

# 复杂路面小尺度行人检测综述<sup>①</sup>



赵书, 陈宁

(西安工程大学 电子信息学院, 西安 710048)

通信作者: 陈宁, E-mail: chennvictor@gmail.com

**摘要:** 行人检测技术是智能交通和智能车辆发展的一个重要方向, 同时也是道路安全的重要保障, 直接影响着车辆控制系统对路况的判断. 在实际应用场景中, 小尺度行人实例占比非常高, 但小尺度行人检测一直是行人检测任务中一个困难且具有挑战性的问题. 而当智能汽车处于复杂的交通环境中时, 小尺度行人的精准检测可以使控制系统有时间提前预警并及时避让, 对于保障汽车安全平稳行驶起重要作用. 随着深度学习的快速发展, 小尺度行人检测技术取得了突破性的进展, 目前该技术的发展处于快速发展时期. 为了进一步促进小尺度行人检测技术的发展, 本文对小尺度行人检测技术的最新方法进行了全面研究. 本文首先分析了小尺度行人检测面临的几大挑战, 并对目前最新的小尺度行人检测网络进行了归类 and 总结. 本文从多尺度表示、上下文信息、新的训练和分类策略、尺度感知和超分辨率 5 个方面对现有的深度学习方法进行了分析和讨论, 其中多尺度学习方法为当前处理小尺度行人检测的主流方法. 同时简要介绍了行人检测常用的评价指标和数据集, 并在 Caltech 等通用数据集上对一些主流方法进行了性能评价. 此外, 本文还对 5 类方法进行了总结和对比. 最后, 本文从多个方面提出了行人检测技术中亟待解决的问题和未来发展的方向和任务.

**关键词:** 目标检测; 深度学习; 卷积神经网络; 小目标; 行人检测

引用格式: 赵书, 陈宁. 复杂路面小尺度行人检测综述. 计算机系统应用, 2022, 31(7): 1-11. <http://www.c-s-a.org.cn/1003-3254/8545.html>

## Review on Small-scale Pedestrian Detection Technology for Complex Pavement

ZHAO Shu, CHEN Ning

(School of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, China)

**Abstract:** Pedestrian detection technology is an important research direction for the development of intelligent transportation and intelligent vehicles, and it is also an important guarantee for road safety, which directly affects the judgment of a vehicle control system on road conditions. In practical application scenarios, small-scale pedestrian instances account for a very high proportion, but small-scale pedestrian detection has always been a challenging problem in pedestrian detection tasks. When an intelligent vehicle is in a complex traffic environment, the precise detection of small-scale pedestrians can make the control system give a warning in advance and help avoid collision in time, which plays an important role in ensuring the safe and stable driving of the vehicle. With the rapid development of deep learning, groundbreaking progress has been made in the fast-growing small-scale pedestrian detection technology. To further promote the development of small-scale pedestrian detection technology, this study conducts comprehensive research on the latest methods of small-scale pedestrian detection technology. To start with, this study analyzes several challenges faced by small-scale pedestrian detection and classifies and summarizes the latest small-scale pedestrian detection networks. The existing deep learning methods are analyzed and discussed from five aspects, namely multi-scale representation, context information, new training and classification strategies, scale perception, and super-resolution.

<sup>①</sup> 基金项目: 国家自然科学基金青年基金 (61902301)

收稿时间: 2021-09-23; 修改时间: 2021-10-25; 采用时间: 2021-10-29; csa 在线出版时间: 2022-05-31

Among them, the multi-scale learning method is the mainstream of small-scale pedestrian detection. Meanwhile, we briefly introduce the commonly used evaluation indicators and datasets for pedestrian detection and evaluate the performance of some mainstream methods on general datasets such as Caltech. In addition, five methods are summarized and compared in this study. Finally, this study proposes the urgent problems to be solved in pedestrian detection technology and the direction and tasks of future development from multiple aspects.

**Key words:** object detection; deep learning; convolutional neural networks (CNN); small targets; pedestrian detection

行人检测是一个预测、定位和标记的过程,在给定的视频或图像中检测是否有行人存在,并返回目标对象的位置信息.此技术作为计算机视觉的一个重要课题有着深刻的研究意义,并且在自动驾驶、视频监控、人机交互等领域都有着广泛的应用.此外,行人检测是目标检测中的一个特例,其研究成果可以对其他目标检测方法起很好的推动作用.

随着现代控制技术和汽车技术的发展,智能汽车已经实现了辅助驾驶乃至自动驾驶功能.行人是交通场景中不可或缺的一部分.同时,道路行人检测也是智能汽车能够应用在各种交通场景中的基本前提.智能汽车通过车载摄像头获取道路信息,然后利用行人检测技术,自动准确的检测出道路前方的行人,及时反馈及预警,这有利于保障行车安全和行人通行安全.近年来,随着智能交通的提出以及无人驾驶等领域的需要,使得行人检测技术有了一个更高精度的要求.当汽车处于复杂的交通环境中时,对于那些尺度较小且分辨率低的行人检测尤为重要,因为远处的小尺度行人更容易被忽视从而引发事故.因此自动驾驶汽车必须具有在一定距离内检测到小尺度行人的能力,以便能够让控制系统可靠且平稳地避免与之碰撞<sup>[1-3]</sup>.

目前行人检测技术已经实现了较高的准确率,但是由于行人尺度变化、低分辨率、遮挡等问题的存在,行人检测仍然受限应用于场景.如何实现小尺度行人的精准检测一直是行人检测任务中的需要解决的关键问题之一.文献[4]中将行人根据像素高度分为近、中、远3个等级对应不同的尺度对象.如图1所示.然而,在许多行人检测应用场景中,小尺度行人的占比更高.据统计,在加州理工学院的行人数据集中,实例高度小于80像素的占83%.人眼在检测小尺度行人时没有太大难度,但现有的行人检测方法在检测大尺度行人时拥有很好的性能,在小尺度行人检测上性能有所下降.



