

基于 YOLO v7 的轻量级红外目标检测算法

陈永麟^{1,2}, 王恒涛^{1,2}, 张 上^{1,2}

(1. 湖北省建筑质量检测装备工程技术研究中心, 湖北 宜昌 443002;

2. 三峡大学 计算机与信息学院, 湖北 宜昌 443002)

摘要: 针对红外图像信噪比低、分辨率不佳、噪声与杂波多等检测难点。提出一种基于 YOLOv7 的轻量化红外图像目标检测算法 ITD-YOLO。首先, ITD-YOLO 算法重设计网络结构, 对特征提取网络与特征融合网络架构重新调整。裁剪掉原网络中深层对应的大感受野, 依据重构后网络特征图输出, 对模型预设锚框进行调节。改变多尺度特征融合中的深层特征与浅层特征的关系, 提高浅层网络提取的细节信息在融合中所占的权重, 提高对较小目标的检测性能; 然后, 在 ELAN 模块中引入 PConv 替换掉常规卷积, 进一步降低模型计算量。其次, 将模型损失函数调整为 PolyLoss 以加速模型收敛, 进一步加强对目标的检测性能; 最后, 使用 SIOU 作为边框损失函数, 增强对目标的定位精度。实验结果表明, ITB-YOLO 能够有效改善检测效果, 在 FLIR 与 OSU 数据集上, 相较于 YOLOv7s 的平均精度均值分别提高 2.27% 与 7.29%。改进后得到的模型体积仅为 17.7MB, 计算量下降 37.11%。与主流算法进行对比, ITD-YOLO 在各项指标均得到了一定程度的提高, 能够满足红外目标实时检测任务。

关键词: 目标检测; 模型轻量化; YOLOv7; PConv; PolyLoss; SIOU

中图分类号: TP391.4 文献标识码: A 文章编号: 1001-8891(2024)12-1380-10

Lightweight Infrared Target Detection Algorithm Based on YOLO v7

CHEN Yonglin^{1,2}, WANG Hengtao^{1,2}, ZHANG Shang^{1,2}

(1. Hubei Province Engineering Technology Research Center for Construction Quality Testing Equipments, China Three Gorges University, Yichang 443002, China; 2. College of Computer and Information, China Three Gorges University, Yichang 443002, China)

Abstract: Aiming at the detection difficulties of infrared images such as low signal-to-noise ratio, poor resolution, and much noise and clutter. We propose a lightweight infrared image target detection algorithm ITD-YOLO based on YOLOv7. Firstly, the ITD-YOLO algorithm redesigns the network structure, and re-adjusts the architecture of the feature extraction network and the feature fusion network. Crop out the large receptive fields corresponding to the deep layers in the original network, and adjust the model preset anchor frames based on the output of the reconstructed network feature map. The relationship between deep and shallow features in multi-scale feature fusion is changed to increase the weight of the detail information extracted by the shallow network in the fusion to improve the detection performance of smaller targets; then, PConv is introduced into the ELAN module to replace the conventional convolution to further reduce the model computation. Next, the model loss function is adjusted to PolyLoss to accelerate the model convergence and further enhance the detection performance for targets; finally, SIOU is used as the edge loss function to enhance the localisation accuracy for targets. The experimental results show that ITB-YOLO can effectively improve the detection effect, and the mean average accuracy is increased by 2.27% and 7.29% compared with YOLOv7s on FLIR and OSU datasets, respectively. The volume of the model obtained after the improvement is only 17.7 MB, and the computation volume decreases by 37.11%. Comparing with the

收稿日期: 2023-04-24; 修订日期: 2023-05-24.

作者简介: 陈永麟 (1999-), 男, 湖北荆门人, 硕士研究生, 研究方向为目标检测, E-mail:1768859718@qq.com。

通信作者: 张上 (1979-), 男, 湖北宜昌人, 副教授, 工学博士, 研究方向为物联网技术、计算机应用技术, E-mail:3011408157@qq.com。

基金项目: 国家级大学生创新创业训练计划 (202111075012), 国家级大学生创新创业训练计划 (202011075013)。

mainstream algorithms, ITD-YOLO has been improved to a certain extent in all the indexes, and can meet the real-time infrared target detection task.

Key words: target detection, model lightweight, YOLOv7, PConv, PolyLoss, SIoU

0 引言

红外成像技术是一项应用前景广阔的技术。其原理就是依靠环境中的红外辐射得到平面图像,将物体反射或辐射的红外能量分布情况转化为二维可见图像。红外成像的最大优势就在于其能够在光强不足或对比度差的环境下提供清晰的图像,其特质使得它能够应对烟雾、雾霾、雨雪等因天气或环境因素造成的视觉影响。红外成像技术在民用与军用领域都已得到大量普及。在民用领域,一些家用汽车都已装上车载热成像系统,基于这项技术的夜视系统大幅提高了汽车在夜间行驶的视觉感知能力。系统辅助驾驶员在复杂环境下对前方路况和障碍物做出预判,提升道路交通安全,保护人们日常出行。在军用领域,红外成像技术也有着重要的应用,已成为现代战争不可或缺的一部分。其在侦查、监视、瞄准、制导等方面的应用要求也愈发严格。

