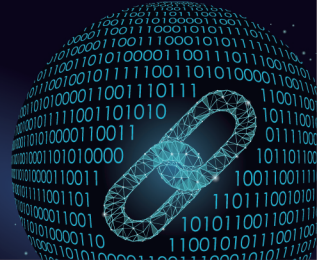




中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION



面向遮挡行人检测的自适应收缩非极大值抑制方法

李翔, 何淼, 罗海波

引用本文:

李翔,何淼,罗海波. 面向遮挡行人检测的自适应收缩非极大值抑制方法[J]. 控制与决策, 2024, 39(7): 2177–2185.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.0084>

您可能感兴趣的其他文章

Articles you may be interested in

多目标小尺度车辆目标检测方法

Multi-target and small-scale vehicle target detection method

控制与决策. 2021, 36(11): 2707–2712 <https://doi.org/10.13195/j.kzyjc.2020.0635>

Anchor-free的尺度自适应行人检测算法

Anchor-free scale adaptive pedestrian detection algorithm

控制与决策. 2021, 36(2): 295–302 <https://doi.org/10.13195/j.kzyjc.2020.0124>

抗遮挡与尺度自适应的改进KCF跟踪算法

Improved KCF tracking algorithm based on anti-occlusion and scale transformation

控制与决策. 2021, 36(2): 457–462 <https://doi.org/10.13195/j.kzyjc.2019.0394>

基于深度学习的行人轨迹预测方法综述

Survey of pedestrian trajectory prediction methods based on deep learning

控制与决策. 2021, 36(12): 2841–2850 <https://doi.org/10.13195/j.kzyjc.2020.1841>

基于MobileNet的多目标跟踪深度学习算法

Deep learning algorithm based on MobileNet for multi-target tracking

控制与决策. 2021, 36(8): 1991–1996 <https://doi.org/10.13195/j.kzyjc.2019.1424>

面向遮挡行人检测的自适应收缩非极大值抑制方法

李翔^{1,2,3}, 何淼^{1,2}, 罗海波^{1,2†}

(1. 中国科学院沈阳自动化研究所 光电信息处理重点实验室, 沈阳 110016;
2. 中国科学院机器人与智能制造创新研究院, 沈阳 110169; 3. 中国科学院大学, 北京 100049)

摘要: 基于视觉的行人检测是目标检测中的重要研究方向。如今主流的基于锚框的行人检测器输出的结果是大量冗余的,在结果输出之前需要对冗余预测进行非极大值抑制,因而非极大值抑制的效果将直接影响检测器的性能。行人检测任务中最大的挑战在于目标之间互相遮挡的现象,而严重重叠的目标使得传统的基于固定阈值的非极大值抑制方法难以在高召回率与低虚警率之间取得平衡。针对以上问题,提出一种根据重叠度自适应收缩预测框的非极大值抑制方法。根据对应目标的重叠度将预测框进行自适应的收缩,以降低预测框之间的重叠度。对收缩后的预测框进行非极大值抑制,可避免高重叠预测框对处理结果的影响。此外,指向性不明确的预测框将影响本算法的性能,为此提出一种中心点排斥损失函数,通过在重叠框的中心点之间施加排斥力来减少介于两目标之间的指向性不明确的模糊预测框数量,增强自适应收缩非极大值抑制算法的效果。仿真实验结果表明,所提算法可以有效提升基于锚框的检测器对重叠行人目标的检测性能。

关键词: 非极大值抑制; 遮挡行人检测; 目标检测; 深度学习; 计算机视觉; 机器学习

中图分类号: TP391 文献标志码: A

DOI: 10.13195/j.kzyjc.2023.0084

引用格式: 李翔,何淼,罗海波. 面向遮挡行人检测的自适应收缩非极大值抑制方法[J]. 控制与决策, 2024, 39(7): 2177-2185.

Adaptive shrinkage non-maximum suppression for occluded pedestrian detection

LI Xiang^{1,2,3}, HE Miao^{1,2}, LUO Hai-bo^{1,2†}

(1. Key Laboratory of Opto-Electronic Information Processing, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China; 2. Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China; 3. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Vision-based pedestrian detection is an important research field in object detection. The outputs of current main-stream anchor-based detectors are redundancy, and the non-maximum suppression (NMS) is necessary before output predictions. Therefore, the effectiveness of NMS will influence detectors performance directly. The biggest challenge in pedestrian detection is the occlusion issue. Heavily overlapping objects make it difficult for NMS to balance between high recall rate and low false positive rate. An adaptive shrinkage NMS (AS-NMS) is proposed to reduce the overlapping degree of occluded objects' boxes by adaptively shrinking the corresponding prediction boxes. Implementing NMS to the shrunken prediction boxes can avoid the impact of occlusion issue. In addition, ambiguous prediction boxes will have negative impact toward AS-NMS, and a center point repulsion loss (CPR loss) is proposed to reduce the numbers of ambiguous boxes hovering between two objects by implement a repulsion force between the center points of overlapping boxes. The experimental results demonstrate that the proposed algorithm can significantly boost the performance of anchor-based detectors towards occluded pedestrian detection.

Keywords: non-maximum suppression; occluded pedestrian detection; object detection; deep learning; computer vision; machine learning

0 引言

行人检测作为计算机视觉的重要课题之一,在智能安防、智慧交通^[1]、自动驾驶^[2]、机器人等领域发挥

着重要的作用^[3-4]。随着深度学习的发展,基于深度神经网络的检测模型^[5-9]在检测的准确性与鲁棒性上大幅超越传统的基于手工特征的检测器^[10],成为行

收稿日期: 2023-01-25; 录用日期: 2023-04-18.

责任编辑: 张国山.

†通讯作者. E-mail: luohb@sia.cn.

