

融合多异构滤波器的轻型弱小目标检测网络

赵菲*, 邓英捷

国防科技大学电子科学学院 ATR 国防科技重点实验室, 湖南 长沙 410073

摘要 针对红外图像信息维度单一且弱小目标因特征不明显而难以检测的问题,将不同结构的多滤波器融入 YOLOv5n 网络,根据增强弱小目标和抑制背景干扰的不同特性分别选择三个异构滤波器作用于网络的多通道输入图像,从而丰富原始图像的信息维度,有效提升后端网络对复杂背景下弱小目标的适应能力;通过添加注意力模块、采用小锚框策略、裁剪网络深层分支等改进措施,在增强 YOLOv5n 网络弱小目标检测能力的同时,进一步减少了计算和存储资源需求。实验结果表明,所提出的算法能够有效检测红外复杂背景中的弱小目标,同时占用存储和计算资源更少,为算法部署在资源受限的嵌入式设备上提供了基础。

关键词 机器视觉; 红外弱小目标; 资源受限; 多异构滤波器; YOLOv5n

中图分类号 TP391.41

文献标志码 A

DOI: 10.3788/AOS221736

1 引言

如何在红外复杂背景下检测弱小目标是红外监视和精确制导系统中的关键问题,其性能直接决定系统任务成败。目标成像尺寸小、能量弱,缺乏形状、纹理特征,同时存在复杂背景的干扰,使得目标检测尤为困难。相比于检测前跟踪(TBD)类方法而言,跟踪前检测(DBT)类方法首先在单帧图像中提取潜在目标,然后利用帧间信息进行目标的判决,计算简单且易于实现,因此广泛应用于工程领域,其中单帧图像中的目标提取是 DBT 类方法能否有效检测目标的关键环节。长期以来,研究者们针对如何在单帧图像中进行弱小目标检测提出了诸多算法,如文献[1]基于 Top-Hat 滤波器增强并检测红外小目标;文献[2]基于增强局部对比度(LCM)提升目标的显著性,同时实现背景杂波抑制和目标信号增强;文献[3-4]基于频域滤波实现对背景的抑制和对目标的增强。总体来讲,传统方法对于红外复杂背景下的弱小目标的检测能力有限,且对于不同场景的泛化能力较弱。

近些年,基于深度网络的目标检测算法相比于传统算法能够获得更好的检测性能与泛化能力,例如 YOLO 系列、SSD、RetinaNet 和 Faster-RCNN 等^[5],这类方法对具有较大尺寸和明显特征的可见光目标具有很好的检测性能,但应用于存在复杂背景干扰的红外弱小目标检测中却很难获得满意的结果,而且网络规

模较大限制了这类方法在资源有限的嵌入式平台中的部署。有学者针对红外弱小目标检测的深度网络算法开展了研究,如文献[6-7]以 YOLO 系列网络为基础,通过适当修改网络结构、添加注意力机制和剪枝等提升了网络对小目标的检测能力并缩小了网络规模;文献[8]通过设计合适锚框并重定义 IoU (Intersection over Union)提升了网络对弱小目标的检测能力;文献[9]通过加入时域能量增强机制提升弱小目标检测性能,但该方法不仅对帧间图像的稳定性有要求,而且需要额外的存储和计算资源。总的来说,红外图像只有灰度一个维度的信息且弱小目标特征有限,对网络结构本身的改进只能在一定程度上提升性能且很难进一步缩小网络规模,因此其在资源有限的嵌入式平台的应用有限。

针对红外图像信息维度单一且弱小目标因特征不明显而难以检测的问题,本文结合传统算法和深度网络算法的优势,首先有侧重地从增强目标、抑制干扰的角度出发,选择多异构滤波器以丰富原始图像信息维度,进而有效提升后端网络对复杂背景下弱小目标的适应能力;随后根据目标弱小的特点,选择了规模较小的 YOLOv5n 网络并通过采用小锚框策略、增加注意力机制、剪除深层次网络等措施对其进行改进,以进一步提升网络对弱小目标的检测能力,同时降低网络实现的资源需求,保证算法具有较好的可部署实现性。

收稿日期: 2022-09-22; 修回日期: 2022-11-22; 录用日期: 2022-12-16; 网络首发日期: 2023-01-06

基金项目: 国家自然科学基金青年基金(61901489)

通信作者: *f_z2020@126.com

2 融合多异构滤波器的轻型弱小目标检测网络

2.1 多异构滤波器的引入

由于红外图像仅有灰度一个维度的信息,在存在复杂背景干扰的情况下仅依靠深度网络本身的特征提取和描述能力很难有效检测不具备明显形状特征的弱小目标。为了能够让网络挖掘图像中有更助于弱小目标检测的信息,势必要求网络具备更大的规模。如果

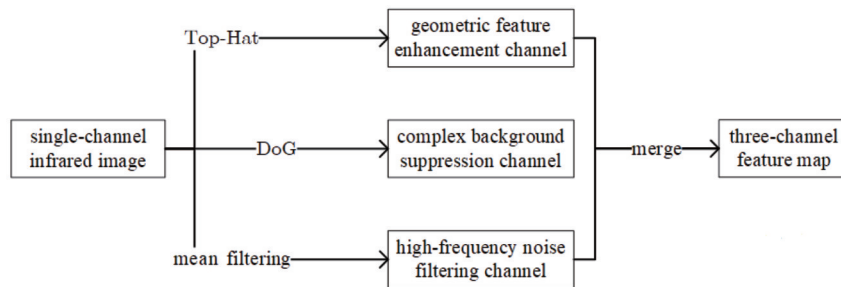


图 1 多异构滤波处理模块

Fig. 1 Multi-heterostructure filtering processing module

本文分别从侧重突显目标、抑制背景干扰以及滤除高频噪声的角度出发,选择了三种不同结构的滤波器,即 Top-Hat 形态学滤波器、高斯差分(DoG)滤波器和均值滤波器。Top-Hat 形态学滤波器是一种非线性滤波器,在模板尺寸略大于目标尺寸时,滤波结果能够保留目标尺寸并凸显弱小目标,考虑到小目标尺寸基本在 5×5 以下,算法中选择 5×5 作为 Top-Hat 的模板尺寸。DoG 滤波器的滤波参数取值服从高斯分布,较小的模板尺寸将会有助于抑制复杂背景影响但同时会削弱目标能量,随尺寸变大对目标能量的保留能力增强但对背景抑制能力减弱,考虑在两种能力中进行折中,算法中选取尺寸为 9×9 的 DoG 滤波模板。均值滤波器具有较好地滤除图像中高频噪声的能力,较大的模板尺寸有助于噪声滤除但会削弱弱小目标信息,较小的模板尺寸有助于弱小目标信息保留但噪声滤除能力相对下降,本文算法选用 3×3 尺寸的均值滤波器以滤除图像中的高频噪声。

图 2 所示为三类具有复杂背景的原红外图像分别经过三个不同特性滤波器通道滤波的结果,图中方框标出的是图像中的真实目标位置。

图 2 中第 1 行为红外原始图像,第 2 行为采用 5×5 尺寸的 Top-Hat 滤波器的滤波结果,第 3 行为采用 9×9 核的 DoG 滤波器的滤波结果,第 4 行为采用 3×3 核的均值滤波器的滤波结果。从图中可以看出,通过三个通道不同结构滤波器的处理,可以得到不同特点的三通道图像。通过采用三个异构滤波器对网络输入层的图像进行处理,原始图像的一维灰度信息被扩展为三个维度并分三个通道输入网络,从而

