DOI:10.19651/j.cnki.emt.2417030

基于阶段特征融合的图像融合行人检测*

葛荣泽[1] 武 一[1,2]
(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: 目前可见光与红外图像融合行人检测算法中存在特征不平衡与特征融合不充分等问题。针对上述问题,提出一种分阶段特征融合可见光-红外图像的行人检测网络 MIFNet。构建的双流网络同时处理可见光与红外输入;设计模态间信息融合模块,改变网络的结构减少特征不平衡造成的影响,提取-注入结构在特征提取的过程中自动学习如何提取多模态全局信息并将其有效地注入可见光与红外特征中,提升网络鲁棒性与特征融合效果。设计并嵌入特征增强融合模块,增强两种模态的独特信息,进一步提升特征融合效果。实验结果表明,算法漏检率仅为 9.74%,与基线算法相比降低了 6%,有效的提升了算法的检测性能。

关键词: 行人检测;双流网络;特征融合
中图分类号: TP391.41;TN919.81 **文献标识码**: A **国家标准学科分类代码**: 510.4

Image fusion pedestrian detection based on stage feature fusion

Ge Rongze[1] Wu Yi[1,2]
(1. School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, China; 2. Electronics and Communication Engineer National Experimental Teaching Demonstration Center, Hebei University of Technology, Tianjin 300401, China)

Abstract: There are problems such as feature imbalance and insufficient feature fusion in the visible and infrared image fusion pedestrian detection algorithm. To address the above problems, we propose a multispectral pedestrian detection network MIFNet with phased feature fusion, a dual-stream network that handles both visible and infrared inputs, an intermodal information fusion module that changes the structure of the network to reduce the impact of feature imbalance, and an extraction-injection structure that automatically learns how to extract multimodal global information during the process of feature extraction and injects it into the visible and infrared features efficiently, which improves the robustness of the network and feature fusion effect. The feature enhancement fusion module is designed and embedded to enhance the unique information of the two modalities to further improve the feature fusion effect. The experimental results show that the leakage rate of the algorithm is only 9.74%, which is 6% lower than that of the baseline algorithm, effectively improving the detection performance of the algorithm.

Keywords: pedestrian detection;two-stream networks;feature fusion

0 引 言

随着行人检测技术不断发展,其在智能驾驶辅助等多方面得到了广泛的应用,行人检测的准确性的重要程度正逐步提升。

主流的行人检测方式基于可见光图像(red green blue, RGB)进行检测[1]。Cai 等[2]基于二阶段算法 Faster R-CNN 提出了 Cascade R-CNN 行人检测算法,采用交并比阈值递增的方式,一定程度上改善了正负样本不平衡的问题,Liu 等[3]受 Cascade R-CNN 算法的启发,提出了 ALFNet,ALFNet 摒弃了候选框生成阶段,通过多步预测、交并比阈值递增方法减少正负样本不平衡影响,在保证检测速度的情况下提高了检测精度。YOLO 系列算法[4]采用预设锚框方案,正负样本匹配方案从跨网格匹配发展到基于锚点的分类与回归的分数加权匹配,提高了算法的准确性。但在某些检测情况中,行人检测效果不佳现象的主要原因并不在于正负样本不平衡。在夜晚照明条件差、眩光严重的复杂现实环境下,传统的行人检测算法会产生比较明显的漏检与误检现象。

