



# 计算机科学

COMPUTER SCIENCE

## 基于软标签和样本权重优化的 Anchor Free 目标检测算法

王灿, 刘永坚, 解庆, 马艳春

引用本文

王灿, 刘永坚, 解庆, 马艳春. 基于软标签和样本权重优化的 Anchor Free 目标检测算法[J]. 计算机科学, 2022, 49(8): 157-164.

WANG Can, LIU Yong-jian, XIE Qing, MA Yan-chun. Anchor Free Object Detection Algorithm Based on Soft Label and Sample Weight Optimization[J]. Computer Science, 2022, 49(8): 157-164.

---

## 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)

[基于边框距离度量的增量目标检测方法](#)

Incremental Object Detection Method Based on Border Distance Measurement

计算机科学, 2022, 49(8): 136-142. <https://doi.org/10.11896/jsjcx.220100132>

[改进 Faster R-CNN 的光学遥感飞机目标检测](#)

Remote Sensing Aircraft Target Detection Based on Improved Faster R-CNN

计算机科学, 2022, 49(6A): 378-383. <https://doi.org/10.11896/jsjcx.210300121>

[基于 GDIoU 损失函数的 YOLOv4 绝缘子高效定位算法](#)

High Performance Insulators Location Scheme Based on YOLOv4 with GDIoU Loss Function

计算机科学, 2022, 49(6A): 412-417. <https://doi.org/10.11896/jsjcx.210600089>

[基于外接圆半径差损失的实时安全帽检测算法](#)

Real-time Helmet Detection Algorithm Based on Circumcircle Radius Difference Loss

计算机科学, 2022, 49(6A): 424-428. <https://doi.org/10.11896/jsjcx.220100252>

[三维城市场景中的小物体检测](#)

Small Object Detection in 3D Urban Scenes

计算机科学, 2022, 49(6): 238-244. <https://doi.org/10.11896/jsjcx.210400174>

# 基于软标签和样本权重优化的 Anchor Free 目标检测算法

王 灿<sup>1,2</sup> 刘永坚<sup>1</sup> 解 庆<sup>1,2</sup> 马艳春<sup>1</sup>

1 武汉理工大学计算机科学与技术学院 武汉 430070

2 青海武汉理工大学文化科技融合产业技术研究院 青海 海东 810600

(cam@whut.edu.cn)

**摘 要** 与 Anchor Based 目标检测算法类似,基于特征点的 Anchor Free 目标检测算法也面临着在正负样本划分中存在模糊样本的问题,即根据特定阈值和特征点位置划分非正即负的训练样本,网络在对特征点位置处在临界值附近的样本进行训练时会产生较大的损失,使得网络将注意力过于集中在这些模糊样本上,降低了网络的整体检测性能。针对此情况,提出从软标签、损失函数和权重优化 3 个方面对基于特征点的 Anchor Free 目标检测算法进行改进,通过充分利用 Center Ness 参数来缓解模糊样本对网络性能的影响,提高目标检测的准确率。为证明所提方法的有效性,分别在经典的 Pascal VOC 数据集和 MS COCO 数据集上使用 FCOS 目标检测器进行对比实验,最终将检测器在 Pascal VOC 数据集上的  $mAP$  提升至 82.16%(提升约 1.31%),在 MS COCO 数据集上的  $AP_{50-95}$  提升至 35.8%(提升约 1.3%)。

**关键词**: 目标检测;模糊样本;Anchor Free;Center Ness;样本权重优化

中图法分类号 TP183

## Anchor Free Object Detection Algorithm Based on Soft Label and Sample Weight Optimization

WANG Can<sup>1,2</sup>, LIU Yong-jian<sup>1</sup>, XIE Qing<sup>1,2</sup> and MA Yan-chun<sup>1</sup>

1 School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China

2 Qinghai-WUT Industrial Technology Research Institute of Culture and Technology Integration, Haidong, Qinghai 810600, China

**Abstract** Similar to the Anchor Based object detection algorithm, the Anchor Free object detection algorithm based on feature points also encounters the problem of ambiguous samples when dividing positive and negative samples. That is, the training samples are divided either positive or negative according to the specific threshold and the position of feature points, and when the model trains samples whose feature point is near the borderline, it will incur great loss, which will make the model pay too much attention to these ambiguous samples and reduce the performance of the model. In view of this situation, this paper proposes to improve the Anchor Free object detection algorithm based on feature points from the three aspects of soft label, loss function and weight optimization. By making full use of Center Ness, the impact of ambiguous samples on network performance is mitigated and the accuracy of object detection is improved. To prove the effectiveness of the proposed method, the FCOS object detector is employed in the comparative experiments on the classical Pascal VOC and MS COCO datasets, respectively. Finally, the  $mAP$  of the detector on Pascal VOC dataset increases to 82.16% (an increase of 1.31%) and the  $AP_{50-95}$  on MS COCO dataset increases to 35.8% (an increase of 1.3%).

**Keywords** Object detection, Ambiguous samples, Anchor Free, Center Ness, Sample weight optimization

## 1 引言

目标检测一直是计算机视觉领域的研究热点之一<sup>[1]</sup>,它既是自动驾驶、机器人视觉<sup>[2]</sup>、医疗检测等应用的理论基础,也是图像分割、目标跟踪、显著性目标检测<sup>[3]</sup>等更复杂的计算机视觉任务的研究基础。在众多的目标检测算法中,基于锚框(Anchor Based)的目标检测算法一直是学者重点研究的

内容,不管是在双阶段还是在单阶段的目标检测器中,锚框(Anchor)都扮演着重要的角色。然而最新的研究成果表明<sup>[4-6]</sup>,无锚框(Anchor Free)的目标检测器也能取得很好的检测效果,并且还能显著降低算法的复杂程度,减少手工设计的超参数。因此,Anchor Free 目标检测算法得到了越来越多的关注。

在众多的 Anchor Free 目标检测算法中,基于特征点的

到稿日期:2021-06-30 返修日期:2021-12-09

基金项目:特色民族文化矢量数字化资源复用与产业创新项目(唐卡壁画矢量数字化标注及流程标准制定)

This work was supported by the Characteristic National Culture Vector Digital Resource Reuse and Industrial Innovation Project(Thangka and Fresco Vector Digital Annotation and Process Standard Formulation).

