,说明分类子网层数少的模型难以很好地拟合数量多的负样本;3)对于应用 HNM 与 AISS 的 RetinaNet 模型,其仅允许数量少的负样本参与训练,当减少分类子网层数后,模型的精确率上升,说明分类子网层数少的模型对数量少的负样本具有更强的泛化能力;4)对于改进后的 RetinaNet 模型,当分类子网络中用于特征整合的卷积层仅保留一层时(见图4),模型的召回率和精确率都达到最优,说明减少分类子网层数能够有效提升改进模型对负样本和正样本的泛化能力。

表1 在 PCB 数据集上网络结构调整的实验结果

Table 1 Experimental results of network structure adjustment on

PCB dataset			
情况	分类子网层数	Recall/%	Precision/%
RetinaNet	5	81.0	95.4
	4	81.7	95.3
	3	82.4	95.3
	2	82.8	94.3
RetinaNet+HNM+ AISS	5	94.6	98.4
	4	94.6	98.9
	3	94.7	98.9
	2	95.1	99.0

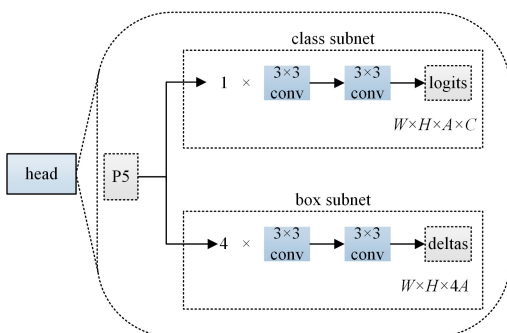


图4 调整后的预测网络

Fig. 4 Adjusted prediction network

### 3 实验验证与分析

为了对比分析,本文数据集选择了单一背景和复杂背景这两类数据集。单一背景数据集为北京大学公开的印刷电路板(Printed Circuit Board, PCB)瑕疵数据集和自建的 LED 点胶缺陷数据集。PCB 瑕疵数据集是由北京大学智能机器人开放实验室公开发布的,该数据集共包含 1386 张图片和 6 种目标,而 LED 点胶缺陷数据集则是由与课题组合作研发半导体表面检测系统的企业所提供,该数据集共包含 400 张图片和 3 种目标。这两类数据集中目标的大小和形状比较规则,背景区域单一重复。复杂背景数据集由香港中文大学研究人员制作的人脸检测数据集 Wider Face,根据事件场景的类型,该数据集可以被划分为 61 个类别,如 Parade, Meeting, Traffic,其中目标大小和目标形状的分布广泛,背景区域非单一重复。

实验使用基于 Pytorch 深度学习框架的开源工具包 Detectron2,工具包版本为 0.6,ResNet50 作为 RetinaNet 的骨干网络,使用 ImageNet 预训练模型。硬件条件如下:17-7700K CPU, Titan X GPU, 32 GB 内存。各实验的超参数如下:batch size 为 4, epoch 为 100, initial learning rate 为 0.001 (当训练至第 60 个和第 80 个 epoch 时,学习率分别除以 10), linear warmup 为初始训练的前 100 步,特征图上每个空间位置铺设单个锚框的长宽为 4 倍步长。

实验从精确率和召回率两个方面评估模型性能,精确率 Precision 和召回率 Recall 的计算式如式(3)、式(4)所示:

$$Precision = TP / (TP + FP) \quad (3)$$

$$Recall = TP / (TP + FN) \quad (4)$$

其中, TP (True Positive) 为被正确检测为正例的个数, FP (False Positive) 表示被错误检测为正例的个数, FN (False Negative) 表示被错误检测为负例的个数,置信分数阈值取 0.3, 交并比阈值取 0.05。

#### 3.1 PCB 瑕疵数据集上的实验

为了便于评估模型性能,本次实验仅选用缺失孔类别的数据,在具体操作中将 115 个样本数据随机划分为 3 份,分别包含 39, 38, 38 个样本,样本图像短边缩放到 320 个像素(长边则按短边缩放比例进行缩放),通过将其其中一份作为验证集而其他两份作为训练集的方式,生成 3 组包含不同训练集和不同验证集的数据,进行 5 次三折交叉验证实验,每组实验的结果取 5 次的平均值。实验的模型评估结果如表 2 所列。

