

基于深度学习的多模态行人检测算法

李晓艳¹, 符惠桐³, 牛文涛¹, 王鹏², 吕志刚¹, 王伟明⁴

(1. 西安工业大学电子信息工程学院, 710021, 西安; 2. 西安工业大学发展规划处, 710021, 西安;
3. 西安翔迅科技有限责任公司, 710068, 西安; 4. 陕西航天技术应用研究院有限公司, 710100, 西安)

摘要: 针对全天候工作的多模态行人检测算法体积大、运算量高、效率不足的问题,提出一种基于深度学习 MBNet 算法搭建的轻量级多模态行人检测算法(G-MBNet)。采用 ResNet18 算法并结合跨阶段链接的思想搭建 CSP-ResNet18 轻量级特征提取网络,以保证检测算法精度;引入轻量级高效通道注意力(ECA)模块来提升特征提取网络对重要特征的专注能力,在引入极少参数的情况下提升算法的检测精度;通过引入轻量级 Ghost 卷积模块来重构 MBNet 算法的特征提取网络,在保证特征提取性能的情况下进一步降低算法的参数与体积,提升算法的检测速度。采用所提的 G-MBNet 算法在 KAIST 行人数据集进行测试,实验结果表明:G-MBNet 算法大小是原始算法的 32.33%,参数量是原始算法的 37.81%,检测速度是原始算法的 1.53 倍;G-MBNet 算法可在保证行人识别精度的情况下有效提升检测速度。

关键词: 行人检测;多模态融合;轻量级网络;Ghost 卷积;注意力机制

中图分类号: TP391.4 **文献标志码:** A

DOI: 10.7652/xjtuxb202210006 **文章编号:** 0253-987X(2022)10-0061-10

Multi-Modal Pedestrian Detection Algorithm Based on Deep Learning

LI Xiaoyan¹, FU Huitong³, NIU Wentao¹, WANG Peng², LÜ Zhigang¹, WANG Weiming⁴

(1. School of Electronic Information Engineering, Xi'an Technological University, Xi'an 710021, China;
2. Development Planning Service, Xi'an Technological University, Xi'an 710021, China;
3. AVIC Xiangxun Technology Co., Ltd., Xi'an 710068, China;
4. Shaanxi Aerospace Technology Application Research Institute Co., Ltd., Xi'an 710100, China)

Abstract: This paper proposes a lightweight multi-modal pedestrian detection algorithm based on deep learning MBNet algorithm to address the problem of 24/7 multi-modal pedestrian detection algorithm having a large size, high computation load and insufficient efficiency. First, the CSP-ResNet18 lightweight feature extraction network is built using the ResNet18 algorithm combined with the idea of cross-stage linking to ensure accuracy of the detection algorithm; then, the lightweight efficient channel attention module is introduced to enhance the feature extraction network's ability to focus on important features, which can improve the detection accuracy of the algorithm by introducing very few parameters; finally, the feature extraction network of MBNet algorithm is reconstructed by introducing a lightweight Ghost convolution module, which can further reduce the parameters required and the size of the algorithm and improve its detection efficiency while ensuring the feature extraction performance. The proposed G-MBNet algorithm is

收稿日期: 2022-05-20。 作者简介: 李晓艳(1982—),女,副教授;王鹏(通信作者),男,教授。 基金项目: 国家自然科学基金资助项目(62171360);陕西省重点研发计划资助项目(2022GY-110)。

网络出版时间: 2022-07-08

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1069.T.20220707.1418.002.html>

tested on the KAIST pedestrian dataset, and the experimental results show that the size of G-MBNet algorithm is 32.33% that of the original algorithm; the number of parameters is 37.81% that of the original algorithm; the detection speed is 1.53 times that of the original algorithm. The experiments verify that G-MBNet algorithm can effectively improve the detection speed while ensuring the pedestrian recognition accuracy.

Keywords: pedestrian detection; multi-modal fusion; lightweight network; Ghost convolution; attention mechanism

行人检测是深度学习中标检测内一项重要的研究内容,广泛应用于安防安检、自动驾驶、人机交互等工业领域^[1]。卷积神经网络的快速发展及应用使目标检测的精度有了质的飞跃,但极大地依赖于卷积神经网络的多参数、多运算以及高功耗^[2]。例如,Faster R-CNN^[3]、SSD^[4]、YOLO^[5]等先进的目标检测算法逐渐被提出,行人检测任务从此变得高效、精准。此外,在军事作战、侦察中,以大规模有生力量的作战已经逐步被机械化、智能化的武器装备所取代,复杂的应用背景、全天候的应用需求、高效的检测效率对目标检测的鲁棒性、检测精度和推理延迟提出了更高的要求^[6]。如何在光照不足的夜间场景下提升行人检测算法的质量是目前研究的重点与难点。

