

基于 YOLO-IDSTD 算法的红外弱小目标检测

蒋昕昊, 蔡伟, 杨志勇, 徐佩伟, 姜波

(火箭军工程大学兵器发射理论与技术国家重点学科实验室, 陕西西安 710025)

摘要: 针对复杂背景下红外弱小目标难以准确快速检测的问题, 提出了一种红外弱小目标轻量化实时检测网络模型 YOLO-IDSTD。首先, 为提高检测速度, 重新设计了特征提取部分的网络结构, 并在输入层后使用 Focus 模块以减少推理时间; 其次, 为增强检测能力, 特征融合部分采用路径聚合网络, 添加了改进的感受野增强模块; 最后, 目标检测部分增加至四尺度检测。在红外弱小目标数据集上进行的对比实验表明, 相较于经典轻量化模型 YOLOv3-tiny, 文中提出的模型召回率提升了 7.57%, 平均检测精度提高了 1.92%, CPU 推理速度提升了 36.1%, 可较好地兼顾精度和速度, 计算量与参数量明显减少, 模型尺寸压缩至 7.27 MB, 减少了对硬件平台运算能力的依赖, 实现了红外弱小目标准确又快速的检测。

关键词: 红外弱小目标; 深度学习; 目标检测; 机器视觉; YOLO

中图分类号: TP391.4 **文献标志码:** A **DOI:** 10.3788/IRLA20210106

Infrared dim and small target detection based on YOLO-IDSTD algorithm

Jiang Xinhao, Cai Wei, Yang Zhiyong, Xu Peiwei, Jiang Bo

(Armament Launch Theory and Technology Key Discipline Laboratory of PRC,
Rocket Force University of Engineering, Xi'an 710025, China)

Abstract: Aiming at the problem that it is difficult to detect infrared dim and small target accurately and quickly in complex background, a lightweight real-time network model YOLO-IDSTD for infrared dim and small target detection was proposed. Firstly, in order to improve the detection speed, the network structure of the feature extraction part was redesigned, and the Focus module was used to reduce the reasoning time after the input layer. Secondly, in order to enhance the detection ability, the path aggregation network was adopted in the feature fusion part and an improved receptive field block was added. Finally, four-scales detection was increased in the target detection part. Compared with the classical lightweight model YOLOv3-tiny on the infrared dim and small target data set, the recall is increased by 7.57%, the average precision is increased by 1.92%, and the CPU reasoning speed is increased by 36.1%. The model can balance accuracy and speed, and the amount of calculation and parameters are significantly reduced. The size of the model is compressed to 7.27 MB, which reduces the dependence on the computing power of the hardware platform and realizes the accurate and fast detection of infrared dim and small targets.

Key words: infrared dim and small target; deep learning; object detection; machine vision; YOLO

收稿日期: 2021-02-20; 修订日期: 2021-05-12

作者简介: 蒋昕昊, 男, 博士生, 主要从事光电防护、计算机视觉方面的研究。

导师简介: 蔡伟, 男, 教授, 博士生导师, 博士, 主要从事光电防护方面的研究。

0 引言

军事领域中,光电对抗已经成为高科技战争决定成败的要害环节。其中,红外搜索和跟踪系统是有效防御敌方空中侦察打击威胁的“视觉中枢”,其目标捕捉速度与准确性直接决定了己方人员装备的生存能力,因此,红外弱小目标快速检测技术一直都是该领域的研究重点。

国际光学工程学会将成像面积小于0.15%,即在256 pixel×256 pixel图像中小于81 pixel的目标称为弱小目标^[1]。而当前的红外弱小目标检测面临着距离远、成像目标小及像素分辨率低、信噪比不高等诸多困难与挑战。

当前的红外弱小目标检测算法可以分为两类:基于目标与背景有效分离的传统检测算法和基于特征提取的深度学习目标检测算法。其中,传统算法高度依赖于红外图像特性的建模,通过分析目标与背景的特性模型,进行二者的有效分离,如参考文献[2]基于人类视觉系统机制发掘具有显著性对比度差异的目标,参考文献[3]基于目标稀疏性将检测问题转化为最优化问题。传统算法虽易于工程化,但不同的检测场景往往需要根据各自特点建立不同的模型,泛化能力较差。

近年来,卷积神经网络(CNN)凭借其特征提取能力,已被广泛应用到目标检测任务当中。目前已经形成了两类完备的CNN目标检测算法:两步检测算法和一步检测算法。两步检测算法的核心思想是生成候选区域,并对区域进行分类,代表性方法是Regions with CNN (R-CNN)^[4]系列算法。此类算法精度较高,但提取候选区域的时间长,实时性不佳。而一步检测算法不需要提取候选区域,采用网格划分的思想,在网格中完成目标检测,经典的一步检测算法有You Only Look Once (YOLO)算法^[5]和Single Shot MultiBox Detector (SSD)算法^[6],这种端到端的方法,在不过分损失精度的情况下,大幅缩短了检测时间。参考文献[7-8]分别基于两步检测算法和一步检测算法,通过增强特征提取能力,提高对小目标的检测精度,但都忽略了检测速度。然而,瞬息万变的战场态势对系统的响应速度要求极高,同时考虑到平台部署的性能约束条件,红外弱小目标的实时检测和模型的轻量化

变得尤为重要。YOLOv3-tiny是YOLO系列的轻量版,参数量的减少明显提升了模型的检测速度,但其特征提取网络深度不够,导致高层次语义信息提取能力不足,影响了检测精度^[9]。

