

DOI:10.19651/j.cnki.emt.2106617

# 基于可见区域的拥挤行人检测

朱肖磊 吴训成 肖子遥 肖永超

(上海工程技术大学 机械与汽车工程学院 上海 201620)

**摘要:** 针对公共场合、拥挤或者背景复杂的情况下,传统的行人检测与跟踪算法效果不佳以及出现误检的问题,提出了一种基于孪生网络的 NMS 算法,可提高拥挤人群下行人检测跟踪的准确率。该方法利用较少遮挡的可见部分,去除多余的检测框,为了获取可见部分,提出了一种双盒模型(DBM)来同时预测行人的全身部分和可见部分,确保整个检测网络中两者之间的对应关系,以便在行人检测任务上实现更好的性能。在 CrowdHuman 数据集上进行了实验验证,实验结果表明,在拥挤情况下对行人检测具有较好的鲁棒性和检测精度且优于其他模型 5% 左右。

**关键词:** 孪生网络;行人检测;NMS

**中图分类号:** TP391.41 **文献标识码:** A **国家标准学科分类代码:** 580.2010

## Crowded pedestrian detection based on non-occluded part

Zhu Xiaolei Wu Xuncheng Xiao Ziyao Xiao Yongchao

(School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

**Abstract:** Aiming at the problems of poor performance of traditional pedestrian detection and tracking algorithms and false detections in public places, crowded or complex background, this paper proposed an NMS algorithm based on the siamese network, which is to improve the detection and tracking of pedestrians in crowded accurately. This method uses less occluded visible parts to remove redundant detection boxes. In order to obtain the visible part, this paper proposed a double-box model (DBM) to predict the whole body part and the visible part of the pedestrian at the same time, which is to ensure the correspondence between the two boxes in the entire network, so as to achieve a better pedestrian detection task performance. This paper has carried out experimental verification on the CrowdHuman dataset, and the experimental results show that it has good robustness and detection accuracy for pedestrian detection in crowded and is better than other models by about 5%.

**Keywords:** siamese network; pedestrian detection; NMS

## 0 引言

在实际各种应用场景中,行人检测都是关键组成部分,如无人驾驶、机器人、智能交通监控等领域。深度卷积神经网络<sup>[1-4]</sup>的迅速发展使得在行人检测的性能等方面也得到了提高且大大优于传统的基于手工特征的方法,但是,在相似人群密集遮挡的情况下,行人的检测与跟踪仍具有很大的挑战性<sup>[5]</sup>,不管在非遮挡行人检测上的表现如何,大多数检测器对严重遮挡行人的检测准确率都是有限的,最近有方法通过注意力机制<sup>[6-8]</sup>、特征变换<sup>[9-10]</sup>等来缓解行人遮挡的问题,然而,这些方法很少考虑到可见区域的应用,这在本质上有助于补偿被严重遮挡的行人所丢失的信息。对于遮挡问题,一般分为两类,即类间遮挡和类内遮挡。在类间遮挡中,其他物体在图片中位置靠前导致人体一部分不可见,比如汽车、建筑等。背景物体的特性会使模型混淆,导

致模型误检率变大。在类内遮挡中,当行人之间存在较大的重叠时,检测框就会无法判断向哪个目标靠近。因此,检测器可能会错误地在重叠区域给出大量的正信号。文献[11-12]则对两个人中间的边界框提出了额外的惩罚,使得这些提议框被紧紧的定位在真实值附近。但即使检测器成功地识别出人群中不同的人类实例,高度重叠的结果也可能被非最大抑制(NMS)的后处理所抑制。较低的 IOU 阈值会导致错过高度重叠的行人,而较高的 IOU 阈值自然会带来更多的误报。这使得现有的行人检测器陷入了一个两难的境地。文献[13]提出一种 Soft-NMS 算法,通过判断置信度高的边界框的重叠区域大小代替直接丢弃高度重叠的边界框的方式,从而达到降低检测分数的效果,但是引入了大量高度重叠的边界框的误检。文献[14]提出了一种 NMS 的动态自适应阈值的方法,首先预测一个密度

收稿日期:2021-05-08

图,之后根据预测的密度为不同的边界框在 NMS 中设置自适应 IOU 阈值,但是,密度估计本身就很难且密度与最优 IOU 阈值的精确匹配也很难确定。此外,边界框预测的不准确往往会导致真实密度与预测边界框的 IOU 不一致。这使得自适应阈值的 NMS 仍然是次优解。