有侧重地通过弱小目标增强、复杂背景干扰抑制以及噪声干扰滤除等多个角度来消除干扰并凸显目标,并分别将处理后的结果分通道输入深度网络进行特征提取和描述,则不仅有助于提升网络对弱小目标的检测能力,而且有利于控制网络的规模。

基于这一思路,本文选择了三种不同结构的滤波器将红外图像处理为特征更丰富的三通道图像,以提升算法在红外复杂背景下对弱小目标检测的适应性。多异构滤波处理模块如图 1 所示。

提升了后端网络对复杂背景下弱小目标的适应能力。

2.2 改进的 YOLOv5n 网络

在考虑检测性能的同时要尽可能地控制网络规模以保证算法在资源受限的信息处理平台上的可实现性。本文选择了 YOLOv5 系列中网络规模最小的 YOLOv5n 作为弱小目标检测的基础网络,并在此基础上针对弱小目标特点对其适应性进行了改进。原始 YOLOv5n 的网络结构中主要包含了以下几个部分。

1) 输入模块。该模块将输入图像的尺寸缩放到网络指定的输入尺寸,训练时使用 Mosaic 方法实现数据增强。

2) 主干网络。该网络采用改进的 CSP-Darknet53,在保证网络检测识别精度未下降的情况下,通过梯度分流、 1×1 卷积等操作,减少计算量,提升推理速度。

3) 提取网络。该网络采用快速空间金字塔池化^[10]丰富特征提取的同时,进一步提升推理速度。

4) 检测网络。该网络包含分类和回归子模块,分类子模块采用 GIOU_Loss^[11] 作为损失函数。GIOU_Loss 的表达式为

$$\begin{cases} V_{\text{GIOU}} = (A \cap B) / (A \cup B) - [C - (A \cup B)] / C, \\ L_{\text{GIOU}} = 1 - V_{\text{GIOU}} \end{cases}, (1)$$

式中: A 为预测框; B 为真实框; C 为同时包含 A 和 B 的最小框; V_{GIOU} 为根据 GIOU 定义得到的 A 与 B 的重叠度量值; L_{GIOU} 为 GIOU 定义下 A 与 B 重叠度量的损失。回归子模块的计算如下:

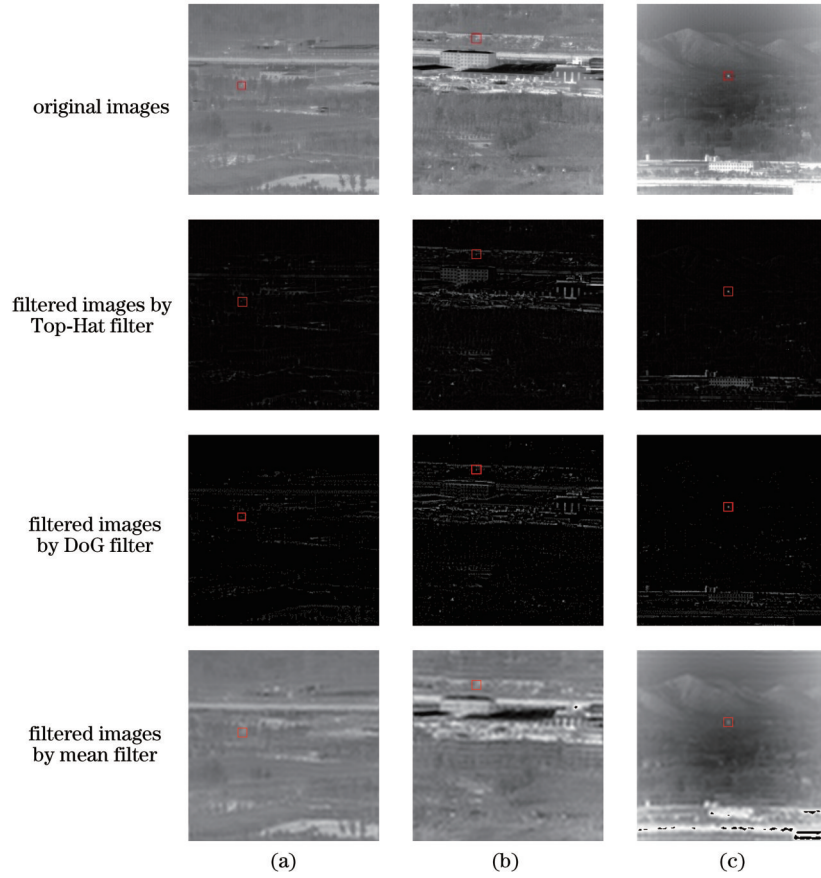


图 2 不同背景的红外原图与三通道滤波结果。(a)场景 1;(b)场景 2;(c)场景 3

Fig. 2 Original images and filtering results of three channels under different backgrounds. (a) Scenario 1; (b) scenario 2; (c) scenario 3

$$\begin{cases} b_x = 2\sigma(t_x) - 0.5 + C_x \\ b_y = 2\sigma(t_y) - 0.5 + C_y \\ b_w = P_w [2\sigma(t_h)]^2 \\ b_h = P_h [2\sigma(t_w)]^2 \end{cases}, \quad (2)$$

式中： (b_x, b_y) 为预测框中心点在全图的坐标； b_w, b_h 分别为预测框的宽和高； (C_x, C_y) 为中心点相对于所在网格原点在全图的坐标； (t_x, t_y) 为预测框中心点相对于所在网络原点的局部坐标； t_w, t_h 分别为预测框的宽和高相对于锚框的缩放比例； P_w, P_h 分别为预设锚框的宽和高； σ 为激活函数。图 3 为目标位置回归示意图，其中预测框如图 3 中矩形框①所示，锚框如图 3 中矩形框②所示，预测框中心点偏移范围如图 3 中区域③所示。

传统的 IoU 的值域为 $[0, 1]$ ，在预测框和真实框不相交时取值为 0，无法进行反向传播。GIoU (Generalized Intersection over Union) 的值域为 $(-1, 1]$ ，当预测框与真实框不相交时取值趋近于 -1 ，仍可以进行反向传播训练，并通过该值衡量两者间距离。

在上述 YOLOv5n 的基础上，本文着重采取了增加注意力机制、采用小锚框策略、剪除对弱小目标特征提取作用不明显的深层次网络，以及优化激活函数等改进措施。

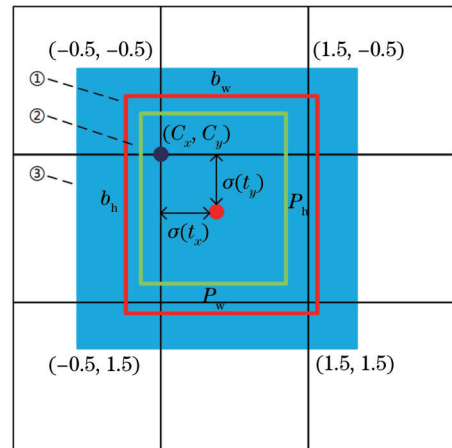


图 3 目标位置回归示意图

Fig. 3 Schematic diagram of target location regression

2.2.1 注意力机制

YOLOV5n 网络在训练过程中，每一层的参数训练都是通过固定学习率作用于当前损失函数梯度来调整自身权重，图像所有区域的特征图计算和调整方式都相同，无法关注到图像中的重点区域。引入注意力机制将使得深度网络在训练时有倾向性地增大感兴趣区域的特征权重，同时抑制无关区域的响应，从而使提取的特征图在后续的目标提取中发挥更大的作用。