针对现实需求,文中提出了更适用于红外弱小目标的轻量化实时检测模型。首先,在特征提取部分,使用聚焦(Focus)结构^[10]添加至输入层用以提升推理速度,又在末端设计了深度可分离卷积残差模块^[11],提高网络深度的同时大幅减少了模型参数;其次,通过四尺度检测的方式,使用路径聚合网络PANet^[12]进行高层语义信息与低层位置信息的深度融合;同时,在输出层前引入了改进的感受野模块(Receptive Field Block, RFB)^[13]用于聚焦感受野中心特征信息。与同类型模型相比发现,该模型可在保证检测能力的同时,大幅提升检测速度,减小模型参数量以及对硬件平台运算能力的依赖。

1 YOLO-IDSTD 目标检测模型

文中提出的YOLO-IDSTD目标检测模型网络结构如图1所示。图片输入后,通过特征提取部分学习特征信息,而后在特征融合部分进行多尺度特征信息融合,目标检测部分完成最后的目标位置与边界框的预测输出。下面分别对三个组成部分的设计思路 and 具体结构进行详细论述。

1.1 特征提取部分

针对板上集成和实时检测的现实需求,设计了一个轻量级深层特征提取部分。网络结构由Focus结构和多阶段卷积池化模块两部分组成。

1.1.1 Focus 结构

在特征提取网络前端添加Focus结构,通过按行和按列周期性间隔采样的方法,将输入的高分辨率图像进行像素级分割分块,对输入图像相邻位置的四个像素进行通道堆叠,使得平面维度的信息扩展到通道维度当中,最终得到四倍于原通道数的输出;而后用1×1大小的普通卷积层对该输出进行卷积操作提取特征,生成新的特征图输入下一特征提取层。Focus结构如图2所示,其中, n 为输入特征图的长和宽, n_c 为卷积核的个数,文中 $n_c=16$ 。分块堆叠操作不涉及参数学习,相当于一次几乎不损失原始特征信息的下采样操作,形成原特征图大小1/4的新特征图,从而

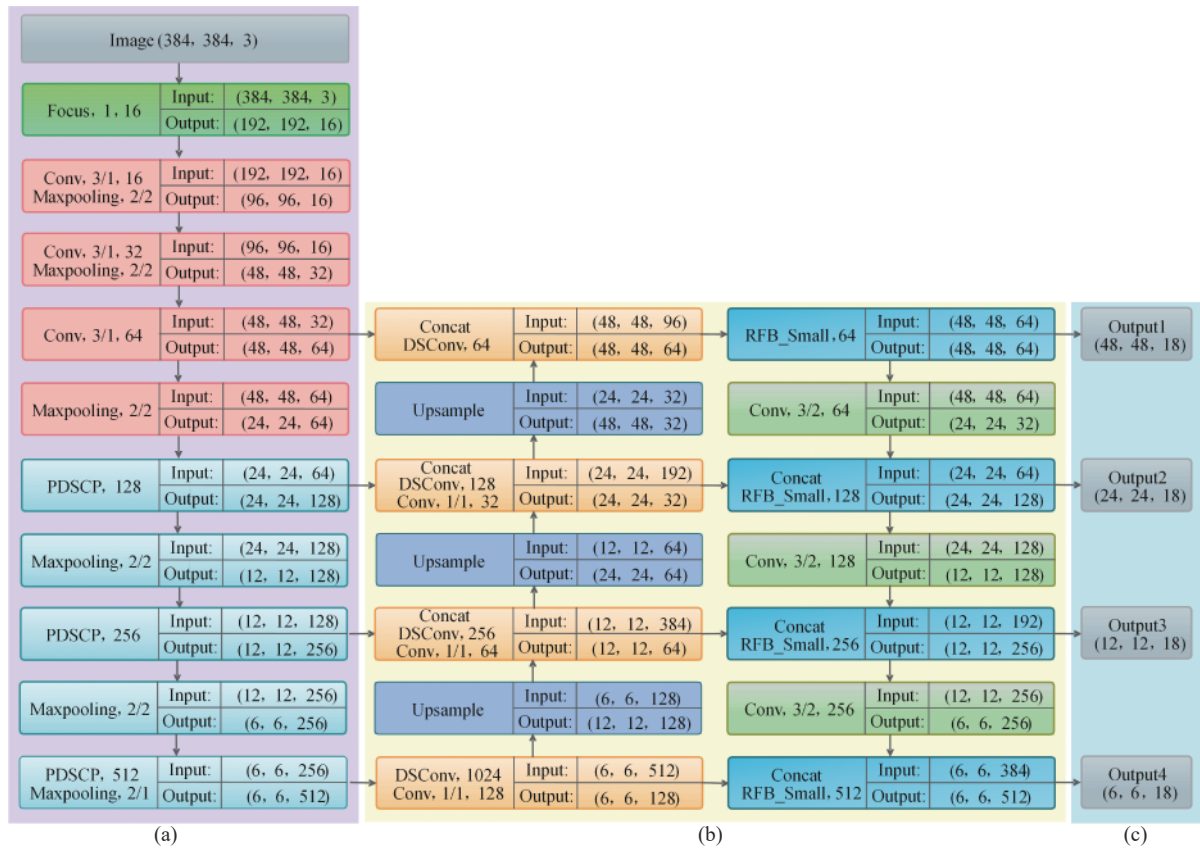


图 1 YOLO-IDSTD 网络结构。(a) 特征提取部分; (b) 特征融合部分; (c) 目标检测部分

Fig.1 YOLO-IDSTD network structure. (a) Feature extraction part; (b) Feature fusion part; (c) Target detection part

