

# 基于注意力机制的红外目标检测方法

顾星, 詹伟达\*, 崔紫薇, 桂婷婷, 石艳丽, 胡家琿

长春理工大学电子信息工程学院, 吉林 长春 130022

**摘要** 由于红外探测器感应波段与可见光不同,不依赖于大气光的反射传播,而是取决于环境中物体自身散发的辐射强度,所以其在雾霾、夜晚等可见度低的条件下往往比可见光具有更好的目标检测效果。针对红外场景中目标检测精度低、实用性差的问题,提出一种基于注意力机制的红外目标检测方法。首先,设计一种轻量化网络结构;其次,采用注意力机制提高网络的特征提取能力;然后,改进迭代特征金字塔结构提高对不同尺度目标的检测能力;最后,在训练过程中引入 complete intersection over union(CIoU)损失函数和梯度均衡机制(GHM)损失函数改善正负样本不平衡问题。与其他算法的对比实验结果表明,所提算法的检测精度和速度显著提高。

**关键词** 成像系统;深度学习;卷积神经网络;目标检测;注意力机制;损失函数

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP213287

## Infrared Target Detection Method Based on Attention Mechanism

Gu Xing, Zhan Weida\*, Cui Ziwei, Gui Tingting, Shi Yanli, Hu Jiahui

School of Electronic Information Engineering, Changchun University of Science and Technology,  
Changchun 130022, Jilin, China

**Abstract** Because the sensing band of the infrared detector is different from the visible light, it does not depend on the reflection and propagation of atmospheric light, but depends on the radiation intensity emitted by the object itself in the environment, so it often has better target detection effect than the visible light under the conditions of low visibility such as haze and night. Aiming at the problems of low accuracy and poor practicality of target detection in infrared scene, an infrared target detection method based on attention mechanism is proposed. First, a lightweight network structure is designed; second, attention mechanism is used to improve the ability of network feature extraction; then, the iterative feature pyramid structure is improved to improve the detection ability of targets with different scales; finally, complete intersection over union (CIoU) loss function and gradient equilibrium mechanism (GHM) loss function are introduced in the training process to improve the imbalance of positive and negative samples. Compared with other algorithms, the experimental results show that the detection accuracy and speed of the proposed algorithm are significantly improved.

**Key words** imaging systems; deep learning; convolutional neural network; target detection; attention mechanism; loss function

## 1 引言

红外探测器与可见光探测器相比,在低照度等恶劣情况下具有更好的探测效果<sup>[1]</sup>。近些年,红外目标检测在安防监控<sup>[2]</sup>、车辆辅助驾驶<sup>[3]</sup>等民用领域及军事侦察<sup>[4]</sup>、实时监控<sup>[5]</sup>等军事领域均有着广泛应用。传统目标检测算法通过人工标注的算子进行特征提取,只能得到浅层次信息,在复杂场景中无法快速且准确地检测<sup>[6]</sup>。基于深度学习的目标检测通过卷积神经网络能更好地进行特征提取,检测精度和速度明显提升。现阶段

目标检测算法主要分为两类:双阶段目标检测与单阶段目标检测。其中,双阶段目标检测算法有 R-CNN<sup>[7]</sup>、SPP-Net<sup>[8]</sup>、Faster-CNN<sup>[9]</sup>等,保证了检测精度,但检测速度较低,单阶段目标检测算法有 SSD<sup>[10]</sup>和 YOLO<sup>[11]</sup>等,其对输入图像进行特征提取,直接回归目标类别概率和位置坐标值,检测速度得到提升,但精度不高。

目前针对红外目标检测的深度学习算法较少,由于红外图像的特性,直接使用通用的深度学习算法检测效果很差<sup>[12]</sup>。为解决上述问题,2019年,Ju等<sup>[13]</sup>提出一种基于交并比(IoU)的数学推导方式,为每个尺

收稿日期: 2021-12-20; 修回日期: 2021-12-23; 录用日期: 2022-02-14; 网络首发日期: 2022-02-24

基金项目: 吉林省发展与改革委员会创新能力建设专项(FG2021236JK)

通信作者: \*zhanweida@cust.edu.cn

度选择候选锚框数量和纵横比尺寸,提高了对不同尺度目标的检测能力。2020年, Lu等<sup>[14]</sup>建立一种基于水平连接融合的三层金字塔网络结构,解决了重叠目标的漏检问题。2021年,许虞俊等<sup>[15]</sup>提出一种基于YOLO系列的轻量化目标检测网络,该网络融合ghost模块并参考squeeze-and-excitation(SE)模块实现检测精度和速度的强平衡。2021年,蒋镛圻等<sup>[16]</sup>提出scSE-IYOLOv4目标检测算法,该算法通过在YOLOv4主干网络中嵌入scSE模块及在颈部加入空间金字塔池化(SPP)模块提高小目标检测精度。

