

基于改进 Shuffle-RetinaNet 的红外车辆检测算法

范晓畅, 梁煜, 张为*

天津大学微电子学院, 天津 300072

摘要 针对当前红外场景下多尺度车辆检测精度欠佳且算法模型复杂度高的问题,提出了基于 Shuffle-RetinaNet 的红外车辆检测算法。该算法以 RetinaNet 网络为基础,并选用 ShuffleNetV2 作为特征提取网络。提出双分支注意力模块,通过双分支结构和自适应融合方法增强网络对红外图像中目标关键特征的提取能力;优化特征融合网络,集成双向交叉尺度连接和快速归一化融合,增强目标多尺度特征的表达能力;设置校准因子增强分类和回归之间的任务交互,提高目标分类和定位的准确性。该算法在自建红外车辆数据集上的检测精度达到 92.9%,参数量为 11.74×10^6 ,浮点计算量为 24.35×10^9 ,同时在公开红外数据集 FLIR ADAS 上也展现出较好的检测性能。实验结果表明:该算法具有较高的检测精度,且模型复杂度低,在红外车辆检测领域具有较高的应用价值。

关键词 目标检测; 红外车辆; 通道注意力模块; 多尺度特征融合; 校准因子

中图分类号 TP183 文献标志码 A

DOI: 10.3788/LOP230713

Infrared Vehicle Detection Algorithm Based on Improved Shuffle-RetinaNet

Fan Xiaochang, Liang Yu, Zhang Wei*

School of Microelectronics, Tianjin University, Tianjin 300072

Abstract In view of the low detection accuracy and high complexity of current multi-scale vehicle detection algorithms in infrared scenes, an infrared vehicle detection algorithm based on Shuffle-RetinaNet is proposed. On the basis of RetinaNet, the algorithm uses ShuffleNetV2 as the feature extraction network. A dual-branch attention module channel attention module is proposed, which adopts the dual-branch structure and adaptive fusion and enhances the ability to extract the key features of the target in infrared images. To optimize the feature fusion, the algorithm integrates cross-scale connection and fast normalized fusion in some feature layers to enhance the multi-scale feature expression. The calibration factor is set to enhance the task interaction of classification and regression, and the accuracy of target classification and locating is increased. A series of experiments are conducted on a self-built infrared vehicle dataset to verify the effectiveness of the proposed algorithm. The detection accuracy of this algorithm for the self-built vehicle dataset is 92.9%, the number of parameters is 11.74×10^6 , and the number of floating-point operations is 24.35×10^9 . The algorithm exhibits better detection performance on the public dataset FLIR ADAS. Experimental results indicate that the algorithm has advantages in detection accuracy and model complexity, giving it good application value in multi-scale vehicle detection tasks in infrared scenes.

Key words object detection; infrared vehicle; channel attention module; multi-scale feature fusion; calibration factor

1 引言

红外成像技术无需借助外部环境光,与可见光相比具有夜视范围广、覆盖全面等优点,适用于一些特殊环境。基于此,红外成像被引入目标检测领域,开始应用于机场安检,红外跟踪系统,红外导航系统乃至现代军事侦察等工作任务中。近年来,红外场景下的车辆

检测更是成为交通、军事等领域的关注重点。然而,红外成像存在分辨率较低,图像目标信号弱等问题,与可见光目标相比纹理、颜色信息不够丰富,这为红外车辆检测任务带来一定挑战。

传统的红外车辆检测技术通常基于经典图像处理方法,通过图像阈值分割、梯度直方图以及支持向量机(SVM)等方法对车辆目标的特征进行描述、分类,进

收稿日期: 2023-02-27; 修回日期: 2023-04-12; 录用日期: 2023-04-20; 网络首发日期: 2023-04-30

通信作者: *tjuzhangwei@tju.edu.cn

而实现检测和定位。Tao 等^[1]基于夜间红外图像,提出了一种结合 SVM 和车辆特征提取的多特征融合算法,该算法可以通过快速细化和分离目标区域,获得准确的目标候选边界框。崔紫薇等^[2]提出了一种基于萤火虫算法的二维 Otsu 阈值分割和基于几何参数约束的连通区域标记相结合的感兴趣区域提取方法,提高了红外场景目标检测中的分割效果和定位精度。Lin 等^[3]提出了一种基于帧间相关性的红外图像序列运动车辆检测方法,对单帧图像进行去噪和锐化增强,根据真实车辆和虚假车辆运动特征的不同,依靠帧间相关性抑制误报。传统检测手段在红外场景的车辆检测中有一定的应用价值,但同时存在检测精度低、泛化能力差等问题,实际应用中通常难以完成检测任务。

近年来,深度学习技术发展迅速,卷积神经网络凭借自身优秀的特征提取能力,在目标检测领域展现出较高的应用价值,逐渐受到研究者青睐。Zhao 等^[4]结合 DenseNet 网络和 YOLOv5s 的特征,将 YOLOv5s 的 C_3 模块中瓶颈模块替换为定制的 Denseblock 模块,提高了网络对红外场景车辆目标的特征提取能力。Li 等^[5]提出了一种基于固态硬盘(SSD)的增量式学习红外车辆检测方法,通过增量学习和训练对检测模型进行微调,改善了误检和漏检问题。李向荣等^[6]基于 YOLOv4 架构提出了一种融合通道注意力机制的多尺度红外目标检测算法,提高了特征信息利用率。顾星等^[7]针对红外小目标检测任务设计了一种轻量化网络结构,改进了迭代特征金字塔结构,并在训练过程中引入完全交并比(CIoU)损失函数和梯度调和机制(GHM)损失函数改善了正负样本的不平衡问题。章琦等^[8]以轻量级主干网络为基础设计了基于上下文调制的跨层间特征融合方式,并引入基于通道和位置的双注意力机制突出特征图中的弱小目标。

综合以往的研究,发现基于深度学习的红外车辆检测算法仍存在一定问题:

1) 特征信息提取能力不足。与可见光相比,该算法红外图像分辨率低,车辆目标信号弱小,关键特征信息稀缺,现有算法通常提取目标特征信息不充分,影响了模型的检测性能。

2) 多尺度车辆检测精度低。实际应用场景中受车辆运动、视野变化影响,检测目标通常尺度不一。现有算法对多尺度车辆目标尤其中小目标检测精度不足,容易出现漏检情况。

3) 实际应用中模型复杂度通常较大,在满足检测精度的同时难以保证较低的模型复杂度,进而影响推理速度,有时甚至无法完成实时检测工作。

针对以上存在的问题,本文提出了一种基于改进 Shuffle-RetinaNet 的红外场景车辆检测算法,选用轻量化网络 ShuffleNetV2 作为特征提取网络,大大降低了网络复杂度。提出双分支注意力模块(DBAM),通过双分支结构和自适应融合方法突出图像中更为关键的车辆特征,增强网络特征提取能力;在原有特征金字塔网络(FPN)中额外增加自底向上的特征聚合路径以丰富车辆目标多尺度特征信息。在训练过程中设置校准因子来度量分类和定位任务的校准程度,更严格地控制分类和回归质量以提升检测效果。该算法可以满足红外场景下车辆检测任务需要,能够实现快速精准的车辆检测。

