DOI：10.19789/j.1004-9398.2023.05.005
文献引用：潘博阳.融合多尺度注意力和分离解耦头的红外弱小目标检测算法[J].首都师范大学学报（自然科学版），2023，44（5）：30-36. PAN B Y. Infrared dim small target detection algorithm with multi-scale attention and decoupling head[J]. Journal of Capital Normal University（Natural Science Edition），2023，44（5）：30-36.

融合多尺度注意力和分离解耦头的红外弱小目标检测算法

潘　博　阳*

（武警指挥学院作战指挥系，天津　300250）

摘要：目前红外弱小目标检测技术广泛应用于预警系统、精确制导、海域监视等军事领域。针对红外弱小目标包含纹理或形状信息有限、传统算法存在较多误检和漏检等问题，以YOLOv5s模型为基础，引入动态小目标数据增强策略、改进网络结构和损失函数，提升模型对于红外弱小目标的检测效果。该方法在红外弱小目标数据集中平均检测精度达到69.2%，参数量为1.64×10^6，浮点计算量为13.5×10^9，实验结果表明，该模型能够有效检测复杂背景下的红外弱小目标，具有一定的工程实践价值。

关键词：目标检测；红外弱小目标；YOLOv5s
中图分类号：TN251；TP391.4　　**文献标识码**：A

Infrared dim small target detection algorithm with multi-scale attention and decoupling head

PAN Boyang*

（Department of Operational Command，Command College of the People's Armed Police Command College of China，Tianjin　300250）

Abstract：At present，infrared dim and small target detection technology is widely used in military fields such as early warning system，precise guidance，sea surveillance，etc. In view of the limited texture or shape information of small and dim infrared targets，and the problems of many false detections and missed detections in traditional algorithms，based on YOLOv5s（you only look once v5 small，YOLOv5s）model，dynamic small target data enhancement strategy is introduced to improve the network structure and loss function，so as to improve the detection effect of the model for small and dim infrared targets. The average precision of the method in the infrared dim small target dataset was 69.2%，and the parameter quantity was 1.64×10^6，floating point calculation amount was 13.5×10^9，the experimental results show that the model can effectively detect infrared dim small targets in complex background，and has certain engineering practice value.

Keywords：object detection；infrared dim and small target；YOLOv5s
CLC：TN251；TP391.4　　**DC**：A

收稿日期：2022-04-07

* 通信作者：panboyang@126.com

0 引 言

与可见光传感器相比,红外成像设备具有全天候探测、穿透烟雾障碍等优势。利用红外成像设备对超远距离目标进行检测,进而辅助作战人员实施远程打击,已经成为军事对抗中重要的组成部分,目前广泛应用于情报侦察、空天对抗等军事领域。近年来研究人员加大了对于该项技术的研究力度[1]。国际光学工程学会将红外弱小目标范围定义为分辨率256像素×256像素的图像中范围小于9像素×9像素的目标[2]。在复杂环境中红外弱小目标信杂比(signal to clutter ratio,SCR)较低,很容易被周围辐射强度较大的噪声所淹没,从而增加检测红外弱小目标的难度。

传统红外弱小目标检测算法通常基于静态背景假设,难以应用于真实场景[3],因此研究人员逐步转向以动态背景为基础的红外弱小目标检测。文献[4-7]采用信号处理方法在红外弱小目标与其邻域之间建立测量对比度模型,利用显著性设置阈值的方法识别潜在目标。虽然上述模型驱动的方法无需训练,计算复杂度相对较低,但是存在许多难以忽视的问题。例如:全局唯一显著性的目标假设在真实场景中并不适用;红外弱小目标较为模糊、对比度较低,许多背景同样满足上述假设导致产生大量误检;参数设置较多使得模型的泛化能力较弱;模型对于场景的高层次理解较少,难以检测极端昏暗条件下的目标。上述种种原因促使算法需要嵌入高层次的语义信息来提升检测效果。

基于卷积神经网络的YOLO(you only look once,YOLO)算法[8]已经在通用目标检测中得到广泛应用,然而在红外弱小目标检测领域还处于起步发展阶段。文献[9]在YOLOv3(you only look once v3,YOLOv3)[10]中嵌入空间注意力机制实现红外小目标检测。文献[11]在YOLOv3的基础上结合跨阶段局部模块和空间金字塔池化模块提高检测精度。然而目前检测效果最好的YOLO系列算法是YOLOv5(you only look once v5,YOLOv5)[12],其性能远优于YOLOv3。文献[13]通过添加注意力机制,改进损失函数和预测框筛选方式,将7组不同复杂背景组成的红外弱小目标数据集分成训练集、验证集和测试集进行实验,平均检测精度(average precision,AP)达到99.6%。然而在实际环境中经常会出现测试场景与训练集合的数据分布完全不同的情况。为有效评估模型的泛化能力,本文实验时选取的训练集和测试集属于不同的数据分布,更加符合真实场景中的应用。针对红外弱小目标检测算法正确率低、漏检率高等问题,本文设计的改进算法以YOLOv5s网络作为基线模型,引入动态小目标数据增强策略、添加基于多尺度注意力的上下文信息融合结构和分离解耦头结构,选取高效交并比损失函数作为辅助训练的损失函数,最后通过消融和对比实验评估该模型的有效性。