通信作者:解庆(felixxq@whut.edu.cn)

检测算法与传统的 Anchor Based 目标检测算法在检测机制上最为接近<sup>[7]</sup>,两者都是对预定义的样本(Anchor 或特征点)进行分类和回归操作,得到最终的预测框。相似的检测机制意味着基于特征点的 Anchor Free 目标检测算法在划分正负样本时面临着和 Anchor Based 目标检测算法类似的模糊样本问题。本文定义模糊样本为:当根据某一参数(Anchor 与真实框之间的 IoU 或特征点与真实框中心点之间的相对距离)和特定阈值硬性地划分非正即负的训练样本时,对应参数值处在阈值附近的样本。在训练阶段,模糊样本会产生较大的损失,使得网络将注意力过分集中于这些质量较差的样本上,从而影响整体的训练效果;在推理阶段,模糊样本会生成很多低质量的预测框,从而降低检测器的整体性能。

在 Anchor Based 目标检测算法中,模糊样本问题大多可以通过设置忽略样本的方式来解决,例如在 RetinaNet<sup>[8]</sup>中,与真实框之间的 IoU 大于 0.5 的 Anchor 被视为正样本,小于 0.4 的 Anchor 被视为负样本,在 0.4~0.5 之间的被视为忽略样本。但是在基于特征点的 Anchor Free 目标检测算法中,特征点的数量远少于 Anchor 的数量,设置忽略样本会大大减少正样本的数量,减小最优可能召回率(Best Possible Recalls, BPR<sup>[9]</sup>),最终影响算法的整体检测效果。因此,如何处理模糊样本带来的干扰成为了基于特征点的 Anchor Free 目标检测算法需要解决的问题。

当前主流的 Anchor Free 目标检测算法大多将网络关注的重心集中于真实框中心区域样本上,以改善模糊样本的问题。FoveaBox<sup>[10]</sup>通过多组对比实验确定了最优的中心采样区域,虽然它获得了一个较好的正负样本判定阈值,但是却并没有直面模糊样本的问题。FCOS<sup>[4]</sup>引入中心度(Center Ness, 表示特征点与目标框中心点相对位置的参数,距离越近,参数值越接近 1)网络预测分支,利用预测的 Center Ness 参数在推理阶段降低模糊样本的置信度,然后通过非极大值抑制(Non-Maximum Suppression, NMS)操作来减少模糊样本生成的低质量预测框,但其并未解决模糊样本在训练阶段产生影响的问题,在之后的工作中,Tian 等<sup>[9]</sup>添加了新的采样规则,重新划定采样区域(根据 FPN 步长和真实框的大小确定),但是与 FoveaBox 类似,改变正负样本判定阈值并没有从根本上改善模糊样本的问题。SAPD<sup>[11]</sup>利用 Center Ness 参数对训练样本进行加权操作,在训练过程中,目标框中心区域的样本权重较大,模糊样本的权重较小,从而减少了模糊样本的损失占比。但是在训练后期,由于分类的目标向量是非正即负的硬编码,中心区域的样本检测效果较好,因此产生的损失较小,成为了容易样本;边缘区域的样本检测效果较差,产生的损失较大,成为了困难样本。由于 Focal Loss(FL)会对困难样本赋予较大权重,因此 Center Ness 加权操作在训练后期的效果会大打折扣。

受此启发,本文充分利用 Center Ness 参数,从软标签、损失函数和权重优化 3 个方面做出改进,以缓解模糊样本对检测器的干扰。本文方法的核心思想也是将网络关注的重心始终集中在真实框中心区域样本上,以改善模糊样本问题。

具体而言,本文通过 Center Ness 参数设置软标签,使得正负样本之间存在一个相对平滑的过渡。此时,训练样本

代表的含义不再是“样本是否包含某个目标”,而是“样本包含某个目标的概率”。借助 Center Ness 参数,本文令目标框中心区域的样本获得较大概率值,边缘区域的样本获得较小概率值,将分类任务变成一个概率回归任务,从源头减小模糊样本的干扰。

由于设置了软标签,分类网络所做的不再是一个纯粹的分类任务,Focal Loss 变得不再适用,因此本文在 Focal Loss 的基础上设计了新的损失函数,并且参考 SAPD 融入了 Center Ness 的权重优化信息,使之更加适合包含软标签的分类损失计算。新的损失函数具备以下优点:1)可以帮助检测器在使用软标签的情况下很好地收敛;2)充分利用 Center Ness 参数进行权重优化,使得中心区域样本在训练过程中的重要性进一步提高;3)保证网络在训练过程中的优化重心能够尽可能集中在中心区域样本上。

本文的主要贡献如下:

(1)使用软标签的方法,从根本上缓解模糊样本对检测器的影响,同时降低了模糊样本的置信度预测值,增加了中心区域样本的置信度预测值,有助于 NMS 抑制质量较差的边缘预测框。

(2)在 Focal Loss 的基础上设计了新的损失函数并添加了权重优化信息,使得检测器在使用软标签进行训练时能够取得很好的检测效果,同时进一步增加了中心区域样本在训练过程中的权重。

(3)分别在 Pascal VOC 数据集和 MS COCO 数据集上设计了对比实验,以证明本文方法的有效性。在相同的实验条件下,本文将 FCOS<sup>[9]</sup>在 Pascal VOC 测试集上的 *mAP* 集上的提升至 82.16%(提升约 1.31%),在 MS COCO 数据集上的 *AP*<sub>50-95</sub> 提升至 35.8%(提升约 1.3%)。

## 2 相关工作

### 2.1 Anchor Based 与 Anchor Free 目标检测算法

Anchor 是通过一定规则预定义的一系列矩形框,其本质是提供给神经网络的先验知识,使得网络能够通过这些先验信息更好地推理出最终的预测信息。因此,不管是在单阶段(YOLOv3<sup>[12]</sup>, YOLOv4<sup>[13]</sup>, EfficientDet<sup>[14]</sup>)还是双阶段(Cascade R-CNN<sup>[15]</sup>, Mask R-CNN<sup>[16]</sup>, Libra R-CNN<sup>[17]</sup>)的目标检测算法中,Anchor 都得到了广泛的使用,并且取得了巨大成功。