图1 小尺度行人样本示例

近年来,越来越多的学者开始研究如何解决小尺度行人的检测问题,也取得了一定的进展,但是仍不满足实际应用的要求.当前,小尺度行人检测技术现处于快速发展期.目前的研究工作与实际应用要求之间的不匹配,使得对小尺度行人检测问题的解决方法进行探索和讨论尤为重要.然而现在有很多讨论小尺度目标检测问题和行人检测问题的综述性文章<sup>[5-8]</sup>,但讨论小尺度行人检测问题的综述却几乎没有.因此本文对小尺度行人检测方法进行了全面的分析和总结,从而更好地促进行人检测技术的进一步发展和提升.

如今现有的行人检测的方法可以分为两类:传统方法和基于深度学习的方法.传统的行人检测方法大多基于手工提取特征+分类器的结构.对传统方法来说,手工提取特征的设计对于检测器性能的好坏至关重要.然而手工特征不能捕捉大规模数据集的多层次表示,因此这类方法受到了特征提取的限制,对类内可变性的鲁棒性较差.

为克服传统手工特征的缺点,文献[9]提出了深度卷积神经网络.相对于传统方法来说,深度学习的提出大大简化了检测的工作.由于深度学习对视觉特征的强大表征能力,使得其现如今在目标检测领域占主导地位.因此,本文分析和探讨了深度学习方法中小尺度行人检测存在的问题和解决方法.首先,以小尺度行

人检测方法的不同思路为依据将现有方法分为5类并介绍其典型模型.除此之外,我们还研究了行人检测的数据集和评估指标.同时,对每一类方法进行对比分析和综合评价,并提出了未来需要解决的问题和努力的方向.

## 1 小尺度行人检测方法

卷积神经网络(CNN)是一类相当具有代表性的深度学习神经网络,也是行人检测中最常用的网络.CNN网络模型主要可分为两类:(1)以RCNN系列为代表的方法,预先回归一次边界框,然后利用骨干网络再进行训练,被称为两阶段检测器,该类方法精度较高但检测时间长;(2)以YOLO、SSD等为代表的只进行一次回归和评分的方法,称为一阶段检测器,该类方法检测速度快但精度稍差.由于CNN会随着网络的加深弱化小尺度对象的特征表示,所以这些模型只适用于粗粒度的分类任务,在细粒度的分类任务如小目标检测和语义分割中有一些限制.因此无论是一阶段检测器还是两阶段检测器,对于小尺度行人的检测都存在几大挑战,如表1所示.

表1 小尺度行人检测存在的挑战

序号	存在的挑战	具体描述
1	特征不足	小尺度行人对象所覆盖的像素少,保留更少的信息,特征表示弱.并且在行人检测技术的典型应用场景中,伴随小尺度行人实例出现的还有低分辨率和背景模糊等问题.
2	信息丢失	经过卷积神经网络中多次卷积和池化操作后,会不可避免的导致小尺度行人语义信息的丢失,同时特征图将更多不必要的背景信息包含到小尺度行人中,从而削弱了小尺度行人的特征表示.
3	感受野不匹配	现有的锚框策略不能合适的匹配小尺度对象的感受野,大多数方法是在一组固定的卷积特征映射上滑动一组固定的窗口来生成多尺度的建议.然而,固定的感受野不能覆盖物体在自然场景中出现的多个尺度,使得对于小尺度行人检测性能较差.
4	不易检测和定位	由于尺寸较小,小尺度行人对象的位置有更多的可能性,例如角落或与其他对象重叠的区域.并且很难区分复杂路面上的小尺度行人和噪声杂波,准确定位他们的边界,这就意味着检测时对定位的精度有更高的要求.

此外,关于小尺度行人检测的经验和知识非常有限,因为大多数先前的工作都是围绕大尺度行人检测

问题展开研究的.因此,有必要对现阶段的小尺度行人检测工作进行梳理,探索有效的解决方法,提高其检测精度.

随着基于深度学习的目标检测技术的发展,许多适用和解决小尺度行人检测难问题的新型网络模型被提出.本文从多尺度表示、上下文信息、训练和分类策略、尺度感知以及超分辨率5个方面总结了现有的小尺度行人检测算法.

### 1.1 基于多尺度表示的方法

对检测和识别任务来说,特征图是精准定位和分类的关键.由于行人较小的尺度和低分辨率表示,行人的位置等细节信息在高层特征图中逐渐丢失.最初解决该问题主要有两种策略:(1)构建图像金字塔<sup>[10]</sup>,将图像缩放到不同比例输入到检测器中,但由于潜在的图像比例很大,计算每个比例下的特征非常耗时,因此这种方法已经很少使用;(2)基于单个特征图生成区域建议,使用不同比例和大小的锚框对应不同的感受野,如Faster R-CNN<sup>[11]</sup>中的多尺度区域建议网络RPN,而此类方法检测小尺度对象的劣势在于,只利用最后一层特征生成建议,使得具有大的感受野的高层特征因为分辨率太低而不能准确的识别到小目标.

针对上述问题,Liu等人提出的SSD分类器首次引入了特征多尺度表示的思想<sup>[12]</sup>,在不同网络分层中提取特征并依次进行边界框回归和分类,SSD利用浅层特征来检测小尺寸对象有一定的提升效果.随后,Cai等人构建了多尺度深度卷积神经网络MS-CNN<sup>[13]</sup>,通过引入反卷积特征上采样替代输入上采样来提高了特征图的分辨率,并使用多层来匹配不同比例的对象,增强了小尺度目标的检测能力.

然而MS-CNN和SSD并没有充分利用网络中的底层信息,直接从网络的高层构建金字塔.为进一步充分利用特征信息,Lin等人进一步结合了单一特征图、集成特征和特征金字塔层次的优点,在Fast RCNN中构建了特征金字塔网络(FPN)<sup>[14]</sup>.FPN采用自顶向下的结构,将低分辨率的高层语义图和高分辨率的低层语义图进行融合,每一层特征都进行独立预测,无需成本但在小尺度行人上有很好的表现.随后又出现了一系列基于FPN改进的方法.由于SSD速度很快但精度不够理想,Li等人提出的FSSD分类器(feature fusion single shot multibox detector)<sup>[15]</sup>将FPN结合到SSD中,将具有不同尺度的多层特征连接在一起,随后下采

样构建新的特征金字塔,大大提升了检测精度.基于FPN,文献[16]中引入了一个跨尺度特征聚合模块,通过融合鲁棒的语义和不同尺度行人的准确定位来增强特征金字塔表示.

