

文章编号:1007-2780(2023)11-1455-13

基于注意力机制的复杂背景下红外弱小目标检测方法研究

刘颖^{1,2}, 孙海江^{1*}, 赵勇先^{1,2}

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院大学, 北京 100049)

摘要:针对复杂场景下红外图像中弱小目标像素占比少、特征细节不明显致使目标特征提取困难、检测准确率低的问题,提出了一种基于注意力机制的复杂背景下红外弱小目标检测方法。该方法以 YOLOv5 网络为基础,设计 SimAMC3 注意力机制模块,优化网络的特征提取层;设计目标检测头,通过增加特征融合层来改变其开始进行特征提取的深度,获得新的弱小目标检测层,使浅层特征层更好地保留弱小目标的空间信息;改进预测框筛选方式,提高距离相近或重叠目标的检测精度。实验选取了两个 SIRST 红外弱小目标图像数据集,对其进行标注并训练。实验结果表明,改进后的算法与原 YOLOv5 算法相比,平均精度均值(mAP)分别提升了 4.8% 和 7.1%,在不同复杂背景下均可有效检测出红外弱小目标,体现了良好的鲁棒性和适应性,可以有效应用于复杂背景中的红外弱小目标检测。

关键词:深度学习;红外弱小目标;目标检测;注意力机制

中图分类号:TP391 文献标识码:A doi:10.37188/CJLCD.2023-0030

Infrared dim-small target detection under complex background based on attention mechanism

LIU Ying^{1,2}, SUN Hai-jiang^{1*}, ZHAO Yong-xian^{1,2}

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The small target in infrared images has fewer pixels and lack of obvious feature details in complex scenes, which make it difficult to extract target features and usually has low detection accuracy. This paper proposes an infrared small target detection method based on attention mechanism under complex background. Based on YOLOv5 (You Only Look Once) network, SimAMC3 attention mechanism module is designed to optimize the feature extraction layer of the network. The target detection head is designed by adding a feature fusion layer to change the depth of feature extraction, a new weak target detection layer can be obtained, so that the shallow feature layer can better retain the spatial information of the weak target.

收稿日期:2023-01-06;修订日期:2023-02-06.

基金项目:吉林省科技发展计划(No. 20200404155YY, No. 20200401091GX)

Supported by Jilin Province Science and Technology Development Plan (No. 20200404155YY, No. 20200401091GX)

*通信联系人, E-mail: sunhaijiang@126.com

Finally, the screening method of prediction box is improved to increase the detection accuracy of objects with close distance or overlapping. In the experiment, two SIRST infrared dim-small target image datasets are selected to label and train them. The experimental results show that compared with the original YOLOv5 algorithm, the improved algorithm improves the mean average accuracy (mAP) by 4.8% and 7.1%. It can effectively detect infrared dim-small targets under different complex backgrounds, reflecting good robustness and adaptability, so it can be effectively applied to detect infrared dim-small target.

Key words: deep learning; infrared dim-small target; target detection; attention mechanism

1 引 言

目标检测是计算机视觉领域中的核心任务之一,也是目标跟踪、图像分割、场景理解等其他复杂任务的基础。与可见光成像系统相比,红外成像系统可以全天时、全天候工作,无需借助外部光源,具有隐蔽性好、不易受恶劣环境影响的优点,在军事、民用、安防等领域应用十分广泛。但由于红外成像系统靠温差成像,与可见光图像相比,红外图像往往成像模糊、易受噪点或条纹噪声的干扰,且图像对比度低。此外,红外图像上的弱小目标像素占比少,国际光学工程学会(SPIE)将红外弱小目标定义为,在一幅(256×256)像素的红外图像中,目标占比不超过(9×9)像素^[1]。其目标分辨率低、特征细节不明显,对比度通常不超过15%^[2],在复杂场景下对红外小目标进行准确检测是一个具有挑战性的难题^[3]。

红外弱小目标检测的多帧型算法通常用于序列图像,综合利用时空域信息进行弱小目标的检测^[4]。传统的多帧红外弱小目标检测算法中,往往会在相邻帧中假设背景为静态或选取某个一致的目标,以此利用目标在图像序列中的时空连续性对目标进行检测^[5]。近年来,预警需求使得单帧探测任务逐渐受到重视。当目标运动轨迹不规则时,传统的多帧型算法很难对随机的运动目标进行准确检测,需要通过提高单帧红外图像的目标检测能力,来增强序列图像的检测准确率,因此,对单帧红外图像目标检测的研究十分重要。传统的单帧红外弱小目标检测算法中,通常假设弱小目标在整幅图像中具有较高的对比度和较好的细节描述,但实际检测中往往只对显著性高的目标具有较好的效果,算法性能取决于设计者的先验知识,且特征设计中参数量有限,目标检测的通用性不强。对于对比度低且背景复

杂的红外图像,检测时易受到噪声的干扰产生虚警,通常需要引入其他方法来进一步降低虚警率,排除噪声干扰。采用传统的目标检测方法很难对单帧红外图像进行稳定、鲁棒、通用的目标检测^[6]。针对以上问题,本文以深度学习网络为基础,设计更适用于复杂背景的单帧红外弱小目标的检测网络。

深度学习为计算机视觉领域中许多具有挑战性的问题引入了有效的、非传统的解决方案^[7],它可以通过网络的训练自动学习特征信息,网络结构灵活,特征提取能力和泛化能力强。与传统目标检测算法相比,基于深度学习的目标检测任务对于不同场景的适应性强,无需手动设计特征且可移植性好,更适用于复杂场景下的目标检测任务,且卷积神经网络中的神经元共享权重参数,可以降低计算机的内存占用^[8]。目前用于目标检测的深度学习网络分为以R-CNN^[9]及其变体Fast R-CNN^[10]、Faster R-CNN^[11]为代表的两阶段目标检测算法和以SSD^[12]、YOLO^[13-15]系列算法为代表的单阶段目标检测算法。对于一般场景中目标分散且不具有重叠或遮挡现象的中、大型目标,这些经典算法的检测效果较好,但对于红外图像中小目标的平均检测精度均值不佳,检测效果远低于一般场景中的中、大型目标。针对此问题,本文以单阶段目标检测网络YOLOv5为基础,设计SimAMC3注意力机制模块,并设计更适合红外弱小目标的目标检测头,最后改进预测框筛选方式。优化后的目标检测网络保留了深度学习方法自动提取特征的优势,同时对红外弱小目标有更强的适应性,可以检测出不同复杂背景中的弱小目标,体现出了良好的鲁棒性和适应性,能有效应用于红外弱小目标的检测。

2 相关模型

2.1 YOLOv5 网络

YOLOv5网络含有4种模型,分别是YOLOv5s、

YOLOv5m、YOLOv5l、YOLOv5x^[16-17]。不同的网络模型所含的C3个数不同,通过两个不同的参数控制网络的深度和特征图的宽度,其中,YOLOv5s的深度和宽度最小,参数量少,适用于不同尺寸小目标的检测,其网络模型如图1所示。