为了进一步地提升行人检测算法在实际复杂环境下的

收稿日期:2024-10-03
* 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

鲁棒性，在输入可见光图像的基础上增添长波红外图像(thermal，T)一同作为输入，形成基于多光谱图像的行人检测。红外光谱图像的信息获取不依赖外部光源，两种互补信息的引入可以有效减少光线条件差对行人检测产生的影响。Li 等[5]提出的 MSDS-RCNN 算法通过给网络添加分割算法来辅助检测算法在训练阶段进行学习，提高了检测性能，同时提供了清洗后的 Kaist 数据集以供研究，虽然加入的额外网络给算法训练增加了负担，但引导网络注意力的思想非常具有启发性。Soonmin 等[6]提出了基于聚合通道特征的行人检测算法，引入了门控模块动态地调整特征，但这种动态调整仅局限于特征通道，两种模态的空间特征并没有得到很好的融合。Song 等[7]收集并讨论了早期融合、中期融合与后期融合对可见光-红外融合行人检测上的效果，其中，中期融合最能发挥多模态协同效果的结果，根据其提出的结论，本文算法也将使用中期融合的方式。Li 等[8]提出的 IAF R-CNN 网络以 Faster R-CNN 构建了双流网络，并使用额外的光照检测网络辅助双流网络进行特征信息融合，引入的光照检测网络虽然进一步地提高了检测算法的融合性能，但模型训练需要有光照信息标注的数据，这增加了数据获取和模型训练的难度。魏明军等[9]提出的 CALNet 通过利用跨层级的特征信息，提升了模型的性能，算法中对阶段特征的利用仅局限于高级语义特征，忽略了包含丰富位置信息和纹理信息的低级特征。

综上，大部分算法在特征提取后进行单次融合并且通过简单的相加或拼接得到融合的结果，单次融合能否有效地融合特征？如果不能，怎样充分地融合提取到的特征？此外，多光谱图像融合算法还存在其他的问题：文献[10]提出双流网络存在训练收敛不平衡的现象，怎么控制算法的两个分支在训练过程中同步收敛？本文针对上述的两个问题，通过扩展一阶段目标检测算法 YOLOv8，提出一种阶段特征融合的多光谱行人检测算法 MIFNet。具体来讲，本文设计 3 部分的创新：将 YOLOv8 改造为双流网络，同时处理可见光图像与红外图像；设计模态间信息融合模块，提取多模态全局信息，使用自适应的注意力机制促进模态间信息的交互，模块采用阶段式分布，实现多阶段的特征融合，帮助模态平衡；提出特征增强融合模块代替模态特征相加操作，充分融合特征。本文算法 Kaist 数据集上实现了优秀的检测性能。

1　本文算法

1.1　YOLOv8 阶段特征融合双流主干网络

通过对模态不平衡现象进行分析，发现多模态模型的优化进程会被性能较好的模态主导，使另一个模态的特征提取分支不能够充分学习。这里假设使用 $\varphi^{RGB}(\theta^{RGB},*)$ 和 $\varphi^{T}(\theta^{T},*)$ 两个编码器对特征进行提取，其中 θ^{RGB} 和 θ^{T} 是编码器的参数，可以得到送入多尺度融合的特征的计算方式：

$$x_i=\varphi^{RGB}(\theta^{RGB},x_i^{RGB})+\varphi^{T}(\theta^{T},x_i^{T}) \tag{1}$$

算法之后的操作使用 $f(.)$ 代指，那么网络的输出就可以写作 $f(x_i)$。将 c 类的输出记为 $f(x_i)_c$，则模型的交叉熵损失 L 可以写为：

$$L=-\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{f(x_i)_c}}{\sum_{k=1}^{M}e^{f(x_i)_k}} \tag{2}$$

通过梯度下降优化法，可见光特征提取分支的编码器的参数更新公式为：

$$\theta_{t+1}^{RGB}=\theta_t^{RGB}-\eta\nabla_{\theta^{RGB}}L(\theta_t^{RGB})=\theta_t^{RGB}-\eta\frac{1}{N}\sum_{i=1}^{N}\frac{\partial L}{\partial f(x_i)}\frac{\partial(\varphi_t^{RGB}(\theta^{RGB},x_i^{RGB}))}{\partial\theta_t^{RGB}} \tag{3}$$