为了提高拥挤行人的检测准确率,本文提出一种基于可见区域的 NMS 算法(V-NMS)。该方法利用行人受遮挡影响较小的可见部分对拥挤行人进行检测,避免了全身框在高度重叠的情况下难以进行 NMS 带来的麻烦,并且相对较低的 IOU 阈值也避免了大量的误检。

总而言之,本文主要有如下 3 点创新:1)融合了孪生网络进行行人检测;2)提出了一种基于可见区域的非极大值抑制的方法,有效地克服了原始的非极大值抑制方法漏检和误检的缺点;3)提出了一个双盒模型,可以同时预测同一行人的全身框和可见框,使得能够通过可见框与全身框之间的对应关系对行人进行检测,从而达到一个优于其他方法的检测精度。

## 1 可见区域拥挤行人检测框架

该部分提出一种孪生网络与非极大值抑制融合的网络,用于检测密集状态下的行人。该可见区域的拥挤行人检测网络框架如图 1 所示,框架由用于特征提取的孪生子网络、用于提议框生成的特征区域提议子网络和 V-NMS 模块组成。

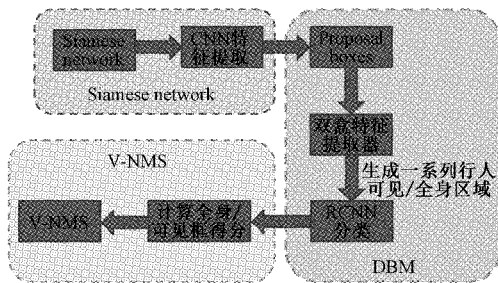


图 1 网络框架

### 1.1 孪生特征提取子网络模块

在孪生网络<sup>[15]</sup>中,本文使用改进后的 Alex-net 网络,将其中的卷积层 2 和卷积层 4 去掉。孪生特征提取子网络包含两个分支,1 个是模板分支,输入为历史帧中的目标块,用  $z$  表示;1 个是检测分支,输入为当前帧的目标块,用  $x$  表示。这两个分支在深度学习网络中共享参数,由于这两个分支在相同的变换中,适合后续的任务。为方便起见,本文将  $\phi(z)$  和  $\phi(x)$  表示为孪生子网络的输出特征图。经过孪生特征网络的特征提取,得到行人全身框与可见框的一个一一对应关系,输入到下一层网络中。

### 1.2 基于 Faster-RCNN 的双盒模型

由孪生特征提取子网络得出一一对应的全身框和可见框,本文提出一种双盒模型(double-box model, DBM),从

而获得行人的可见部分,同时可以预测行人的全身框和可见框。因此,双盒模型在标准的 Faster-RCNN 的基础上进行了修改,首先对行人的全身框和可见框的特征区域提取网络进行修改,区域提取网络首先生成一组对应每个行人的全身区域和可见区域的边界框;其次对提议框特征提取器进行修改,提议框特征提取器提取每组边界框的特征,并融合全身框和可见框的特征,以提供每一个行人的特征集成表示;最后对 RCNN 进行修改,将集成表示的形式输入到修改后的 RCNN 中,该修改后的 RCNN 对每组边界框进行分类,从而进一步完善预测的全身框和可见框。最终可以获得具有一一对应关系的全身框和可见框,这将有利于特征区域非极大值抑制的使用。

#### 1) 双盒区域提议子网络

双盒区域提议子网络的任务是生成具有一一对应关系的全身框和可见框,如图 2 所示。由于行人的全身区域和可见区域通常有很高的重合度,因此,可以从同一个锚框回归行人的全身框和可见框,此外,从同一个锚框回归的全身框和可见框之间会有内在的对应关系。每个行人样本都用两个边界框注释,这两个边界框分别指定其全身区域和可见部分区域,带注释的全身框  $F$  和对应的可见框  $V$  构成对  $Q = (F, V)$ , 本文将  $Q$  作为模型的真实值的基准,本文参考文献[16]中的提议匹配方法。对于行人锚框  $A$ , 如果满足式(1)的要求,即当锚框  $A$  的全身框 IOU 阈值大于  $\alpha_1$  且可见框的 IOU 阈值大于  $\beta_1$  时,则认为锚框  $A$  与真实值对  $Q = (F, V)$  是正匹配,式(2)和(3)分别为全身框和可见框 IOU 阈值的计算公式。