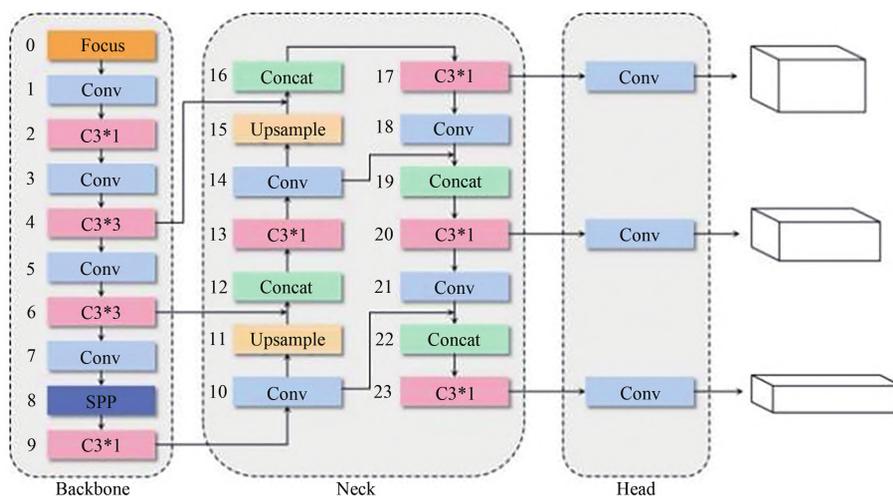


图1 YOLOv5s网络模型结构图

Fig. 1 YOLOv5 network model structure

YOLOv5s模型主要由特征提取网络Backbone、特征融合网络Neck以及检测头Head组成,主干特征提取网络用来进行目标特征的提取,包含1个Focus层、4个卷积层、4个C3层和1个SPP层。Focus结构对图片进行切片操作,可以使图片在下采样的过程中减少信息的丢失,并且可以减少参数计算量。卷积层对图像进行特征提取。C3结构是对残差特征进行学习的主要模块,包含3个标准卷积层和多个BottleNeck模块,其结构分为两支,其中一支仅经过一个标准卷积,另一只经过一层卷积和多个BottleNeck层,最终两个支路进行融合。C3结构在降低网络参数数量的同时,可以增强网络对目标特征的学习和提取能力,防止网络因层数增加而产生梯度消失或梯度爆炸的现象。空间金字塔池化SPP层,通过池化核大小分别为 1×1 、 5×5 、 9×9 和 13×13 的4个最大池化层,实现不同尺度的特征融合。特征融合网络采用FPN+PAN的特征金字塔结构。FPN是一个自上而下的网络结构,将高层特征逐渐传递下去,而PAN是一个自下而上的网络,将低层的特征信息向上传播。两者将不同尺度的特征进行融合,可以有效提高模型对不同尺寸目

标的检测能力。检测头部分用于进行分类回归预测,包含3个检测头,分别用来检测大、中、小目标。在训练阶段采用了CIOU_Loss和NMS预测框筛选方式。

除此之外,YOLOv5网络在进行特征提取前,会对图像进行一系列预处理操作,如马赛克数据增强、自适应锚框计算以及自适应图片缩放,提升检测的平均精度均值。

2.2 YOLOv5适用于红外弱小目标检测的原因

首先,对于序列红外图像,一些传统的目标检测方法有帧差法、背景减除法等。它们利用帧间的动静差异来检测移动的弱小目标,适用于背景相对静止情况下的运动目标检测。对于单帧红外图像,每张图像背景不同且不像序列图像那样具有时间关联性。一些常规的传统算法如高通滤波、形态学等方法,常常需要通过手工调整阈值来将目标与背景分离,从而实现红外弱小目标的检测。对于不同复杂程度的背景来说,难以设定同一阈值来满足所有场景的分离需求。

其次,深度学习算法从本质上来说,利用神经元学习多层次的抽象表示,隐式地利用了上下文信息。除此之外,YOLOv5算法还具有两个适合

红外弱小目标检测的设计。一是在输入端加入了 Mosaic 数据增强,丰富了检测数据集,增加了很多小目标,使网络的鲁棒性更好。并且能有效解决模型训练中“小目标不如大目标那样准确地被检测到”的问题。二是特征融合部分的 FPN+PAN 的特征金字塔结构,将低层特征层中更显著的位置信息和高层特征层中更丰富的语义信息进行融合,从而充分利用全局上下文信息更加有效地提取弱小目标的特征,提高目标被检测到的准确率,尤其对于小目标的检测性能的提升更为显著^[18]。

最后,本文优化后的网络引入了 SimAM 注意力机制,它通过能量函数找到具有“空间抑制现象”的神经元,即需要被关注的目标像素所在区域,来获取红外弱小目标的空间位置信息,增强了图像的局部上下文信息关联。

3 红外弱小目标检测网络

YOLOv5 网络模型与经典的两阶段网络相比,具有更佳检测速度,同时其检测精度在众多单

阶段检测网络中也脱颖而出,故本文以 YOLOv5 网络为基础进行改进,使其更加适用于弱小目标在不同的红外复杂场景下的检测。图 2 展示了优化后的红外弱小目标检测框架。红外图像输入目标检测网络中,首先进行马赛克数据增强、自适应锚框计算以及自适应图像缩放的预处理操作。其次经过含有 SimAMC3 注意力机制模块的主干特征提取网络,使模型可以更好地提取弱小目标的空间位置信息;然后经过特征融合网络,通过增加网络的特征融合层,改变其开始特征提取的深度,从第一个 SimAMC3 模型开始提取特征,以此得到新的弱小目标检测层,使浅层特征层更好地保留弱小目标的空间信息,从而增强弱小目标的空间关联性,避免空间位置信息过度丢失;最后,改进预测框筛选方式,将 DIOU^[19] 作为抑制准则,将预测框与真实框的重叠区域,以及两框中心之间的距离共同作为抑制冗余框的限制条件,不仅可以提高对距离较近或具有重叠遮挡现象的小目标的检测精度,还可以解决原方法存在的目标回归不准确的问题,加快模型的收敛速度。

