

一种基于锚框质量分布的动态标签分配策略

王璐璐^{1,2}, 陈东方^{1,2}, 王晓峰¹

(1. 武汉科技大学 计算机科学与技术学院, 武汉 430065;

2. 武汉科技大学 智能信息处理与实时工业系统湖北省重点实验室, 武汉 430065)

摘要: 传统的标签分配策略根据锚框与 ground-truth 之间的交并比(IoU)是否超过阈值来划分正负样本,但确定 IoU 阈值需要依靠大量的经验和实验,并且在不同的网络结构中需要重新设定新的阈值;其次固定的阈值无法保证正负样本数量变化的平缓,过多或过少的正样本都将影响网络训练的稳定性。为解决这些问题,提出一种自适应标签分配策略。根据中心先验原则设计中心权重,通过联合分类得分和定位得分表示锚框质量,组成一个统一的锚框评分方案便于简化划分步骤。为了使网络能够根据训练程度自适应调整 IoU 阈值,保证网络训练的稳定性,利用锚框质量高斯函数模拟锚框总体质量得分的分布情况,使用极大似然估计算法获取最优参数,根据统计结果动态预测最佳 IoU 阈值。实验结果表明,该算法在基于锚框和基于无锚框的检测算法中均使检测性能得到有效提升,在 PASCAL VOC 数据集上分别提升 3.1 和 6.6 个百分点,并且可以有效降低漏检率。

关键词: 标签分配;目标检测;深度学习;样本;交并比

开放科学(资源服务)标志码(OSID):



中文引用格式: 王璐璐,陈东方,王晓峰.一种基于锚框质量分布的动态标签分配策略[J].计算机工程,2023,49(4): 85-91,100.

英文引用格式: WANG L L, CHEN D F, WANG X F. A dynamic label assignment strategy based on quality distribution of anchor[J]. Computer Engineering, 2023, 49(4): 85-91, 100.

A Dynamic Label Assignment Strategy Based on Quality Distribution of Anchor

WANG Lulu^{1,2}, CHEN Dongfang^{1,2}, WANG Xiaofeng¹

(1. School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China;

2. Hubei Province Key Laboratory of Intelligent Information Processing and Real-Time Industrial System,

Wuhan University of Science and Technology, Wuhan 430065, China)

[Abstract] Traditional label assignment strategies divide positive and negative samples according to whether the Intersection over Union(IoU) between the anchor and ground-truth exceeds the threshold. However, determining the IoU threshold requires considerable experience and experimentation, and new thresholds must be set for different network structures. Moreover, a fixed threshold cannot guarantee the consistent partitioning of positive and negative samples, and too many or too few positive samples will affect the reliability of network training. To solve these problems, we propose an Adaptive Label Assignment Strategy (ALAS). First, the center weights are designed according to the center prior principle; the anchor quality is represented by the joint classification and localization scores; and a unified anchor scoring scheme is developed to simplify the partition. To enable the network to adaptively adjust the IoU threshold according to the degree of training and ensure the consistency of network training, the anchor quality Gaussian function is used to simulate the distribution of the overall quality score of the anchor. The maximum likelihood estimation algorithm is used to obtain optimal parameters, and then the optimal IoU threshold is dynamically predicted based on statistical results. The experimental results show that the algorithm can effectively improve both anchor-based and non-anchor-based detection algorithms, and it improves these algorithms by 3.1 and 6.6 percentage points, respectively, on the PASCAL VOC dataset. Therefore, it can effectively reduce missed detection rates.

[Key words] label assignment; object detection; deep learning; sample; Intersection over Union(IoU)

DOI: 10.19678/j.issn.1000-3428.0064346

基金项目: 国家自然科学基金(61572381, 61273225)。

作者简介: 王璐璐(1999—),女,硕士,主研方向为深度学习、目标检测;陈东方,教授、博士;王晓峰,副教授、博士。

收稿日期: 2022-03-31 **修回日期:** 2022-05-17 **E-mail:** 2387013946@qq.com

0 概述

目标检测是计算机视觉领域中一个重要的研究方向,其在人脸识别、无人驾驶、医学图像处理等方面都有广泛的应用。在目标检测过程中,一张图片中会产生成千上万的锚框,而待检测目标可能只有几个,所以需要为锚框匹配相应的目标框,让网络知道每个锚框需要预测的目标是哪个,为锚框匹配 ground-truth 的过程就是标签分配。标签分配的结果决定了模型的学习目标,进而决定模型的性能^[1]。检测算法主要依据交并比(Intersection over Union, IoU)^[2]的大小划分正负样本,但是如果设置比较低的 IoU 阈值,会得到过多的噪声样本,导致网络的识别能力降低,影响检测精度,而如果 IoU 阈值设置得过高,则只能保留到高质量样本,过滤掉太多中等质量的样本,可能会产生过拟合现象^[3]。设置合理的 IoU 阈值能够提高检测精度,避免过拟合问题。

