

一种基于 SSD 与 FRN 相结合的密集连接行人检测算法

冯婷婷, 葛华勇, 孙家慧

(东华大学 信息科学与技术学院, 上海 201620)

摘要: 现实场景行人的复杂性和多样性使得行人检测成为计算机视觉领域中一个既具有研究价值又极具挑战性的热门课题, 为提高其准确性, 提出一种基于 SSD(Single Shot Multibox Detector)与 FRN(Filter Response Normalization)相结合的密集连接行人检测算法, 将串联式的 SSD 基础网络修改为引入上下文语义信息的多层融合的密集连接的 FRN 网络结构, 运用聚类思想设置适宜行人尺度的候选框, 并且根据行人尺寸的统计分布规律调整不同检测层的缩放因子, 从而实现端到端训练。在融合数据集和 VOC2007TEST 数据集上验证该模型的性能, 相比于 SSD 方法, 该方法准确率 AP(Average Precision)分别提高 5.8%、2.9%, 具有更高的准确性和鲁棒性。

关键词: 行人检测; 密集连接; 聚类算法; SSD; FRN

中图分类号: TP301.6

文献标识码: A

DOI: 10.19358/j.issn.2096-5133.2020.12.010

引用格式: 冯婷婷, 葛华勇, 孙家慧. 一种基于 SSD 与 FRN 相结合的密集连接行人检测算法[J]. 信息技术与网络安全, 2020, 39(12): 56-60, 66.

A densely connected pedestrian detection algorithm based on the combination of SSD and FRN

Feng Tingting, Ge Huayong, Sun Jiahui

(School of Information Science and Technology, Donghua University, Shanghai 201620, China)

Abstract: The complexity and diversity of pedestrians in real scenes make pedestrian detection a hot topic with both research value and challenge in the field of computer vision. In order to improve its accuracy, this paper proposes a densely connected pedestrian detection algorithm based on the combination of SSD and FRN, modifies the serial SSD basic network into a multi-layered densely-connected FRN network structure that introduces contextual semantic information, and uses clustering ideas to set candidate boxes suitable for pedestrian scale, and adjusts the scaling factors of different detection layers according to the statistical distribution law of pedestrian size, so as to achieve end-to-end training. The performance of the model is verified on the fusion dataset and the VOC2007TEST dataset. Compared with the SSD method, the accuracy of the method AP is improved by 5.8% and 2.9% respectively, with higher accuracy and robustness.

Key words: pedestrian detection; dense connection; clustering algorithm; SSD; FRN

0 引言

行人检测作为计算机视觉技术的重要分支和智能化产品的核心技术, 受到了学术界和工业界的广泛关注, 其能够从图像或视频中识别出行人, 并给出其具体的位置, 在车辆辅助驾驶和行人重识别技术等方面有巨大的研究价值和应用前景。行人检测作为车辆辅助驾驶技术中不可或缺的一部分, 可以及时检测出车辆前方的行人并针对实际状况及时提醒司机或者紧急制动, 从而避免交通事故的发

生; 在刑侦工作中, 刑侦人员经常要浏览多个摄像头中的视频, 此时先进行行人检测判断视频中是否存在行人, 把视频中的行人筛选出来, 再利用行人重识别技术查找某个特定的行人在哪些摄像头曾经出现过, 可为刑侦工作带来便利。

近十几年间, 基于深度学习的行人检测技术取得了巨大进步, 能够自动学习从图像像素中提取的基于边缘的低级特征和基于语义信息的高级特征。其分为两阶段检测算法和单阶段检测算法。在两阶

段检测算法中,文献[1]提出了基于区域的卷积神经网络(Region based Convolutional Neural Network, R-CNN),文献[2]提出了空间金字塔池化(Spatial Pyramid Pooling, SPP)网络,文献[3]提出了快速基于区域的卷积网络方法(Fast-RCNN),文献[4-5]提出了Faster-RCNN。这些目标检测算法的训练过程步骤繁琐,检测速度慢,没有达到实时的检测标准。基于此,以 REDMON J^[6]提出的统一实时目标检测框架(You only look once, Yolo)和以 Liu Wei^[7]提出的单阶段多尺度检测器(Single Shot MultiBox Detector, SSD)框架为代表的单阶段检测算法由此产生。Yolo 存在定位精度、召回率等较低的问题,泛化能力相对较弱,为了解决该算法的缺陷,2016年 Liu Wei 等提出 SSD 算法进行多尺度检测,在保证速度的同时提高了检测精度,但是其对于小目标检测不精准,加之在实际生活中,由于行人穿着、姿态、尺度、视角、光照和复杂背景等多方面原因,在检测精度及速度方面的提高仍是研究重点。由此针对行人多尺度问题,本文提出一种 FRN 提升模型性能的密集连接的 SSD 行人检测算法,尝试引入不依赖批尺寸大小的上下文语义信息的多层特征融合的密集连接网络结构,结合行人检测任务特点进行优化与改进。