单一的可见光传感器难以适应微光及无光环境,无法完成复杂环境下的行人检测工作,多传感器的信息融合有助于完成不同信息间的互补,从而更好地适应全天候的工作^[7]。可见光和红外数据的多光谱融合有效地提升了算法对于环境的感知能力,使夜间高精度行人检测成为了可能,各国学者也在积极进行研究。文献[8]由聚合通道特征的方法扩展而来,提出可见光和红外信息融合的聚合通道特征(ACF)+热通道+热通道直方图方向梯度的方法,并创建了KAIST多光谱行人检测数据集^[9]。该方法对可见光和红外光图像进行了图像配准,可有效地进行特征融合,但是该方法为传统方法,检测效率及准确率都很不理想。文献[10]首次将可见光图像和红外图像进行图像深度特征的融合,相较于传统方法在准确率上有了极大提升。文献[11]进一步对可见光和红外图像的特征级融合进行了探索,提出了更加高效的中层特征融合方案。文献[12]基于中层特征级融合方法,结合Faster R-CNN中候选区建议网络(RPN)^[13-14]的思想,最终采用提高决策树(BDT)分类的方法,进一步提升了行人检测精度。文献[15]基于Faster R-CNN目标检测算法,在不同的阶段进行可见光和红外图像深层特征的融合,并且提出一种感知光照导向的融合算法,实现了全天候行人的精确检测,但是效

率依然不高。文献[16]讨论了多模态融合检测过程中存在的模态不平衡的问题,并搭建了MBNet网络以解决模态不平衡的问题,极大地提升了算法的检测效率,但算法体积庞大,推理效率有较大的优化空间。综上所述可知,多模态行人检测算法大多数存在推理效率不高、算法体积庞大的问题。

本文在MBNet算法的基础上,提出了一种针对可见光和红外多模态融合的轻量级行人检测算法(G-MBNet),在保证检测精度的情况下,有效地压缩了算法体积、提升了算法推理效率。本文以ResNet18^[17]为基础网络,结合跨链接^[18]的思想搭建跨阶段部分残差网络(CSP-ResNet18),并且在残差结构内容引入轻量级注意力机制ECA^[19]提升算法性能,最终通过Ghost轻量级模块^[20]进行算法重构,减少算法参数,提升算法效率。

1 MBNet 多模态行人检测算法

MBNet多模态行人检测算法由特征提取结构、光照感知结构和特征对齐模块共3部分组成,如图1所示。

在特征提取阶段融入了差分模态感知融合(DMAF)模块以解决特征模态不平衡问题。该模块以差分放大电路为原型,通过两个模态特征之间进行互补,以一种模态去增强另外一种模态,从而可以提高算法对另一个模态特征信息的敏感程度,提升两个通道间信息交互的程度。DMAF模块具体公式为

$$F_w = \frac{F_{w1} + F_{w2}}{2} + \frac{F_{r1} - F_{r2}}{2} = \frac{F_{w1} + F_{r1}}{2} + \frac{F_{w2} - F_{r2}}{2} \quad (1)$$

$$F_r = \frac{F_{r1} + F_{r2}}{2} + \frac{F_{w1} - F_{w2}}{2} = \frac{F_{r1} + F_{w1}}{2} + \frac{F_{r2} - F_{w2}}{2} \quad (2)$$

式中: F_w 和 F_r 分别表示模态增强后的可见光和红外图像特征; F_{w1} 和 F_{w2} 分别表示模态增强前的通道1和通道2的可见光图像特征; F_{r1} 和 F_{r2} 分别表示模态增强前的通道1和通道2的红外图像特征。

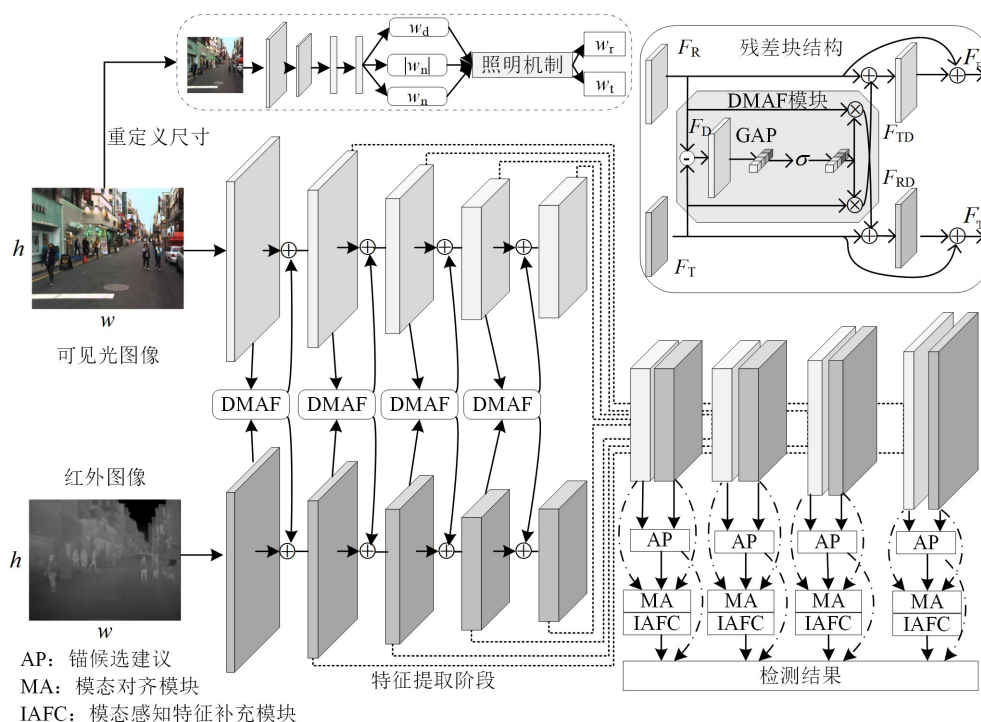


图 1 MBNet 多模态行人检测算法

Fig. 1 MBNet multi-modal pedestrian detection algorithm

差分模态很好地反映了两个模态之间的差异。为了更好地让差分信息融入至对应的通道中,对差分信息进行平均全局池化、激活以及点乘操作,之后与对应通道进行相加得到新的融合特征,公式为

$$F'_R = F_R + f(F_R \oplus F_{TD}) = F_R + f(F_R \oplus (\sigma(\text{GAP}(F_D)) \otimes F_R)) \quad (3)$$

$$F'_T = F_T + f(F_T \oplus F_{RD}) = F_T + f(F_T \oplus (\sigma(\text{GAP}(F_D)) \otimes F_T)) \quad (4)$$