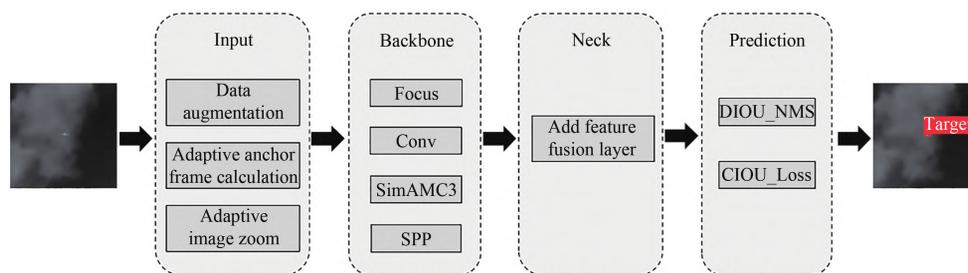


图 2 优化后的红外弱小目标检测框架图

Fig. 2 Optimized infrared dim-small target detection frame

3.1 SimAMC3 注意力机制模块的设计

注意力机制 (Attention Mechanism) 类似于人类视觉所特有的大脑信号处理机制,即视觉注意力机制。在目标检测任务中,注意力机制首先对图像进行扫描,通过生成的权重来判断每个区域的重要程度,从而使网络将注意力资源投入到目标所在的区域,获取与目标相关的信息并抑制冗余的背景和杂波干扰。由于弱小目标在图像中像素占比少,背景信息庞大,通过注意力机制可以帮助网络从复杂红外图像中快速筛选出弱小目标的特征信息,增强局部上下文信息的关联性,更准确地提取目标的空间位置。

常见的注意力机制通常分为通道注意力机制、空间注意力机制和通道与空间混合的注意力机制。通道注意力机制沿着通道维度进行特征细化操作,生成一维权重,对图像的不同通道分别进行处理,而对所有空间位置同等对待。空间注意力机制沿着空间维度进行特征细化操作,生成二维权重,对图像的不同位置分别进行处理,而对所有通道同等对待。它们只能沿着通道或空间维度细化特征,从而限制了学习注意力权重的灵活性,这些权重在通道和空间上都是不同的。

SimAM^[20] 是一个简单有效的、具有三维权重

的注意力模块。与现有的空间注意力模块和通道注意力模块不同,它在不向原始网络添加参数的情况下,通过提出的能量函数来发掘每个神经元的重要性,从而推断特征图的三维注意力权重。

能量越低,当前神经元与周围神经元的区别越大,重要性越高。SimAM 模块与其他注意力模块相比更加灵活,并且仍然保持轻量级。常见的注意力机制如图 3 所示。

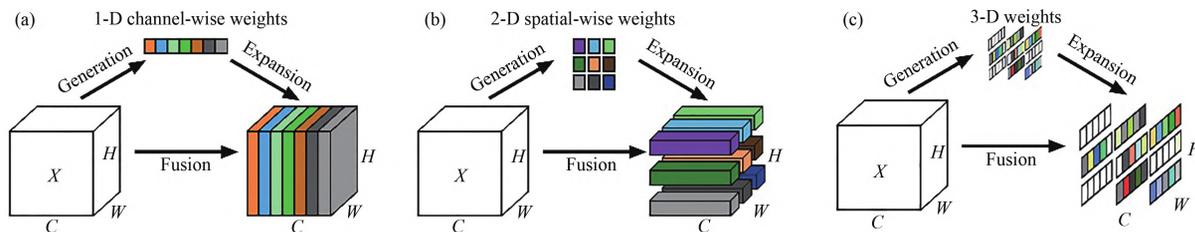


图 3 注意力机制原理图。(a)通道注意力机制;(b)空间注意力机制;(c)SimAM 注意力机制。

Fig. 3 Schematic diagram of the attention mechanism. (a) Channel attention module; (b) Spatial attention module; (c) SimAM attention module.

SimAM 注意力机制的具体实现原理如下:在视觉神经科学中,当某个神经元与周围的神经元放电模式不同时,其往往含有较大的信息量,此外,这种神经元也可能抑制其周围神经元的活动,Webb 等人称之为空间抑制现象^[21]。在视觉处理中,表现出空间抑制现象的神经元重要程度更高,即被处理的优先级应更高。为了找到这些具有优先级的神经元,SimAM 注意力机制给每个神经元定义了能量函数,用能量函数来测量这种神经元与其他神经元之间的线性可分性,如式(1)所示:

其中: $\hat{t} = \omega_t t + b_t, \hat{x}_i = \omega_i x_i + b_i$ 是 t 和 x_i 的线性变换; $M = H \times W$ 是通道上的神经元数量,每个通道都有 M 个能量函数; ω_i 和 b_i 是权重。

$$e_i(\omega_i, b_i, y, x_i) = (y_i - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_o - \hat{x}_i)^2, \quad (1)$$

引入设计的 SimAMC3 注意力机制模块后的网络整体结构图和 SimAMC3 注意力机制模块结构图分别如图 4 和图 5 所示。SimAMC3 注意力机制模块在主干特征提取部分,优化网络的特征提取层,通过能量函数找到表现出具有空间抑制现象的神经元,使网络能够更有效地提取红外弱小目标的空间位置信息,抑制复杂背景及噪声等冗余信息的干扰。SimAM 注意力本身参数数量为零,在不给原网络增加任何参数数量的情况下,有效提升了弱小目标的检测准确率。

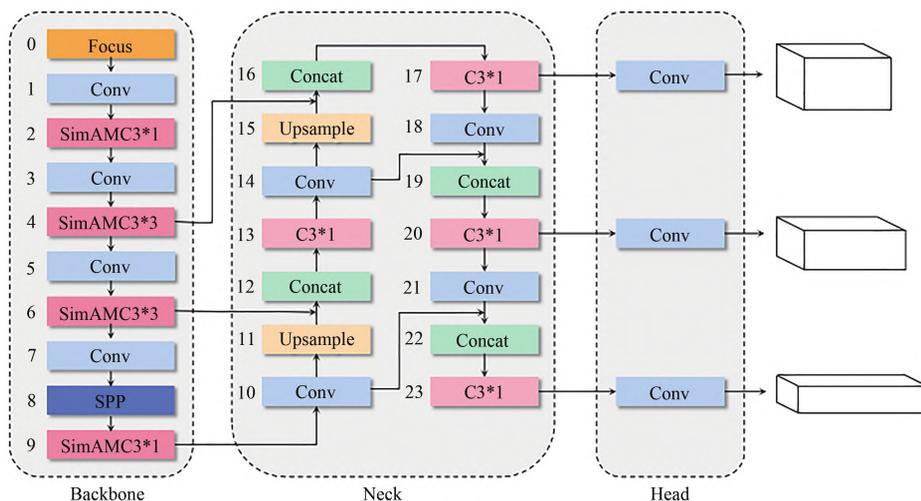


图 4 引入 SimAMC3 模块后的网络结构图

Fig. 4 Structure diagram of network after SimAMC3 module is introduced

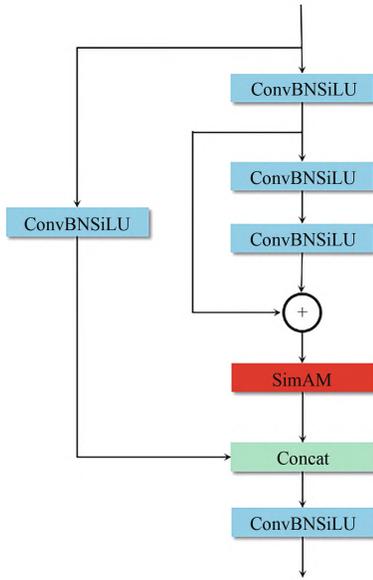


图 5 SimAMC3模块结构图
Fig. 5 SimAMC3 module

3.2 设计目标检测头

随着主干特征提取网络层数的增加,下采样倍数增加,感受野也随之增大,模型能够学习到更丰富的语义信息。此方法对于中、大型目标的检测具有较好的效果,但对于像素占比少、对比