红外特征提取分支的公式同理。其中 η 是学习率。

根据式(3)可以发现，可见光特征提取分支的参数优化与红外特征提取分支除了 $\frac{\partial L}{\partial f(x_i)}$，内部几乎没有相关性。因此单模态编码器几乎无法根据彼此的反馈进行调整，当其中一个模态性能更好时，对 $\frac{\partial L}{\partial f(x_i)}$ 的贡献会更大，而 $\frac{\partial L}{\partial f(x_i)}$ 部分损失的降低会导致另一个尚未收敛的模态特征提取分支参数更新受阻，造成双流分支收敛不平衡的现象。

针对这种现象提出了使用 YOLOv8 改造的基于分阶段特征融合双流骨干网络，如图 1 所示。首先，将两幅图像输入双分支网络，两个分支均由 YOLOv8 的特征提取部分构成。其次，在双流特征提取过程中，分别在网络的四次特征提取后的位置嵌入模态间信息融合模块。最后，将后 3 个阶段得到的特征信息送入特征增强融合模块进行增强，并将结果输出到网络的颈部网络中进行多尺度特征融合。阶段嵌入的融合模块使网络的两个分支在特征提取的过程中产生了相关性，计算损失与梯度反向传递的过程中可以相互影响，能够有效改善两分支间在训练时产生的收敛速度不平衡的问题。

1.2　模态间信息融合模块(intermodal information fusion module)

已有的工作为了提高多模态信息的融合效果不断探索复杂的融合方式，但并未充分考虑利用提取到的全局模态特征信息，将融合的特征信息简单切分后加回原模态特征，这样得到的特征中一部分是原模态的信息，另一部分是另一种分布的全局模态信息，这可能会对后续的特征提取产生影响。

于是提出了提取-注入结构的多光谱特征融合模块，称为模态间信息融合模块。模块接受可见光与红外两种模态的信息，提取模态间的全局信息，将提取到的全局信息自适应地补充到两种模态特征当中。该模块分为两部分，分别进行模态间全局特征信息的提取与全局特征信息的重新注入。

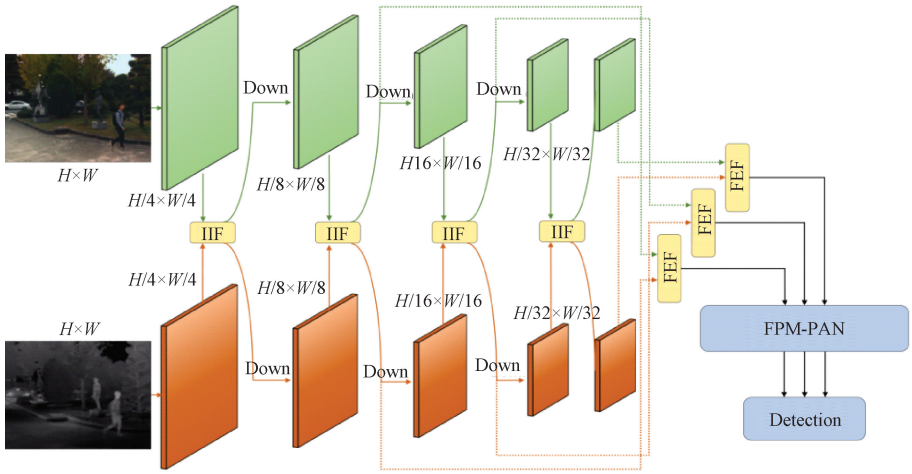

图 1 模型整体结构

Fig. 1 Overall structure of the model

模块如图 2 所示,提取模态间全局特征信息的部分包括拼接操作、多层重参数化卷积。具体来说,先将输入的多模态特征 F_{rgb}、F_t 沿通道维度进行拼接,拼接后的特征图 F_{align}(channel=sum(C_rgb, C_t))包含两种模态的所有特征信息,之后 RepBlock 卷积块对 F_{align} 提取模态间全局特征信息 F_{fea}(channel=C_mid),RepBlock 包含的结构重参数化卷积可以丰富特征融合过程的梯度流,使两种特征信息充分融合,中间通道 C_mid 是一个可调节的参数,以适应模型不同位置的通道尺寸。以上操作用公式表示为:

$$F_{fea} = RepBlock(cat(F_{RGB}, F_T)) \tag{4}$$

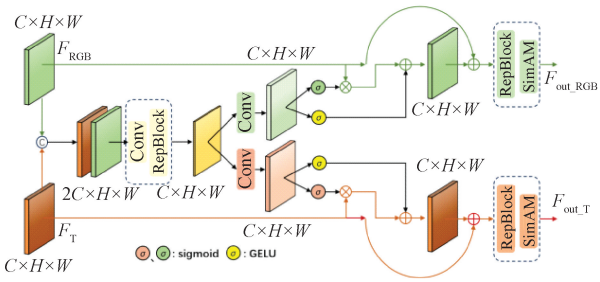

图 2 模态间信息融合模块

Fig. 2 Intermodal information fusion module

为了将模态间特征信息更有效地注入 F_{rgb}、F_t 两个模态特征,本文使用两个相互独立的 1×1 卷积对 F_{fea} 进行独立的特征编码,生成对应 F_{rgb}、F_t 的全局信息特征 F_{global_rgb}(channel=C_rgb)、F_{global_ir}(channel=C_t),这一操作允许网络自主学习应该将全局信息中的哪些部分增强到红外和可将光模态特征中。然后将 F_{global_i} 经过 sigmoid 函数生成权值用以增强模态特征 F_i,增强后的 F_i 与使用激活函数激活的 F_{global_i} 相加得到融合特征 F'_{out_rgb}、F'_{out_t},原始的通道特征与经过激活函数处理的信息进行融合,全局特征能够更平滑地注入回原模态,可以得到更加丰富和有效的特征表示。融合之后,使用 RepBlock 卷积块与 SimAM 注意力模块来对融合特征进一步的提取和信息融合,得到输出 F_{out_i}。以上步骤用公式表示为:

$$\begin{aligned} F_{global\_i} &= Conv_i(F_{fea}) \\ F'_{out\_i} &= F_i + F_i * sigmoid(F_{global\_i}) + \sigma(F_{global\_i}) \\ F_{out\_i} &= SimAM(RepBlock(F'_{out\_i})) \end{aligned} \tag{5}$$

模态间全局信息提取与全局特征信息注入两部分的配合,可以自主地识别到另一模态中的有效信息并在不破坏原特征的情况下实现高效的信息融合,能够更好的应对多变的检测环境,提升模态融合效果。

1.3 特征增强融合模块(feature enhancement fusion module)

图像融合行人检测网络中的颈部网络需要一张拥有可见光与红外两种模态特征信息的特征图作为输入,大部分算法的处理方法是简单的将两张特征图逐元素相加,这样做虽然可以保证得到的特征图包含两种模态的特征信息,但这种简单的相加并不能够很好的融合两种模态,融合两种模态的模块除了做到无损的融合两种模态信息之外还应当能够自主地发现两种模态的差异特征信息并对应地在融合特征图中进行增强,引导后续的检测网络利用各模态独特的特征信息。

为解决上述问题,设计了特征增强融合模块,该模块旨在无损且不增加额外参数的前提下增强可见光与红外特征图中的差异特征,提高特征融合的效果。

如图 3 所示,特征增强融合模块首先将接收到的可见光与红外模态特征互相做差,差异部分反映了差异特征的位置信息,然后对两个差异特征进行全局平均池化并应用范围为−1~1 的 Tanh 激活函数将其处理为差异信息通道权向量,之后分别使用差异通道权向量对可见光、红外原特征通道进行加权处理,这样就得到了两种模态各自的独特信息,最终将 4 种特征信息进行相加,就得到了增强后的融合信息。用公式表示为:

$$F_{out}=F_{RGB}+F_{RGB}(Tanh(AvgPooling(F_{RGB}-F_T)))+F_T+F_T*(Tanh(AvgPooling(F_T-F_{RGB}))) \quad (6)$$