表2 PCB 瑕疵数据集的实验结果

Table 2 Experimental results of PCB defect dataset

情况	HNM	AISS	分类子网层数	Recall/%	Precision/%
1	×	×	5	81.0	95.4
2	✓	×	5	85.4	95.5
3	✓	✓	5	94.6	98.4
4	✓	✓	2	95.1	99.0

#### 3.2 LED 点胶缺陷数据集上的实验

为了便于评估模型性能,本次实验仅使用气泡类别的数据,实际操作中,将 151 个样本数据随机划分为 3 份,分别包含 51, 50, 50 个样本,样本图像短边缩放到 320 个像素(长边

则按短边缩放比例进行缩放),通过将其中一份作为验证集而其他两份作为训练集的方式,生成3组包含不同训练集和不同验证集的数据,进行5次三折交叉验证实验,每组实验的结果取5次的平均值。实验的模型评估结果如表3所列。

表3 LED点胶缺陷数据集的实验结果

Table 3 Experimental results of LED dispensing defect dataset

情况	HNM	AISS	分类子网层数	Recall/%	Precision/%
1	×	×	5	97.1	94.7
2	√	×	5	97.6	94.6
3	√	√	5	98.7	94.8
4	√	√	2	98.9	95.1

### 3.3 Wider Face 人脸检测数据集上的实验

为了便于评估模型性能,鉴于 Wider Face 数据集中 Parade 事件类别的人脸数据具有背景复杂、样本规模适中的特点,本次实验仅使用 Parade 事件类别的数据,实际操作中将461个样本数据随机划分为3份,分别包含154,153,153个样本,样本图像短边缩放到640个像素(长边则按短边缩放比例进行缩放),通过将其中一份作为验证集而其他两份作为训练集的方式,生成3组包含不同训练集和不同验证集的数据,进行5次三折交叉验证实验,每组实验的结果取5次的平均值。实验的模型评估结果如表4所列。

表4 Wider Face 数据集的实验结果

Table 4 Experimental results of Wider Face dataset

情况	HNM	AISS	分类子网层数	Recall/%	Precision/%
1	×	×	5	60.0	64.8
2	√	×	5	61.1	62.7
3	√	√	5	60.4	63.2
4	√	√	2	60.2	61.0

### 3.4 实验分析与结论

#### (1) 单一背景数据集上的实验分析与结论

在公开的 PCB 数据集和自建的 LED 数据集上,使用改进策略进行了验证实验,验证结果分别记录于表2和表3,对比表2和表3中的结果可以发现,各种改进策略对模型检测准确度的影响是基本一致的,因此可以得出如下分析与结论:

1)对比情况1和情况2,应用难负样本挖掘,对于单一背景的检测任务,模型的召回率都有所提升,而模型的精确率则没有明显改变,说明难负样本挖掘策略在单一背景下可以有效减轻正负样本不均衡的影响,在不影响模型对负样本学习的情况下,提升模型对正样本的学习效果。

2)对比情况2和情况3,在单一背景的检测任务上,应用自适应忽略样本选择策略显著提升了模型的召回率和精确率,说明自适应忽略样本选择策略可以通过减少与正样本或目标接近的混淆样本,来有效提升模型的训练效果。

3)对比情况3和情况4,在单一背景的检测任务上,当仅选择少量合适的负样本参与训练时,采用浅层分类子网络来进一步提升模型的召回率和精确率,说明在简单检测任务和少量训练样本的情况下,浅层分类子网络具有更好的泛化能力。

#### (2) 复杂背景数据集上的实验分析与结论

根据单一背景数据集上的实验分析与结论,可知改进

方法在单一背景检测任务是具备有效性和实用性的。但改进方法实际上主要是基于单一背景检测任务中存在大量简单负样本这一特点,因此改进方法在预期情况下是不适用于复杂背景检测任务的。为了进一步验证改进方法的适用范围,本文还完成了改进方法在公开的复杂背景的 Wider Face 人脸检测数据集上的实验,其实验结果记录于表4。根据表4中的结果,可以得出如下结论:

1)对比情况1和情况2,应用难负样本挖掘,对于复杂背景的检测任务,虽然模型的召回率有所提升,但模型的精确率也显著降低,说明难负样本挖掘策略在复杂背景下仅允许少量难负样本参与训练,会在严重影响模型对负样本学习的情况下,提升模型对正样本的学习效果。