1 SSD 算法框架及原理

SSD 采用 VGG16 空洞卷积作为基础模型,在其后添加额外的卷积层,通过在不同尺度的特征图下对可能存在的目标进行预测,包括边框回归和分类,最后通过非极大值抑制算法(Non-Maximum Suppression, NMS)从候选框中筛选出置信度最高、存在目标的区域,从而降低预测框的冗余度^[7]。SSD 网络模型结构如图 1 所示。

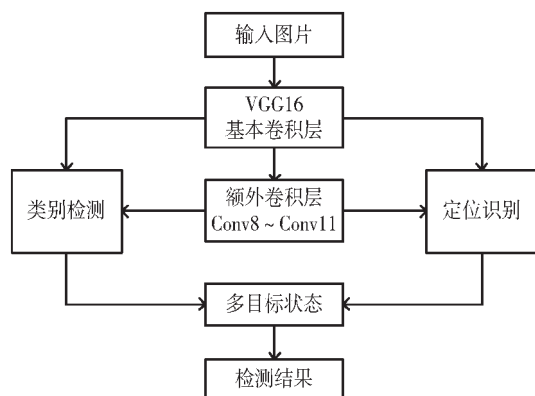


图 1 SSD 算法框架

不同特征图具有不同感受野,SSD 算法的核心之一就是利用特征金字塔结构在不同特征层上设置不同的候选框以适应不同大小的目标,从而进行分类和边框回归。在训练时,为了使默认框与真实框相匹配,SSD 采用困难样本挖掘技术,将正、负样本比例设置为 1:3 以得到更快的权值优化和更稳定的训练;在测试阶段,直接预测每个默认框的位置偏移量和属于每种类别的概率,再通过 NMS 筛选出最佳默认框,即最终的检测结果。使用 NMS 算法前后检测框对比如图 2 所示,其步骤可总结为:首先对输入集合 E 中的窗口按照置信度由高到低排列,选择置信度最高的窗口,计算它与剩余所有窗口的交并比(Intersection Over Union, IOU);其次删除 IOU 大于阈值的窗口;最后将 E 中剩余检测框再次排序,重复上述步骤,直至 E 中的窗口都被处理后,得到输出窗口集合 F 。

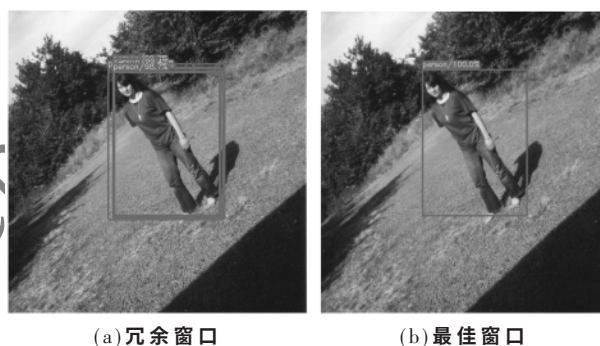


图 2 使用 NMS 算法前后检测框对比

2 基于 SSD 的 FRN 密集连接行人检测算法

由于卷积神经网络的池化操作,不同特征层具有不同特点,低层特征分辨率高,包含更多位置、细节信息,其语义性低、噪声多;高层特征具有更强的语义信息,但分辨率低,对细节的感知能力较差。因此本文针对多尺度检测问题,将串联式的基础网络修改为引入上下文语义信息的多层特征融合的 FRN 密集连接结构,需要对 SSD 模型进行双线性插值之后添加 FRN 归一化层再进行特征融合,从而实现不同特征图的多尺度检测。