式中: F'_R 和 F'_T 分别代表经过差分放大后的可见光和红外通道特征; f 表示残差映射; GAP 表示全局平均池化; σ 表示激活函数; $\oplus$ 、 $\otimes$ 分别表示逐元素相加和相乘操作; F_{TD} 表示可见光互补特征; F_{RD} 表示红外互补特征。

此外,光照感知特征对齐模块可以使算法适应不同的光照条件,并且在区域建议阶段后对齐两个模态的特征。其中,光照感知值通过一个微型神经网络捕获照明机制获得,只采用 RGB 图像进行光照感知,得到的感知系数作为可见光和红外特征的权重系数并进行加权融合,得到最终的特征值。由于可见光和红外图像并不是完美校准,所以通过形态对齐模块用于预测两个模态之间的偏移量,以此进

行模态之间的校准。此外,紧接着基于光照感知特征补充阶段对可变形锚框和置信度分数进行微调,提高检测精度。

总体来说,特征提取阶段计算量极大,推理效率不足,所以本文主要对特征提取阶段的网络进行轻量化设计,以压缩算法体积、加快整体算法的推理效率。

2 G-MBNet 轻量化行人检测算法

本文基于 MBNet 多模态行人检测检测算法,通过特征提取网络构建、网络模型优化、网络模型重构等方法,在保证检测精度的情况下对检测算法进行轻量化设计,最终的轻量化设计方案如图 2 所示。首先,选用 MBNet 算法的 ResNet50 特征提取网络作为基础特征提取网络;其次,在特征提取阶段采用更加轻量的 ResNet18 特征提取网络,并结合跨阶段连接的方式搭建 CSP-ResNet18 算法,同时在残差中嵌入 ECA 轻量级注意力机制,以很小的计算代价提升轻量级特征提取网络性能;最终,利用即插即用的轻量级模块重构特征提取网络,基于冗余思想而设计的 Ghost 卷积可进一步压缩算法体积,加快算法效率。

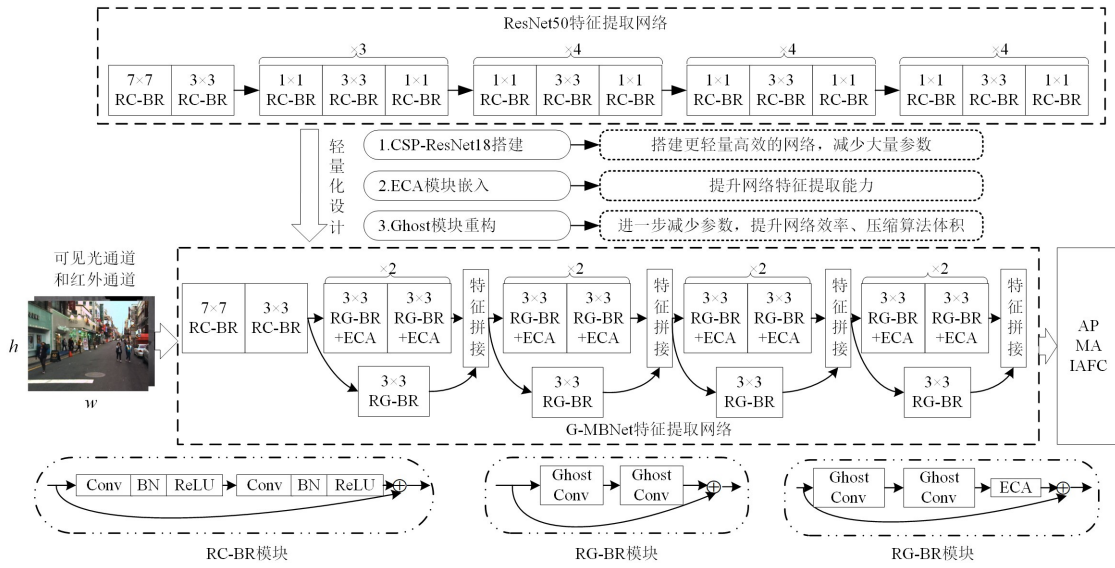


图 2 轻量化方案示意

Fig. 2 Schematic diagram of lightweight scheme

2.1 CSPResNet18 特征提取网络

在 MBNet 网络中,两个特征采集通道均采用 ResNet50 网络,网络模型较大且推理时延较长,所以本文选择最轻量的 ResNet18 结构作为基础网络进行优化,并且保证识别精度。ResNet 网络的残差结构如图 3 所示。