人检测领域的主导方法。

为了尽可能捕获任意大小、形状与尺度的目标,先进的基于锚框的检测器^[11-15]会在特征图上铺设密集的先验锚框。由于一个目标可能匹配多个锚框,模型的输出结果中将包含大量的冗余预测框,因此需要对模型输出结果进行非极大值抑制(non-maximum suppress, NMS)后处理。NMS的作用是删除预测结果中的冗余预测,仅为每个目标保留一个高质量预测。以双阶段检测模型Faster-RCNN^[16]为例,其通过区域提议网络(region proposal networks, RPN)产生大量提议框,并在此基础上通过感兴趣区域池化(region of interest pooling, RoI pooling)来提取实例特征。然而,因为目标会匹配多个提议框,从而产生大量重复的冗余预测框,所以NMS后处理的效果将对检测器的性能产生重要影响。传统NMS方法根据预测框之间的重叠度来判定各预测框是否归属于同一目标,即重叠度大于阈值的判定为属于同一目标,否则判定为属于不同目标。在稀疏目标检测中,由于目标之间很少有重叠现象,传统的NMS方法已经能够满足性能要求。然而在行人检测场景中,行人目标成群出现且随机移动,行人之间互相遮挡重叠的现象使得目标的预测框之间具有很高的重叠度。在传统NMS过程中,当分属于两个目标的预测框的重叠度大于阈值时,得分较低的预测框会被错误抑制,造成召回率下降。如果为了保留高重叠的预测框而提高NMS阈值,同属于一个目标的冗余预测框不能完全做到紧凑一致,较高的NMS阈值会使得松散的预测框被错误保留,造成虚警率上升。因此,传统NMS方法面对目标重叠场景时的问题在于其无法同时在高召回率与低虚警率之间取得平衡。针对上述问题,Gahlert等^[17]和Huang等^[18]分别提出VG-NMS和PBM。他们的基本思想是通过网络预测目标的可见部分,并将可见部分预测框应用于NMS。由于可见部分几乎不会重叠,可以避免目标重叠对NMS带来的干扰。Xu等^[19]提出利用目标的可见部分框与全身框来生成beta分布来表示该目标,并在beta分布空间内完成对冗余预测的抑制。上述方法都需要依赖额外的可见部分标注框,对数据集标注要求高。Liu等^[20]提出一种自适应非极大值抑制算法Adaptive NMS。该算法根据目标的重叠程度来动态调整对应的NMS阈值,以解决高重叠目标预测框被错误抑制的问题。该方法中的目标密度通过在网络中添加额外的轻量化网络来预测。因为预测的目标密度值将直接作为NMS阈值,所以该方法的性能对预测密度和预测框自身的精度均比较

敏感。

本文针对传统NMS无法同时适用于目标高度重叠的场景与目标稀疏的场景这一问题,提出一种基于自适应缩放的非极大值抑制(AS-NMS)方法。本方法根据预测框对应目标与其他目标的重叠度自适应设定一个收缩系数。在NMS之前,将对应预测框按照该系数进行收缩。收缩后的预测框之间的重叠度被缩小,高重叠预测框被转化成稀疏预测框。此时对收缩后的预测框进行非极大值抑制,可以缓解目标重叠对非极大值造成的干扰。

此外,由于检测模型输出预测框并不能做到完全紧凑,在高重叠目标之间会形成指向性不明确的模糊预测框。这些模糊预测框十分松散,容易被NMS过程遗漏,形成虚警预测。当预测框被收缩后,各框之间的重叠度进一步被缩小,松散预测框更容易形成虚警预测。为了进一步提升各预测框对于其所属目标的指向性,避免模糊预测对于NMS过程的干扰,本文提出一种中心点排斥损失函数(center-point repulsion loss, CPR loss),以提升高度重叠预测框中心点之间的区分度,避免缩小后的预测框指向性不明确的问题,缓解预测框收缩后虚警率上升的问题。

1 NMS处理高重叠目标时的难点分析

传统NMS方法流程概括如下:首先,根据预测框置信度得分对所有预测框进行排序,并将得分最高的预测框作为保留框,其余作为候选框。接下来,根据候选预测框与保留框之间的重叠度(intersection over union, IoU)来判定候选框与保留框是否属于同一目标:若IoU大于阈值,则判定保留框与当前候选框属于同一目标,候选框被抑制;若IoU小于阈值,则判定两者属于不同目标,候选框予以保留。当所有与保留框重叠度大于阈值的候选框都被抑制后,将当前保留框保存,并重新从剩余的预测框(不包括上一步中的保留框)中选取置信度分数最高分预测框作为新的保留框,然后重复上述过程。当剩余预测框两两之间IoU均小于阈值时停止非极大值抑制过程,保存的预测框则作为最终结果被输出。

传统NMS方法无法同时处理稀疏目标与高重叠目标的原因在于难以找到同时适用于稀疏目标与高重叠目标的NMS阈值 t 。如图1(a)所示,图中绿色矩形框为Faster-RCNN网络输出的未经非极大值抑制的预测框。可以看到,无论是左侧重叠场景还是右侧稀疏场景,网络为每个目标都输出了冗余的预测框。为了方便描述,将图1(a)进行简化,仅从中筛选出部分预测框,如图1(b)所示。