因为在 SSD 中对 Conv4_3、Conv_7、Conv8_2、Conv9_2、Conv10_2、Conv11_2 这 6 个不同大小的特征层进行多尺度检测,它们每层卷积运算输入仅为与之相连的上一卷积层,并未考虑之前的卷积层,也并未考虑该卷积层之后的其他卷积层,丧失了大量的特征信息,因此进行特征融合需要首先将卷积

特征图规范化,统一为相同尺寸大小,从而进行不同特征层的相互融合,使其得到的特征图包含更多的语义信息和空间位置信息。文献[8]将 SSD 中 Conv4_3 与 Conv5_3 反卷积得到的特征图进行融合,文献[9]将 Faster RCNN 中 Conv3_3、Conv4_3 与 Conv5_3 三层特征图相融合,本文采取不同的采样策略融合不同层的特征,即将不相邻的卷积层进行融合,从而减少冗余。SSD 改进模型结构如图 3 所示,其步骤可总结为:首先将 Conv4_3 特征层进行双线性插值再进行 FRN 归一化,得到的特征图分别与 Conv8_2、Conv9_2、Conv10_2、Conv11_2 融合;其次将 Conv11_2 特征层采用相同的方法分别与 Conv9_2、Conv8_2、Conv7、Conv4_3 融合;最后将经过拼接得到的融合特征图进行检测,其低层特征层可增加更多的语义特征,高层特征层可包含更多空间信息,鲁棒性更强,更具有判别能力。

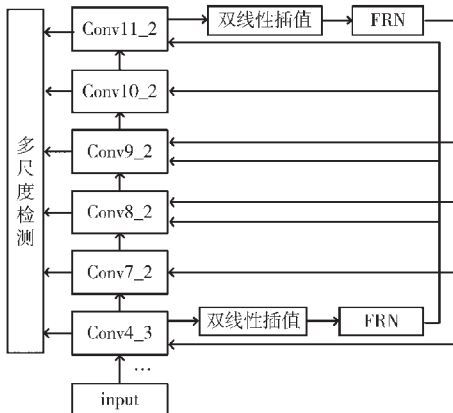


图 3 SSD 改进模型结构图

FRN 归一化层包括归一化函数 FRN 和激活函数 TLU (Threshold Linear Normalize),其操作是在 (H, W) 维度上的,FRN 层结构示意图如图 4 所示,其中 x 是一个 $N(H \times W)$ 维度向量, v^2 是 x 的二次范数的平均值, ε 是一个极小的正常量, γ 、 β 、 τ 是可学习参量。其步骤可总结为:首先对每个样本的每个通道单独进行归一化,没有对批尺寸 batch 的依赖,并且在 batch 比较大的时候,其性能还优于批处理归一化 (Batch Normalization, BN);其次在 FRN 之后连接阈值化激活函数 TLU,这样做的原因是 FRN 缺少去均值操作,在其后若连接激活函数 ReLU,可能会使得归一化后的结果任意地偏移 0,对模型训练结果和性能不利。

SSD 中每层特征图宽高比 r 如表 1 所示,但是

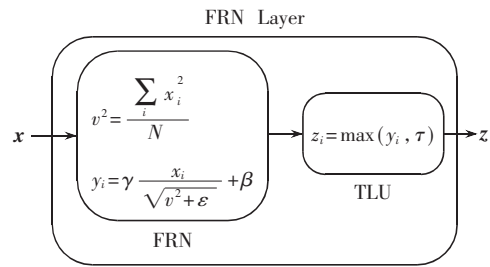


图 4 FRN 层结构示意图

表 1 SSD 各层特征图宽高比 r

特征层	r
Conv4_3	(1.0, 2.0, 0.5, 1')
Conv7	(1.0, 2.0, 3.0, 0.5, 0.333 3, 1')
Conv8_2	(1.0, 2.0, 3.0, 0.5, 0.333 3, 1')
Conv9_2	(1.0, 2.0, 3.0, 0.5, 0.333 3, 1')
Conv10_2	(1.0, 2.0, 0.5, 1')
Conv11_2	(1.0, 2.0, 0.5, 1')

对于行人检测绝大多数情况下,行人图像信息基本都是高度大于宽度,但由于部分数据标签的制作错误,则将融合数据集经过聚类后得到的候选框分布范围示意图如图 5 所示,其横坐标为行人高度,单位为像素,纵坐标为行人宽高比。根据聚类思想本文每层特征图宽高比 r 如表 2 所示。另外在 SSD 中对