Cao等人认为FPN等网络中小尺度行人的语义水平还不够高,于是提出了一种多分支高级网络MHN-D将底层特征转化为高级语义特征<sup>[17]</sup>.MHN-D的分支具有不同的空间分辨率和感受野,适合检测多个尺度的行人.MHN-D网络的分支之间采用跨层连接来提高检测性能,并利用空洞卷积增加特征图的分辨率,为小尺寸行人的定位保留更多的空间信息.

图2展示了多种方法结构的对比.具体来说,多尺度表示就是一种将包含丰富语义信息的深层特征和具有详细位置信息的浅层特征相结合的策略.这类策略有效地避免了特征图的重复计算,并提升了检测精度.并且多尺度表示的方法也从基于多个单层特征预测趋向于多层特征融合,目前此类方法已经成为克服小尺度行人检测中信息丢失和感受野不匹配问题的主流方法.

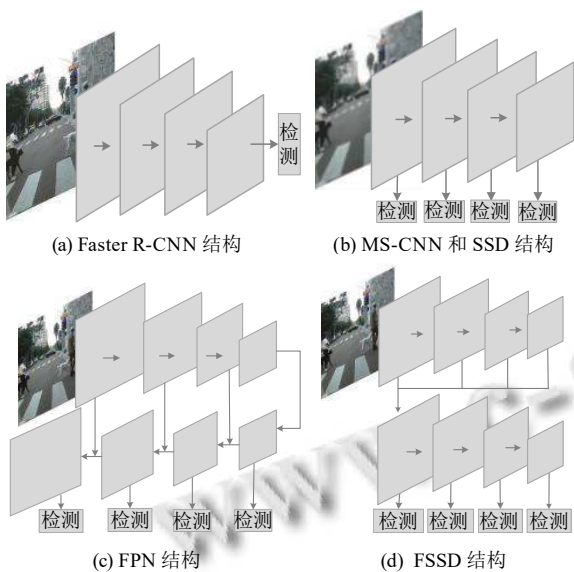


图2 多尺度特征学习网络结构对比

## 1.2 基于上下文信息的方法

上下文信息充分利用视觉对象及其共存环境之间关系,在目标检测中起着重要作用.文献[18]已经证明背景等信息对于小规模目标检测是有帮助的.大尺度的行人可以给检测器提供足够的感兴趣区域特征,相比之下,从小尺度行人对象中提取的感兴趣区域特征非常少.因此提取上下文信息作为原始感兴趣区域特

征的补充可以有效地解决小尺度行人特征不足和定位不准确的问题.

然而像FPN等网络简单地将不同层的语义图连接起来隐式的学习上下文信息,所得到的组合特征还不足够丰富.因此明确的挖掘上下文信息具有一定的意义.为了进一步改善检测性能,文献[19]提出了一种扩展特征金字塔的FPN++框架,在FPN检测器的头部引入上下文感知检测模块,利用上下文信息进行分类和回归.文献[20]提出了一种上下文感知的DIF R-CNN行人检测方法,通过集成反卷积模块引入额外的上下文信息.具体地说,网络将反卷积层和初始特征图相结合生成用于收集附加信息的合成特征图,为检测提供更多的视觉细节和语义上下文表示.ALNet<sup>[21]</sup>中的卷积预测块CPB将残差学习和多尺度上下文语义信息结合到一起,构建了一个上下文编码块.而文献[22]认为简单地将不同强度的语义特征图结合起来,会导致语义不一致从而不能充分的提取到对象周围的上下文信息.对此,提出了一个利用可分离大核卷积作为横向连接的语义转换模块,将弱语义特征先经过3个不同分支的特征映射后进行连接用于跳跃层融合,缓解语义不一致性并提取更多的上下文信息.

表2展示了几种上下文信息不同引入方式.与多尺度表示的方法的初衷相似,基于上下文信息的方法也旨在给检测网络提供更多的信息.获取上下文信息主要通过不同特征层之间的跳跃连接实现,此外引入空洞卷积可以获得更多的信息.所获取的上下文信息主要为兴趣域附近的信息,通过学习对象和周围环境之间的关系来提升检测效果.

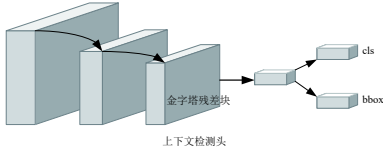
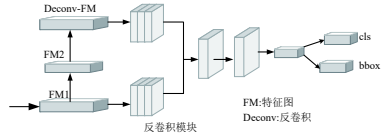
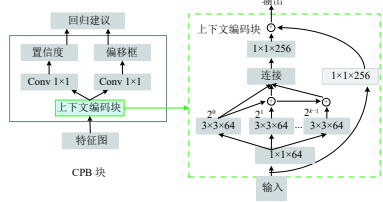
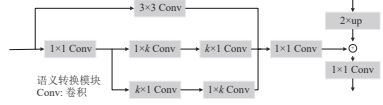
## 1.3 基于新的训练和分类策略的方法

网络训练和分类器的性能对检测器有很大的影响,分类器对行人检测的精准度起着决定性的作用,不同的训练策略则影响着分类器的能力.使用不同分类策略对不同尺度的特征图进行分类,有助于增强分类器对低像素特征的敏感性,提升检测器对小尺度行人的检测和分类能力.

随着CNN的在目标检测上的成功,行人检测经历了基于CNN特征方法和基于端到端的方法两个阶段.复杂性感知网络(CompACT)通过利用手工特征和CNN的特性,在每个阶段对准确性和复杂性进行权衡<sup>[23]</sup>,先用简单的特征筛选出行人可能存在的位置,将高复杂性的特征用于级联的后期.训练用于检测不同

尺度行人的分类子网络,对于增强检测器处理低像素行人的能力非常有效.而在文献 [24] 中研究发现 Faster R-CNN 中的区域提议网络 (RPN) 作为一个独立的行人检测器表现良好,但在下游分类器中由于小尺度行人的影响导致性能降低.因此他们提出的 RPN+BF 方法,在共享的高分辨率卷积特征地图上训练级联增强森林分类器,充分挖掘小尺度行人的特征,不受预训练网络结构的限制.在 F-DNN 中,不再使用单个的下游分类器而是利用多个并行的深度分类器结合软过滤器来进一步验证每个建议 [25].