图1 重叠目标与稀疏目标预测框示意图

图1(b)中属于同一目标的预测框用相同颜色表示. 图1(b)中两个紫色框之间的IoU值为 $I_p = 0.64$, 两个红色框之间的IoU值为 $I_r = 0.85$, 红色框与蓝色框之间最大IoU值 $I_{rb} = 0.7$. 对于右侧紫色稀疏目标预测框, 其不与任何其他行人目标具备遮挡关系, 此时为了抑制其中一个紫色框并保留另一个作为最终预测, NMS 阈值取值范围的下限最低可以设为0, 上限则为两个紫色预测框的IoU值, 即紫色框的NMS 阈值 t_p 的取值范围为 $t_p \in [0, 0.64)$. 对于一般稀疏目标, 通常设置NMS 阈值 $t = 0.5$, 即重叠程度大于0.5的框会被判定为属于同一目标而被抑制. 而对于高度重叠目标而言, 如图1(b)中左侧红色与蓝色框标注目标所示, 为了在抑制冗余红色框的同时又能保留属于另一目标的蓝色框不被错误抑制, 其NMS 阈值 t_r 的取值范围为 $t_r \in [0.7, 0.85)$. 由此可见, 目标的NMS 阈值 t 的取值范围上限是由其预测框紧凑度决定的, 而其下限则由当前目标与其他目标的重叠程度决定. 当属于同一目标的预测框越紧凑, 两两之间的IoU值越大时, 其 t 的取值上限就越接近1. 相反, 当其预测框越松散时, 其 t 的上限越小. 而当前目标与其他目标重叠程度越高时, 其NMS 阈值范围的下限越接近1, 意味着此时 t 的取值范围越窄. 以上例子中, 紫色稀疏目标的NMS 阈值取值范围为 $[0, 0.64)$, 红色高重叠目标的NMS 阈值范围则为 $[0.7, 0.85)$, 两者之间不存在交集, 意味着不存在一个固定阈值能同时满足紫色目标与红色目标的非极大值抑制需求. 当 $t < 0.64$ 时, 蓝色框会被错误抑制, 造成召回率的下降; 当 $t > 0.7$ 时, 两个紫色框都会被保留作为输出, 它们其中之一将成为虚警预测, 增加预测结果的虚警率. 正因为紫色框代表的稀疏目标预测框与红蓝框代表的重叠目标预测框所适用的NMS 阈值的取值范围会出现没有交集的情况, 使得召回率与虚警率之间难以取得平衡, 这成为NMS 方法处理密集遮挡目标时的主要问题所在.

2 自适应收缩NMS

2.1 方法原理

本文有如下观察: 在固定中心点的前提下, 将中心点重合的矩形框缩小相同的比例, 矩形框之间的IoU值不发生改变. 如图2(a)所示, 两矩形框中心点重合, IoU为0.85. 固定中心点将矩形框同时缩小50%, 缩小后的两个矩形IoU仍然是0.85. 对于中心点不重合的矩形框, 如图2(b)所示, 固定中心点并将边长同时缩小50%之后, 它们之间的IoU值由原来的0.7缩小至0.3. 基于上述原理, 本文提出一种自适应收缩非极大值抑制方法AS-NMS.

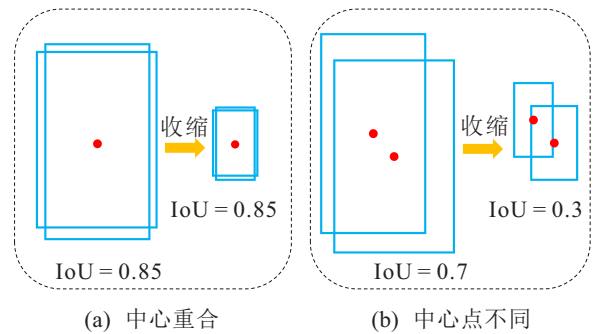


图2 矩形框收缩前后重叠度变化示意图

一般情况下, 属于同一目标的预测框的中心点互相之间更加靠近, 而不同目标预测框的中心点之间则更加远离. 因此, 当预测框边长被按照某一比例同时缩小之后, 属于同一目标的预测框彼此之间重叠度的减小幅度要小于不同目标预测框之间重叠度的减小幅度. 如图3所示, 图中红色框与蓝色框分属于两个相互重叠的目标. 当图中矩形框都被按比例缩小50%之后, 同色(同类)矩形框之间的重叠度变化不大, 而红色与蓝色矩形框之间的重叠度减小. 由前文可知, 不同目标重叠度小, 意味着其预测框NMS 阈值下限更接近0. 因此, 红色框与蓝色框被缩小之后, 它们各自的NMS 阈值取值范围的下限相比于收缩前将有所下降. 又因为同色(同类)框彼此之间重叠度变化并不大, 所以他们各自NMS 阈值范围的上限变化也不大. 因此, 收缩之后图中重叠目标的NMS 阈值的取值范围被扩大. 同理, 图中紫色稀疏目标框经过缩小后, 彼此之间重叠度变化也不明显, 所以其NMS 阈值的取值范围变化也较小. 因此, 当图中预测框经过等比例缩小后, 高重叠目标的预测框NMS 阈值取值范围被拓宽, 稀疏目标NMS 取值范围则变化不大, 稀疏目标与高重叠目标的NMS 阈值的取值范围将产生交集. 从该交集中选取的NMS 阈值可以同时适用于图3中的稀疏目标与高重叠目标. 实际应用中, 利用收缩后的预测框进行NMS 操作. 由于原始预测框与

收缩框是一一对应的,可以利用AS-NMS输出的预测框索引值对原始预测框进行索引,索引后的预测框即为最终的输出预测。

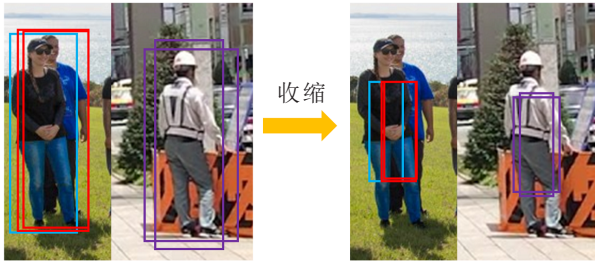


图3 预测框收缩示意图

综上所述,AS-NMS通过将预测框按一定比例缩小,降低各目标预测框的重叠度,从而拓宽重叠目标NMS阈值的取值范围,使得高重叠目标的NMS阈值的取值范围与一般稀疏目标NMS阈值的取值范围产生交集.在该交集中选取阈值对收缩后的预测框进行NMS处理,并利用输出的索引值索引原始预测框,得到最终输出。