2 算法设计与原理

2.1 算法整体结构

RetinaNet 算法是由 Lin 等^[9]提出的单阶段目标检测网络,其网络结构灵活,检测精度超越经典双阶段网络,在目标检测领域极具代表性,因而将其作为所提算法的基础网络。然而 RetinaNet 模型复杂度较高,限制了其在红外车辆检测领域的广泛应用。基于此,引入 ShuffleNetV2 作为改进网络模型的主干网络,大幅度降低了模型的复杂度。基于 Shuffle-RetinaNet 改进后的算法整体结构图如图 1 所示。针对红外场景下目标

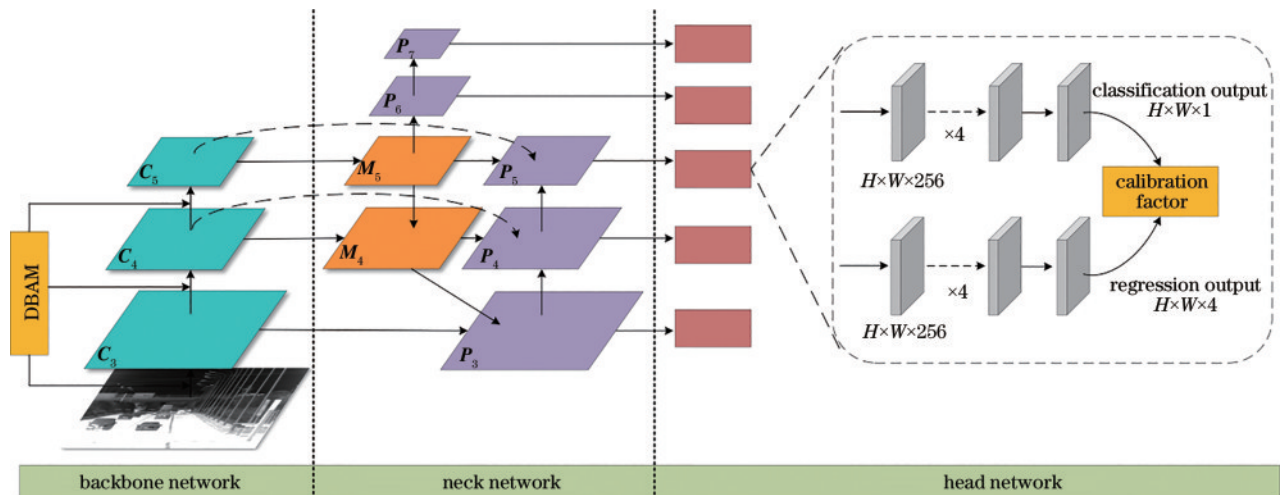


图 1 基于改进 Shuffle-RetinaNet 的红外车辆检测算法整体结构

Fig. 1 Overall architecture of infrared vehicle detection algorithm based on improved Shuffle-RetinaNet

特征信息稀缺的问题,综合考量模型复杂度与性能,设计了 DBAM 通道注意力模块并添加到 ShuffleNetV2 主干网络的 ShuffleBlock 瓶颈层中,以提高网络对红外场景车辆的关键特征筛选和提取能力。

主干网络提供了更为充足的车辆关键特征信息,但在实际检测应用中目标不断运动,距离远近不一,算法对多尺度目标特征的表达还需进一步增强。针对此问题,在原有特征融合基础上增加了自底向上的聚合路径来进一步增强尺度特征表达,并在部分特征层同一尺度的原始输入与最终输出节点之间添加横向连接以融合更多特征。图 1 中,特征融合网络的输入特征层为 C_3, C_4, C_5 ,通过改进融合后生成 P_3, P_4, P_5 ,同时对 M_5 特征层进行两次步长为 2 的 3×3 卷积得到高语义信息的 P_6, P_7 ,最终形成了颈部网络的输出。

检测网络部分如图 1 虚线框所示,由两条分支构成,分别用于分类预测任务和回归预测任务,两分支输入特征图均为 $H \times W \times 256$ 。分类分支网络利用 5 次 3×3 卷积操作得到 $H \times W \times 1$ 的特征图,统计了每个采样点的分类分数。回归分支网络同样利用 5 次 3×3 卷积操作得到 $H \times W \times 4$ 的特征图,统计了每个采样点关于位置信息的四维向量。针对检测阶段分类和回归任务缺乏信息交互的问题,在训练过程中设置校准因子作为分类分数和回归分数的高阶组合,明确度量两个任务的校准程度。将校准因子集成到样本分配策略和原有的损失函数中,促进分类损失和回归损失在训练过程中的同步优化。

2.2 通道注意力模块

与可见光相比,红外图像的分辨率较低,图像目标信号弱,因此从较少的特征中筛选出有效的特征信息至关重要。目前很多目标检测算法通过引入通道注意力机制^[10],帮助网络进行资源分配。基于此,设计了

DBAM 并通过全局平均池化和全局最大值池化并行的策略,综合利用红外图像的全局信息和局部信息。同时,采用双分支自适应融合的策略动态调整两条分支的通道权重。在 ShuffleNetV2 的 ShuffleBlock 瓶颈层中加入了 DBAM 以增强网络的特征提取能力。

全局平均池化操作可以很好地提取全局信息,却难以兼顾局部关键信息,而全局最大值池化可以很好地弥补该缺陷,因此采用双分支结构,在注意力模块中以两种不同池化方式设置两条支路以更全面地聚合图像中的特征信息。借鉴 Efficient channel attention (ECA) 思想,采用动态一维卷积生成通道权值,避免全连接层带来的大量参数,同时更好地完成跨通道间的信息交互^[11-12]。设计了双分支自适应融合的方法,动态地调整两种池化分支在融合中所占权重比。其具体实现方式如图 2 所示,首先对高、宽和通道数分别为 H, W, C 的输入特征在通道维度上进行全局平均池化和全局最大值池化操作,再通过卷积核大小为 k 的快速一维卷积生成两分支的通道权值矩阵 M_{avg} 和 M_{max} 。两通道权值矩阵通过自适应融合结构完成通道权值汇总,并与原输入特征经过逐元素乘积,将权值映射到 $H \times W \times C$ 的特征图中。

图 2 中,卷积核大小 k 代表着跨通道交互的范围,其与 C 之间存在非线性的映射关系。遵循 ECA 原文思想将 γ 和 b 设置为 2 和 1,从而保证网络最好的跨通道信息交互效果。

$$k = \psi(C) = \left\lfloor \frac{\text{lb } C}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}}, \quad (1)$$

式中: $\psi(C)$ 为 k 与 C 之间的非线性关系函数; $\lfloor \cdot \rfloor_{\text{odd}}$ 为最邻近奇数; γ, b 为此非线性关系中的 2 个可调参数,用以保证卷积核大小 k 与通道 C 之间的最佳匹配关系。

