

基于多尺度加权特征融合网络的地铁行人目标检测算法

董小伟^① 韩悦^{*①} 张正^① 曲洪斌^② 高国飞^③ 陈明钊^③ 李博^①

^①(北方工业大学信息学院 北京 100144)

^②(中国石油管道局工程有限公司国际事业部 北京 065000)

^③(北京城建设计发展集团股份有限公司城市轨道交通绿色与安全建造技术国家工程实验室 北京 100037)

摘 要: 随着地铁乘客的大量增加, 实时准确地监测地铁站内客流量对于保证乘客安全具有重要意义。针对地铁场景复杂、行人目标小等特点, 该文提出了多尺度加权特征融合(MWF)网络, 实现地铁客流量的精准实时监测。在数据预处理阶段, 该文提出过采样目标增强算法, 对小目标占比不足的图片进行拼接处理, 增加小目标在训练时的迭代频率。其次, 在单镜头多核检测器(SSD)网络基础上添加了基于VGG16网络的特征提取层, 将不同尺度的特征层以不同方式进行加权融合, 并选出最优的特征融合方式。最终, 结合小目标过采样增强算法, 得到多尺度加权特征融合模型。实验证明, 该方法与SSD网络相比, 在保证实时性的同时, 检测精度提升了5.82%。

关键词: 目标检测; 小目标; 深度网络; 加权特征融合

中图分类号: TN911.73

文献标识码: A

文章编号: 1009-5896(2021)07-2113-08

DOI: 10.11999/JEIT200450

Metro Pedestrian Detection Algorithm Based on Multi-scale Weighted Feature Fusion Network

DONG Xiaowei^① HAN Yue^① ZHANG Zheng^① QU Hongbin^②

GAO Guofei^③ CHEN Mingdian^③ LI Bo^①

^①(School of Information Science and Technology, North China University of Technology, Beijing 100144, China)

^②(International Business Department, China Petroleum Pipeline Engineering Co., Ltd., Beijing 065000, China)

^③(Beijing Urban Construction Design and Development Group Co., Ltd., National Engineering Laboratory for Green and Safe Construction Technology of Urban Rail Transit, Beijing 100037, China)

Abstract: With the large increase of passengers in metro stations, precise and real-time monitoring of passenger flow in subway stations is of great significance for ensuring passenger safety. Based on the features of complicated subway scenes and small pedestrian targets, a Multi-scale Weighted Feature (MWF) fusion network to achieve accurate real-time monitoring of subway passengers is proposed. In the data preprocessing stage, an oversampling target enhancement algorithm is proposed to stitch the pictures with an insufficient proportion of small targets to increase the iteration frequency of small targets during training. Secondly, feature extraction layers based on the VGG16 network are added to the Single Shot multibox Detector (SSD) network. The feature layers of different scales are weighted and fused in different ways, and the optimal feature fusion method is selected. Finally, combined with the small target oversampling enhancement algorithm, a multi-scale weighted feature fusion model is obtained. Experiments show that the detection accuracy of this method has improved by 5.82 percent compared with the SSD network and doesn't reduce the speed of detection.

Key words: Target detection; Small target; Deep network; Weighted feature fusion

收稿日期: 2020-06-02; 改回日期: 2020-10-18; 网络出版: 2020-10-21

*通信作者: 韩悦 hanyue_428@163.com

基金项目: 北京市自然科学基金(4192002), 北方工业大学科研启动基金

Foundation Items: Beijing Natural Science Foundation (4192002), The Scientific Research Foundation of North University of Technology

1 引言

近年来,公共安全已成为热点问题被社会各界高度关注,城市的各个角落均安装了较为密集的监控装置,确保市民的出行安全。地铁交通受外界环境干扰小,行车速度稳定,是市民主要的出行方式之一,但地铁多建设于地下,环境复杂具有特殊性。当发生突发事件时,人员疏散和救援困难,极易发生踩踏事故。因此需要更精准的行人检测技术,对地铁中的复杂环境进行实时监控,避免群体突发性事件,确保市民安全。随着深度学习的发展,行人检测任务已经取得了很大的进步,但地铁中由于摄像头架设角度高、人流量大且不断运动,因此拍摄行人目标小、遮挡等问题极大地影响了检测结果。