上述工作从不同方向提升对红外图像的目标检测性能,但在自动驾驶等对检测速度要求高的领域以及需要更高检测精度来实现侦察的军事领域,检测精度和速度需要进一步提高。鉴于上述问题,本文以YOLOv3<sup>[17]</sup>为基础提出一种红外目标检测算法。设计一种轻量化网络作为主干网络,减小模型参数及计算量;在主干网络中加入注意力模块,提高整个网络特征提取能力;通过基于迭代特征金字塔结构的多尺度目标检测网络结构,提高网络对不同尺度目标的检测精度;引入complete intersection over union(CIoU)<sup>[18]</sup>和GHM-C<sup>[19]</sup>损失函数改善正负样本不平衡问题。实验结果表明,所提算法在红外场景中的检测速度和精度显著提高。

## 2 主干网络

### 2.1 注意力机制模块

人类视觉注意力机制可以在大量视觉信息中选出需要关注的目标区域,极大地提高了处理信息的效率和准确性。借鉴人类视觉的选择注意力机制,深度学习中的注意力机制聚焦感兴趣区域,抑制无用信息,特别是在目标检测领域,可以提高目标检测精度。所设计的并行注意力机制模块(PAM)如图1所示。

现有的注意力机制分为通道注意力和空间注意力。通道注意力关注全局信息,而空间注意力关注局部信息,两者互为补充<sup>[20]</sup>。文献[20]指出通道注意力的优先级略优于空间优先级,但是大量实验结果表明

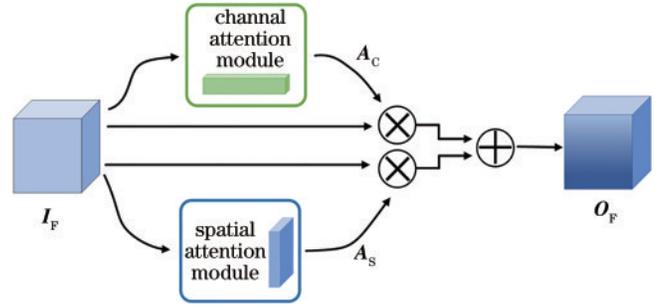


图1 PAM模块

Fig. 1 PAM module

该优先级会受注意力模块本身结构和主干网络结构的影响。本研究选择并行注意力机制,将通道与空间注意力设为相同优先级。将特征图分别输入通道与空间注意力得到两个特征图,相加得到最终特征图。将两个注意力结合得到的注意力机制模块能更有效地提取特征,进一步提高检测精度。

$$I'_F = A_C \times I_F, \quad (1)$$

$$I''_F = A_S \times I_F, \quad (2)$$

$$O_F = I'_F + I''_F, \quad (3)$$

式中: $I_F$ 为输入特征图; $\times$ 为元素乘法; $A_C$ 为通道注意力权重系数; $A_S$ 为空间注意力权重系数; $I'_F$ 为经过通道注意力的特征图输出; $I''_F$ 为经过空间注意力的特征图输出; $O_F$ 为特征图经过PAM模块的最终输出。

#### 2.1.1 通道注意力模块

通道注意力中通道维度不变,基于空间维度进行操作,能更好地关注每个通道上的特征,如图2所示。维度为 $C \times W \times H$ 的特征图分别经过最大池化和平均池化操作将维度压缩为 $C \times 1 \times 1$ ,使得一维特征图具有全局感受野,采用两种池化操作可以提取到更全面、丰富的高层特征。再依次进入卷积核为 $1 \times 1$ 、 $3 \times 3$ 、 $1 \times 1$ 的3个卷积块中,经过通道数压缩、特征提取、通道数恢复等一系列操作,在计算量较小的前提下,完成对高层特征的进一步提取。最终将输出相加并使用Sigmoid激活,得到 $A_C$ 。通过与初始特征图相乘,完成在通道维度上的特征重标定。

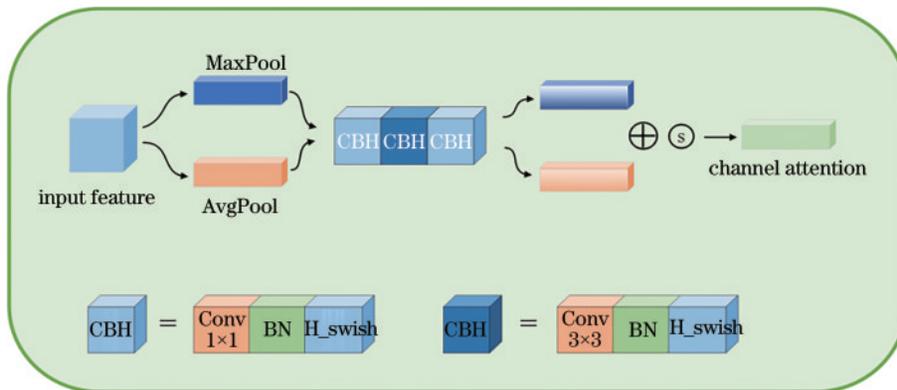


图2 通道注意力模块

Fig. 2 Channel attention module

2.1.2 空间注意力模块

空间注意力模块保持空间维度不变,基于通道维度进行操作,能更好地关注空间维度特征,如图 3 所示。特征图基于通道分别进行平均池化和最大池化操作,避免单一池化丢失过多信息。维度为  $C \times W \times H$