基于以上考虑,本文在 YOLOv5n 的主干网络中加入轻量化的卷积注意力模块(CBAM)^[12]以提升网络能力,即在卷积层输出的结果先通过通道注意力模块(CAM)增大目标相关特征在特征图中的权重,再经

过空间注意力模块(SAM)使得该加权特征能够持续保留到后续用于检测的较深层网络中。CAM 和 SAM 的操作定义分别如下:

$$\begin{cases} M_{CAM}(f) = \sigma \{ \text{MLP}[\text{AvgPool}(f)] + \text{MLP}[\text{MaxPool}(f)] \} \\ M_{SAM}(f) = \sigma \{ \text{Conv}[\text{AvgPool}(f); \text{MaxPool}(f)] \} \end{cases}, \quad (3)$$

式中: $M_{CAM}(f)$ 为特征图 f 经过 CAM 的结果; $M_{SAM}(f)$ 为特征图 f 经过 SAM 的结果; f 为当前输入特征图; $\text{AvgPool}(f)$ 为对特征图 f 进行平均池化;

$\text{MaxPool}(f)$ 为对特征图 f 进行最大池化; $\text{MLP}(\cdot)$ 为多层感知器过程; $\text{Conv}(\cdot)$ 为卷积过程。

CBAM 的示意图如图 4 所示。

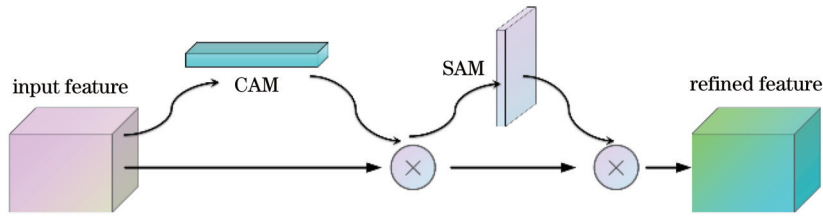


图 4 CBAM 示意图

Fig. 4 Schematic diagram of CBAM

2.2.2 小锚框策略

在标准 YOLOV5n 网络中,利用第 P17、P20 和 P23 层的特征图进行目标检测,而在提取目标的过程中,预设了多种目标尺寸并通过这些预设锚框在原始图像中搜索并框定目标。考虑到本文检测目标尺寸较小,因此本文通过缩小锚框尺寸来提高小目标检测能力。

锚框使用宽、高两个参数表征目标在原始图像上的大致尺寸和形状。浅层网络具有较大尺寸的特征图且包含丰富的表层信息,有利于检测小目标,本文将 YOLOv5n 网络中有益于小目标检测的 P3 层的锚框尺寸调整为 [5, 6, 6, 8, 9, 11], 从而使网络更好地适应小目标的特点。

2.2.3 裁剪策略

在深度网络中,浅层网络的感知视野区域较小,主要用来提取目标的局部区域特征,包括灰度、局部形状等。深层网络的感知视野区域较大,主要用来提取目标更为抽象的全局语义特征,包括目标轮廓、目标物体属性等,对目标大小、位置、方向等均不敏感。浅层网络与深层网络相结合可以提取不同目标、不同层次的特征,以提高网络对各类目标的适应性。而对于小目标检测的应用场景,深层网络所提取的特征对检测效果提升有限,甚至在检测模块中,其权重较大可能会干扰最终的检测结果。

基于以上分析,考虑到本文检测对象为红外弱小目标,经过主干网络多层特征提取后,深层网络几乎不包含小目标特征,因此对标准 YOLOv5n 深层网络进

行裁剪,去除 P5 至 P23 相关网络结构,只使用浅层的 P3 至 P17 和 P4 至 P20 的输出特征进行检测。

2.2.4 优化激活函数

计算过程中需要大量重复使用激活函数,为加速激活函数处理,采用多项式函数拟合:

$$\begin{cases} f_{\text{sigmoid}} = 1/[1 + \exp(x)] \\ f_{\text{fast_sigmoid}} = 0.5 \times [x|x|/(0.99 + x^2) + 1] \end{cases}, \quad (4)$$

式中: f_{sigmoid} 为 sigmoid 函数; $f_{\text{fast_sigmoid}}$ 为 fast_sigmoid 函数。

图 5 所示为 sigmoid 函数和 fast_sigmoid 函数的曲线对比,由此可以看出 fast_sigmoid 可以较好地代替 sigmoid 函数。sigmoid 函数带有复杂的指数运算,而 fast_sigmoid 只含有简单的四则运算,从而能够提升计

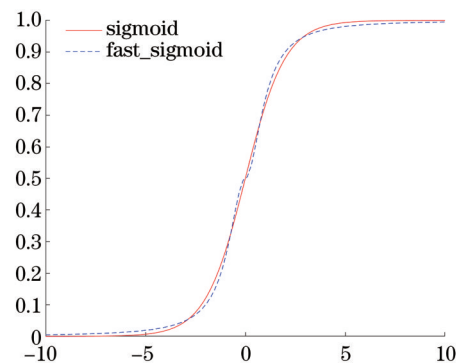


图 5 sigmoid 函数和 fast_sigmoid 函数的曲线对比

Fig. 5 Comparison of curves of sigmoid function and fast_sigmoid function

算效率,在 Python 环境下,执行 100000 次 fast_sigmoid 耗时 49.2 ms,执行 100000 次 sigmoid 耗时 62.6 ms,因此 fast_sigmoid 的计算效率提升约 21.4%。

图 6 为本文所提出的红外弱小目标检测算法的整体结构,算法流程如下:

1)对网络输入层的红外原始图像进行多异构滤波处理,扩展原始图像的信息维度,并将三通道的滤波结果拼接为三通道图像作为网络输入;

2)将上述多滤波器的结果输入到改进 YOLOv5n 网络,在主干网络中每个 CSP 模块后额外添加 CBAM,以提升网络对弱小目标特征的提取能力;在

检测模块中,舍弃深度较深的 P23 层特征图,并以小锚框循环遍历网络深度较浅的 P17 层和 P20 层特征图来进行目标检测和非极大值抑制,得出坐标位置及置信度,实现对目标的检测识别。其中 CBS 由卷积(Conv)、批归一化(BN)、激活函数(SiLu)过程构成。Resunit 为模仿 Resnet 网络中残差结构所构造的具有更深表达能力的模块。CSPX 由三个卷积层和 X 个 Resunit 模块拼接(Concat)组成。C3CBAMX 由 CSPX 模块和 CBAM 模块构成。SPPF 表示快速空间金字塔池化过程。MaxPool 表示最大池化。UpSample 表示上采样过程。