表 2 上下文信息方法对比

网络名称	网络结构	优点
FPN++		适用性强,可移植性高
DIF R-CNN		精准度高,鲁棒性高
ALFNet		网络结构简单,检测速度快
LBST		漏检率低,定位准确性高

Zhang 等人考虑到行人身体形状的不对称性,提出了一种用于小尺度行人的非对称多级网 AMS-Net [26]. 检测过程中网络根据行人身体型设计了非对称卷积来捕捉行人身体的紧凑特征,并采用非对称锚框来生成矩形建议.而 Song 等人 [27] 研究发现大部分方法都过于依赖训练集的标注,而标注不精准则会影响检测性能.于是他们提出了一种基于拓扑线定位的方法如图 3 所示,通过建立不同尺度行人拓扑信息作为标注用于训练阶段,同时利用聚集相邻帧的特征获取时间信息,可以自动适应小尺度行人.

此外,行人检测有两种主要方法即目标检测和语义分割,这两种方法本质上具有一定的相关性.目标检测在定位对象时表现良好,但缺乏对象边界的信息.而

语义分割在区分之间的像素边界方面表现得很好,因此,另一种思路是利用语义分割来提升检测和分类的准确性,并且已经取得了很好的性能.



图 3 TLL 拓扑标注方法

文献 [28] 提出了一种分割注入的两阶段网络 (SDS RCNN), 在顶层网络中增加一个语义分割分支,并在训练期间注入语义信息,将语义分割作为行人检测辅助信息与区域检测结合,优化下游行人检测器. SDS-RCNN 只是增加了语义分割分支,并没有将语义分割结果直接用于行人检测.而 SSA-CNN 从具有不同分辨率的多个网络层执行语义分割 [29], 将各种尺度粒度的语义信息集成到共享的特征映射中,给检测提供像素级的分类信息,提高对小尺度行人的分类能力.

基于训练和分类策略的方法采用不同的训练方法和分类器,以最终获得更适用于小尺度行人的检测器.所设计的训练分类策略需满足检测器对各种场景的适应性,并且研究发现将多种计算机任务如目标检测、语义分割等结合起来用于行人检测,会得到更加丰富的信息从而会大幅度的提高单独任务的精准度.

### 1.4 基于尺度感知的方法

由于特征不足及感受野不匹配导致网络发生了漏检是小尺度行人检测效果差的主要原因之一.在其他方法中,不同尺度的行人被网络进行统一检测.然而,不同尺度的行人之间有着巨大的类内差异,比如,大尺度行人有着丰富的细节信息,而小尺度行人往往模糊不清.为了解决行人间的类内差异问题,有些方法采用“分治”的思想,充分利用不同尺度实例下的显著特征.具体来说,网络将不同尺度的行人分开处理,分别捕获特定尺度下的特征从而实现检测特定范围尺度下的实例.

为解决行人的类内间距,SAF RCNN [30] 的规模感知模型将大、小尺度行人作两个检测任务在两个子

网络中完成, 根据建议的高度设置不同尺度子网络的权重, 并采用一个感知加权层来融合检测结果作为最终的输出结果. 但由于 SAF RCNN 仅利用最后一层卷积检测行人, 使得小尺度行人检测没有达到最好的性能. 对此, Han 等人提出了一种小规模感知网络 SSN, 充分利用卷积层来提高检测性能<sup>[31]</sup>. 为了生成更有效的检测小规模行人的建议区域, 提出了一种尺度建议网络. 该方法将不同的卷积层与反卷积合并, 获得每个

特征点描述更详细的特征图, 尺度建议网络用于生成一些更有利于捕捉小尺度对象的建议区域. 而在文献 [32] 的 GDFL 方法中, 如图 4 所示, 使用尺度感知的行人注意模块和放大-缩小模块 (ZIZOM) 来实现了更稳定的行人检测. 他们将细粒度的注意力掩膜编码加入到卷积中构成注意模块来引导检测器聚焦行人区域, 而放大-缩小模块则去探索丰富的上下文语义信息和本地细节, 以进一步减轻对小尺度目标的检测.

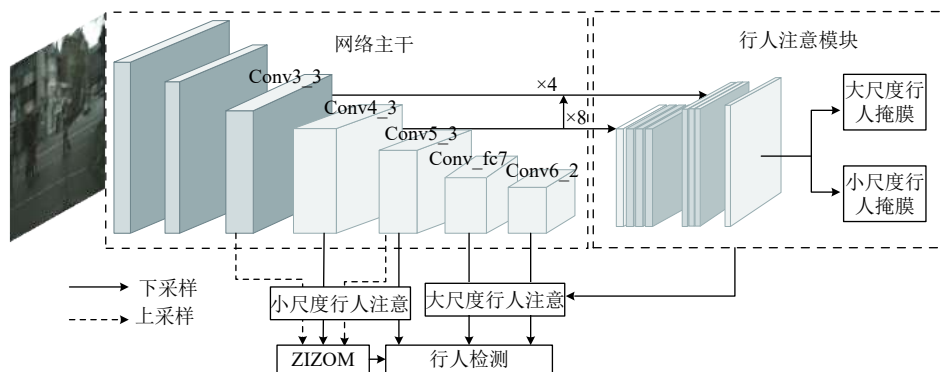


图4 GDFL 框架

基于尺度感知的方法将不同尺度上的行人视为不同的子类别, 让网络拥有相对独立的执行小尺度行人的检测, 使得在整体检测过程中避免漏检情况的发生.

### 1.5 基于超分辨率的方法

超分辨率方法旨在提高原有图像的分辨率, 精细的细节信息对于对象实例的检测和定位至关重要. 对于小尺度行人具有覆盖像素少、分辨率低的特点, 最初的方法通过简单的对图像和特征图进行上采样处理, 这样也许有效但也可能会造成伪影等其他问题, 小目标对象可能仍然会模糊不清、难以检测.