传统的标签分配是基于交并比的手工分配策略,如需要手动调整 IoU 阈值、提前设定锚框等。近年来关于标签自动化分配的方案越来越多:经典的基于锚框的算法 RetinaNet^[4]根据锚框和 ground-truth 之间的交并比确定正负样本;FreeAnchor^[5]在它的基础上,通过极大似然估计的思想构建损失函数去除人工参与指定锚框的过程,使网络能够自主匹配锚框和标签框;基于无锚框结构的 FCOS^[6]网络根据目标中心区域和物体的尺度范围确定正负样本;在 FCOS 的基础上,ATSS^[7]通过自适应训练样本选择的方法实现正负样本划分;Autoassign^[8]跳出以往非正既负的分配方式和监督原则,认为每个位置都有正样本属性和负样本属性,在优化过程中,这些样本将会同时受到正样本和负样本的监督,利用正负样本权重平衡配比。

虽然当前有很多高效的正负样本划分方法,但 FSAP^[9]、FreeAnchor 等都是通过增加额外的卷积层来划分正负样本,然而改变网络结构将会导致其他检测网络无法直接使用这类划分方法。此外,在训练后期,随着定位能力提高,锚框平均 IoU 的数值增大,这会导致网络分配到的正样本数量大幅增加,进而影响训练的稳定性。而采用 K-means^[10]、混合高斯模型^[11]等聚类方法划分正负样本会提高计算复杂度,降低网络训练测试速度。

针对以上问题,本文提出一个新的自适应标签分配策略(Adaptive Label Assignment Strategy, ALAS)。为统一计算锚框分数,设计中心权重函数限定样本点位于 ground-truth 的中心附近,同时联合分类和定位分数,采用统一的加权机制设计锚框质量得分函数。在此基础上,将标签分配建模为一个概率分布的过程,利用高斯模型计算锚框的整体质量分数,通过最大似然估计获取最优参数,使得模型能够根据锚框得分情况动态学习如何分配标签。最后,在

MS COCO2017^[12]数据集和5个主干网络设置下进行对比实验,验证所提方法的有效性。

1 自适应标签分配策略

自适应标签分配策略设计应考虑以下两点:1)应能够衡量锚框的定位和分类质量;2)将锚框划分为正样本和负样本的过程应是自适应的,这样就不需要人工设计固定的超参数。

1.1 锚框评分方案

首先需要设计一个能够反映网络定位和分类能力的锚框评分方案。图1为 FCOS 网络训练迭代过程中正样本点的可视化结果(彩色效果见《计算机工程》官网 HTML 版本,下同),迭代次数增加后网络定位准确度增加,而这时热力图显示正样本点更多地集中在物体中点附近,可知在训练阶段关注目标框的中心区域可以使检测器达到更好的效果。



图1 迭代过程中置信度权重的演化

Fig.1 Confidence weight evolution during iteration process

利用 L2^[13]损失函数对离群点的敏感性实现上述中心先验的要求,记真实边界框的中心点由 (g_x, g_y) 表示,样本点坐标由 (b_x, b_y) 表示,则本文的中心加权函数如式(1)所示:

$$W_{\text{center}} = e^{-\sum_{i \in \{x, y\}} L_2(b_i - g_i)} \quad (1)$$

在式(1)中, W_{center} 与 L_2 呈负相关,离群点的中心权重相对其他样本点较小。

为了使评分方案能够同时评估定位和分类的质量,可以将定位得分和分类得分相乘。直接将网络中分类分支输出的置信度作为分类得分,用中心权重(式(1))表示定位得分,因为定位分支输出的是偏移量而不是一个分数,所以利用中心权重使评分方案更加简洁。本文将锚框评分方案定义为中心权重与分类得分的乘积,如式(2)所示:

$$Q = W_{\text{center}} \times S_{\text{cls}} = e^{-\sum_{i \in \{x, y\}} L_2(b_i - g_i)} \times S_{\text{cls}} \quad (2)$$

其中: S_{cls} 表示分类得分; W_{center} 是样本点的中心权重。本文取质量函数 Q 的负对数,得到:

$$-\ln Q = -\ln W_{\text{center}} - \ln S_{\text{cls}} = \sum_{i \in \{x, y\}} L_2(b_i - g_i) + L_{\text{cls}} \quad (3)$$

$$Q = e^{-\sum_{i \in \{x, y\}} L_2(b_i - g_i)} - L_{\text{cls}} \quad (4)$$

中心权重可以规范样本中心点的位置,但锚框的尺度对框定范围也有影响,不能仅使用中心权重代表定位质量。如图2(a)和图2(b)所示,当样本点与 ground-truth 的中心点距离很近时,锚框过大或过小都会使定位不准确,然而在这种情况下它们的中心权重很高。在图2(a)和图2(b)中都有较高的中心