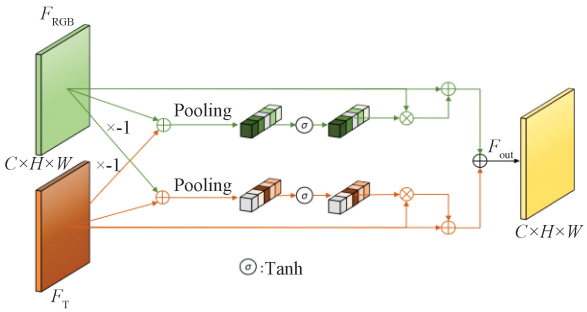

图 3　特征增强融合模块

Fig. 3　Feature enhancement fusion module

整个流程并不涉及任何需要网络学习的参数，因此不会给网络训练增加负担，并且该模块在没有丢失任何原始特征信息的情况下实现了对差异特征信息的增强，使颈部网络更好地区分并利用两种模态提供的差异特征，为检测头提供更有效的特征信息，能够有效提升网络的检测性能。

2　实验结果及分析

2.1　实验数据集及评价标准

本文所选用的实验数据集是 Kaist 多光谱行人检测数据集，数据集是包含配对可见光图像、红外图像的大型行人检测数据集，数据均由配准后的车载摄像头采集标注获得，包含了校区、居住区及郊区等多种现实交通场景中的白天与黑夜，拥有眩光、低光照、遮挡等多种复杂环境下的行人目标。原始的 Kaist 数据集包含大量无意义的重复帧，本文所使用的是由浙江大学 Li 等清洗后的 Kaist 数据集，训练集每 3 张取一张，去掉不包含行人、像素小于 50 的行人及被严重遮挡的行人图像，测试集每 20 张取一张，保留负样本。最终得到 7 601 对训练集图片，2 252 对测试集图片。本文随机选取训练集的 10%，即 760 对图片作为验证集。

对数平均漏检率(log average miss-rate，LAMR)作为一种能够用来量化漏检率与每图像误检曲线(miss rate versus false positives per image，MR-FPPI)的指标，能够综合地评判检测器的性能，该指标越小代表检测器性能越高，故本文使用 LAMR 作为定量评价指标并对整体、白天、黑夜 3 种场景进行评估。

2.2　实验环境及配置

本文使用 Pytorch 框架搭建模型进行训练及测试，使用 SGD 优化器，动量设置为 0.937，权重衰减系数设置为 0.000 5，所有输入的图像被调整至 640×640 进行训练，Batch-size 设置为 16，epoch 设置为 110，学习率设置为 0.01，以余弦退火方式进行衰减。实验所用运行环境如表 1 所示。

表 1　实验环境及配置

Table 1　Experimental environment and configuration

名称	实验参数
GPU	NVIDIA GeForce RTX A5000
CPU	Intel(R) Xeon(R) Platinum 8350C CPU@2.60 GHz
操作系统	Ubuntu 20.04.4 LTS
Python	Python 3.9.18
Pytorch	1.12.1+cu116

2.3　消融实验

为证明本文所提出算法结构及模块的有效性，设计了如表 2 所示的各模块的消融实验。

表 2　消融实验定量结果

Table 2　Quantitative results of ablation experiments　%

序号	Baseline	Backbone	IIF	FEF	LAMR(all)	LAMR(day)	LAMR(night)
1	√	×	×	×	15.52	17.44	12.28
2	√	√	×	×	13.02	14.33	10.54
3	√	√	√	×	10.69	12.37	8.03
4	√	√	√	√	9.74	10.91	7.7