使得整体网络模型在几乎不损失精度的同时, 计算量显著减少, 训练及推理速度明显提高。

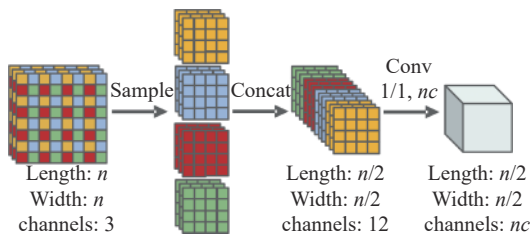


图 2 Focus 结构图

Fig.2 Structure of Focus

1.1.2 多阶段卷积池化模块

考虑到红外弱小目标的目标特性, 特征提取网络不应过于复杂; 且应控制参数量与计算复杂度以适用于板上集成。因此, 在 Focus 结构之后, 特征图仅通过六次卷积与池化操作提取特征信息, 前三个低层特征提取模块各使用一组普通的 3×3 大小的卷积层和池化层, 后三个高层特征提取模块使用重新设计的深度可分离卷积残差模块代替普通卷积层, 并将其称之为

为 PDSCP 结构。

PDSCP 结构如图 3 所示, 设计目的是提高网络模型的轻量化程度, 同时保留网络的特征提取能力。首先将输入特征图进行 1×1 的逐点卷积 (PWConv) 操作, 起到增强梯度传播的作用, 而后使用深度可分离卷积 (DSCConv) 进行特征提取。DSCConv 是深度卷积 (DWConv) 和逐点卷积的组合形式, 可以在几乎不损失精度的同时, 有效减少参数量。但层数的加深容易产生梯度消失现象, 因此再将逐点卷积后的特征图与深度可分离卷积后的特征图进行通道上的叠加, 最后再次通过逐点卷积进一步增强通道间的信息融合。

下面分析 PDSCP 结构的轻量化程度, 设输入特征图大小为 $M \times M \times N$, 则输出特征图大小为 $M \times M \times 2N$ 。常用的普通卷积层尺寸为 3×3 , 卷积层参数量为 $3 \times 3 \times N \times 2N$, 计算量为 $3 \times 3 \times N \times 2N \times M \times M$; 而 PDSCP 结构的总参数量为 $N \times N + 3 \times 3 \times N + N \times 2N + 3N \times 2N$, 总计算量为 $N \times N \times M \times M + 3 \times 3 \times N \times M \times M + N \times 2N \times M \times M + 3N \times 2N \times M \times M$, 两种结构的参数量和计算量比值均为 $\frac{1}{2} + \frac{1}{2N}$ 。

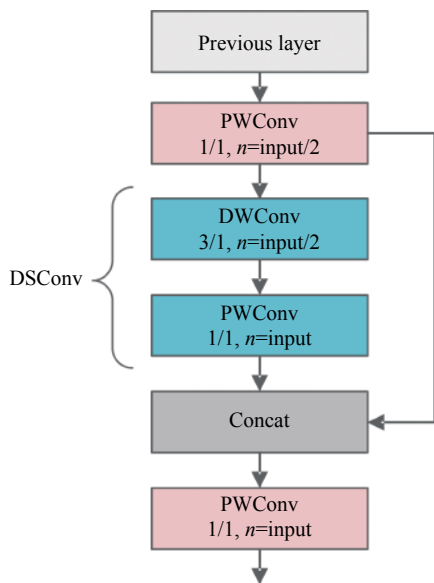


图 3 PDSCP 结构图

Fig.3 Structure of PDSCP

而在 YOLO-IDSTD 中, $N \geq 64$, 故文中设计的 PDSCP 结构层数更深, 但参数量和计算量均约缩减为 3×3 卷积层的 $1/2$ 。

YOLO-IDSTD 模型特征提取部分各层的参数量和计算量如表 1 所示。

表 1 特征提取部分各层的参数量和计算量

Tab.1 Each layer's parameters and FLOPs of feature extraction part

| No. | Name | Parameter | FLOPs |
|-----|---------------|----------------------|--------------------|
| 1 | Focus, 1, 16 | 224×10^6 | 33.0×10^6 |
| 2 | Conv, 3/1, 16 | 2336×10^6 | 86.1×10^6 |
| 3 | Conv, 3/1, 32 | 4672×10^6 | 43.1×10^6 |
| 4 | Conv, 3/1, 64 | 18560×10^6 | 42.8×10^6 |
| 5 | PDSCP, 128 | 38016×10^6 | 21.9×10^6 |
| 6 | PDSCP, 256 | 149760×10^6 | 21.6×10^6 |
| 7 | PDSCP, 512 | 594432×10^6 | 21.4×10^6 |

1.2 特征融合部分

红外弱小目标具有尺寸小、有效检测特征少的特点, 每次下采样都会使目标在最后的特征图上的尺度越来越小, 运行到特征提取的末端, 虽然可以较好地提取高层语义信息, 但过高的下采样率会造成特征信息的过度丢失。如果仅通过降低下采样率的方法减少信息丢失, 又会使得高层语义信息提取不充分, 影响检测精度。

针对上述问题, YOLO-IDSTD 网络模型通过使用 PANet 进行特征深度融合, 同时将改进的感受野模块替换输出层前的卷积层, 提升特征提取能力。结合图 1, 下面对特征融合部分的三个阶段进行介绍。

1.2.1 自下而上的特征融合

首先对特征提取部分学习到的不同层次的特征图进行自下而上的特征融合, 使用 1×1 的卷积层进行特征图通道数目的缩减, 而后进行上采样完成尺度变换, 进而与同尺度的低层特征图拼接 (Concat), 以实现高层与低层特征图的融合。

融合后特征图的大小分别为输入图片的 $1/64$ 、 $1/32$ 、 $1/16$ 和 $1/8$, 从而实现了各个尺度的高低层特征信息的充分利用, 提高了低层特征的利用率, 增强了弱小目标的检测能力。

