激光与光电子学进展 先进成像

基于光照感知权重融合的多模态行人检测算法

刘珂琪[1]，董绵绵[1*]，郜辉[1]，吕志刚[1]，郭宝亿[2]，庞敏[3]

[1]西安工业大学电子信息工程学院，陕西 西安 710021；

[2]西安工业大学本科生院，陕西 西安 710021；

[3]北京微电子技术研究所，北京 100000

摘要 针对现有利用可见光与红外模态融合的行人目标检测算法在全天候环境下漏检率高的问题，提出一种基于光照感知权重融合的多模态行人目标检测算法。首先，使用引入高效通道注意力(ECA)机制模块的ResNet50作为特征提取网络，分别提取两个模态的特征；其次，对现有光照加权感知融合策略进行改进，通过设计一种新的光照感知加权融合机制获取可见光与红外模态的对应权重，并进行加权融合得到融合特征，从而降低算法的检测漏检率；最后，将从特征网络最后一层提取的多模态特征和生成的融合特征共同送入到检测网络，完成行人目标检测。实验结果表明，所提算法在KAIST数据集下具有良好的检测性能，在全天候下对行人目标的检测漏检率为11.16%。

关键词 多模态图像融合；注意力机制；光照感知权重融合；行人检测

中图分类号 TP391.41 **文献标志码** A **DOI**：10.3788/LOP222528

Multi-Modal Pedestrian Detection Algorithm Based on Illumination Perception Weight Fusion

Liu Keqi[1], Dong Mianmian[1*], Gao Hui [1], Zhigang Lü[1], Guo Baoyi[2], Pang Min[3]

[1]*School of Electronic and Information Engineering, Xi'an Technological University, Xi'an* 710021, *Shaanxi, China*;

[2]*Undergraduate College, Xi'an Technological University, Xi'an* 710021, *Shaanxi, China*;

[3]*Beijing Institute of Microelectronics Technology, Beijing* 100000, *China*

Abstract Existing pedestrian target detection algorithm based on visible light and infrared modal fusion has a high missed detection rate in all-weather environment. In this paper, we propose a novel multi-modal pedestrian target detection algorithm based on illumination perception weight fusion to solve this problem. First, ResNet50, incorporating an efficient channel attention (ECA) mechanism module, was used as a feature extraction network to extract the features of both visible light and infrared modes, respectively. Second, the existing illumination weighted sensing fusion strategy was improved. A new illumination weighted sensing fusion mechanism was designed to attain the corresponding weights of the visible light and infrared modes, and weighted fusion was performed to achieve fusion features to reduce the missed detection rate of the algorithm. Finally, the multi-modal features extracted from the last layer of the feature network and the generated fusion features were fed into the detection network to accomplish the detection of pedestrian targets. Experimental results show that the proposed algorithm has an excellent detection performance on the KAIST dataset, and the missed detection rate for pedestrian targets in all-weather is 11.16%.

Key words multi-modal image fusion; attention mechanism; illumination perception weight fusion; pedestrian detection

1 引 言

行人检测作为目标检测中的重要组成部分，目前已被广泛应用在军事和民用等多个领域[1]。传统的行人检测算法一般是基于可见光单模态或红外单模态提出的。其中可见光模态能够提供图像更丰富的纹理和

收稿日期：2022-09-12；**修回日期**：2022-10-14；**录用日期**：2022-11-08；**网络首发日期**：2022-11-21

基金项目：国家自然科学基金(62171360)、陕西省科技厅重点研发计划(2022GY-110)、西安工业大学校长基金面上培育项目(XGPY200217)、西安市智能兵器重点实验室(2019220514SYS020CG042)、2022年度陕西高校青年创新团队项目

通信作者：*dong_mm@aliyun.com

细节信息，但在光照条件不佳的情况下，算法无法有效地检测到图像中的行人目标，从而导致在夜间环境下的检测性能低下[2]。红外模态根据红外传感器的特性可以获取到目标的热信息，在夜晚光照不足的场景中也可以有效获取到图像中的目标特征信息，但存在目标细节信息获取不充分的问题。因此，根据可见光与红外模态之间的互补性，结合两种模态之间的特征信息，从而获得更全面的图像信息，使图像中的目标特征得到进一步加强。近年来基于可见光与红外的多模态图像融合算法被大量提出，在军事和民用领域均得到了更好的应用[3]。

国内外研究学者对基于图像融合的行人检测已做了不少研究。文献[4]研究了深度特征在不同融合时期对行人目标检测结果的影响，发现对特征提取网络的中间层特征进行融合后能够使可见光与红外特征信息得到有效利用，但存在算法模型运行速度较慢的问题。之后，文献[5]在中间层特征融合的基础上，使用RPN+BDT网络进行特征提取与分类，提高了行人检测的性能，但该方法对行人目标的检测性能还有进一步提升空间。文献[6]虽然采用中间层特征堆叠策略，在后续的检测输入中既用到融合特征，又用到原始两模态的最后一层特征，但在复杂环境下的检测效果不佳。由于上述仅选取某个阶段特征进行融合的方法会造成其他特征层的特征浪费，影响目标的检测识别结果，之后，文献[7]提出了一种基于多模态多尺度的特征融合方式，通过对两种模态的特征在不同尺度下进行融合，以提高对目标的检测性能，但是这种方法在获取融合特征时仍采用直接堆叠策略，无法自适应地应对外界光照环境的变化。文献[8]考虑到特征堆叠融合策略容易造成特征信息冗余的问题，在融合时加入其他特征层信息，使两种模态的特征得到进一步增强，但未考虑到两种模态之间的关系。文献[9]考虑到可见光与红外图像两种模态融合时在不同光照环境下所占权重不同的问题，设计了光照感知模块，用来模拟不同光照情况，虽然在一定程度上提高了模型的适用性，但目标的检测结果极大地依赖于光照感知模块。

针对以上问题，本文利用可见光图像中的光照信息，通过设计一种新的适用性强的光照感知融合模块，提出了基于光照感知权重融合的多模态行人目标检测算法，并在特征提取网络结构中加入注意力机制模块，有效提高了全天候下对行人目标的检测质量。除此之外，本文将特征提取网络最后一层生成的可见光与红外特征送入到检测网络，避免融合机制不稳定的问题，从而更大程度上增强算法的检测性能。