早期目标检测方法大多基于手工特征,在底层特征中构建复杂模型提升精度,但人工提取的特征对于复杂场景而言鲁棒性不足。随着卷积神经网络(Convolutional Neural Network, CNN)、区域建议神经网络(Region-CNN, R-CNN)^[1]等深度学习模型的提出,以卷积网络为代表的目标检测方法迎来了前所未有的飞速发展,卷积神经网络从原始像素中学习高级语义特征,对背景噪声中具有复杂姿势的行人目标具有更强的识别力。目前主流的目标检测算法主要分为基于区域的两阶段算法,如Ren等人^[2]提出结合目标候选框的检测模型更快速区域卷积网络(Faster Region-CNN, Faster-RCNN),设计了候选区域提取网络(Region Proposal Network, RPN),提高了候选框提取质量。另一种是基于回归的单阶段目标检测算法如只需看一次网络(You Only Look Once, YOLO)^[3]、单镜头多核检测器(Single Shot multibox Detector, SSD)^[4],这类算法只经过一次前馈网络,因此检测速度更快,实时性更强。2016年Liu等人^[4]提出了SSD方法,结合多尺度的概念在6个不同特征尺度上进行预测且设定了多种尺度的检测框,设计了不同尺寸不同长宽比的默认框,其采用全卷积网络进行连接,通过多个

卷积层的信息来检测不同尺寸的目标。SSD算法虽然检测精度高、速度快,但是它仅将含有较少语义信息的底层特征图用于小目标检测,比如在以VGG16作为主干网络的检测模型中所选用的最低卷积层是Conv4,这样一些具有更低级别的信息层就被忽略掉了,包含的特征信息少,因此对于小中目标的检测效果不佳。针对这一问题,特征金字塔网络(Feature Pyramid Networks, FPN)^[5]设计了一种自顶向下的连接结构,同时利用低层特征的高分辨率以及高层特征的语义信息,通过对特征层的融合得到最终的输出结果。

本文结合FPN^[5]的特征融合思想构建了适用于地铁车站的行人头肩检测网络,在原来的SSD网络的基础上增加多尺度加权特征融合模块,充分利用低层网络信息提高对于小目标的检测性能。同时,针对地铁行人数据中小目标样本数量少、分布不均衡的特点,本文提出了一种针对小目标数据的过采样增强算法,有效扩充了小目标样本数量,增加了小目标的训练权重。结合以上两种策略,并通过大量试验验证,本文提出多尺度加权特征融合网络(Multi-scale Weighted Feature fusion Single Shot multibox Detector, MWFSSD),在保证速度的前提下,有效提高了小目标的检测精度,提升了整体模型的检测性能。

2 多尺度特征加权融合模型

2.1 SSD网络

SSD模型是采用VGG为特征提取网络的端到端卷积网络,采用多尺度方法进行目标检测,在保证准确率的同时,可以更快更高效地完成实时检测任务,图1是SSD网络流程图。原始SSD模型在进行地铁行人头肩目标检测时出现以下问题:大量小尺寸目标难以识别;一些背景区域如人像海报等被误识成目标对象;目标的重复识别问题。针对以上问题,本文改进了SSD模型,将相互独立的特征层进行特征融合,充分利用上下文信息^[6]。

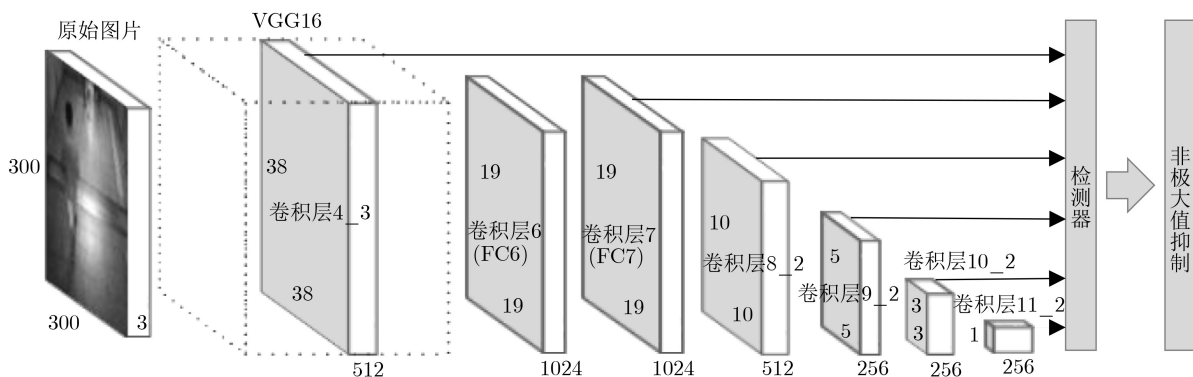


图1 SSD网络检测流程

2.2 多尺度加权特征融合网络MWFSSD

SSD目标检测算法经过多个下采样层来提取特征，而每经过一次下采样，网络的分辨率就会被压缩，从而损失不同分辨率的特征图信息，这样使得网络在最后很难提取到针对小目标的特征，另外，每个特征图对小目标的贡献程度不同，网络浅层提取目标边缘和纹理等细节特征，网络深层提取物体轮廓的特征。根据FPN，浅层特征图对小目标检测更有利，因此要想提高小目标检测精度就要增加底层特征信息，针对这一情况，本文改进了特征提取网络，选取VGG16的4个不同尺度的特征层分配权重进行加权融合处理，使每个特征图都能融合不同特征尺度的信息^[7,8]。特征融合的结果由式(1)得到

$$F_{\text{concat}} = \Theta_f \left(\sum_{i=1}^4 x_i \sigma_i w_i \right) \quad (1)$$