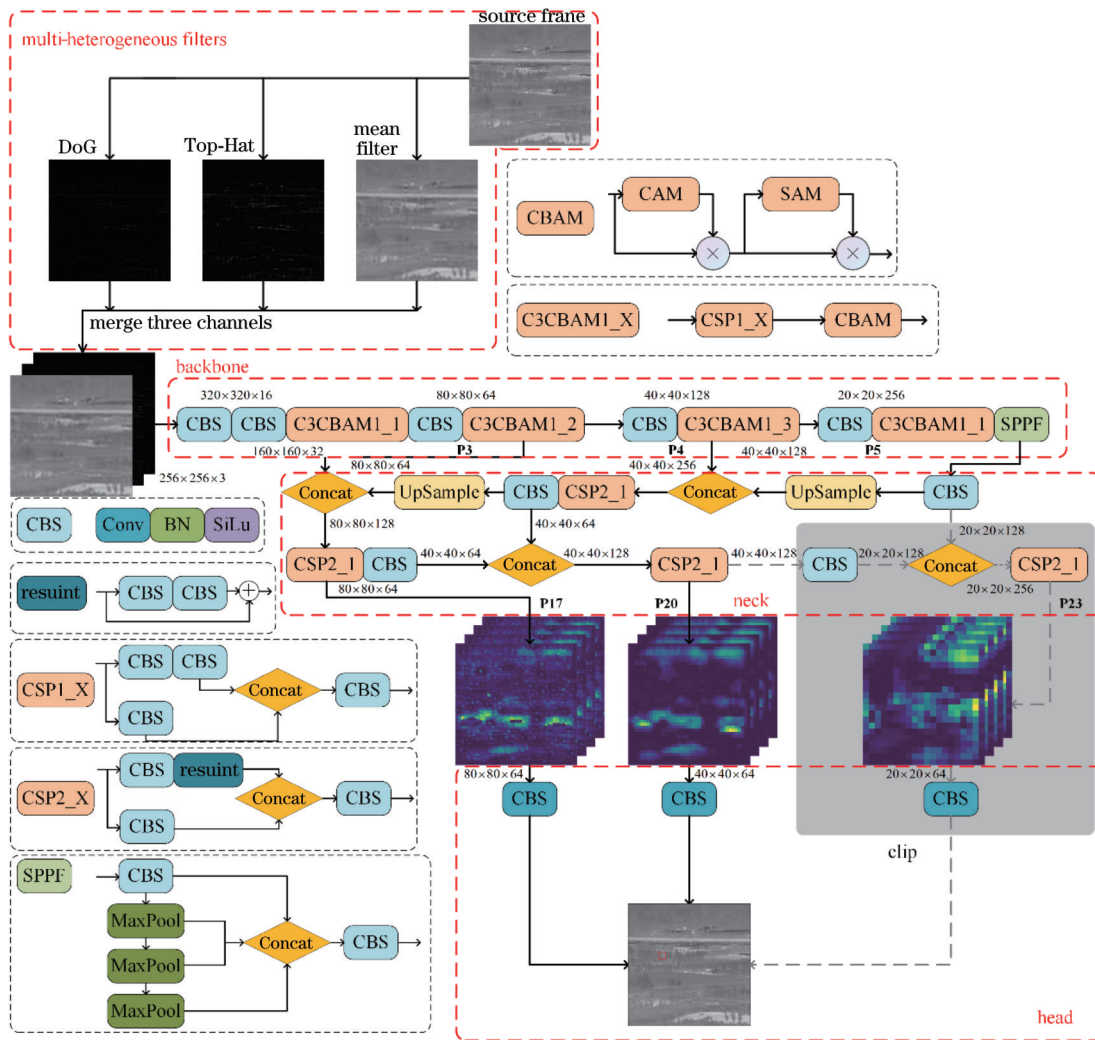


图 6 所提方法的结构框图

Fig. 6 Structure diagram of proposed method

3 实验结果

3.1 实验结果与性能比较

为验证算法的性能,本文选取地/空背景下红外图像弱小飞机目标检测跟踪数据集^[13]对算法性能进行对比实验。将本文算法与 YOLO-IDSTD^[6]、YOLOv3-CSP^[7]、TSDF^[8]和 STFBDF^[9]等采用深度网络进行弱

小目标检测的算法进行了性能比较,其中 YOLO-IDSTD、YOLOv3-CSP、STFBDF 针对的是红外弱小目标检测,实验用数据集来源于文献^[13];YOLOv3-CSP 主要针对可见光图像中的小目标检测,实验中也将其在文献^[13]的数据集中进行了训练。考虑到对比实验的广泛性,本文同时也参照文献^[9]所列实验条件,与 MEGA^[14]、FGFA^[15]、Faster-RCNN^[16]、RDN/

RDN_base^[17]、DFP^[18]、3DCNN^[19]、Efficientdet_b0^[20]等典型深度网络目标检测算法进行了性能比较。所有算法的训练与测试均在 Intel I7 6700 K+16 G DDR3+ GTX1080TI 8 G 显存+Ubuntu16.04 LTS 平台上进行对比实验。

实验选用的弱小目标数据集共有 16177 张图像, 包含 22 个目标子集和多种不同类型的复杂场景。经过统计, 数据集中目标局部信噪比小于 1.5 的图像约

占总数量的 35.6%。实验时选取 70% 的图像作为训练集, 30% 的图像作为测试集。图 7 为从数据集中抽取出的有代表性的不同背景下的原始图像, 图中使用方框标识了真实的目标。将本文算法与上述提到的算法分别在训练集上进行训练, 训练采用学习率为 0.01, 余弦退火参数为 0.01, 训练交并比阈值 iou_t 为 0.2, 计算平均精确度 (AP) 使用的交并比阈值为 0.5, 标签与锚框的尺寸比阈值 anchor_t 为 2.0。

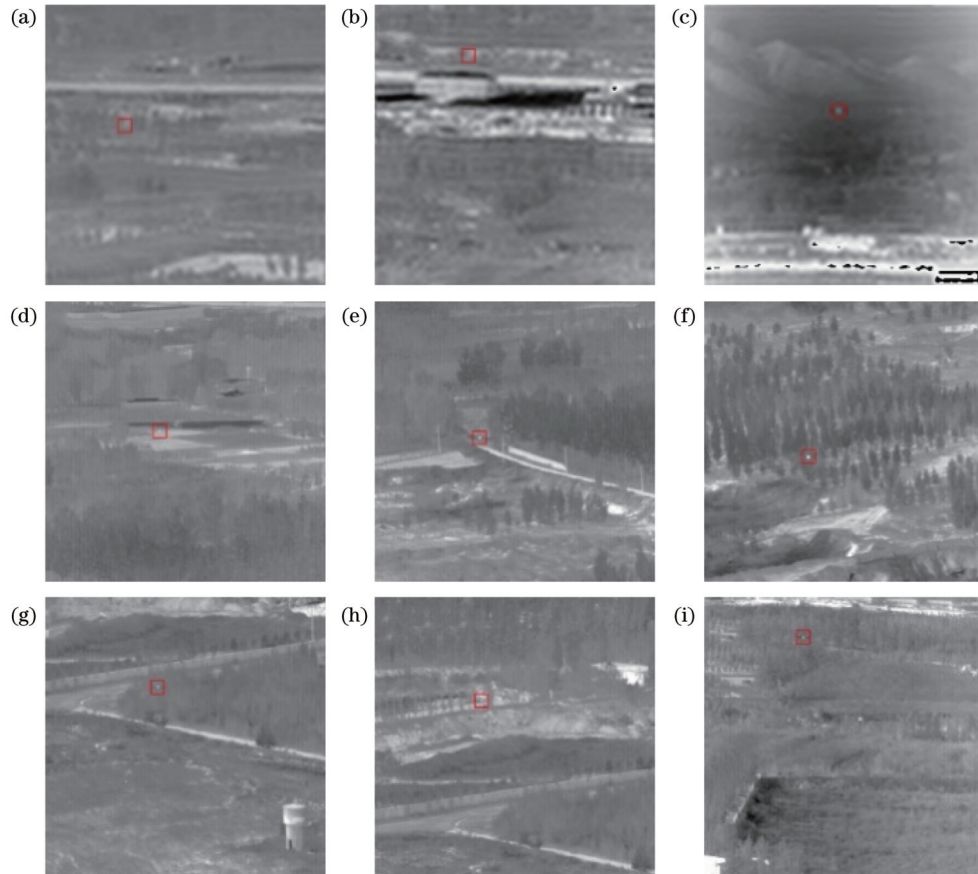


图 7 数据集原始图像。(a)场景 1;(b)场景 2;(c)场景 3;(d)场景 4;(e)场景 5;(f)场景 6;(g)场景 7;(h)场景 8;(i)场景 9

