DOI:10.19651/j.cnki.emt.2211842

基于多阶段交叉信息融合的多光谱行人检测*

孙 昆[1] 武 一[1,2] 牛雅睿[1] 卢 昊[1] 赵 普[1]
(1.河北工业大学电子信息工程学院 天津 300401; 2.河北工业大学电子与通信工程国家级实验教学示范中心 天津 300401)

摘 要: 针对多光谱行人检测中双模态特征融合不充分、特征融合质量低的问题,提出一种基于多阶段交叉信息融合的多光谱行人检测算法。算法首先通过双流骨干网络分别对可见光图像和红外图像进行特征提取;设计交叉信息融合模块并多阶段嵌入双流骨干网络中引导双模态特征融合,实现双模态特征信息的充分融合;引入条件卷积对融合后的特征信息进行动态处理,改善融合信息的质量,最终提升算法的检测性能。实验结果表明,算法的漏检率仅为10.41%,较原算法降低了10%,显著提升了算法的检测性能。

关键词: 多光谱行人检测;模态融合;CondConv
中图分类号: TP391.41 **文献标识码:** A **国家标准学科分类代码:** 510.4

Multi-spectral pedestrian detection based on multi-stage cross information fusion

Sun Kun[1] Wu Yi[1,2] Niu Yarui[1] Lu Hao[1] Zhao Pu[1]
(1. School of Electronics Information Engineering, Hebei University of Technology, Tianjin 300401, China;
2. Electronics and Communication Engineering National Experimental Teaching Demonstration Center, Hebei University of Technology, Tianjin 300401, China)

Abstract: To solve the problems of insufficient bimodal feature fusion and low quality of feature fusion in multispectral pedestrian detection, a multispectral pedestrian detection algorithm based on multistage cross information fusion is proposed. Firstly, the algorithm extracts the features of visible and infrared images through the dual stream backbone network; The cross information fusion module is designed and embedded in the dual stream backbone network in multiple stages to guide the bimodal feature fusion, so as to achieve full fusion of bimodal feature information; Conditional convolution is introduced to dynamically process the fused feature information to improve the quality of the fused information and ultimately improve the detection performance of the algorithm. Experimental results show that the missing rate of the algorithm is only 10.41%, which is 10% lower than the original algorithm, and the detection performance of the algorithm is significantly improved.

Keywords: multispectral pedestrian detection; modal fusion; CondConv

0 引 言

行人检测作为计算机视觉的一个重要研究分支,在自动驾驶、辅助安全驾驶、智能安防等方面发挥着重要作用[1]。随着深度学习的发展,行人检测性能得到了很大的发展。

基于可见光图像的传统行人检测方法是通过手工设计的特征提取算子提取行人特征,借助提取到的行人特征实现行人检测,经典的算法有 Haar 特征算子[2]结合支持向量机(support vector machine, SVM)[3]、梯度特征直方图(histogram of oriented gradients, HOG)[4]结合 SVM、HOG 结合自适应提升方法(adaptive boosting, AdaBoost)[5]等。随着深度学习的发展,目标检测的性能不断提升,基于可见光图像的行人检测借助两阶段行人检测与一阶段行人检测的性能提升取得了有效的发展。但由于检测场景的多样性和复杂性,行人检测性能仍有待提高。例如,在自动驾驶的场景中,会出现雨天、雾天、甚至晚上照明条件差的情况,基于可见光图像的行人检测在这种复杂场景下会出现严重的漏检与误检。