因此, 一些方法试图通过生成超分辨率图像进行检测, 与用二次插值等方法重新调整大小生成的图像相比, 超分辨率图像更清晰, 包含更详细的信息, 这些图像的特征图包含足够的信息来区分它们和背景. 由 Pang 等人提出的用于检测小规模行人的 JCS-Net, 将分类任务和超分辨率任务集成在一个统一的框架中<sup>[33]</sup>, 旨在利用大规模行人和相应的小规模行人之间的关系来帮助恢复小规模行人的详细信息, 从而提高检测小规模行人的性能. Wu 等人开发了一种模拟方法来增强小尺度行人的表示<sup>[34]</sup>. 他们构造了一个 SML (self-mimic learning) 组件来改善小尺度行人的检测性能, 通

过强制小尺度行人模拟学习大规模行人的特征来丰富和增强自身的表示.

近年来, 生成对抗网络 GAN 在合成图像方面也显示出了巨大的优势, 可以很好地应用于小物体检测. 生成对抗网络有两个子网络组成, 生成器网络和鉴别器网络, 两者通过博弈来实现一个更好的图像生成效果. 最初人们常用 GAN 网络生成更多的图像以达到数据增强的目的, 近几年, 有些研究尝试利用 GAN 进行小目标对象的超分辨率重建, 并取得了很好的成效.

Li 等人首次将生成对抗网络应用在解决小目标检测问题上, 并在行人检测上表现出了很好的性能. 他们构建了一个感知生成对抗网络 (perceptual GAN) 为小目标生成超分辨表示<sup>[35]</sup>, 由生成器和感知鉴别器两个子网络组成. 生成器基于残差网络, 从浅层特征引入细粒度细节, 将小目标的表示增强为能够提供高检测精度的超分辨表示. 感知鉴别器的一个分支区分“真假”, 另一个分支对生成的细粒度细节的质量和优势提供指导. 在文献 [36] 中发现利用 GAN 一步生成小尺度行人的高分辨图像质量并不是很好, 因此他们将原来的超分辨率网络改进为两个生成器级联, 第 1 个生成器

生成粗略的超分辨率图像,第2个生成器进一步生成精细的超分辨率图像.生成对抗网络的优势在于能够与任意检测器相结合,不需要进行额外的处理操作.

细节对于对象实例定位至关重要,此类方法尝试利用不同尺度之间的内在关系,将原始图像的低分辨率重建为更高的分辨率,旨在将低分辨率小尺度对象的表示增强,生成更高分辨更具有细节信息的类似于大尺度对象的特征信息,从而有效提高检测性能.

## 2 数据集与方法评估

在过去的工作中,人们对行人检测技术的探索不仅限于方法上的研究,许多组织机构还提供了一些用

于评价性能的数据集和基准.

### 2.1 行人数据集

众所周知,全面且丰富的数据集能够推动计算机视觉的发展.行人数据集如表3所示,主要分为早期行人数据集和现代行人数据集.早期的行人数据集相对较小,主要基于传统方法使用,如MIT<sup>[37]</sup>、INRIA<sup>[38]</sup>、Daimler<sup>[39]</sup>等.CrowdHuman<sup>[40]</sup>和EuroCity<sup>[41]</sup>是最新发布的数据集.目前,主流的行人检测数据集主要有3个:Caltech、KITTI<sup>[42]</sup>和CityPerson<sup>[43]</sup>.这3个数据集在数量上呈现了更大的规模,具有更完整的标注信息和更好的标注效果,并且包含了遮挡和多尺度场景,因此应用更加广泛.表3总结了这3个行人数据集.

表3 行人数据集概要

数据集	来源	图片数量	行人	分辨率	时间范围	具体描述
MIT	IJCV2000	—	924	64×128	白天	最早的公开数据集,包含正面、背面两个角度的正样本.
INRIA	CVPR2005	2120	1774	640×480	白天	早期流行的数据集,由GRAZ 01数据集和网络上一张图片构成.
Daimler	PAMI2009	29k	72k	640×480	白天	记录一天不同时刻的行人,由灰色图像组成.
Caltech	PAMI2010	250k	289k	640×480	白天	一个标准完整的行人数据集,数量最大.
KITTI	CVPR2012	15k	9k	1240×376	白天	自动驾驶场景下的最大数据集,包括3D标注,主要用于自动驾驶场景.
CityPerson	CVPR2017	5k	32k	2048×1024	白天	包含了27个城市、3个季节、各种天气状况和更常见的人群的图像.
CrowdHuman	PAMI2019	24k	552k	—	白天	用于拥挤人体检测的数据集,图像来自互联网,具有更高的拥挤度和更大的尺度变化.
EuroCity	PAMI2019	47k	219k	1920×1024	白天、夜间	多个欧洲城市白天和夜间拍摄的城市场景,包含3种不同的人类别.

Caltech数据集是在一辆驾驶在洛杉矶不同街道的汽车所拍摄的约10个小时的11段行驶视频中提取的,是最完整的行人检测基准之一.其中前6个视频用于训练,后5个视频用于测试.数据集相较于以往的数据集提升了两个数量级的样本,并且数据集还详细标注了包围盒和遮挡标签,图像间具有时间对应关系.KITTI是自动驾驶场景下的数据集,包含市区、乡村和高速公路等地区采集的180GB真实图像数据,每张图像中最多达30个行人,并且数据集的行人比例变化很大(从25像素到300像素的高度).二维目标检测任务包含汽车、行人和自行车3大类,评估分为easy、moderate和hard3个层次.CityPerson是在城市街道场景中建立的行人数据集.相较于前两个数据集都是在一个城市记录的,该数据集汇集了27个城市的街道场景,注释也更具有多样性.它包含2975幅训练图像和500幅验

证图像,以及1575幅测试图像.

### 2.2 评价指标

评价一个检测器的检测能力需要由相应的评价指标来体现的,一个好的评价标准对于性能检测来说至关重要.目前,行人检测器性能的评估是基于其在数据集上的表现.现在流行的行人检测的评价指标有两个:MR-FFPI和MR<sup>-2</sup>.

MR-FFPI (miss rate versus false positives per image) 曲线以单帧评估的方式更加适合行人检测的评价指标.通过改变检测置信阈值,可以在对数空间中绘制每幅图像的误检率(FFPI).给定一个检测置信度阈值,误检率(MR)可以通过真阳性数( $N_{tp}$ )和基础真值数( $N_g$ )来计算:

$$MR = 1 - N_{tp}/N_g \quad (1)$$

而FFPI可以通过将误检数量除以图像数量来计算.