2　多模态图像融合的行人检测算法

2.1　网络结构

由于行人目标检测对算法的实时性和精确度都有较高的要求，考虑到目标检测输入的多源性和检测目标的单一性，提出了基于光照感知权重融合的多模态行人目标检测算法，选取SSD网络框架[10]作为双模态检测网络的基础框架。原始的SSD算法采用VGG16网络作为特征提取网络，但该网络结构简单，网络层数较浅，提取目标的深层次信息时具有一定的局限性。与VGG16网络相比，ResNet50特征提取网络[11]增加了网络的深度，能够提取到更深层次的目标特征信息，提高了模型的检测精度，同时又减少了模型的参数量，避免了在模型训练过程中网络深度增加造成的梯度爆炸或梯度消失的问题，有效提高网络模型的检测性能。因此，本文选择ResNet50作为SSD算法的特征提取网络，网络结构如图1所示。

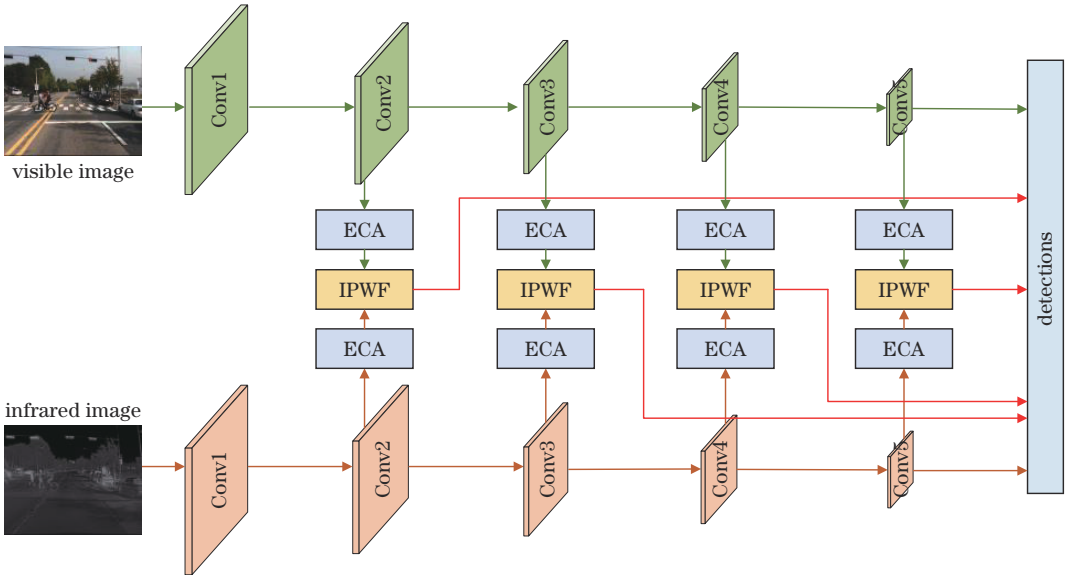

图1　所提网络结构

Fig. 1　Structure of the proposed network

首先，将对应的可见光与红外图像分别送入到ResNet50特征提取网络中提取特征；其次，将提取出的可见光与红外特征分别输入高效通道注意力（ECA）机制模块，提升对特征的表达能力；接着，经过注意力机制的可见光与红外特征传输至光照感知权重融合（IPWF）模块，获取可见光与红外图像的权重，并对对应的特征进行加权融合，生成融合特征；最后，将不同阶段生成的融合特征与特征提取网络最后一层生成的可见光和红外特征共同送入到检测网络中进行行人目标检测。图1中Conv1～Conv5分别表示ResNet50网络中的第1层至第5层卷积层。

2.2　注意力机制模块

注意力机制模块在网络模型中引入可学习权重，使模型更多地关注目标特征区域，从而有效降低噪声和一些冗余信息在双模态特征提取中的影响[12]。目前注意力机制模块主要是以Squeeze and Excitation（SE）注意力机制模块为代表的通道注意力机制模块及以convolutional block attention module（CBAM）[13]为代表的结合通道与空间的注意力机制模块。其中SE通道注意力机制主要采用一种降维操作来控制模型的复杂性，但同时也影响了通道注意力的预测结果。CBAM在通道注意力的基础上增加了空间注意力，使模型的检测精度更高，但精度越高，模型的复杂度就越高，计算量也越大，影响模型的效率。因此，从模型的精度与效率两方面综合考虑，本文选用在SE注意力机制基础上改进的ECA模块[14]，该模块不仅能够有效提升模型的检测性能，并且由于该注意力机制的轻量化设计，仅增加了少量的参数量。

在特征提取网络提取出可见光与红外特征后，给可见光与红外双模态流分别加入ECA机制模块，用来增强关注的特征[15]。ECA机制模块的结构如图2所示。

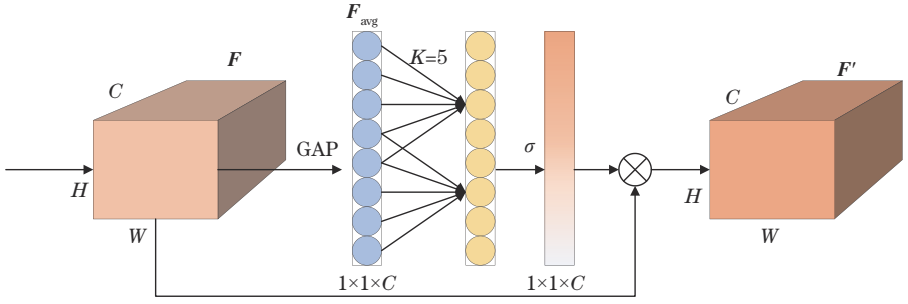

图2　ECA机制模块的结构

Fig. 2　Structure of ECA mechanism module

由于在ECA机制模块中对可见光与红外特征做相同操作，在此，将可见光与红外特征均记为$\boldsymbol{F}$。输入的可见光与红外特征图$\boldsymbol{F}$记为$\boldsymbol{F}\in\mathbb{R}^{C\times H\times W}$。ECA机制首先对$\boldsymbol{F}$进行全局平均池化（GAP），得到特征描述符，其大小为$C\times1\times1$，记为$\boldsymbol{F}_{\mathrm{avg}}\in\mathbb{R}^{C\times1\times1}$；其次，利用卷积核为$K$的一维卷积获取局部跨通道信息；接着，通过Sigmoid激活函数得到通道权重；最后，将得到的权重值与原始特征信息逐通道相乘，获取到最终特征$\boldsymbol{F}'\in\mathbb{R}^{C\times H\times W}$，用于后续的特征融合。

其中，一维卷积的卷积核K由具体的通道参数自适应决定，计算公式为

$$K=\varphi(C)=\left|\frac{\log_2 C}{\gamma}+\frac{b}{\gamma}\right|_{\mathrm{odd}},\tag{1}$$