权重,但是IoU较低,进一步利用IoU来细致规范锚框的定位就可以将这类锚框从高精度质量锚框的队列中筛选出来,如图2(c)所示,同时在拥有较高的中心权重和IoU时才能高质量地预测到物体的位置。

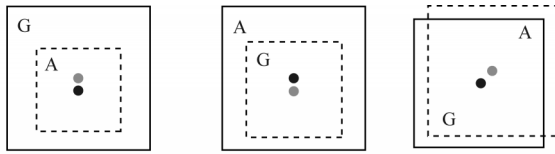


图2 中心权重定位的局限性

Fig.2 Limitations of center weight positioning

具体来说,预测框与对应 ground-truth 框之间的 IoU 不仅用于计算 IoU 损失,还可以作为分类的标签。利用 IoU 指导分类训练,即把 one-hot 分类标签的 0 或 1 替换成对应的 IoU,再计算分类损失,可以将 IoU 作为定位质量使其与分类相结合, IoU 较低即定位能力很差时,它的分类分数也会比较低。进一步将定位与分类相结合使网络可以同时筛选出定位错误但分类正确的边界框和定位准确但分类错误的边界框。

由于将离散的分类标签(0或1)替换成了连续的 IoU,因此需要将离散型的 Focal Loss 改为连续型的质量焦点损失(Quality Focal Loss, QFL):

$$F_{\text{Focal Loss}} = -y \ln y' - (1-y) \ln(1-y') = \begin{cases} -\ln y', y=1 \\ -\ln(1-y'), y=0 \end{cases} \quad (5)$$

$$Q_{\text{QFL}}(p) = L_{\text{cls}} = -\left[|y-p|^{\gamma} y \ln p + |y-p|^{\gamma} (1-y) \ln(1-p) \right] \quad (6)$$

其中: p 为预测结果; γ 的取值为2; y 为标签即样本点与对应 ground-truth 之间的 IoU 值, $y \in [0, 1]$ 。具体来说, $y=0$ 表示定位分数为0的负样本, $0 < y \leq 1$ 表示 IoU 分数为 y 的正样本。在训练过程中,预测框与 ground-truth 框之间的 IoU 得分在 0 到 1 之间动态变化。QFL 依旧保留分类的向量,但是对应类别位置的置信度的物理含义不再是分类得分,而是质量预测的分数。

如表1所示,ALAS算法利用QFL作为分类损失函数比使用Focal Loss提高了0.2个百分点。若使用Focal Loss,算法将按照中心先验原则衡量锚框的定位质量。而QFL联合IoU,使ALAS可以在中心先验的基础上进一步考虑到锚框框定范围的准确性。

表1 Focal Loss和QFL对网络结果的影响

Table 1 Effect of Focal Loss and QFL

on network results		%
计算方式	AP	
Focal Loss		78.6
QFL		78.8

经过优化后,最终的锚框质量评价模块如下:

$$Q = e^{-\sum_{i \in \{x,y\}} L_2(b_i - g_i)} - Q_{\text{QFL}} \quad (7)$$

1.2 正负样本划分策略

单纯的锚框质量评价方案只是一个锚框得分的计算方式,依旧无法动态调整 IoU 阈值。为了动态掌握网络的训练状态,更方便模型根据总体质量得分分布情况划分正负样本,本文使用高斯模型模拟锚框评分的分布情况,引入一个参数可学习的锚框质量高斯函数 G (如式(8)所示),根据不同的数据分布自动调整 IoU 阈值。通过 MLE 最大化锚框的质量得分 Q 的概率计算最优参数。

$$G(Q|x, g, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(Q-\mu)^2}{2\sigma^2}\right) \quad (8)$$

其中: Q 表示式(7)中的质量得分; μ 和 σ 是可学习的参数。一开始,参数 μ 会被初始化为0, $\sigma=1$ 。给定一组锚框得分,可以使用最大似然估计算法优化函数 G 的概率,计算最优参数。本文将参数 μ 和 σ 的总和作为 IoU 阈值。

若基于聚类思想采用混合高斯模型模拟锚框得分情况,则无法使用最大似然估计为其求解,而需要通过迭代的方法划分代表正负样本的两个高斯模型,迭代求解的过程将大大影响网络训练和检测的速度。本文方法可直接利用MLE求解,在参数量、计算复杂度和检测速度上都得到了进一步优化。