2.2 自适应收缩比例 γ

上文中,为举例说明,将收缩比例 γ 一律设定为50%.在实际应用中,本文提出根据目标的重叠度来自适应设定收缩比例.对于稀疏目标,由于其NMS阈值取值范围下限已经接近为0,收缩稀疏目标的预测框并不能继续降低其NMS阈值的下限.而同类预测框也不完全是中心点重合的,因此过度的收缩还可能造成同类框重叠度下降,使其阈值取值范围上限降低.因此对于稀疏目标,收缩的比例应当尽可能小.而对于高重叠目标,若收缩的比例过小,不同目标预测框之间重叠度变化不大,使得NMS取值范围不能得到有效扩大.因此,本文提出根据目标的重叠度自适应调整收缩比例.目标 i 的收缩比例 γ_i 的表达式如下:

$$\gamma_i = \exp\left(-\frac{d_i^2}{\sigma}\right). \quad (1)$$

其中: d_i 为目标 i 的密度,含义是目标 i 和与其重叠度最大的其他目标框之间的IoU值,表示为

$$d_i = \max_{b_j \in G, i \neq j} \text{iou}(b_i, b_j). \quad (2)$$

b_i 表示目标 i 的标注框, $\text{iou}(x, y)$ 计算框 x 与框 y 之间的IoU值, G 表示所有目标框的集合.由式(1)可知,目标框的收缩比例 γ 随着其密度 d 的增大逐渐减小.这意味着对高重叠框的收缩幅度要大于稀疏框的收缩幅度.当 d_i 为0时,表明没有任何其他目标与目标 i 重叠,此时 $\gamma_i = 1$,意味着不对该框进行收缩.当 d_i 逐渐增大时, γ_i 逐渐减小,此时对应框将被缩小,以拓展

其NMS阈值的取值范围. σ 是敏感度参数,用以控制收缩比例 γ_i 随密度 d_i 的变化而变化的幅度.图4展示了在不同 σ 下,收缩系数 γ_i 随着目标密度 d_i 的变化曲线.从中可以看出,在相同敏感度参数 σ 下,收缩系数 γ_i 随着目标密度 d_i 的增大而逐渐减少,并且随着 σ 逐渐增大,曲线逐渐趋于平缓,即收缩系数 γ_i 对目标密度 d_i 变化的敏感度逐渐降低.图5则展示了在敏感参数 $\sigma = 0.5$ 时,不同重叠度的目标框收缩前后的对比图.可以看到,当目标框与其他目标框没有交集时,目标的密度为0,此时收缩系数为1.0,即不收缩.当目标之间互相遮挡时,目标的密度值大于0,此时收缩系数小于1.0,且密度越大,收缩系数越接近0.

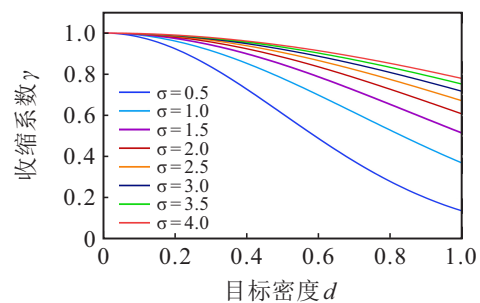


图4 不同 σ 下收缩系数 γ 随目标密度 d 变化的曲线



(a) 原始目标框 (b) 收缩后目标框

图5 $\sigma = 0.5$ 时不同密度目标框收缩结果示意图

每个预测框对应的密度 d_i 可以采用Adaptive NMS中的方法进行预测输出.由于本文目标在于提出新的NMS方法,在实验中直接采用标签信息如式(2)所示计算每个预测框的密度值.实际应用中采用Adaptive NMS文中方法预测密度值即可. AS-NMS算法流程如算法1所示.

算法1 AS-NMS算法流程.

输入: $B = \{b_1, \dots, b_N\}$, $S = \{s_1, \dots, s_N\}$, $D = \{d_1, \dots, d_N\}$, N_t, σ // B 是包含所有初始预测框的列表, S 是包含所有预测框置信度的列表, D 是包含所

有预测框密度预测值的列表, N_t 是非极大值抑制的阈值, σ 是敏感参数.

输出: F, S .

- 1) $F \leftarrow \{\}$
- 2) while $B \neq \emptyset$ do
- 3) $m \leftarrow \arg \max S$
- 4) $M \leftarrow b_m$
- 5) $F \leftarrow F \cup M; B \leftarrow B - M$
- 6) $d_M \leftarrow d_m$
- 7) $\gamma_M = \exp\left(\frac{-d_M^2}{\sigma}\right)$
- 8) $M^s \xleftarrow{\gamma_M} M$ #将框 M 按比例 γ_M 缩小
- 9) for (b_i, d_i) in (B, D) do
- 10) $\gamma_i = \exp\left(\frac{-d_i^2}{\sigma}\right)$
- 11) $b_i^s \xleftarrow{\gamma_i} b_i$ #将框 b_i 按比例 γ_i 缩小
- 12) if $\text{iou}(M^s, b_i^s) \geq N_t$ then
- 13) $B \leftarrow B - b_i; S \leftarrow S - s_i; D \leftarrow D - d_i$
- 14) end if
- 15) end for
- 16) end while

2.3 中心点排斥损失函数 CPR loss

对于本文提出的 AS-NMS 方法而言, 属于同一目标的预测框的紧凑性十分重要. 若同类预测框紧凑性较低, 则经过自适应收缩之后, 其 NMS 阈值取值范围的上限将大幅下降. 如果同类目标预测框过于松散, 则可能出现收缩后的同类预测框互相之间 IoU 值为 0 的情况, 极易造成假阳率的上升.

对于重叠度较高的两个目标, 它们的标注框中心点通常也会十分靠近. 如果它们的预测框较为松散, 则可能出现预测框指向性不明确的情况, 即无法分辨各个松散的预测框属于哪个目标. 通过观察检测器的输出结果, 可以发现在高度重叠的两个目标之间经常会出现介于两者之间的模糊预测. 若此时对这些松散的预测框进行收缩, 其重叠度因为收缩而被降低以致低于 NMS 阈值, 此时进行 NMS 操作, 这些介于两目标之间的松散预测框可能会被错误保留, 形成虚警预测. 为了提升高重叠目标预测框的指向性, 减少目标之间模糊预测框数量, 本文提出一种中心点排斥损失函数 CPR loss.