度低的弱小目标来说,网络层数的增加会导致空间位置信息的丢失,如果不能有效利用浅层特征层提取的信息,则容易出现漏检和误检的现象,致使检测准确率下降。

为了更好地保留弱小目标的空间位置信息,本文设计了更适合红外弱小目标的目标检测头,具体实现为如下两点。首先,改变特征融合网络开始进行特征提取的深度,通过优化特征融合网络,在目标检测头部增加一个弱小目标检测层,优化后的网络结构如图 6 所示。首先,在网络的特征融合 Neck 部分增加了 7 层,分别为两个卷积层、两个 C3 层、两个特征融合层和一个上采样层,以此来改变网络开始进行特征提取的深度,将主干特征提取网络的第二层 SimAMC3 层提取到的特征进行特征融合,得到的结果送至新增加的弱小目标检测头。优化后的特征融合网络可以更好地利用全局上下文信息来提取弱小目标的特征。其次,在新增加的弱小目标检测头和原小目标检测头前分别增加了一层 SimAM 注意力机制,使网络更好地聚焦于小目标的检测。经实验验证,融入更适合红外弱小目标的检测头后,能够显著提高检测的平均精度均值。

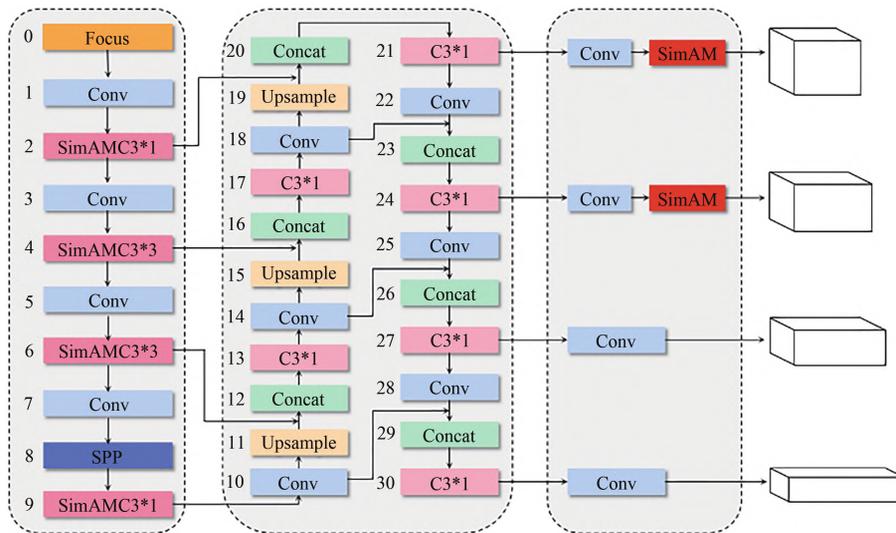


图 6 优化后的红外弱小目标检测网络模型图

Fig. 6 Optimized infrared dim-small target detection network model

3.3 改进预测框筛选方式

原 YOLOv5 网络采用普通的非极大值抑制 NMS (Non Maximum Suppression) 方法。NMS 通过筛选出局部极大值得到最优解,在实际运算

中,利用交并比 IOU (Intersection Over Union) 抑制图片中冗余的预测框,IOU 的评判标准为真实框与预测框的重叠区域,无法判断两者没有交集的情况,且不能精确地反映两者重合度大小,对

于大目标易产生误检现象,而对于小目标易产生漏检现象。

本文将非极大值抑制准则改为 DIOU (Distance Intersection Over Union), 采用 DIOU_NMS 非极大值抑制, 同时考虑尺度、重叠率以及目标与锚框之间的距离, 能够直接将两个边界框之间的距离最小化, 解决原抑制准则对目标框回归不准确、不稳定的问题, 训练过程不易发散, 且收敛速度更快, 可以提高对距离相近和被遮挡目标的检测能力, 提升模型的检测精度, 如式(2)所示:

$$\text{DIOU} = \text{IoU} - \frac{\rho^2(b, b^{\text{gt}})}{c^2} = \text{IoU} - \frac{d^2}{c^2},$$

$$-1 \leq \text{DIOU} \leq 1, \quad (2)$$

其中: $b = (x, y, w, h)$, x, y, w, h 分别为预测框的中心坐标和长宽; $b^{\text{gt}} = (x^{\text{gt}}, y^{\text{gt}}, w^{\text{gt}}, h^{\text{gt}})$, $x^{\text{gt}}, y^{\text{gt}}, w^{\text{gt}}, h^{\text{gt}}$ 分别为真实框的中心坐标和长宽; d 是预测框与真实框中心点的距离, c 是两框最小外接矩形的对角线长度, $\rho(\cdot)$ 是欧几里得距离。DIOU 原理如图 7 所示, 其中绿色框为预测框, 黑色框为真实框, 灰色虚线框为预测框与真实框的最小外接矩形。

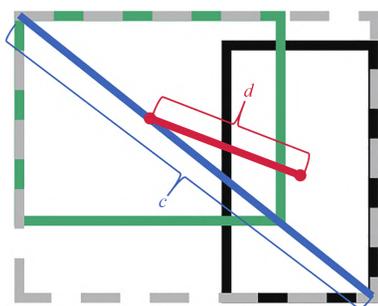


图 7 DIOU 原理图

Fig. 7 Schematic diagram of DIOU

4 实验与分析

4.1 数据集

深度学习网络结构灵活、可自动提取特征、数据处理能力强。与传统的目标检测方法相比, 具有性能高、实际应用广等优点, 更适用于不同复杂红外场景中弱小目标的检测。但深度学习网络的检测效果需要大量数据支持, 因此数据集的质量、数量、以及场景的多样化, 对网络鲁棒性和泛化能力有很大影响。