其中， x_i 是用于融合的原特征层， w_i 为每一层特征图的权重， σ_i 为激活函数， i 代表4个不同的特征层。

本文在VGG16网络中选择conv3_3, conv4_3, fc_7和conv7_2作为特征提取层；在融合之前每一个原始特征层都需要经过一个类似于激活函数的变换函数 σ_i ，对特征图的尺寸进行统一，使用 1×1 的卷积核来降低输入通道，FC7和Conv7_2使用线性插值的方法，将所有特征图转换成相同尺寸($38 \times 38 \times 256$)^[9]。最终为每一层特征图分配权重

w_i 后经过 Θ_f 融合函数进行Concat操作^[10]，为保证特征数据尺度一致性，对不同尺度特征进行归一化处理，添加了批标准化层 (Batch Normalization, BN)，其作用是在网络训练时，使得每一层的输入保持相同分布，防止过拟合，简化调参过程，以提高收敛速度。最后，将处理后的特征图送入SSD网络中，更好地提取出小目标的位置及类别信息^[11-16]。多尺度加权特征融合网络原理图如图2所示。

3 实验结果分析

3.1 实验数据及其预处理

本实验的硬件配置采用Intel E5-2603 v4处理器，NVIDIA Titan XP显卡，软件环境是Ubuntu16.04, Python3.7和TensorFlow深度学习框架。

本文实验所用的地铁行人数据集采集自北京地铁主要换乘站，共43954个头肩样本以及3933个难分样本，包含了地铁中的很多场景如车站、出入口、上下扶梯等，并且涵盖了地铁运营各个时间段的客流情况。图3为训练集及测试集中部分样本图像。

地铁行人样本库中，行人目标在镜头的近处和远处尺寸差别很大，这就给准确检测带来了困难。COCO^[17]数据集是2014年微软提出的专门用于计算机视觉领域较权威的数据集，其对于面积小于 32×32 的目标，就认为是小目标；对于中目标的定义是 $32 \times 32 < \text{面积} < 96 \times 96$ 。图4展示了COCO数据