$$\text{IoU}(A, F) \geq \alpha_1 \text{ and } \text{IoU}(A, V) \geq \beta_1 \quad (1)$$

$$\text{IoU}(A, F) = \frac{\text{Area}(A \cap F)}{\text{Area}(A \cup F)} \quad (2)$$

$$\text{IoU}(A, V) = \frac{\text{Area}(A \cap V)}{\text{Area}(V)} \quad (3)$$

式中:  $\alpha_1$  和  $\beta_1$  分别是全身框和可见框的正阈值,根据本文的实验,当  $\alpha_1 = 0.7$  和  $\beta_1 = 0.7$  时,全身框和可见框模型表现最佳。

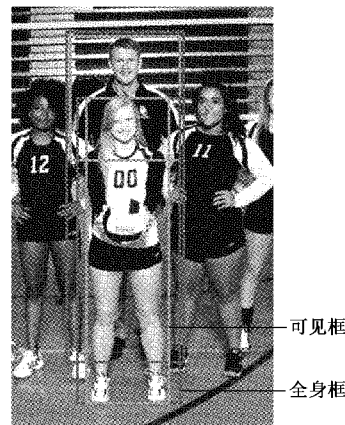


图 2 可见框与全身框对应关系

双盒区域提议网络的详细架构沿用了 Faster-RCNN<sup>[17]</sup>中的 RPN,唯一的区别在于输出层,除了预测全身区域提议框和可见区域提议框的位置,双盒区域提议网络还会预测每组匹配的得分,以显示该组匹配是否是当前行人的可见框与全身框的匹配。因此,对于每个密集的锚点,双盒区域提议网络会产生长度为 10 的结果  $(R_f, R_v, S)$ , 这里的  $R_f$  和  $R_v$  分别是全身框和可见框的 4 个偏移量向量  $(f_x, f_y, f_w, f_h)$  和  $(v_x, v_y, v_w, v_h)$ , 分别对应着全身框和可见框的真值。S 是 softmax 归一化后的二维置信度向量  $(S_+, S_-)$ , 训练中使用的损失函数与标准的 RPN 中的损失函数相同。

### 2) 双盒提议框特征提取器

由于双盒区域提议网络会产生全身框和可见框的一一对应关系,因此,本文提出了双盒提议框特征提取器,从而有效地整合来自这两种提议框的信息。有利于对行人全身区域和行人可见区域进行检测。

如图 3 所示,该图显示了两种提议框的功能集成方法,一种直接的方法是在 ROI Align 之后直接将全身框和可见框的特征向量  $F_f$  和  $F_v$  堆叠,经过 ROI Align 之后,由于改进后的 RCNN 网络中只使用了  $f_c$  层,因此不需要对两个提议框的特征进行空间对齐。实验结果表明,直接将特征进行线性融合的方法可以有效的提高性能。

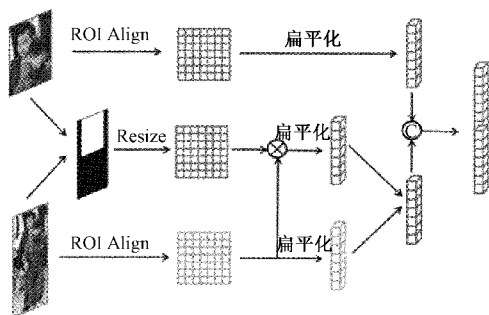


图 3 双盒提议框特征提取器特征融合方式

特征融合的第 2 种方法是基于注意力机制,该机制突出显示信息量更大的可见区域。具体来说,对于每对提议,本文都会生成一个可见部分的 mask,该 mask 显示是否要将全身框中的某个像素显示在可见框内,如果全身框中的某个像素也位于可见框中,则给注意力 mask 中的这个像素赋值 1,可见框之外的像素在注意 mask 中都设置为 0。然后,在 ROI Align 之后将注意 mask 调整为与提议特征相同的大小,即  $7 \times 7$ 。然后,本文以元素的方式将全身提议特征  $F_f$  与 mask 相乘,得到可见框的 mask 的注意力特征  $F_m$ ,最后,本文将  $F_v$  和  $F_m$  连接起来以获得该对的最终集成特征。