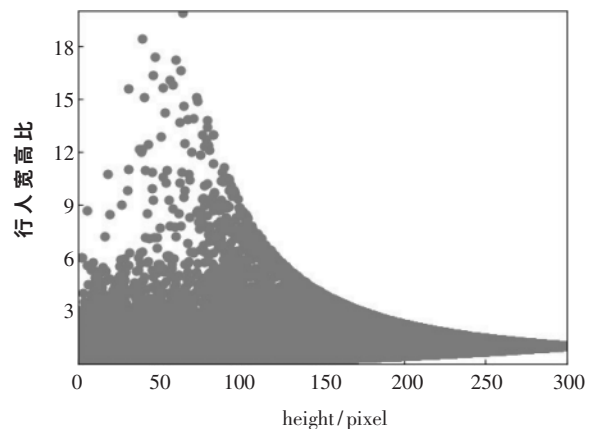


图 5 候选框分布范围示意图

表 2 改进 SSD 各层特征图宽高比 r

特征层	r
Conv4_3	(0.2, 0.41, 0.6, 1.0')
Conv7	(0.2, 0.41, 0.6, 1.0')
Conv8_2	(0.3, 0.41, 0.7, 1.1)
Conv9_2	(0.3, 0.41, 0.7, 1.1)
Conv10_2	(0.3, 0.41, 0.7, 1.1)
Conv11_2	(0.7, 1.4)

于 r 取 1' 的时候, 本文取不同的缩放因子 $r' = \{[0.8], [0.8], [0.8], [0.9], [1.0], [1.0]\}$, 对于行人检测, 一共会产生 48 490 个结果。

3 实验与结果分析

为了验证改进 SSD 算法对行人检测的有效性, 通过融合数据集与 VOC2007TEST 数据集进行评估测试, 并将测试结果与其他方法的结果进行对比与分析。

3.1 数据集

分别在融合数据集(包括 VOC2007、VOC2012、COCO、SYSU、PRW)与 VOC2007TEST 数据集上评估了本文方法, 其中融合数据集一共 3 776 张行人图片, SYSU 数据集包括公园、地铁等地方的行人, PRW 数据集是中学、大学校内的摄像头连拍的一些行人图片, VOC2007、VOC2012、COCO 中仅提取行人的部分照片; VOC2007TEST 中仅是行人的照片, 共 2 097 张。

3.2 测试结果及分析

模型训练结束后通过测试集评估模型性能。融合数据集与仅含行人的 VOC2007TEST 数据集的查准率-查全率 (Precision Recall, PR) 曲线分别如图 6 和图 7 所示, 其横坐标为查全率, 纵坐标为查准率, 曲线下方的面积值将作为行人检测的最终指标 AP, AP 值越高表示算法的行人检测性能越好, 即 PR 曲线越靠近右上角其模型性能越好。对于呈交叉状态的曲线图, 往往不能直观地反映算法性能, 此时可通过平衡点即 $P=R$ (P 表示查准率, R 表示查全率) 时取值的大小来说明学习器的性能, 其值越大, 性能越好。融合数据集与 VOC2007TEST 数据集测试结果对比如图 8 所示。