近些年计算机视觉技术的不断发展,在目标检测方向,基于深度学习的目标检测算法不断涌现。传统的基于机器学习的目标检测方法存在两个很大的缺陷:其一是使用效率不高的基于滑动窗口的区域选择策略。其二是手动提取特征缺乏鲁棒性。对物体实施精准的、实时的检测就成了关键问题,深度学习的出现很好地解决了以上问题。目前基于深度学习的目标检测算法大致分为4类,第一类是两阶段深度学习算法,以R-CNN系列^[1-3]为代表的基于候选区域算法将检测过程划分为两步。首先通过区域选择方法生成候选区域,然后对样本提取特征并回归分类。这类算法通过区域建议将检测问题转化为对区域内图片的分类问题。其准确度相对较高但是速度却不太理想。第二类是单阶段深度学习算法,以YOLO系列^[4-6]和SSD系列^[7]为代表的基于回归的算法实现了端对端对目标进行检测。直接使用整幅图像预测物体类别和边界框位置。该类算法常常具备轻量化、速度快、泛化能力好等优点饱受用户青睐。第三类是无锚框目标检测算法,以CenterNet^[8-9]为代表的基于关键点的算法通过热力图来预测物体中心点,利用获取到的特征预测边界框的尺寸。由于其抛弃了NMS(非极大抑制)的过程,所以它相当简洁。第四类是基于注意力机制^[10]的目标检测算法,Transformer骨干网络使用

自注意力编码特征,其对于图像全局信息的感知能力是卷积神经网络所不具备的。但同时CNN对局部信息的提取能力又是Transformer所不具备的,所以将两者相结合是未来的研究趋势。

对比传统目标检测方法^[11-13],深度学习高度依赖数据集,精度很大程度上取决于数据集的规模和质量,但其强大的特征提取能力是它最大的优势。

在红外目标检测与算法轻量化上,许多学者为提高在各种困难环境下的检测效果,提出了许多有创新性的方法。贺顺等人^[14]根据人眼的特性,通过增强图像局部对比度,提出一种基于IHBf的红外小目标检测方法。李飏^[15]等人为抑制背景同时消除更多噪声,提出一种自适应梯度倒数滤波算法,用于实现红外弱小目标场景背景抑制。李向荣等人^[16]在YOLOv4架构中融入通道注意力机制,重新构建多尺度特征融合模块,提升了算法的检测精度。Bao等人^[17]在YOLOv7基础上设计新特征融合模块与损失函数,提高了在红外物体纹理不清晰条件下的检测能力。Li等人^[18]通过双三次插值方法将红外图像上采样,以应对卷积过程中的纹理特征丢失问题。Hong等人^[19]在YOLOv5网络中引入了深度可分离卷积和ECA机制,很好地实现了模型的轻量化。

为解决目标检测技术在复杂背景下的红外图像检测^[20-21]效果不佳、算法效率不高、硬件条件苛刻等诸多问题,本文着力于改进相关算法,提出一种基于YOLOv7的红外图像目标检测算法,主要工作如下:

①设计轻量化网络算法ITD-YOLO(infrared target detector-YOLO),调整网络感受野,改变多尺度融合关系,增大浅层网络特征层权重。

②将ELAN网络部分卷积改进为PConv,进一步实现网络轻量化目标。

③引入PolyLoss计算置信度与分类损失,增大对难分类样本的处理能力,提升召回率。

④使用SIoU增强锚定框对目标的定位精度,在训练阶段实现模型的更快收敛。

⑤在红外图像数据集FLIR与OSU上检验本文所提出的改进算法性能。

1 ITD-YOLO 介绍

YOLOv7提出了一种新的高效层聚合网络

ELAN (efficient layer aggregation network), 通过控制梯度传播路径的长度, 使网络更高效地学习与收敛, 解决了基于串联的深度模型缩放过程中一些层的输入宽度被改变的问题。应用梯度路径设计策略设计的 ELAN 模块结构如图 1 所示。

ELAN-W 与 ELAN 的不同点就在于第二条分支选取输出叠加的个数不同。ELAN-W 结构如图 2 所示。

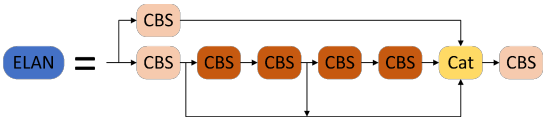


图 1 ELAN 网络结构

Fig.1 ELAN network structure

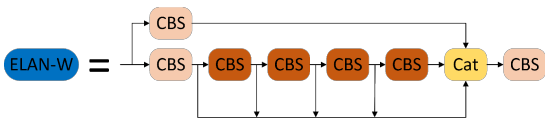


图 2 ELAN-W 网络结构

Fig.2 ELAN-W network structure

由于 YOLOv7 网络层数有 415 层且参数量巨大, 在实际目标检测过程中, 检测速度方面有很大提升空间。为实现对复杂背景下红外图像中的目标进行精准识别与定位, 故本文设计 ITD-YOLO (ITD: infrared target detector, 红外目标检测器)。

针对待检测目标小且特征不明显的特点, 网络参数量与计算量大导致模型推理时间长。在 YOLOv7 基础网络结构上改进与优化, 将原 YOLOv7 网络结构轻量化, 对结构中的模块关系进行重新设计, 替换效率更高的模块, 删除冗余模块。本算法将轻量化网络结构、SIoU 和 PolyLoss 三个关键点进行融合。首先, 为增强网络的特征提取能力, 调节特征提取网络中多尺度的感受野权重, 对特征网络进行轻量化剪裁; 针对重新设计的主干网络, 构建新的特征融合网络结构。剔除冗余的 Neck 层模块。对于边框回归损失函数, 使用 SIoU 优化边界框回归任务中的样本不平衡问题, 加速网络的收敛并且提高模型回归精度。最后, 使用 PolyLoss 用于训练密集目标检测器, 超越常用的交叉熵和焦点损失, 很好地提升了目标检测的精确度。如图 3 所示为 ITD-YOLO 网络架构。