为了将行人检测扩展到全时段场景下,在可见光行人

收稿日期:2022-10-25
* 基金项目:国家自然科学基金(51977059)、河北省自然科学基金(E2020202042)项目资助

检测方法的基础上加入长波红外图像作为补充信息形成基于多光谱图像的行人检测。长波红外传感器通过对人体热辐射的检测,可以有效避免在雾天、雨天、夜间等光线较差的条件下行人检测的影响。Wagner 等[6]首次将深度卷积神经网络用于多光谱行人检测算法,文献将 R-CNN 作为检测模型,研究了前期融合和后期融合两种不同策略对于多光谱行人检测器性能的影响;Liu 等[7]提出一种中期融合方式;Kieu 等[8]通过自适应任务模块,改善网络对热域的自适应性,实现较好的检测性能;Kong 等[9]为了减小多光谱行人检测算法的误检率,引入 BDT 分类器(boosted decision trees classifier)[10]来辅助检测网络;Li 等[11]根据图像的照明亮度对双模态信息进行自适应的融合;Zhang 等[12]针对多光谱数据不对齐问题提出区域对齐模块实现端到端的弱对齐检测算法;Zheng 等[13]以 SSD[14]模型为基本检测框架,VGG16 为特征提取网络,探究了多光谱行人检测器的特征融合结构。但是目前多光谱行人检测中对于红外和可见光两种模态的融合时机大多为单层融合,忽略了剩余特征层的信息;大多的融合策略多是采用通道堆叠和简单相加,忽略了两种模态信息互补依赖关系,因此导致多光谱行人检测检测性能有所不足。

针对上述多光谱行人检测中所存在的问题,本文扩展了 YOLOv5 算法,提出一种多阶段交叉信息融合的多光谱行人检测网络。具体来讲,就是本文重新设计了 YOLOv5 的主干网络,将其设计为双流主干网络,用来分别提取可见光图像和红外图像的特征信息;设计并嵌入交叉信息融合模块,用以促进两种模态信息交互与融合;在多尺度融合阶段引入条件卷积(conditionally parameterized convolutions, CondConv[15])来减少冗余信息对检测结果的影响。最终本文算法在 Kaist[16]数据集上实现了更优的检测性能。

1　YOLOv5 算法

YOLOv5 是 YOLO[17]系列最新的目标检测算法,属于一阶段目标检测算法,它在取得高精度的同时也保持了实时性的检测速度。整体网络结构如图 1 所示,由骨干网络(Backbone)、颈部(Neck)、检测头(Head)三部分组成。其中骨干网络主要对输入图片进行特征信息提取,颈部将网络的浅层空间信息与深层的语义信息进行融合。网络的骨干部分主要由 Focus 模块,C3 模块以及 SPP 模块构成。Focus 模块将输入数据进行切片后拼接操作,最后再进行卷积,相比普通卷积操作,输入图片经过 Focus 模块会保留了更完整的图片下采样信息;C3 模块作为骨干网络中瓶颈层构成部分,借鉴了跨阶段局部网络(cross stage partial network, CSPNet[18])的网络结构,其结构分为两个分支,分别经过多个残差网络层(Bottleneck)和标准卷积层(Conv),最后将两个分支拼接。C3 模块可以较好地减少信息集成过程中重复的梯度信息,增强网络的学习能力;SPP 模块在骨干网络(Backbone)的末端,组合使用多尺度的最大池化层,大幅度提升了感受野,进一步提高了模型的精度。在 Neck 部分,YOLOv5 采用了特征金字塔(feature pyramid networks, FPN[19])结合路径聚合网络(path aggregation network, PAN[20])的结构,FPN 自顶向下,将高层的特征信息通过上采样的方式进行强语义特征的传达,而 PAN 则自底向上传达强的定位特征,二者结合进行特征聚合,得到富含语义信息与空间信息的特征图。然后将充分融合后的信息输入 Head,生成原图像中目标的预测边界框和类别,得到最终的检测结果。