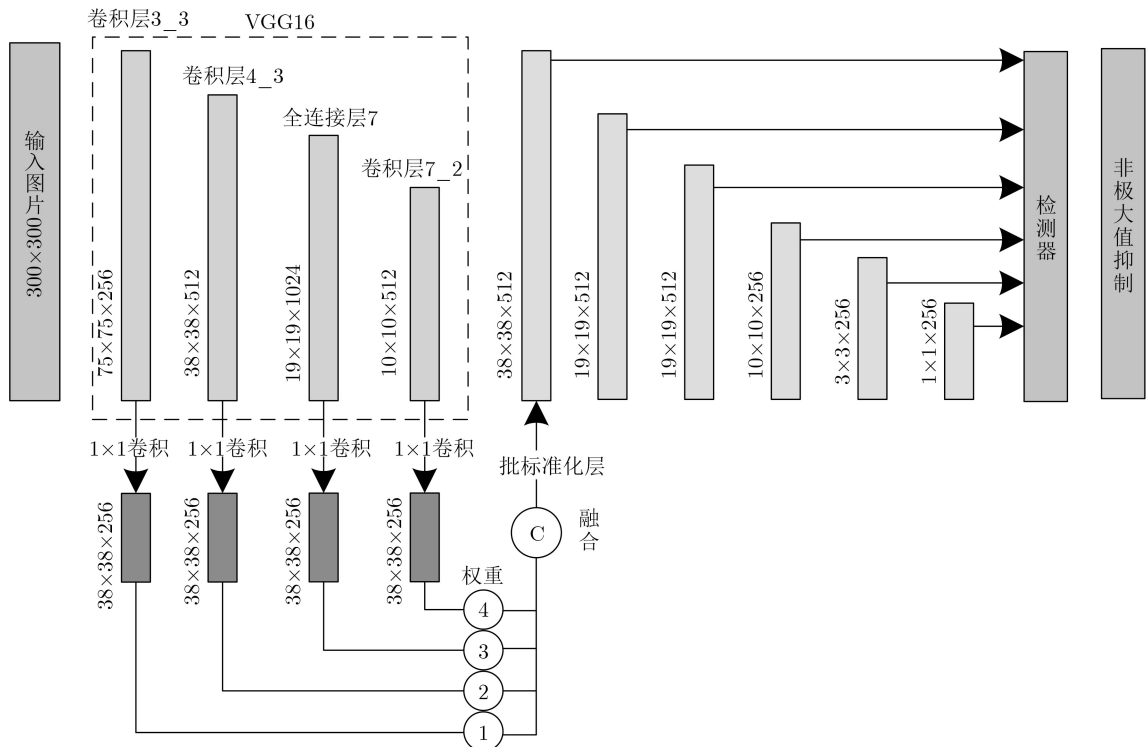


图2 MWFSSD检测流程图



图3 地铁行人样本库示例

集和地铁行人数据集目标尺寸的分布情况,其中横纵坐标分别代表目标框高和宽的尺寸,可以看出COCO数据集图4(a)中不同尺寸目标分布均匀,大中小目标都有,而我们的地铁行人数据集图4(b)目标尺寸

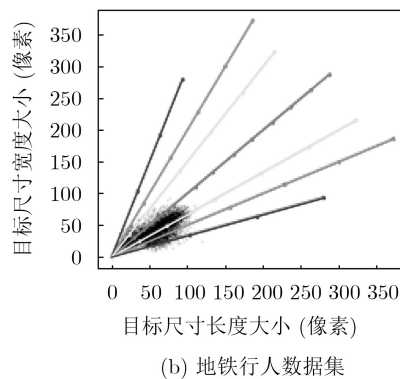
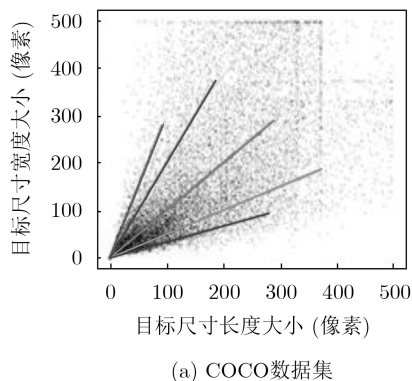


图4 COCO数据集和地铁行人数据集目标尺寸分布

图5是小目标过采样增强算法的原理图。首先统计出地铁行人数据集中每幅图片中的小目标(尺寸小于 64×64)占总目标的比例 R_s ,若比例小于阈值 τ ,则说明小目标占比小,应该增加小目标的数量。因此,本文设计了小目标增强预处理模块,将这些图片做拼接处理,并添加新的目标框节点坐标到新的数据文件中最终与原图片一起送入设计好的检测网络进行训练,指导网络更新^[18]。假设原始图片高和宽分别为 h 和 w ,则拼接后的图片高和宽变成 $2h$, $2w$ 。设原始图片中目标位置为 $(X_{\min}, Y_{\min}, X_{\max}, Y_{\max})$ 则该目标扩充后的4个新目标坐标分别为

$$(X_{\min}, Y_{\min}, X_{\max}, Y_{\max}) \quad (2)$$

$$(X_{\min}+w, Y_{\min}, X_{\max}+w, Y_{\max}) \quad (3)$$

$$(X_{\min}, Y_{\min}+h, X_{\max}, Y_{\max}+h) \quad (4)$$

$$(X_{\min}+w, Y_{\min}+h, X_{\max}+w, Y_{\max}+h) \quad (5)$$

3.2 模型的评价指标

本文采用准确率(P)、召回率(R)、平均精度均

集中在 96×96 以内的小、中尺寸目标,几乎不存在较大的目标,这就需要我们的模型对于小尺寸目标具有更强的检测性能。

小目标在地铁行人数据集中非常常见,而它们在不同图像中的分布是不可预测的。在整个数据集中,包含小目标的图片数量极少,然而却存在个别图像中几乎所有的目标都是小目标。这种数据分布的不平衡严重妨碍了训练进程。小目标数量少且尺寸相对较小,因此在训练中对于网络的优化作用较小,这种不平衡的网络参数优化导致检测性能不佳。根据这一发现,本文提出了一种数据增强的方法,使用过采样方法构造并扩充小目标样本,将小目标占比低的图像尺寸缩小到原来的 $1/4$,并将4张缩小后的图像拼接成与原图像尺寸相同的图片送入网络,从而增加小目标数据的权重。