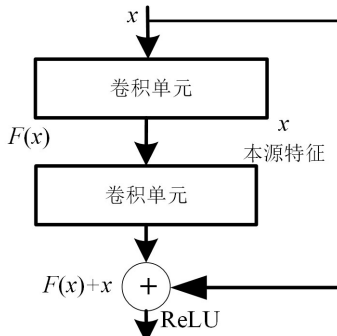


图 3 残差结构示意

Fig. 3 Schematic diagram of residual structure

由于 ResNet18 网络结构的特征提取能力随着网络深度的减少而减弱,所以本文参考借鉴 CSPNet 网络的跨阶段部分连接网络,对 ResNet18 进行网络结构优化,跨阶段部分 CSP 网络结构如图 4 所示。

前向预测过程可表示为

$$X_k = W_k * [X'_0, X_1, \dots, X_{k-1}] \quad (5)$$

$$X_T = W_T * [X'_0, X_1, \dots, X_k] \quad (6)$$

$$X_U = W_U * [X'_0, X_T] \quad (7)$$

式中: $*$ 表示卷积操作; X_k 为第 k 层的输出特征; W_k 为第 k 层的权重; $[X'_0, X_1, \dots, X_{k-1}]$ 表示将输

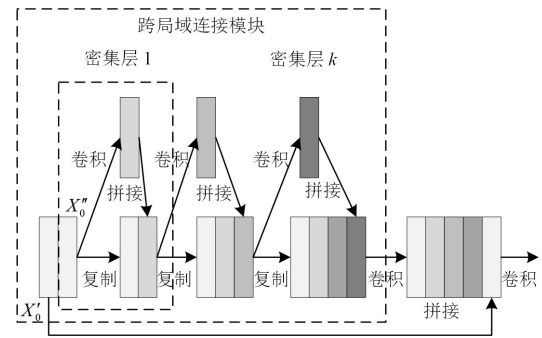


图 4 CSP 结构示意图

Fig. 4 Diagram of CSP structure

入特征 $X'_0, X_1, \dots, X_{k-1}$ 进行拼接操作; X'_0, X_0 为输入特征 X_0 的两个部分, X'_0 分别经过式(5)、(6)两个卷积层后与 X_0 进行拼接操作,最后通过卷积操作输出特征 X_U ,从而完成该阶段的前向预测。

CSP 结构有效地丰富了网络优化中的梯度信息,增强了卷积神经网络的学习能力,可以在减少参数、增加效率的同时,保证网络的准确性,并且有效地降低了内存成本。最终,基于 ResNet 的 CSP 结构如图 5 所示。

2.2 ECA 轻量级注意力机制

人类视觉在快速地进行全局图像扫描后,可以关注到一些重点目标区域,对需要注意的地方投入更多的精力,以获取所需的细节信息。与此同时,会降低对其他区域投入的权重。通过这种方式,可以极大地提升图像识别精度与处理效率。研究人员受此启发后发现:应对识别重要的特征数据提升权

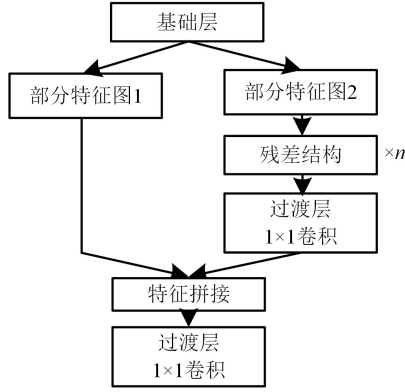


图 5 CSPResNet 结构

Fig. 5 CSPResNet structure diagram

重,增强重要特征对识别结果的影响;应压制不重要数据的权重,减弱其在网络中的影响力。

轻量级 ECA 通道注意力机制以 SE 注意力机制^[21]为原型,但其没有使用降维操作,从而避免了降维而引起的信息丢失,同时进行了适当的跨信道交互,在显著降低模型复杂度的同时保证了模型性能。此外,自适应卷积核的方法确定了局部信道的覆盖范围,十分高效。ECA 模块示意如图 6 所示。

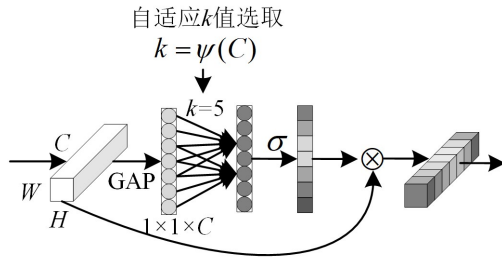


图 6 ECA 通道注意力示意

Fig. 6 ECA channel attention diagram

ECA 模块相较传统的注意力机制避免了全连接操作,并且使用了一个 $k \times C$ 的矩阵 $\mathbf{W}_k$ 用于学习通道间的注意力信息,矩阵 $\mathbf{W}_k$ 可表示为

$$\begin{bmatrix} \omega^{1,1} & \cdots & \omega^{1,k} & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \omega^{2,2} & \cdots & \omega^{2,k+1} & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \omega^{C,C-k+1} & \cdots & \omega^{C,C} \end{bmatrix} \quad (8)$$

矩阵 $\mathbf{W}_k$ 将可学习的权重分为 k 个一组,避免了不同组之间的完全独立。此时,输入权重 y_i 以及 k 个相邻通道之间的信息计算式为

$$\omega_i = \sigma \sum_{j=1}^k \omega_i^j y_i^j, \quad y_i^j \in \Omega_i^k \quad (9)$$

式中 Ω_i^k 表示 y_i 的 k 个相邻通道的集合。

为了让模块更加高效,采用权重共享的方式,公式为

$$\omega_i = \sigma \sum_{j=1}^k \omega_i^j y_i^j, \quad y_i^j \in \Omega_i^k \quad (10)$$