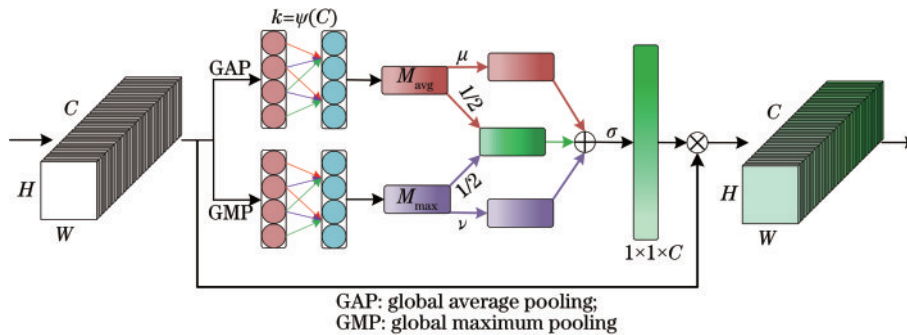


图 2 DBAM 模块结构图

Fig. 2 Structure of DBAM

两条特征分支通过快速一维卷积生成通道注意力权值 M_{avg} 和 M_{max} ,并采用自适应融合方法得到最终汇总的通道权值矩阵 M_F :

$$M_F = \delta \left[\frac{1}{2} (M_{\text{avg}} \oplus M_{\text{max}}) \oplus \mu M_{\text{avg}} \oplus \nu M_{\text{max}} \right], \quad (2)$$

式中: δ 为 Sigmoid 激活函数; μ, ν 为 2 个浮点类可学习

参数,随网络模型动态学习,初始均值设为 1,为防止两参数衰减为 0,在融合中加入两支路各 1/2 的权重以防止特征信息丢失; \oplus 为矩阵逐元素相加。汇总后的通道权重与原输入特征 F 逐元素相乘,最终得到经过通道注意力模块校正后的新特征 F' :

$$F' = M_F \otimes F, \quad (3)$$

式中： \otimes 为矩阵逐元素相乘。

与其他经典注意力机制相比，采用双池化分支更好地聚合了图像中全局和局部关键特征，丰富了图像信息。借鉴ECA思想采用一维动态卷积生成通道权值，以更少的参数量实现了局部跨通道的信息交互。自适应融合的双分支结构可以动态地调整两种池化的信息权重，使网络能够自适应地采纳不同分支对图像的信息关注度。通过添加DBAM注意力模块，增强了网络对于红外图像关键特征信息的提取效果，弥补了ShuffleNetV2特征提取能力不足的缺陷，更有助于后续车辆检测任务。

2.3 特征融合网络

在红外场景下的车辆检测任务中，车辆目标以及成像设备通常处于相对运动状态，时远时近，成像

目标尺度不一，这对网络的多尺度检测性能提出了更高的要求。目前经典的多尺度车辆检测算法，构建思路通常围绕FPN结构，如图3(a)所示，其在一定程度上可以融合不同特征层之间的信息以增强特征表达。然而，在红外场景下车辆目标的特征信息较少且各尺度特征信息不一致，仅靠一条自上而下的单向信息流难以有效地增强各尺度特征。PANet在FPN的基础上额外增加了一条自底向上的路径用以特征增强，如图3(b)所示，但是对所有节点不加区分的聚合会增加颈部网络的计算负担，且融合效率较低^[13]。另一方面，在调整特征图大小的过程中，传统的上采样方法（例如均值插值法等）无法很好地表达图像的细节信息，丢失了红外图像目标更为细腻的特征。

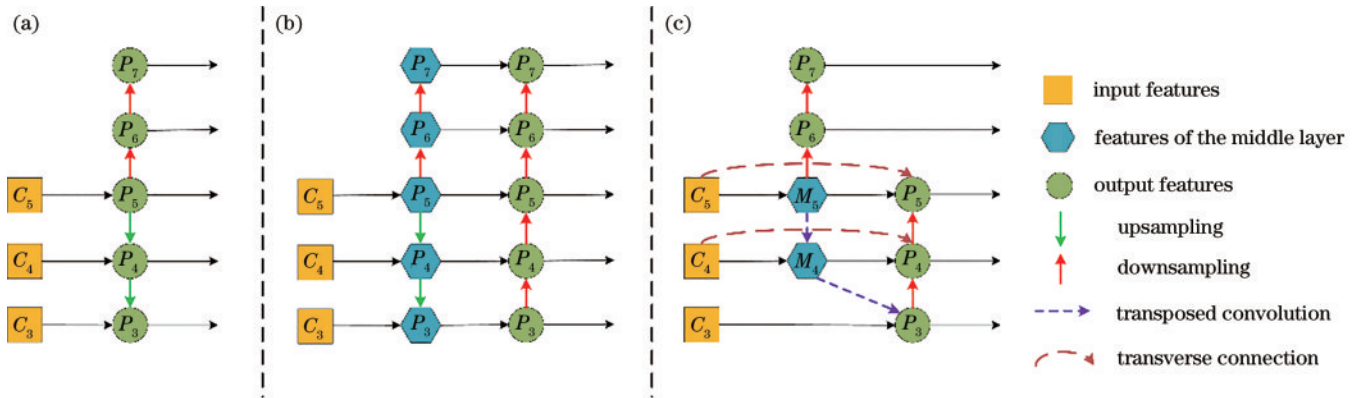


图3 不同的特征融合网络。(a)传统FPN；(b)PANet；(c)本文改进网络

Fig. 3 Different feature network design. (a) Conventional FPN; (b) PANet; (c) our network design

针对以上问题，研究了红外图像车辆特征以及多尺度特征层间关系，受BiFPN^[14]启发，在部分特征层额外增加了自底向上的路径来增强特征，并在同一尺度特征层的原始输入与最终输出节点之间添加横向连接，以少量的计算开支为代价融合更多的当前层特征。另一方面，在调整特征图大小的过程中，将原有的上采样过程全部替换为转置卷积^[15]，以达到更充分还原目标细节特征的效果。最终特征融合网络结构如图3(c)所示， C_3 、 C_4 、 C_5 为ShuffleNetV2网络的输出特征，3层特征图全部经过 1×1 卷积进行通道变换，而后经过转置卷积和层间相加操作生成 P_3 、 M_4 和 M_5 。同时，由 M_5 连续进行两次步长为2的 3×3 卷积操作生成 P_6 和 P_7 ，从 P_3 开始增加自底向上的聚合路径，并与同尺度层输入横向连接，最终生成 P_4 、 P_5 。

BiFPN中指出，只有一条输入边的节点对网络贡献较小，因此为了减少颈部网络计算量和提高融合效率，仅针对特征信息较为丰富的 P_4 、 P_5 两特征层进行双向交叉的特征聚合。不同的输入特征具有不同的分辨率，其对输出特征的贡献不同，使用快速归一化融合，为每个特征输入添加额外权重，这对特征信息稀缺的红外图像非常适用。以第4层为例，基于快速归一化