值(mean Average Precision, mAP)、每秒帧率(frame per second, fps)对算法进行性能评价。其中,平均精度均值mAP定义为所有类别检测的平均准确度,通常用mAP作为评估算法精度的最终指标。目标被成功检测出认定为正样本,未被成功检测出则认为是负样本。IoU阈值设为0.5,大于0.5则认为是检测正确,记为 T ;否则记为 F 。实验结果中,IoU小于等于0.5的记为FP,IoU大于0.5的记为TP。

准确率计算公式为

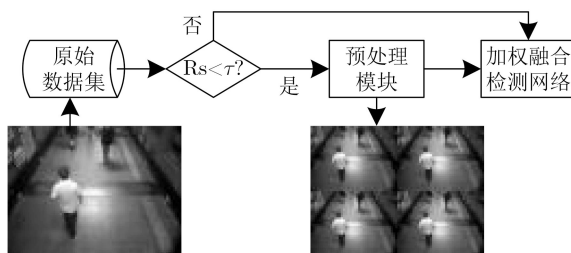


图5 小目标过采样增强算法

$$P = \frac{T_P}{T_P + F_P} \quad (6)$$

召回率计算公式为

$$R = \frac{T_P}{T_P + F_N} \quad (7)$$

其中, T_P 是正样本检测出的数量, F_P 是负样本检测出的数量, F_N 是正样本被误识为负样本的数量。

本文AP采用VOC2007中的11-point计算方法, mAP为AP的平均值

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, \dots, 1\}} \rho_{\text{interp}(r)} \quad (8)$$

$$\rho_{\text{interp}(r)} = \max_{\tilde{r}: \tilde{r} \geq r} \rho(\tilde{r}) \quad (9)$$

$$\text{mAP} = \frac{1}{|Q_R|} \sum_{q \in Q_R} AP(q) \quad (10)$$

其中, $\rho(\tilde{r})$ 是召回率满足 $\tilde{r} \geq r$ 时的最大精度, Q_R 为类别数, 这里为背景和头肩目标两类。

除了检测准确度, 目标检测算法的另外一个重要性能指标是速度, 只有速度快, 才能实现实时检测, 这在地铁行人检测应用中极其重要。评估速度的常用指标是每秒帧率, 即每秒内可以处理的图片数量。在实际应用中, 每秒帧数不能低于20 fps, 否则会影响录像视频的流畅度^[19]。

模型的时间复杂度可以用浮点运算次数(floating-point operations)即FLOPS衡量。单个卷积层的时间复杂度计算公式如式(11)所示

$$\text{Time} \sim O(M^2 \times K^2 \times C_{\text{in}} \times C_{\text{out}}) \quad (11)$$

其中, M 表示每个卷积层的时间复杂度由卷积后输出特征图的大小 M^2 , 卷积核的大小 K^2 , 输入通道数与输出通道数即 C_{in} 与 C_{out} 共同决定。

最终, 经过计算, MWFSSD相对于SSD网络而言, 时间复杂度增加了31.31GFLOPs, 没有大幅增加网络的时间开销。

3.3 特征层选择结果

本文实验均基于地铁行人头肩数据集, 选用VGG16网络的Conv3_3(Conv3), Conv4_3(Conv4), FC7(Conv7), Conv7_2(Conv8), 4层特征层进行不同方式融合, 与原始SSD网络进行对比, 得到不同特征层融合后的模型及其模型精度mAP的对比实验结果。部分实验结果如表1所示。

由表1可知, 当由Conv3, Conv4, Conv7, Conv8 4层特征图一起融合后的模型精度可以达到86.48, 比原始SSD高出2.58, 说明融合底层特征的信息对模型精度的提升有很大影响。

3.4 本文网络MWFSSD实验结果

本文在Conv3+Conv4+Conv7+Conv8的融合模型基础上, 分别为4个特征层分配不同权重 w_1, w_2, w_3, w_4 再进行融合, 并且 $w_1 + w_2 + w_3 + w_4 = 1$, 融合公式如式(1)。

本文做了大量实验, 验证了不同权重分配方式的检测精度。表2记录了本文选择的分配权重与无权重进行比较的检测结果。结果证明, 浅层特征更有利于中、小目标检测, 加入权重进行融合比不加权重融合模型mAP提高了1.44, 有效提高了模型精度。

最终, 结合小目标过采样增强算法与多尺度特征加权融合网络进行实验验证, 表3记录了本文模型与主流目标检测模型SSD, Faster-RCNN, YoloV3在精度和速度上的对比结果。本文所提模型在检测精度上较原SSD模型提升了5.82, 相较于其他模型也有不同程度的提高, 速度上SSD每秒检测43帧图像, 本文模型MWFSSD每秒检测32帧图像, 速度有轻微下降是因为模型深度有所增加, 但与精度较高的Faster-RCNN相比, 速度上有明显优势, 并且仍能满足实时性的要求。图6是部分SSD与