但是,Anchor 的使用也在一定程度上限制了目标检测技术的发展,它增加了算法的复杂度和计算开销,同时也降低了检测器在不同数据集上的通用性。为了解决引入 Anchor 所带来的问题,学者们开始研究 Anchor Free 目标检测算法,并且取得了较大进展。

当前主流的 Anchor Free 目标检测算法大致分为 3 类:基于关键点的检测算法<sup>[5,18]</sup>、基于 Transformer 的检测算法<sup>[6]</sup>和基于特征点的检测算法。基于关键点的检测算法将目标检测转换为关键点检测问题,网络输出目标的关键点信息以及将这些关键点联系在一起的嵌入向量,从而确定目标的类别和位置。基于 Transformer 的检测算法则充分发挥注意力机制的作用,通过大量编码器和解码器的组合使用,检测出

图片中的目标信息。这两类 Anchor Free 目标检测算法和传统 Anchor Based 目标检测算法在检测机制上存在较大差异,因此在一定程度上避免了划分正负样本时引入模糊样本的问题。正如前文所述,基于特征点的 Anchor Free 目标检测算法和 Anchor Based 目标检测算法都是在一定先验知识的基础上进行分类和回归操作,进而获得预测结果,因此这两种算法都存在类似的模糊样本问题。

## 2.2 软标签、损失函数与权重优化

软标签是神经网络中的常用技术。Zhang 等<sup>[19]</sup>使用标签平滑将分类样本由 0 和 1 组成的硬标签改为接近 0 和 1 的软标签,避免分类网络的过拟合。Li 等<sup>[20]</sup>引入清洁度指标,根据 Anchor 的清洁度设置样本标签,从而降低了噪声样本的预测值。由于清洁度主要根据样本的损失值进行设置,因此该方法本质上是对困难样本和容易样本的信息利用。本文使用 Center Ness 参数来设置软标签,以减少边缘区域模糊样本的干扰,本质上则是对训练样本相对位置信息的融合利用。

目标检测网络的损失函数一般由分类损失函数和边框回归损失函数两部分组成。分类损失函数包括 BCELoss, Focal Loss, DRLoss<sup>[21]</sup>等;边框回归损失函数包括 Smooth L1 Loss<sup>[22]</sup>, IoULoss<sup>[23]</sup>, GIoULoss<sup>[24]</sup>, CIoULoss<sup>[25]</sup>等。

在计算损失的过程中通常伴随着权重优化操作。Focal

Loss 在 BCELoss 的基础上融入了正负样本和难易样本的权重信息,以改善单阶段目标检测中的正负样本不均衡的问题。困难负样本挖掘(Hard Negative Mining<sup>[26]</sup>)技术可被视为将困难负样本和正样本的权重设置为 1,容易负样本的权重设置为 0,通过硬性的权重分配来更好地训练网络。

本文在 Focal Loss 的基础上进行改进,使之更适合计算存在软标签和 Center Ness 加权的分类损失,同时在 CIoU Loss 的基础上添加了 Center Ness 的权重优化操作,以减少模糊样本对回归网络的干扰。

## 3 软标签及样本权重优化方法

### 3.1 检测流程

基于特征点的 Anchor Free 目标检测算法可以视为在待检测的图片上平铺了一系列的特征点(卷积网络在下采样的过程中自动生成),网络对这些特征点进行分类和回归操作以获得最终的检测结果。图 1 给出了算法的整体工作流程。其中  $H_L$  和  $W_L$  分别表示特征金字塔第  $L$  层特征图的高和宽,  $N_L$  表示第  $L$  层特征图包含的特征点总数(即  $N_L = H_L \times W_L$ ),  $C$  表示检测的类别总数,  $l, t, r, b$  分别指特征点距真实框左、上、右、下边界的距离,  $N$  表示所有特征点总数(即各层特征点总数之和)。

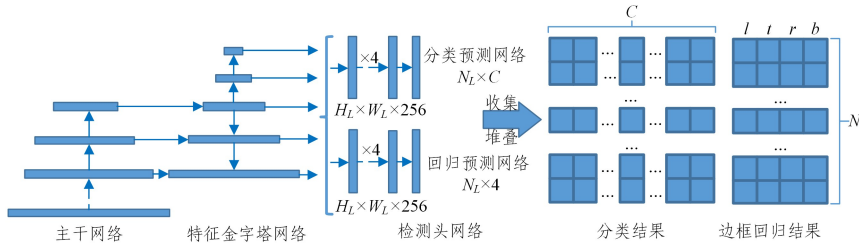


图 1 经典 Anchor Free 目标检测的整体流程

Fig. 1 Overall process of typical Anchor Free object detection

在输入待检测的图片后,算法首先会将图片送进主干网络(常见的主干网络包括 ResNet<sup>[27]</sup>, ResNeXt<sup>[28]</sup>, EfficientNet<sup>[29]</sup>等,本文实验选用 ResNet50)提取特征;然后选择特定的特征图送进特征金字塔网络(常见的特征金字塔网络包括 FPN<sup>[30]</sup>, BiFPN<sup>[14]</sup>, NasFPN<sup>[31]</sup>等,本文实验选用 FPN)进行特征融合,并生成特征金字塔;最后将特征金字塔送进检测头网络,输出金字塔各层的检测结果。为了方便运算,网络会将特征金字塔各层的检测结果进行汇总,通过堆叠获得  $N \times C$  的分类向量和  $N \times 4$  的边框回归向量。后续的损失计算(训练阶段)和非极大值抑制(推理阶段)等操作都将在分类向量和边框回归向量的基础上进行。

### 3.2 Center Ness 软标签

分类网络的输出经过处理后最终会转化为  $N \times C$  的特征向量,  $N$  表示特征点个数,  $C$  代表待检测的类别数。每一个特征点对应一个训练样本,令第  $i$  个特征点对应的训练样本为  $cls_i$ , 则  $cls_i$  是一个长度为  $C$  的一维向量。

当使用硬标签设置  $cls_i$  时,  $cls_i$  表示样本  $i$  是否对应一个真实框,其设置规则为:样本  $i$  是负样本,则  $\|cls_i\|_{L1} = 0$ ;样本  $i$  是正样本且对应的真实框类别为  $c$ , 则  $\|cls_i\|_{L1} = 1$ , 且

$cls_i[c] = 1$ 。训练时,中心区域样本和边缘区域的模糊样本具有同等地位,在训练后期,模糊样本产生较大的损失,使得网络开始将注意力转向模糊样本,以寻求全局最优。