融合的特征融合可表示为

$$M_4 = \text{Conv} \left[\frac{w_1 \times C_4 + w_2 \times \text{Resize}(M_5)}{w_1 + w_2 + \epsilon} \right], \quad (4)$$

$$P_4 = \text{Conv} \left[\frac{w'_1 \times C_4 + w'_2 \times M_4 + w'_3 \times \text{Resize}(P_3)}{w'_1 + w'_2 + w'_3 + \epsilon} \right], \quad (5)$$

式中： Conv 为卷积操作； w_1 、 w_2 、 w'_1 、 w'_2 、 w'_3 为快速归一化融合中赋予各特征输入的额外权重； C_4 为主干网络的第4层输出特征层； M_5 、 M_4 分别为特征融合网络第5层和第4层的中间特征； P_4 、 P_3 分别为特征融合网络第4层、第3层的输出特征，其中自顶向下路径中的 Resize 通过转置卷积来实现，同时设置固定值参数 $\epsilon = 0.0001$ 来保证融合过程中数值的稳定。

与原有的FPN网络相比，优化后的特征融合网络额外增加了自底向上的聚合路径，更好地增强了目标尺度特征。综合考量红外图像特征信息稀缺和各尺度特征信息不均衡的问题，在 P_4 、 P_5 层集成了双向交叉尺度连接和快速归一化融合，更好地权衡了特征融合效果和计算成本之间的关系，兼顾轻量化的同时实现了多尺度特征的进一步强化。同时，转置卷积能更深层地挖掘红外图像的像素特征，加强图像细粒度识别，进

一步弥补红外车辆目标模糊、分辨率低的不足。

2.4 校准因子

目标检测工作通常被设计为两个独立分支,其中分类任务旨在学习目标区别于其他物体的关键、显著部分的特征,而回归任务着眼于精确定位整个待识别对象及其边界。双分支可能会导致两个任务之间缺乏交互,在训练的过程中分类损失和回归损失无法朝着最优方向同步优化,两任务所得到的最佳锚框不一致,且会因目标的形状多样、目标遮挡等问题出现较大差异^[16]。如图 4 所示,以真值框作为参考,分类置信度最高的锚框与真值框的交并比(IoU)非常低,然而回归效果较好的锚框,其分类置信度又很低。由此可以看出这种分类和回归之间任务的不一致严重影响了模型的检测效果。同时,常用的非极大值抑制(NMS)只保留重叠锚框中分类置信度最大的框,忽略定位精度^[17],但事实上高质量的锚框不仅意味着精确的分类,还意味着精准的定位^[18-19],两任务的预测结果应该是接近一致的。因此,训练样本分配中也应该更为关注那些分类和定位一致性高的锚框。

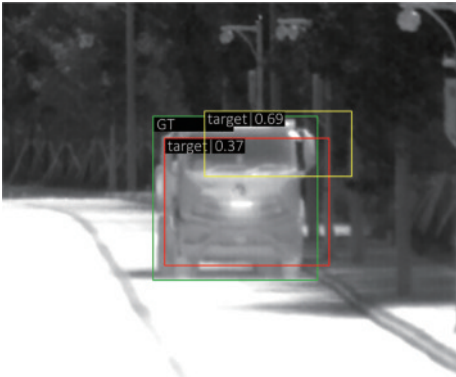


图 4 分类回归任务不一致问题

Fig. 4 Inconsistency of classification and regression

红外车辆检测应用于实际场景,例如导航、跟踪乃至军事侦察任务,通常对目标检测精度有着很高的要求。然而,红外图像中的车辆目标形状不规则且易遮挡,很大程度上影响了边界框回归与分类两任务预测结果的一致性,这导致推理阶段的目标检测精度难以满足要求。为更好地解决此类问题,所提算法针对检测器部分设置了校准因子来明确度量两个任务的校准程度,并将校准因子集成到样本分配策略和损失函数两个部分中,以动态细化每个锚框上的预测。校准因子的计算过程可表示为

$$t = s^\alpha \times u^\beta, \quad (6)$$

式中: t 为计算得出的校准因子; s 、 u 分别为锚框的分类得分和 IoU,设置 α 和 β 来控制两个任务得分对校准因子的影响。在样本分配策略中,选择 t 最大的 m 个锚框作为正样本,其余锚框作为负样本,通过此分配策略来关注两任务一致性更高的样本框。在损失函数的调

整中,以 focal loss 作为分类损失来缓解正负样本不均衡情况,并在此基础上用 t 替换原有的正样本锚框标签值来提高校准度高的锚框分类分数,最终分类损失 L_{cls} 可表示为

$$L_{cls} = \sum_{i=1}^{N_{pos}} |t_i - s_i|^\gamma \text{BCE}(s_i, t_i) + \sum_{j=1}^{N_{neg}} s_j^\gamma \text{BCE}(s_j, 0), \quad (7)$$

式中: L_{cls} 为最终得到的分类损失;BCE为二分类交叉熵损失函数; i 为每个实例对应的 N_{pos} 个正样本中的第 i 个锚框; j 为 N_{neg} 个负样本中的第 j 个锚框; s_i 、 s_j 均为预测值; t_i 为由第 i 个锚框计算得到的校准因子值; γ 为聚焦参数。

与分类损失类似,基于 t 对每个锚框的回归损失进行重新加权,得到最终回归损失 L_{reg} , L_{reg} 可表示为

$$L_{reg} = \sum_{i=1}^{N_{pos}} t_i L_{GIoU}(b_i, b'_i), \quad (8)$$

式中: L_{GIoU} 为GIoU回归损失; b_i 、 b'_i 分别为预测边界框和真实框。

经实验表明,取样参数 m 的波动对模型检测最终结果影响极小,根据实际场景设置为9,而校准因子参数 α 和 β 设置为1.0与8时模型的目标检测效果最好,具体实验结果见3.3节。通过引入校准因子,从样本分配和损失计算2个角度优化检测器。将校准程度作为样本分配依据,改善了传统NMS对定位精度的忽略问题。同时,作为分类和回归计算的高阶组合,校准因子鼓励网络在训练过程中从联合优化的角度来评价样本框质量,很大程度上改善了双分支检测任务缺乏交互的问题,使网络在训练过程中更严格地控制预测框质量,从而提升模型的检测效果。

3 实验与结果分析

3.1 数据集及实验环境

红外车辆检测在军事领域具有研究价值,大型红外数据集较为稀缺,实验数据由长波红外军工摄像头拍摄,部署在顶楼及卡车顶部等位置分别拍摄天桥以及街道等不同位置的车辆目标,数据涵盖了不同时间段及天气情况以保证外界环境的客观多样性,同时包含了不同角度和尺寸的车辆目标。其中,为便于区分目标尺度,将目标检测框面积占图像面积比小于2%的目标定为小目标,占比大于2%且小于5%为中目标,占比大于5%定为大目标。基于以上数据,根据MS COCO数据集格式创建了图像信息丰富的红外车辆检测数据集,数据集来源于112段红外场景视频,共包含10807张图片,图片为JPEG格式,分辨率为640 pixel×512 pixel,位深为24 bit。整个数据集中共包含小目标数量20187个,中目标数量18663个,大目标数量13884个。其中,训练集8213张,测试集2594张。训练集和测试集中各自包含了不同场景以更好地验证模型的泛化能力。图5为数据集中的部分图像。