基于锚框的检测器, 如 Faster-RCNN, 会为每个目标匹配多个提议框 (proposals), 所以每个目标都会有多个预测输出. 综上所述, 中心点重合的矩形框经过收缩之后重叠度不改变. 所以在 CPR loss 中, 只需要提高各预测框中心点的指向性即可, 即让不属于同一目标的预测框中心点尽量远离, 如图 6 所示.

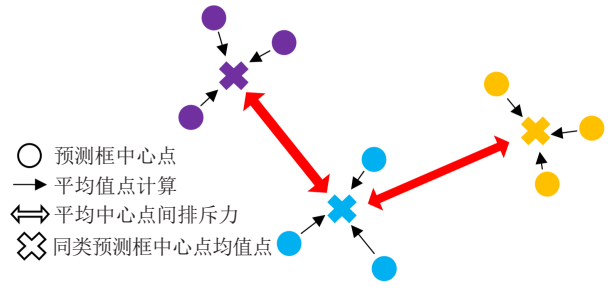


图 6 中心点排斥损失函数示意图

在训练过程中, 首先根据各个预测框所属目标的不同, 对所有预测框进行分组. 将属于目标 i 的预测框中心点分组记为 $C_i = \{c_1, c_2, \dots, c_{N_i}\}$, N_i 表示属于目标 i 的预测框的数量, $c_j = (x_j, y_j)$ 表示第 j 个预测框的中心点坐标. 接下来, 计算各中心点坐标分组的均值

$$\bar{C}_i = \frac{1}{N_i} \sum_{n=1}^{N_i} c_n = (\bar{x}_i, \bar{y}_i), \quad (3)$$

则目标 i 与目标 j 所属预测框的中心点之间的距离为

$$m_{ij} = \sqrt{(\bar{x}_i - \bar{x}_j)^2 + (\bar{y}_i - \bar{y}_j)^2}. \quad (4)$$

然后, 利用各目标标注框, 计算各目标中心点之间应当保持的最小距离. 令 $C_i^t = (x_i^t, y_i^t)$ 表示目标 i 的标注框中心点坐标, 且另有一目标 j 与其重叠, 记为 $C_j^t = (x_j^t, y_j^t)$. 目标 i 与目标 j 标注框中心点之间的欧氏距离为

$$m_{ij}^t = \sqrt{(x_i^t - x_j^t)^2 + (y_i^t - y_j^t)^2}. \quad (5)$$

CPR loss 表达式如下:

$$L_{\text{CPR}} = \sum_{i=1}^M \sum_{j=1}^M \max(m_{ij} - m_{ij}^t, 0) \times \mathbf{1}^{\text{iou}(b_i, b_j) > f} \times \mathbf{1}^{\text{iou}(b_i, b_j) = d_i}. \quad (6)$$

其中

$$\mathbf{1}^F = \begin{cases} 1, & F = \text{True}; \\ 0, & \text{otherwise}. \end{cases} \quad (7)$$

CPR 损失函数在重叠目标预测框中心点之间施加一个排斥力, 使得各目标的预测框平均中心点间的距离不低于对应目标标注框之间中心点距离, 以提升重叠目标中心点的指向性. $\mathbf{1}^F$ 项的作用是进行困难样本挖掘, 以使该函数发挥更好的效果. $\mathbf{1}^{\text{iou}(b_i, b_j) > f}$ 的含义是, 只在重叠度大于 f 的两个样本之间施加这一排斥力, 而对于重叠度较小的两目标, 则不计入损失. 对于一些稀疏目标而言, 施加该排斥力并不能带来正向作用, 因此该项可以使优化过程更加专注于提升高重叠目标中心点的指向性. $\mathbf{1}^{\text{iou}(b_i, b_j) = d_i}$ 的含义则是, 若目标 j 是与目标 i 重叠度最大的目标, 则在

它们之间施加排斥力,否则不计入损失.其中 d_i 为目标 i 的密度,定义如式(2)所示.该项使得该损失函数只在重叠度最大的两个预测框之间施加排斥力,即假如目标 i 同时与目标 j 和目标 z 具有重叠关系,而 $iou(b_i, b_j) > iou(b_i, b_z)$,则CPR损失函数只在目标 i 与目标 j 的预测框之间施加排斥力,使得CPR函数更加专注于对高重叠目标预测框的指向性的提升.

3 仿真实验

3.1 实验数据集

本文采用CrowdHuman^[21]数据集进行训练与测试.该数据集被分成训练集、验证集和测试集,分别包含15 000张、4 370张与5 000张图片.本文在训练集上训练模型,在验证集上进行测试.在CrowdHuman中,平均每张图像中包含22.64个行人目标,且平均每张图像包含高度重叠(与任一其他目标之间IoU>0.5)的行人2.4个.因此该数据集上的训练与测试的结果可以很好地展示模型对于高重叠目标的检测性能.

3.2 实验评价指标

本文共采用3个评价指标,其中两个是行人检测常用评价指标,分别是精度均值AP和平均对数漏检率 MR^{-2} .另一个是用来衡量密集目标检测能力的指标JI系数.AP同时反映了检测结果的准确率(Precision)和召回率(Recall).以召回率为横坐标,准确率为纵坐标,二者形成的曲线成为P-R曲线,而曲线与坐标轴所包围区域的面积即为AP值,如下所示:

$$AP = \int_0^1 P(R)dR \quad (8)$$

其中: R 为召回率, $P(R)$ 为准确率随召回率变化的函数.AP值越大代表检测器性能越好.召回率与准确率分别如下所示:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (9)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (10)$$

其中TP、FP、FN分别为真阳、假阳和假阴样本的数量.与AP类似,以单张图片假阳率(false positive per image, FPPI)为横坐标,丢失率(miss rate, MR)为纵坐标,而这形成的曲线与坐标轴所围区域面积就是 MR^{-2} .丢失率 $MR = 1 - \text{Recall}$,因此 MR^{-2} 同时反映检测结果的召回率与假阳(虚警)率. MR^{-2} 对虚警率十分敏感,其值越小代表检测器性能越好.JI则是检测结果中预测框集合与真值框集合的重叠程度,用来衡量检测器的计数能力,常被用在密集目标检测中,其计算方法如下所示:

$$JI = \frac{|P \cap G|}{|P| + |G| - |P \cap G|}. \quad (11)$$

其中: P 为预测框集合, G 为基准框集合, $|x|$ 为集合 x 的元素个数.JI越大代表性能越好.