本文旨在使网络在训练阶段将注意力尽可能地集中在中心区域样本上,同时,在推理阶段,中心区域样本输出的目标置信度要大于模糊样本。因此,本文将根据特征点与目标框中心点的相对位置设置  $cls_i$ 。

本文使用 Center Ness 参数表示特征点与目标框中心点之间的相对位置,令  $center\_ness_i$  表示第  $i$  个特征点的 Center Ness 取值,当样本  $i$  为负样本时,  $center\_ness_i = 0$ ;当样本  $i$  为正样本时,令  $l, t, r, b$  分别表示特征点距真实框左边界、上边界、右边界和下边界的距离,  $center\_ness_i$  的计算式如式(1)所示:

$$center\_ness_i = \sqrt{\frac{\min(l, r) \times \min(t, b)}{\max(l, r) \times \max(t, b)}} \quad (1)$$

因为正样本必然位于目标框的内部,所以  $center\_ness_i$  是一个取值范围在 0~1 之间的实数,并且特征点对应位置与真实框中心点距离越近,  $center\_ness_i$  越接近 1,当两点位置重合时,  $center\_ness_i = 1$ ;特征点对应位置与真实框中心点距离越远,  $center\_ness_i$  越接近 0,当特征点对应位置落在边框上时,

$center\_ness_i = 0$ 。使用 Center Ness 参数设置  $cls_i$  的规则为: 样本  $i$  是负样本, 则  $\|cls_i\|_{L1} = 0$ ; 样本  $i$  是正样本且对应真实框类别为  $c$ , 则  $\|cls_i\|_{L1} = center\_ness_i$ , 且  $cls_i[c] = center\_ness_i$ 。

使用 Center Ness 参数设置软标签, 使得正样本与负样本之间存在一个平缓的过渡过程, 而且在训练阶段能够起到平滑标签的作用, 减少网络的过拟合, 同时避免训练后期模糊样本出现太大的损失。在推理阶段, 网络对中心区域样本输出的置信度预测值较大, 对模糊样本输出的置信度预测值较小, 通过 NMS 操作即可进一步减少模糊样本生成的质量较差的预测框。

### 3.3 分类损失函数

基于特征点的 Anchor Free 目标检测算法大多会采用 Focal Loss 来解决训练过程中正负样本不均衡的问题。针对多分类的 Focal Loss 计算式如式(2)~式(5)所示:

$$pt = C \times \hat{C} + (1 - C) \times (1 - \hat{C}) \quad (2)$$

$$bce\_loss = -C \times \log(\hat{C}) - (1 - C) \times \log(1 - \hat{C}) \quad (3)$$

$$\alpha_f = \alpha_1 \times C + (1 - C) \times (1 - \alpha_1) \quad (4)$$

$$FL(C, \hat{C}) = bce\_loss \times \alpha_f \times (1 - pt)^{\gamma_1} \quad (5)$$

其中,  $C$  表示分类的目标向量,  $\hat{C}$  表示分类网络输出的预测向量,  $\alpha_1$  和  $\gamma_1$  是 Focal Loss 的超参数。  $bce\_loss$  是标准的交叉熵损失函数 (BCELoss),  $\alpha_f$  主要负责平衡正负样本之间的权重,  $(1 - pt)^{\gamma_1}$  主要负责抑制容易样本的权重。在使用软标签后, 分类问题转换成概率预测的回归问题, 正样本的目标值不再是 1, 而是 0~1 之间的实数。此时,  $C$  和  $1 - C$  无法准确地起到指示函数的作用,  $\log(\hat{C})$  和  $\log(1 - \hat{C})$  也无法很好地衡量目标向量和预测向量之间的损失, 因此纯粹用于分类的 BCE-Loss 不再适用; 同理,  $\alpha_f$  无法很好地平衡正负样本的权重,  $(1 - pt)^{\gamma_1}$  也无法很好地达到困难样本挖掘的效果。

出现上述情况的原因在于, 在计算损失的过程中, Focal Loss 只使用预测向量参与计算损失, 目标向量仅作为指示函数使用。为使软标签发挥其应有的作用, 本文在 Focal Loss 的基础上设计了新的损失函数 Soft Focal Loss (SFL)。Soft Focal Loss 统一了正负样本的损失计算过程, 能够更加明确地度量目标向量和预测向量之间的差异, 同时能够处理好正负样本不均衡的问题, 其计算式如式(6)~式(9)所示:

$$spt = |C - \hat{C}| \quad (6)$$

$$\log\_loss = -\log(1 - spt) \quad (7)$$

$$cnw = \alpha_2 \times cn + 1 \quad (8)$$

$$SFL(C, \hat{C}, cn) = \log\_loss \times cnw \times spt^{\gamma_2} \times \beta \quad (9)$$

其中,  $cn$  表示训练样本对应的 *centerness* 向量, 根据式(1)计算得出,  $\alpha_2$ ,  $\gamma_2$  和  $\beta$  是超参数。式(6)中,  $spt$  是用来度量目标向量和预测向量之间差异的向量, 向量中元素的取值范围在 0~1 之间, 预测值和目标值的差异越大, 元素值越接近 1; 差异越小, 元素值越接近 0。使用  $\log\_loss$  根据  $spt$  计算软标签的损失, 一方面可以避免使用指示函数, 统一正负样本的计算过程; 另一方面, 由于预测值和目标值都参与了最终的损失

运算, 因此能够更加明确地度量目标向量和预测向量之间的距离。此外, 当采用硬编码时,  $\log\_loss$  和  $bce\_loss$  等价。

$spt^{\gamma_2}$  主要用来处理困难样本和容易样本之间的权重分配问题, 原理和 Focal Loss 相同。容易样本对应的  $spt$  取值接近 0, 困难样本对应的  $spt$  取值接近 1, 当  $\gamma_2$  的取值大于 1 时, 容易样本的  $spt^{\gamma_2}$  将会进一步趋近于 0。Soft Focal Loss 通过  $spt^{\gamma_2}$  来减小容易样本的权重, 从而达到挖掘困难样本的效果。此外, 当采用硬标签时,  $spt^{\gamma_2}$  和 Focal Loss 中的  $(1 - pt)^{\gamma_1}$  等价。实验中,  $\gamma_2$  的取值固定为 2.0。