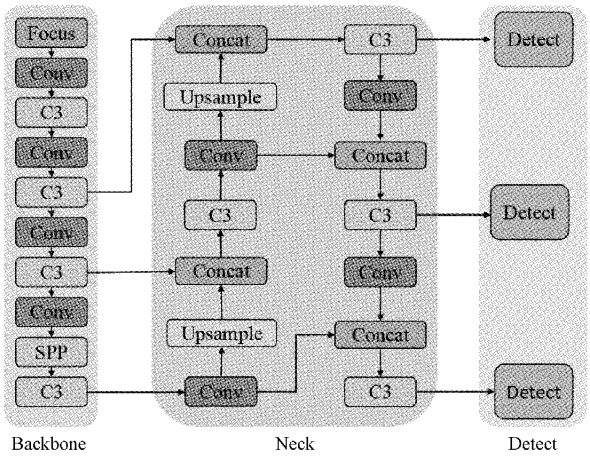

图 1　YOLOv5 网络结构

2　本文算法

2.1　双流主干网络

为了实现多光谱行人检测网络,本文采用双流主干网络分别对可见光图像与红外图像进行特征提取,如图 2 所示,将可见光图像和与之相对应的红外图像送入双流网络中提取特征,主干部分保留 Focus 切片结构,以减少图像特征的信息损失,分别在主根网络的四个阶段进行两种模态信息交叉融合,模态融合的位置分别在 C3 结构之后,并且将后三个阶段的融合结果输出到网络的 Neck 部位进行多尺度特征融合。在双流主干网络的中间层进行多阶段的特征融合,用以解决单层融合受限于选定层的特征,不能充分利用两种模态的特征信息,无法进行充分融合。

2.2　交叉信息融合模块

可见光图像在光照条件好的情况下,可以提供丰富的颜色、纹理等底层信息,红外图像在夜晚等光照条件不足的情况下,包含清晰的轮廓信息。两者的互补性可以有效应对全时段下的行人检测。传统的融合方法包括简单的特征相加和在通道维度进行特征堆叠。由于不同模态的特征空间存在不一致问题,不同模态的特征表现有着较大的差异,如果按照固定权值进行 1∶1 的简单相加,会破坏独立的特征,影响原始的信息表达;而在通道维度进行拼接堆叠实现融合不能充分利用两种模态之间互补关联信息,并且这种融合方法会使得通道维度成倍增加,导致后续的网络计算量骤增。

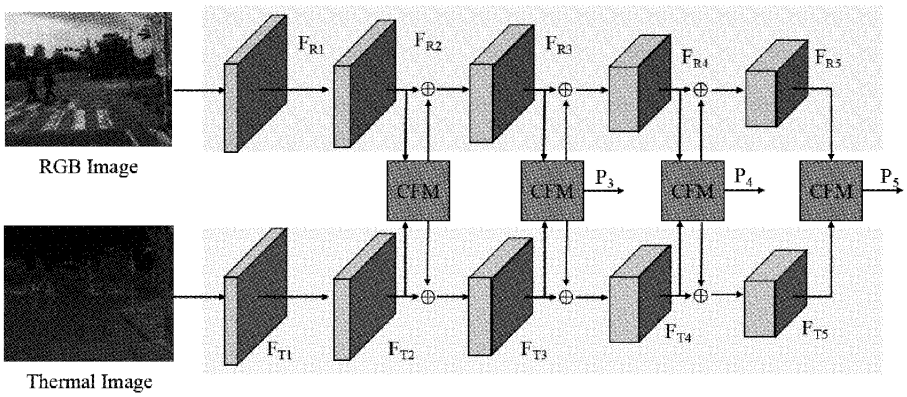

图 2　双流主干网络结构

为了充分利用红外和可见光两种模态特征信息，解决简单融合所引起的问题，实现更优的模态融合，本文提出信息交叉融合模块（cross information fusion module，CFM），通过对共同特征保留原有的特征信息表达，并使用差分模态信息增强双模态所独有的特征信息。如图 3(a)所示，交叉信息融合模块包括两个子模块，分别用来提取双模态间的共同特征与双模态间的差分特征，输出为共同特征与差分特征的堆叠。