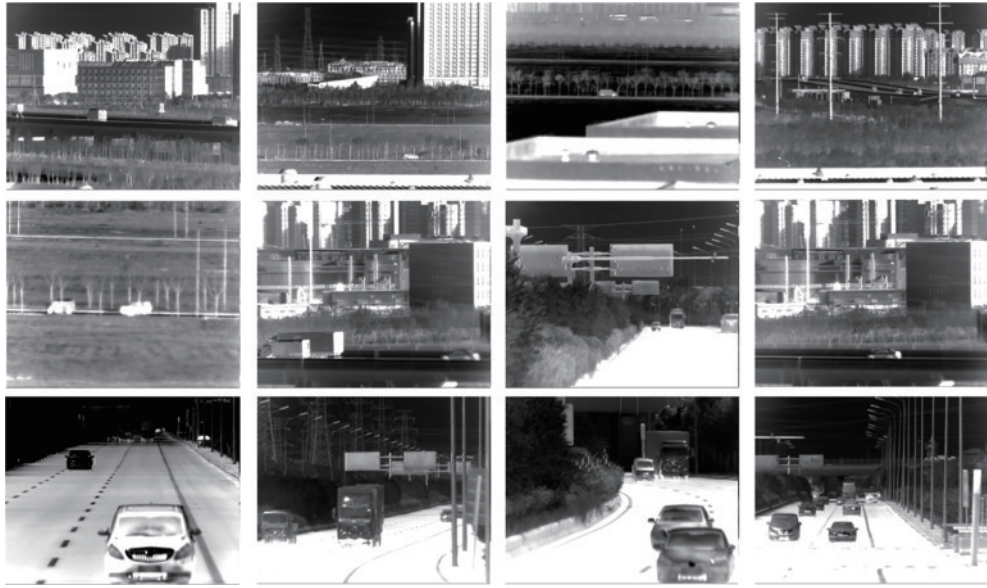


图 5 数据集中部分红外车辆图像
Fig. 5 Partial infrared vehicle images in the dataset

实验平台配置如表 1 所示。训练过程中采用随机梯度下降优化器,训练周期数为 12,初始学习率为 0.001,批处理数为 4,学习率在第 8 个和第 11 个训练周期衰减为当前学习率的 10%,权重衰减和动量分别设置为 0.0001 和 0.9。此外,在训练过程中对输入的红外图像进行随机翻转、缩放、附加高斯噪声以提高模

表 1 实验平台配置

Table 1 Experimental platform configuration

Platform and tools	Version
Operating system	Ubuntu 16.0.7
Graphic processing unit	NVIDIA Quadro RTX 8000
Graphics driver	510.47.03
Programing language	Python 3.7
Framework	PyTorch 1.7.1

表 2 轻量级主干网络对比

Table 2 Comparison of lightweight backbones

Backbone	Number of parameters / 10^6	FLOPs / 10^9	Speed / (frame \cdot s $^{-1}$)	AP ₅₀ / %
ResNet50	36.10	65.35	14.2	87.4
ShuffleNetV1 1.0 \times (g=3)	11.91	27.32	29.8	84.9
ShuffleNetV2 1.0 \times (g=3)	11.15	24.09	32.4	86.9
MobileNetV2	14.01	26.97	28.4	85.1
MobileNetV3	14.98	24.03	30.3	86.1

由表 2 可知,在 RetinaNet 基础上将主干网络 ResNet50 替换为轻量级网络后,网络模型参数量和计算量明显减少,推理速度得到较高的提升。将 4 种轻量级网络进行对比,其中 ShuffleNetV2 的模型复杂度达到最低 (11.15×10^6),与原主干网络 ResNet50 相比降低近 25×10^6 。其推理速度达到 $32.4 \text{ frame} \cdot \text{s}^{-1}$,同样快于其他网络,计算量为 24.09×10^9 ,仅比

型的鲁棒性。

3.2 轻量级主干网络对比

使用主流的轻量级主干网络 ShuffleNetV1、ShuffleNetV2、MobileNetV2、MobileNetV3 替换 RetinaNet 原主干网络 ResNet50,在自建的红外车辆数据集上进行对比实验来衡量各主干网络的综合性能,以选出最适用于本课题研究的网络。实验结果如表 2 所示,以模型参数量、浮点计算量(FLOPs)、推理速度和检测精度作为综合评价指标。其中,模型参数量表示整个网络模型中包含的参数数量,直接决定模型所占内存大小;浮点计算量为计算模型中加法和乘法的运算次数,用以反映模型对计算资源的消耗;推理速度用每秒内算法检测图像帧数来表示;检测精度 AP₅₀ 表示将 IoU 阈值设置为 0.5 时对应的平均精度值,能较好地反映模型的多尺度检测效果。

MobileNetV3 高了 0.06×10^9 ,检测精度 AP₅₀ 达到 86.9%,高于其他轻量级网络。综合考虑模型复杂度、推理速度与检测精度,ShuffleNetV2 更加满足红外车辆检测需求,与其他轻量级网络相比具有一定优势。

3.3 消融实验

设计了消融实验进行自对比,实验中以 RetinaNet 和 Shuffle-RetinaNet(将 RetinaNet 原主干网络替换为

ShuffleNetV2)作为实验对比的标准模型,以 COCO 评价标准中的 AP_{50} 作为模型主要评价指标,同时以分别针对小目标、中目标和大目标 AP 值的 AP_s 、 AP_m 、 AP_l 等指标来评价网络针对不同尺度车辆目标

的检测效果,以模型参数量、浮点计算量和检测帧率来反映模型的复杂度和推理速度。表 3、表 4 分别记录了整体消融实验结果和多尺度检测精度消融实验结果。

表 3 整体消融实验结果

Table 3 Overall ablation experimental results

Method	DBAM	Improved feature network	Calibration factor	$AP_{50} / \%$	Number of parameters / 10^6	FLOPs / 10^9	Speed / (frames·s ⁻¹)
RetinaNet				87.4	36.10	65.35	14.2
Shuffle-RetinaNet				86.9	11.15	24.09	32.4
	✓			89.1	11.44	24.21	31.7
	✓	✓		91.1	11.74	24.35	31.1
Ours	✓	✓	✓	92.9	11.74	24.35	30.9

表 4 多尺度检测精度消融实验结果

Table 4 Ablation study results of multi-scale detection accuracy

Method	DBAM	Improved feature network	Calibration factor	$AP_s / \%$	$AP_m / \%$	$AP_l / \%$
RetinaNet				45.2	59.5	66.5
Shuffle-RetinaNet				42.9	55.7	64.7
	✓			43.1	55.9	65.3
	✓	✓		44.9	58.4	65.9
Ours	✓	✓	✓	45.7	59.9	66.7