$cnw$ (center ness weight)的作用有两个, 一是通过参数  $\alpha_2$  平衡正负样本之间的权重, 使得正样本的权重在  $1 \sim \alpha_2 + 1$  之间变化, 负样本的权重保持为 1, 在功能上和 Focal Loss 中的  $\alpha_1$  类似。本文使用线性函数设置  $cnw$ , 目的是让 Soft Focal Loss 和 Focal Loss 尽可能地保持一致。当不采用 Center Ness 进行加权操作时, 适当调整 SFL 的参数后,  $cnw$  和  $\alpha_f$  在计算上等价。实验中, 本文将  $\alpha_2$  固定为 3.0。  $cnw$  的第二个作用是在正样本内部设置权重, 这点将在下一节中进行介绍。

最后, 为了避免使用软标签导致计算的分类损失过低, 使得分类损失和边框回归损失之间存在不平衡, 本文设置超参数  $\beta$  来放大分类损失, 实验中  $\beta$  的取值固定为 3.0。从总体来看, 如果不使用软标签和 Center Ness 加权操作, 那么在适当调整参数后, Soft Focal Loss 将和 Focal Loss 等价。

### 3.4 Center Ness 加权

$cnw$  是线性函数, 配合 Center Ness 参数的性质, 即可令正样本的权重由中心区域向边缘区域逐渐下降, 使得检测器在训练期间将注意力更多地集中在中心区域样本上。同时, 正样本的最低权重为 1, 和负样本权重相同, 保证了正样本到负样本之间权重变化的连续性。此外, 本文使用软标签配合 Soft Focal Loss 进行损失计算, 在一定程度上改善了模糊样本在训练后期转化为困难样本的问题, 使得 Center Ness 权重优化在训练期间能够持续发挥作用。

目标检测网络主要由分类网络和回归网络两个部分组成, 而模糊样本同样会影响回归网络的训练效果, 因此本文将  $cnw$  的加权操作同时用于回归网络的损失计算, 在加速网络的收敛的同时, 迫使网络更加关注中心区域样本的边框预测质量, 模糊样本生成的低质量预测框最终会被 NMS 操作抑制。边框回归损失函数如式(10)所示:

$$Loss_{reg} = RegLoss(\mathbf{R}, \hat{\mathbf{R}}) \times cnw \quad (10)$$

其中,  $\mathbf{R}$  是回归网络的目标向量,  $\hat{\mathbf{R}}$  是回归网络的预测向量,  $RegLoss$  表示通用的边框回归损失函数, 实验使用 CIoU Loss 进行计算。

## 4 实验

### 4.1 实验数据与评估指标

Pascal VOC 和 MS COCO 是目标检测领域最常用的两个标准数据集。Pascal VOC 数据集包含 Pascal VOC 2012 和 Pascal VOC 2007 两个版本, 需要检测 20 种不同类别的目标。MS COCO 数据集包含 MS COCO 2014 和 MS COCO 2017

两个版本,需要检测 80 种不同类别的目标。本文在 Pascal VOC 和 MS COCO 两个数据集上分别进行了对比实验。在 Pascal VOC 数据集上,本文将 Pascal VOC 2007 和 Pascal VOC 2012 的训练集合并后进行训练,并使用 Pascal VOC 2007 的测试集验证算法效果,训练集包含 21 667 张图片,测试集包含 4 952 张图片。在 MS COCO 数据集上,本文直接使用 MS COCO 2017 的训练集进行训练,使用 MS COCO 2017 的测试集验证算法效果,训练集包含 118 287 张图片,测试集包含 5 000 张图片。

目标检测算法通常使用平均准确率(Average Precision, AP)指标对检测效果进行评估。设置判定阈值  $t$ ,预测框与真实框之间的 IoU 大于  $t$  的预测框被视为真阳性样本(TP),IoU 小于  $t$  的预测框被视为假阳性样本(FP),漏检的真实框被视为假阴性样本(FN)。目标检测中,准确率(Precision,  $P$ )和召回率(Recall,  $R$ )的计算式如式(11)和式(12)所示:

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

根据准确率和召回率绘制 PR 曲线,平均准确率 AP 即为 PR 曲线与坐标轴之间的面积。

Pascal VOC 数据集分别计算 20 个不同类别目标的 AP 值(判定阈值  $t$  为 0.5),然后取平均值( $mAP$ )作为最终的评估指标。MS COCO 数据集分别计算判定阈值  $t$  在 0.5, 0.55, ..., 0.95 下所有类别目标的  $mAP$  值,然后取这些  $mAP$  值的平均值作为最终的评估指标( $AP_{50-95}$ )。此外,为了充分表现检测器的性能,MS COCO 数据集的评估指标还给出了判定阈值  $t$  为 0.5 的  $mAP$  值( $AP_{50}$ )、判定阈值为 0.75 的  $mAP$  值( $AP_{75}$ )、小目标(目标面积小于  $32^2$ )的  $AP_{50-95}$  值( $AP_S$ )、中目标(目标面积在  $32^2 \sim 96^2$  之间)的  $AP_{50-95}$  值( $AP_M$ )、大目标(目标面积大于  $96^2$ )的  $AP_{50-95}$  值( $AP_L$ )。

## 4.2 实验设置与实验环境

本文实验均在 FCOS<sup>[9]</sup> 检测器的基础上进行,受计算资源的限制,本文对输入图片的大小、Batch Size 等超参数进行了修改,为不失公平性,所有实验的超参数完全一致,除特别声明外,本文其他超参数与 FCOS<sup>[9]</sup> 保持一致。检测器输入图片的大小为  $896 \times 896$ (图片等比例缩放,较短边使用随机的纯色进行填充,避免对图片进行拉伸而造成失真),数据增强技术包括通用的左右翻转、颜色抖动和空间抖动。检测器的主干特征提取网络选用在 ImageNet 上预训练过的 Res-Net50,特征融合网络选用 FPN。对于每张输入的图片,检测器将会输出 16 709 个特征点信息进行分类和回归操作。

由于 FCOS 的采样规则限制了参与训练的正样本数量(大目标只对固定大小的中心区域进行采样,正样本数量太少),导致软标签无法起到很好的正负样本过渡功能,因此本文方法使用 FoveaBox 的中心采样(Center Sample, CS)规则,中心采样区域的阈值为 0.7。