Fig. 7 Raw images in dataset. (a) Scenario 1; (b) scenario 2; (c) scenario 3; (d) scenario 4; (e) scenario 5; (f) scenario 6; (g) scenario 7; (h) scenario 8; (i) scenario 9

将本文算法与上述提到的算法分别在训练集上进行训练, 随后在测试集上进行测试并统计相应的性能指标。实验中将各算法的精准率-召回率 (PR) 曲线和 AP 作为检测性能的对比指标; 同时也将算法执行帧率以及网络的参数规模大小进行了对比, 并用来衡量算法在算力和存储资源有限的处理平台上部署的难易程度。算法 PR 曲线对比如图 8 所示, AP、计算帧率和网络参数文件大小 (单位为 MB) 对比如表 1 所示。

从表 1 的结果对比可以看出, 在检测性能方面, 本文所提出算法的性能获得了第二高的 AP 值, 性能比文献 [9] 中所提算法的略低 1.4%, 比第三名高 3% 以上。其中文献 [9] 的算法需要在时域上进行能量增强操作, 对于图像序列的帧间稳定性有很高要求且需要额外的存储资源来存储图像序列, 而本文算法的处理

流程相对简单且易实现, 具有较强的实用性。与文献 [6-8] 中所提算法相比, 本文算法有 6% 以上的性能提升, 同时通过图 8 的 PR 曲线对比也能看出, 本文算法的 PR 曲线轮廓与检测性能最好的算法曲线均位于靠外侧的区间, 代表本文算法能够在以检测精确率和召回率为衡量指标时获得较好的性能。网络的规模以及计算帧率代表着算法能够在资源受限的信息处理平台上部署的难易程度, 网络规模越小、执行速度越快, 其在资源受限的处理平台上才有可能消耗更少的存储、计算资源并获得更快的处理速度, 这是网络能否实时处理、是否易于部署应用的重要指标。从表 1 可以看出, 本文算法以最小的网络规模 (3 MB) 实现了 416 frame/s 图像的最快处理速度; 从网络规模来看, 本文算法的网络规模为效率排第二的文献 [7] 中算法

表 1 不同深度学习方法的检测效果与规模

Table 1 Detection effects and sizes of different deep learning methods

Method	AP@0.5	Speed / (frame·s ⁻¹)	Size /MB
MEGA	0.8564	10	160
Fast-RCNN	0.8331	71	30
RDN_base	0.8231	37	124
FGFA	0.8227	43	332
TSDF	0.8078	22	27
DFE	0.7842	112	16
RDN	0.7653	30	125
3DCNN	0.5210	67	91
YOLO-IDSTD	0.8242	327	7.3
STFBDF	0.9025	7	49.7
Efficientdet_b0	0.1150	233	15.8
YOLOv3-CSP	0.7690	374	8.8
Proposed algorithm	0.8880	416	3.0

的一半以下。与检测性能最好的文献[9]相比,本文算法的执行效率约为文献[9]中算法的60倍,网络参数规模约为其1/16。

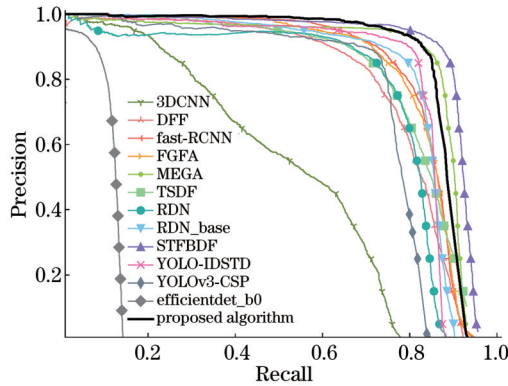


图 8 不同检测方法的 PR 曲线对比

Fig. 8 Comparison of PR curves of different detection methods

上述实验结果证明,本文算法通过将多异构滤波器分别作用于网络的多通道输入图像,扩展了原始红外图像的信息维度,有效提升了后端网络对复杂背景下弱小目标的适应能力,使得网络在参数规模最小、运算效率最高的情况下,依然可以保持很好的检测性能。

3.2 参数选择对算法性能的影响

算法部分处理流程中有特定的参数需要选择,本文在实验过程中分析了参数对算法性能的影响并对参数的选择进行了实验,具体如下所述。

1) 多异构滤波的有效性及其参数选择

算法分别从凸显目标、抑制背景干扰以及滤除高频噪声的角度出发,采用多异构滤波器对网络输入层原始图像进行了多通道的信息扩展。通过实验发现,不加入多异构滤波扩充原始图像信息而直接使用改进

YOLOv5n网络进行弱小目标检测的AP约为80%,低于本文算法约9%。同时本文对比了加入多异构滤波器前后用于检测的P3、P4层特征图的变化,可视化结果如图9所示。图9(a)为原始图像,图9(b)、9(c)分别为P3层在引入多异构滤波器前、后的特征图,图9(d)、9(e)分别为P4层在引入多异构滤波器前、后的特征图,图中用方框标识了目标所在的位置。通过对比可以看出,在网络前添加了多异构滤波器之后,目标在特征图中得到了有效增强,这进一步证明了算法采用多异构滤波进行信息扩充的有效性。

滤波器模板尺寸参数选择的影响在算法原理章节已介绍,实验中以目标信杂比增益(SCRG)为目标凸显能力的衡量来选择Top-Hat滤波器模板尺寸,以背景抑制系数(BSF)作为复杂背景抑制能力的衡量来选择DoG的模板尺寸,以及以图像信息差异(IID)作为噪声滤除效果的衡量来选择均值滤波器的模板尺寸,其中SCRG、BSF以及IID的定义分别如下:

$$\begin{cases} V_{SCR} = (\mu_{bk} - \mu_{ob}) / \sigma_{bk} \\ V_{SCRG} = 20 \lg(V_{SCR, out} / V_{SCR, in}) \\ V_{BSF} = 20 \lg(\sigma_{in} / \sigma_{out}) \\ V_{IID} = \text{abs} \left\{ \sum_i^w \sum_j^h [f_{in}(i, j) - f_{out}(i, j)] \right\} \end{cases}, \quad (5)$$

式中: μ_{bk} 为图像背景灰度均值; μ_{ob} 为图像中目标灰度均值; σ_{bk} 为图像背景灰度标准差; σ_{in} 、 σ_{out} 分别表示处理前和处理后图像的标准差; V_{SCR} 为当前图像的信杂比; $V_{SCR, in}$ 、 $V_{SCR, out}$ 分别为处理前和处理后图像的信杂比; $f_{in}(i, j)$ 、 $f_{out}(i, j)$ 分别表示处理前和处理后图像在坐标 (i, j) 处的灰度值; $\text{abs}(\cdot)$ 表示针对数据集中的数据取绝对值,统计衡量指标平均值后发现,在Top-Hat、DoG和均值滤波中分别选择 5×5 、 9×9 及 3×3 模板尺寸会获得最好的效果。

2) 小锚框尺寸与注意力机制选择

本文使用了针对小目标尺寸特性的小锚框,从而使得网络更加适合检测小目标,同时对于网络而言,浅层的网络更有利于提取目标的表层特征(如轮廓,尺寸等),因此小锚框的尺寸与目标尺寸相匹配更有利于目标表层特征的提取。