3.3 实验设置

本文采用带有FPN^[22]的Faster-RCNN检测器作为实验对象来验证所提算法的有效性.训练时,设置单批图像数量为16,总共在训练集上迭代30轮.采用随机梯度下降优化器SGD,初始学习率设置为0.02,分别在第24轮与第27轮时乘以0.1与0.01.权重衰减设置为0.0001,动量设为0.9.训练与测试时,图像最短边都调整为800像素.所有实验均基于Pytorch1.7.1框架,采用4张NVIDIA RTX3090Ti GPU.

3.4 实验结果分析

表1给出了本算法基于Faster-RCNN检测器的消融实验.NMS阈值设为0.5,AS-NMS中设置 $\sigma = 2.5$,CPR loss中设置 $f = 0.4$.

表1 基于Faster-RCNN的消融实验

CPR loss	NMS	Adaptive NMS	AS-NMS	AP	MR^{-2}	JI
	✓			87.27	44.60	79.21
✓	✓			87.97	44.32	80.27
✓		✓		88.18	44.69	80.09
✓			✓	88.45	44.38	80.59

由表1结果可以看出,在同样使用NMS的情况下,采用CPR损失函数训练的模型测试结果相比于基线模型,在AP上提升了0.70%,在JI上提升了1.06%,同时在MR上降低了0.28%.JI增加意味着模型对于人群的计数能力有所提升,侧面反映了对模型召回率的提升.这是因为CPR损失函数增加了高重叠预测框的指向性,可以避免一些预测框由于过于靠近相邻目标而被NMS过程错误抑制.同时,预测框指向性提升带来的模糊预测数量的减少,可以有效降低虚警率, MR^{-2} 随之降低.相比于NMS,本文提出的AS-NMS进一步将AP提升了0.45%,将JI提升0.32%,同时对 MR^{-2} 几乎不产生影响.这表明AS-NMS在不增加假阳率的同时提升了模型对目标的召回能力与检测精度.当高重叠预测框被缩小之后,重叠度降低,从而使其NMS阈值范围得到扩大.当实验采用的NMS阈值0.5正好被包含在拓展后的NMS阈值范围内时,原本会被错误抑制的高重叠框此时将得到保留,从而使召回率得到提升.综合而言,相比于基线模型,本文所提算法将AP提升了1.18%,将JI提升了1.38%,还将 MR^{-2} 降低了0.32%.此外,从AS-NMS与Adaptive NMS的对

比结果可以看出, AS-NMS处理的结果在AP方面比Adaptive NMS大0.37%。同时, Adaptive NMS对MR和JI都产生了负面影响, 即使MR⁻²增大了0.37%, 使JI减少了0.18%。这是因为Adaptive NMS直接采用目标的密度值作为NMS阈值, 对预测框的紧凑程度更加敏感。当预测框的紧凑度不理想时, Adaptive NMS将产生较大的误差, 从而影响处理的效果。AS-NMS通过缩放扩大NMS阈值的取值范围, 对于预测框的紧凑程度具有更好的宽容度。

表2结果展示了CPR loss中参数*f*对检测性能的影响, 其中结果均采用NMS方法处理得到。可以看到, 当*f* = 0.4时, 模型取得了最优的综合性能, 即在AP与MR⁻²两项参数上取得最优。虽然*f* = 0.45时JI最优, 但其对应的MR⁻²与*f* = 0.4时相差0.43%, 因此综合而言*f* = 0.4时综合性能更好。当*f*低于0.4时, 意味着CPR loss在低重叠的目标中点之间施加了排斥力, 而这会分散优化过程对于高重叠目标的注意力。当*f*大于0.4时, 部分CPR loss只对严重重叠的预测框施加排斥力, 使得部分高重叠框被忽略, 造成CPR loss效果减弱。

表2 CPR loss中参数*f*对性能影响的实验结果

<i>f</i>	AP	MR ⁻²	JI
baseline	87.97	44.60	79.21
0.30	87.86	44.93	80.11
0.35	87.95	44.54	80.31
0.40	87.97	44.32	80.27
0.45	87.89	44.75	80.40
0.50	87.97	45.01	80.36

此外, 为了直观说明CPR loss对模糊预测框数量的抑制作用, 在验证集上统计了中心点与对应匹配目标距离大于与其他周围目标距离的预测框的数量。若一个预测框的中心点更加靠近周围目标而非自身所匹配的目标, 那么该预测框可以被归为指向性不明确的模糊预测。表3所列的是对于高重叠目标(密度大于0.45), 采用*f* = 0.4的CPR loss前后模糊预测框数量在预测框总数量中占比的变化情况。可以看到, CPR loss使得模糊预测框占比下降了0.55%, 验证了CPR loss对指向性不明确的模糊预测具有抑制作用。

表3 CPR loss对高重叠目标的模糊预测框数量的影响

CPR loss	模糊预测数量/预测框总数量
	23.01%
✓	22.46%

本文还对参数σ对检测性能的影响进行了探

究。σ越小, 则对应的γ越小, 意味着预测框被收缩的幅度越大。实验结果如表4所示, 其中baseline(基线结果)为加入CPR损失(*f* = 0.4)后训练的网络, 且输出结果由NMS方法处理得出。表中所有NMS阈值均为0.5。由表4中结果可知, 当σ较小时, 由于预测框被收缩幅度过大, 造成同类预测框之间的重叠度减小幅度较大, 此时NMS阈值的取值范围上限被大幅缩小。当NMS阈值上限小于0.5时, 采用0.5作为NMS阈值将产生大量虚警预测。如表4所示, MR⁻²大幅上升。当σ取值逐渐上升, 预测框收缩幅度减小时, 虚警预测增加的问题被逐渐改善。将表4中结果可视化, 如图7所示, 其中阴影区域表示最优σ的取值区间。在图7(a)与图7(c)阴影区域表示的区间中, AP与JI分别取得最优, 同时MR⁻²在此区间中也趋于平稳。当σ = 2.5时, 3项指标的综合性能相对最好。