### 3) 双盒 RCNN

由上述双盒提议框特征提取器产生的集成特征用作双盒 RCNN 的输入,双盒 RCNN 用于根据提议框来进行全身和可见的行人框的预测。双盒 RCNN 的详细架构是基

于 Faster-RCNN 中的 RCNN 模块进行了如下修改,输入为双盒提议框特征提取器集成特征,双盒 RCNN 包含两个分叉的分支,分别位于两个共享层之后,用于全身边界框和可见边界框的预测。每个分支都具有与标准 Faster-RCNN 完全相同的体系结构和输出。和双盒 RPN 相同,双盒 RCNN 中的本质问题是如何将提议框分配给真实值,本文使用了一个和双盒 RPN 中的锚分配方法十分相似的策略。具体而言,对于一组全身和可见提议框注释  $Q = (F, V)$ , 如果提案  $P$  可见框和全身框的阈值满足式(4)的要求。

$$\text{IoU}(P_f, F) \geq \alpha_2 \text{ and } \text{IoU}(P_v, V) \geq \beta_2 \quad (4)$$

其中,一对提案  $P = (P_f, P_v)$  为正,则该提议为正。经过多次实验验证,  $\alpha_2, \beta_2$  的最佳值为 0.5, 0.5。训练中采用与标准的 Faster-RCNN 相同的损失函数。

综上所述,双盒模型对 Faster-RCNN 的主要修改几乎没有带来额外的计算,但却带来了大量的性能提升。下一部分的实验结果将详细说明所提出的模型的有效性。

### 1.3 可见区域的非极大值抑制

考虑到现有的标准的 NMS 和自适应阈值的 NMS 很难得到精确的匹配,从而本文提出一种基于可见区域的 NMS 的方法。基于可见区域的 NMS 与标准的 NMS 的方法最大的不同点在于 IOU 值的计算。该方法不是直接计算出两个完整的边界框的 IOU 作为他们的重叠度,而是利用可见区域之间的 IOU 来判断两个全身的边界框是否重叠。当定位的是两个不同行人的边界框,虽然两个全身的边界框之间的 IOU 较大,但是可见区域之间的 IOU 通常较低,相反,当两个边界框定位的是同一个行人时,全身区域和可见区域都将具有较大的 IOU,因此,采用两个可见框之间的 IOU 作为指标效果较好,可以通过可见区域的重叠度来判断两个全身框是否属于同一个行人。因此,在可见区域的基础上,一个相对较低的 IOU 阈值可以充分的去除定位同一行人的冗余框,也避免了大量的误检。

## 2 实 验

### 2.1 数据集

本文选用 CrowdHuman 数据集, CrowdHuman 数据集数据量比较大,训练集 15 000 张,测试集 5 000 张,验证集 4 370 张。训练集和验证集中共有 470 000 个实例,约每张图片包含 23 个人,同时包含各种各样的遮挡, CrowdHuman 为每个人类实例提供了 3 类边界框注释:头部边界框、人类可见区域边界框和人体全身体边界框注释。本文所有的实验都在全身和可见区域的环境下进行的,在训练集上训练模型,并在验证集上评估模型。

对于 CrowdHuman 数据集,本文采用 Faster-RCNN 和 RetinaNet<sup>[18]</sup> 作为基准检测器,采用特征金字塔网络和 Resnet-50 作为骨干网络,为了提取更精确的特征,本文采用 ROI Align 进行特征提取。将全身和可见物体的锚框纵横比设置为  $[0.5, 1, 2]$ 。由于 CrowdHuman 数据集中的图

像具有各种形状,因此本文将调整大小,使得短边为800 pixel,而长边小于1400 pixel。

## 2.2 评估指标

本文出于不同的目的采用平均精度 AP 和 MR 两个标准,平均精度是最流行的检测标准,AP 反映了检测结果的精度和查全率,在本文的实验中,本文根据实验数据发现 AP 对召回分数更为敏感,尤其是在 CrowdHuman 等拥挤的数据集上。AP 越大性能越好;MR 是 $[10^{-2}, 1]$ 中对每幅图像假阳性的对数平均丢失率的缩写,通常用于行人检测。MR 对假阳性非常敏感,尤其是置信度高的假阳性将严重损害 MR 的比例。MR 越小则性能越好。