基于以上考虑,在保持其他条件不变的情况下,本文对比了不同尺寸的小锚框对于算法的影响,选择了5组锚框的尺寸(尺寸分别为 $[10, 13, 16, 30, 33, 23]$, $[7, 8, 8, 9, 10, 13]$, $[5, 6, 6, 8, 9, 11]$, $[3, 5, 5, 6, 6, 8]$, $[2, 4, 5, 3, 4, 6]$)进行了PR曲线的对比。从图10中PR曲线可以看出,选用不同的锚框尺寸确实对网络的检测性能存在一定影响。通过对比,本文最终选择了 $[5, 6, 6, 8, 9, 11]$ 尺寸的小锚框,其PR曲线明显优于原始尺寸锚框的PR曲线,且AP值增加了约3%,检测性能得到明显提升。

类似地,本文通过实验评估了注意力机制对网络

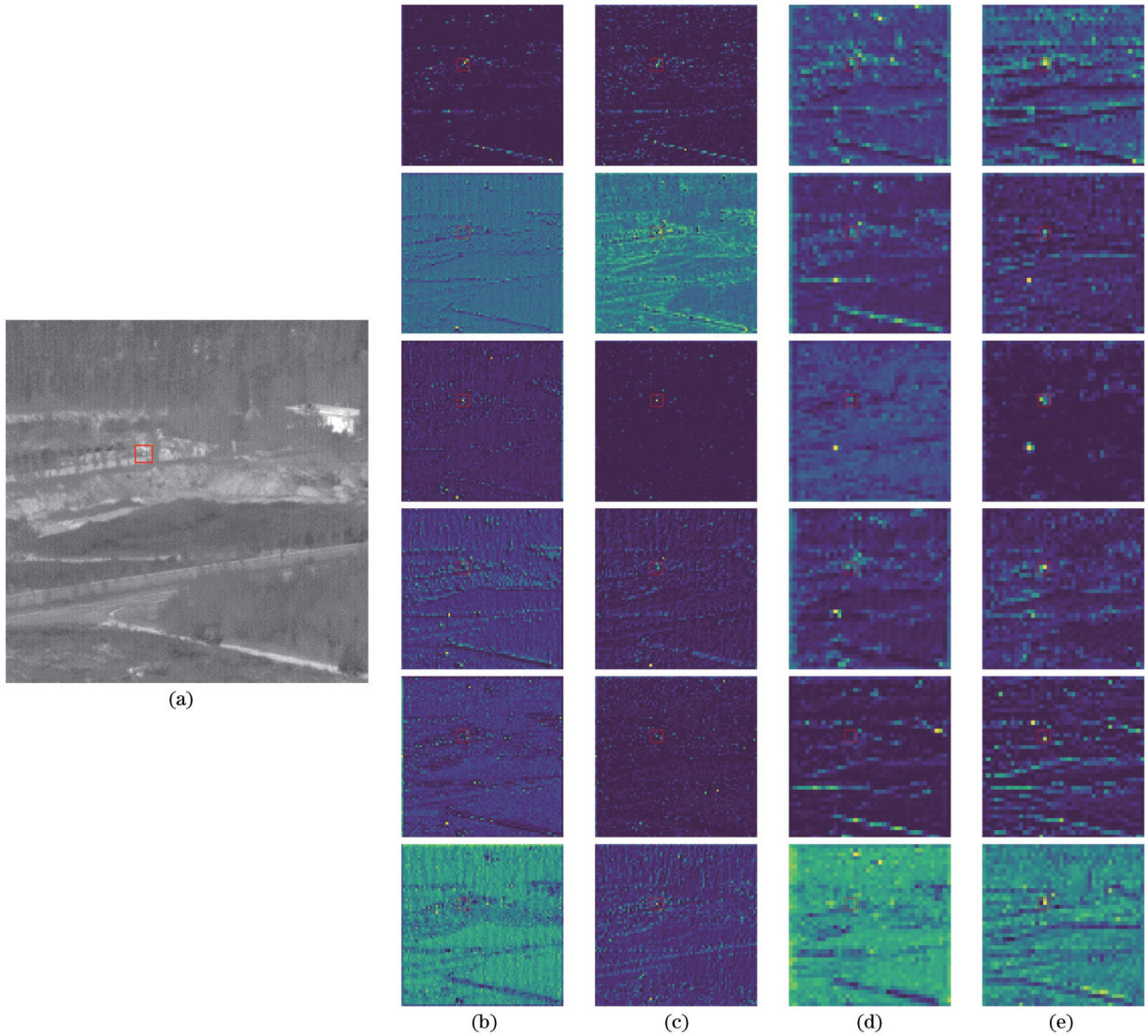


图 9 使用多异构滤波前后浅层特征图变化。(a)原始图像;(b)处理前 P3 层特征图;(c)处理后 P3 层特征图;(d)处理前 P4 层特征图;(e)处理后 P4 层特征图

Fig. 9 Changes of shallow feature maps before and after multi-heterostructure filtering. (a) Original images; (b) P3 feature maps before processing; (c) P3 feature maps after processing; (d) P4 feature maps before processing; (e) P4 feature maps after processing

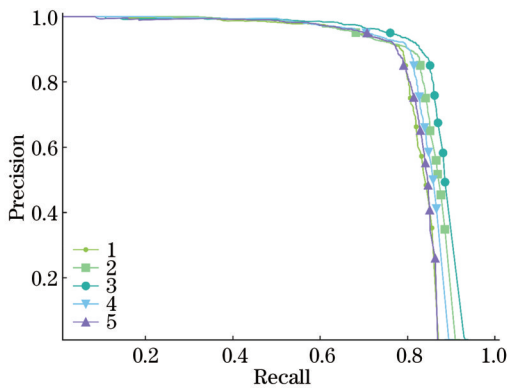


图 10 不同锚框尺寸对 PR 结果的影响

Fig. 10 Effect of different anchor sizes on PR results

特征表征能力的增强。通过对比实验发现,在主干网络中加入轻量的 CBAM 注意力机制能够增大网络相

应特征层中小目标的特征权重。

图 11(a)为原始图像,图 11(b)、11(c)分别为网络用于检测的 P3 层引入注意力机制前、后的部分通道的特征图,图 11(d)、11(e)分别为网络用于检测的 P4 层引入注意力机制前、后的部分通道的特征图。图 11 中用方框标识了目标所在的位置。从图中可以看出,在引入注意力机制后,目标在特征图中得到了一定程度的增强,同时背景噪声得到了一定的抑制。引入注意力机制前后的 PR 性能曲线如图 12 所示。

3) 网络裁剪对性能的影响

高层网络提取到的特征更加抽象,对于小目标检测并没有太多作用,因此本文在 YOLOv5n 中裁剪了 P5 层输出到 P23 层的整个分支结构,这可以在缩小网络规模的同时排除深层网络特征对检测的干扰。裁剪后网络规模减小了约 24%,裁剪前后的 PR 性能曲线