算法1描述了所提方法对输入图像的工作原理。对于图像上的每一个 ground-truth 标签框 g , 首先找出它的候选正样本。如算法1第3~7行描述,在金字塔的每一层选取 k 个质量评分最高的锚框。之后根据这些候选框的质量得分计算出锚框质量高斯函数 G , 见算法第9行。在算法第10行,通过极大值似然估计计算出 σ 和 μ 两个参数,根据这些统计数据计算 ground-truth g 的 IoU 阈值 $t_g = \sigma_g + \mu_g$ 。最后只需要根据算法第12~16行选择那些 IoU 大于或等于 IoU 阈值的候选框作为正样本,如果一个锚框被分配给多个 ground-truth,则会选择 IoU 最高的那个,其余都是负样本。

算法1 自适应标签分配策略 ALAS

输入 ground-truth 标签框 G , 所有的锚框 A , 金字塔第 i 层的锚框 A_i , 金字塔层数 L , 每层金字塔选取的候选框数 K

输出 正样本 P , 负样本 N

1. $P \leftarrow \emptyset, N \leftarrow \emptyset, C_g \leftarrow \emptyset$
2. for $g \in G$
3. for $i = 1$ to L do
4. $W_{\text{center}} \leftarrow$ 计算中心权重
5. $Q_i \leftarrow$ 计算锚框质量(A_i, g)
6. $S_i \leftarrow$ 根据 Q_i 选择 K 个候选框
7. $C_g = C_g \cup S_i$
8. end for
9. $G(Q|x, g, \theta) \leftarrow$ 计算 G
10. 通过最大似然估计计算参数 σ_g, μ_g
11. 为 ground-truth g 计算 IoU 阈值: $t_g = \sigma_g + \mu_g$
12. for each candidate $c \in C_g$ do

```
13.if IoU(c, g) ≥ tg
14.P = PUC
15.end if
16.end for
17.end for
18.N = A - P
19.return P, N
```

在算法1中, IoU 阈值的大小是根据模型参数条件自适应确定的, 能够根据质量得分的总体分布动态调整参数和 IoU 阈值, 这与之前的方法^[14-15]有所不同, 文献^[14-15]方法只考虑到锚框的定位能力, 忽略了网络预测能力提升对正负样本划分带来的影响。本文通过 ALAS 联合分类和定位动态模拟质量分布情况自适应调整 IoU 阈值, 改善了检测效果。

2 实验结果与分析

为验证所提方法的有效性, 在 MS COCO 2017 和 PASCAL VOC^[16] 两个公共数据集上进行实验。

MS COCO2017 数据集共包含 80 个类别, 训练集包含约 118 000 张图片, val 集包含约 5 000 张图片, test-dev 集包含约 20 000 张图片。本文在 test-dev 2017 上进行实验并与现有的目标检测算法比较, 然后在 val 集上进行数据分析。

PASCAL VOC 数据集包含 20 个类别, 其中训练图片 22 136 张, 测试图片 4 952 张, 实验结果遵循 VOC 数据集的指标, 其中平均精度表示该类别在 IoU 阈值为 0.5 时的平均精度。同时, 在 VOC 2012 和 VOC 2017 上进行消融实验和对比实验。

2.1 超参数设置

ALAS 只涉及一个 top-K 超参数, 在本节中进行多个实验研究 top-K 的取值对检测器效果的影响。top-K 表示从每个金字塔层级中选择的候选框个数。

表2为 top-K 取值为 9~21 时的效果对比。可以看出: 太小的 K 值会导致准确率明显下降, 因为过少的候选正样本会导致漏检也会造成样本统计不准确; 太大会导致过多低质量的候选框, 造成精度损失。如表2所示, 当 top-K 的取值在 13~21 的范围内时 AP 变化不敏感, 实验结果显示 top-K 是一个稳健的超参数。经过详细比较, 当 K 取值为 17 时效果最好, 达到了 78.8% 的准确率。

表2 ALAS使用不同K值的结果

Table 2 Results of using different number of K on the ALAS

K 值	AP/%
7	77.7
9	77.8
10	78.0
13	78.6
15	78.7
17	78.8
19	78.6
21	78.4

2.2 联合分类损失效果分析

由于本文研究的主要目的是简化锚框分配的步骤, 在实现动态调整阈值的同时, 还需要重点关注模型的大小和训练速度, 为了更好地评价网络性能优劣, 设计实验对比网络的参数量和计算复杂度。