表4 AS-NMS中敏感度参数σ对性能影响的实验结果

σ	AP	MR ⁻²	JI
baseline	87.97	44.60	79.21
0.5	86.38	65.21	75.66
1.0	88.34	51.16	79.71
1.5	88.41	45.91	80.41
2.0	88.53	44.69	80.58
2.5	88.45	44.38	80.59
3.0	88.39	44.30	80.58
3.5	88.33	44.29	80.54
4.0	88.29	44.30	80.53
4.5	88.26	44.30	80.49
5.0	88.24	44.29	80.46

表5展示了在不同检测器上应用AS-NMS的实验结果。可以看到, 相较于Faster-RCNN, AS-NMS对YOLOv3^[23]的性能提升是更加显著的。与NMS相比, AS-NMS将YOLOv3的AP提升了2.92%, 将JI提升了3.05%, 将MR⁻²降低了0.80%。与Adaptive NMS相比, AS-NMS对模型性能的提升同样是更加显著的。虽然MR⁻²上升了0.16%, 但是AP和JI分别提升了1.04%和1.13%。然而, AS-NMS对RetinaNet^[24]的提升却是有限的, 对应的MR⁻²甚至出现了1.68%的上升。这是因为, 表中所列的3种检测器所输出的预测框紧凑性是不同的。如图8展示的是表中3种检测器对同一张图片的预测结果, 图中的预测框均未进行非极大值抑制。可以明显看到, 相比于Faster-RCNN与YOLOv3, RetinaNet的预测结果是十分松散的。由上文可知, 对松散的预测框进行收缩, 同目标预测框重叠度也会被大幅减小, 从而造成虚警率上升。但是与Adaptive NMS相比, AS-NMS仍然输

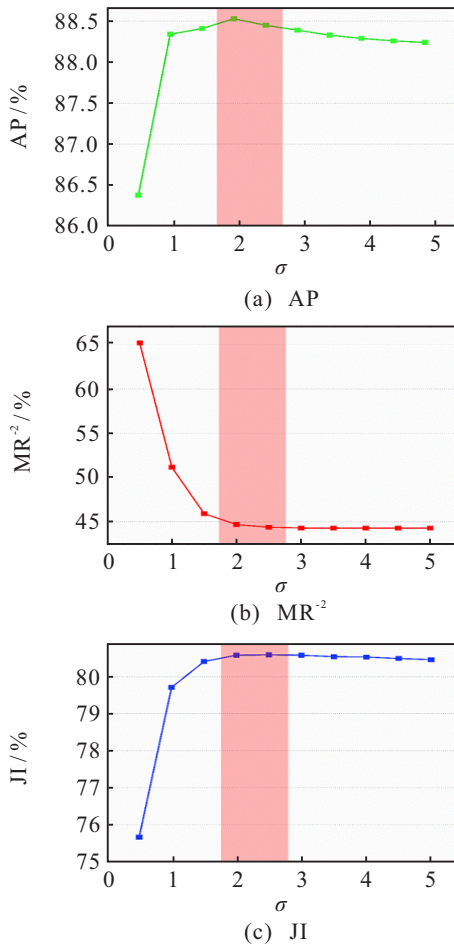


图7 敏感系数σ对检测性能的影响

表5 AS-NMS在不同检测器上的实验结果

model	NMS	Adaptive NMS	AS-NMS	AP	MR ⁻²	JI
Faster-RCNN	✓			87.97	44.32	80.27
		✓		88.18	44.69	80.09
			✓	88.45	44.38	80.59
RetinaNet	✓			81.26	53.62	73.78
		✓		81.01	57.85	73.49
			✓	81.40	55.30	73.73
YOLOv3	✓			85.01	50.39	71.88
		✓		86.89	49.43	73.80
			✓	87.93	49.59	74.93

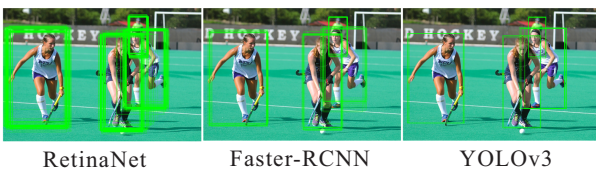


图8 不同检测器输出预测框紧凑度对比

出了更好的结果. Adaptive NMS直接采用密度值作为阈值,使得其对预测框松散程度更加敏感.表5的结果表明了,相比于Adaptive NMS而言,AS-NMS对于预测框的松散程度具有更好的宽容度.同时还说

明了预测框越紧凑,AS-NMS表现越好.

3.5 定性结果

图9展示了实验的定性结果,其中图9(a)中绿色点为预测框中心点,图9(b)中红色框为被NMS错误抑制的重叠目标的预测框.从图9(a)中椭圆区域可以看出,基线结果中出现大量重叠目标预测框中心点指向性不明确的情况. CPR loss的加入有效增加了高重叠预测框中心点之间的距离,提升预测框的指向性.图9(b)则展示了AS-NMS相较于NMS,可以防止高重叠目标的预测框被错误抑制.



(a) CPR损失函数对预测框中心点指向性提升效果 (b) AS-NMS与NMS对高重叠目标作用的对比

图9 定性结果展示

4 结论

针对传统非极大值抑制在处理高重叠目标预测框时不能做到高召回率与低虚警率兼顾的问题,本文提出了一种自适应收缩非极大值抑制方法AS-NMS. 根据预测框对应目标的密度来自适应收缩预测框,达到减小预测框重叠度从而拓宽其NMS阈值取值范围的目的. 同时,为了提升重叠目标预测框的区分度与指向性,提出一种中心点排斥损失函数CPR loss,在属于不同目标的重叠目标预测框的中心点之间施加一个排斥力,以减少模糊预测框的数量. 实验结果表明,本文所提出的算法可以有效提升遮挡行人检测的性能,可以在不增加虚警率的同时提升预测结果的精度与召回率.