式中：C为通道数；$|\cdot|_{\mathrm{odd}}$为取最邻近奇数；γ和b均为超参数，分别为2和1。

除此之外，为了更好地评估注意力机制模块对行人目标检测性能的影响，选用CBAM来验证特征提取网络的性能。实验分析部分验证了与CBAM相比，ECA机制模块具有更好性能的结论。

2.3　IPWF模块

可见光与红外特征被注意力机制模块增强后，传输到所提IPWF模块中进行加权融合。该融合模块在减少网络特征叠加造成的特征冗余同时，满足在不同光照条件下可见光与红外图像的自适应融合要求，从而在全天候下能够良好地检测行人。本文在可见光与红外特征经过的注意力机制后的每一层都加入IPWF模块，充分利用每一层特征，进一步加强两种模态中对重要信息的学习。IPWF模块是通过设计的微型神经网络来获取可见光与红外模态对应权重的[16]，在不增加整体网络结构复杂度的同时，利用权重加权生成融合特征，用于后续的检测。IPWF模块结构如图3所示。

在图3中，由于需要在不同光照条件下进行可见光与红外图像融合，而红外图像不能够反映环境中的光照条件，选用可见光特征作为IPWF网络的输入，通过对可见光特征进行加权操作，得到白天和夜晚的光照预测权重值w_{d}和w_{n}；然后在光照机制中重新调整w_{d}和w_{n}，生成更可靠的可见光权重w_{r}与红外权重w_{t}；对其与原始特征对应相乘并级联，得到最终的融合特征。该模块结构主要由权重生成和光照机制加权融合2部分组成。

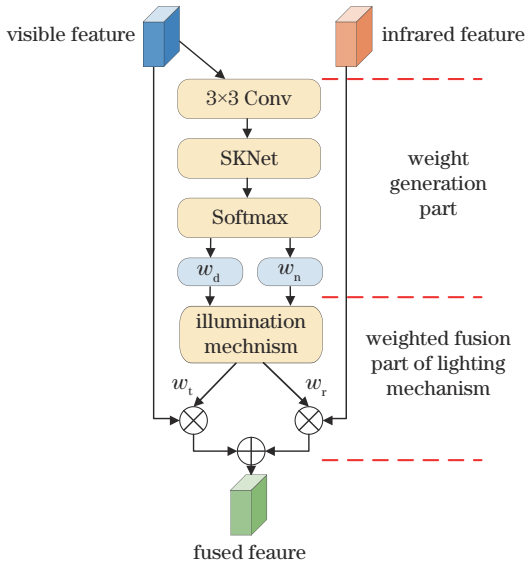

图 3　IPWF 模块的结构

Fig. 3　Structure of IPWF module

2.3.1　权重生成部分

权重生成部分通过 1 个 3×3 卷积层、1 个 SKNet 层[17]和全连接层的组合完成权重的生成。其中 SKNet 层主要对不同的图像生成不同的卷积核，以便于网络能够根据不同图像的特征选择合适的卷积核，从而使得到的特征具有更加丰富的感受野，进一步增强网络模型的鲁棒性[18]，结构如图 4 所示。

SKNet 层主要有分解、融合和选择 3 个步骤。

1）分解。对输入的特征图 $\boldsymbol{X}\in\mathbb{R}^{C\times H\times W}$ 分别进行 3×3 和 5×5 两个不同卷积核的卷积操作，得到特征图 $\boldsymbol{F}_1:\boldsymbol{X}\rightarrow\boldsymbol{U}_1\in\mathbb{R}^{C\times H\times W}$ 和 $\boldsymbol{F}_2:\boldsymbol{X}\rightarrow\boldsymbol{U}_2\in\mathbb{R}^{C\times H\times W}$。

2）融合。通过对特征图 $\boldsymbol{U}_1$ 和 $\boldsymbol{U}_2$ 进行求和操作，得到融合不同感受野的特征图 $\boldsymbol{U}$，

$$\boldsymbol{U}=\boldsymbol{U}_1+\boldsymbol{U}_2\text{。}\tag{2}$$

然后通过 GAP 操作，得到具有全局信息的特征 $\boldsymbol{S}_c$：

$$\boldsymbol{S}_c=F_{\mathrm{GAP}}(\boldsymbol{U})=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}U(i,j)\text{。}\tag{3}$$

再利用全连接（FC）层进行线性变换，得到特征 $\boldsymbol{z}$，有效降低特征维度，提高效率，公式为

$$\boldsymbol{z}=F_{\mathrm{FC}}(\boldsymbol{s})=\delta\left[\beta(\boldsymbol{W}_s)\right],\tag{4}$$

式中：δ 为 ReLU 激活函数；β 为 BN 层；$\boldsymbol{W}_s\in\mathbb{R}^{d\times c}$。

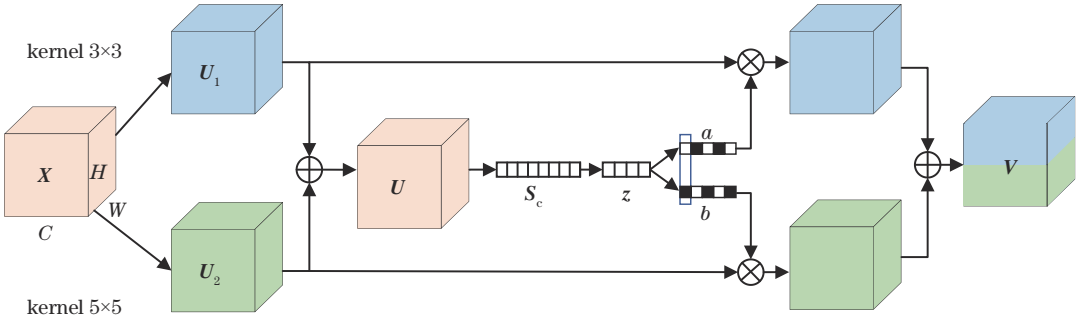

图 4　SKNet 结构

Fig. 4　SKNet structure

3）选择。经过融合操作，可自适应得到不同空间尺度的权重信息，对 $\boldsymbol{z}$ 进行 Softmax 函数计算，可得到 a 和 b 权重通道特征层，公式为

$$\begin{cases}a=\dfrac{\mathrm{e}^{Az}}{\mathrm{e}^{Az}+\mathrm{e}^{Bz}}\\ b=\dfrac{\mathrm{e}^{Bz}}{\mathrm{e}^{Az}+\mathrm{e}^{Bz}}\end{cases}\text{。}\tag{5}$$

再进行加权操作，得到向量 $\boldsymbol{V}$，公式为

$$\begin{cases}\boldsymbol{V}=a\boldsymbol{U}_1+b\boldsymbol{U}_2\\ a+b=1\end{cases}\text{。}\tag{6}$$