2)对比情况2和情况3可以发现,在复杂背景的检测任务上,应用自适应忽略样本选择策略,模型的召回率降低了0.7%,而模型的精确率则提升了0.5%,说明自适应忽略样本选择策略使模型选择了更合适的负样本,提升了模型对负样本的学习。

3)对比情况3和情况4可以发现,在复杂背景的检测任务上,当仅选择少量合适的负样本参与训练时,采用浅层分类子网络降低了模型的召回率和精确率,说明对于复杂检测任务,即使在少量训练样本的情况下,仍然需要表征能力更强的深层分类子网络。

综上,本文改进方法使用难负样本挖掘策略和自适应忽略样本选择策略使模型采得少量且适合参与训练的负样本,有效解决了 RetinaNet 在单一背景目标检测中的极端正负样本不均衡问题,同时简化 RetinaNet 分类子网络,降低了模型在少量训练样本下的过拟合风险。但是在复杂背景的检测任务上,本文改进方法的检测效果有所降低,说明其仅适用于单一背景目标检测。

**结束语** 本文在分析目标检测算法 RetinaNet 原理和单一背景目标检测特点的基础上,以利用单一背景下大量简单重复负样本为出发点,使用难负样本挖掘策略来缓解单一背景下的样本不均衡问题,提出并采用自适应忽略样本选择策略来避免与标记目标接近的负样本混淆模型的训练,减少了分类子网络的层数,从而降低少量训练样本下的过拟合风险。在北京大学 PCB 缺失孔数据集和自建 LED 气泡数据集上进行了实验验证,实验结果表明,在单一背景目标检测任务中,难负样本挖掘策略可以显著提升模型对正样本的学习效果,自适应忽略样本选择策略有利于模型更好地拟合正样本和负样本,减少分类子网络的层数,使模型在少量训练样本上得到更正确的收敛。以上策略都有效提高了模型在单一背景目标检测上的检测水平。但同时我们也认识到,难负样本挖掘策略仅选择相对正样本固定比例的高损失负样本,丢失了一些损失低但对负样本学习重要的训练样本,因此未来的工作将探索如何更好地选择对负样本学习重要的训练样本,以便进一步提高单一背景下的目标检测精度。

### 参考文献

[1] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-

- time object detection with region proposal networks[C]// Advances in Neural Information Processing Systems. 2015;91-99.
- [2] DAI J, LI Y, HE K, et al. R-fcn: Object detection via region-based fully convolutional networks[J]. Advances in Neural Information Processing Systems, 2016, 29: 379-387.
- [3] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv: 2004. 10934, 2020.
- [4] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]// Proceedings of the IEEE/ CVF International Conference on Computer Vision. 2021; 2778-2788.
- [5] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021[J]. arXiv: 2107. 08430, 2021.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]// European Conference on Computer Vision. Cham: Springer, 2016; 21-37.
- [7] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017; 2980-2988.
- [8] LAW H, DENG J. Cornernet: Detecting objects as paired keypoints[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018; 734-750.
- [9] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint triplets for object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019; 6569-6578.
- [10] TIAN Z, SHEN C, CHEN H, et al. Fcos: Fully convolutional one-stage object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019; 9627-9636.
- [11] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021; 10012-10022.
- [12] ZHANG S, WEN L, BIAN X, et al. Single-shot refinement neural network for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018; 4203-4212.
- [13] KONG T, SUN F, YAO A, et al. Ron: Reverse connection with objectness prior networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017; 5936-5944.
- [14] ZHANG S, CHI C, YAO Y, et al. Bridging the Gap Between Anchor-based and Anchor-free Detection via Adaptive Training Sample Selection[J]. arXiv: 1912. 02424, 2019.
- [15] 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. 2016; 770-778.
- [16] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017; 2117-2125.
- [17] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015; 3431-3440.
- [18] ZHANG S, BENENSON R, SCHIELE B. Citypersons: A diverse dataset for pedestrian detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017; 3213-3221.
- [19] ZHUGE Y. The Quest for Machine Learning[M]. Beijing: Posts and Telecommunications Press, 2018.



**ZHOU Bo**, born in 1981, Ph.D, associate professor. His main research interests include digital terrain analysis and object detection.



**LUO Yuetong**, born in 1978, Ph.D, professor. His main research interests include image processing and scientific visualization.

(责任编辑:喻黎)