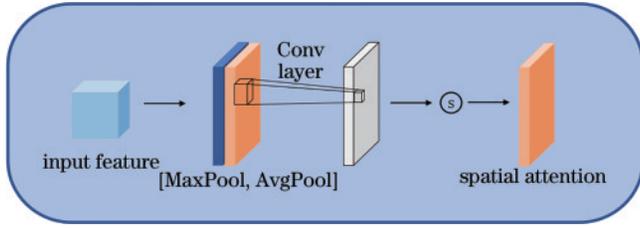


图 3 空间注意力模块

Fig. 3 Spatial attention module

的特征图经过池化后变为两个维度为  $1 \times W \times H$  的特征图,仅用一个通道的特征图表示全局特征,得到语义信息丰富的特征。然后通过基于通道的拼接操作变为  $2 \times W \times H$  的特征图。最后用 1 个卷积核进行卷积操作,将通道数重新压缩为 1 并使用 Sigmoid 激活得到  $A_s$ ,与初始特征图相乘,完成空间维度上的特征重标定。

2.2 主干网络

本研究设计了一种瓶颈模块 Pbottleneck,其具体结构如图 4 所示,引入残差结构,采用  $1 \times 1$  卷积层扩张通道数、 $3 \times 3$  卷积层提取特征、 $1 \times 1$  卷积层压缩通道数的操作,减小计算量和参数量,同时加入 PAM 注意力机制模块来提高网络的特征提取能力。

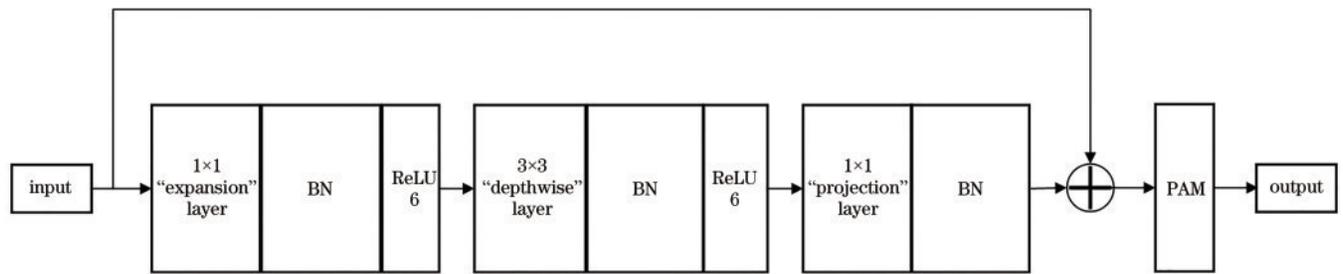


图 4 Pbottleneck 模块

Fig. 4 Pbottleneck module

为提高检测速度和精度,以 MobileNet V2<sup>[21]</sup>为基础设计了一种轻量化网络作为主干网络,网络结构如表 1 所示,其中  $t$  是输入通道的倍增系数, $n$  是模块重复次数, $c$  是输出通道数, $s$  是步长。该主干网络可在减小计算量和参数前提下,提高网络的特征提取能力。

表 1 主干网络结构

Table 1 Backbone network structure

Input	Operator	$t$	$c$	$n$	$s$
$224^2 \times 3$	Conv2D3×3		32	1	2
$112^2 \times 32$	Pbottleneck	1	16	1	1
$112^2 \times 16$	Pbottleneck	6	24	2	2
$56^2 \times 24$	Pbottleneck	6	32	3	2
$28^2 \times 32$	Pbottleneck	6	64	4	1
$14^2 \times 64$	Pbottleneck	6	96	3	2
$14^2 \times 96$	Pbottleneck	6	160	3	2
$7^2 \times 160$	Pbottleneck	6	320	1	1
$7^2 \times 320$	Conv2D1×1		1280	1	1
$7^2 \times 1280$	AvgPool7×7			1	
$1 \times 1 \times 1280$	Conv2D1×1		1280		

3 基于多注意力机制的红外目标检测算法

所提红外目标检测算法的总体网络框架如图 5 所示,解码器网络用于将主干网络提取的特征进行解码,

得到预测框和分类概率的回归值。采用轻量化网络作为主干网络,并加入注意力机制模块提高网络的特征提取能力;将多尺度检测方式改为迭代特征金字塔网络 (Recursive-FPN) 并将检测层从 3 层增加到 4 层,通过增加下采样 4 倍的检测层来检测更小目标(小目标是指目标像素点小于  $32 \times 32$  或目标大小小于原始图像尺寸的 0.1 的目标);使用 CIoU 损失函数作为位置损失函数加快模型收敛,提高预测框的准确率,同时加入 GHM-C 损失函数作为分类损失函数解决正负样本比例不均衡的问题。

3.1 多尺度检测结构

现实场景中,图像不可能只含有单一尺度的目标。为提高不同尺度的检测能力,常采用多尺度特征检测。通过不同的特征金字塔层提供不同尺度、不同分辨率图像的特征信息。通过特征金字塔进行图像融合,提取目标特征信息,充分挖掘图像潜力。对迭代特征金字塔结构 (RFP)<sup>[22]</sup> 进行改进实现多尺度检测,将检测层由 3 层增加到 4 层。RFP 是在 FPN 的基础上实现的,将传统 FPN 融合后的输出,经过反馈结构直接从检测头再输入主干网络进行二次迭代,其网络框架如图 6 所示。