表 1 特征层融合精度结果比较

方法	预训练	融合层	mAP
SSD300	×	None	64.89
SSD300	✓	None	83.90
MWFSSD300	×	Conv4+Conv7+ Conv8	82.79
MWFSSD300	✓	Conv4+Conv7+ Conv8	85.12
MWFSSD300	✓	Conv3+Conv7	83.67
MWFSSD300	✓	Conv7+ Conv8	82.10
MWFSSD300	✓	Conv3+Conv4+Conv7+ Conv8	86.48
MWFSSD300	✓	Conv4+Conv7+ Conv8	85.22

表 2 不同权重分配的检测结果

权重分配方式	w_1	w_2	w_3	w_4	mAP
0	0	0	0	0	86.48
1	0.4	0.3	0.2	0.1	86.39
2	0.5	0.3	0.1	0.1	87.92
3	0.2	0.2	0.3	0.3	86.44

表 3 MWFSSD与主流检测方法检测结果对比

网络	mAP(%)	fps(frame/s)
MWFSSD	89.72	32
SSD	83.90	43
Faster-RCNN	88.86	20
YoloV3	86.50	38

MWFSSD检测结果对比图,可以看出本文方法有效改善了地铁行人检测中漏检、错检的问题。实验

证明,我们的方法在提高精度的同时,可以满足实时性需求^[20,21]。



图6 SSD与MWFSSD检测效果对比图

4 结论

本文针对地铁行人检测中小尺寸目标数量不足、分布不均衡的特点提出了针对小目标数据的过采样增强算法,根据尺度比例增加小目标的利用率。同时,针对地铁行人目标小的问题,提出了多尺度加权融合算法,并进行大量实验确定最优参数,提高对小目标的检测精度。结果显示,本文所提地铁行人多尺度加权特征融合算法在保证实时性检测的同时,检测精度mAP达到89.72%,相比原始SSD算法提升了5.82%,能够更准确地实时监测地铁中的客流量。