## 2.3 CrowdHuman 的结果

### 1) 主要结果

为了彻底评估本文提出的方法的性能,本文在 CrowdHuman 数据集上进行了大量实验,并根据以上两个评估指标对性能进行了评估。选择 MR 作为主要指标。图3所示为基准、无孪生网络和本文模型在 CrowdHuman 验证子集上提出的方法的性能。为了公平比较,图3中所有模型在超参数上都具有相同的设置。如图4所示,本文重新实现的 FPN 金字塔模型<sup>[19]</sup>基线在全身检测时达到了54.96%的 MR,在可见身体检测时达到了47.52%的 MR,本文还验证了孪生网络的有效性,在无孪生网络的情况下,相较于本文方法,在全身检测和可见身体检测 MR 分别降低了4.29%和7.3%,因此,足以验证本文方法的有效性。基于优化后的基准,本文方法分别带来 MR、AP 和 Recall 显著的4.4%,5.46%和3.77%的改进,这显著证明了本文方法的能力。本文还单独分析了提议的模块的贡献,并逐步用本文的模块替换基线模型中的组件进行相关的实验和分析。

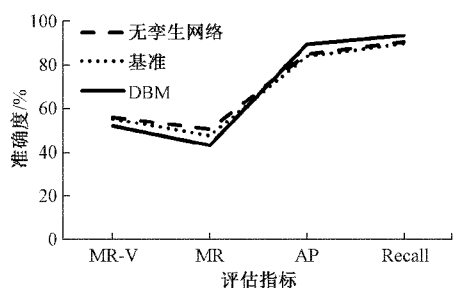


图4 基准和模型在验证子集上的性能

### 2) 不同超参数设置的影响

为了评估基于双盒的样本单元在整个网络中的功效,本文建立了一个简单的双盒模型版本,该模型简单地用带有一组全身框和可见框的RPN和RCNN替换了Faster-RCNN中的RPN和RCNN。

在非双盒模型(non-double box model, NBM)中,来自RPN的全身提案和可见提案被馈送到两个单独的RCNN模块中。这两个单独的RCNN模块分别负责全身部分和可见部分区域检测。每个RCNN模块具有与Faster-

RCNN中标准模块相同的体系结构。本文在非双盒模型上进行了一组实验,以找出最合理的超参数的值。为了使非双盒模型具有与基线一致的超参数设置, $\alpha_1$ 和 $\alpha_2$ 分别固定为0.7和0.5。全面研究了不同的 $\beta_1$ 和 $\beta_2$ 的影响,结果如表1所示。由实验结果可知,当 $\beta_1 = 0.7$ 和 $\beta_2 = 0.5$ 时,非双盒模型的效果很好。当 $\alpha_1$ 和 $\alpha_2$ 固定时, $\beta_1$ 和 $\beta_2$ 控制本文基于双盒的样本单元中训练时可见区域和全身区域之间的平衡。当 $\beta_1$ 和 $\beta_2$ 较小时,许多质量较低的可见提议也会被计数,这会损害非双盒模型在可见检测结果上的性能。较大的 $\beta_1$ 和 $\beta_2$ 排除了堆叠后不清晰的可见提议,同时,这种设置会排除一些堆叠后良好的全身阳性训练示例,从而导致不良的全身检测结果。如表1所示, $\beta_1 = 0.7$ 和 $\beta_2 = 0.5$ 在两个注释类别之间实现了良好的权衡。因此,本文使用 $\alpha_1 = 0.7, \alpha_2 = 0.5, \beta_1 = 0.7$ 和 $\beta_2 = 0.5$ 。

表1 不同的 $\beta_1, \beta_2$ 对非双盒模型检测效果的影响

方法	$\beta_1$	$\beta_2$	MR-V	MR	MR-V+MR
基准	—	—	54.96	47.52	102.48
	0.8	0.5	54.12	47.28	101.40
	0.6	0.5	54.27	46.13	100.40
NBM	0.7	0.5	54.58	45.27	99.85
	0.7	0.6	54.60	47.33	101.93
	0.7	0.4	55.59	44.92	100.51