现有的红外小目标检测的开源数据集非常稀少, 大多数传统方法都是在其内部数据集上进行评估, 仅发布了少数可供深度学习网络进行训练的红外小目标数据集。Dai 等人^[22]用高质量图像和标签构建了第一个真正的单帧红外弱小目标数据集 NUAA-SIRST。该数据集共有 427 张红外图像, 包含 480 个目标, 约 55% 的目标仅占图像面积的 0.02%。在给定分辨率为 300×300 的图像中, 目标仅为 (3×3) 像素。NUAA-SIRST 数据集由数百个图像序列中选取最具代表性的图像构建而成, 目标亮度微弱、对比度低、缺乏颜色和纹理信息, 且隐藏在云层、城市、海洋等复杂的背景环境中, 受到强光源、片状云层、海面及天空边缘的干扰, 许多目标人眼甚至难以辨识, 检测难度较大。Li 等人^[23]为丰富单帧红外弱小目标数据集的目标大小、类型和复杂场景, 开发了大规模红外小目标数据集 NUDT-SIRST。该数据集有 5 个主要场景, 分别为云层、城市、海洋、旷野和高亮场景, 涵盖了各种目标。其所有场景都是通过将仿真的红外小目标与真实红外背景渲染合成的, 共包含 1 327 张红外图像, 分辨率均为 256×256 。图 8 和图 9 分别展示了以上两个数据集所包含的典型场景图像, 表 1

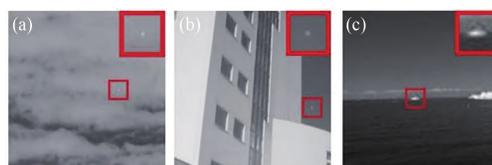


图 8 NUAA-SIRST 数据集典型场景图像示例。(a) 云层背景; (b) 城市背景; (c) 海面背景。

Fig. 8 Sample images of typical scenes in NUAA-SIRST. (a) Cloud background; (b) City background; (c) Sea background.



图 9 NUDT-SIRST 数据集典型场景图像示例。(a) 云层背景; (b) 城市背景; (c) 海面背景; (d) 旷野背景; (e) 高亮背景。

Fig. 9 Sample images of typical scenes in NUDT-SIRST. (a) Cloud background; (b) City background; (c) Sea background; (d) Field background; (e) Highlight background.

表 1 本文所选数据集的主要特征

Tab. 1 Main characteristics of the data set selected in this paper

数据集	数据类型	图像描述	图像数量
NUAA-SIRST	真实红外图像	背景: 云层、城市、海面; 目标: 亮度微弱, 有强光源干扰	427
NUDT-SIRST	合成红外图像	背景: 云层、城市、海面、旷野、高亮场景; 目标: 亮度微弱, 有强光源干扰	1 327

列出了数据集的主要特征。

4.2 评价标准

评价指标用于衡量目标检测网络的性能, 可以反应出所用算法的实用性和有效性。本文采用的衡量指标为: 平均精度均值 (Mean Average Precision, mAP)、准确率 (Precision, P) 和召回率 (Recall, R), 来验证网络模型的检测性能和整体效果。如表 2 所示, 在对目标进行分类时, 可将样例划分为 4 种情况。其中: TP 代表将正类目标正确预测为正, FP 代表将负类目标错误的预测为正, FN 代表将负类目标正确判定为负。

表 2 样例划分

Tab. 2 Sample classification

正类	负类
被检测到 True Positives (TP)	False Positives (FP)
未被检测到 False Negatives (FN)	True Negatives (TN)

平均精度均值 mAP 通过对目标分类及定位性能的综合评价反映网络模型的检测性能。AP 是某类目标的检测精度, mAP 是所有类目标的检测精度, 不同类别的 AP 取算术平均值, 即为 mAP, 其值越大, 表明当前网络模型的检测效果越好, 其公式如式 (3) 所示。mAP@0.5 表示当交并比 IOU 设为 0.5 时, 对不同类别的 AP 取算术平均值得到的平均精度均值。

$$AP = \int p(r) dr. \quad (3)$$

准确率 (Precision), 又名查准率。代表将正类目标正确预测为正类占全部预测为正类目标的比例, 其公式如式 (4) 所示:

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (4)$$

召回率 (Recall), 又名查全率。代表将正类目标正确预测为正类占全部实际为正类目标的比例, 其公式如式 (5) 所示:

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (5)$$

因此, 准确率可以衡量网络模型对样本进行正确分类的能力, 召回率可以反映网络模型能否将正样本全部检测出的能力。

4.3 训练参数及配置环境

为了更好地验证优化网络的检测性能, 首先根据 NUAA-SIRST 和 NUDT-SIRST 两个数据集给出的 mask 掩码, 采用 LabelImg 工具对其进行标注。然后将两个数据集分别按照 6:2:2 的比例进行划分, 作为实验的训练集、验证集和测试集。实际划分情况如表 3 所示。

表 3 红外弱小目标数据集划分

Tab. 3 Infrared dim-small target data set classification

Dataset	Total	Training	Val	Test
NUAA-SIRST	427	257	85	85
NUDT-SIRST	1 327	797	265	265

本实验在 Ubuntu20.04 运行, 用 Pytorch 深度学习框架搭载网络训练环境。以 5.0 版本的 YOLOv5s 网络模型为 Baseline, 网络训练权重采用 yolov5s.pt, 输入图像尺寸为 640×640, 使用 0.01 的初始学习率, 衰减权重设置为 0.0005, 批尺寸 (batch size) 为 16, 训练批次 (epoch) 设为 100。实验所用的具体配置如表 4 所示。图 10 展示了 NUDT-SIRST 数据集训练过程的损失下降曲线, 网络训练初期的损失下降较快, 在第 25 轮左右达到拐点, 之后的损失下降梯度减缓, 在第 70 轮左右趋于稳定, 最终损失收敛在 0.0002 左右。

表 4 实验配置

Tab. 4 Experimental configuration

Parameter	Configuration
OS	Ubuntu 20.04
Processor	Intel(R) Core(TM) i7-9700K
RAM	64.0 GB
GPU	NVIDIA TITAN RTX
GPU Acceleration	CUDA 11.2

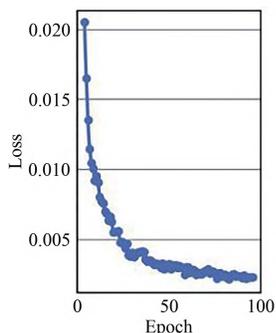


图 10 损失下降曲线图

Fig. 10 Loss decline curve

4.4 数据分析

4.4.1 消融实验

为充分验证本文设计的检测网络对不同复杂场景的适用性以及红外弱小目标检测的有效性,分别对 NUAA-SIRST 和 NUDT-SIRST 数据集进行了消融实验,其检测效果如表 5 所示。

第 1 组为原始 YOLOv5 网络模型;第 2 组为在原网络中添加了设计的 SimAMC3 模块;第 3 组在第 2 组的基础上,引入设计的目标检测头;第 4 组在第 3 组的基础上,修改预测框的筛选方式,将抑制准则改为 DIUO,为最终的优化网络模型。