为了验证添加注意力模块改进 ShuffleNetV2 网络结构后模型性能的提升,针对 DBAM 模块部分进行了消融实验。如表 3 所示,在主干网络中采用 DBAM 模块后,网络检测精度 AP_{50} 从 86.9% 提升到了 89.1%,超过了原基础网络 RetinaNet 的检测精度,而网络模型的参数量仅增加了 0.29×10^6 ,计算量也只增加了 0.12×10^9 ,推理速度较改进之前降低了 $0.7 \text{ frame} \cdot \text{s}^{-1}$,几乎不影响网络实时检测效果。实验结果表明,在 ShuffleNetV2 网络中添加 DBAM 模块仅为网络增加了很小的模型复杂度,但带来了较大的检测精度提升,证明双分支自适应融合结构可以更全面地获取全局和局部信息,增强有效特征的信息表达,能够弥补 ShuffleNetV2 对整体特征信息提取能力不足的缺陷,从而显著地提升检测效果。

针对改进后的特征融合网络进行消融实验,实验结果如表 3、表 4 所示。结果表明,与原有的传统 FPN 相比,改进后的特征融合网络的整体检测精度 AP_{50} 提高了 2.0 个百分点,模型参数量和计算量仅分别增加了 0.3×10^6 和 0.14×10^9 ,推理速度也仅降低了 $0.6 \text{ frame} \cdot \text{s}^{-1}$,说明改进后的特征融合网络可以有效提升对红外车辆目标的整体检测精度,且只带来较小的模型复杂度。表 4 结果显示,对特征融合网络进行改进后, AP_s 、 AP_m 、 AP_l 指标分别提高了 1.8 个百分点、2.5 个百分点和 0.6 个百分点。其中,中、小目标精度涨幅明显,大目标精度也有一定提升。结果表明:改进后的特征融合网络能够更好地融合目标特征信息,有效增强网络对多尺度特征的表达能力,从而提高多尺

度目标的检测效果。

在以上改进的基础上,引入校准因子,从样本分配和损失计算两个角度优化检测器。为探索出校准效果最好的校准因子参数 α 和 β ,在 $\alpha \in [0.5, 1.5]$, $\beta \in [6, 10]$ 两个区间进行间隔粗采样,不同参数设置的实验结果如表 5 所示。结果表明: α 、 β 分别为 1.0、8 时,校准因子对模型的检测效果提升最大,与引入校准因子之前相比, AP_{50} 得到了 1.8 个百分点的提升。图 6 为引入校准因子前后的检测效果对比,由图 6 可知,改进后的检测结果置信度有所提高,边界预测框更为准确,这说明校准因子在一定程度上能加强双分支检测任务间信息交互,进而提升检测效果。

表 5 校准因子对比实验结果

Table 5 Results of calibration factor comparison

Parameter value of calibration factor	$AP / \%$	$AP_{50} / \%$
$\alpha=0.5, \beta=6$	56.2	91.0
$\alpha=0.5, \beta=8$	56.2	91.1
$\alpha=0.5, \beta=10$	56.1	91.0
$\alpha=1.0, \beta=6$	56.9	92.4
$\alpha=1.0, \beta=8$	57.4	92.9
$\alpha=1.0, \beta=10$	57.4	92.5
$\alpha=1.5, \beta=6$	56.3	91.4
$\alpha=1.5, \beta=8$	57.1	92.0
$\alpha=1.5, \beta=10$	56.8	91.8

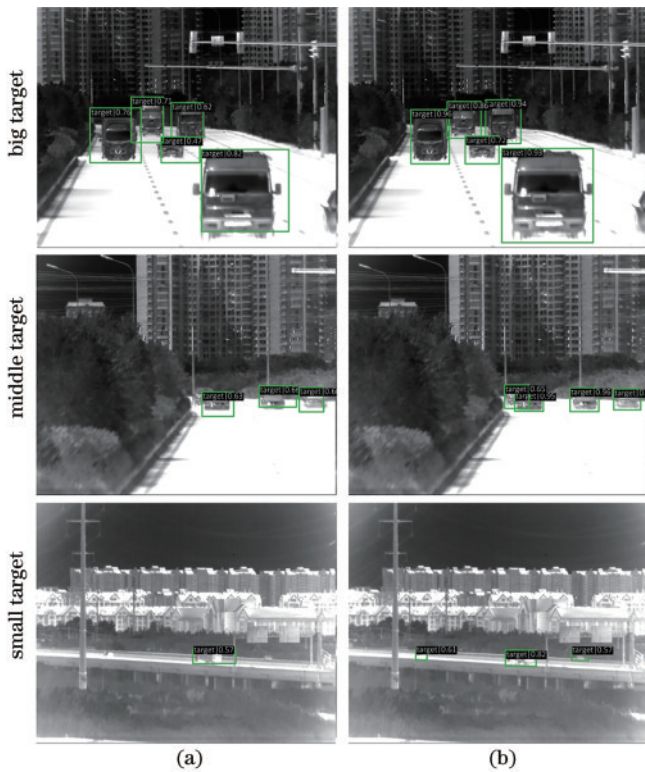


图6 引入校准因子检测效果对比。(a)改进前;(b)改进后
Fig. 6 Comparison of detection results before and after introducing calibration factor. (a) Before improvement; (b) after improvement

与其他场景下的各种类别目标相比,红外场景下的车辆目标具有其独有的特征共性(包括红外视野下的车辆轮廓、纹理特征、遮挡情况等),在检测任务中对两任务信息交互程度以及锚框对齐度量要求是相似的,因此对红外场景下的车辆检测任务而言,检测提升效果最好的校准因子不会随着待检测数据变化而变化。为了使用更充分的实验来验证探索出的最佳校准因子对红外车辆目标的泛化能力,基于 Shuffle-RetinaNet 基础网络,仍然在 $\alpha \in [0.5, 1.5]$, $\beta \in [6, 10]$ 两个区间内采样,在自建数据集以及红外公开数据集 FLIR ADAS 上进行车辆检测实验,以 AP_{50} 作为精度评价指标。实验结果如表 6 所示。

实验结果显示,在基础网络 Shuffle-RetinaNet 上,自建数据集和公开数据集的 AP_{50} 分别为 86.9% 和 87.6%。在调整校准因子过程中,两红外车辆数据集均在 α 、 β 分别取值 1.0、8 时达到最高检测精度,分别为 89.1%、89.7%,验证了最佳校准因子对红外车辆目标的泛化能力。

通过以上改进,所提网络检测精度超过了原始的 RetinaNet,而模型复杂度大幅度降低,推理速度提升明显。与 Shuffle-RetinaNet 相比,改进后的算法仅增加少量参数和计算量,检测精度得到明显提升。为更加直观地展示所提算法改进效果,将所提算法与 Shuffle-RetinaNet 进行红外车辆检测的效果比较,结

表 6 Shuffle-RetinaNet 上的校准因子对比实验

Table 6 Results of calibration factor comparison on Shuffle-RetinaNet

Parameter value of calibration factor	Shuffle-RetinaNet	
	Self-built dataset $AP_{50} / \%$	FLIR ADAS $AP_{50} / \%$
No calibration factor	86.9	87.6
$\alpha=0.5, \beta=6$	87.1	87.9
$\alpha=0.5, \beta=8$	87.2	87.8
$\alpha=0.5, \beta=10$	87.1	87.9
$\alpha=1.0, \beta=6$	88.9	89.5
$\alpha=1.0, \beta=8$	89.1	89.7
$\alpha=1.0, \beta=10$	89.0	89.5
$\alpha=1.5, \beta=6$	88.6	89.1
$\alpha=1.5, \beta=8$	88.7	89.2
$\alpha=1.5, \beta=10$	88.4	89.0