从图3可以看出: 传统检测器 Faster RCNN 的参数量大于其他模型, 而 SSD 算法的计算复杂度远高于其他网络; FreeAnchor 的参数量和时间复杂度是 ALSA 的 1.2 倍, 针对标签分配的改进算法中参数量最少的 PAA 模型的参数量也是 ALSA 的 1.1 倍; 本文方法的总体参数量和计算量都达到了与 ATSS 模型相当的水平。

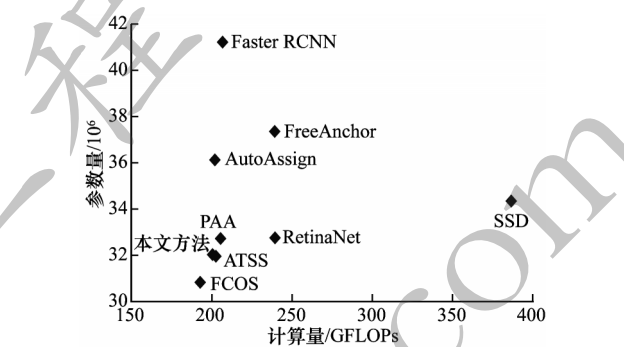


图3 参数量与计算量对比

Fig.3 Comparison of parameters and computation

表3所示为本文方法分别应用在基于有锚框和基于无锚框的目标检测算法的速度。可以看出: 以 FCOS 和 RetinaNet 为基础网络时, 训练过程中每次迭代分别仅增加 0.04 s 和 0.02 s, 而在检测过程中, 检测单张图片分别增加 0.006 s 和 0.005 s, 相应的检测精度分别提高了 6.6 和 3.1 个百分点, 自适应标签分配算法在对训练和检测速度影响细微的情况下, 能有效提高检测的精确度, 提高网络的效率。

表3 ALAS与PAA方法速度与精度对比

Table 3 Speed and accuracy comparison of ALAS and PAA methods

目标检测算法	训练速度 (s·iter ⁻¹)	检测时间/s	AP/%
FCOS	0.36	0.212	72.2
RetinaNet	0.41	0.224	74.3
FCOS+ALAS	0.40	0.218	78.8
RetinaNet+ALAS	0.43	0.229	77.4
FCOS+PAA	0.52	0.267	78.6
RetinaNet+PAA	0.58	0.294	76.2

与采用聚类思想利用高斯混合模型分离正负样本的 PAA 方法相比, 本文的 ALAS 方法简化迭代过程在训练和检测的速度上提升了 23%, 同时在检测的准确度上提高了 0.8 个百分点。

2.3 与主流目标检测网络效果对比

为了验证 ALAS 的有效性, 分别在 MS COCO2017 和 PASCAL VOC(07+12) 数据集上进行训练和测试, 同时与其他目标检测网络进行对比和定量

分析。在 MS COCO 测试子集上对本文模型与其他的目标检测进行比较,如表 4 所示。为了不失公平性,除 top- K 外,本文其他超参数与 FCOS 保持一致。选用 ResNet-101 作为检测器的主干特征提取网络,特征融合网络选用 FPN。从表 4 可以看出:在 MS COCO 数据集中,在 ResNet-101^[17]作为主干网络的情况下,本文方法比相同网络结构的密集检测器 FCOS 平均提高了 2.2 个百分点,与经典网络结构 RetinaNet、FSAF、

NoiseAnchor 相比也有不低于 1 个百分点的提升,在目标检测任务上取得了不错的性能提升;与同样改进标签分配策略的 FreeAnchor 相比也有 0.7 个百分点的提升;通过使用更深的主干网络 ResNext-101 进一步将精确度提高到 44.2%。利用可变性卷积(DCN)^[18]改进 ResNet101 主干网络以及检测器的最后一层后,可以将检测网络的 AP 提高到 46.4%,优于 FCOS、RetinaNet、RepPoints 等目前常用的目标检测网络。