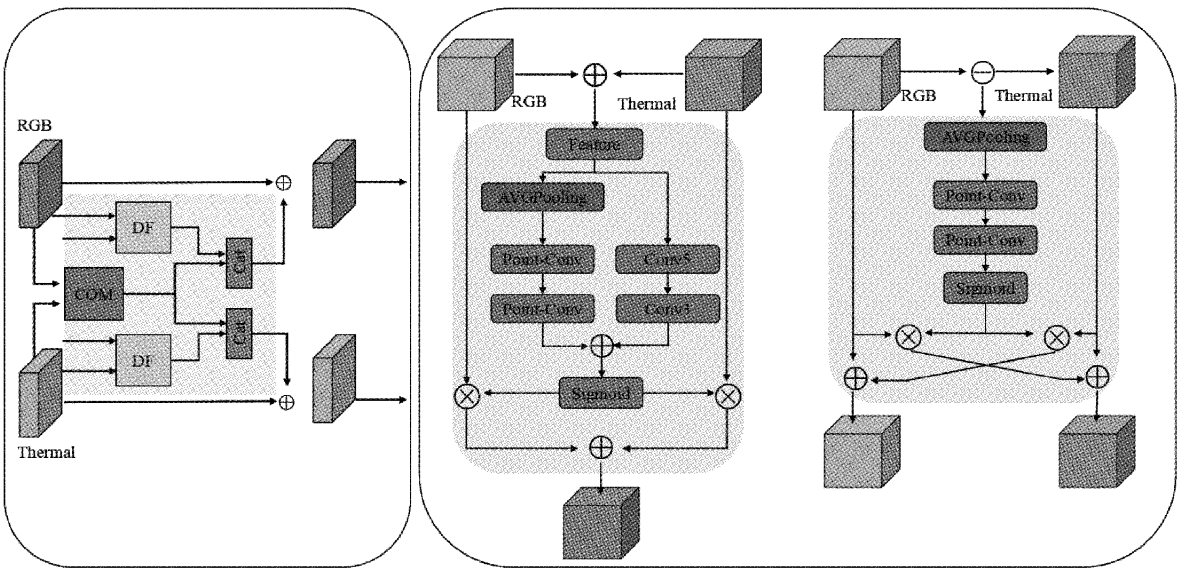

图 3　交叉信息融合模块

共同特征提取模块是基于注意力机制设计，旨在通过网络自适应学习到模态间的互补权值，最终进行加权融合。如图 3(b)左所示，首先将红外特征和可见光特征进行简单相加，相加后的特征图包含两种模态的全部信息，然后分别通过点卷积和自适应全局平均池化对其进行特征编码，点卷积操作相当于对相加后的特征信息进行一种局部的信息抽取，而自适应全局池化则是对特征信息进行一种全局范围内的信息编码，结合这两种操作，我们可以获得局部区域内特征信息，并且可以获取到特征信息在长距离范围的依赖性。将局部信息与全局信息进行相加，通过 sigmoid 函数生成权值，通过权值对可见光特征和红外特征进行互补式加权，实现优势互补的特征信息融合，最终所得到的特征图包含了两种模态所共有的一部分特征。用公式可以表示为：

$$F = F_R + F_T \tag{1}$$

$$W = Sigmoid(g(F) + l(F)) \tag{2}$$

$$F^r = F_R \times W + F_T \times (1 - W) \tag{3}$$

其中，F_R 与 F_T 分别表示可见光特征和红外特征，W 为网络学习到的融合权值，最终通过式(3)实现互补式融合，由于权值是通过网络模型自适应学习到的，因此并不会破坏原有模态的特征信息表达。

由于红外特征和可见光特征都会有自己所独有的特征信息，网络模型可以建模两种模态之间的相互依赖性，通过差分特征信息模块可以挖掘出模态间的特征差异，从而提高网络对于来自另一模态的特征信息的敏感性。如图 3 中右图所示，差分特征信息模块首先获取双流主干网

络的特征信息,然后之间相减获得差分特征,通过适应全局池化对差分特征进行全局编码,并使用两层的 1×1 卷积对其进行通道之间的信息交互,由 sigmoid 函数激活,获得权值向量,将得到的权值向量与其对应的模态特征逐通道相乘获得加权后的差分特征,最后将此特征图最为互补信息与对应的原特征图相加得到新的特征图,作为最终的差分特征进行输出。用公式可以表示为:

$$F_R^r = F_R + \sigma(GAP(F_D)) \otimes F_T \tag{4}$$

$$F_T^r = F_T + \sigma(GAP(F_D)) \otimes F_R \tag{5}$$

通过共同特征提取模块与差分特征信息模块分别提取两种模态的共同特征和不同模态独有的特征,不仅保留了共同特征的信息表达,并且通过差分特征增强了对不同模态的敏感性,结合这两个模块,可以充分利用可见光特征信息和红外特征信息,并且两个子模块并非简单的 1∶1 加权,而是通过网络模型自适应学习所实现的自适应互补融合,可以适应场景的动态变化,实现更优的模态融合。