由表 5 可以看出,第 2 组引入 SimAMC3 模块后,NUAA 和 NUDT 数据集的 mAP 分别提升了 1.9% 和 1.6%,说明本文设计的 SimAMC3 模块引入主干特征提取网络后,可以有效抑制复杂背景及噪声等冗余信息的干扰,对红外弱小目标的特征提取有一定的帮助。第 3 组引入设计的目标检测头后,两组数据的 mAP 分别提升了 4.7% 和 2.7%,均有显著提升。这是因为在进行特征提取时,弱小目标本身极少的特征信息会随着网络的加深而丢失,从浅层网络开始进行特征提取能够更好地保留小目标的空间位置信息,这也表明,红外弱小目标更适合用相对较浅的网络来检测。第 4 组在网络的预测部分,将预测框筛选方式的抑制准则换为 DIUO,优化后的 mAP 均提升了 0.5%。DIUO 抑制准则主要针对距离相近或存在遮挡情况的多目标,由于两个数据集中多数图像为单目标,只有少部分图像中存在多目标距离相近的情况,故改进预测框的筛选方式后,mAP 的提升较少。最终优化后的网络与原始 YOLOv5 网络相比,其 mAP 分别提升了 4.8% 和 7.1%,取得了较好的检测效果。

表 5 消融实验

Tab. 5 Ablation experiments

Group	SimAMC3	Add Head	DIUO_NMS	NUAA-SIRST			NUDT-SIRST		
				P/%	R/%	mAP@0.5	P/%	R/%	mAP@0.5
1				88.7	88.8	91.2	93.7	92.1	94.3
2	✓			91.7	90.8	93.1	95.6	93.9	95.9
3	✓	✓		95.8	97.9	97.8	97.2	96.4	98.6
4	✓	✓	✓	96.9	97.7	98.3	97.9	97.1	99.1

经过实验验证,采用 DIUO_NMS 非极大值抑制能够有效地检测出距离相近的两个小目标,部分结果如图 11 所示。其中(a)为采用以 IOU 为抑制

准则的 NMS 方法得到的检测结果,错误地将两个距离相近的目标判断成一个目标,存在误检现象;(b)为采用以 DIUO 为抑制准则的 DIUO_NMS 方法得到的检测结果,正确地检测出了两个弱小目标,提高了目标的定位精度,使检测结果更加准确。

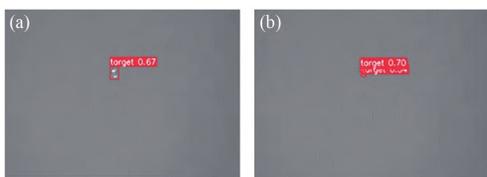


图 11 检测结果对比图。(a) IOU 抑制准则;(b) DIUO 抑制准则。

Fig. 11 Comparison of test results. (a) IOU inhibition criteria; (b) DIUO inhibition criteria.

4.4.2 对比实验

4.4.2.1 传统算法实验效果

本文在用于单帧红外弱小目标检测的传统算法中选取了 3 种经典的算法进行对比实验。实验效果如图 12 所示,其中(a)为标注出目标的原图,(b)为形态学算法中的顶帽运算,(c)为局部对比

度算法中的 MLCM (Multiscale Local Contrast Method)^[24], (d) 为以超完备稀疏表示法为基础的 IPI (Infrared Patch-Image) 算法^[25]。其中顶帽运算在具体实现中通过控制卷积核的大小来实现不同程度的背景抑制。MLCM 利用所提出的局部对比度量获得输入图像的局部对比度图, 之后采用自适应阈值来分割目标。IPI 利用背景的非局部自相关性, 将目标检测转化为低秩矩阵和稀疏矩阵的优化问题。

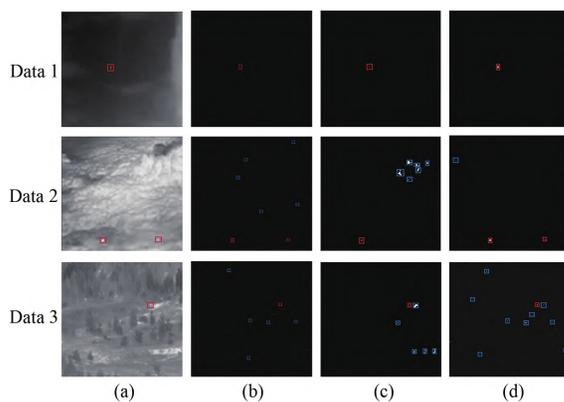


图 12 传统算法实验效果图。(a) 标注目标的原图;(b) 顶帽运算;(c) MLCM;(d) IPI。

Fig. 12 Experimental effect of traditional algorithm. (a) Original image with labelled targets; (b) Top Hat; (c) MLCM; (d) IPI.

这些算法的普遍问题在于容易受高亮点噪声的干扰, 往往需要在红外弱小目标图像检测前加入一定的预处理步骤, 或辅以其他算法来抑制与目标相似噪声的干扰。而且当调整参数使其满足某一背景的目标检测需求时, 在另一复杂背景中的检测效果却不佳, 如图 12 所示, 其中红色框标注为目标, 蓝色框标注为噪点。对于背景较为单一的 Data1 来说抑制效果较好, 但在 Data2 云

层背景和 Data3 旷野背景中, 3 种算法均受到了不同程度的高亮噪声干扰, 且部分目标本身经过算法处理之后亮度减弱。对于单帧的红外数据集来说, 普适性不强。

4.4.2.2 深度学习算法对比实验

本文方法可从训练数据中自动获取最优权重参数, 实现端到端的红外弱小目标检测。为了进一步验证所提算法的有效性, 将本文优化算法与原 YOLOv5 网络模型以及经典的深度学习目标检测网络 Faster R-CNN、SSD、YOLOv3 进行了对比实验, 实验效果如表 6 所示。

从表 6 可以看出, 与经典的目标检测网络 Faster R-CNN、SSD、YOLOv3 相比, 本文优化网络的 mAP 在 NUAA 和 NUDT 数据集上分别提升了 62.0%、32.0%、15.7% 和 55.5%、26.5%、14.1%, 充分验证了所提网络模型对复杂背景下红外弱小目标检测的有效性, 且优化网络的准确率、召回率、平均精度均值都有明显的提升, 能很好地查准、查全目标, 具有良好的鲁棒性。

最终优化的红外弱小目标检测模型在两个数据集上的平均精度均值、准确率和召回率如图 13 所示。从图中可知, NUDT 数据集在训练约 40 个 epoch 后, 目标检测的平均精度均值、准确率和召回率均趋于稳定, NUAA 数据集在约第 70 个 epoch 后趋于稳定, 网络模型能够快速收敛。