在图 6 和图 7 中, 由于各图两曲线呈明显的交

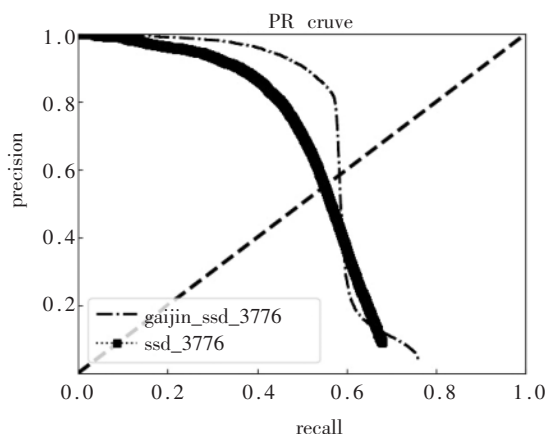


图 6 融合数据集测试结果

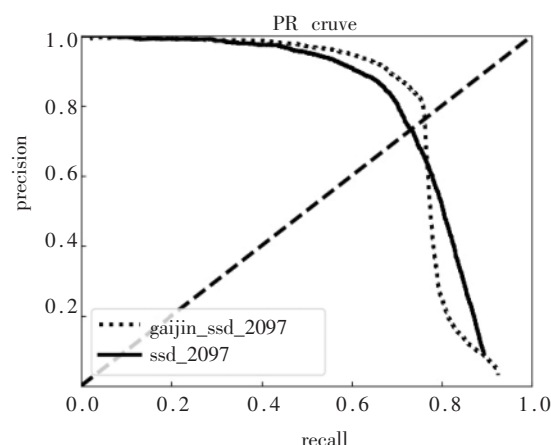


图 7 VOC2007TEST 数据集测试结果

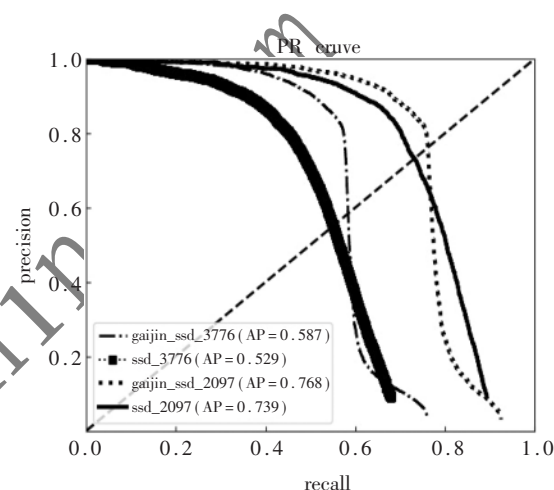


图 8 融合数据集与 VOC2007TEST 数据集测试结果对比

叉状态, 同时改进 SSD 算法平衡点与 SSD 算法的平衡点相比较而言较大, 所以改进 SSD 算法性能较好。根据图 8 知, 融合数据集在改进 SSD 算法的 AP 值高于 SSD 算法 5.8 个百分点, VOC2007TEST 数据集在改进 SSD 算法的 AP 值高于 SSD 算法 2.9 个百分点。

本文方法还与文献[8]、[9]所示方法进行对比, 结果如表 3 所示。

表 3 不同检测方法 AP 对比

方法	KITTI	VOC2007TEST	融合数据集
文献[8]原方法	50.14		
文献[8]改进法	66.87		
文献[9]原方法		69.9	
文献[9]改进法		72.8	
本文 SSD		73.9	52.9
本文 SSD 改进		76.8	58.7

文献[8]中 SSD 在 KITTI 数据集上表现一般,而本文中融合数据集在 SSD 上的表现也同样如此;文献[9]中 Faster RCNN 在 VOC2007TEST 数据集上 AP 为 69.9%,其改进方法在相同数据集上 AP 达到 72.8%,而由本文得到的 VOC2007TEST 数据集 PR 曲线轮廓能够完全包住融合数据集 PR 曲线,体现了 SSD 在复杂背景与小目标检测场景下的不足。本文在 SSD 检测框架基础上,其网络结构由原来的串联连接变换为引入上下文语义信息的提升模型性能的 FRN 密集连接特征融合结构,使得各检测层既包含了低层特征层所具有的更多轮廓、梯度的细节信息,同时又包含高层特征层所具有的更强的语义信息,并且消除了模型对批尺寸的依赖,使其在融合数据集上的 AP 由 52.9% 增加至 58.7%,提高了 5.8%,在仅含行人的 VOC2007TEST 数据集上 AP 由 73.9% 增加至 76.8%,提高了 2.9%。

不同于 VOC2007TEST 数据集,融合数据集包含遮挡较严重的多尺度行人小目标,当行人目标较小、较模糊时,SSD 算法会出现漏检、错检问题,而本文的方法在这方面改进很多。SSD 算法改进前后测试对比图如图 9 所示。

4 结论

本文首先阐述了 SSD 算法理论基础及行人检测算法,然后详细介绍了改进思路及成果,吸取了双线性插值和特征融合的研究思路,将 SSD 模型中串联式的基础网络修改为引入上下文语义信息的多层特征融合的 FRN 密集连接网络结构,即将经过双线性插值和 FRN 归一化的特征层 Conv4_3 分别与 Conv8_2、Conv9_2、Conv10_2、Conv11_2 融合,同时