网络的训练主要由两个阶段组成。第一阶段,冻结主干特征提取网络的参数,训练 25 个 Epoch(MS COCO 上为 5 个

Epoch),Batch Size 为 16,使用余弦退火学习率衰减算法将学习率由  $1 \times 10^{-3}$  逐步衰减到  $1 \times 10^{-4}$ 。第二阶段,解冻主干特征提取网络的参数,训练 50 个 Epoch(MS COCO 上为 18 个 Epoch),Batch Size 为 8,使用余弦退火学习率衰减算法将学习率由  $5 \times 10^{-4}$  逐步衰减到  $1 \times 10^{-5}$ 。

本文实验均在搭载了 CPU E5-2673 的服务器上进行,操作系统为 Ubuntu 18.04,编程语言 Python 3.7.7,深度学习框架为 pytorch 1.5.0,训练时使用两块 12 GB 的 NVIDIA 1080Ti GPU,对应的 CUDA 版本为 10.2。

## 4.3 对比实验

本文分别使用传统训练方法和本文训练方法在 Pascal VOC 数据集和 MS COCO 数据集上训练 FCOS 检测器。由于本文方法更换了 FCOS 的采样规则,为避免采样规则的变化对实验结果造成干扰,充分证明本文方法的有效性,本文添加了 FCOS 在中心采样规则下的实验结果作为对照。

FCOS 检测器在 Pascal VOC 数据集上的实验结果如表 1 所列。表中第一行是使用 FCOS 的采样方法和传统训练方法得到的结果,第二行是使用中心采样规则 and 传统训练方法得到的结果,第三行是使用中心采样规则配合本文训练方法得到的结果。可以看出,在使用了中心采样规则后,FCOS 的  $mAP$  下降了 0.71%,但是在使用本文方法进行训练后, $mAP$  得到了提升并最终超过原文方法训练的结果。总体而言,本文方法使得检测器对大多数类别目标的检测效果都有所提升,最终的  $mAP$  比原文方法训练的 FCOS 提升了 1.31%,说明本文方法通过减少模糊样本的干扰,提高了检测器的准确率,证明了本文方法的有效性。

表 1 Pascal VOC 数据集上的训练结果

Table 1 Training results on Pascal VOC dataset

(单位:%)

methods	FCOS	FCOS (CS)	本文方法 (CS)	methods	FCOS	FCOS (CS)	本文方法 (CS)
Plane	88.85	90.18	92.47	Table	71.23	73.75	74.02
Bicycle	86.90	84.82	87.04	Dog	91.33	90.90	91.91
Bird	85.62	84.97	87.02	Horse	85.06	83.69	85.33
Boat	69.04	70.78	72.30	Motorbike	85.86	81.60	86.39
Bottle	70.24	69.08	70.49	Person	86.20	85.30	85.61
Bus	87.31	85.89	90.66	Plant	54.79	54.28	57.20
Car	90.21	90.07	90.58	Sheep	86.58	85.73	88.10
Cat	93.16	93.98	94.00	Sofa	69.71	70.19	73.59
Chair	58.06	56.64	57.37	Train	88.90	86.98	89.64
Cow	86.71	84.89	89.09	Tv	81.12	86.98	80.33
				$mAP$	80.85	80.14	82.16

图 2 给出了 FCOS 在 Pascal VOC 数据集上训练的过程中,检测器准确率的变化曲线。可以看出,在冻结主干网络的预训练阶段(前 25 个 Epoch),本文方法可以使得网络更快收敛,主要原因是利用 Center Ness 参数进行权重优化后,中心区域样本得到了更多的关注,比模糊样本更快地进行拟合,并且,软标签使这部分样本拥有更高的置信度,使得这部分高质量的预测框在 NMS 操作中得以保留,最终使检测器在前期取得了更好的检测效果。在解冻主干网络之后(后 50 个 Epoch),本文方法最终稳定在一个更好的检测结果上,主要原因是本文方法在训练后期仍然能够将训练的重点

保持在中心区域样本上,在训练阶段和推理阶段同时减少了边缘区域模糊样本的干扰,因此提高了检测器的整体检测效果。

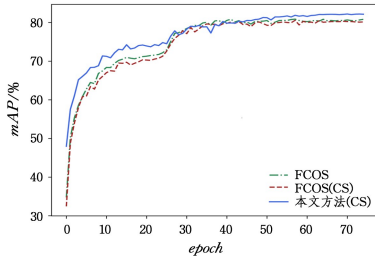


图2 检测器 mAP 的变化曲线

Fig. 2 Change curve of mAP of detector

FCOS 检测器在 MS COCO 数据集上的训练结果如表 2 所列。可以看出,在使用了中心采样规则后,FCOS 的  $AP_{50-95}$  下降了 0.4%,但是在配合本文方法进行训练后, $AP_{50-95}$  获得了提升并超过基线方法 1.3%,整体情况与 Pascal VOC 数据集上的表现类似,在证明了本文方法的有效性的同时,也表明本文方法能够使 FCOS 在不同数据集上的检测效果获得一致提升。

表 2 MS COCO 数据集上的训练结果

Table 2 Training results on MS COCO dataset (单位:%)

	FCOS	FCOS (CS)	本文方法(CS)
$AP_{50-95}$	34.5	34.1	35.8
$AP_{50}$	54.6	53.6	55.3
$AP_{75}$	36.2	35.9	38.4
$AP_S$	18.5	16.8	19.2
$AP_M$	38.4	38.6	40.5
$AP_L$	45.4	45.1	48.1

此外,本文方法的  $AP_{50}$  指标提升了 0.7%, $AP_{75}$  指标提升了 2.2%,表明本文方法在不同判定阈值下的指标均有提升,且在较为严格的指标上,提升效果更为明显。其原因是本文方法迫使网络更加关注中心区域样本,而这部分样本产生的预测框大多质量较高,因此检测器最终的定位精度得到了一定的提升。

本文方法的  $AP_S$  指标提升了 0.7%, $AP_M$  指标提升了 2.1%, $AP_L$  指标提升了 2.7%,表明本文方法在不同尺度目标的检测上,效果均有提升,且对于中大型目标的提升效果更为明显,但是对于小目标的检测效果提升较差。其主要原因是中心采样方法使得小目标的召回率受到影响,也说明本文方法更适合对大中型目标的检测。