果如图 7 所示。由图 7 可知,所提算法能够很好地完成红外场景下多尺度车辆检测任务,在不同场景、不同角度的拍摄条件下,所提算法检测结果的置信度更高,定位准确度提升明显,与原始 RetinaNet 和 Shuffle-RetinaNet 相比,所提算法能更好地满足实际车辆检测任务需要。

3.4 算法对比与分析

为进一步验证所提算法的综合性能,在保持自建数据集、实验环境及参数设置相同的基础上,选择了几种常用于红外车辆检测的经典目标检测算法,并对每种算法的不同版本网络模型进行测试,挑选出综合性能最好的网络模型版本,与所提算法进行实验对比。实验具体围绕 AP_{50} 、参数量、浮点计算量和推理速度 4 个指标进行对比评估,对比实验结果如表 7 所示。所提算法的 AP_{50} 达到了 92.9%,且模型复杂度较低,推理速度较快,各项指标都超过了作为基础网络的 RetinaNet;与其他算法相比,所提算法仅在模型参数量以及推理速度上不及 YOLOv5s,且两指标差距均较小,但检测精度却高于其 4.6 个百分点。综合对比各项指标,所提算法表现出更强的综合性能。

为进一步验证所提算法对红外场景车辆检测任务的鲁棒性,在红外公开数据集 FLIR ADAS 上对所提算法进行测试,并与其他经典的红外检测算法进行对比,采用 AP 和模型参数量来衡量算法的检测性能,对比结果如表 8 所示。鉴于以车辆目标作为检测对象,表 8 中仅展示公开数据集中车辆检测效果,其他算法的实验数据均来源于原文献,其中“-”表示原文献中未统计相关数据。由表 8 可知,在公开的红外车辆数据上,所提算法的检测精度达到 91.7%,与经典的红外车辆检测算法相比有一定优势,且模型复杂度相对较低。对比结果表明,所提算法鲁棒性较强,在公开红外数据上其同样能很好地胜任车辆检测工作。

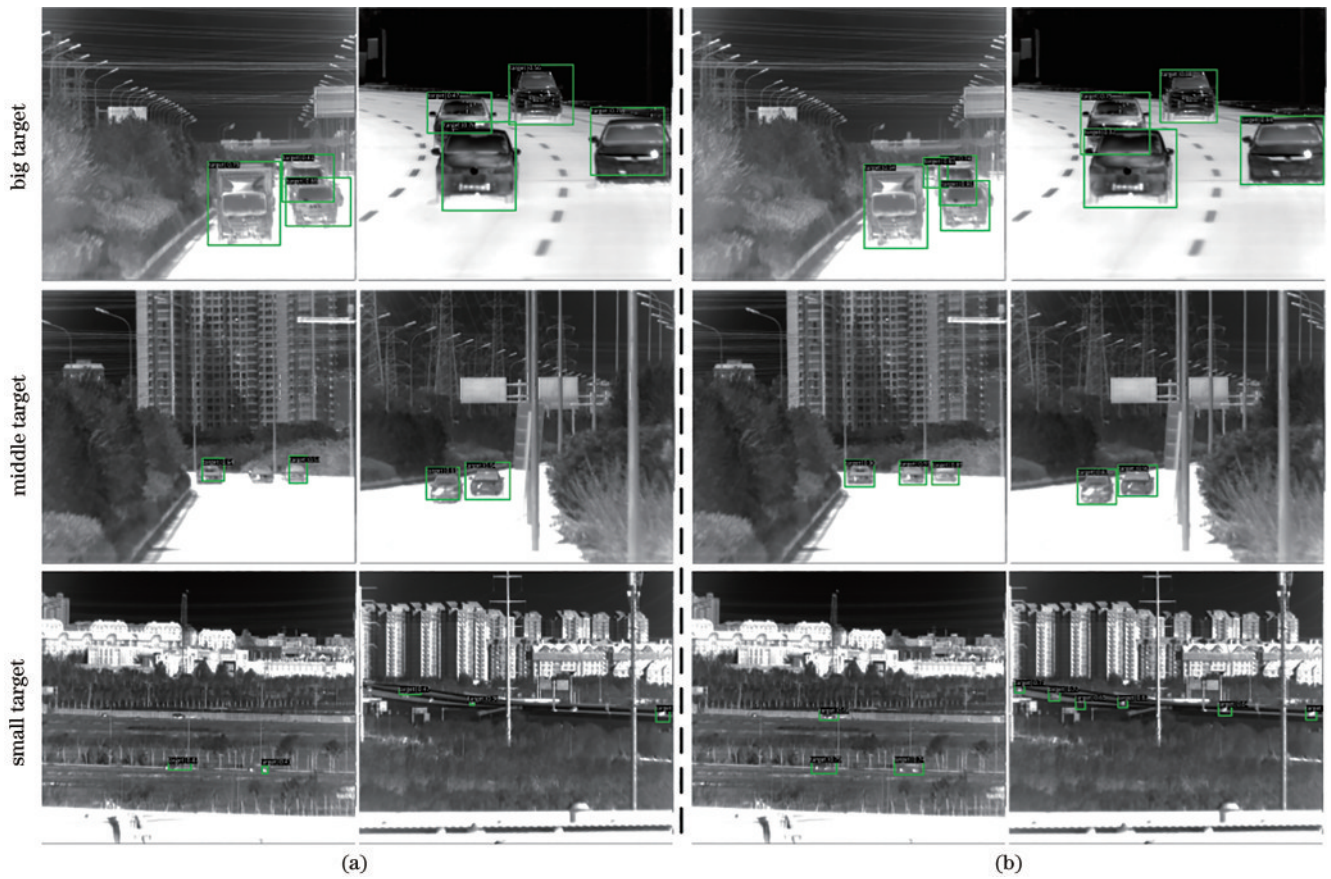


图 7 Shuffle-RetinaNet在改进前后的检测效果对比。(a)改进前;(b)改进后

Fig. 7 Comparison of detection effect of Shuffle-RetinaNet before and after improvement. (a) Before improvement; (b) after improvement

表 7 与经典目标检测算法的对比实验结果

Table 7 Comparison with classical object detection algorithms

Method	AP ₅₀ / %	Number of parameters / 10 ⁶	FLOPs / 10 ⁹	Speed / (frames ⁻¹)
Faster RCNN	88.6	41.12	67.75	12.4
RetinaNet	87.4	36.10	65.35	14.2
YOLOv5s	88.3	7.21	26.67	31.6
SSD512	86.2	36.04	60.85	18.1
Ours	92.9	11.74	24.35	30.9