式(11)可以直接通过一个 $k \times k$ 的一维卷积实现,较为简单。 k 的选择十分重要,跨信道交互作用的覆盖范围应该与特征维度 C 成正比。映射关系可表示为

$$C = \varphi(k) = 2^{k-b} \quad (11)$$

此时, k 的表达式为

$$k = \psi(C) = \left\lceil \frac{\lg C}{\gamma} + \frac{b}{\gamma} \right\rceil_{\text{odd}} \quad (12)$$

式中: $\lceil \cdot \rceil_{\text{odd}}$ 表示取最接近的奇数;超参数 γ 和 b 根据经验设置为 2 和 1。

最终,ECA 模块引入 ResNet 的位置如图 7 所示。

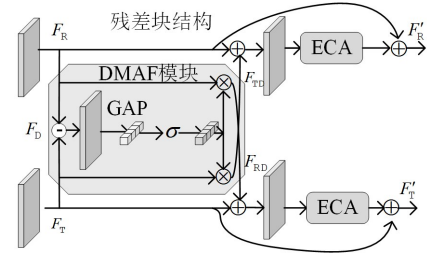


图 7 ECA 模块插入位置

Fig. 7 The ECA module insertion position

2.3 基于 Ghost 模块重构

Han 等^[20]认为,深度学习中对冗余信息的利用也很关键,由此基于对相似冗余信息的重复利用提出 Ghost 轻量级卷积模块。传统卷积和 Ghost 卷积模块卷积示意图如图 8 所示。

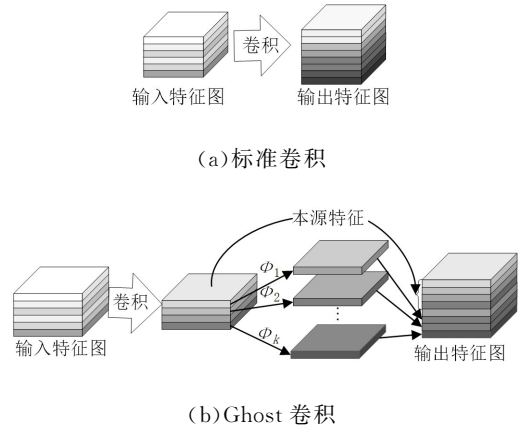


图 8 标准卷积与 Ghost 卷积

Fig. 8 Standard convolution and Ghost convolution

传统卷积操作计算了大量的近似特征,使最终的特征图存在过多重复信息,所以冗余信息的生成操作可以用更廉价的线性变换完成。Ghost 模块基于该思想完成冗余信息的计算。首先,通过传统的标准卷积完成 m 层本源特征的生成,公式为

$$\mathbf{Y}' = \mathbf{X} * f + b \quad (13)$$

式中: b 为偏置项;输出特征图 $\mathbf{Y}' \in \mathbb{R}^{h' \times w' \times m}$ 。接着,基于 $\mathbf{Y}'$ 进行线性映射

$$y_{ij} = \varphi_{i,j}(y'_i), \forall i = 1, \dots, m, \forall j = 1, \dots, s \quad (14)$$

式中: $y'_i \in \mathbf{Y}'$; $\varphi_{i,j}$ 表示本源特征 i 的第 j 次线性变换。

$\varphi_{i,j}$ 可以生成对应的 Ghost 特征 y_{ij} , 可见 y'_i 可以生成一个或多个 Ghost 特征。当 $j = s$ 时, $\varphi_{i,j}$ 为普通的卷积操作,最终获得 $n = ms$ 个特征图,其中 n 为卷积操作最终的通道数量。

传统卷积操作的浮点型运算量为 $n \times h' \times w' \times c \times k \times k$, 其中, c 为输入的通道数, Ghost 卷积由传统卷积操作结合 $m(s-1) = n/s(s-1)$ 次线性操作两部分,线性变换部分采用的卷积核平均尺寸为 $d \times d$, 故理论加速比率 R 为

$$R = \frac{nh'w'ckk}{(n/s)h'w'ckk + (s-1)(n/s)h'w'dd} = \frac{ckk}{s^{-1}ckk + ((s-1)/s)dd} = \frac{sc}{s+c-1} \approx s \quad (15)$$

由于 $d \times d = k \times k$, $s \ll c$, 所以可以认为, Ghost 卷积相比传统卷积,运算速度提升了 s 倍。

3 实验与分析

3.1 实验环境

实验测试环境为 Ubuntu 16.04 操作系统,由 Pytorch 深度学习框架实现 G-MBNet 多模态轻量化行人检测算法。服务器显卡型号为 NVIDIA RTX2080 Ti。

3.2 实验数据集及预训练

本文采用 KAIST 行人数据集,该数据集包含 95 328 对对齐的可见光和红外图像对,共 103 128 个边界框,覆盖 1 182 个独特的行人,图像尺寸为 640×512 像素。此外,测试集包含 2 252 帧从视频中每 20 帧采集的图像,其中,1 455 帧白天拍摄图像,797 帧夜间拍摄图像。评价指标为平均缺失率,表示每张图片的平均误检率(FPPI)在 $[0.01, 1]$ 范围内的对数空间均匀取 9 个点时的缺失率均值,即 FPPI 分别为 0.010 0、0.017 8、0.031 6、0.056 2、0.100 0、0.177 8、0.316 2、0.562 3、1.000 时的平均缺失率。平