表 4 MS COCO2017数据集上的测试结果对比

Table 4 Comparison of test results on MS COCO2017 dataset							%
方法	Backbone	AP	AP50	AP75	APs	APm	API
RetinaNet	ResNet-101	39.1	59.1	42.3	21.8	42.7	50.2
DetNet ^[19]	DetNet-59	40.3	62.1	43.8	23.6	42.6	50.0
Soft-NMS ^[20]	ResNet-101	40.8	62.4	44.9	23.0	43.4	53.2
FCOS	ResNet-101	41.5	60.7	45.0	24.4	44.8	51.6
NoisyAnchor ^[21]	ResNet-101	41.8	61.1	44.9	23.4	44.9	52.9
FSAF ^[22]	ResNet-101	42.8	63.1	46.5	27.8	45.5	53.2
FreeAnchor	ResNet-101	43.1	62.2	46.4	24.5	46.1	54.8
SAPD ^[23]	ResNet-101	43.5	63.6	46.5	24.9	46.8	54.6
MAL ^[24]	ResNet-101	43.6	61.6	47.1	25.0	46.9	55.8
ATSS	ResNet-101	43.6	62.1	47.4	26.1	47.0	53.6
FCOS	ResNeXt-64x4d-101	43.2	62.8	46.6	26.5	46.2	53.3
NoisyAnchor	ResNeXt-101	44.1	63.8	47.5	26.0	47.4	55.0
RepPoints ^[25]	ResNet101-DCN	45.0	66.1	49.0	26.6	48.6	57.5
ATSS	ResNet101-DCN	46.3	64.7	50.4	27.7	49.8	58.4
FreeAnchor	ResNeXt-64x4d-101	44.9	64.3	48.5	26.8	48.3	55.9
ATSS	ResNeXt-64x4d-101	45.6	64.6	49.7	28.5	48.9	55.6
ALAS	ResNet-101	43.8	62.8	47.3	26.4	47.2	53.8
	ResNeXt-101	44.2	64.0	47.1	26.2	47.6	54.9
	ResNet101-DCN	46.4	64.6	50.0	28.7	49.9	58.6

为了验证 ALAS 的鲁棒性,在 PASCAL VOC (07+12)数据集上与多个经典的目标检测网络进行了逐类别对比,如表 5 所示。由表 5 可以看出:自适应锚框分配策略相比于 ATSS 总体精确度提高了 0.8 个百分点,高于 Faster RCNN、SSD、FCOS 网络

1.3 个百分点以上;对 20 个类别的检测准确度详细比较,自适应标签分配策略在鸟类、汽车、狗类、显示器等 10 个类别效果突出;总体上,在小目标上的检测效果要高于 Faster RCNN、FCOS、ATSS、SSD 等检测网络。

表 5 VOC 数据集上的测试结果对比

Table 5 Comparison of test results on VOC dataset												%
类别	Faster RCNN	SSD300 ^[26]	FCOS	ATSS	本文方法	类别	Faster RCNN	SSD300 ^[26]	FCOS	ATSS	本文方法	
飞机	76.5	79.5	76.4	83.9	86.2	马	84.6	87.5	79.8	85.8	86.0	
自行车	79.0	83.9	79.4	81.9	82.7	摩托车	77.5	83.9	76.7	82.8	83.6	
鸟类	70.9	76.0	70.9	76.0	77.5	人	76.7	79.4	77.3	83.6	83.4	
船	66.5	69.6	63.3	68.3	69.6	植物	38.8	52.3	42.4	50.6	51.4	
瓶子	52.1	50.5	57.0	68.4	83.9	羊	73.6	77.9	77.3	81.4	84.0	
公交车	83.1	87.0	77.7	83.7	87.0	沙发	73.9	79.5	70.5	71.5	74.5	
轿车	84.7	85.7	82.8	87.0	88.2	火车	83.0	87.6	81.3	83.1	84.2	
猫	86.4	88.1	85.0	88.2	61.7	显示器	72.6	76.8	68.2	77.1	78.6	
椅子	52.0	60.3	54.2	62.2	61.7	狗	84.8	86.1	79.3	85.4	86.3	
牛	81.9	81.5	80.6	85.3	84.4	AP	73.2	77.5	72.2	78.0	78.8	
餐桌	65.7	77.0	63.9	74.4	72.4							

分别在 MS COCO 和 PASCAL VOC 数据集中对自适应标签分配策略与 FCOS 和 ATSS 网络进行可视化对比。由图 4 和图 5 可以看出,在物体被部分遮挡以及