2.3　条件卷积

多光谱行人检测网络中 Neck 部分的特征融合了两种模态的特征信息,具有复杂性和多变性,需要网络对其有很高的适应性,由于标准卷积的参数是所有数据共享的,因此单一尺度的标准卷积无法很好应的对这种融合信息,为了处理这种信息,需要增加网络的容量,一般会引入更多的卷积层,但是这样会引入大量的网络参数,增加计算量,并且需要对网络的宽度,深度,多分支等进行精心设计,会引入更多的超参。为了使网络能够应对这种难题,并保持网络模型的简洁与实时性,在网络的多尺度融合阶段引入条件卷积。

如图 4 所示为不同卷积的线性组合的结构,图 5 为条件卷积的结构,两者本质相同的,但是条件卷积只需要进行一次卷积,运算更为简便,会大大的减小计算开销。条件卷积通过输入计算卷积核参数,在一层中将多个卷积核参数化,也就是集成多个分支,使得一个卷积核参数包含多套权重,最终加权求和,最终的输出可以表示为:

$$Output(x) = \sigma((\alpha_1 \cdot W_1 + \cdots + \alpha_n \cdot W_n) \times x) \tag{6}$$

其中,x 是来自上一层的输入,σ 为激活函数,α 是可训练的权重数,由路由函数计算得到,路由函数本质上是一个注意力机制,通过将注意力机制作用到卷积权重上,路由函数的计算包括 3 个步骤分别为全局池化、全连接层、Sigmoid 激活函数,用公式可以表示为:

$$\alpha = r(x) = sigmoid(Fc(avgpool(x))) \tag{7}$$

将条件卷积引入到网络模型的颈部加强多尺度特征融合。针对骨干网络输入的融合后的特征信息,动态的生成不同的卷积核系数,自适应的做卷积核的融合,可以充分利用融合后的模态信息,从而为决策层提供更有支持的特征信息,实现更有的检测性能。

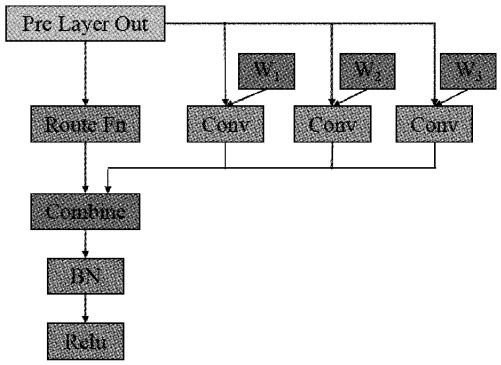

图 4　多组卷积线性组合结构

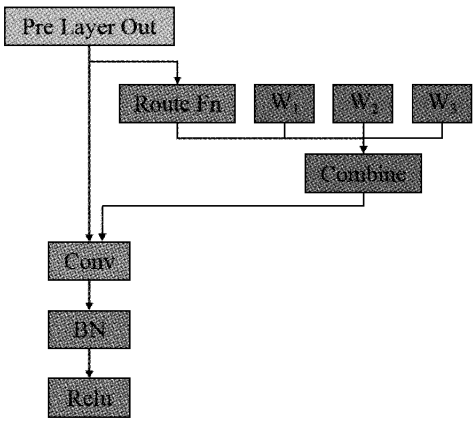

图 5　条件卷积结构

3　实验结果及分析

3.1　实验数据集及评价标准

本文的实验是在 Kaist 多光谱行人检测数据集上进行的。Kaist 是唯一一个具有对齐良好的可见光/热对的大型数据集,其中包含白天和晚上捕获的视频。数据集采用车辆摄像头采集的多光谱图像对进行配准,包括校园、城镇和公路场景中的昼夜图像。它包含多尺度、遮挡和照明不足等复杂条件下的行人目标。使用文献[21]中改进后的训练标注和文献[22]中的测试标注,每两帧从训练视频中提取样本,排除严重遮挡和小的人体实例(<50 pixel)。最终的训练集包含 7 601 张图像。测试集由每 20 帧采样 2 252 个图像对组成。