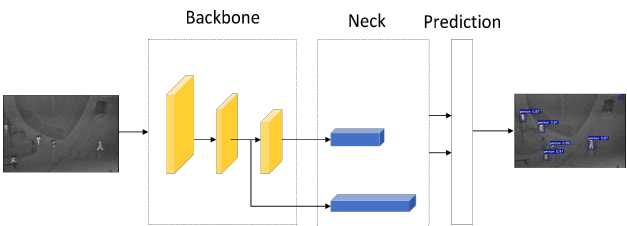


图 3 ITD-YOLO 系统架构图

Fig.3 ITD-YOLO system architecture

2 ITD-YOLO 架构

2.1 网络结构重设计

由于红外图像的特质, 图像中的目标常具有特征不明显、轮廓模糊、像素点少等特点。ITD-YOLO 重新设计原模型。YOLOv7 使用 3 个尺度实现对不同大小的目标进行检测, 输出特征图尺寸分别为 20 pixel×20 pixel、40 pixel×40 pixel、80 pixel×80 pixel。其中 80 pixel×80 pixel 用于检测小目标, 20 pixel×20 pixel 用于检测大目标。本模型根据实际需要, 剔除用于检测较大尺寸目标的 20 pixel×20 pixel 尺度输出。

实际检测过程中, 红外图像中占比最大的是无效背景信息, 而待检测目标仅占据其中较少部分。输入图片在 YOLOv7 原网络结构中经过层层卷积, 目标的特征信息与位置信息逐渐减弱, 感受野过大导致受背景噪声影响程度较深。基于此点可以考虑适当减小网络的深度, 提高浅层含有更强空间位置信息的特征权重, 以提高对小且模糊的目标检测能力。ITD-YOLO 对主干特征提取网络 CSPDarkNet 进行重构, 删除 20 pixel×20 pixel 尺寸的特征图输出路径。重构 CSPDarkNet 结构如表 1 所示。

重构 CSPDarkNet 结构优势如下:

- ①删除大量低效的模块, 浅层网络提取出的细节特征权重得到提升, 模型得到很大程度的精简。
- ②模型整体复杂度和计算量大幅下降, 缩减了模型体积, 推理时间下降, 模型效率提升。
- ③改进后网络参数量缩减至 YOLOv7 原模型的 25.91%, 由 13371808 降至 3464608。

表 1 重构 CSPDarkNet 结构

Table 1 Reconfiguration of the CSPDarkNet structure				
Module	Parameters	Channel	Kernel size	Output
CBS	928	32	(3,3)	640×640
CBS	18560	64	(3,3)	320×320
CBS	36992	64	(3,3)	320×320
CBS	73984	128	(3,3)	160×160
ELAN-P	108800	256		80×80
MP-1	213760	256		80×80
ELAN-P	432640	512		80×80
MP-1	853504	512		40×40
ELAN-P	1725440	1024		40×40

ITD-YOLO 输出特征图大小调整为 40 pixel×40 pixel 与 80 pixel×80 pixel。对应特征提取网络的调

整, 将特征融合网络进行重新设计, 在网络结构中, CBS 由卷积、批归一化和激活函数组成, SPPCSPC 为空间金字塔池化模块, 使用 MP-2 在实现下采样同时进行维度扩增, REP 为重参数化卷积。融合后得到的特征图大小为输入图片的 1/16 和 1/8, 提高了对浅层细节特征的关注度, 提高了浅层语义信息的权重, 减少过高的下采样率伴随的特征信息丢失问题。改进后特征融合网络结构如图 4 所示。

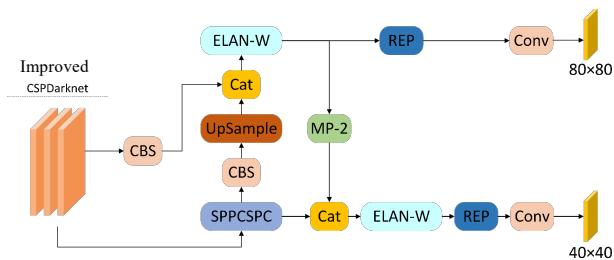


图 4 特征融合网络结构
Fig.4 Feature fusion network structure

2.2 锚框参数优化

在模型训练之前, 模型设置的默认锚框参数与两个数据集中目标的 BPR (Best Possible Recall) 均小于 0.98, 表明预设大小与实验所用数据集不匹配。YOLOv7 的自适应锚框机制采用 K-Means 算法聚类数据集自动生成与之适应的锚框。本算法经过改进, 特征图输出由原模型的三尺度变为二尺度, 锚框个数由 9 变为 6, 每个特征图对应 3 种不同大小的锚框。

经过 K-Means 聚类得到结果, 然后使用遗传算法分析, 在 FLIR 与 OSU 得到 BPR 分别为 0.9993 与 1.0000。相应地对锚框的大小进行重新调整。通过试验发现, 选择合适的锚框大小能有效加快模型收敛速度、减小训练所需时间和提高目标定位准确度。改进后生成的锚框参数如表 2 所示。

表 2 锚框分配表

Table 2 Table of anchor box assignments		
Feature map	40×40	80×80
Receptive field	Medium	Small
FLIR	(15,16)	(23,61)
	(14,31)	(54,44)
	(30,26)	(102,86)
OSU	(30,40)	(33,44)
	(36,47)	(37,50)
	(39,52)	(42,55)

2.3 部分卷积

为了设计快速的神经网络, 许多改进都集中于减少浮点运算次数 (FLOPs) 上。但实际上, 浮点运算次数的减小不一定能带来同等水平的延迟减小。主要

原因在于每秒进行的浮点运算量较少, 普通卷积操作对内存的频繁访问。为了实现更快的网络, 在主干网络引入一种新的卷积^[22], 名为部分卷积 (PConv, Partial Convolution), 减少冗余计算和内存访问次数。部分卷积只在少数几个输入通道上应用过滤器, 而对其余几个通道不做处理。常规卷积与纵深/分组卷积的结构如图 5 所示。