图 6 中,第 1、3 部分为主干网络,为下采样过程,如图 7(a) 所示;图 6 第 2、4 部分为解码器网络,为上采样过程,如图 7(b) 所示,虚框区域对应每个上采样层和下采样层。具体实现如下:取第 3 个 Pbottleneck 输出

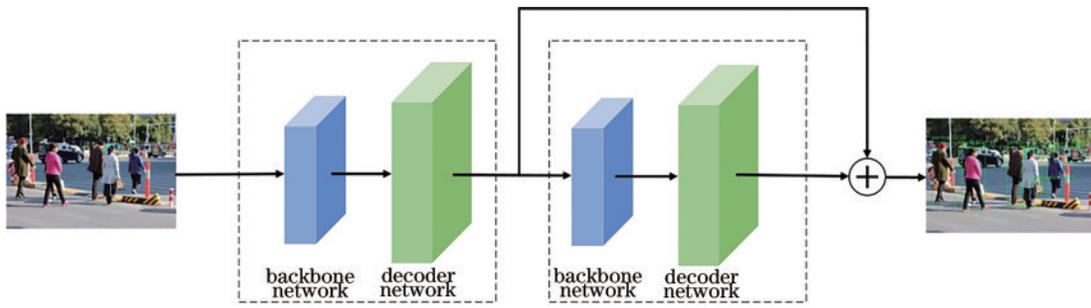


图 5 所提算法的总体网络框架

Fig. 5 Overall network framework of proposed algorithm

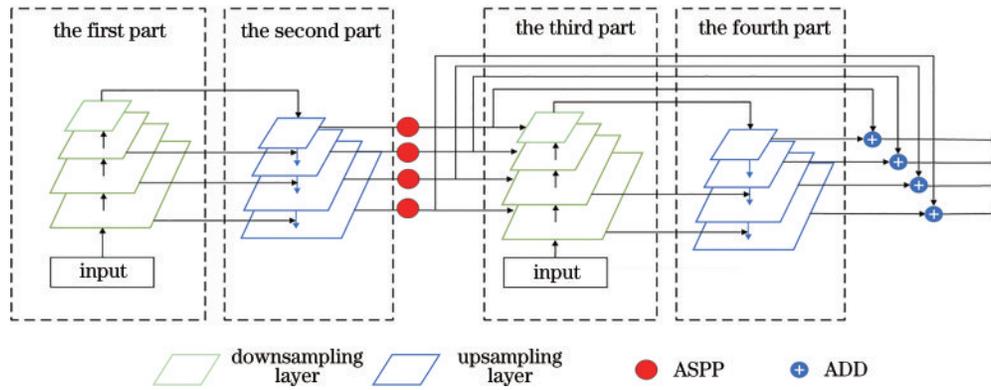


图 6 多尺度检测网络框架

Fig. 6 Multiscale detection network architecture

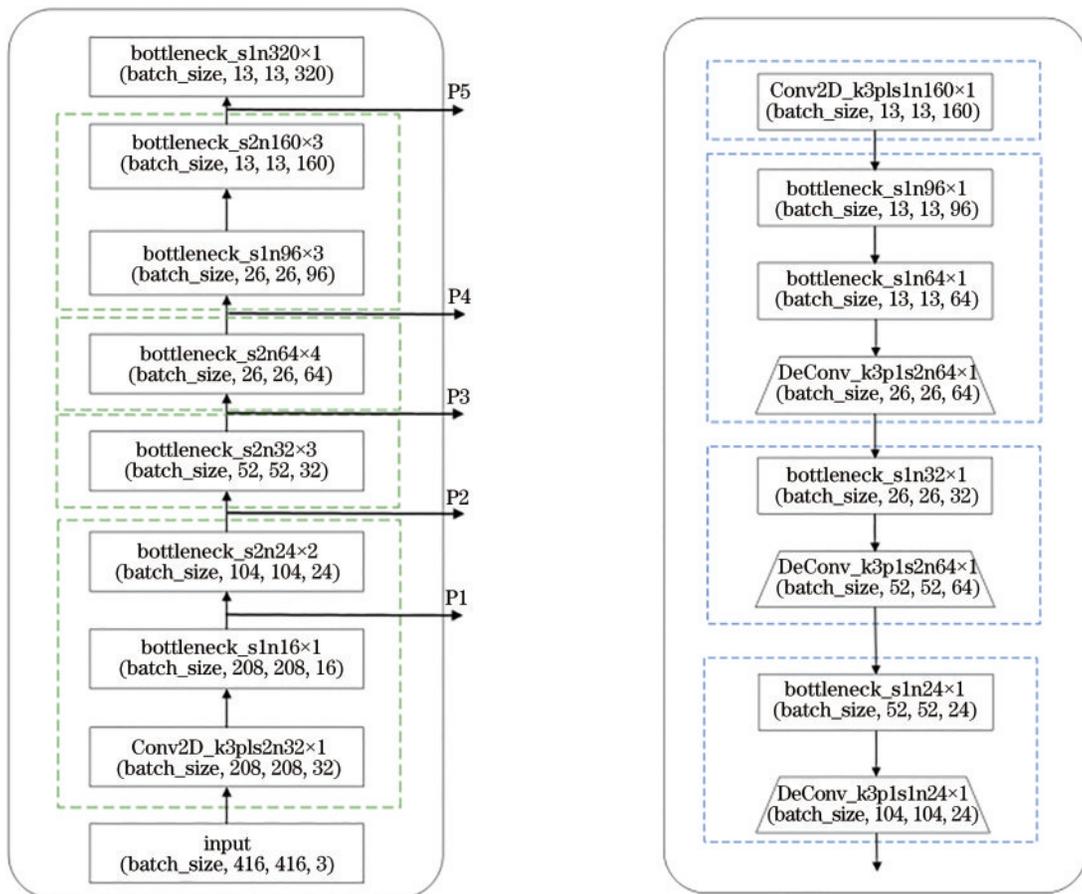


图 7 主干网络和解码器网络。(a)主干网络;(b)解码器网络

Fig. 7 Backbone network and decoder network. (a) Backbone network; (b) decoder network

层为 P2, 第 6 个 P bottleneck 输出层为 P3, 第 10 个 P bottleneck 输出层为 P4, 第 16 个 P bottleneck 为 P5, 构建 P2、P3、P4、P5 等 4 个不同尺度检测层; 下采样层是由 P bottleneck 模块和反卷积层组成的。第 1 个解码器网络输出  $13 \times 13$ 、 $26 \times 26$ 、 $52 \times 52$ 、 $104 \times 104$  这 4 个特征层。将  $13 \times 13$  的特征层输入主干网络, 经过第 1 个上采样输出  $26 \times 26$ , 与第 1 部分输出的  $26 \times 26$  特征图相加输出为两部分, 一部分继续上采样, 再与  $52 \times 52$  的特征图融合; 一部分作为第 2 部分的输出。以此类推, 图像经过第 1、2 部分输出 4 个特征层。4 个特征层通过空洞空间金字塔池化 (ASPP)<sup>[23]</sup> 来实现更好的特征融合。4 个检测层分别重新再经过主干网络和解码器网络, 与上面所述完全一致, 然后将第 1 次经过主干网络与解码器网络的输出与最终输出相加, 整个网络最终得到  $13 \times 13$ 、 $26 \times 26$ 、 $52 \times 52$ 、 $104 \times 104$  等 4 个检测层, 进行不同尺度的目标检测。

### 3.2 损失函数

#### 3.2.1 位置损失函数

位置损失包含预测中心点距离和宽高损失, 本研究采用 CIoU<sup>[18]</sup> 损失函数:

$$L_{\text{CIoU}} = 1 - I_{\text{IoU}} + \frac{P^2(b, b^{\text{gt}})}{r^2} + \alpha v, \quad (4)$$

$$v = \frac{4}{\pi} \left( \arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2, \quad (5)$$

式中:  $w^{\text{gt}}$ 、 $h^{\text{gt}}$ 、 $w$ 、 $h$  分别表示目标框的宽、高和预测框的宽、高;  $b$ 、 $b^{\text{gt}}$  分别表示预测框的中心点与目标框的中心点;  $P^2(b, b^{\text{gt}})$  表示预测框与目标框之间中心距离的平方;  $r$  表示把预测框和目标框包含在内的最小封闭区域的对角线长度;  $\alpha$  为  $v$  的权重系数;  $I_{\text{IoU}}$  为预测框与真实框的交并比。CIoU 不仅解决了检测框和真实框没有重叠时的问题, 还增加了对检测框尺度和长宽的判断, 对位置损失的判断有着更好的效果。

#### 3.2.2 类别损失函数

针对正负样本比例不均衡问题, 引入 GHM-C 损失函数<sup>[19]</sup> 作为类别损失函数。梯度均衡机制 (GHM) 把类别不平衡归结为难易样本不平衡, 难易样本不平衡归结为梯度不平衡。GHM-C 作为动态损失函数, 不受数据变化和模型更新的影响。类别损失函数 GHM-C 的表达式为

$$L_{\text{GHM-C}} = \frac{1}{N} \sum_{i=1}^N \beta_i L_{\text{CE}}(p_i, p_i^*) = \sum_{i=1}^N \frac{L_{\text{CE}}(p_i, p_i^*)}{G_D(g_i)}, \quad (6)$$

$$L_{\text{CE}}(p, p^*) = \begin{cases} -\log_2 p & p^* = 1 \\ -\log_2 (1-p) & p^* = 0 \end{cases}, \quad (7)$$

$$\beta_i = \frac{N}{G_D(g_i)}, \quad (8)$$

式中:  $G_D(g)$  为单位模长在  $g$  附近的样本数量;  $\beta_i$  为密度协调参数;  $N$  是样本数量;  $p$  是预测样本类别概率;  $p^*$  是类真实标签;  $g$  为样本真实值与预测值之间距离;  $g_i$

为第  $i$  个样本的梯度。

## 4 实验结果与分析

### 4.1 实验细节

实验采用 COCO 数据集为训练数据集, 该数据集共 67618 个目标, 包含 22372 位行人, 41260 辆汽车, 3986 辆单车。采用 FLIR 公司提供的红外场景数据集作为测试集, 包括 5579 位行人, 5432 辆汽车, 471 辆单车, 共计 11682 个目标。