为了更好的评价本文算法检测性能的优劣,采用采用多光谱行人检测领域通用的对数平均漏检率(log-average miss rate, MR)作为定量评价标准,MR 的值越小,说明算法的检测性能更优,并且分别评价白天时段的 MR 与夜晚时段的 MR。

3.2　实验环境及配置

本文实验需要较好的硬件配置以及 GPU 加速运算。其中模型的搭建、训练和结果的测试均在 Pytorch 框架下完成,实验所需运行环境具体如表 1。在模型训练中,使用

SGD 作为优化器，并且动量 momentum 设为 0.9，权重衰减系数设置为 0.000 5，在所有模型的训练中，对输入图片的尺寸进行调整，统一设置其大小为 640×640，batch-size 大小设置为 32，epoch 大小设置为 100，并且学习率衰减采用余弦退火方式，使用了原 YOLOv5 网络的权值文件作为预训练权值。

表 1　模态对比实验　　%

方法	MR	MR_day	MR_night
RGB	31.65	26.48	44.83
Thermal	24.41	30.85	10.94
RGB+Thermal	10.41	11.97	7.26

3.3　消融实验

为了证明本文算法使用双模态的优越性以及各模块的有效性，设计了双模态与单模态之间的对比实验和各模块之间的消融实验。

为了显示双模态行人检测的优越性和鲁棒性，本文分别使用 YOLOv5 算法实现基于 RGB 的行人检测和基于红外图像的行人检测，并且与使用双模态的本文算法进行对比。如表 1 所示基于 RGB 图像的行人检测在夜晚的行人平均漏检率为 44.83%，算法在夜晚表现很差；而基于红外图像的行人检测在白天的表现比较差。由于 RGB 图受光照条件的影响比较大，在夜晚时段，RGB 图像并无可用的行人特征；而红外图像由于其热成像原理，在白天时段背景物体也有一定的热量，因此图像的噪声信息比较多。本文算法基于双模态信息可以同时利用 RGB 图像与红外图像的优势，得到全时段且检测性能更优的行人检测算法。

为了验证本文算法融合位置的有效性，设计相应的对比实验来验证融合位置的有效性。如表 2 所示，当只在 p5 位置进行单层的交叉信息融合时，算法的漏检率略微上升，而逐步添加 p4、p3 位置进行交叉信息的融合时，算法的漏检率整体变化不大，但是在此基础上在 p2 位置添加交叉信息融合，整体的漏检率都有明显的降低。由于本文设计交叉信息融合模块本质是一种差分融合机制，用以增加两种模态之间的相互敏感性，因此为了获得高质量的融合特征，需要多阶段的添加交叉信息融合模块，循序渐进的引导网络融合，最终获得高质量的融合特征，提高网络的检测性能。

表 2　融合位置对比实验

序号	P5	P4	P3	P2	MR/%	MR_day/%	MR_night/%
0	×	×	×	×	12.36	14.16	8.67
1	√	×	×	×	12.82	15.29	8.35
2	√	√	×	×	12.30	14.36	8.58
3	√	√	√	×	12.06	13.55	9.65
4	√	√	√	√	11.23	13.58	6.58

消融实验包括双流骨干网络、交叉信息融合模块以及条件卷积。为了显示本文双流网络比其他算法的处理方式有效，本文选取最优的中期融合方案作为 Baseline 算法，在 YOLOv5 主干网络的第 2 个 C3 位置进行模态融合；为了证明交叉信息融合模块的有效性，在双流网络的基础上在每一个 C3 的位置加入交叉信息融合模块；最后在网络模型的 Neck 结构中引入条件卷积 CondConv。消融实验结果如表 3 所示，在使用双流骨干网络后，检测性能大幅提升，说明多阶段融合方法的有效性；交叉信息融合模块的加入使整体的检测性能都有一个点的提升；条件卷积的加入对白天的检测性能提升较高，虽然夜晚的漏检有微升，但是整体的漏检率下降明显。通过消融实验结果表明本文算法算法设计的模块均有效，且对算法的提升效果十分显著。

表 3　消融实验

序号	Baseline	Backbone	CFM	CondConv	MR/%	MR_day/%	MR_night/%
0	√	×	×	×	21.76	27.33	9.86
1	√	√	×	×	12.36	14.16	8.67
2	√	√	√	×	11.23	13.58	6.58
3	√	√	√	√	10.41	11.97	7.26