1.2.2 自上而下的路径增强

在自下而上的特征融合之后, 进行自上而下的路径增强, 使用步长为 2 的卷积进行下采样操作, 完成尺度变换后, 与同尺度的特征图拼接, 进行维度上的叠加。缩短了低层特征的信息传递路径, 无需通过多次特征提取便可直接输入特征融合网络, 增强了各层之间的信息传递效率。

1.2.3 改进的感受野模块

人类视觉感受野机制更加关注目标中心, 感受野内特征的重要性由中心到边缘递减。RFB 模块通过模拟人类视觉感受野增强机制, 在增大感受野的同时提升感受野中心低层特征的重要性, 从而提升了特征提取能力, 更有利于检测红外弱小目标这类中心位置信息重要的目标。

由于红外弱小目标尺寸较小, 过大的感受野会纳入过多的背景信息, 影响检测精度。RFB-Small 是该模块适用于小目标的版本, 感受野增强的程度相对较小, 但实验结果显示, 对于红外弱小目标检测仍然适用性不强。因此, 文中对 RFB-Small 进行了针对性改进, 结合实验效果, 减小最大感受野增强分支的空洞卷积膨胀系数, 将依次通过 1×1 和 3×3 卷积层处理后的特征图输入至膨胀系数为 3 的空洞卷积层, 减小该分支感受野, 从而减少背景信息的干扰。改进后的 RFB-Small 模块如图 4 所示, 将其替换掉输出层前的卷积层以增强小目标的检测能力。

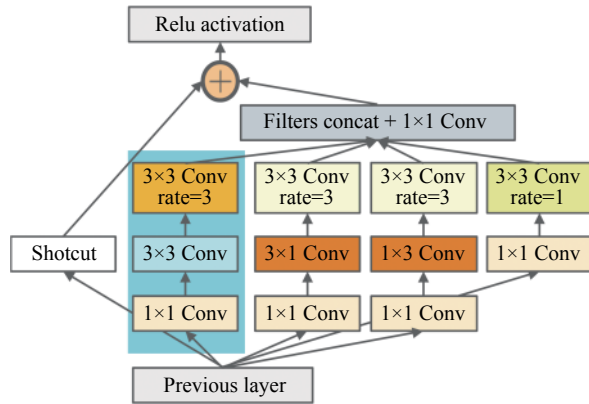


图 4 改进后的 RFB-Small 模块

Fig.4 Improved RFB-Small block

1.3 目标检测部分

考虑到红外弱小目标所占的像素尺寸小,常用的三尺度检测无法满足弱小目标的检测输出。故 YOLO-IDSTD 网络模型增加了上采样次数,利用了更低层的特征信息。最终,多尺度融合特征图通过四尺度目标检测部分实现目标位置与边界框的预测输出。

损失函数方面,由于红外弱小目标检测是单类别检测,故损失函数仅由坐标位置误差和置信度误差组成,数学表达式为:

$$L=L_{\text{coord}}+L_{\text{conf}} \quad (1)$$

式中: L_{coord} 为坐标位置误差,选择二分类交叉熵损失函数(Binary Cross Entropy Loss with Logits)以增加数值稳定性; L_{conf} 为坐标位置误差,选择 Complete-IOUloss(CIOUloss)^[14]作为损失函数,使得框目标回归变得更加稳定。

结合上述分析,文中提出的 YOLO-IDSTD 目标检测模型可兼顾轻量化与特征提取能力,具有高低层特征信息融合度高、低层特征信息利用率高的优点,更有利于检测红外弱小目标。

2 实验结果与分析

2.1 准备工作

实验训练测试阶段使用的硬件平台配置如表 2 所示,使用 Pytorch 深度学习开发框架。

表 2 实验平台配置

Tab.2 Configuration of experimental platform

| Name | Related configurations |
|--------------------|------------------------------|
| GPU | NVIDIA quadro GV100 |
| CPU | sInter Xeon silver 4210/128G |
| GPU memory size | 32G |
| Operating systems | Win10 |
| Computing platform | CUDA11.0 |
| CPU(test) | Inter Core i7 10700/16G |

2.1.1 数据集预处理

文中使用的数据集为公开红外弱小目标数据集^[15]。该数据集共有 22 个不同场景的子数据集,采集了涵盖地面背景、天空背景和地空交界背景复杂环境下的无人机飞行红外序列图像,图像尺寸为 256 pixel×256 pixel。图 5 展示了数据集中满足红外弱小目标定义的部分单帧图像,右上角为放大后的目标。

原始数据集中有 7 个能清晰辨别形状、不满足定义子数据集,因此,文中首先进行了数据筛选,仅保留如图 6 所示满足弱小目标定义子数据集,共 15 个子数据集,13282 张图像,其中 13 个子数据集目标尺寸小于 5 pixel×5 pixel,两个子数据集目标尺寸小于 7 pixel×7 pixel。