为了进一步证明本文方法的有效性,本文将本文方法训练的 FCOS 与其他主流的目标检测器进行了横向对比,实验数据如表 3 所列。除特别说明外,表中所有检测器的网络配置均保持一致,主干特征提取网络为 ResNet50(YOLOv3 的主干网络为 DarkNet53),特征融合网络为 FPN。表 3 中的数据表明,与这些主流的目标检测算法相比,本文方法在 Pascal VOC 数据集上取得了最优的检测效果,在 MS COCO 数据集上取得了较好的检测效果,证明了本文方法的有效性。

表 3 本文方法与其他主流检测器的效果对比

Table 3 Comparison between the proposed method and other main stream detector (单位:%)

Algorithm	Anchor Free	Pascal VOC (mAP)	MS COCO ( $AP_{50-95}$ )
Faster RCNN	否	77.77	36.5
YOLOv3	否	80.25	33.0
RetinaNet	否	75.50	35.7
FoveaBox	是	76.6	36.4
FCOS	是	80.85	34.5
本文方法	是	82.16	35.8

图 3 给出了传统方法训练的 FCOS 和本文方法训练的 FCOS 对指定图片进行目标检测的结果,其中,第 1 和第 3 行是传统方法训练后的检测结果,第 2 和第 4 行是本文方法训练后的检测结果。可以看出,本文方法通过缓解模糊样本的干扰,有效地减少了检测器错检、漏检的情况,同时对检测器的定位精度也有一定的提升。

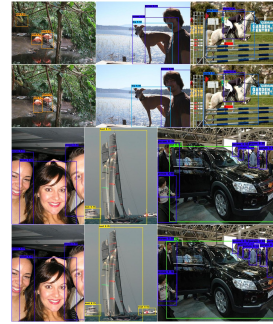


图 3 FCOS 与本文方法的检测结果对比

Fig. 3 Comparison of detection results between FCOS and the proposed method

#### 4.4 损失的归一化分布

为了直观地展现本文训练方法与传统训练方法之间的本质区别,本文分别使用这两种方法对 FCOS 目标检测器在 Pascal VOC 数据集上进行预训练(即在冻结主干网络的情况下训练 25 个 Epoch),然后使用预训练的网络对指定图片进行目标检测,并计算网络输出结果的归一化损失,实验结果如图 4 所示。图中,左侧是待检测的图片,中间是两种训练方法获得的检测器的检测结果,右边的概率热图表示样本损失归一化后的相对大小,颜色越亮,代表损失相对越大。第一行是使用原文方法训练的结果( $mAP$  为 72.19%),第二行是使用本文方法训练的结果( $mAP$  为 74.82%)。



图 4 使用不同方法训练 FCOS 的归一化损失分布

Fig. 4 Normalized loss distribution of FCOS with different train methods

可以看出,在网络预训练的后期,两种训练方法得到的检测器都能够很好地完成目标检测任务,但是在传统的训练

方法中,检测效果较差的模糊样本产生的损失较大,因此网络开始将注意力集中在模糊样本上,期望所有的训练样本都拥有较好的检测结果,力求达到全局最优。而在使用本文方法进行训练后,中心区域样本产生的损失仍然大于边缘区域的模糊样本,说明网络的训练重点仍然保持在中心区域的样本上,从而减少了模糊样本在训练阶段对检测器的干扰。

#### 4.5 消融实验

本节将通过 2 组消融实验分别证明共同使用软标签、Soft Focal Loss 和权重优化的必要性,以及权重优化方法对分类网络和回归网络的性能的影响。实验均在 FCOS 目标检测器的基础上进行,数据集为 Pascal VOC,采样规则为中心采样(采样区域为 0.7),评估指标为 mAP。

图 5 给出了使用传统方法训练、仅添加软标签进行训练、仅添加 Center Ness 权重优化进行训练以及同时使用软标签、Center Ness 权重优化和 Soft Focal Loss 进行训练的过程中,检测器准确率的变化曲线。

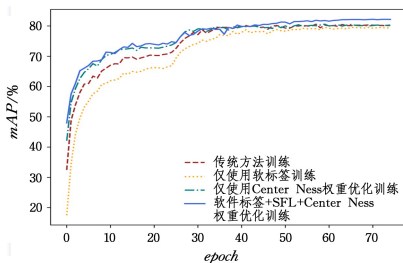


图 5 不同方法训练 FCOS 的过程对比

Fig. 5 Comparison of FCOS training process with different methods

当仅使用软标签进行训练时,检测器前期的收敛速度和最终的准确率均低于传统方法训练的结果(基线方法),主要原因是使用了软标签后,分类网络负责的不再是单纯的分类任务,而是一种概率的回归任务。由于分类网络训练的目标向量不是标准的独热编码,因此常用的分类损失函数 Focal Loss 不再适用,最终导致检测器的整体检测性能受到影响。

当仅使用 Center Ness 权重优化进行训练时,检测器的前期收敛速度明显快于基线方法,但是最终的准确率却和基线方法相差很小。其原因是 Center Ness 权重优化使得网络在训练前期能够将注意力集中于中心区域样本,减少了边缘区域模糊样本的影响,但是在训练后期,模糊样本因产生了较大损失而变成困难样本,Focal Loss 对困难样本分配了更大的权重,使得 Center Ness 加权难以发挥应有的效果。

当同时使用软标签、权重优化和 Soft Focal Loss 时,检测器前期的收敛速度和最终的检测效果均优于基线方法,主要原因是本文的训练方法能够使网络在训练过程中始终关注着中心区域样本。在训练前期,中心区域样本能更早收敛并拥有更高的置信度,因此网络的整体效果提升更快;在训练后期,边缘区域模糊样本不会产生太大损失,也就不会干扰检测器的整体性能,因此最终的检测效果更好。

表 4 列出了将 Center Ness 权重优化分别作用于分类网络和回归网络的实验结果,表中“是”表示对该网络使用了权重优化,“否”表示未对该网络使用权重优化。实验均在使用了 Soft Focal Loss 和 Center Ness 软标签的 FCOS 检测器上进行,除是否使用 Center Ness 加权操作外,其他配置均保持一致。