可见光图像特征在经过 SKNet 层后，有效加强了行人目标的特征，抑制了图像中的其他特征信息，最后将输出的特征信息输入到全连接层，通过全连接层预测出可见光图像白天和夜晚的权重值 w_d 和 w_n。

2.3.2　光照机制加权融合

由于现有的基于光照加权的融合策略将生成的白天和夜晚权重作为可见光与红外特征的权重值，可见光与红外特征加权融合后生成融合特征。这种方法虽然能够解决不同模态在不同光照条件下自适应加权融合的问题，但同时又存在融合机制适用性不强的问题，不能保证甚至会降低对行人目标的检测性能，对应情况如图 5 所示。

行人目标出现在白天光照良好的树荫下时，可见光图像中的特征不明显，而从红外图像中能清晰地分辨出行人目标，此时对可见光与红外图像进行融合时，应提供红外图像特征更多的权重值，才能使融合效果更好。现有的光照加权融合方法在白天光照良好的情况下会使白天条件下的目标权重变大，夜晚条件下的目标权重变小，从而赋给可见光特征的权值变大，红外特征的权值变小，与实际所期望的权重相反，影响融合效果。因此，针对这一问题，本文提出了一种改进的光照机制，在光照机制中对从可见光图像获取到的白天与黑夜权重重新进行调整，以得到适用性更强的可见光与红外权重，从而再进行加权融合。

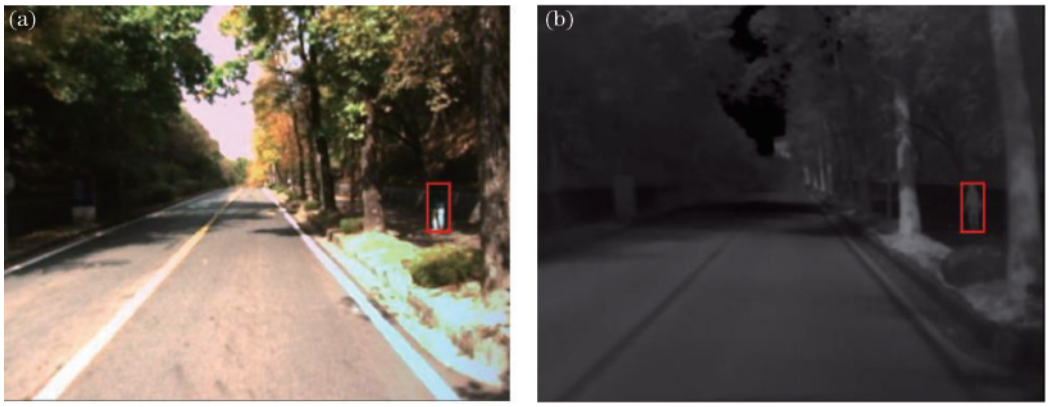

图 5　树荫下可见光与红外图像对。(a)可见光图像;(b)红外图像

Fig. 5　Visible and infrared image pair under shade trees. (a) Visible light image; (b) infrared image

在光照机制中对经全连接层生成的白天与夜晚的权重值 w_{d} 和 w_{n} 进行重新调整,得到可见光与红外图像的预权重值 w_{r}' 和 w_{t}'。光照机制调整权重的计算公式为

$$\begin{cases} w_{\mathrm{r}}'=\left[(w_{\mathrm{d}}-w_{\mathrm{n}})/2\right]\cdot b+1/2 \\ w_{\mathrm{t}}'=1-w_{\mathrm{r}}' \\ b=\alpha_w\cdot|w|+\gamma_w \end{cases}。\tag{7}$$

为了使可见光与红外特征的权重值更接近对应图像的描述,先将可见光与红外的初始权重均定义为 0.5,再从 0.5 开始学习权重的偏差,b 为权重的偏差,其中 $|w|\in[0,1]$。α_w 和 γ_w 为可学习的参数,其初始化值分别为 1 和 0。

然后,为了增强光照机制的鲁棒性,通过 Sigmoid 函数对得到的可见光预权重值进行进一步调整,得到最终的可见光权重 w_{r},此时的红外权重值 w_{t} 为总光照值 1 减去可见光权重的结果。δ 表示 Sigmoid 函数操作,表达式为

$$\delta(x)=\frac{1}{1+\mathrm{e}^{-x}}。\tag{8}$$

可见光权重 w_{r} 和红外权重 w_{t} 的计算公式分别为

$$\begin{cases} w_{\mathrm{r}}=\delta\left(w_{\mathrm{r}}'\right) \\ w_{\mathrm{t}}=1-w_{\mathrm{r}} \end{cases}。\tag{9}$$

最后,将得到的可见光与红外图像权重值分别与对应的原始特征相乘并进行级联,得到最终的融合特征。

2.4　损失函数和优化方法

网络的整体损失函数是由分类损失 L_{cls}、回归损失 L_{reg} 和光照损失 L_{I} 构成的,与文献[16]一致,表达式为

$$L=L_{\mathrm{I}}+L_{\mathrm{cls0}}+L_{\mathrm{reg0}}+L_{\mathrm{cls1}}+L_{\mathrm{reg1}},\tag{10}$$

式中:L_{cls0} 和 L_{reg0} 为可见光分支输出的分类损失和候选框回归损失;L_{cls1} 和 L_{reg1} 为红外分支输出的分类损失和候选框回归损失。

光照损失 L_{I} 使用交叉熵损失函数来优化网络[19],表达式为

$$L_{\mathrm{I}}=-\hat{\omega}_{\mathrm{d}}\cdot\log\omega_{\mathrm{d}}-\hat{\omega}_{\mathrm{n}}\cdot\log\omega_{\mathrm{n}},\tag{11}$$

式中:$\hat{\omega}_{\mathrm{d}}$ 和 $\hat{\omega}_{\mathrm{n}}$ 分别为白天与夜晚的真实标签,当训练图像处于白天场景时,$\hat{\omega}_{\mathrm{d}}=1$,$\hat{\omega}_{\mathrm{n}}=0$;训练图像处于夜晚场景时,$\hat{\omega}_{\mathrm{d}}=0$,$\hat{\omega}_{\mathrm{n}}=1$。$w_{\mathrm{d}}$ 和 w_{n} 是最后一层全连接层经过 Softmax 激活后,光照强度预测网络输出的白天预测权重和夜晚预测权重。

分类损失 L_{cls} 设定为聚焦损失函数,公式为

$$L_{\mathrm{cls}}=-\alpha\sum_{i\in S_+}(1-q_i)\log q_i-(1-\alpha)\sum_{i\in S_-}q_i^{\gamma}\log(1-q_i),\tag{12}$$