由于该数据集仅提供目标中心点位置,而当前的

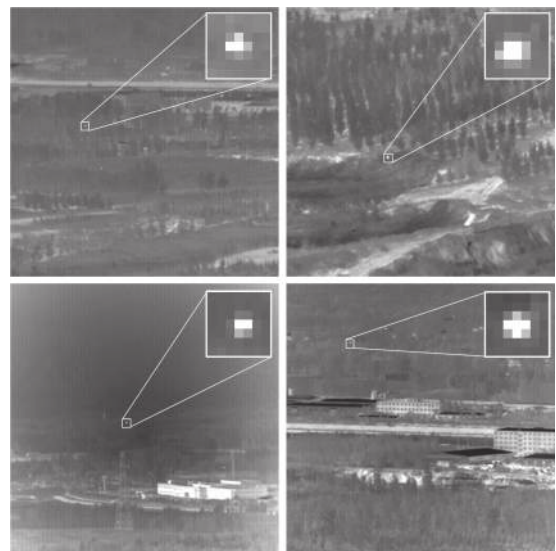


图 5 数据集部分图像

Fig.5 Some images of data set



图 6 数据集中的典型红外弱小目标

Fig.6 Typical infrared dim and small targets in data set

卷积神经网络对于点目标很难进行有效的特征提取,故需要增强数据集输入给网络模型的有效特征范围。

然而手动标注弱小目标误差极大,因此,将点目标统一用扩展框进行标注,使网络模型能够更好地学习到目标与背景的差异性以及目标的有效特征。用 5 pixel×5 pixel 的扩展框标注目标尺寸较小的 13 个子数据集,用 7 pixel×7 pixel 的扩展框标注两个子数据集中稍大的目标。转化为 VOC 数据格式,两种扩展框的数量统计如表 3 所示。

不分扩展框大小,将该数据集随机划分为 11954

表 3 扩展框的数量统计

Tab.3 Statistics of extension box

| Size of extension box | Number of datasets | Number of images |
|-----------------------|--------------------|------------------|
| 5 pixel×5 pixel | 13 | 12484 |
| 7 pixel×7 pixel | 2 | 798 |

张图片组成的训练集、1195 张图片组成的验证集和 133 张图片组成的测试集。同时,为验证模型的综合检测能力,文中在 OSU Thermal Pedestrian Database^[16] (红外行人小目标数据集)和 FLIR Thermal Datasets^[17] (FLIR 红外数据集)上均做了对比实验。

2.1.2 实验参数设置

通过实验对超参数进行了选取,优化算法采用随机梯度下降法,初始学习率设为 0.01,引入动量因子并设置为 0.937,权重衰减设置为 0.0005。在各数据集训练时的模型参数设置总结如表 4 所示。

表 4 实验参数设置

Tab.4 Setting of experimental parameters

| Parameter | Infrared dim and small targets datasets | Thermal Pedestrian Database | FLIR Thermal Datasets |
|------------------|---|-----------------------------|-----------------------|
| Class number | 1 | 1 | 3 |
| Epoch | 500 | 500 | 500 |
| Batch size | 64 | 4 | 64 |
| Image size | 384×384 | 320×320 | 512×512 |
| Batch size(test) | 1 | 1 | 1 |

2.1.3 评价指标的选取

深度学习中检测准确率 (Precision rate) 和目标召回率 (Recall rate) 的定义如下:

$$P = \frac{n_t}{n_t + n_f} \quad (2)$$

$$R = \frac{n_t}{n_r} \quad (3)$$

式中: n_t 为算法检测到的真实弱小目标的数目; n_r 为图像中实际存在的真实弱小目标的数目; n_f 为算法误检出的虚假目标的数目。

在侦察告警领域中,召回率体现了该系统将全部真实目标检测出来的能力,该指标对于提升己方防护能力来说尤为重要,因此笔者选取召回率作为首要的评价指标。

同时笔者也选取平均精度 AP 和模型推理时间作为评价指标。考虑到侦察预警系统无显卡,且 CPU 计算能力有限,故将模型推理时间转移到性能低的 CPU 上进行测试,取 CPU 处理测试集中单张图片 (Batchsize 设置为 1) 的平均值进行对比。

2.2 在红外弱小目标数据集上的实验效果对比

为了验证文中提出的 YOLO-IDSTD 网络模型对红外弱小目标检测所具有的针对性,在处理后的红外弱小目标数据集上进行了对比实验。对比算法涵盖范围广泛,其中,SSD300、YOLOv3 和 Mobilenet-SSD 为目标检测经典模型,Efficientdet b0^[18]、Centernet-ResNet50^[19]、YOLOv4-tiny^[20]、YOLOv5s 是最新公布的高性能快速检测模型,YOLOv3-tiny 是广泛应用轻量化网络模型。实验结果如表 5 所示。

由表 5 可知,相较于网络结构更为复杂的 YOLOv3 网络,YOLO-IDSTD 仅在检测准确率上损失了约 9 个百分点,这意味着使用复杂网络检测红外弱小目标时虚警率更低,但检测无法满足高召回率。SSD300 虽网络提取能力较强,但网络过深导致弱小目标特征丢失严重。实验表明:复杂模型对检测特征稀少的红外弱小目标适用性不强,且参数量大,检测时间长,不适用于板上集成和实时检测红外弱小目标。

表 5 不同检测方法的精度与效率

Tab.5 Precision and efficiency of different detection methods

| Method | Precision rate | Recall rate | AP@0.5 | Parameter | Model size/MB | GFLOPs | Detection time/ms |
|--------------------|----------------|---------------|---------------|-------------------------------------|---------------|-------------|-------------------|
| YOLOv3-384 | 0.7371 | 0.8182 | 0.8123 | 61.6×10^6 | 117.7 | 155.2 | 364.8 |
| SSD300 | 0.3664 | 0.7585 | 0.5170 | 23.7×10^6 | 90.6 | 35.2 | 370.4 |
| Mobilenet-SSD | 0.5241 | 0.5111 | 0.3300 | 6.3×10^6 | 24.0 | 1.14 | 66.8 |
| Efficientdet b0 | 0.5948 | 0.0589 | 0.0999 | 3.9×10^6 | 15.1 | 2.5 | 73.8 |
| Centernet-ResNet50 | 0.8323 | 0.6156 | 0.6843 | 32.6×10^6 | 124.8 | 3.8 | 30.3 |
| YOLOv5s-384 | 0.7310 | 0.8029 | 0.7957 | 7.3×10^6 | 16.6 | 17.0 | 98.5 |
| YOLOv4-tiny--384 | 0.6713 | 0.7847 | 0.8195 | 6.2×10^6 | 12.6 | 16.5 | 80.1 |
| YOLOv3-tiny-384 | 0.6780 | 0.7652 | 0.8050 | 8.9×10^6 | 14.2 | 12.9 | 78.5 |
| YOLO-IDSTD | 0.6405 | 0.8409 | 0.8242 | 3.7×10^6 | 7.3 | 3.0 | 50.2 |