Log-average miss rate ( $MR^{-2}$ ), 对数平均误检率常用来总结检测器的性能, 并作为结果对比的参考. 计算方法是将对数坐标下, 从  $[0.01, 1]$  区间内均匀的取 9 个  $FPPI$  值并计算  $FPPI$  对应的 9 个 miss rate 值的平均值  $MR^{-2}$ , 较低的  $MR^{-2}$  反映了更好的结果.

表 4 和表 5 中展示了一些经典的模型在 KITTI 和 CityPerson 数据集上的检测结果. KITTI 数据集包含了大量不同尺度的行人, 更具有挑战性. 如表 4、表 5 所示, 主流算法在 KITTI 数据集上的检测性能均低于 CityPerson 数据集. 表 4 对算法的平均检测精度进行统计, 在 easy 子数据集中, MHN-D 算法精度最高达到 85.81%, 实时性也差强人意; AMS-Net 算法检测速度最快, 其检测精度也十分理想; 在 mid 子数据集中, MHN-D 稍逊于 MS-J 都达到了 74% 以上的精度, 相较于在 easy 子集上的表现, 算法精度都降低了 10% 左右; 而在 hard 子数据集中, 不同算法的检测精度之间差距缩小, MS-J 取得了最佳性能. 表 5 对算法的误检率进行统计, 算法整体都实现了较高的精准度, SML 方法在两个子集上均达到了最低的错误率.

表 4 不同方法在 KITTI 数据集上的平均检测精度 (%)

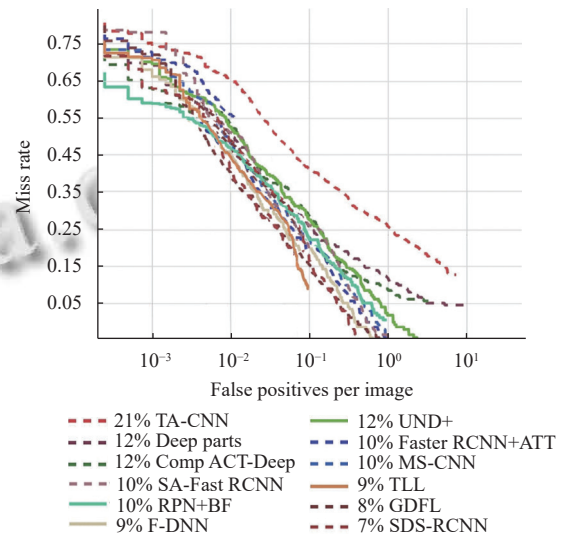
方法	主干网	Easy	Mid	Hard	速度
CompACT-Deep	VGG16	70.69	58.74	52.71	1.00
SAF RCNN	VGG16	77.93	65.01	60.42	—
RPN+BF	—	75.58	61.29	56.08	0.60
SDS-RCNN	—	—	63.05	—	0.21
MS-CNN	VGG16	83.92	73.70	68.31	0.40
AMS-Net	—	84.57	68.57	66.97	0.19
GDFL	VGG16	84.61	68.62	66.86	0.27
MHN-D	VGG16	85.81	74.60	68.94	0.39
MS-J[24]	VGG16	85.62	74.99	69.65	—

表 5 深度学习方法在 CityPerson 数据集上的检测情况 ( $MR^{-2}$ )

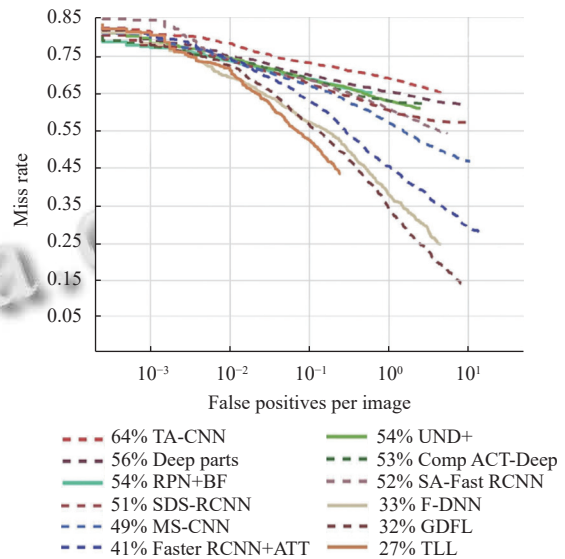
方法	年份	主干网	Reasonable	Small
Faster-RCNN	2015	VGG16	15.4	—
MS-CNN	2016	VGG16	13.32	15.86
ALFNet	2018	ResNet-50	12.0	19.0
TLL	2018	ResNet-50	15.5	—
MHN-D	2018	VGG16	12.92	17.24
SML	2020	ResNet-50	10.6	16.3

图 5 分别展示了算法在加州理工学院数据集中 reasonable 子集和 small 子集上不同方法的检测结果. 对于 reasonable 子集, 不同算法准确率之间差异较小, 但在 small 子集上所有算法的精度都有一个大幅度的下降, 且算法间差距也逐渐变大, 证实了小规模行人的

存在是当前行人检测算法的主要瓶颈之一. 而相对于 TA-CNN<sup>[44]</sup>、DeepParts<sup>[45]</sup>、UDN+<sup>[46]</sup> 算法, SAF RCNN 等在两个子集上都有着更好的表现, 其中 TLL 实现了最优性能. 同时比发现, 同种方法在不同数据集上训练所展现出的性能存在一定的差异, 数据集的选择对检测器的性能也产生一定的影响.



(a) Reasonable 子集



(b) Small 子集

图 5 不同深度学习算法在 Caltech 上两个子集的表现

### 2.3 方法评价

目前小尺度行人检测需要研究的核心问题是如何让行人的特征表达包含更多的语义信息, 这实质上对提高检测性能起着至关重要的作用. 前面所介绍和讨论的解决小尺度行人检测问题的解决思路也主要是



围绕小尺度行人的特征表示展开的,并且5类方法的侧重点也有所不同,多尺度表示和上下文信息着手于主干网络产生的特征图,而训练和分类策略以及尺度感知方法则侧重对训练和检测过程的改进.基于超分辨率的方法旨在对图像的恢复及重建,相对于其他方法来说更直接和可观.表6对深度学习中的5种方法进行对比和分析.