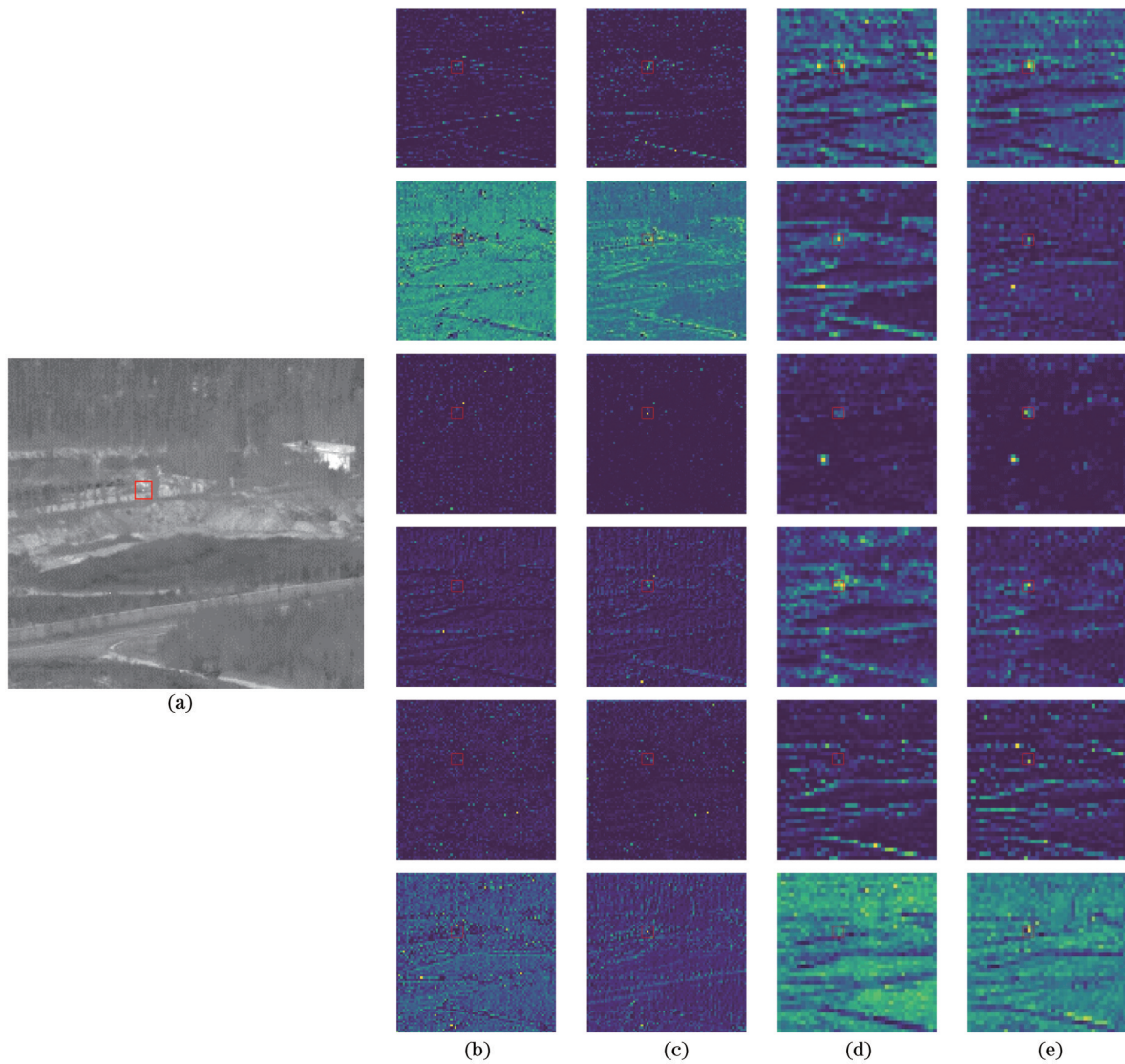


图 11 使用注意力机制前后浅层特征图变化。(a)原始图像;(b)处理前 P3 层特征图;(c)处理后 P3 层特征图;(d)处理前 P4 层特征图;(e)处理后 P4 层特征图

Fig. 11 Changes of shallow feature maps before and after using attention mechanism. (a) Original images; (b) P3 feature maps before processing; (c) P3 feature maps after processing; (d) P4 feature maps before processing; (e) P4 feature maps after processing

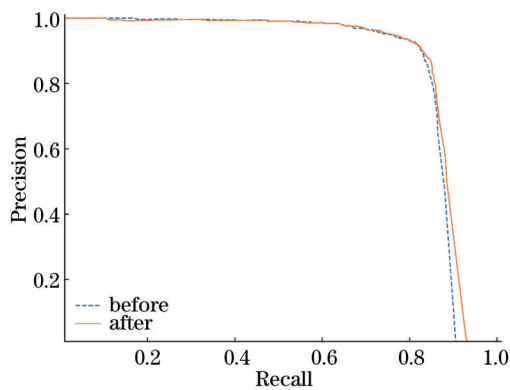


图 12 使用注意力机制前后的 PR 变化曲线

Fig. 12 Change curves of PR before and after using attention mechanism

381 frame/s 提升至 416 frame/s, 算法的实用性得到了进一步增强。

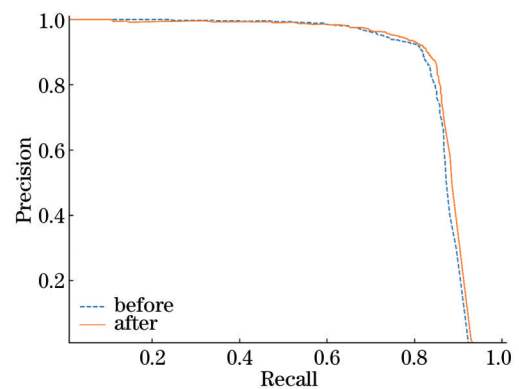


图 13 裁剪前后 PR 变化曲线

Fig. 13 Change curves of PR before and after clipping

如图 13 所示, 由图中可以看出, 裁剪网络并未降低网络的检测性能, 同时实验表明算法的执行速度由

4 结 论

为提升复杂背景下红外弱小目标的检测性能并增强其部署实用性,本文提出了一种融合多异构滤波器的轻型弱小目标检测网络算法。所提算法首先以扩展原始图像信息维度为目的,采用多个不同结构的滤波器分别作用于网络的多通道输入图像,有效提升了后端网络对复杂背景下弱小目标的适应能力;随后选择了规模较小的 YOLOv5n 网络并通过采用小锚框策略、增加注意力机制、剪除对小目标特征提取作用不明显的深层次网络等措施对其进行改进,从而在进一步提升网络对红外复杂背景下的弱小目标检测性能的同时降低网络对存储和计算资源的需求。实验对比结果表明,本文算法在有效检测红外复杂场景中红外弱小目标的同时,具备更小的网络规模和更高的执行效率,更加容易实现在资源受限的信息处理平台中的部署。