Mobilenet-SSD 和 Efficientdet b0 虽然网络结构更加简洁、更加轻量化,但实验证明此类模型不具备红外弱小目标的准确检测能力。Centernet-ResNet50 作为无框检测的代表算法,检测速度大幅缩短,但检测准确性仍有提升空间。相较于 YOLOv4-tiny、YOLOv5s 和 YOLOv3-tiny, YOLO-IDSTD 虽然准确率稍有下

降,但是召回率、平均精度和单张图片的检测推理时间,这三项更能表明红外弱小目标检测性能的指标明显提升。与 YOLOv3-tiny baseline 网络相比,召回率提升了 7.57%,平均精度 AP@0.5 提高约 2%,与此同时,CPU 测试单张图片的时间减少了 28.3 ms,速度提升 36.1%。图 7 为四种 AP@0.5 高于 0.8 的网络模型

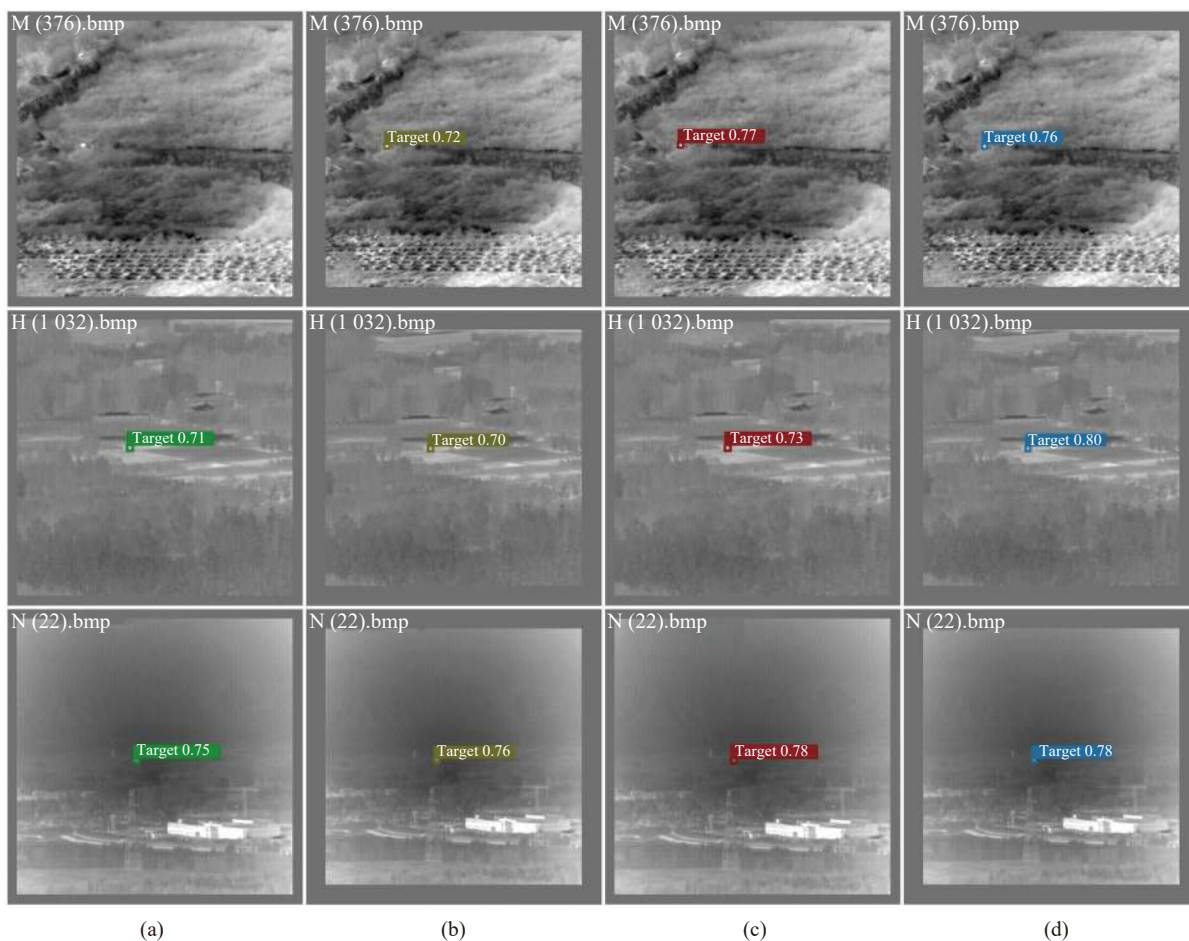


图 7 典型红外弱小目标检测结果对比。(a) YOLOv3; (b) YOLOv4-tiny; (c) YOLOv3-tiny; (d) YOLO-IDSTD

Fig.7 Comparison of detection results of typical infrared dim and small targets. (a) YOLOv3; (b) YOLOv4-tiny; (c) YOLOv3-tiny; (d) YOLO-IDSTD