表6 5种方法对比

方法类型	解决的问题				方法思想概述
	特征不足	信息丢失	感受野不匹配	不易检测和定位	
多尺度表示	—	√	√	—	将深层特征和浅层特征相结合
上下文信息	√	√	—	√	提取上下文信息作为区域特征的补充
新的训练和分类策略	√	—	—	√	提高分类器对低像素特征的敏感性,提高分类器的性能
尺度感知	√	—	√	√	将不同尺度的行人分开处理
超分辨率	√	√	—	√	补偿细节,生成更高分辨率的图像

(1) 基于多尺度表示的方法是处理小尺度和多尺度问题最典型和有效的方法.但该类方法在提高性能的同时会不可避免的增加计算负担,因此如何实现准确性和复杂性之间的良好平衡需要更进一步地去探索.目前,处理小尺度行人的大多数方法都基于FPN进行适当调整和改进作为检测网络的基本架构.

(2) 上下文信息是近几年的一个研究热点,很多方法会利用上下文信息作为辅助信息,但需要注意的问题是并不是所有的上下文信息都有效,因此需要对如何有效控制环境信息的传递进一步研究.

(3) 在基于训练和分类策略的方法中,进行多任务联合学习是一个很好的方向.同时采用多个计算机视觉任务可以获得更丰富的信息.因此如何有效地利用多任务联合学习和优化来提高小尺度行人检测的性能是未来的一个研究重点.

(4) 基于尺度感知的方法对于小尺度行人检测是有效的,其关键在于如何准确划分不同尺度的行人,以获得更多的小尺度行人建议.

(5) 基于超分辨率的方法是近几年解决小尺度行人检测问题的一个新的发展方向.该类方法目前主要

大多基于GAN网络,无需设计特定的架构,最大的挑战在于GAN很难训练,因此如何在生成器和鉴别器之间实现良好的平衡是未来需要探索的方向.

### 3 结论与展望

行人检测是计算机视觉中一个重要且具有挑战性的任务,在取得巨大进步的同时也伴随着很多问题的产生.本文针对小尺度行人的检测所面临的挑战进行了剖析,并对小尺度行人检测方法分类讨论.当前多尺度表示、上下文信息、不同的训练和分类策略、尺度感知、超分辨率5种策略在解决小尺度行人检测问题上均取得了不错的成绩,并且具有很好的发展前景.

尽管行人检测技术已经取得了较大的进展,但目前仍存在很多的问题亟待解决,主要包括:

(1) 缺乏检测数据集和基准:在一定程度上,深度学习的性能是通过大量的数据提升的.像在广泛使用Caltech数据集中,“小”对象类中的许多对象实例占据了图像的很大一部分.为了更好地评估小目标检测算法的性能,需要专门用于小目标检测的大规模数据集,因此,建立大规模的小尺度行人目标数据集和相应的基准是行人检测领域的一个研究方向.

(2) 多变化融合问题:在实际应用场景中,小尺度问题和遮挡问题大多同时存在,并且天气、光线等也会给检测带来影响.目前大多数的方法只针对于单个问题的进行改善,因此,有必要进一步研究更加具有鲁棒性的检测算法处理融合的多变化问题.

(3) 行人多姿态问题:目前行人检测对象多为直立行人,当前技术对于识别一些特殊的行人状态,如坐、蹲、骑等比较困难.因此,有必要深入挖掘多模态行人的共同特征,以增强行人检测器的泛化能力.

(4) 检测实时性问题:现有的行人检测方法大多侧重于提高检测精度,而忽略了效率.而应用在驾驶/监视场景中时,设备的计算资源有限,同时还需要达到实时检测的速度要求,从而来满足实际应用的需求.因此,有必要对嵌入式设备的轻量化和实时行人检测方法进行研究.

### 参考文献

- 陈政宏,李爱娟,邱绪云,等.智能车环境视觉感知及其关键技术研究现状.河北科技大学学报,2019,40(1):15-23.
- 徐凯恺,马岩,钱旭,等.自动驾驶场景的尺度感知实时行