参 考 文 献

- [1] Dehghani A, Parsayan A. Small target detection and tracking based on the background elimination and Kalman filter[C]//2015 The International Symposium on Artificial Intelligence and Signal Processing (AISP), March 3-5, 2015, Mashhad, Iran. New York: IEEE Press, 2015: 328-333.
- [2] 袁明, 宋延嵩, 张梓祺, 等. 基于增强局部对比度的红外弱小目标检测方法[J]. 激光与光电子学进展, 2023, 60(4): 0410001. Yuan M, Song Y S, Zhang Z Q, et al. Infrared dim small target detection method based on enhanced local contrast[J]. Laser & Optoelectronics Progress, 2023, 60(4): 0410001.
- [3] Qi S X, Ma J, Li H, et al. Infrared small target enhancement via phase spectrum of quaternion Fourier transform[J]. Infrared Physics & Technology, 2014, 62: 50-58.
- [4] Wang H, Xin Y H. Wavelet-based contourlet transform and kurtosis map for infrared small target detection in complex background[J]. Sensors, 2020, 20(3): 755.
- [5] 张静, 农昌瑞, 杨智勇. 基于卷积神经网络的目标检测算法综述[J]. 兵器装备工程学报, 2022, 43(6): 37-47. Zhang J, Nong C R, Yang Z Y. Review of object detection algorithms based on convolutional neural network[J]. Journal of Ordnance Equipment Engineering, 2022, 43(6): 37-47.
- [6] 蒋昕昊, 蔡伟, 杨志勇, 等. 基于 YOLO-IDSTD 算法的红外弱小目标检测[J]. 红外与激光工程, 2022, 51(3): 20210106. Jiang X H, Cai W, Yang Z Y, et al. Infrared dim and small target detection based on YOLO-IDSTD algorithm[J]. Infrared and Laser Engineering, 2022, 51(3): 20210106.
- [7] 张官荣, 陈相, 赵玉, 等. 面向小目标检测的轻量化 YOLOv3 算法[J]. 激光与光电子学进展, 2022, 59(16): 1610008. Zhang G R, Chen X, Zhao Y, et al. Lightweight YOLOv3 algorithm for small object detection[J]. Laser & Optoelectronics Progress, 2022, 59(16): 1610008.
- [8] Du J M, Lu H Z, Hu M F, et al. CNN-based infrared dim small target detection algorithm using target-oriented shallow-deep features and effective small anchor[J]. IET Image Processing, 2021, 15(1): 1-15.
- [9] Du J M, Lu H Z, Zhang L P, et al. A spatial-temporal feature-based detection framework for infrared dim small target[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-12.
- [10] 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.
- [11] Rezatofighi H, Tsoi N, Gwak J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 658-666.
- [12] 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.
- [13] 回丙伟, 宋志勇, 范红旗, 等. 地/空背景下红外图像弱小飞机目标检测跟踪数据集[J]. 中国科学数据, 2020, 5(3): 291-302. Hui B W, Song Z Y, Fan H Q, et al. A dataset for infrared detection and tracking of dim-small aircraft targets under ground/air background[J]. China Scientific Data, 2020, 5(3): 291-302.
- [14] Chen Y H, Cao Y, Hu H, et al. Memory enhanced global-local aggregation for video 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: 10334-10343.
- [15] Zhu X Z, Wang Y J, Dai J F, et al. Flow-guided feature aggregation for video object detection[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 408-417.
- [16] 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.
- [17] Deng J J, Pan Y W, Yao T, et al. Relation distillation networks for video object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 7022-7031.
- [18] Zhu X Z, Xiong Y W, Dai J F, et al. Deep feature flow for video recognition[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 4141-4150.
- [19] Zhang L, Zhu G M, Shen P Y, et al. Learning spatiotemporal features using 3DCNN and convolutional LSTM for gesture recognition[C]//2017 IEEE International Conference on Computer Vision Workshops, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 3120-3128.
- [20] 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.

Light Dim Small Target Detection Network with Multi-Heterogeneous Filters

Zhao Fei*, Deng Yingjie

*National Key Laboratory of Science and Technology on ATR, College of Electronic Science and Technology,
National University of Defense Technology, Changsha 410073, Hunan, China*

Objective Dim small target detection in infrared images with complex backgrounds is a key technology for precise guidance systems and infrared surveillance systems, and the detection performance directly determines the success or failure of tasks. As a result, it has become a hot topic, and different detection methods have been presented. Compared with traditional algorithms, deep network algorithms have achieved remarkable results in many aspects in recent years, and some frameworks designed based on existing deep networks have been applied to detect the dim small target. Although these methods can improve the detection performance of small targets by modifying the network structure because the infrared images have only information of one dimension and limited features in small targets, it is difficult to obtain satisfactory results when the deep network is directly applied to detect dim small targets in the complex infrared background, and the large network scale makes it difficult to deploy the above methods on the embedded platform with constrained resources.

Methods In view of the characteristics of single information dimension in infrared images and inconspicuous features of dim small targets, this study enriches the information of original images and incorporates multiple filters with different structures into the YOLOv5n network. In this study, three filters with different structures, namely the Top Hat filter, difference of Gaussian filter (DoG), and mean filter, are selected from the perspective of highlighting targets, suppressing backgrounds, and filtering high-frequency noises. By introducing three heterogeneous filters to process the images in the input layer of the network, the one-dimensional gray information of the original image is expanded into three dimensions, and then they are fed to the network through three channels, which improves the adaptability of the network to dim small targets in complex backgrounds.

YOLOv5n network is selected in this study and improved as follows. 1) In order to make the deep network improve the feature weight of the region of interest and suppress the response of the unrelated region during training, the lightweight convolutional block attention module (CBAM) is added to the backbone of YOLOv5n so that the extracted feature map can play a greater role in the subsequent target extraction. The output in the convolution layer first passes through the channel attention module (CAM) to improve the weight of target-related features and then through the spatial attention module (SAM), which enables the weighted feature to remain in the deeper network. 2) In the standard YOLOv5n network, target detection is carried out using the feature maps of P17, P20, and P23 layers. In the process of target extraction, targets are searched and selected through preset anchor boxes of different sizes. Since the shallow network has a feature map with a large size and contains rich original information, it is conducive to small target detection. Therefore, this study adjusts the size of the anchor in P3 layer to [5, 6, 6, 8, 9, 11], which is beneficial to small target detection. 3) The perception field of view of the shallow network is small, which is conducive to extracting the local features of the target. The deep network has a large perception field of view, which is mainly used to extract the global features of the target. For the application scenario of small target detection, the features extracted by the deep network are limited and may even interfere with the final detection results. After multi-layer feature extraction of the backbone network, the deep network almost does not contain small target features, so the standard YOLOv5n network is cropped to remove P5–P23 layers, and only P3–P17 and P4–P20 output features are used for detection. By adding attention modules, adopting small anchor strategies, and cutting deep branches of the network, the dim small target detection performance of the YOLOv5n network is improved, and the consumption of computational and storage resources is reduced.

Results and Discussions In order to verify the performance of the algorithm, this study selects the dim small target detection and tracking infrared dataset against the ground/air background. Multiple deep network algorithms dedicated to small target detection are selected for comparison. Furthermore, the classical target detection network algorithms which are modified for small target detection are selected. In terms of detection performance, the proposed algorithm obtains the second-highest average precision (AP) value of 0.888, which is 1.4% lower than the highest value and 3% higher than the third-highest value. In terms of network size and computational efficiency, the proposed algorithm achieves the fastest processing speed of 416 frame/s at the smallest network size (3 MB), and the network size is half that of the algorithm in Ref. [7]. Compared with the algorithm with the best detection performance, the proposed algorithm performs approximately 60 times more efficiently, and the network size is approximately 1/16. This study analyzes the performance

gains of improvement measures, such as introducing multi-heterogeneous filters, adding attention modules and small anchor box strategies, and cropping deep networks. The experimental results show that the proposed algorithm can still maintain an excellent detection performance with the smallest parameter size and the highest operational efficiency.

Conclusions In order to improve the detection performance of dim small targets and enhance the deployment ability of algorithms, a light dim small target detection network with multi-heterogeneous filters is proposed. Experimental results show that the proposed algorithm can detect dim small targets in the complex infrared background effectively. In addition, fewer computational and storage resources are consumed, which lays a foundation for deployment on the embedded platform with constrained resources.

Key words machine vision; infrared dim small target; limited resources; multi-heterogeneous filters; YOLOv5n