### 3) 双盒提议框特征提取器的影响

双盒模型是在NBM的基础上配备双盒提议框特征提取器,如图5所示,本文对双盒模型和非双盒模型进行了比较,从图5中可以看出,1.2节中提出的两种特征集成方法均对MR-V和MR带来了显著改进。比较简单的一种方法是直接合并全身提议和可见提议的功能,可将MR-V和MR分别提高2.93%和2.25%,而带有注意机制的双盒提议框特征提取器则表现出更好的性能。双盒模型和非双盒模型之间的如此大的差距完全证明了本文提出的双盒提议框特征提取器模块可以有效地从基于双盒的样本单元中提取和集成特征。此外,从图5可以了解到,双盒提议框特征提取器通过改善模型在可见区域上的性能,使得基于可见区域的非极大值抑制方法的性能更好,这说明了双盒提议框特征提取器模块的必要性。

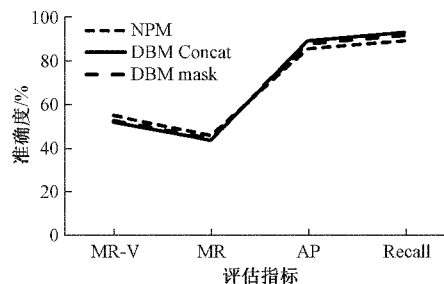


图5 双盒模型和非双盒模型对比

## 4) MR-V 与 V-NMS 之间的关系

为了证明 V-NMS 的有效性,本文用 V-NMS 代替了原始的 NMS。从表 2 的 3 组实验结果中可以得出,在所有设置下,V-NMS 能同时提高 AP 和 Recall 的性能,而当 V-NMS 用于非双盒模型时会使 MR 恶化。为此本文将介绍 MR-V 和 V-NMS 两者之间的关系。因为 V-NMS 是根据两个可见区域边界框的 IOU 来判断两个全身框是否重叠,因此可见区域的预测的精度至关重要。然而,本文认为与预测的可见边界框的绝对定位质量相比,捕获两个人类实例之间的重叠程度的相对定位质量更为重要。

本文通过简单的例子验证上述观点,假设所有可见区域的预测与全身区域的预测完全相同,在这种情况下,MR-V 将很差。但是,在 V-NMS 期间使用此类可见区域预测会导致与原始 NMS 完全相同的全身检测结果,但不如 MR-V 差。该示例明确表明,对于 V-NMS 的全身预测而言,不良的 MR-V 不一定会导致不良的 MR。更具体地讲,尽管较低的 MR-V 可能可以更好地模拟每个实例对之间的重叠,这将进一步有益于全身检测,综上所述,但是 MR-V 的绝对值不是决定全身检测性能的决定性因素。如表 2 所示,本文的实验结果也证明了这一点。随着 MR-V 的降低,V-NMS 的性能变得更好,而 MR-V 低于 52% 足以对通过 V-NMS 进行全身检测产生积极影响。

表 2 MR-V 与 V-NMS 之间的关系

方法	$\beta_1$	$\beta_2$	MR-V	MR	AP	Recall
NBM	0.7	0.4	55.59	44.92	86.53	88.85
NBM+V-NMS	0.7	0.4		45.98	88.64	93.33
NBM	0.7	0.5	54.58	45.27	85.95	88.20
NBM+V-NMS	0.7	0.5		46.51	89.03	93.82
DBM	0.7	0.5	51.74	45.11	85.96	89.40
DBM+V-NMS	0.7	0.5		42.96	89.07	93.51

## 5) 关于 V-NMS 的进一步分析

图 3 实验结果表明,V-NMS 可以显著改善 AP 和 Recall,而 MR 增益不明显。这种现象是由 MR 和 AP 之间的差异引起的。MR 和 AP 之间的主要区别在于感兴趣区域的预测分数范围。MR 只关心分数高于最高得分的假阳性的预测边界框。相反,AP 将评分在 0~1 之间的所有检测结果都考虑在内。因此,仅一小部分预测结果会影响 MR。如引言中所述,在拥挤的情况下,检测器趋向于生成许多高分的假阳性,这些假阳性很难通过 NMS 删除。因此,结果中的大量高分误报使每幅图像中的假阳性 FPPI (false positive per image) 迅速达到 1。为了证明这一点,本文计算了所有测试图像中得分最高的假阳性的平均得分。事实证明,该值非常高,有时甚至超过 0.9。这种现象表明,MR 只会受到得分高于 0.9 的边界框的影响。因此,MR 反映了检测结果中得分较高的部分的性能,而 AP 则