在使用改进后的主干网络结构后，解决了训练过程中收敛不平衡的现象，使算法的全天平均漏检率由 15.52%降为 13.04%，黑夜与白天的平均对数漏检率均有大幅降低。在添加模态间信息融合模块对两种模态进行信息融合增强后，全天 LAMR 进一步降低至 10.69%，充分说明了模态间信息融合模块有效地提取了全局模态特征信息，并有效地注回两种模态中，对可见光-红外融合的帮助巨大。在增加特征增强融合模块后，LAMR(all)进一步降低了 0.95%，且白天和夜晚的场景中检测性能均有提高，可以说明特征增强融合模块对两种不同模态信息的融合比其他融合方式的融合更充分，能够显著地提高算法的检测性能。

2.4　对比实验

为证明本文算法具有优越性，选取可见光-红外融合行人检测领域内其他优秀的算法进行对比实验，如表 3 所示，对比单模态网络，本文算法的效果是其 3.5～4.9 倍，充

分展示出了多模态算法的优势。本文提出的算法相较于 Kaist 作者提出的算法提升了 37.5%;相较 Faster R-CNN-F、IAF-R-CNN,本文算法检测性能提升明显;得益于全阶段融合结构,本文算法对比单层融合算法 CIAN、IT-MN、RFA,全天平均误检率降低 4%以上,说明全阶段融合结构明显优于单层融合机构;和使用 transformer 注意力机制进行融合的算法 CFT、MASNet 相比,本文算法能够取得 1.13%~3.8%的提升,充分说明本文所使用的类注意力融合机制在融合表现上优于 transformer;和二阶段检测算法 IAF-R-CNN、MSDS-R-CNN 以及 AR-CNN 相比,本文算法作为一阶段检测算法,在没有引入额外的光照信息的情况下优于 IAF-R-CNN、MSDS-R-CNN,检测效果与 AR-CNN 相当,可以说明本文算法具有自适应学习可见光与红外特征的功能。

表 3 Kaist 数据集定量对比结果

Table 3 Quantitative comparison results of Kaist dataset %

方法	数据	LAMR(all)	LAMR(day)	LAMR(night)
YOLOv8m	RGB	34.01	36.86	27.31
YOLOv8m	Thermal	47.75	55.68	31.02
Faster R-CNN-F[11]	RGB+T	47.00	—	—
IAF-R-CNN	RGB+T	16.22	13.94	18.28
RFA[12]	RGB+T	14.61	16.78	10.21
IT-MN[13]	RGB+T	14.19	14.30	13.98
CIAN[14]	RGB+T	14.12	14.77	11.13
CFT[15]	RGB+T	13.54	16.39	8.43
MSDS-R-CNN	RGB+T	11.34	10.60	13.73
MSANet[16]	RGB+T	10.87	13.26	6.25
FRFPD[17]	RGB+T	10.79	—	—
FCE-R-CNN[18]	RGB+T	10.62	12.31	6.92
AR-CNN[19]	RGB+T	9.34	9.94	8.38
MIFNet(our)	RGB+T	9.74	10.91	7.7

为证明本文算法的实用性,统计了各算法进行推理的速度,具体如表 4 所示。本文算法对比以推理速度见长的 YOLOv8 算法,本文算法的推理速度是其 3.1 倍左右,但仍保持在实时推理的范围内,而对比众多的多模态网络,本文算法在推理时间上有绝对的优势。由于采用了结构重参数化思想设计模态间信息融合模块,使其在推理过程中可以在保持训练阶段效果的情况下减少计算量,配合无参增强融合模块,本文算法能够保证实时的推理速度、优秀的检测精度,具有部署在嵌入式平台上进行实际应用的潜力。

表 4 推理速度定量对比结果

Table 4 Quantitative comparison results of inference speed

方法	平台	Runtime/ms
YOLOv8m	TITAN X	16
Faster R-CNN-F	MATLAB	2 730
IAF R-CNN	TITAN X	210
CIAN	1080Ti	70
MSDS-R-CNN	TITAN X	220
AR-CNN	1080Ti	120
MSANet	RTX3090	110
CFT	TITAN X	100
MIFNet(本文)	TITAN X	50