检测数据集中不同场景红外弱小目标的可视化结果, 可以发现, YOLO-IDSTD 网络模型平均置信度最高, 检测更加准确。

因此, 结合检测能力和推理速度两方面分析, 文中提出的 YOLO-IDSTD 网络模型更为轻量化, 检测红外弱小目标可以做到低漏警率, 且实现了较好的实

时性检测效果。

2.3 在红外弱小目标数据集上的消融实验

为了验证相较于 YOLOv3-tiny baseline 网络, YOLO-IDSTD 网络模型中各个关键模块对检测性能的影响, 在红外弱小目标数据集上进行了消融实验, 实验结果如表 6 所示。

表 6 YOLO-IDSTD 的消融实验

Tab.6 Ablation experiment of YOLO-IDSTD

| Improve the detection speed | | Improve the detection accuracy | | | | Recall rate | AP@0.5 | Model size/MB | Detection time/ms |
|-----------------------------|------------|--------------------------------|------------|-----------------------------|----------------|---------------|---------------|---------------|-------------------|
| YOLOv3-tiny baseline | With Focus | With PDSCP | With PANet | With Four-scales prediction | With RFB-Small | | | | |
| √ | | | | | | 0.7879 | 0.7771 | 16.59 | 78.5 |
| √ | √ | | | | | 0.7576 | 0.7235 | 16.59 | 35.4 |
| √ | √ | √ | | | | 0.7652 | 0.7342 | 3.65 | 26.5 |
| √ | √ | √ | √ | | | 0.7652 | 0.7604 | 9.11 | 31.7 |
| √ | √ | √ | √ | √ | | 0.8258 | 0.8037 | 9.22 | 36.9 |
| √ | √ | √ | √ | √ | √ | 0.8409 | 0.8242 | 7.27 | 50.2 |

可以看出, 通过添加 Focus 结构和重新设计的 PDSCP 深度可分离残差结构减小了特征提取部分的参数量, 使得 CPU 检测单张图像的速度显著提升, 模型大小锐减 78%, 轻量化程度最高, 但是精度较 baseline 网络有一定程度上的下降; 通过借鉴使用 PANet 网络结构, 增强了不同尺度特征图的融合, 在几乎不损失速度的情况下提高了 3.4% 的检测精度; 增加至四尺度预测后, 低层特征利用率更高, 可以大幅度提高小目标的检测能力, 召回率及 AP 值均已突破原网络; 最后, 将改进后的 RFB-Small 模块更换输出层前的原有卷积层, 提升感受野中心低层特征的重要性, 在减小模型大小的同时, 对召回率及 AP 值均有明显的提高效果。最终使得 YOLO-IDSTD 网络可以鲁棒又快速地检测红外弱小目标。

2.4 在红外小目标数据集上的实验效果对比

为了验证文中提出的模型针对红外波段小尺寸目标的综合检测能力, 在 OSU Thermal Pedestrian Database 和 FLIR Thermal Datasets 两个红外小目标数据集上做了对比实验。

OSU Thermal Pedestrian Database 共有 284 张 360 pixel×240 pixel 的红外图像, 检测目标为道路上的行人小目标。FLIR Thermal Datasets 是用于无人驾驶

汽车的红外数据集, 是车上配备的红外成像设备采集到的不同复杂场景的道路行驶序列图像, 标注目标包含人、自行车和汽车三类, 共 10228 张 640 pixel×512 pixel 的红外图像。两个数据集中成像尺寸小于 1% 的小目标占绝大多数, 故对算法验证有一定的参考价值。与同类型轻量化模型进行对比实验。实验结果如表 7 所示。

两组实验数据证明, YOLO-IDSTD 在红外波段小目标数据集上也具有较好的检测能力。在红外小行人目标数据集中, 由于目标所占像素均小于整幅图像的 1%, 故 YOLO-IDSTD 可以很好地满足小尺寸行人目标的检测, 精度和速度优于其他算法。但在 FLIR 红外数据集中, 由于环境多样、目标尺寸大小不一, YOLO-IDSTD 网络的优势发挥不明显。模型更为复杂、参数量更多的 YOLOv5s 在检测精度上领先于 YOLO-IDSTD, 但在文中重点关注的实时性方面, YOLO-IDSTD 显著优于其他算法, 同时 mAP 较 YOLOv3-tiny 提升 3.4%, 召回率提升 2.6%, 证明该模型具有较好的红外波段小尺寸目标的综合检测能力。

图 8 为 YOLO-IDSTD 在 OSU Thermal Pedestrian Database 上的检测结果。图 9 为在 FLIR Thermal Datasets 上的检测结果, 可以看出, YOLO-IDSTD 模型

表 7 在两组红外小目标数据集上的对比实验

Tab.7 Comparative experiments on two sets of infrared small target datasets

| Method | OSU Thermal Pedestrian Database | | | FLIR Thermal Datasets | | | | | |
|-----------------|---------------------------------|--------|-----------------------|-----------------------|---------|--------------------|---------------------|-----------------|-----------------------|
| | Recall rate | AP@0.5 | Detection time/ ms | Recall rate | mAP@0.5 | AP@0.5 (person) | AP@0.5 (bicycle) | AP@0.5 (car) | Detection time/ ms |
| Efficientdet b0 | 0.8723 | 0.8651 | 90.5 | 0.3374 | 0.4943 | 0.444 | 0.435 | 0.604 | 160.8 |
| YOLOv5s | 0.9909 | 0.9860 | 69.6 | 0.7706 | 0.7441 | 0.799 | 0.563 | 0.870 | 122.6 |
| YOLOv3-tiny | 1 | 0.9875 | 53.2 | 0.6906 | 0.6334 | 0.641 | 0.449 | 0.810 | 98.4 |
| YOLO-IDSTD | 1 | 0.9899 | 42.9 | 0.7166 | 0.6676 | 0.724 | 0.448 | 0.831 | 60.7 |