实验基于 Window 10 操作系统的 Pytorch1.4.0 深度学习框架, 编程语言为 Python; CPU 型号为 Intel Core i9-9900k@3.6 GHz, GPU 型号为 NVIDIA RTX 3090, 运行内存为 16 GB。在训练的过程中, 采用 Adam 优化器对网络参数进行优化, 在训练的初始阶段, 冻结骨干网络参数, 仅对分类头进行训练, 学习率设置为 0.001, batch size 为 16, 训练 250 轮。之后将骨干网络进行解冻, 对整个网络进行训练, 学习率设置为 0.0001, batch size 为 4, 训练 200 轮。

### 4.2 所提算法与其他算法实验结果对比

为验证所提算法的有效性, 采用相同的实验环境和红外图像测试集, 在 4 种情景中测试 4 种算法的效果, 分别为 YOLOv3<sup>[17]</sup>、E-YOLO<sup>[15]</sup>、scSE-IYOLOv4<sup>[16]</sup>、所提算法。图 8 为 4 个场景下的 4 种算法检测结果图, 4 种算法在 FLIR 测试集实验结果如表 2 所示。同时为了验证所提算法在不同数据集上的检测效果, 其余实验细节不变, 仅改用 TNO 数据集作为测试数据集, 4 种算法在 TNO 数据集的测试结果如表 3 所示。

从图 8 可以看出, YOLOv3 对小目标检测效果很差, 漏检率高。其他先进算法的检测精度和速度得到了一定程度的提升, 但在红外场景下的检测效果还有待加强, 尤其在小目标检测方面出现漏检的情况较多。而所提算法在红外场景中对一些较小的目标也有着很好的效果, 漏检率低, 对目标有着更准确的检测。

由表 2、表 3 可知, 所提算法在 FLIR 数据集 mAP 达到 81.4%, 比 YOLOv3 的 mAP 高 5.8 个百分点, 在 TNO 数据集 mAP 达到 88.1%, 比 YOLOv3 高近 4 个百分点。通过加入注意力机制模块, 改进特征金字塔结构等方式提高了检测精度。由于使用轻量化网络模型作为主干网络, 参数和计算量减小。所提算法模型大小为 80.5 MB, 在 FLIR 数据集上的 FPS 达到 56.6, 在 TNO 数据集上的 FPS 达到 62.8。与对比算法相比, 所提算法模型参数量小, 在检测精度和速度上均有一定程度的提高。

### 4.3 消融实验

为了更好地明确各个模块对网络的影响, 设置 3 个消融实验。消融实验采用与上述在 FLIR 数据集上进行对比实验相同的实验环境, “√”为包含该模块, “×”为不包含该模块。