参考文献

- [1] GIRSHICK R. Fast R-CNN[C]. 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1440–1448. doi: [10.1109/ICCV.2015.169](https://doi.org/10.1109/ICCV.2015.169).
- [2] REN Shaoqing, HE Kaiming, 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. doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [3] 王殿伟, 何衍辉, 李大湘, 等. 改进的YOLOv3红外视频图像行人检测算法[J]. 西安邮电大学学报, 2018, 23(4): 48–52, 67. doi: [10.13682/j.issn.2095-6533.2018.04.008](https://doi.org/10.13682/j.issn.2095-6533.2018.04.008).
WANG Dianwei, HE Yanhui, LI Daxiang, *et al*. An improved infrared video image pedestrian detection algorithm[J]. *Journal of Xi'an University of Posts and Telecommunications*, 2018, 23(4): 48–52, 67. doi: [10.13682/j.issn.2095-6533.2018.04.008](https://doi.org/10.13682/j.issn.2095-6533.2018.04.008).
- [4] LIU Wei, ANGUELOV D, ERHAN D, *et al*. SSD: Single shot MultiBox detector[J]. *Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands*, 2016: 21–37. doi: [10.1007/978-3-319-46448-0_2](https://doi.org/10.1007/978-3-319-46448-0_2).
- [5] 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, USA, 2017: 936–944. doi: [10.1109/CVPR.2017.106](https://doi.org/10.1109/CVPR.2017.106).
- [6] 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. doi: [10.16208/j.issn1000-7024.2018.11.032](https://doi.org/10.16208/j.issn1000-7024.2018.11.032).
- LÜ Junqi, QIU Weigen, ZHANG Lichen, *et al.* Multi-scale convolutional feature fusion for pedestrian detection[J]. *Computer Engineering and Design*, 2018, 39(11): 3481–3485. doi: [10.16208/j.issn1000-7024.2018.11.032](https://doi.org/10.16208/j.issn1000-7024.2018.11.032).
- [7] 张文明, 姚振飞, 高雅昆, 等. 一种平衡准确性以及高效性的显著性目标检测深度卷积网络模型[J]. 电子与信息学报, 2020, 42(5): 1201–1208. doi: [10.11999/JEIT190229](https://doi.org/10.11999/JEIT190229).
- ZHANG Wenming, YAO Zhenfei, GAO Yakun, *et al.* A deep convolutional network for saliency object detection with balanced accuracy and high efficiency[J]. *Journal of Electronics & Information Technology*, 2020, 42(5): 1201–1208. doi: [10.11999/JEIT190229](https://doi.org/10.11999/JEIT190229).
- [8] 刘晴, 唐林波, 赵保军, 等. 基于自适应多特征融合的均值迁移红外目标跟踪[J]. 电子与信息学报, 2012, 34(5): 1137–1141. doi: [10.3724/SP.J.1146.2011.01077](https://doi.org/10.3724/SP.J.1146.2011.01077).
- LIU Qing, TANG Linbo, ZHAO Baojun, *et al.* Infrared target tracking based on adaptive multiple features fusion and mean shift[J]. *Journal of Electronics & Information Technology*, 2012, 34(5): 1137–1141. doi: [10.3724/SP.J.1146.2011.01077](https://doi.org/10.3724/SP.J.1146.2011.01077).
- [9] 颜伟, 耿路, 周雷, 等. 基于海情和三次样条插值算法的舰船雷达散射截面优化分析方法[J]. 电子与信息学报, 2018, 40(3): 579–586. doi: [10.11999/JEIT170562](https://doi.org/10.11999/JEIT170562).
- YAN Wei, GENG Lu, ZHOU Lei, *et al.* Optimization analysis method on ship RCS based on sea conditions and cubic spline interpolation algorithm[J]. *Journal of Electronics & Information Technology*, 2018, 40(3): 579–586. doi: [10.11999/JEIT170562](https://doi.org/10.11999/JEIT170562).
- [10] 邓苗, 张基宏, 柳伟, 等. 基于全变分的权值优化的多尺度变换图像融合[J]. 电子与信息学报, 2013, 35(7): 1657–1663. doi: [10.3724/SP.J.1146.2012.01183](https://doi.org/10.3724/SP.J.1146.2012.01183).
- DENG Miao, ZHANG Jihong, LIU Wei, *et al.* A total variation-based lowpass weight function optimization in multiscale image fusion[J]. *Journal of Electronics & Information Technology*, 2013, 35(7): 1657–1663. doi: [10.3724/SP.J.1146.2012.01183](https://doi.org/10.3724/SP.J.1146.2012.01183).
- [11] 李秋华, 李吉成, 沈振康. 基于多尺度特征融合的红外图像小目标检测[J]. 系统工程与电子技术, 2005, 27(9): 1557–1560. doi: [10.3321/j.issn:1001-506X.2005.09.018](https://doi.org/10.3321/j.issn:1001-506X.2005.09.018).
- LI Qihua, LI Jicheng, and SHEN Zhenkang. IR image small target detection based on multi-scale feature fusion[J]. *Systems Engineering and Electronics*, 2005, 27(9): 1557–1560. doi: [10.3321/j.issn:1001-506X.2005.09.018](https://doi.org/10.3321/j.issn:1001-506X.2005.09.018).
- [12] 姜文涛, 张驰, 张晟翀, 等. 多尺度特征图融合的目标检测[J]. 中国图象图形学报, 2019, 24(11): 1918–1931. doi: [10.11834/jig.190021](https://doi.org/10.11834/jig.190021).
- JIANG Wentao, ZHANG Chi, ZHANG Shengchong, *et al.* Multiscale feature map fusion algorithm for target detection[J]. *Journal of Image and Graphics*, 2019, 24(11): 1918–1931. doi: [10.11834/jig.190021](https://doi.org/10.11834/jig.190021).
- [13] 王瑶, 王正勇, 何小海, 等. 基于多尺度训练库与多特征融合的人脸识别[J]. 电视技术, 2015, 39(1): 121–126. doi: [10.16280/j.videoe.2015.01.031](https://doi.org/10.16280/j.videoe.2015.01.031).
- WANG Yao, WANG Zhengyong, HE Xiaohai, *et al.* Face recognition by features fusion based on multiscale training set[J]. *Video Engineering*, 2015, 39(1): 121–126. doi: [10.16280/j.videoe.2015.01.031](https://doi.org/10.16280/j.videoe.2015.01.031).
- [14] 余春艳, 徐小丹, 钟诗俊. 面向显著性目标检测的SSD改进模型[J]. 电子与信息学报, 2018, 40(11): 2554–2561. doi: [10.11999/JEIT180118](https://doi.org/10.11999/JEIT180118).
- YU Chunyan, XU Xiaodan, and ZHONG Shijun. An improved SSD model for saliency object detection[J]. *Journal of Electronics & Information Technology*, 2018, 40(11): 2554–2561. doi: [10.11999/JEIT180118](https://doi.org/10.11999/JEIT180118).
- [15] 孙彦景, 石楦开, 云霄, 等. 基于多层卷积特征的自适应决策融合目标跟踪算法[J]. 电子与信息学报, 2019, 41(10): 2464–2470. doi: [10.11999/JEIT180971](https://doi.org/10.11999/JEIT180971).
- SUN Yanjing, SHI Yunkai, YUN Xiao, *et al.* Adaptive strategy fusion target tracking based on multi-layer convolutional features[J]. *Journal of Electronics & Information Technology*, 2019, 41(10): 2464–2470. doi: [10.11999/JEIT180971](https://doi.org/10.11999/JEIT180971).
- [16] 张思宇, 张轶. 基于多尺度特征融合的小目标行人检测[J]. 计算机工程与科学, 2019, 41(9): 1627–1634. doi: [10.3969/j.issn.1007-130X.2019.09.014](https://doi.org/10.3969/j.issn.1007-130X.2019.09.014).
- ZHANG Siyu and ZHANG Yi. Small target pedestrian detection based on multi-scale feature fusion[J]. *Computer Engineering and Science*, 2019, 41(9): 1627–1634. doi: [10.3969/j.issn.1007-130X.2019.09.014](https://doi.org/10.3969/j.issn.1007-130X.2019.09.014).
- [17] DOLLAR P, WOJEK C, SCHIELE B, *et al.* Pedestrian detection: An evaluation of the state of the art[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(4): 743–761. doi: [10.1109/TPAMI.2011.155](https://doi.org/10.1109/TPAMI.2011.155).
- [18] 汪荣贵, 韩梦雅, 杨娟, 等. 多级注意力特征网络的小样本学习[J]. 电子与信息学报, 2020, 42(3): 772–778. doi: [10.11999/JEIT190242](https://doi.org/10.11999/JEIT190242).
- WANG Ronggui, HAN Mengya, YANG Juan, *et al.* Multi-level attention feature network for few-shot learning[J]. *Journal of Electronics & Information Technology*, 2020, 42(3): 772–778. doi: [10.11999/JEIT190242](https://doi.org/10.11999/JEIT190242).
- [19] 代科学, 李国辉, 涂丹, 等. 监控视频运动目标检测减背景技术的研究现状和展望[J]. 中国图象图形学报, 2007, 11(7): 919–927. doi: [10.3969/j.issn.1006-8961.2006.07.002](https://doi.org/10.3969/j.issn.1006-8961.2006.07.002).
- DAI Kexue, LI Guohui, TU Dan, *et al.* Prospects and current studies on background subtraction techniques for moving objects detection from surveillance video[J]. *Journal of Image and Graphics*, 2007, 11(7): 919–927. doi: [10.3969/j.issn.1006-8961.2006.07.002](https://doi.org/10.3969/j.issn.1006-8961.2006.07.002).