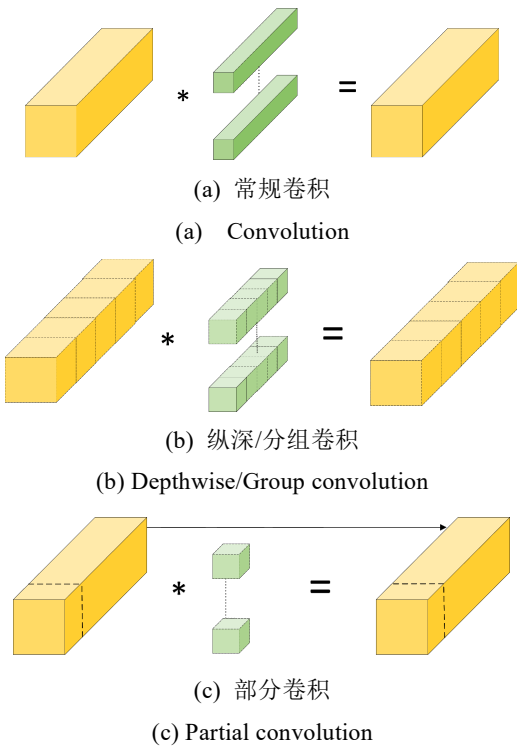


图 5 三种卷积结构对比

Figure 5 Comparison of 3 types of convolution structure
如图 5(c)所示, 它利用特征映射中的冗余, 系统地在一部分输入通道上应用常规卷积 (Conv), 同时保持其余通道不变。

部分卷积相比常规卷积与纵深/分组卷积具有更低的 FLOPs 和更高的 FLOPS, 其中 FLOPS 是每秒浮点运算的简称, 作为有效计算速度的一个衡量标准。部分卷积相比普通卷积更好地利用了设备的计算能力, 在空间特征提取上也很有效。本算法将其引入模型主干网络的 ELAN 模块, 得到的 ELAN-P 结构如图 6 所示。

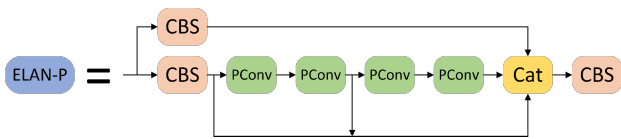


图 6 ELAN-P 网络结构

Fig.6 Structure of the ELAN-P network

2.4 PolyLoss

YOLOv7 原模型使用 BCE loss 作为分类与置信

度损失函数,但其存在无法平衡正负样本,收敛速度慢,模型训练效果不佳等缺点。本节方法将常用分类损失函数(例如交叉损失和焦点损失)分解为一系列加权多项式基。分解的形式如下: $\sum_{j=1}^{\infty} \alpha_j (1-P_t)^j$

其中多项式系数 $\alpha_j \in R^+$, 是目标类标签的预测概率。每个基础多项式 $(1-P_t)^j$ 由多项式系数 α_j 加权,以便实际应用中灵活调整。当 $\alpha_j = 1/j$ 时,其等效于常用的交叉熵损失,即通过水平移动多项式系数,将交叉熵损失和焦点损失解释为 PolyLoss^[23]的特例。在实际应用中,针对不同的任务和数据集灵活调整多项式系数 α_j ,就能达到优于交叉熵损失和焦点损失的效果,公式如下:

$$L_{\text{poly}} = \alpha_1(1-P_t) + \alpha_2(1-P_t)^2 + \cdots + \alpha_N(1-P_t)^N + \cdots =$$

$$\sum_{j=1}^{\infty} \alpha_j (1-P_t)^j \quad (1)$$

利用 PolyLoss 框架设计新损失函数,对多项式系数进行调整,但是过多的系数同时改变,成本很高且结果不一定最优。方法选择扰乱交叉熵损失中的主要多项式系数,余下部分不变。并损失公式表示为 Poly- N , 其中 N 代表将被调整的前导系数的数量。Poly- N 公式如下:

$$L_{\text{Poly-}N} = (\varepsilon_1 + 1)(1-P_t) + \cdots + (\varepsilon_N + 1/N)(1-P_t)^N + 1(N+1)(1-P_t)^{N+1} + \cdots = -\log(P_t) + \sum_{j=1}^N \varepsilon_j (1-P_t)^j \quad (2)$$

当仅对交叉熵损失中的第一个多项式系数做调整时,得到函数 Poly-1。在绝大数情况下,前导多项式在训练时贡献大量的交叉熵梯度,这也表明了第一多项式 $(1-P_t)$ 与其后无限多项相比有更大价值。

本算法引入 PolyLoss,加速了模型收敛,使模型更多地关注困难且高价值的样本,提升检测精度与召回率。

2.5 SIoU

以前的损失函数在定义时未考虑过真实框与预测框之间的方向, SIoU^[24]引入真实框与预测框的向量角度作为评价指标。SIoU 损失函数由角度、距离、形状和 IoU 四个方面构成。

SIoU 损失函数如下:

$$L_{\text{SIoU}} = 1 - L_{\text{IoU}} + \frac{\Delta + \Omega}{2} \quad (3)$$

式中: L_{IoU} 为交并比损失; Δ 表示预测框与真实框的距离损失; Ω 是定义的形状损失。

形状损失由锚框宽度、锚框高度和参数 τ 来计算,并用参数 τ 控制对形状损失的关注程度。

$$\Omega = \sum_{t=w,h} \left(1 - e^{-\alpha_t}\right)^{\tau} \quad (4)$$

考虑到角度因素,距离损失被重新定义。当两框位置持平时,距离损失在损失中的占比减小。当两框中心点位置夹角趋近于 45° 时,距离损失的贡献就越大。

$$\Delta = \sum_{t=x,y} \left(1 - e^{-(2-\mu)D_t}\right) \quad (5)$$

$$D_x = \left(\frac{\text{box}_{c_x}^{\text{gt}} - \text{box}_{c_x}}{c_w}\right)^2, D_y = \left(\frac{\text{box}_{c_y}^{\text{gt}} - \text{box}_{c_y}}{c_h}\right)^2 \quad (6)$$