为了能够更直观地对模型的检测性能进行分析, 在 NUAA 和 NUDT 数据集上选取了部分具有代表性的数据。图 14 展示了各网络模型的可视化检测效果, 其中, Data1 为云层背景, 目标被云层遮挡, 检测时易受到块状云层的干扰; Data2 为城市背景, 目标与楼宇有相对距离, 但目标亮度相对于楼宇来说并不突出; Data3 为海面背景, 在海天交汇处含有两个距离较近的舰船小目标;

表 6 对比实验

Tab. 6 Contrast experiment

Network	NUAA-SIRST			NUDT-SIRST		
	P/%	R/%	mAP@0.5/%	P/%	R/%	mAP@0.5/%
Faster R-CNN	33.6	46.4	36.3	39.7	49.2	43.6
SSD	62.8	54.3	66.3	70.5	66.1	72.6
YOLOv3	78.9	71.2	82.6	82.9	80.6	85.0
YOLOv5	88.7	88.8	91.2	93.7	92.1	94.3
Ours	96.9	97.7	98.3	97.9	97.1	99.1

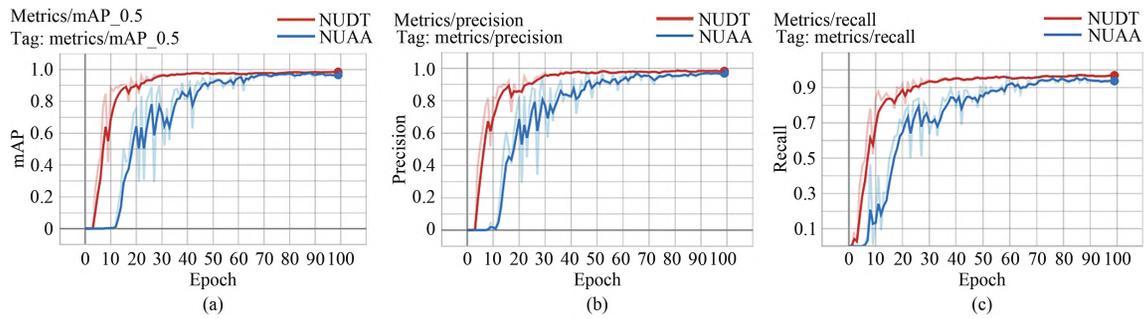


图 13 训练模型参数图。(a) 平均精度均值;(b) 准确率;(c) 召回率。

Fig. 13 Training model parameter. (a) Mean average precision; (b) Precision; (c) Recall.

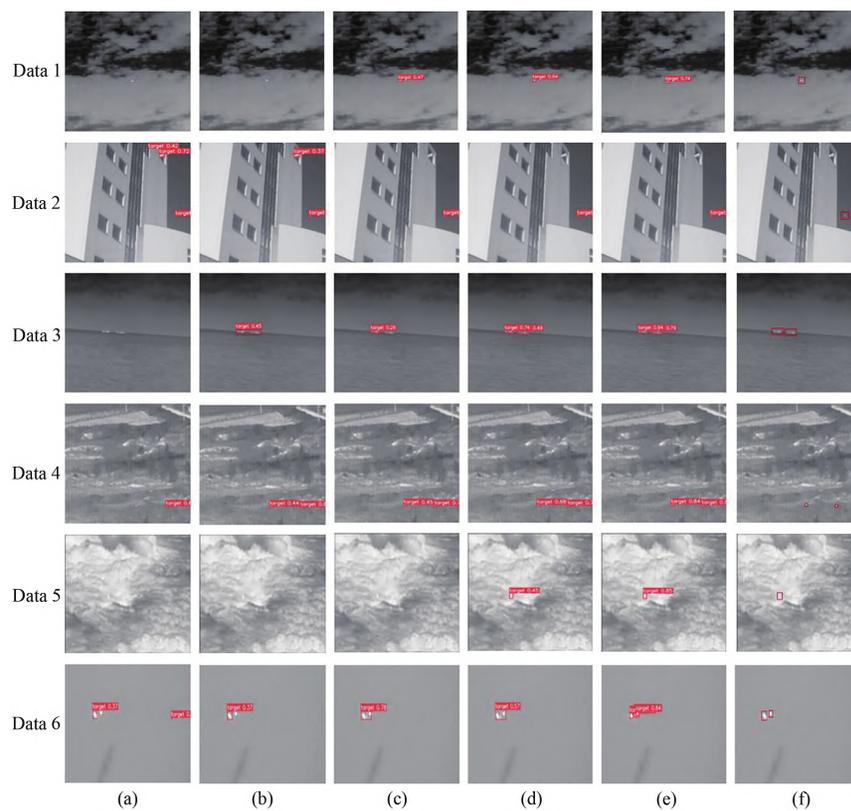


图 14 各网络模型可视化检测效果图。(a) Faster R-CNN;(b) SSD;(c) YOLOv3;(d) YOLOv5;(e) 本文优化网络;(f) 标注目标的原图。

Fig. 14 Visual test effect diagram of each network model. (a) Faster R-CNN; (b) SSD; (c) YOLOv3; (d) YOLOv5; (e) Ours; (f) Original image with labelled targets.

Data4为旷野背景,弱小目标呈点状,图像对比度低、背景复杂且有条纹噪声干扰。Data5为高亮背景,目标淹没在云层中,且与周围云层亮度相近。Data6为距离相近的多目标,目标亮度大,背景相对单一。从图 13 的检测结果可知,Faster R-CNN和 SSD 网络模型易受图像中亮度高的像素点影响而产生虚警,并且当图像背景复杂、目标像素占比极少或多目标距离较近时,存在漏检的

现象。其原因在于,Faster R-CNN 只用卷积网络的最后一层来进行预测,其经过多层卷积的特征提取之后,得到的特征图非常小,导致弱小目标的空间位置信息丢失,不利于后续的检测回归。特征提取网络的低层特征层往往能够保留相对高分辨率的弱小目标特征,而 SSD 网络没有有效利用低层特征层提取到的信息,所以在相对复杂的背景中容易产生漏检现象。YOLOv3和 YOLOv5 的

检测效果显著优于 Faster R-CNN 和 SSD 网络,其原因在于充分利用了浅层特征信息,对浅层和深层特征层进行了特征融合,并设置了 3 个不同的检测尺度,能够根据数据集的实际情况检测不同大小的目标。相比之下,YOLOv5 网络输入端加入了数据增强来丰富数据集,并且采用了更佳的损失函数来优化模型参数。但对背景复杂或距离相近的小目标来说,两者均存在漏检现象。

本文设计的目标检测网络相对于以上网络的平均精度均值有明显提升,检测效果与标签基本一致,且漏检率、误检率低。其原因在于,设计的 SimAMC3 模块使网络在进行特征提取时,可以有效地提取与弱小目标相关的特征信息并舍弃无关信息,并且设计了更适用于红外弱小目标的目标检测头,使网络能够充分利用浅层特征层提取到的相对高分辨率的弱小目标空间位置信息,避免随着卷积层数的加深而丢失信息,提高模型的检测精度。最后,修改了预测框的筛选方式,使网络能够精确检测出距离较近的弱小目标。