2.5 定性分析

图 4 对比了 MIFNet 与 Baseline 算法行人检测的热力图效果,处于画面右侧的行人与背景区分度较低,不易被检测,MIFNet 特征融合的更充分,关注点明显更加集中在行人上。图像经过模态间信息融合模块后的特征可视化图像如图 5 所示,可见光模态特征融合了红外图像中清晰简洁的行人位置信息;红外模态特征融合了可见光图像中行人丰富的纹理细节,证明模块能够利用全局模态信息,有针对性地优化补足两种模态分支的特征。

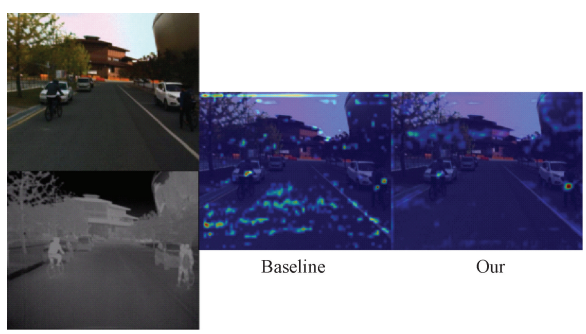

图 4 MIFNet 热力图

Fig. 4 Heat map of MIFNet

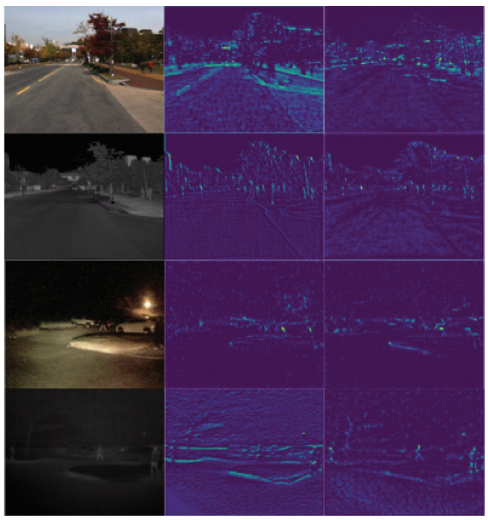

图 5　IIF 模块输入(融合前特征)和输出(融合后特征)的可视化

Fig. 5　Visualization of IIF module inputs (pre-fusion features) and outputs (post-fusion features)

MIFNet 与 Baseline 算法、单模态输入算法的检测结果对比如图 6 所示。从图中可以看出对比单模态输入算法和基线算法，MIFNet 产生的漏检和误检更少，在白天场景和黑夜场景中均能够利用增加的红外模态信息提高检测效果，在白天情况下给出更好地包裹行人目标的检测框，在黑夜状态下检测出更多处于无光照位置的行人目标。由此定性得出，本文算法能够有效的结合两种模态图像的输入信息，实现更加准确、更加鲁棒的行人检测。

3　结　　论

本文针对特征提取双分支网络结构存在的收敛速度不一致问题，采用阶段特征融合的双流网络结构，配合两个设计的特征融合模块，设计出了基于特征阶段融合的 RGB-T 图像融合检测算法 MIFNet。实验表明，本文提出的算法可以有效利用多模态输入信息，使网络保持较高的推理速度、较好的鲁棒性与较好的检测精度。但由于双分支网络结构参数量相较于常规单分支检测算法仍属偏大，在对检测速度有较高要求的场景中仍然略显吃力。在后续工作中要继续探索在保证模型性能的前提下，加快模型检测速度的方法。

图 6　Kaist 数据集可视化对比结果(正确检出：绿色框，漏检：蓝色框，误检：红色框)

Fig. 6　Visual comparison results of Kaist dataset (correctly detected: green box, missed: blue box, false: red box)

参考文献

[1]　梁天添，杨淞淇，钱振明. 基于改进 YOLOv8s 的恶劣天气车辆行人检测方法[J]. 电子测量技术，2024，47(9)：112-119.
LIANG T T, YANG S Q, QIAN ZH M. Improved YOLOv8s-based vehicle and pedestrian detection method for adverse weather conditions[J]. Electronic Measurement Technology, 2024, 47(9): 112-119.