2.6 系统算法详解

ITD-YOLO 算法设计如算法 1 所示,对网络进行轻量化后融入 SIoU、PolyLoss 进行复杂背景下的红外图像目标检测。检测算法训练过程伪代码如下。

算法 1:

前期工作: 数据集输入、训练参数设定

训练次数: epoch

While (模型未收敛 or $T < \text{epoch}$)

①Backbone 网络

使用 CBS、MP-1、ELAN-P 模块提取图像特征:

$$\begin{cases} x_1 = \text{MP-1}(\text{CBS}) \\ x_2 = \text{ELAN-P}(x_1) \end{cases}$$

②Neck 网络

经过上采样与下采样改变图像尺寸,通过 Concat 进行通道数的合并,即特征融合。

计算融合特征:

$$x_3 = \text{Concat}(x_2)$$

③预测层

使用 PolyLoss 来加速模型收敛。使用 SIoU 提高锚框定位的精确度。

SIoU 损失函数:

$$L_{\text{SIoU}} = 1 - L_{\text{IoU}} + \frac{\Delta + \Omega}{2}$$

预测融合:

$$x_4 = \text{PolyLoss}(\text{SIoU}(x_3))$$

End While

输出: 模型训练收敛的网络模型及权重。

2.7 ITD-YOLO 结构

ITD-YOLO 算法结构如图 7 所示,与可见光相比,红外图像特征信息不明显,为实现更佳的检测效果,重新设计原 YOLOv7 的网络结构并进行轻量化处理。过深的网络易导致图像特征信息丢失过多,加上红外图像本身的特质,越浅层的网络更能提取到有

效特征。新模型剔除最后一次特征提取的两种模块:MP-1与ELAN,其后的特征融合网络也随之重新设计。将原模型的 $20\text{pixel}\times 20\text{pixel}$ 、 $40\text{pixel}\times 40\text{pixel}$ 和 $80\text{pixel}\times 80\text{pixel}$ 的三尺度输出缩减为 $40\text{pixel}\times 40\text{pixel}$ 与 $80\text{pixel}\times 80\text{pixel}$ 的二尺度。同时将ELAN模块中通道数为3的常规卷积替换为PConv,进一步减小模型参数量与计算量。该方法在提高浅层的感受

野权重的同时,模型的计算量与参数量也得到大幅降低。其中不同颜色的CBS模块表示卷积核的大小和步长不同:蓝色用于改变通道偶数、黄色用于提取特征、棕色用于下采样,ELAN-P为本文改进的模块,MP-1进行下采样同时不改变维度数。改进的网络结构,有效提升检测精度与速度,且模型体积也得到缩减。