均缺失率越低,则证明算法性能越好。

此外,为了使小网络有更加合理的初始权重、更加有效地进行收敛,本文提取 COCO 数据集^[22]中行人目标数据对 G-MBNet 以及中间网络进行预训练,从而获得更合理的初始化权重,实验结果如表 1 所示。

表 1 基于预训练权重的对比实验结果

Table 1 Comparative experiment based on pre-training model

算法	是否使用 预训练权重	时长/h	平均缺失率/%
MBNet	否	38.2	9.69
	是	25.4	8.13
G-MBNet	否	39.5	13.96
	是	26.2	13.74

由表 1 可以看出,未使用预训练权重的时长为使用预训练权重的 1.5 倍,使用预训练权重能够使时间成本大幅度降低。并且,使用预训练权重的漏检率优于未使用预训练权重的。在 MBNet 中,漏检率降低了 1.56%,在 G-MBNet 中,漏检率降低了 0.22%,证明了预训练权重的重要性。此外,MBNet 在未使用预训练权重时,红外通道出现了梯度爆炸的现象,预训练权重的引入使整个网络的训练更加稳定,证明了合理初始化权重在网络训练时的必要性。

3.3 消融实验

MBNet(基础算法)、CSPR-MBNet(在 MBNet 中引入 CSP 结构搭建的 ResNet18 网络 CSP-ResNet18)、CSPR-ECA-MBNet(引入 ECA 注意力模块后的骨干网络算法)、G-MBNet、G-MBNet-CBAM(在 Ghost-MBNet 中引入 CBAM^[23] 注意力机制)、G-MBNet-SE(在 Ghost-MBNet 中引入 SE 注意力机制)、DW-ECA-MBNet(在 MBNet 中引入深度可分离卷积(DW)^[24-25]与 ECA 注意力机制)算法的对比如表 2 所示。

由表 2 可以看出,CSPR-MBNet 相较于原始的 MBNet 参数减少了 55.51%,并且算法体积也减少了 62.38%,推理效率由原来的 52 ms 提升至 38 ms,效率提升 1.37 倍,精度仅下降了 4.58%。在引入 ECA 注意力机制后,算法参数、体积以及推理时延基本没有发生变化,但是精度相较于 MBNet 仅下降了 3.95%,证明了 ECA 模块的轻量以及对于检测精度有较大的提升。此外,本文对比了不同注意力机制,如 CBAM 与 SE 模块。在引入 CBAM 与 SE 后,由于引入的模块数量较少,所以算法的体积与运算量

变化不大。根据实验结果来看,轻量级 ECA 模块 具有更好的漏检率以及推理时延,有较高的性价比。

表 2 各轻量级算法的对比实验结果

Table 2 Comparison *experiment* of each lightweight model

算法	参数量/ 10^6	算法体积/MB	平均缺失率/%	时延/ms
MBNet	67.12	305.68	8.13	52
CSPR-MBNet	29.86	114.99	12.71	38
CSPR-ECA-MBNet	29.88	115.19	12.08	39
G-MBNet	25.38	98.82	13.74	34
G-MBNet-CBAM	25.59	104.86	14.94	39
G-MBNet-SE	25.59	104.53	16.19	45
DW-ECA-MBNet	21.89	84.68	18.52	32

注:表中加粗数据为最优值。

在 Ghost 卷积模块重构后,本文将 Ghost 模块与深度可分离卷积 DW 模块进行了性能比较,在使用 DW 卷积模块进行网络重构后发现,推理时延有较大的提升,优于 Ghost 重构网络的,但是漏检率相较于 G-MBNet 具有较大的差距。权衡之下可知,Ghost 卷积模块兼顾了精度与速度,具有较大的优势,证明了 Ghost 模块在本文方案中的适用性。最终,本文对比了搭建的各个轻量级算法的损失值曲线,如图 9 所示。

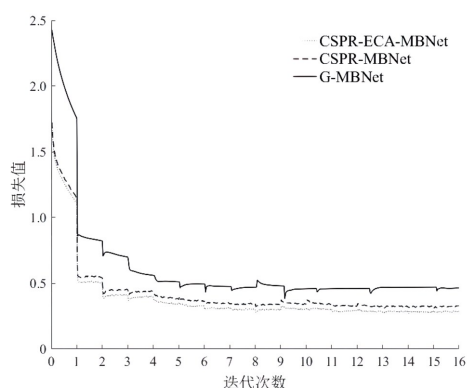


图 9 各轻量级算法损失值曲线的对比

Fig. 9 Comparison of loss value curves of various light-weight algorithms

由图 9 可以看出,各轻量级算法均有优秀的收敛性能。其中,CSPR-ECA-MBNet 收敛性能最佳,在第 9 次迭代后损失值趋于稳定。CSPR-MBNet 由于缺少 ECA 的支持,整体损失略微逊色于 CSPR-ECA-MBNet,也于第 9 次迭代后趋于稳定。G-MBNet 收敛性能较弱,不过在前 9 个迭代数也进行了快速收敛,同样达到了较好的收敛效果。

白天、夜晚情况下的 G-MBNet 行人检测结果如图 10 和图 11 所示。白天情况下的检测结果如下:从图 10(a)可以看出,当光线较好且人员可视效

果较好时,G-MBNet 具有良好的检测效果,相较于真值标签所发生的偏移小且精度良好;从图 10(b)可以看出,当光线较弱且目标较小时,发生了一个标注框标注多个目标的现象;从图 10(c)可以看出,当场景较为复杂且目标较小时,目标框发生了较大的偏移,检测效果较差。图 10(b)和 10(c)检测效果不佳的原因是特征提取网络深度的减少削弱了其特征提取能力。