[2]　CAI Z W, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154-6162.

[3]　LIU W, LIAO SH C, HU W D, et al. Learning efficient single-stage pedestrian detectors by asymptotic localization fitting[C]. European Conference on Computer Vision (ECCV), 2018: 618-634.

[4]　REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]. Computer Vision & Pattern Recognition. IEEE, 2016.

[5]　LI CH Y, SONG D, TONG R F, et al. Multispectral pedestrian detection via simultaneous detection and segmentation [J]. ArXiv preprint arXiv: 1808.04818, 2018.

[6]　WOO S, HWANG S, JANG H D, et al. Gated

bidirectional feature pyramid network for accurate one-shot detection[J]. Machine Vision and Applications, 2019, 30: 543-555.

[7] SONG K C, ZHAO Y, HUANG L M, et al. RGB-T image analysis technology and application: A survey[J]. Engineering Applications of Artificial Intelligence, 2023, DOI: 10.1016/j.engappai.2023.105919.

[8] LI CH 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.

[9] 魏明军,魏帅,刘亚志,等.基于跨层级注意力学习的RGB-T显著目标检测[J].郑州大学学报(理学版), 2024, DOI: 10.13705/j.issn.1671-6841.2023163.
WEI M J, WEI SH, LIU Y ZH, et al. RGB-T salient object detection based on cross-level attention learning[J]. Journal of Zhengzhou University (Science Edition), 2024, DOI: 10.13705/j.issn.1671-6841.2023163.

[10] PENG X K, WEI Y K, DENG A D, et al. Balanced multimodal learning via on-the-fly gradient modulation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 8238-8247.

[11] ZHANG L, LIU Z, CHEN X, et al. The cross-modality disparity problem in multispectral pedestrian detection[J]. ArXiv preprint arXiv:1901.02645, 2019.

[12] YANG X P, LI ZH H, LIU Y, et al. Research on pedestrian detection based on multimodal infor-mation fusion [J]. Information Technology and Control, 2023, 52(4): 1045-1057.

[13] ZHUANG Y F, PU Z Y, HU J, et al. Illumination and temperature-aware multispectral networks for edge-computing-enabled pedestrian detection [J]. IEEE Transactions on Network Science and Engineering, 2021, 9(3): 1282-1295.

[14] ZHANG L, LIU ZH Y, ZHANG SH F, et al. Cross-modality interactive attention network for multispectral pedestrian detection [J]. Information Fusion, 2019, 50: 20-29.

[15] FANG Q Y, HAN D P, WANG ZH K. Cross-modality fusion transformer for multispectral object detection [J]. ArXiv preprint arXiv: 2111.00273, 2021.

[16] YOU SH, XIE X D, FENG Y J, et al. Multi-scale aggregation transformers for multispectral object detection[J]. IEEE Signal Processing Letters, 2023, DOI: 10.1109/LSP.2023.3309578.

[17] FENG Y, LUO E B, LU H, et al. Cross-modality feature fusion for night pedestrian detection [J]. Frontiers in Physics, 2024, DOI: 10.3389/fphy.2024.1356248.

[18] NIE L ZH, LU M H, HE ZH W, et al. Multispectral pedestrian detection based on feature complementation and enhancement [J]. IET Intelligent Transport Systems, 2024, DOI: 10.1049/itr2.12562.

[19] ZHANG L, ZHU X Y, CHEN X Y, et al. Weakly aligned cross-modal learning for multispectral pedestrian detection [C]. IEEE/CVF International Conference on Computer Vision, 2019: 5127-5137.

作者简介

葛荣泽,硕士研究生,主要研究方向为计算机视觉。
E-mail:gerongzehbgydx@163.com

武一(通信作者),教授,硕士研究生导师,主要研究方向为智能控制系统研究与应用。
E-mail:wuyihbgydx@163.com