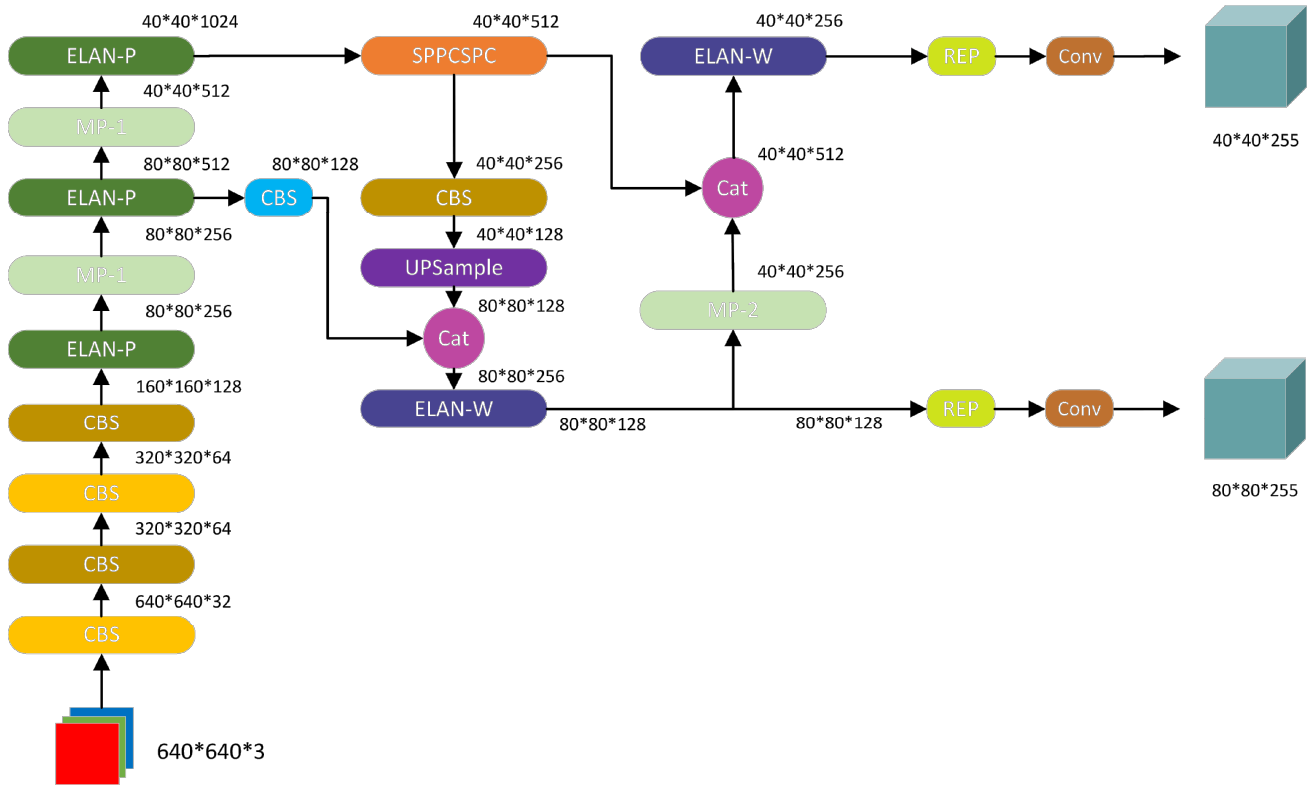


图7 ITD-YOLO 算法结构

Fig.7 Detail of the ITD-YOLO algorithm structure

3 实验及结果分析

3.1 实验数据集

为验证改进算法在复杂环境下红外目标的检测效果,实验使用在不同环境下的红外图像数据集FLIR与OSU进行算法测试。

FLIR热成像数据集是由FLIR公司收集并标注的一个自动驾驶开源热成像数据集,用于开发使用卷积神经网络(CNN)的目标检测系统。图像通过安装在车辆上的热成像和可视摄像头获取的,其中包含10 000多张注释热图像,尺寸均为 $640\text{pixel}\times 512\text{pixel}$ 。场景涵盖白天与黑夜,有行人、汽车和自行车三类注释。

OSU热图像数据集场景为白天,包含雨水天气,拍摄自俄亥俄州立大学校园内人行横道,总计284张图片,984人被标记出来,平均每张图3~4人。图

像尺寸为 $360\text{pixel}\times 240\text{pixel}$ 。

3.2 评价指标

为验证本文所提出的改进算法ITD-YOLO的性能。实验采用平均精度(average precision, AP)、召回率(recall, R)、平均精度均值(mean average precision, mAP)、模型参数量(parameters)、模型体积(model volume)和浮点型计算量(FLOPs)为评价指标。

精确度反映模型检测出来的目标是真实目标的比例,TP是模型预测正确的正样本,FP是模型预测错误的负样本。精确度的表达式如下:

$$P = \frac{TP}{TP + FP} \quad (7)$$

召回率反映总样本中被模型预测正确的比例,FN为模型预测错误的正样本。召回率的表达式如下:

$$R = \frac{TP}{TP + FN}$$

(8)

平均精度代表模型对于这类目标的检测准确度，通常以 P - R 图中曲线求积分来计算，平均精度的表达式如下：

$$AP = \int_0^1 p(r) \, dr$$

(9)

平均精度均值就是所有类别的平均 AP 值，在多类别目标检测中用于衡量识别精度。平均精度均值的表达式如下：

$$mAP = \frac{1}{k} \sum_{i=1}^k AP_i$$

(10)

3.3 消融实验

为验证各项改进在实际应用中的有效性，消融实

表 3 消融实验

Table 3 Ablation experiments

Algorithm	Reconstruction +PConv	PolyLoss	SIoU	FLIR			OSU			Volume /MB	Parameters	GFLOPs
				P	R	mAP	P	R	mAP			
				/ (%)	/ (%)	/ (%)	/ (%)	/ (%)	/ (%)			
YOLOv7s				86.10	84.12	89.97	89.49	91.01	88.22	71.3	37207344	105.1
A	√			86.54	85.30	90.39	91.46	91.45	92.03	17.7	9152256	66.10
B		√		87.33	85.89	90.94	92.52	92.35	92.75	71.3	37207344	105.1
C			√	86.62	84.68	90.52	91.43	92.06	91.82	71.3	37207344	105.1
D	√	√		86.69	85.29	90.74	94.65	90.53	93.55	17.7	9152256	66.10
E		√	√	87.21	86.36	91.23	95.41	91.27	94.28	71.3	37207344	105.1
ITD-YOLO	√	√	√	88.15	86.94	92.02	96.33	93.17	94.65	17.7	9152256	66.10

3.4 对比实验

对比算法选择自 YOLOv5 发布至今的主流 YOLO 系列算法。YOLOv5 在前代的基础上进行改进，将主干网络中的 CSP 模块拓展到 Neck 层中，并且使用 SPPF 结构替换原有的 SPP 结构，锚框的聚类操作使用程序自适应计算。两年后发布的 YOLOv6 首次引入 RepVGG 结构，主干网络更换为 EfficientRep，在 Neck 层提出 Rep-PAN，使用 RepBlock 替换原 CSP-Block，重新设计解耦头。YOLOv7 由原 v4 开发团队发布，引入了模型结构重参化，使用动态标签分配方法。YOLOv8 在 v5 的基础上使用梯度流信息丰富的 C2f 模块替换掉原有的 C3，还有通过目标物体的中心点来定位的 Anchor free 策略和 Task Aligned Assigner 的正负样本匹配方法。

为验证各项改进在实际应用中的有效性，对比实验在 FLIR 与 OSU 数据集上进行，对比目标为当下

1386

验在 FLIR 与 OSU 数据集上进行。以 YOLOv7 为基准算法，实验使用平均精度、召回率、平均精度均值、模型体积、模型参数量、模型计算量为评价指标。消融实验结果如表 3 所示。

经过实验验证可得出结论：网络重构和轻量化处理成功缩减了模型体积，模型参数量和计算机大幅下降。模型在引入 PolyLoss 后，检测精度有显著提高。模型在引入 SIoU 边框回归损失函数后，召回率得到一定提升。经过实验进一步证实，PolyLoss 和 SIoU 可以在不增加模型大小的情况下，提升其性能。对比 YOLOv7s，ITD-YOLO 模型体积缩小 53.6MB，参数量减少 28055088，计算量降低 39 GFLOPs。算法在 FLIR 数据集上平均精度均值提升 2.27%，在 OSU 数据集上平均精度均值提升 7.29%，各项性能指标都有明显提升，验证了 ITD-YOLO 算法的先进性。