测量了所有检测到的边界框的性能。这种差异导致 V-NMS 的 MR 和 AP 的增益之间存在较大差异。总之,尽管 V-NMS 并未在 MR 上带来大的改进,但它极大地提高了 AP。与原始 NMS 相比,AP 的改进有力地证明了 V-NMS 不仅可以保留更多的真实例结果,而且可以引入更少的假阳性结果。

## 3 结 论

本文主要创新在于融合了孪生网络进行行人检测,首先对数据集进行特征提取,接着提出了 V-NMS 来有效地去除拥挤情况下由于冗余框带来的误报,V-NMS 通过两个可见区域边界框的 IOU 来判断两个全身框是否重合,为了验证这一想法,本文提出了一种双盒模型来预测全身框和可见框,本文在 CrowdHuman 数据集上进行了实验验证,最终结果表明,该方法可以达到最优的结果,有力的证明了该方法的优越性。

## 参考文献

- [1] 崔少华,汪徐德,王江涛,等. 高斯建模和卷积神经网络联合的红外视频行人检测方法[J]. 电子测量与仪器学报,2020,34(5):140-148.
- [2] 孙锐,王慧慧,叶子豪. 融合深度感知特征与核极限学习机的行人检测[J]. 电子测量与仪器学报,2019,33(2):39-47.
- [3] 刘丹,马同伟. 结合语义信息的行人检测方法[J]. 电子测量与仪器学报,2019,33(1):54-60.
- [4] 白中浩,王鹏辉,李智强. 基于 Stixel-world 及特征融合的双目立体视觉行人检测[J]. 仪器仪表学报,2017,38(11):2822-2829.
- [5] LEIBE B, SEEMANN E, SCHIELE B. Pedestrian detection in crowded scenes[C]. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR'05), IEEE, 2005, 1:878-885.
- [6] PANG Y, XIE J, KHAN M H, et al. Mask-guided attention network for occluded pedestrian detection[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 4967-4975.
- [7] ZHANG S, CHEN D, YANG J, et al. Guided attention in CNNs for occluded pedestrian detection and re-identification [J]. International Journal of Computer Vision, 2021, 129(6): 1875-1892.
- [8] ZHANG S, YANG J, SCHIELE B. Occluded pedestrian detection through guided attention in cnns[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:6995-7003.
- [9] CAO J, PANG Y, LI X. Learning multilayer channel features for pedestrian detection[J]. IEEE Transactions on Image Processing, 2017, 26(7): 3210-3220.

- [10] SONG T, SUN L, XIE D, et al. Small-scale pedestrian detection based on topological line localization and temporal feature aggregation [C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 536-551.
- [11] WANG X, XIAO T, JIANG Y, et al. Repulsion loss: Detecting pedestrians in a crowd[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7774-7783.
- [12] ZHANG S, WEN L, BIAN X, et al. Occlusion-aware R-CNN: Detecting pedestrians in a crowd [C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 637-653.
- [13] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS improving object detection with one line of code [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017:5561-5569.
- [14] LIU S, HUANG D, WANG Y. Adaptive NMS: Refining pedestrian detection in a crowd [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019:6459-6468.
- [15] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking [C]. European Conference on Computer Vision, Springer, Cham, 2016:850-865.
- [16] ZHOU C, YUAN J. Bi-box regression for pedestrian detection and occlusion estimation[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018:135-151.
- [17] 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, 2016, 39(6): 1137-1149.
- [18] GORDEN D, FARHADI A, FOX D. Rc3: Real-time recurrent regression networks for object tracking[J]. IEEE Robot. Autom. Lett., 2017, 3(2):788-795.
- [19] 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.

### 作者简介

朱肖磊, 硕士研究生, 主要研究方向为计算机视觉、图像处理等。

E-mail: zhuxl96@163.com

吴训成(通信作者), 博士, 教授, 主要研究方向为智能网联汽车技术等。

E-mail: 18861120776@163.com

肖子遥, 硕士研究生, 主要研究方向为计算机视觉、图像处理等。

E-mail: xiaoziyao2019@163.com

肖永超, 硕士研究生, 主要研究方向为计算机视觉、图像处理等。

E-mail: 1667416126@qq.com