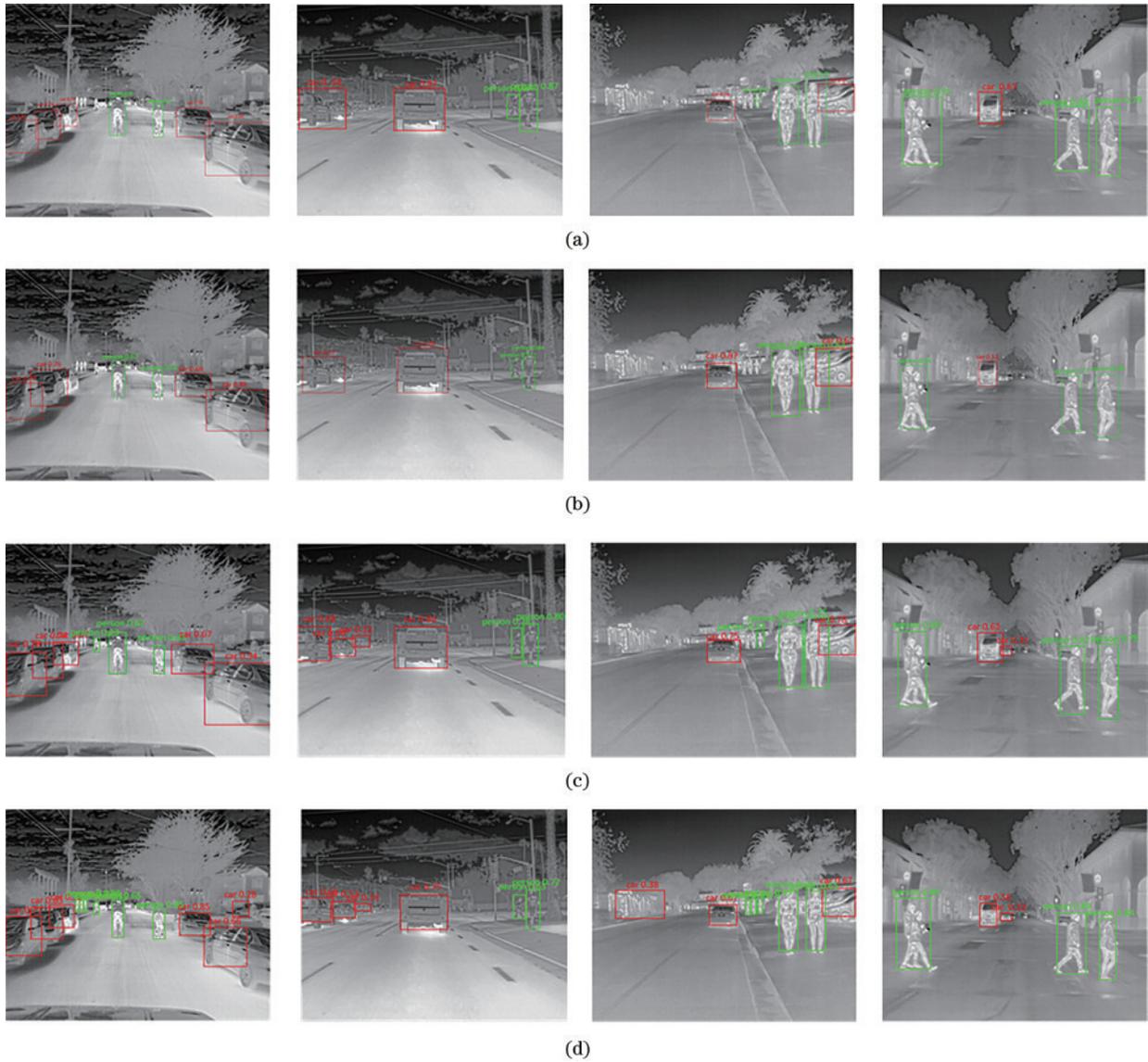


图 8 检测结果图:(a)YOLO v3;(b) E-YOLO;(c)scSE-IYOLOv4;(d)所提算法

Fig. 8 Graphs of detection results. (a) YOLO v3; (b) E-YOLO; (c) scSE-IYOLOv4; (d) proposed algorithm

表 2 4 种算法在 FLIR 数据集上的实验结果

Table 2 Experimental results of four algorithms on FLIR dataset

Algorithm	Model size / MB	mAP / % (IoU:0.5)	FPS
YOLOv3	89.4	75.6	54.7
E-YOLO	85.3	77.3	55.6
scSE-IYOLOv4	93.1	79.3	53.9
Proposed algorithm	80.5	81.4	56.6

表 3 4 种算法在 TNO 数据集上的实验结果

Table 3 Experimental results of four algorithms on TNO dataset

Algorithm	Model size / MB	mAP / % (IoU:0.5)	FPS
YOLOv3	89.4	84.2	60.5
E-YOLO	85.3	85.7	62.3
scSE-IYOLOv4	93.1	87.4	61.7
Proposed algorithm	80.5	88.1	62.8

#### 4.3.1 消融实验 1

针对所设计的注意力机制模块,在所提算法基础上,仅改变 P bottleneck 模块中注意力机制模块,消融实验结果如表 4 所示。为明确所设计注意力模块的有效性,设置仅有通道注意力(group 1)、仅有空间注意力(group 2)、CBAM(group 3)<sup>[20]</sup>、PAM(group 4)等 4 组实验进行对比。CBAM 也是一种同时采用空间和通道注意力的模块,其将特征图依次通过通道注意力、空间注意力实现特征重标定。

实验结果表明:仅有通道注意力或空间注意力均会提高整个网络模型的特征提取能力,两者都在一定程度提高了网络模型的检测精度;将空间注意力和通道注意力结合,CBAM 和 PAM 均牺牲较少检测速度,获得了比使用单一注意力机制更好的检测精度;相较于 CBAM,PAM 模块使得所提红外目标检测模型的检测精度和速度均有提升。

表 4 消融实验 1 结果对比

Table 4 Comparison of results of ablation experiment 1

Experiment grouping	Only-channel	Only-spatial	CBAM	PAM	mAP / %	FPS
group 1	✓	×	×	×	77.9	57.4
group 2	×	✓	×	×	78.5	56.8
group 3	×	×	✓	×	80.6	56.1
group 4	×	×	×	✓	81.4	56.5

#### 4.3.2 消融实验 2

针对所设计的多尺度检测结构,在所提算法基础上,仅改变多尺度检测结构,设置 FPN(group 1)、RFP(group 2)、T-FPN(group 3)等 3 组实验。使用的 RFP 模块为迭代一次的特征金字塔结构, T-FPN 表示迭代两次的特征金字塔结构,实验结果如表 5 所示。

表 5 消融实验 2 结果对比

Table 5 Comparison of results of ablation experiment 2

Experiment grouping	FPN	RFP	T-FPN	mAP / %	FPS
group 1	✓	×	×	77.6	57.9
group 2	×	✓	×	81.4	56.6
group 3	×	×	✓	82.5	52.8

从消融实验 2 结果发现,迭代次数越多,目标检测精度越高,检测速度越低。迭代两次的金字塔结构的 mAP 比 RPN 高不到 1 个百分点,但 FPS 降低近 4,综合考虑选择 RPN 作为特征金字塔结构。RPN 作为多尺度检测模块可以在牺牲较少检测速度的前提下,提高检测精度。

#### 4.3.3 消融实验 3

针对所设计的损失函数,在所提算法基础上,仅改变损失函数,消融实验结果如表 6 所示。YOLOv3 由位置损失、置信度损失、类别损失组成。位置损失为中心点  $(x, y)$  产生的二元交叉熵(BCE)损失和预测宽高产生的均方误差(MSE)损失,置信度损失和类别损失函数为 BCE 损失。表 6 中:CIoU 表示将 YOLOv3 位置损失换成 CIoU,其余不变;GHM 表示将 YOLOv3 类别损失函数换成 GHM-C,其余不变;CIoU+GHM 表示将位置损失换成 CIoU,类别损失函数换成 GHM-C,置信度损失换为 BCE 损失。