式中:q_i 是样本 i 的正概率;S_+ 和 S_- 分别代表正确和错误的候选框;α 和 γ 均为聚焦参数,分别设定为 0.25 和 2[20]。

回归损失 L_{reg} 使用 smooth L1 函数进行优化,表示为

$$L_{\mathrm{reg}}=\sum_{m\in\{c_x,c_y,w,h\}}\mathrm{smooth}_{\mathrm{L1}}\left(l_i^m-\hat{g}_j^m\right),\tag{13}$$

$$\mathrm{smooth}_{\mathrm{L1}}(x)=\begin{cases} 0.5x^2, & |x|<1 \\ |x|-0.5, & |x|\geqslant 1 \end{cases},\tag{14}$$

式中:(c_x,c_y) 是边界框的中心点坐标;w 和 h 分别为边界框的宽和高;l_i^m 和 $\hat{g}_j^m$ 是第 i 个回归框和第 j 个真实框的偏移量;m 表示目标框信息,包含目标框的中心点、高和宽。$\hat{g}_j^m$ 定义为

$$\begin{cases} \hat{g}_j^{c_x}=\left(g_j^{c_x}-d_i^{c_x}\right)/d_i^w \\ \hat{g}_j^{c_y}=\left(g_j^{c_y}-d_i^{c_y}\right)/d_i^h \\ \hat{g}_j^w=\log\dfrac{g_j^w}{d_i^w} \\ \hat{g}_j^h=\log\dfrac{g_j^h}{d_i^h} \end{cases},\tag{15}$$

式中:$\hat{g}_j^{c_x}$、$\hat{g}_j^{c_y}$、$\hat{g}_j^w$、$\hat{g}_j^h$ 分别为真实框 $g_j^{c_x}$、$g_j^{c_y}$、g_j^w、g_j^h 的偏移量;$d_j^{c_x}$、$d_j^{c_y}$、d_j^w、d_j^h 分别为默认框的中心、宽、高。

3　分析与讨论

3.1　实验配置与参数设置

所提算法是在Keras深度学习框架下完成的，实验环境为Python 3.5，CUDA 10.0，Ubuntu 18.04，硬件配置为RTX 2080Ti。实验中采用小批量梯度下降法（MBGD）进行训练，设置每批训练图像数batch size为4，模型共迭代了16个epoch，初始学习率设置为0.0001。

3.2　数据集与评价指标

使用公开的KAIST数据集。数据集是车载摄像头采集的配准多光谱图像对，包含了在校园、城镇和公路等场景中的白天和夜晚图像，共有12个子数据集（set00～set02、set06～set08采集时间为白天，set03～set05、set09～set11采集时间为夜晚），图像大小为640×512。在训练集中去除数据集中没有行人目标的样本，最终筛选到8963对可见光与红外图像用于训练。测试数据集与原数据集划分一致，其中全时段测试集共2252对图片，包含白天场景下1455对图片和夜晚场景下797对图片。

利用平均漏检率（MR）和对数坐标下MR-FPPI曲线[21]进行实验验证，并评估算法的性能，其中MR值越小，MR-FPPI曲线越低，表明算法性能越优。

3.3　消融实验结果分析

3.3.1　融合策略消融实验

对可见光与红外两种图像的特征进行融合，在KAIST测试集下用不同的融合策略进行测试分析，从而得到具有最佳性能的行人目标检测方法[22]。表1为不同融合策略下的网络性能。其中直接堆叠融合（ZJDD）为其他多模态行人目标检测算法常用的融合策略；TIWF表示传统光照加权融合策略；IPWF表示所提光照感知权重融合策略；在IPWF的基础上，分别加入CBAM、SE和ECA三种注意力机制模块，表示为IPWF＋CBAM、IPWF＋SE和IPWF＋ECA；在TIWF的基础上，加入ECA机制模块，表示为TIWF＋ECA。对它们与所提算法进行对比。图6为不同融合策略的MR-FPPI曲线。

由表1和图6可知，所提IPWF＋ECA融合策略具有最优性能。其中在MR方面：与ZJDD方法相比，所提IPWF策略的漏检率下降了2.92个百分点，有较大的性能提升；在此基础上通过加入不同的注意力机制模块来提升网络性能，但根据实验结果来看，加入CBAM和SE注意力机制后反而使MR分别上升了0.59个百分点和0.43个百分点，而加入ECA机制使MR降低了1.58个百分点，对多模态行人目标检测提供了正向帮助；同时，与加入ECA机制的传统TIWF策略相比，虽然传统TIWF策略有较低的MR，但所提算法能够获得更优的性能，MR比传统方法下降了0.95个百分点。在模型大小方面：引入光照感知权重融合模块之前，多模态行人目标检测网络模型的大小为305.7×10^6，引入光照感知权重融合模块之后模型大小为320.5×10^6，模型大小仅增加了4.84%，基本不影响算法的实时性；同时，在光照感知权重融合模块的基础上增加的CBAM和SE注意力机制与ECA机制相比，增加ECA机制的网络模型的大小基本没有发生变化；并且与在传统光照加权融合策略的基础上增加的ECA机制模型相比，所提算法的模型大小与之相差不大。

表1　不同融合策略的性能比较

Table 1　Performance comparison of different fusion policies

Fusion strategy	MR /%			Model size /10^6
	All the time	Day	Night	
ZJDD	15.66	17.85	11.39	305.7
IPWF	12.74	14.69	9.51	320.5
IPWF＋CBAM	13.33	14.81	10.17	346.3
IPWF＋SE	13.17	14.50	9.86	346.0
TIWF＋ECA	12.11	13.65	9.45	317.7
IPWF＋ECA	11.16	12.40	8.46	320.7

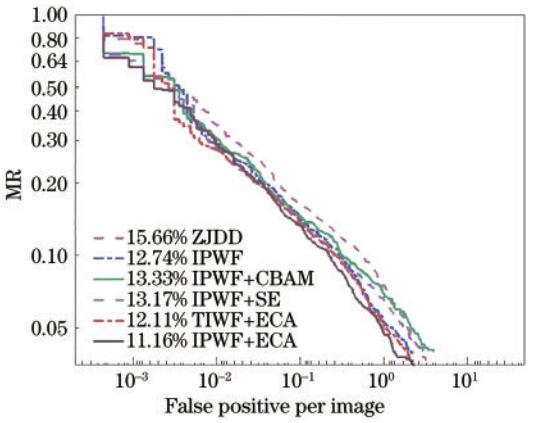

图6　不同融合策略的MR-FPPI曲线

Fig. 6　MR-FPPI curves of different fusion strategies

综上分析，所提光照感知权重融合模块在模型体积增加不大的情况下，能较大地降低行人目标检测的漏检率，且在引入ECA机制模块后，不仅能够保持模型体积基本不变，且能进一步提升算法的检测性能，证明了所提算法对行人目标检测的有效性。