待检测物体较小的情况下,自适应标签分配更容易正确检测出物体的位置,能够有效检测出 FCOS 和 ATSS 网络都未检测出的目标,降低漏检率。

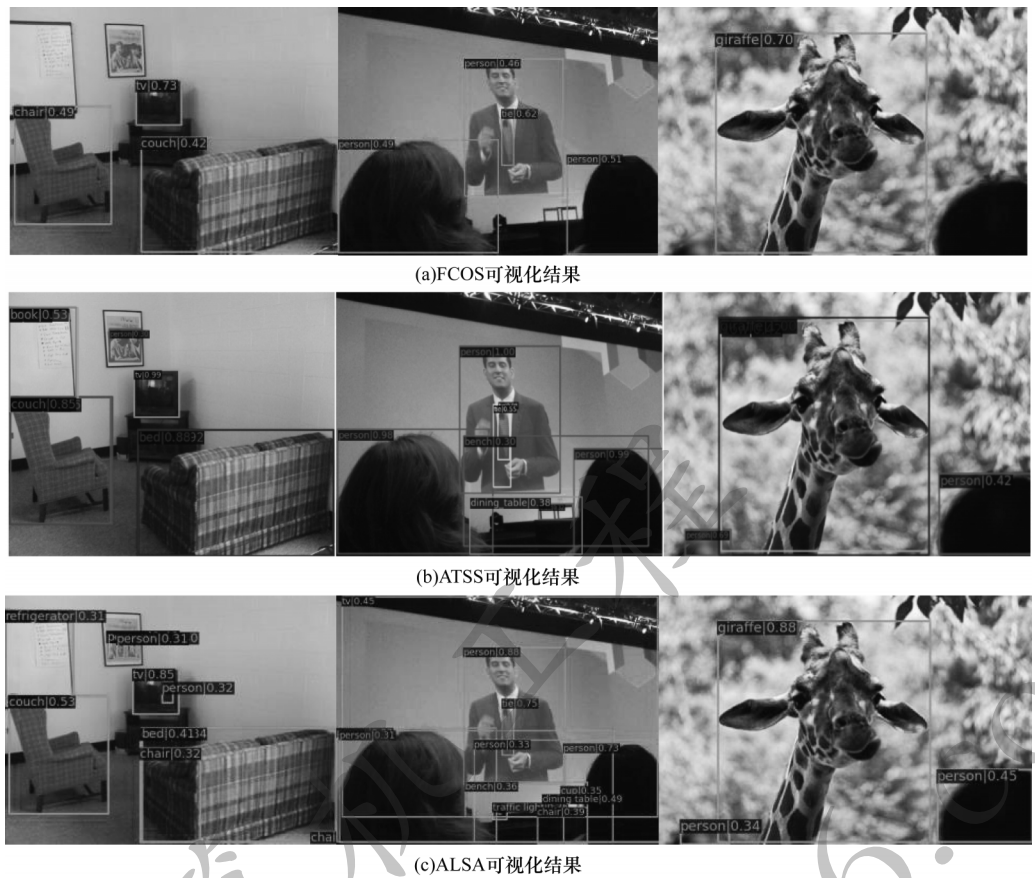


图4 在MS COCO数据集上的可视化结果对比
Fig.4 Comparison of visualization results on MS COCO dataset

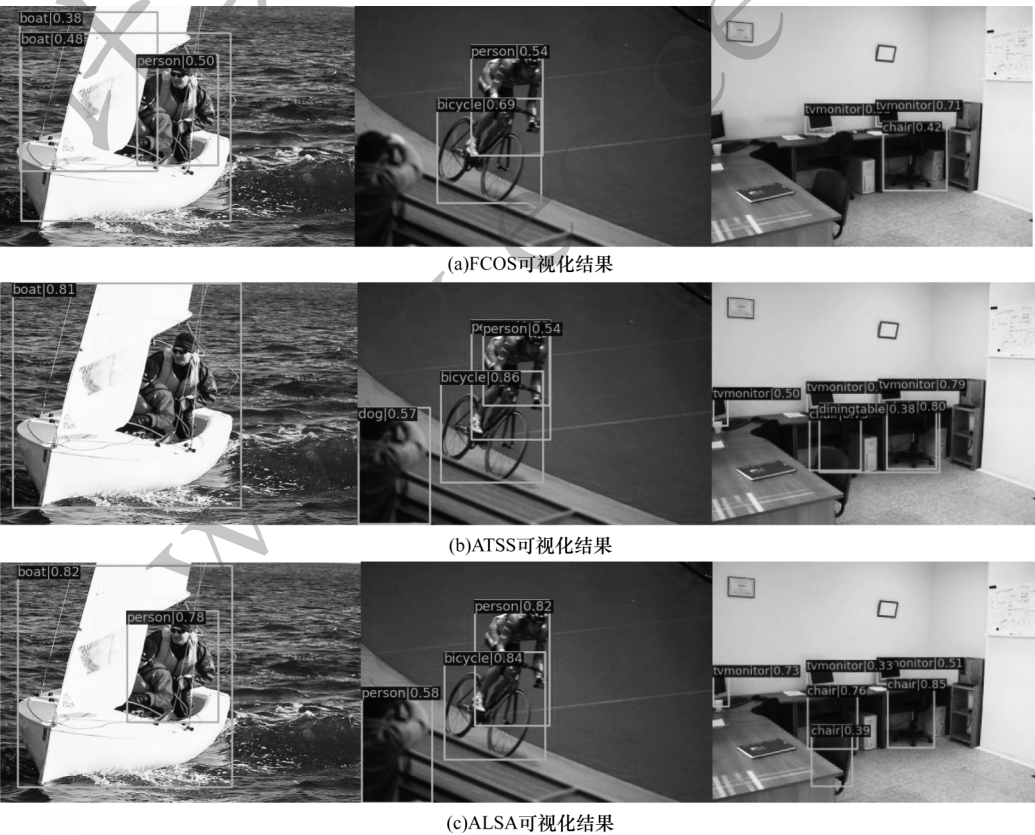


图5 在PASCAL VOC数据集上的可视化结果对比
Fig.5 Comparison of visualization results on PASCAL VOC dataset