- 人检测. 中国图象图形学报, 2021, 26(1): 93–100.
- 3 龚安, 李承前, 牛博. 基于卷积神经网络的实时行人检测方法. 计算机系统应用, 2017, 26(9): 215–218. [doi: [10.15888/j.cnki.csa.005943](https://doi.org/10.15888/j.cnki.csa.005943)]
  - 4 Dollár P, Wojek C, Schiele B, *et al.* Pedestrian detection: A benchmark. Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009. 304–311.
  - 5 高新波, 莫梦竟成, 汪海涛, 等. 小目标检测研究进展. 数据采集与处理, 2021, 36(3): 391–417.
  - 6 陈科圻, 朱志亮, 邓小明, 等. 多尺度目标检测的深度学习研究综述. 软件学报, 2021, 32(4): 1201–1227. [doi: [10.13328/j.cnki.jos.006166](https://doi.org/10.13328/j.cnki.jos.006166)]
  - 7 陈涛, 路红. 基于深度学习的行人检测技术研究. 软件导刊, 2021, 20(1): 76–80.
  - 8 耿艺宁, 刘帅师, 刘泰廷, 等. 基于计算机视觉的行人检测技术综述. 计算机应用, 2021, 41(S1): 43–50.
  - 9 Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017, 60(6): 1097–1105.
  - 10 Adelson EH, Anderson CH, Bergen JR, *et al.* Pyramid methods in image processing. RCA Engineer, 1984, 29(6): 33–41.
  - 11 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)]
  - 12 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.
  - 13 Cai ZW, Fan QF, Feris RS, *et al.* A unified multi-scale deep convolutional neural network for fast object detection. 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 354–370.
  - 14 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 936–944.
  - 15 Li ZX, Zhou FQ. FSSD: Feature fusion single shot multibox detector. arXiv: 1712.00960, 2017.
  - 16 Zhang XW, Cao S, Chen CZ. Scale-aware hierarchical detection network for pedestrian detection. IEEE Access, 2020, 8: 94429–94439. [doi: [10.1109/ACCESS.2020.2995321](https://doi.org/10.1109/ACCESS.2020.2995321)]
  - 17 Cao JL, Pang YW, Li XL. Exploring multi-branch and high-level semantic networks for improving pedestrian detection. arXiv: 1804.00872, 2018.
  - 18 Chen LC, Papandreou G, Schroff F, *et al.* Rethinking atrous convolution for semantic image segmentation. arXiv: 1706.05587, 2017.
  - 19 Hu JH, Jin L. FPN++: A simple baseline for pedestrian detection. 2019 IEEE International Conference on Multimedia and Expo. Shanghai: IEEE, 2019. 1138–1143.
  - 20 Xie H, Chen YF, Shin H. Context-aware pedestrian detection especially for small-sized instances with deconvolution integrated faster RCNN (DIF R-CNN). Applied Intelligence, 2019, 49(3): 1200–1211. [doi: [10.1007/s10489-018-1326-8](https://doi.org/10.1007/s10489-018-1326-8)]
  - 21 Liu W, Liao SC, Hu WD. Efficient single-stage pedestrian detector by asymptotic localization fitting and multi-scale context encoding. IEEE Transactions on Image Processing, 2019, 29: 1413–1425.
  - 22 Cao JL, Pang YW, *et al.* Taking a look at small-scale pedestrians and occluded pedestrians. IEEE Transactions on Image Processing, 2020, 29: 3143–3152. [doi: [10.1109/TIP.2019.2957927](https://doi.org/10.1109/TIP.2019.2957927)]
  - 23 Cai ZW, Saberian M, Vasconcelos N. Learning complexity-aware cascades for deep pedestrian detection. Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 3361–3369.
  - 24 Zhang LL, Lin L, Liang XD, *et al.* Is faster R-CNN doing well for pedestrian detection? 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 443–457.
  - 25 Du XZ, El-Khamy M, Lee J, *et al.* Fused DNN: A deep neural network fusion approach to fast and robust pedestrian detection. 2017 IEEE Winter Conference on Applications of Computer Vision. Santa Rosa: IEEE, 2017. 953–961.
  - 26 Zhang S, Yang XS, Liu YX, *et al.* Asymmetric multi-stage CNNs for small-scale pedestrian detection. Neurocomputing, 2020, 409: 12–26. [doi: [10.1016/j.neucom.2020.05.019](https://doi.org/10.1016/j.neucom.2020.05.019)]
  - 27 Song T, Sun LY, Xie D, *et al.* Small-scale pedestrian detection based on somatic topology localization and temporal feature aggregation. arXiv: 1807.01438, 2018.
  - 28 Brazil G, Yin X, Liu XM. Illuminating pedestrians via simultaneous detection and segmentation. IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 4960–4969.
  - 29 Zhou CJ, Wu MQ, Lam SK. SSA-CNN: Semantic self-attention CNN for pedestrian detection. arXiv: 1902.09080, 2019.
  - 30 Li JN, Liang XD, Shen SM, *et al.* Scale-aware fast R-CNN for pedestrian detection. IEEE transactions on Multimedia, 2018, 20(4): 985–996.

- 31 Han B, Wang YH, Yang Z, *et al.* Small-scale pedestrian detection based on deep neural network. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 21(7): 3046–3055. [doi: [10.1109/TITS.2019.2923752](https://doi.org/10.1109/TITS.2019.2923752)]
- 32 Lin CZ, Lu JW, Wang G, *et al.* Graininess-aware deep feature learning for pedestrian detection. *Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018. 745–761.
- 33 Pang YW, Cao JL, Wang J, *et al.* JCS-Net: Joint classification and super-resolution network for small-scale pedestrian detection in surveillance images. *IEEE Transactions on Information Forensics and Security*, 2019, 14(12): 3322–3331. [doi: [10.1109/TIFS.2019.2916592](https://doi.org/10.1109/TIFS.2019.2916592)]
- 34 Wu JL, Zhou CL, Zhang Q, *et al.* Self-mimic learning for small-scale pedestrian detection. *Proceedings of the 28th ACM International Conference on Multimedia*. Seattle: ACM, 2020. 2012–2020.
- 35 Li JN, Liang XD, Wei YC, *et al.* Perceptual generative adversarial networks for small object detection. *2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 1951–1959.
- 36 Yin RH. Multi-resolution generative adversarial networks for tiny-scale pedestrian detection. *2019 IEEE International Conference on Image Processing (ICIP)*. Taipei: IEEE, 2019. 1665–1669.
- 37 Papageorgiou C, Poggio T. A trainable system for object detection. *International Journal of Computer Vision*, 2000, 38(1): 15–33. [doi: [10.1023/A:1008162616689](https://doi.org/10.1023/A:1008162616689)]
- 38 Dalal N, Triggs B. Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego: IEEE, 2005. 886–893.
- 39 Enzweiler M, Gavrila DM. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(12): 2179–2195. [doi: [10.1109/TPAMI.2008.260](https://doi.org/10.1109/TPAMI.2008.260)]
- 40 Shao S, Zhao ZJ, Li BX, *et al.* Crowdhuman: A benchmark for detecting human in a crowd. arXiv: 1805.00123, 2018.
- 41 Braun M, Krebs S, Flohr F, *et al.* Eurocity persons: A novel benchmark for person detection in traffic scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(8): 1844–1861. [doi: [10.1109/TPAMI.2019.2897684](https://doi.org/10.1109/TPAMI.2019.2897684)]
- 42 Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite. *2012 IEEE Conference on Computer Vision and Pattern Recognition*. Providence: IEEE, 2012. 3354–3361.
- 43 Zhang SS, Benenson R, Schiele B. CityPersons: A diverse dataset for pedestrian detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 4457–4465.
- 44 Tian YL, Luo P, Wang XG, *et al.* Pedestrian detection aided by deep learning semantic tasks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015. 5079–5087.
- 45 Tian YL, Luo P, Wang XG, *et al.* Deep learning strong parts for pedestrian detection. *Proceedings of the IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015. 1904–1912.
- 46 Ouyang WL, Zhou H, Li HS, *et al.* Jointly learning deep features, deformable parts, occlusion and classification for pedestrian detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(8): 1874–1887. [doi: [10.1109/TPAMI.2017.2738645](https://doi.org/10.1109/TPAMI.2017.2738645)]

(校对责编: 牛欣悦)