3.4　对比实验

本文算法与其他先进算法的对比如表 4 所示，本文算法在全天、白天、夜晚分别取得了 10.41%、11.97%、7.62%的漏检率。与 KAIST 数据集所给的 Baseline 算法相比，提升了 37%；对比 Halfway Fusion 与 Late Fusion 的单层融合算法检测性能大幅领先；对比 TC-Det 这种使用 RGB 模态辅助红外模态的算法，有 17%的性能提升；在与 AR-CNN 算法相比时，本文算法在没有引入额外数据和标签的情况下，取得了与 AR-CNN 相当的成绩。因此对比实验可以充分说明本文算法的有效性和先进性。并且其他主流算法为了追求检测精度，基于两阶段目标检测算法实现多光谱行人检测，使得算法参数量过大，无法达到实时性。而本文算法通过精准测算得到参数量为 13.87 M，最终得到的模型大小仅为 37.8 MB，使得算法在边缘终端的部署成为可能，并且算法的 FPS 可以达到 33.20，完全达到实时检测的要求。因此本文算法在检测精度与检测速度方面均达到先进水平。

表 4　对比实验　　%

方法	MR	MR_day	MR_night
ACF+C+T[16]	47.32	42.47	56.17
Late Fusion[6]	43.80	46.15	37.00
Half Fusion[7]	25.75	24.88	26.59
TC-Det[8]	27.11	34.81	10.31
RPN+BF[9]	18.29	19.57	16.27
LAF[10]	15.73	14.55	18.26
AR-CNN[11]	10.43	11.34	8.85
本文算法	10.41	11.97	7.62

3.5　定性分析

本文算法的检测结果如图 6 和 7 所示,其中图 6 为本文算法在白天时段的检测结果,图 7 为本文算法在夜晚时段的检测结果,第 1 列为原图,第 2 列与第 3 列分别为本文算法在 RGB 图像和红外图像上的检测结果。由检测结果图可以看出,在背景复杂或者光照条件差的情况下,RGB 图像中可用的行人特征很少、目标几乎不可见,但借助本文算法可以准确检测出行人区域。可以定性得出在面对行人的遮挡、小目标行人、复杂背景、夜晚等场景时,本文算法具有优秀的检测性能与鲁棒性。

图 6　白天时段检测结果图

图 7　夜晚时段检测结果图

4　结　　论

针对目前多光谱行人检测算法中的模态融合不充分、融合特征质量差的问题，本文提出基于多阶段交叉信息融合的多光谱行人检测算法。实验结果表明，本文算法提出的交叉信息融合模块可以有效地引导网络实现模态融合，充分利用不同模态的特征信息，提高了算法的鲁棒性，取得了比较优秀的检测性能。但是目前的算法仍限制于驾驶场景，无法满足实际应用中多场景问题，为了提高算法对各种行人检测场景的适应性，还需要对算法进一步优化。并且在后续工作中，需要进一步探索加快模型的运行速度和提升检测性能的结构和方法。

参考文献

[1] 于波，刘畅. 基于改进 SSD 算法的行人检测方法[J]. 电子测量技术，2021，44(12)：24-28.

[2] OREN M，PAPAGEORGIOU C，SINHA P，et al. Pedestrian detection using wavelet templates[C]. Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition，Los Alamitos：IEEE Press，1997：193-199.

[3] CORTES C，VAPNIK V. Support-vector networks[J]. Machine Learning，1995，20(3)：273-297.

[4] DALAL N，TRIGGS B. Histograms of oriented gradients for human detection[C]. Proc of the IEEE Conference on Computer Vision and Pattern Recognition，Los Alamitos：IEEE Press，2005：886-893.

[5] VIOLA P，JONES M. Rapid object detection using a boosted cascade of simple features[C]. Proc of the IEEE Conference on Computer Vision and Pattern Recognition，Los Alamitos：IEEE Press，2001：511-518.

[6] WAGNER J，FISCHER V，HERMAN M，et al. Multispectral pedestrian detection using deep fusion convolutional neural networks[C]. Proc of the 24th European Symposium on Artificial Neural Networks，Computational Intelligence and Machine Learning，Belgium：DBLP，2016：509-514.