表 8 与经典红外车辆检测算法的对比实验结果

Table 8 Comparison with classical infrared vehicle detection algorithms

Method	AP / %	Number of parameters / 10 ⁶
Algorithms 1 ^[20]	76.57	20.60
Algorithms 2 ^[6]	85.00	70.53
Algorithms 3 ^[21]	91.30	—
Algorithms 4 ^[22]	90.50	8.10
Ours	91.70	11.74

4 结 论

提出了一种基于改进 Shuffle-RetinaNet 网络的红

外车辆检测算法。以 ShuffleNetV2 作为特征提取网络,有效降低了模型的复杂度,提升了网络的轻量化效果;设计了 DBAM 模块,改进 ShuffleNetV2 的 ShuffleBlock 瓶颈层,增强了其对红外车辆目标关键特征信息的提取能力,提升了检测效果;改进了特征融合网络结构,采用双向交叉尺度连接和快速归一化融合方法,增强了多尺度目标特征的表达能力。设置了校准因子,改善了双分支检测任务间缺乏交互的问题,进一步提升了检测效果。实验结果表明:所提算法具有较好的检测效果和较高的推理速度,在红外场景下检测车辆具有一定的优势。未来将在算法中引入车辆的时序信息,利用车辆的运动特征进一步提升算法的检测效果。

参 考 文 献

- [1] Xue T, Zhang Z Q, Ma W N, et al. Nighttime pedestrian and vehicle detection based on a fast saliency and multifeature fusion algorithm for infrared images[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(9): 16741-16751.
- [2] 崔紫薇. 红外图像目标检测方法研究[D]. 长春: 长春理工大学, 2022: 36-45.
Cui Z W. Research on target detection method of infrared

- image[D]. Changchun: Changchun University of Science and Technology, 2022: 36-45.
- [3] Lin J, Tan Y H, Tian J W. Inter-frame correlation based on moving vehicle target detection in infrared image sequences[C]//2017 International Conference on Computer Technology, Electronics and Communication (ICCTEC), December 19-21, 2017, Dalian, China. New York: IEEE Press, 2019: 1187-1191.
- [4] Zhao H X, Liang Z R, Cai D H, et al. An improved method for infrared vehicle and pedestrian detection based on YOLOv5s[C]//2022 International Conference on Machine Learning, Cloud Computing and Intelligent Mining (MLCCIM), August 5-7, 2022, Xiamen, China. New York: IEEE Press, 2022: 377-383.
- [5] Li H, Yuan J, Liu H L, et al. Incremental learning of infrared vehicle detection method based on SSD[C]//2020 IEEE 20th International Conference on Communication Technology (ICCT), October 28-31, 2020, Nanning, China. New York: IEEE Press, 2020: 1423-1426.
- [6] 李向荣, 孙立辉. 融合注意力机制的多尺度红外目标检测[J]. 红外技术, 2022, 44(6): 152-161.
Li X R, Sun L H. Multi-scale infrared target detection based on attention mechanism[J]. Infrared Technology, 2022, 44(6): 152-161.
- [7] 顾星, 詹伟达, 崔紫薇, 等. 基于注意力机制的红外目标检测方法[J]. 激光与光电子学进展, 2023, 60(10): 1011002.
Gu X, Zhan W D, Cui Z W, et al. Infrared target detection method based on attention mechanism[J]. Laser & Optoelectronics Progress, 2023, 60(10): 1011002.
- [8] 章琦, 朱鸿泰, 程虎, 等. 轻量级红外弱小目标检测算法[J]. 激光与光电子学进展, 2022, 59(16): 1611007.
Zhang Q, Zhu H T, Cheng H, et al. Lightweight infrared dim and small target detection algorithm[J]. Laser & Optoelectronics Progress, 2022, 59(16): 1611007.
- [9] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2999-3007.
- [10] 金梅, 李义辉, 张立国, 等. 基于注意力机制改进的轻量级目标检测算法[J]. 激光与光电子学进展, 2023, 60(4): 0415008.
Jin M, Li Y H, Zhang L G, et al. Attention mechanism-based improved lightweight target detection algorithm[J]. Laser & Optoelectronics Progress, 2023, 60(4): 0415008.
- [11] 丁子轩, 张娟, 李想, 等. 基于注意力引导的轻量级图像超分辨率算法[J]. 激光与光电子学进展, 2023, 60(14): 1410010.
Ding Z X, Zhang J, Li X, et al. A lightweight image super-resolution algorithm based on attention guidance[J]. Laser & Optoelectronics Progress, 2023, 60(14): 1410010.
- [12] Ma C, Liu S H, Zhang H G, et al. ECA-based drive mechanism for exception handling in single and small batch MES[C]//International Conference on Software Intelligence Technologies and Applications & International Conference on Frontiers of Internet of Things 2014, December 4-6, 2014, Hsinchu, Taiwan, China. London: Institution of Engineering and Technology, 2014: 221-226.
- [13] 张润梅, 毕利君, 汪方斌, 等. 多尺度特征融合与锚框自适应的目标检测算法[J]. 激光与光电子学进展, 2022, 59(12): 1215019.
Zhang R M, Bi L J, Wang F B, et al. Multi-scale feature fusion and anchor frame adaptive target detection algorithm[J]. Laser & Optoelectronics Progress, 2022, 59(12): 1215019.
- [14] Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10778-10787.
- [15] 朱子健, 刘琪, 陈红芬, 等. 基于并行融合网络的航拍红外车辆小目标检测方法[J]. 光子学报, 2022, 51(2): 0210001.
Zhu Z J, Liu Q, Chen H F, et al. Infrared small vehicle detection based on parallel fusion network[J]. Acta Photonica Sinica, 2022, 51(2): 0210001.
- [16] Wang K Y, Zhang L. Reconcile prediction consistency for balanced object detection[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2022: 3611-3620.
- [17] Gao Z T, Wang L M, Wu G S. Mutual supervision for dense object detection[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2022: 3621-3630.
- [18] Kim K, Lee H S. Probabilistic anchor assignment with IoU prediction for object detection[M]//Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12370: 355-371.
- [19] Li X, Wang W, Wu L, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection[M]//Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual. [S.l.: s.n.], 2020: 21002-21012.
- [20] 郭勇, 张凯. 基于特征增强的快速红外目标检测[J]. 无线电工程, 2023, 53(1): 47-55.
Guo Y, Zhang K. Fast infrared target detection based on feature enhancement[J]. Radio Engineering, 2023, 53(1): 47-55.
- [21] 赵明, 张浩然. 一种基于跨域融合网络的红外目标检测方法[J]. 光子学报, 2021, 50(11): 1110001.
Zhao M, Zhang H R. An infrared object detection method based on cross-domain fusion network[J]. Acta Photonica Sinica, 2021, 50(11): 1110001.
- [22] 李强龙, 周新文, 位梦恩, 等. 基于条形池化和注意力机制的街道场景红外目标检测算法[J]. 计算机工程, 2022, 44(7): 37-50.
Li Q L, Zhou X W, Wei M E, et al. Infrared target detection algorithm based strip pooling and attention mechanism in street scene[J]. Computer Engineering, 2022, 44(7): 37-50.