3 结束语

本文提出一种自适应标签分配策略(ALAS),将标签分配作为一个概率分布的似然估计过程。该策略的核心是确定对模型有利的正样本和负样本划分方法,从而提高检测器的精度,比启发式IoU硬性分配或非概率分配策略更容易训练,并能保证网络的稳定性。在PASCAL VOC数据集上,本文方法对多个类别的检测准确率都有了明显提升,且最终的mAP指标提升至78.8%;在具有挑战性的MS COCO数据集上利用各种主干网络进行大量实验,并在ResNet-101作为主干网络时将AP提升至43.8%。下一步将充分利用本文联合分类定位得分的思想,对检测网络的后处理过程进行优化。

参考文献

- [1] 张晨阳,黄腾,吴壮壮. 基于K-Means聚类与深度学习的RGB-D SLAM算法[J]. 计算机工程,2022,48(1):236-244,252.
ZHANG C Y, HUANG T, WU Z Z. RGB-D SLAM algorithm based on K-means clustering and deep learning[J]. Computer Engineering, 2022, 48(1): 236-244, 252. (in Chinese)
- [2] JIANG B, LUO R, MAO J, et al. Acquisition of localization confidence for accurate object detection[C]//Proceedings of European Conference on Computer Vision. New York, USA: ACM Press, 2018: 816-832.
- [3] 包俊,刘宏哲. 融合可变形卷积网络的鱼眼图像目标检测[J]. 计算机工程,2021,47(4):248-255.
BAO J, LIU H Z. Object detection in fisheye images combining deformable convolutional networks [J]. Computer Engineering, 2021, 47(4): 248-255. (in Chinese)
- [4] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 2999-3007.
- [5] ZHANG X S, WAN F, LIU C, et al. FreeAnchor: learning to match anchors for visual object detection[EB/OL]. [2022-02-10]. <https://arxiv.org/abs/1909.02466>.
- [6] TIAN Z, SHEN C H, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2020: 9626-9635.
- [7] ZHANG S F, CHI C, YAO Y Q, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2020: 9756-9765.
- [8] ZHU B J, WANG J F, JIANG Z K, et al. AutoAssign: differentiable label assignment for dense object detection [EB/OL]. [2022-02-10]. <https://arxiv.org/abs/2007.03496>.
- [9] ZHU C C, HE Y H, SAVVIDES M. Feature selective anchor-free module for single-shot object detection[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2020: 840-849.
- [10] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 6517-6525.
- [11] KIM K, LEE H S. Probabilistic anchor assignment with IoU prediction for object detection [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2020: 355-371.
- [12] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2014: 740-755.
- [13] BÜHLMANN P, YU B. Boosting with the L2 Loss[J]. Journal of the American Statistical Association, 2003, 98(462): 324-339.
- [14] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 779-788.
- [15] REN S Q, HE K M, 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.
- [16] EVERINGHAM M, VAN Gool L, WILLIAMS C K I, et al. The pascal Visual Object Classes (VOC) challenge [J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [17] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 770-778.
- [18] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 764-773.
- [19] LI Z M, PENG C, YU G, et al. DetNet: design backbone for object detection[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 339-354.
- [20] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS—improving object detection with one line of code [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 5562-5570.
- [21] LI H D, WU Z X, ZHU C, et al. Learning from noisy anchors for one-stage object detection[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2020: 10585-10594.
- [22] ZHU C C, HE Y H, SAVVIDES M. Feature selective anchor-free module for single-shot object detection[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2020: 840-849.
- [23] ZHU C C, CHEN F Y, SHEN Z Q, et al. Soft anchor-point object detection[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2020: 91-107.

(下转第100页)

(上接第91页)

- [24] KE W, ZHANG T L, HUANG Z Y, et al. Multiple anchor learning for visual object detection [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA : IEEE Press, 2020: 10203-10212.
- [25] YANG Z, LIU S H, HU H, et al. RepPoints: point set representation for object detection [C]//Proceedings of

- IEEE/CVF International Conference on Computer Vision. Washington D. C. , USA : IEEE Press, 2020: 9656-9665.
- [26] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 21-37.

编辑 金胡考