[7] LIU J J，ZHANG S T，WANG S，et al. Multispectral deep neural networks for pedestrian detection[C]. Proc of the British Machine Vision Conference，Los Alamitos：IEEE Press，2016：731-733.

[8] KIEU M，BAGDANOV D，BERTINI M，et al. Task-conditioned domain adaptation for pedestrian detection in thermal imagery[C]. Proc of the 16th European Conference on Conputer Vision. Glasgow，UK：Springer，2020. 1-17.

[9] KONIG D，ADAM M，JARVERS C，et al. Fully convolutional region proposal networks for multispectral person detection[C]. Proc of IEEE Conference on Computer Vision and Pattern Recognition Workshops，Piscataway：IEEE Press，2017：243-250.

[10] 郭景峰，米浦波，刘国华. 决策树算法的并行性研究[J]. 计算机工程，2002(8)：77-78.

[11] LI C Y，SONG D，TONG R F，et al. Illumination-aware faster R-CNN for robust multispectral pedestrian detection[J]. Pattern Recognition，2019，85：161-171.

[12] ZHANG L，ZHU X Y，CHEN X Y，et al. Weakly aligned cross-modal learning for multispectral pedestrian detection[C]. Proc of IEEE International Conference on Computer Vision，Piscataway：IEEE Press，2019：5126-5136.

[13] ZHENG Y，IZZAT I H，ZIAEE S. GFD-SSD：Gated fusion double SSD for multispectral pedestrian detection[J]. ArXiv Preprint，2019，ArXiv：1903.06999.

[14] LIU W，ANGUELOV D，ERHAN D，et al. SSD：Single shot multibox detector[C]. Proc of the 14th European Conference on Computer Vision，Amsterdam：Springer，2016：21-37.

[15] YANG B，BENDER G，LE Q V，et al. Condconv：Conditionally parameterized convolutions for efficient inference[J]. Advances in Neural Information Processing Systems，2019，DOI：10.48550/arXiv.1904.04971.

[16] HWANG S，PARK J，KIM N，et al. Multispectral pedestrian detection：Benchmark dataset and baseline[C]. Proc of IEEE Conference on Computer Vision and Pattern Recognition，Boston：IEEE，2015：1037-1045.

[17] REDMON J，DIVVALA S，GIRSHICK R，et al. You only look once：Unified，real-time object detection[C]. Proc of IEEE Conference on Computer Vision and Pattern Recognition，Piscataway：IEEE Press，2016：779-788.

[18] WANG Y，MARK L Y，WU Y，et al. CSPNet：A new backbone that can enhance learning capability of CNN[C]. Proc of the IEEE Conference on Computer Vision and Pattern Recognition，New York：IEEE Press，2020：1571-1580.

[19] LIN T，DOLLAR P，GIRSHICK R，et al. Feature pyramid networks for object detection[C]. Proc of IEEE Conference on Computer Vision and Pattern

Recognition, New York: IEEE, 2017: 936-944.

[20] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]. Proc of the IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE Press, 2018: 8759-8768.

[21] GHOSE D, DESAI S M, BHATTACHARYA S, et al. Pedestrian detection in thermal images using saliency maps[C]. Proc of IEEE Conference on Computer Vision and Pattern Recognition Workshops. Washington, USA: IEEE, 2019: 434-443.

[22] BAEK J, HONG S, KIM J, et al. Efficient pedestrian detection at nighttime using a thermal camera[J]. Sensors, 2017, 17(8): 1850.

作者简介

孙昆,硕士研究生,主要研究方向为智能控制系统研究与应用。

E-mail:skypai1011@163.com

武一(通信作者),博士,教授,主要研究方向为智能控制系统研究与应用。

E-mail:wuyihbgydx@163.com

牛雅睿,硕士研究生,主要研究方向为计算机视觉。

E-mail:nyr98446@163.com

卢昊,硕士研究生,主要研究方向为无人系统智能感知。

E-mail:lh0102251@outlook.com

赵普,硕士研究生,主要研究方向为计算机视觉。

E-mail:zp_dling@163.com