参考文献(References)

[1] 金沙沙, 龙伟, 胡灵犀, 等. 多目标检测与跟踪算法在智能交通监控系统中的研究进展[J]. 控制与决策, 2023, 38(4): 890-901. (Jin S S, Long W, Hu L X, et al. Research progress of detection and multi-object tracking algorithm in intelligent traffic monitoring system[J]. Control and Decision, 2023, 38(4): 890-901.)

[2] 任柯燕, 谷美颖, 袁正谦, 等. 自动驾驶3D目标检测研究综述[J]. 控制与决策, 2023, 38(4): 865-889. (Ren K Y, Gu M Y, Yuan Z Q, et al. 3D object detection

- algorithms in autonomous driving: A review[J]. *Control and Decision*, 2023, 38(4): 865-889.)
- [3] 罗艳, 张重阳, 田永鸿, 等. 深度学习行人检测方法综述[J]. *中国图象图形学报*, 2022(7): 2094-2111.
(Luo Y, Zhang C Y, Tian Y H, et al. An overview of deep learning based pedestrian detection algorithms[J]. *Journal of Image and Graphics*, 2022(7): 2094-2111.)
- [4] 祝文斌, 苑晶, 朱书豪, 等. 低光照场景下基于序列增强的移动机器人人体检测与姿态识别[J]. *机器人*, 2022, 44(3): 299-309.
(Zhu W B, Yuan J, Zhu S H, et al. Sequence-enhancement-based human detection and posture recognition of mobile robots in low illumination scenes[J]. *Robot*, 2022, 44(3): 299-309.)
- [5] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, 2014: 580-587.
- [6] Girshick R. Fast R-CNN[C]. 2015 IEEE International Conference on Computer Vision. Santiago, 2016: 1440-1448.
- [7] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multiBox detector[J/OL]. 2016, arXiv: 1512.02325.
- [8] Tian Z, Shen C H, Chen H, et al. FCOS: Fully convolutional one-stage object detection[C]. 2019 IEEE/CVF International Conference on Computer Vision. Seoul, 2020: 9626-9635.
- [9] 王红梅, 王晓鸽, 王晓燕. 基于深度学习的复杂背景下目标检测[J]. *控制与决策*, 2022, 37(12): 3115-3121.
(Wang H M, Wang X G, Wang X Y. Target detection under complex background based on deep learning[J]. *Control and Decision*, 2022, 37(12): 3115-3121.)
- [10] 姬东飞, 丁学明. 基于特定区域去相关的行人检测算法[J]. *信息与控制*, 2020, 49(4): 404-413.
(Ji D F, Ding X M. Specific region decorrelation feature channel on pedestrian detection[J]. *Information and Control*, 2020, 49(4): 404-413.)
- [11] 钱惠敏, 陈纬, 马宜龙, 等. 基于 ResNet34_D 改进 YOLOv3 模型的行人检测算法[J]. *控制与决策*, 2022, 37(7): 1713-1720.
(Qian H M, Chen W, Ma Y L, et al. Pedestrian detection based on developed YOLOv3 with ResNet34_D[J]. *Control and Decision*, 2022, 37(7): 1713-1720.)
- [12] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[J/OL]. 2017, arXiv: 1703.06870.
- [13] Chu X G, Zheng A L, Zhang X Y, et al. Detection in crowded scenes: One proposal, multiple predictions[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, 2020: 12211-12220.
- [14] 邹梓吟, 盖绍彦, 达飞鹏, 等. 基于注意力机制的遮挡行人检测算法[J]. *光学学报*, 2021, 41(15): 1515001.
(Zou Z Y, Gai S Y, Da F P, et al. An algorithm for detecting occluded pedestrians based on attention mechanism[J]. *Acta Optica Sinica*, 2021, 41(15): 1515001.)
- [15] 李翔, 何淼, 罗海波. 一种面向遮挡行人检测的改进 YOLOv3 算法[J]. *光学学报*, 2022, 42(14): 1415003.
(Li X, He M, Luo H B. An improved YOLOv3 algorithm for occlusion pedestrian detection[J]. *Acta Optica Sinica*, 2022, 42(14): 1415003.)
- [16] 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.
- [17] Gahlert N, Hanselmann N, Franke U, et al. Visibility guided NMS: Efficient boosting of amodal object detection in crowded traffic scenes[J/OL]. 2020, arXiv: 2006.08547.
- [18] Huang X, Ge Z, Jie Z Q, et al. NMS by representative region: Towards crowded pedestrian detection by proposal pairing[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, 2020: 10747-10756.
- [19] Xu Z X, Li B H, Yuan Y, et al. Beta R-CNN: Looking into pedestrian detection from another perspective[J/OL]. 2020, arXiv: 2210.12758.
- [20] Liu S T, Huang D, Wang Y H. Adaptive NMS: Refining pedestrian detection in a crowd[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2020: 6452-6461.
- [21] Shao S, Zhao Z J, Li B X, et al. CrowdHuman: A benchmark for detecting human in a crowd[J/OL]. 2018, arXiv: 1805.00123.
- [22] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 936-944.
- [23] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J/OL]. 2018, arXiv: 1804.02767.
- [24] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]. 2017 IEEE International Conference on Computer Vision. Venice, 2017: 2999-3007.

作者简介

李翔(1993—), 男, 博士生, 从事图像处理、深度学习、行人检测、目标检测等研究, E-mail: lixiang184@mails.ucas.ac.cn;

何淼(1992—), 男, 副研究员, 博士, 从事图像处理、深度学习、目标检测、目标跟踪等研究, E-mail: hemiao@sia.cn;

罗海波(1967—), 男, 研究员, 博士生导师, 博士, 从事实时图像处理、自动目标识别、自动目标检测、先进制导技术等研究, E-mail: luohb@sia.cn.