- [20] 陈勇, 刘曦, 刘焕淋. 基于特征通道和空间联合注意机制的遮挡行人检测方法[J]. 电子与信息学报, 2020, 42(6): 1486–1493. doi: [10.11999/JEIT190606](https://doi.org/10.11999/JEIT190606).

CHEN Yong, LIU Xi, and LIU Huanlin. Occluded pedestrian detection based on joint attention mechanism of channel-wise and spatial information[J]. *Journal of Electronics & Information Technology*, 2020, 42(6): 1486–1493. doi: [10.11999/JEIT190606](https://doi.org/10.11999/JEIT190606).

- [21] 贺丰收, 何友, 刘准钊, 等. 卷积神经网络在雷达自动目标识别中的研究进展[J]. 电子与信息学报, 2020, 42(1): 119–131. doi: [10.11999/JEIT180899](https://doi.org/10.11999/JEIT180899).

HE Fengshou, HE You, LIU Zhunga, *et al.* Research and development on applications of convolutional neural networks of radar automatic target recognition[J]. *Journal*

of Electronics & Information Technology, 2020, 42(1): 119–131. doi: [10.11999/JEIT180899](https://doi.org/10.11999/JEIT180899).

董小伟: 女, 1978年生, 博士, 研究方向为高速信号处理.

韩 悦: 女, 1996年生, 硕士生, 研究方向为人工智能与图像处理.

张 正: 男, 1983年生, 副研究员, 研究方向为人工智能与图像处理.

曲洪斌: 男, 1976年生, 工程师, 研究方向为信息化应用.

高国飞: 男, 1983年生, 高级工程师, 研究方向为城市轨道交通安全.

陈明钿: 男, 1991年生, 工程师, 研究方向为城市轨道交通安全.

李 博: 男, 1995年生, 硕士生, 研究方向为人工智能与图像处理.

责任编辑: 余 蓉