3.3.2　损失函数消融实验

本文的损失函数由分类损失L_{cls}、回归损失L_{reg}和光照损失L_I组成。为了证明所提损失函数的有效性，对损失函数也进行了消融实验。其中分类损失与光照损失交叉采用聚焦损失（focal loss）和交叉熵损失（cross entropy loss），回归损失分别使用smooth L1和CIOU损失函数，构成了不同的损失策略，用√表示模型在训练过程中包含的对应损失函数。损失函数消融实验结果如表2所示。

由表2可以看出：在未加入光照损失之前，所提算

表 2　损失函数消融实验结果

Table 2　Experimental results of loss function ablation

Ablation strategy	L_{cls}		L_{reg}		L_I		MR /%
	Focal loss	Cross entropy loss	Smooth L1	CIOU	Focal loss	Cross entropy loss	
Strategy 1	√		√				11.87
Strategy 2	√			√			12.06
Strategy 3		√	√				12.17
Strategy 4		√		√			12.55
Strategy 5		√	√		√		11.60
Strategy 6	√		√			√	11.16

法利用分类损失和回归损失均能够获得较低的行人目标检测漏检率；在分类损失相同的情况下，与 CIOU 损失函数相比，采用 smooth L1 损失函数后获得了更低的行人目标检测漏检率。因此，在消融策略 5 与策略 6 中选用 smooth L1 损失函数作为回归损失，同时，引入了光照损失，当光照损失采用交叉熵损失函数，分类损失采用聚焦损失时，行人目标检测漏检率达到了最低。

3.4　不同算法对比实验结果

所提算法通过光照感知权重融合模块有效提高了行人目标检测性能，并在此基础上通过对增加的不同注意力机制模块进行实验分析，得到效果最佳的注意力机制模块，进一步增强了算法的检测性能。为了证明所提算法的有效性，对所提算法与现有效果较优的 6 种方法 ACF＋T＋THOG[23]、Halfway-fusion[4]、Fusion-RPN[5]、IAF-RCNN[9]、IATDNN＋IAMSS[8]、GFR[24]在 KAIST 多模态行人检测数据集上进行了 MR 值和检测速度的比较。比较结果如表 3 所示。图 7 为不同算法的 MR-FPPI 曲线。

表 3　不同算法的性能比较

Table 3　Performance comparison of different algorithms

Algorithm	MR /%			Speed /(frame·s^{-1})
	All the time	Day	Night	
ACF＋T＋THOG[23]	47.32	42.57	56.17	0.27
Halfway-fusion[4]	25.75	24.88	26.59	0.43
Fusion-RPN[5]	18.29	19.57	16.27	0.80
IAF-RCNN[9]	15.73	14.55	18.26	0.21
IATDNN＋IAMSS[8]	14.95	14.67	15.72	0.25
GFR[24]	11.51	12.64	10.63	0.12
Proposed algorithm	11.16	12.40	8.46	0.08

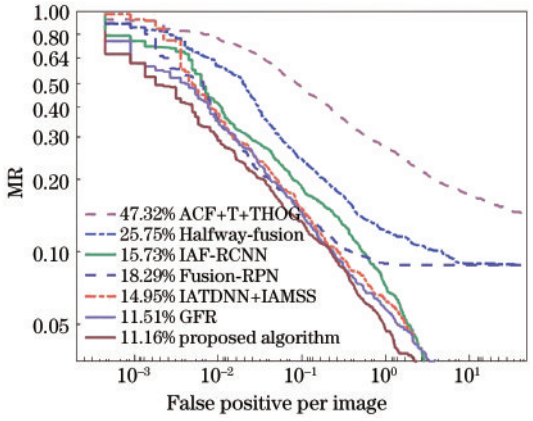

图 7　不同算法的 MR-FPPI 曲线

Fig. 7　MR-FPPI curves of different algorithms

通过表 3 和图 7 中不同算法在 MR 和检测速度指标上的性能分析，所提算法所具有的光照感知权重融合机制和注意力机制增强了行人目标检测性能，在全时段、白天以及黑夜三种条件下均获得了最低的 MR 值，优于现有的其他算法，并且检测速度也很快。

除此之外，还检测了不同算法分别在 LLVIP 数据集和 M^{3}FD 数据集上的结果，其中 LLVIP 数据集中的图像大部分是处在黑夜环境下的，M^{3}FD 数据集中存在较多的雨天、雾天等环境下的图像，因此，此仅考虑全天条件下的漏检率，不区分白天与黑夜两种条件。不同算法在 LLVIP 数据集和 M^{3}FD 数据集上的漏检率和检测速度如表 4 所示。

根据表 4 可以看出，在检测漏检率上，由于 LLVIP 数据集中存在较多骑车的行人和遮挡下的行人目标，M^{3}FD 数据集中较多图像受到雨雾天气的影响，不同算法在这两种数据集下的行人目标检测性能均有所下降，但所提算法与其他算法相比，仍取得了最低的检测漏检率。在检测速度上，所提算法在不同数据集下均具有较好的检测速率。

表 4　不同算法在 LLVIP 数据集和 M^3FD 数据集上的性能比较

Table 4　Performance comparison of different algorithms on LLVIP dataset and M^3FD dataset

Algorithm	LLVIP dataset		M^3FD dataset	
	MR /%	Speed /(frame·s^{-1})	MR /%	Speed /(frame·s^{-1})
ACF+T+THOG[23]	59.40	0.32	51.20	0.30
Halfway-fusion[4]	34.40	0.47	28.94	0.43
Fusion-RPN[5]	26.15	0.81	21.10	0.80
IAF-RCNN[9]	23.70	0.20	18.28	0.21
IATDNN+IAMSS[8]	24.41	0.25	18.16	0.26
GFR[24]	22.10	0.14	17.37	0.12
Proposed algorithm	20.17	0.09	16.07	0.08

3.5　实验检测结果

为了更加直观地验证所提算法的有效性，给出所提算法在白天场景和夜晚场景下的行人目标检测结果，如图 8 和图 9 所示。其中图 8(a)和图 9(a)中行人目标在图像中的原始标注框位置用红色框表示，图 8(b)和图 9(b)中行人目标的检测框位置用绿色框表示。