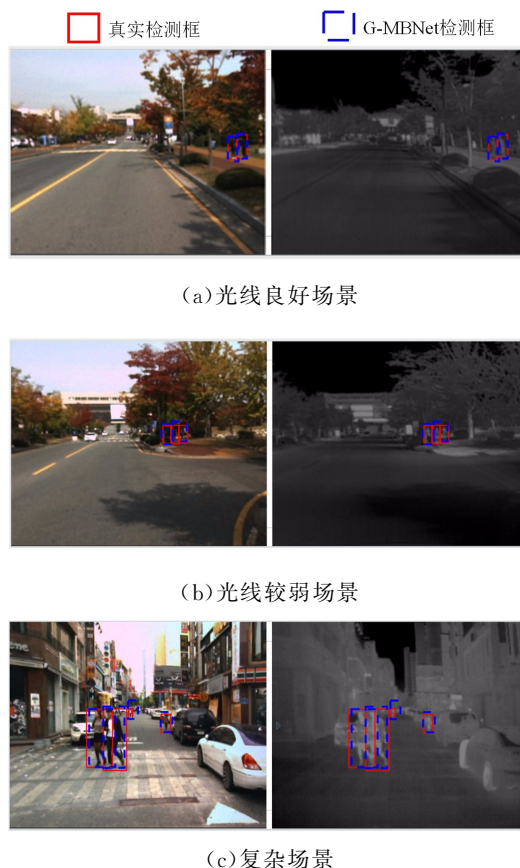


图 10 G-MBNet 在 KAIST-test 数据集中白天情况下的检测结果

Fig. 10 G-MBNet results in KAIST-test dataset in daytime

夜晚情况下的检测结果如下:从图 11(a)可以看出,当场景简单时,G-MBNet 检测效果较为理想;从图 11(b)和 11(c)可以看出,在面对小目标场景及复杂场景时,出现了小目标漏检的情况,检测效果较差。

综上所述可知,经过轻量化设计后的网络在进行多模态行人检测时具有良好的表现。但是,在针对远距离的小目标时,特征提取能力较弱,没有表现出较好的检测精度,在多个小目标在一起或者目标框较小时,出现了目标框定位较大、多目标在一个标注框内,甚至漏检的情况。总体而言,本文算法在多模态目标检测阶段检测速度较快、检测精度较高。为了更好地进行算法验证,本文在下一小节对不同算法进行了精度对比。

3.4 算法对比实验

本文算法与其他先进算法在 KAIST 数据集上的对比结果如表 3 所示。可以看出,经过轻量化后的行人检测算法在检测速度上有了极大提升。虽然精度相较于 MBNet 有部分损失,但在检测精度和检测时延上依然优于其他算法,并且速度达到了最优。由此,本文所设计的轻量级多模态行人检测算法 G-MBNet 的有效性得到了证明。

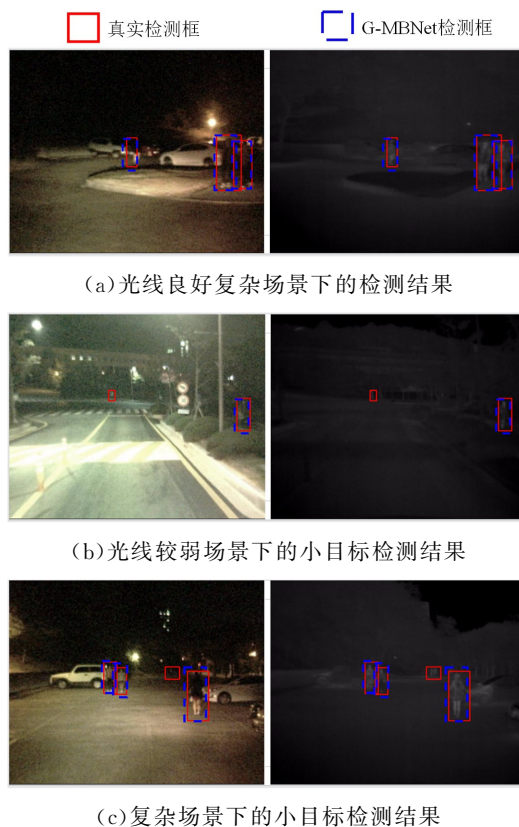


图 11 G-MBNet 在 KAIST-test 数据集中夜晚情况下的检测结果

Fig. 11 G-MBNet results in KAIST-test dataset in night

表 3 不同先进算法的对比结果

Table 3 Comparison of different SOT models

算法	平均缺失率/%			软件或硬件平台	时延/ms
	全部	白天	夜晚		
ACF ^[8]	47.32	42.57	56.17	MATLAB	273
Halfway Fusion ^[10]	25.75	24.88	26.59	TITAN X	430
Fusion RPN+BF ^[11]	18.29	19.57	16.27	MATLAB	800
IAFR-CNN ^[12]	15.73	14.55	18.26	TITAN X	210
CIAN ^[15]	14.12	14.77	11.13	1080 Ti	70
MBNet ^[16]	8.13	8.28	7.86	2080 Ti	52
G-MBNet	13.74	14.06	13.21	2080 Ti	34