主流的 YOLO 系列目标检测算法。实验以平均精度、模型体积和推理时间为评价指标。对比结果见表 4、表 5。在 FLIR 数据集上实验得出：改进算法的平均检测精度均值为 92.02%，在对比算法中最优。在 OSU 数据集上实验得出：改进算法的平均检测精度均值为 94.65%，为对比算法中最优。从模型体积上看，ITD-YOLO 仅有 17.7 MB，相比于 YOLOv6s、YOLOv7s 和 YOLOv8s 算法体积分别降低了 54.3%、75.2%、21.3%。从推理时间上看，改进模型在 YOLOv7s 基础上下降了 15.8%。YOLOv5 在 FLIR 数据集上 mAP 仅有 66.67%，由于模型根据 COCO 数据集进行设计，缺乏对小目标的检测能力，当面对检测难度大的红外数据集且目标较小时难以应对。综上所述，ITD-YOLO 在各种算法对比中具有一定优势。其轻量化加高精度的特点能使其被轻松和高效地部署在硬件资源有限的平台上。

表4 FLIR 数据集算法对比

Table 4 Comparison of algorithms for the FLIR dataset			
Algorithm	Volume	mAP	Inference time
	/(MB)	/(%)	/ms
YOLOv5s	13.7	66.67	5.8
YOLOv6s	38.7	86.25	10.9
YOLOv7s	71.3	89.97	13.3
YOLOv8s	22.5	88.74	5.7
ITD-YOLO	17.7	92.02	11.2

3.5 实验效果与分析

为检验 ITD-YOLO 算法在真实运行过程中的检测效果,分别使用 FLIR 数据集与 OSU 数据集各自的测试集图片进行检测。最终检测效果如图 8 所示,编号(a)~(d)为 FLIR 数据集图片,编号(e)~(h)为 OSU 数据集图片。左列为 YOLOv7 算法检测结果,右列为 ITD-YOLO 检测结果。观察图片可知,原始

YOLOv7s 算法在面对较小目标或背景复杂的情况下无能为力,而改进方法能够在纹理不清晰或目标密集的图像下展现出更强的检测性能。由此可得出结论:ITD-YOLO 算法成功对 YOLOv7 进行了改进,提高了算法在面对困难样本时的处理能力,误检率与漏检率都得到下降。

表5 OSU 数据集算法对比

Table 5 Comparison of algorithms for OSU datasets			
Algorithm	Volume	mAP	Inference time
	/(MB)	/(%)	/ms
YOLOv5s	13.7	92.61	5.8
YOLOv6s	38.7	90.84	10.9
YOLOv7s	71.3	88.22	13.3
YOLOv8s	22.5	93.12	5.7
ITD-YOLO	17.7	94.65	11.2



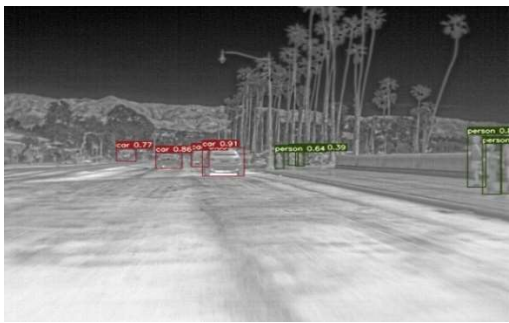
(a) YOLOv7s 对小目标的检测效果
(a) Effect of YOLOv7s in detecting small targets



(b) ITD-YOLO 对小目标检测效果
(b) Effect of ITD-YOLO in detecting small targets



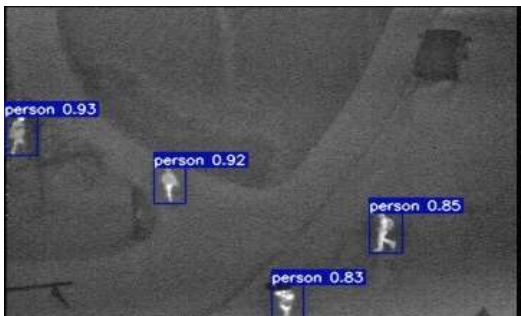
(c) YOLOv7s 在复杂背景下的检测效果
(c) YOLOv7s detection effect in complex backgrounds