5 结 论

本文所设计的网络模型主要针对复杂场景

下单帧红外图像中弱小目标像素占比少、特征细节不明显致使目标特征提取困难、检测准确率低的问题,首先,设计 SimAMC3 注意力机制模块,优化网络的特征提取层,加入该模块后网络的平均精度均值在 NUAA 和 NUDT 数据集上分别提升了 1.9% 和 1.6%,提高了模型的特征提取能力和检测精度。其次,设计更适合红外弱小目标的目标检测头,通过增加网络的特征融合层,改变其开始特征提取的深度,从第一个 SimAMC3 模型开始提取特征,以此得到新的弱小目标检测层,使浅层特征层更好地保留弱小目标的空间信息,从而避免弱小目标空间位置信息过度丢失。最后,改进预测框筛选方式,以 DIOU 为抑制准则,将预测框与真实框的重叠区域,以及两框中心之间的距离共同作为抑制冗余框的限制条件,解决原方法回归不准确的问题。本文算法与 Faster R-CNN、SSD、YOLOv3、YOLOv5 相比,平均精度均值(mAP)在 NUAA 和 NUDT 数据集上分别提升了 62.0%、32.0%、15.7% 和 55.5%、26.5%、14.1%,在不同复杂背景下均可有效检测出红外弱小目标,体现出了良好的鲁棒性和适应性,可以有效应用于红外弱小目标的检测。

参 考 文 献:

- [1] ZHANG W, CONG M Y, WANG L P. Algorithms for optical weak small targets detection and tracking [C]. International Conference on Neural Networks and Signal Processing, 2003. Nanjing: IEEE, 2003: 643-647.
- [2] 杨昶,徐长彬,马玉莹,等. 低信噪比下的红外弱小目标检测算法研究综述[J]. 激光与红外,2019,49(6):643-649. YANG Y, XU C B, MA Y Y, et al. A review of infrared dim small target detection algorithms with low SNR [J]. *Laser & Infrared*, 2019, 49(6): 643-649. (in Chinese)
- [3] 赵鹏鹏,李庶中,李迅,等. 融合视觉显著性和局部熵的红外弱小目标检测[J]. 中国光学,2022,15(2):267-275. ZHAO P P, LI S Z, LI X, et al. Infrared dim small target detection based on visual saliency and local entropy [J]. *Chinese Optics*, 2022, 15(2): 267-275. (in Chinese)
- [4] 吴言枫,王延杰,孙海江,等. 复杂动背景下的“低小慢”目标检测技术[J]. 中国光学,2019,12(4):853-865. WU Y F, WANG Y J, SUN H J, et al. LSS-target detection in complex sky backgrounds [J]. *Chinese Optics*, 2019, 12(4): 853-865. (in Chinese)
- [5] DAI Y M, WU Y Q, ZHOU F, et al. Attentional local contrast networks for infrared small target detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(11): 9813-9824.
- [6] 侯旺,孙晓亮,尚洋,等. 红外弱小目标检测技术研究现状与发展趋势[J]. 红外技术,2015,37(1):1-10. HOU W, SUN X L, SHANG Y, et al. Present state and perspectives of small infrared targets detection technology [J]. *Infrared Technology*, 2015, 37(1): 1-10. (in Chinese)
- [7] ZUO C, QIAN J M, FENG S J, et al. Deep learning in optical metrology: a review [J]. *Light: Science & Applications*, 2022, 11(1): 39.
- [8] SITU G H. Deep holography [J]. *Light: Advanced Manufacturing*, 2022, 3(2): 278-300.
- [9] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic

- segmentation [C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014: 580-587
- [10] GIRSHICK R. Fast R-CNN [C]//*Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago: IEEE, 2015: 1440-1448.
- [11] 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.
- [12] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: Single shot MultiBox detector [C]//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016: 21-37.
- [13] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: unified, real-time object detection [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 779-788.
- [14] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2017: 6517-6525.
- [15] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J/OL]. *arXiv*, 2018: 1804.02767.
- [16] WU T H, WANG T W, LIU Y Q. Real-time vehicle and distance detection based on improved YOLO v5 network [C]. 2021 3rd World Symposium on Artificial Intelligence (WSAI). Guangzhou: IEEE, 2021: 24-28.
- [17] SONG X Y, GU W. Multi-objective real-time vehicle detection method based on YOLOv5 [C]. 2021 International Symposium on Artificial Intelligence and its Application on Media (ISAIAM). Xi'an: IEEE, 2021: 142-145.
- [18] 黄龙飞. 上下文信息在目标检测算法中的应用[J]. *电子世界*, 2020(2): 43-44.
HUANG L F. Application of context information in target detection algorithm [J]. *Electronics World*, 2020(2): 43-44. (in Chinese)
- [19] ZHENG Z H, WANG P, LIU W, *et al.* Distance-IoU loss: faster and better learning for bounding box regression [C]//*Proceedings of the 34th AAAI Conference on Artificial Intelligence*. New York: AAAI, 2020: 12993-13000.
- [20] YANG L X, ZHANG R Y, LI L D, *et al.* SimAM: a simple, parameter-free attention module for convolutional neural networks [C]//*Proceedings of the 38th International Conference on Machine Learning*. PMLR, 2021: 11863-11874.
- [21] WEBB B S, DHURUV N T, SOLOMON S G, *et al.* Early and late mechanisms of surround suppression in striate cortex of macaque [J]. *Journal of Neuroscience*, 2005, 25(50): 11666-11675.
- [22] DAI Y M, WU Y Q, ZHOU F, *et al.* Asymmetric contextual modulation for infrared small target detection [C]//*Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa: IEEE, 2021: 949-958.
- [23] LI B Y, XIAO C, WANG L G, *et al.* Dense nested attention network for infrared small target detection [J]. *IEEE Transactions on Image Processing*, 2023, 32: 1745-1758.
- [24] CHEN C L P, LI H, WEI Y T, *et al.* A local contrast method for small infrared target detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(1): 574-581.
- [25] GAO C Q, MENG D Y, WANG Y T, *et al.* Infrared patch-image model for small target detection in a single image [J]. *IEEE Transactions on Image Processing*, 2013, 22(12): 4996-5009.

作者简介:



刘 颖(2000—),女,湖南邵阳人,硕士研究生,2020年于海南大学获得学士学位,主要从事深度学习与红外弱小检测方面的研究。E-mail:liuying200930@163.com



孙海江(1980—),男,吉林长春人,博士,研究员,2012年于中国科学院长春光学精密机械与物理研究所获得博士学位,主要从事空间在轨图像处理与机器视觉技术方面的研究。E-mail:sunhaijiang@126.com