在图 8 白天场景下：针对左侧光照良好的街区环境下的行人目标，所提算法能准确地检测到，且生成的检测框与原始标注框位置有较少的偏移量；针对右侧

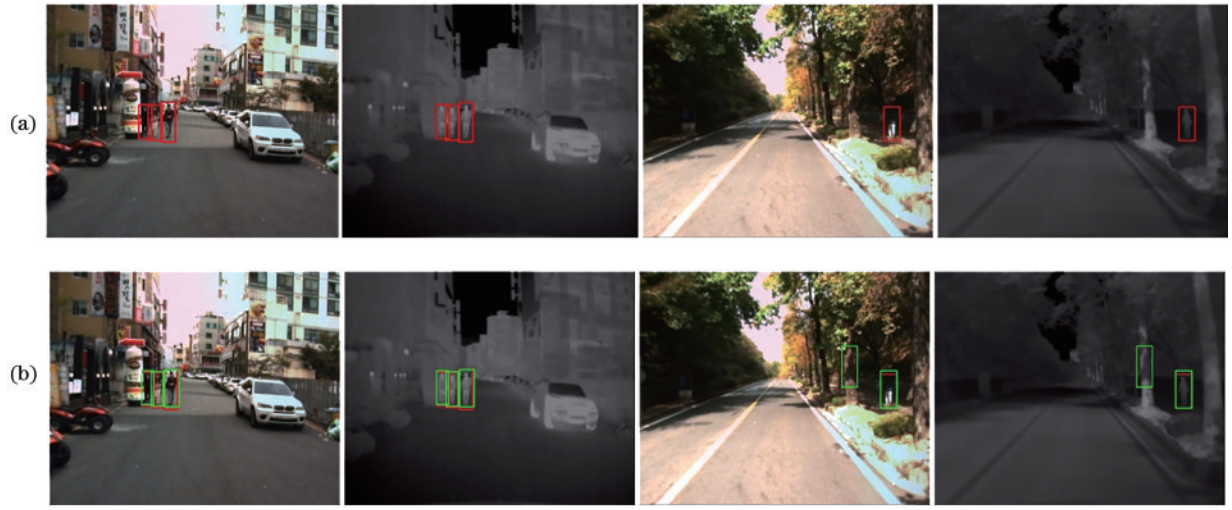

图 8　白天场景下行人目标检测结果。(a)原始标注；(b)检测结果

Fig. 8　Pedestrian target detection results in day time scenes. (a) Original annotation; (b) test result

图 9　夜晚场景下行人目标检测结果。(a)原始标注；(b)检测结果

Fig. 9　Pedestrian target detection results in night time scenes. (a) Original annotation; (b) test result

光照良好的图像，在行人目标出现在树荫导致可见光图像不易分辨的情况下，所提算法仍能够有效检测出行人目标，但存在将旁边的树木误检为行人目标的现象。在图 9 夜晚场景下：针对左侧夜间较为昏暗且场景较为简单的环境，所提算法能够将行人目标精确检测出且检测框与原始标注框位置相差不大，但对于较小的行人目标，仅检测出部分目标，出现目标漏检情况，且检测到的目标框与原始框位置有较大差异；针对右侧夜间灯光明亮且场景较为复杂的街区环境，所提算法均能够将图像中出现的目标检测出，但对于路边出现的其他物体，由于具有一定的热量，在红外图像中表现出与行人目标相似的特征，会将其误检为行人目标。

除此之外，在 LLVIP 数据集和 M^{3}FD 数据集上分别进行了不同条件下的实验实例测试，以验证所提算法的泛化性能，测试结果如图 10 和图 11 所示。其中红色框表示目标原始标注框，绿色框表示目标的检测框。

图 10 LLVIP 数据集下行人目标检测结果。(a)白天条件检测结果；(b)夜晚条件检测结果

Fig. 10 Pedestrian target detection results on LLVIP dataset. (a) Detection results of daytime condition; (b) detection results of night condition

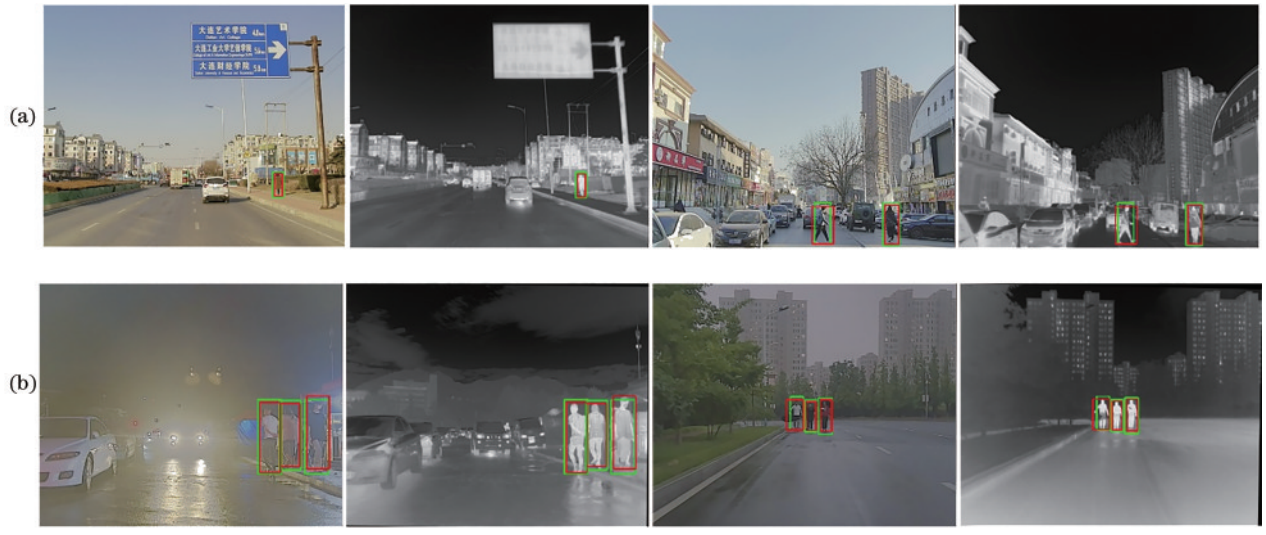

图 11 M^{3}FD 数据集下行人目标检测结果。(a)白天条件检测结果；(b)夜晚条件检测结果

Fig. 11 Pedestrian target detection results on M^{3}FD dataset. (a) Detection results of daytime condition; (b) detection results of night condition

在图 10 LLVIP 数据集中，对于检测到的行人目标，其目标的检测框与原始标注框位置相差不大，但出现未能检测到在白天条件下骑自行车的行人的问题，主要原因在于骑车的行人姿态与直立行走的行人不相同，存在目标特征的差异性。在图 11 M^{3}FD 数据集中，所提算法在白天光照较强的条件下能准确检测到路边小尺寸的行人目标，而对于夜间光照较为昏暗条件下的远处小尺寸行人目标，目标检测框与原始标注框位置存在略微的偏移。但总的来说，所提算法在不同数据集不同光照条件下均能够有效检测出图像中的