表 4 的实验结果表明,将 Center Ness 权重优化操作分别作用于分类网络和回归网络均能在一定程度上提升检测性能,说明权重优化确实可以减小模糊样本对分类网络和回归网络的干扰。因此,当分类网络和回归网络共同使用 Center Ness 权重优化时,检测器的性能得到了进一步的提升。

表 4 权重优化分别作用于分类网络和回归网络的实验结果对比  
Table 4 Experimental results comparison of weight optimization on classification network and regression network respectively

(单位:%)			
分类网络	回归网络	预训练结果	训练结果
否	否	72.10	80.96
是	否	73.44	81.55
否	是	73.26	81.23
是	是	74.82	82.16

此外,将 Center Ness 权重优化操作单独作用于分类网络比将其单独作用于回归网络获得了更大的收益,这是因为权重优化在回归网络中仅需负责减少模糊样本的干扰,而在分类网络中还需负责平衡正负样本的权重,这说明了权重优化对分类网络来说更为重要。

**结束语** 本文充分利用 Center Ness 参数,在软标签、损失函数和权重优化 3 个方面对基于特征点的 Anchor Free 目标检测算法进行了改进,以改善根据某一阈值划分非正即负的训练样本时存在的模糊样本问题。本文方法的核心思想是将网络关注的重心集中在真实框中心区域样本,以改善模糊样本的问题。本文使用 Center Ness 参数设置软标签和进行权重优化是为了迫使检测器在训练阶段始终关注目标框的中心区域,而使用 Soft Focal Loss 计算损失则是为了保证使用了软标签的分类网络能够很好地收敛,且 Center Ness 的加权操作能够持续发挥作用。

在 Pascal VOC 数据集上,本文的训练方法使得 FCOS 对多个类别的预测准确率都有了明显提升,且最终的 mAP 指标提升至 82.16%(提升了 1.31%)。在 MS COCO 数据集上,本文方法在不同的评估指标上均有提升,且最终的 AP<sub>50-95</sub> 指标提升至 35.8%(提升了 1.3%)。实验证明了本文方法的有效性。

从 MS COCO 数据集上的实验结果来看,本文方法对小目标的检测效果提升较小,分析认为主要原因是中心采样方法降低了小目标的召回率,使得检测器对小目标的检测受到影响。因此,使用更加合理的采样方法将有可能进一步提升目标检测器效果。接下来的工作会在采样方法上进行更多的研究,以进一步提高算法的检测精度。

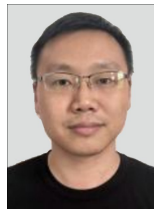
#### 参考文献

- [1] CHEN K Q, ZHU Z L, DENG X M, et al. Deep Learning for Multi-Scale Object Detection: A survey[J]. Journal of Software, 2021, 32(4): 1201-1227.
- [2] LI S P, LI C L, HAN J B, et al. Application of Binocular Vision Single Step Multi-target Detection Method for Robot Grasping [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2021, 38(5): 68-74.
- [3] XUAN D D, WANG J, WANG Z. Salient target detection based on high-level priori semantics [J]. Journal of Chongqing Univer-

- sity of Posts and Telecommunications(Natural Science Edition), 2020,32(2):304-312.
- [4] 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. 2019:9626-9635.
- [5] 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.
- [6] CARION N, MASSA F, SYNNAEVE G, et al. End-to-End Object Detection with Transformers [C] // European Conference on Computer Vision. Cham: Springer, 2020:213-229.
- [7] ZHANG S, CHI C, YAO Y, et al. Bridging the Gap Between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:9756-9765.
- [8] 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. 2017:2999-3007.
- [9] TIAN Z, SHEN C, CHEN H, et al. FCOS: A Simple and Strong Anchor-Free Object Detector[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 1922-1933.
- [10] KONG T, SUN F, LIU H, et al. FoveaBox: Beyond Anchor-Based Object Detection [J]. IEEE Transactions on Image Processing, 2020, 29: 7389-7398.
- [11] ZHU C, CHEN F, SHEN Z, et al. Soft Anchor-Point Object Detection[C] // European Conference on Computer Vision. Cham: Springer, 2020:91-107.
- [12] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement [J]. arXiv:1804.02767, 2018.
- [13] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection [J]. arXiv:2004.10934, 2020.
- [14] TAN M, PANG R, LE Q V. EfficientDet: Scalable and Efficient Object Detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:10778-10787.
- [15] CAI Z, VASCONCELOS N. Cascade R-CNN: Delving Into High Quality Object Detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:6154-6162.
- [16] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C] // Proceedings of the IEEE International Conference on Computer Vision. 2017:2980-2988.
- [17] PANG J, CHEN K, SHI J, et al. Libra R-CNN: Towards Balanced Learning for Object Detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:821-830.
- [18] LAW H, DENG J. CornerNet: Detecting Objects as Paired Keypoints [C] // Proceedings of the European Conference on Computer Vision(ECCV). 2018:765-781.
- [19] ZHANG Z, HE T, ZHANG H, et al. Bag of Freebies for Training Object Detection Neural Networks [J]. arXiv:1902.04103, 2019.
- [20] LI H, WU Z, ZHU C, et al. Learning From Noisy Anchors for One-Stage Object Detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:10585-10594.
- [21] QIAN Q, CHEN L, LI H, et al. DR Loss: Improving Object Detection by Distributional Ranking [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:12161-12169.
- [22] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [23] YU J, JIANG Y, WANG Z, et al. UnitBox: An Advanced Object Detection Network [C] // Proceedings of the 24th ACM International Conference on Multimedia. 2016:516-520.
- [24] REZATOFI GHI H, TSOI N, GWAK J Y, et al. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:658-666.
- [25] ZHENG Z, WANG P, LIU W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2020:12993-13000.
- [26] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector [C] // European Conference on Computer Vision. Cham: Springer, 2016:21-37.
- [27] 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.
- [28] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated Residual Transformations for Deep Neural Networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:1492-1500.
- [29] TAN M, LE Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks [C] // International Conference on Machine Learning. PMLR, 2019:6105-6114.
- [30] 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.
- [31] GHIASI G, LIN T Y, LE Q V. NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:7036-7045.



**WANG Can**, born in 1994, postgraduate. His main research interests include object detection and so on.



**XIE Qing**, born in 1986, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include information retrieval and machine learning.