表 6 消融实验 3 结果对比

Table 6 Comparison of results of ablation experiment 3

Experiment grouping	Loss	CIoU	GHM	mAP / %	FPS
group 1	✓	×	×	77.9	55.7
group 2	×	✓	×	78.2	56.2
group 3	×	×	✓	78.5	56.1
group 4	×	✓	✓	81.4	56.6

从实验结果可以看出:以 CIoU+GHM 作为损失函数时 mAP 达 81.4%,FPS 达 56.6,检测精度和检测速度最高,更加适合红外场景的目标检测。

## 5 结 论

为了提高红外场景的目标检测精度和速度,提出一种基于注意力机制的目标检测算法。采用一种轻量化网络并改进迭代特征金字塔结构,有效地减少了网络的参数量和计算量。同时,通过注意力机制模块来提高整个模型特征提取能力,并引入 CIoU 和 GHM-C 损失函数解决了正负样本不平衡问题。大量实验结果表明,所提算法与其他对比算法相比,在检测精度和速度均有优势,能够有效地对红外目标进行检测。如何提高严重遮挡情况下的漏检现象是下一阶段研究的方向。

## 参 考 文 献

- [1] Zhu D P, Zhan W D, Jiang Y C, et al. MIFFuse: a multi-level feature fusion network for infrared and visible images[J]. IEEE Access, 2021, 9: 130778-130792.
- [2] 张涛, 张乐. 一种基于多尺度特征融合的目标检测算法[J]. 激光与光电子学进展, 2021, 58(2): 0215003. Zhang T, Zhang L. Multiscale feature fusion-based object detection algorithm[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0215003.
- [3] 程叶群, 王艳, 范裕莹, 等. 基于卷积神经网络的轻量化目标检测网络[J]. 激光与光电子学进展, 2021, 58(16): 1610023. Cheng Y Q, Wang Y, Fan Y Y, et al. Lightweight object detection network based on convolutional neural network[J]. Laser & Optoelectronics Progress, 2021, 58(16): 1610023.
- [4] 王永顺, 贾文杰, 王晨飞, 等. 基于改进 YOLOv3 的车辆识别方法[J]. 激光与光电子学进展, 2021, 58(16): 1610010. Wang Y S, Jia W J, Wang C F, et al. Vehicle recognition method based on improved YOLOv3 algorithm[J]. Laser & Optoelectronics Progress, 2021, 58(16): 1610010.
- [5] Jiang Y C, Liu Y Q, Zhan W D, et al. Lightweight dual-stream residual network for single image super-resolution[J]. IEEE Access, 2021, 9: 129890-129901.
- [6] 李一男. 深度学习目标检测方法研究综述[J]. 中国新通信, 2021, 23(9): 159-160. Li Y N. A survey of research on deep learning target detection methods[J]. China New Telecommunications, 2021, 23(9): 159-160.
- [7] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [8] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid

- pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [9] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [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] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [12] 秦鹏, 唐川明, 刘云峰, 等. 基于改进 YOLOv3 的红外目标检测方法[J]. 计算机工程, 2022, 48(3): 211-219.
- Qin P, Tang C M, Liu Y F, et al. Infrared target detection method based on improved YOLOv3[J]. Computer Engineering, 2022, 48(3): 211-219.
- [13] Ju M, Luo H B, Wang Z B, et al. The application of improved YOLO V3 in multi-scale target detection[J]. Applied Sciences, 2019, 9(18): 3775.
- [14] Lu L P, Li H S, Ding Z, et al. An improved target detection method based on multiscale features fusion[J]. Microwave and Optical Technology Letters, 2020, 62(9): 3051-3059.
- [15] 许虞俊, 李晨. 基于 YOLO 优化的轻量级目标检测网络[J]. 计算机科学, 2021, 48(S2): 265-269.
- Xu Y J, Li C. Light-weight object detection network optimized based on YOLO family[J]. Computer Science, 2021, 48(S2): 265-269.
- [16] 蒋榕圻, 彭月平, 谢文宣, 等. 嵌入 scSE 模块的改进 YOLOv4 小目标检测算法[J]. 图学学报, 2021, 42(4): 546-555.
- Jiang R Q, Peng Y P, Xie W X, et al. Improved YOLOv4 small target detection algorithm with embedded scSE module[J]. Journal of Graphics, 2021, 42(4): 546-555.
- [17] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-08-08)[2021-12-09]. <https://arxiv.org/abs/1804.02767>.
- [18] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [19] Li B Y, Liu Y, Wang X G. Gradient harmonized single-stage detector[EB/OL]. (2018-11-13)[2021-12-09]. <http://arxiv.org/abs/1811.05181>.
- [20] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [21] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.
- [22] Qiao S Y, Chen L C, Yuille A. DetectoRS: detecting objects with recursive feature pyramid and switchable atrous convolution[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 10208-10219.
- [23] Yang M K, Yu K, Zhang C, et al. DenseASPP for semantic segmentation in street scenes[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 3684-3692.