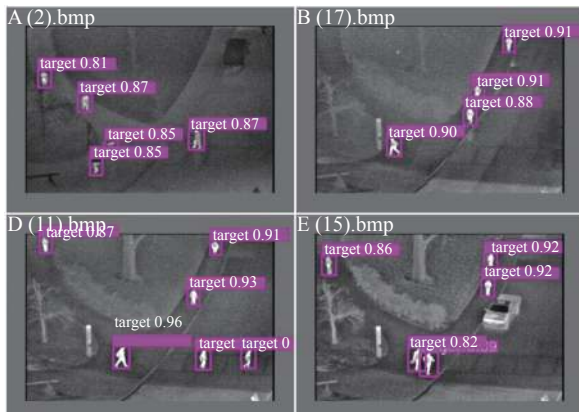


图 8 在 OSU Thermal Pedestrian Database 上的检测结果
Fig.8 Test results on OSU Thermal Pedestrian Database



图 9 在 FLIR Thermal Datasets 上的检测结果
Fig.9 Test results on FLIR Thermal Datasets

可以以较高的置信度检测出图中的红外小目标。

3 结 论

针对同时提高复杂背景下红外弱小目标检测速度和检测准确性的难题,文中提出了一种适用性更强

的 YOLO-IDSTD 网络模型。在速度提升方面,输入层后添加了 Focus 结构,设计了轻量化的 PDSCP 深度可分离残差结构,显著降低了网络模型的计算量与参数量;在精度提升方面,借鉴 PANet 网络设计了特征融合部分,且增加至四尺度检测,同时在输出层前添加改进后的感受野增强模块,做到了不同尺度特征信息的深度融合。多组对比实验可以表明,文中提出的 YOLO-IDSTD 网络模型在红外弱小目标数据集上召回率可达 84.09%,CPU 处理单张图像的时间缩短至 50.2 ms,可在保证良好检测精度的同时,显著提升推理速度,达到了实时性准确检测的设计目的。同时,实验证明该模型对于红外波段的小尺寸目标也具有较好的检测能力。

参考文献:

- [1] Zhang W, Cong M, Wang L. Algorithms for optical weak small targets detection and tracking: Review[C]//International Conference on Neural Networks and Signal Processing, Proceedings of the 2003. IEEE, 2003.
- [2] Du P, Hamdulla A. Infrared small target detection using homogeneity-weighted local contrast measure [J]. *IEEE Geoece and Remote Sensing Letters*, 2020, 17(3): 514-518.
- [3] Pan S D, Zhang S, Zhao M, et al. Infrared small target detection based onweighted scene prior [J]. *Journal of Infrared and Millimeter Waves*, 2019, 38(5): 633-641. (in Chinese)
- [4] 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. New York: IEEE, 2014: 580-587.
- [5] 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 (CVPR). New York: IEEE, 2016: 779-788.

- [6] 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.
- [7] Zhao Xiaofeng, Xu Mingyang, Wang Danpiao, et al. Infrared camouflage detection method for special vehicles based on improved SSD [J]. *Infrared and Laser Engineering*, 2019, 48(11): 1104003. (in Chinese)
- [8] Ju M R, Luo H B, Wang Z B, et al. Improved YOLO v3 algorithm and its application in small target detection [J]. *Acta Optica Sinica*, 2019, 39(7): 0715004. (in Chinese)
- [9] Qi H Y, Xu T H, Wang G, et al. MYOLOv3-Tiny: A new convolutional neural network architecture for real-time detection of track fasteners [J]. *Computers in Industry*, 2020, 123: 103303.
- [10] Glenn Jocher, Alex Stoken, Jirka Borovec, et al. Ultralytics/yolov5: v3.1-Bug Fixes and Performance Improvements (Version v3.1) [EB/OL]. (2020-10-29)[2021-01-19]. <https://zenodo.org/record/4154370>.
- [11] Howard J, Zhu M L, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-05-17)[2021-01-19]. <https://arxiv.org/abs/1704.04861>.
- [12] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 8759-8768.
- [13] Liu S T, Huang D, Wang Y H. Receptive field block net for accurate and fast object detection[C]//Proceedings of the European Conference on Computer Vision(ECCV). New York: IEEE, 2018: 385-400.
- [14] Zheng Z, Wang P, Liu W, et al. Distance-IoU Loss: Faster and better learning for bounding box regression[C]//AAAI Conference on Artificial Intelligence, 2020.
- [15] Hui B W, Song Z Y, Fan H Q, et al. A dataset for infrared image dim-small aircraft target detection and tracking under ground/air background [DB/OL]. Science Data Bank, 2019. (2019-10-28)[2021-01-19]. <http://www.dx.doi.org/10.11922/sciencedb.902>.
- [16] Davis J W, Keck M A. A two-stage template approach to person detection in thermal imagery[C]//Application of Computer Vision, WACV/MOTION'05, IEEE Workshops on IEEE, 2005.
- [17] FREE Teledyne FLIR thermal dataset for algorithm training [EB/OL]. (2019-06-18)[2021-01-19]. <https://www.flir.cn/oem/adas/adas-dataset-form/>.
- [18] Tan M X, Pang R M, Le Q V. EfficientDet: Scalable and efficient object detection[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 10781-10790.
- [19] Zhou X, Wang D, Krähenbühl P. Objects as points [EB/OL]. (2019-04-16)[2021-01-19]. <https://doi.org/10.48550/arXiv.1904.07850>.
- [20] Wang C Y, Bochkovskiy A, Liao H. Scaled-YOLOv4: Scaling cross stage partial network [EB/OL]. (2020-11-16)[2021-03-31]. <https://arxiv.org/pdf/2011.08036.pdf>.