(d) ITD-YOLO 在复杂背景下的检测效果
(d) ITD-YOLO detection effect in complex backgrounds



(e) YOLOv7s 在夜间下的检测效果
(e) YOLOv7s detection effect at night



(f) ITD-YOLO 在夜间下的检测效果
(f) Detection effect of ITD-YOLO at night

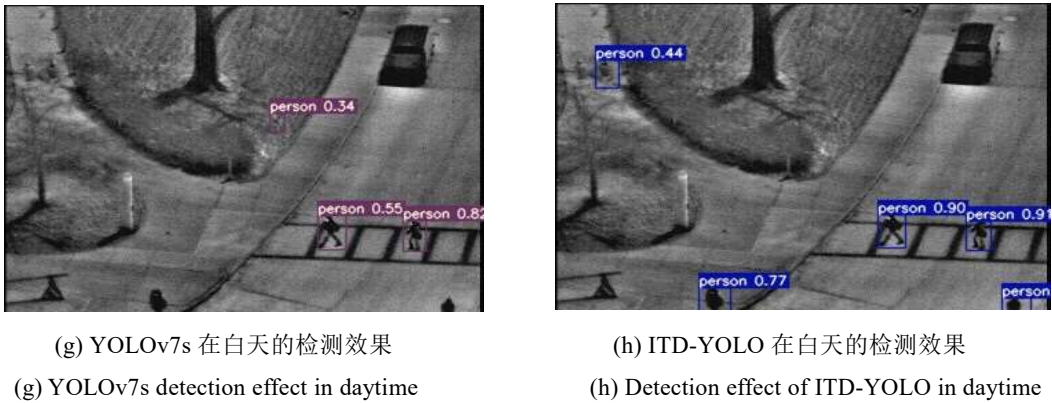


图8 检测效果对比

Fig.8 Comparison of detection results

4 结论

针对红外目标检测过程中常常面临的痛点,本文提出一种基于 YOLOv7 的改进算法。本算法成功实现轻量化目标,助力技术在硬件资源平台有限情况下的应用,同时很好地提高了在红外图像中目标检测的精度。本算法重新构建网络模型,生成新的感受野关系,提升浅层细节特征的利用率。其次,引入一种新型卷积,进一步推进轻量化。再次,使用 PolyLoss 替换原有置信度与分类损失函数。最后,引入性能更优的边框损失函数 SIoU。为验证方法的有效性,实验在公开数据集 FLIR 与 OSU 上完成。结果显示,本算法的平均精度均值在两个数据集上相较于 YOLOv7s 提升 2.27%与 7.29%。相较于前代算法,在 FLIR 数据集上, YOLOv5s、YOLOv6s 和 YOLOv8s 分别提升 38.02%、6.69%和 3.70%。在 OSU 数据集上, 分别提升 2.20%、4.19%和 1.64%。改进算法的模型体积仅为 17.7MB, 计算量下降 37.11%。因此, ITD-YOLO 具有良好性能,能满足对红外目标的高质量检测。

参考文献:

[1] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV), 2015: 1440-1448. DOI: 10.1109/ICCV.2015.169.

[2] REN S. Faster r-CNN: towards real-time object detection with region proposal networks[J]. arxiv preprint arxiv:1506.01497, 2015.

[3] HE K, Gkioxari G, Dollár P, et al. Mask r- CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.

[4] Bochkovskiy A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[5] Redmon J, Farhadi A. Yolov3: an incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[6] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.

[7] LIU W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//Computer Vision—ECCV 2016, 2016: 21-37.

[8] DUAN K W, BAI S, XIE L X, et al. Centernet: keypoint triplets for object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6569-6578.

[9] TANG C W, LIU C L, CHIU P S. HRCenterNet: an anchorless approach to Chinese character segmentation in historical documents[C]//2020 IEEE International Conference on Big Data (Big Data), 2020: 1924-1930.

[10] Vaswani A. Attention is all you need[J/OL]. Advances in Neural Information Processing Systems, 2017: <https://doi.org/10.48550/arXiv.1706.03762>

[11] 王恒涛, 张上, 陈想, 等. 轻量化无人机航拍目标检测算法[J]. 电子测量技术, 2022, 45(19): 167-174.

WANG Hengtai, ZHANG Shang, CHEN Xiang, et al. Lightweight target detection algorithm for drone aerial photography[J]. Electronic Measurement Technology, 2022, 45(19): 167-174.

[12] 王恒涛, 张上. 轻量化 SAR 图像舰船目标检测算法[J]. 电光与控制, 2023, 30(5): 99-104, 110.

WANG Hengtai, ZHANG Shang. Lightweight SAR image ship target detection algorithm[J]. Electro-Optics and Control, 2023, 30(5): 99-104, 110.

[13] 黄磊, 杨媛, 杨成煜, 等. FS-YOLOv5: 轻量化红外目标检测方法[J]. 计算机工程与应用, 2023, 59(9): 215-224.

HUANG Lei, YANG Yuan, YANG Chengyu, et al. FS-YOLOv5: lightweight infrared target detection method[J]. Computer Engineering and Applications, 2023, 59(9): 215-224.

[14] 贺顺, 谢永妮, 杨志伟, 等. 基于 IHBF 的增强局部对比度红外小目标检测方法[J]. 红外技术, 2022, 44(11): 1132-1138.

HE Shun, XIE Yongni, YANG Zhiwei, et al. Enhanced local contrast infrared small target detection method based on IHBF[J]. Infrared

Technology, 2022, **44**(11): 1132-1138.

[15] 李飏, 徐智勇, 王琛, 等. 基于自适应梯度倒数滤波红外弱小目标场景背景抑制[J]. 光电工程, 2021, **48**(8): 47-58.

LI Biao, XU Zhiyong, WANG Chen, et al. Adaptive gradient reciprocal filtering for infrared dim and small target scene background suppression[J]. *Opto-Electronic Engineering*, 2021, **48**(8): 47-58.

[16] 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, 2023, **45**(7):746-754.

LI Xiangrong, SUN Lihui. Multi-scale infrared target detection with attention mechanism fusion[J]. *Infrared Technology*, 2023, **45**(7): 746-754.

[17] BAO C, CAO J, HAO Q, et al. Dual-YOLO architecture from infrared and visible images for object detection[J]. *Sensors*, 2023, **23**(6): 2934.

[18] LI L, JIANG L, ZHANG J, et al. A complete YOLO-based ship detection method for thermal infrared remote sensing images under complex backgrounds[J]. *Remote Sensing*, 2022, **14**(7): 1534.

[19] HONG R, WANG X, FANG Y, et al. Yolo-light: remote straw-burning smoke detection based on depthwise separable convolution and channel attention mechanisms[J]. *Applied Sciences*, 2023, **13**(9): 5690.

[20] 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, 2023, **49**(8): 310-320.

LI Qianglong, ZHOU Xinwen, WEI Meng'en, et al. Infrared target detection algorithm in street scene based on stripe pooling and attention mechanism[J]. *Computer Engineering*, 2023, **49**(8): 310-320.

[21] 李杨, 武连全, 杨海涛, 等. 一种无人机视角下的小目标检测算法[J]. 红外技术, 2023, **45**(9): 925-931.

LI Yang, WU Lianquan, YANG Haitao, et al. A small target detection algorithm from drone perspective[J]. *Infrared Technology*, 2023, **45**(9): 925-931.

[22] CHEN J, KAO S, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[J]. arXiv preprint arXiv:2303.03667, 2023.

[23] LENG Z, TAN M, LIU C, et al. Polyloss: a polynomial expansion perspective of classification loss functions[J]. arXiv preprint arXiv:2204.12511, 2022.

[24] Gevorgyan Z. SIOU loss: More powerful learning for bounding box regression[J]. arXiv preprint arXiv:2205.12740, 2022.