Conv11_2 特征层也采用相同的方法分别与 Conv9_2、Conv8_2、Conv7、Conv4_3 融合,这一改进保证了提取特征的鲁棒性及其精度;在 FRN 归一化层中添加 TLU 激活函数而不是 ReLU,不仅避免产生很多零值,而且消除了模型训练对批尺寸的依赖,这对于模型性能提升是至关重要的;在行人候选框这一模块中,根据聚类思想设置了不同特征层不同的宽高比从而进行行人检测。通过实验表明,本文提出的检测模型的精度相对于 SSD 模型有很大的提升,其准确率在融合数据集与仅含行人的 VOC2007TEST 数据集上分别提升 5.8%、2.9%。在未来的研究工作中,将继续改进网络模型,增强特征信息的共享,提高计算效率,使得速度与精度达到优异的平衡。

参考文献

- [1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Computer Vision and Pattern Recognition, 2014: 580-587.
- [2] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [3] GIRSHICK R. Fast R-CNN[C]. International Conference on Computer Vision, 2015: 1440-1448.
- [4] 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.



(a) SSD 算法测试结果



(b) SSD 改进算法测试结果

图 9 SSD 算法改进前后测试对比图

(下转第 66 页)

效的支持。

经多次调试验证后,该全景拼接软件系统稳定性和实时性满足设计要求,并顺利通过验收。但是对于视频拼接的质量,如对图像色差进一步调整以实现更大对比度下的流畅过渡、整个系统的可靠性、处理速度的提升以实现 30 帧以上的视频处理等问题,还有待进一步研究。

参考文献

- [1] 王娟,师军,吴宪祥.图像拼接技术综述[J].计算机应用研究,2008(7):1940-1943,1947.
- [2] PERAZZI F,SORKINE-HOMUNG A,ZIMMER H,et al. Panoramic video from unstructured camera arrays[J]. Computer Graphics Forum,2015,34(2):57-68.
- [3] BROWN M,LOWE D G. Automatic panoramic image stitching using invariant features[J]. International Journal of Computer Vision,2007(74):59-73.
- [4] LI J,WANG Z,LAI S,et al. Parallax-tolerant image

stitching based on robust elastic warping[J]. IEEE Transactions on Multimedia,2018,20(7):1672-1687.

- [5] 张浩.基于 DSP 的高清全景图像拼接的研究与实现[D].成都:电子科技大学,2018.
- [6] BRADSKI G,KAHLER A. Learning OpenCV[M].[s.l.]: O'Reilly Media, Inc.,2008.
- [7] 罗莹,宋利,解蓉,等.全景媒体的系统架构研究综述[J].电信科学,2018,34(2):88-98.
- [8] 黄凯奇,陈晓棠,康运锋,等.智能视频监控技术综述[J].计算机学报,2015,38(6):1093-1118.

(收稿日期:2020-07-14)

作者简介:

方谔言(1997-),男,硕士,主要研究方向:智能信息处理。

陈贤富(1963-),男,副教授,主要研究方向:复杂系统及认知计算。

(上接第 60 页)

- [5] SUN X D,WU P C,HOI S C H,et al. Face detection using deep learning: an improved Faster-RCNN approach[J]. Neurocomputing,2018,299:42-50.
- [6] REDMON J,DIVVALA S K,GIRSHICK R,et al. You only look once: unified, real-time object detection[C]. Computer Vision and Pattern Recognition,2016:779-788.
- [7] LIU W,ANGUELOV D,ERHAN D,et al. SSD: single shot multibox detector[C]. European Conference on Computer Vision,2016:21-37.
- [8] 朱敏.基于 SSD 的行人检测方法[J].科技创新,

2017(36):75-76.

- [9] 吕俊奇,邱卫根,张立臣,等.多层卷积特征融合的行人检测[J].计算机工程与设计,2018,39(11):3481-3485.

(收稿日期:2020-07-09)

作者简介:

冯婷婷(1993-),女,硕士研究生,主要研究方向:深度学习、目标检测等。

葛华勇(1976-),女,硕士生导师,副教授,主要研究方向:数字图像处理、信号与信息处理、信息安全等。

孙家慧(1997-),女,硕士研究生,主要研究方向:深度学习、目标检测等。

版权声明

经作者授权，本论文版权和信息网络传播权归属于《信息技术与网络安全》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《信息技术与网络安全》编辑部
中国电子信息产业集团有限公司第六研究所