行人目标，也证明了所提算法具有较好的泛化能力。

4 结 论

针对目前基于可见光与红外多模态图像融合的行人检测算法漏检率高的问题，提出了基于光照感知权重融合的多模态行人检测网络，适应全天候及各种交通场景变化。首先在特征提取网络中加入注意力机制来增强特征；其次提出一种光照感知权重融合模块，自适应地学习可见光与红外特征的权重并进行加权融合；最后将融合特征与两种模态的最后一层特征共同送入到检测网络中进行检测，完成行人目标的检测。实验结果表明，与现有的其他先进算法相比，所提算法具有较低的漏检率，在全天候下不同光照场景对行人目标具有良好的检测性能。但对于场景图像中小尺寸和骑车的行人目标，所提算法检测性能不佳，易出现漏检、误检、行人检测框与原始标注框差异较大的情况，后续需要针对这些复杂条件进一步改进算法，进一步提高算法的性能。

参 考 文 献

[1] 李翔，何淼，罗海波. 一种面向遮挡行人检测的改进YOLOv3算法[J]. 光学学报，2022，42(14)：1415003.
Li X，He M，Luo H B. Occluded pedestrian detection algorithm based on improved YOLOv3[J]. Acta Optica Sinica，2022，42(14)：1415003.

[2] 冷阳. 可见光/红外图像特征级融合目标识别研究[D]. 南京：南京航空航天大学，2019.
Leng Y. Research on target recognition based on feature-level fusion of visible/infrared images[D]. Nanjing：Nanjing University of Aeronautics and Astronautics，2019.

[3] Ma J Y，Yu W，Liang P W，et al. FusionGAN：a generative adversarial network for infrared and visible image fusion[J]. Information Fusion，2019，48：11-26.

[4] Liu J J，Zhang S T，Wang S，et al. Multispectral deep neural networks for pedestrian detection[C]//Proceedings of the British Machine Vision Conference 2016，September 19-22，2016，York，UK. Guildford：British Machine Vision Association，2016.

[5] 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，July 21-26，2017，Honolulu，HI，USA. New York：IEEE Press，2017：243-250.

[6] Cao Y P，Guan D Y，Huang W L，et al. Pedestrian detection with unsupervised multispectral feature learning using deep neural networks[J]. Information Fusion，2019，46：206-217.

[7] Lee Y，Bui T D，Shin J. Pedestrian detection based on deep fusion network using feature correlation[C]//2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)，November 12-15，2018，Honolulu，HI，USA. New York：IEEE Press，2018：694-699.

[8] Guan D Y，Cao Y P，Yang J X，et al. Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection[J]. Information Fusion，2019，50：148-157.

[9] 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.

[10] Liu W，Anguelov D，Erhan D，et al. SSD：single shot MultiBox detector[M]//Leibe B，Matas J，Sebe N，et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham：Springer，2016，9905：21-37.

[11] He K M，Zhang X Y，Ren S Q，et al. Identity mappings in deep residual networks[M]//Leibe B，Matas J，Sebe N，et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham：Springer，2016，9908：630-645.

[12] 李凯，林宇舜，吴晓琳，等. 基于多尺度融合与注意力机制的小目标车辆检测[J]. 浙江大学学报(工学版)，2022，56(11)：2241-2250.
Li K，Lin Y X，Wu X L，et al. Small target vehicle detection based on multi-scale fusion and attention mechanism[J]. Journal of Zhejiang University (Engineering Science)，2022，56(11)：2241-2250.

[13] 余子航. 全时段交通场景下多光谱行人检测算法研究[D]. 无锡：江南大学，2021.
Yu Z H. Research on multispectral pedestrian detection algorithm in all-time traffic scene[D]. Wuxi：Jiangnan University，2021.

[14] Wang Q L，Wu B G，Zhu P F，et al. ECA-net：efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)，June 13-19，2020，Seattle，WA，USA. New York：IEEE Press，2019：11531-11539.

[15] 邹梓吟，盖绍彦，达飞鹏，等. 基于注意力机制的遮挡行人检测算法[J]. 光学学报，2021，41(15)：1515001.
Zou Z Y，Gai S Y，Da F P，et al. Occluded pedestrian detection algorithm based on attention mechanism[J]. Acta Optica Sinica，2021，41(15)：1515001.

[16] Zhou K L，Chen L S，Cao X. Improving multispectral pedestrian detection by addressing modality imbalance problems[M]//Vedaldi A，Bischof H，Brox T，et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham：Springer，2020，12363：787-803.

[17] 王战涛，张策，王晓田. 基于YOLOV$_3$的改进目标检测识别算法[J]. 上海航天(中英文)，2021，38(6)：60-70.
Wang Z T，Zhang C，Wang X T. Improved target detection and recognition algorithm based on YOLOV$_3$[J]. Aerospace Shanghai (Chinese & English)，2021，38(6)：60-70.

[18] 何自芬，陈光晨，陈俊松，等. 多尺度特征融合轻量化夜间红外行人实时检测[J]. 中国激光，2022，49(17)：1709002.
He Z F，Chen G C，Chen J S，et al. Multi-scale feature fusion lightweight infrared pedestrian real-time detection at night[J]. Chinese Journal of Lasers，2022，49(17)：1709002.

[19] 童靖然. 基于多模态数据的目标检测与追踪[D]. 无锡：

江南大学, 2019.
Tong J R. Video target detection and tracking based on multimodal data[D]. Wuxi: Jiangnan University, 2019.

[20] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.

[21] 陈莹, 朱宇. 模态自适应权值学习机制下的多光谱行人检测网络[J]. 光学 精密工程, 2020, 28(12): 2700-2709.
Chen Y, Zhu Y. Multispectral pedestrian detection network under modal adaptive weight learning mechanism [J]. Optics and Precision Engineering, 2020, 28(12): 2700-2709.

[22] 施政, 毛力, 孙俊. 基于YOLO的多模态加权融合行人检测算法[J]. 计算机工程, 2021, 47(8): 234-242.
Shi Z, Mao L, Sun J. YOLO-based multi-modal weighted fusion pedestrian detection algorithm[J]. Computer Engineering, 2021, 47(8): 234-242.

[23] 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, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 1037-1045.

[24] Zhang H, Fromont E, Lefevre S, et al. Multispectral fusion for object detection with cyclic fuse-and-refine blocks[C]//2020 IEEE International Conference on Image Processing, October 25-28, 2020, Abu Dhabi, United Arab Emirates. New York: IEEE Press, 2020: 276-280.