注:表中加粗数据为最优值。

4 结 论

目前,主流的多模态行人检测算法大多数存在检测速度不足的问题,本文提出一种基于 Ghost 模块的搭建的 G-MBNet 轻量级多模态行人检测算法。该算法选取轻量级骨干网络 ResNet18,借鉴跨阶段部分 CSP 构架的思想重构 CSP-ResNet18 网络,在减少参数以及算法体积的基础上提升了 Res-

Net18 算法的检测性能。其次,引入 ECA 通道注意力机制。通过引入极少的参数提升了网络关注特征通道中重要的部分,从而提升了算法的检测精度。最终,通过 Ghost 轻量级卷积模块重构网络,减少了网络不必要的卷积操作,提升了整体算法的检测效率。实验结果表明,本文提出的轻量级多模态行人检测算法 G-MBNet 有效。相较于 MBNet, G-MBNet 体积更小、预测速度更快、精度损失更小。

本文提出的轻量化特征提取网络可以在其他检测方案中直接使用,为深度学习算法的部署提供了有效方案。但是,本文方案仅考虑了 MBNet 算法特征提取部分算法参数的减少,其他部分依旧存在较多的冗余操作,如光照感知部分的全连接层、复杂的特征对齐模块等,可以在未来的研究中提出更加高效的光照感知和特征对齐操作。

参考文献:

- [1] 童靖然,毛力,孙俊. 特征金字塔融合的多模态行人检测算法 [J]. 计算机工程与应用, 2019, 55(19): 214-222.
TONG Jingran, MAO Li, SUN Jun. Multimodal pedestrian detection algorithm based on fusion feature pyramids [J]. Computer Engineering and Applications, 2019, 55(19): 214-222.
- [2] 符惠桐,王鹏,李晓艳,等. 面向移动目标识别的轻量化网络模型 [J]. 西安交通大学学报, 2021, 55(7): 124-131.
FU Huitong, WANG Peng, LI Xiaoyan, et al. Light-weight network model for moving object recognition [J]. Journal of Xi'an Jiaotong University, 2021, 55(7): 124-131.
- [3] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]//Proceedings of the 2015 Conference on Advances in Neural Information Processing Systems. Vancouver, Canada: NIPS, 2015: 91-99.
- [4] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot multiBox detector [C]//Computer Vision-ECCV 2016. Cham, Germany: Springer International Publishing, 2016: 21-37.
- [5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 779-788.
- [6] 乔梦雨,王鹏,吴娇,等. 面向陆战场目标识别的轻量级卷积神经网络 [J]. 计算机科学, 2020, 47(5): 161-165.
QIAO Mengyu, WANG Peng, WU Jiao, et al. Light-weight convolutional neural networks for land battle target recognition [J]. Computer Science, 2020, 47(5): 161-165.
- [7] 程腾,孙磊,侯登超,等. 基于特征融合的多层次多模态目标检测 [J]. 汽车工程, 2021, 43(11): 1602-1610.
CHENG Teng, SUN Lei, HOU Dengchao, et al. Multi-level and multi-modal target detection based on feature fusion [J]. Automotive Engineering, 2021, 43(11): 1602-1610.
- [8] HWANG S, PARK J, KIM N, et al. Multispectral pedestrian detection: benchmark dataset and baseline [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2015: 1037-1045.
- [9] CHOI Y, KIM N, HWANG S, et al. KAIST multi-spectral day/night data set for autonomous and assisted driving [J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(3): 934-948.
- [10] LIU Jingjing, ZHANG Shaoting, WANG Shu, et al. Multispectral deep neural networks for pedestrian detection [EB/OL]. [2022-03-01]. <https://arxiv.org/abs/1611.02644>.
- [11] KÖNIG D, ADAM M, JARVERS C, et al. Fully convolutional region proposal networks for multispectral person detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2017: 243-250.
- [12] LI Chengyang, SONG Dan, TONG Ruofeng, et al. Illumination-aware faster R-CNN for robust multispectral pedestrian detection [J]. Pattern Recognition, 2019, 85: 161-171.
- [13] HE Kaiming, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 2980-2988.
- [14] GIRSHICK R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2015: 1440-1448.
- [15] ZHANG Lu, LIU Zhiyong, ZHANG Shifeng, et al. Cross-modality interactive attention network for multispectral pedestrian detection [J]. Information Fusion, 2019, 50: 20-29.
- [16] ZHOU Kailai, CHEN Linsen, CAO Xun. Improving multispectral pedestrian detection by addressing modality imbalance problems [C]//Computer Vision-ECCV 2020. Cham, Germany: Springer International Publishing, 2020: 787-803.
- [17] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.

- [18] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2020: 1571-1580.
- [19] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 11531-11539.
- [20] HAN Kai, WANG Yunhe, TIAN Qi, et al. GhostNet: more features from cheap operations [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 1577-1586.
- [21] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7132-7141.
- [22] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C]//Computer Vision-ECCV 2014. Cham, Germany: Springer International Publishing, 2014: 740-755.
- [23] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]//Computer Vision-ECCV 2018. Cham, Germany: Springer International Publishing, 2018: 3-19.
- [24] HOWARD A G, ZHU Menglong, CHEN Bo, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2022-03-01]. <http://arxiv.org/abs/1704.04861>.
- [25] SANDLER M, HOWARD A, ZHU Menglong, et al. MobileNetV2: inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 4510-4520.

(编辑 陶晴 刘杨)