1 YOLOv5s模型

YOLOv5s模型的结构如图1所示,主要包括骨干网络单元、特征融合单元和检测头单元3个模块。骨干网络单元通过组合不同卷积层和激活函数实现权值共享和非线性变换,主要由焦点层(focus)、卷积归一化激活层(convolution + batch normalization+ silu,CBS)、具有3组卷积的跨阶段局部层(简称C3层)、空间金字塔池化层(spatial pyramid pooling,SPP)、上采样层和拼接层组成。其中:焦点层通过切片操作将特征分辨率缩小为原始数据的1/4;CBS层由卷积层、批标准化层(batch normalization,BN)和S型线性单元(sigmoid linear unit,SiLU)激活函数组成;C3层由3组卷积层和跨阶段局部层(cross stage partial,CSP)组成,跨阶段局部层[14]借助交叉计算将反向传播过程的梯度融合至特定的特征单元;空间金字塔池化层[15]将特征池化为不同的尺寸,利用卷积变换为固定维度的特征后进行拼接。为充分利用多尺度特征映射图的优势,特征融合单元通过上采样和拼接操作实现自底向上的特征传递。检测头单元将真实目标分别映射到小、中和大3种尺度的特征映射图,借助非极大值剔除重叠的预测框,将分类置信度最高的预测框作为模型输出结果。

2 改进YOLOv5s单阶段目标检测算法

2.1 改进数据增强

目前公开的红外图像数据集中数据规模较小、小目标数量稀少,然而卷积神经网络需要海量样本数据进行训练才能提高模型检测精度。为缓解上述问题,本文提出一种新的动态小目标数据增强策略,在训练的每次迭代中随机选择该批处理数据中的图像,将其包含的目标按照小目标范围动态调整,随机复制到该批次其他图像的目标周围,在不破坏原有数据分布的情况下,增加红外弱小目标的

多样性。训练过程中结合使用动态小目标数据增强策略和Mosaic数据增强策略,以提升模型对于小目标的拟合能力。

2.2　改进网络结构

2.2.1　基于多尺度注意力的上下文信息融合结构

YOLOv5s模型分别使用16和32倍下采样特征映射图检测中和大尺度目标,通过对数据集分析后发现红外弱小目标所占比例较小,大部分小于16像素×16像素,因此本文舍弃16和32倍下采样特征映射图,保留8倍下采样特征映射图,同时添加4倍下采样特征映射图作为骨干网络单元的输出。

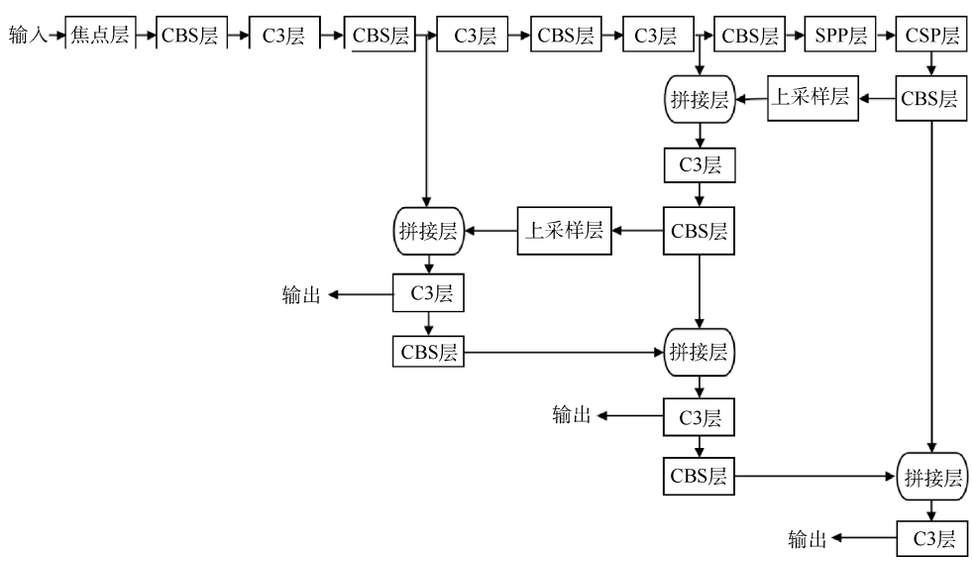
单阶段目标检测算法输出多组不同尺度特征映射图对应检测不同分辨率的目标。通常大尺度特征映射图包含语义信息较多,小尺度特征映射图边缘信息较多,为综合运用不同尺度特征映射图的互补能力,Liu等[16]在特征金字塔网络(feature pyramid networks,FPN)[17]的基础上添加自下向上的聚合路径,验证路径聚合网络(path aggregation network,PAN)对于多尺度特征融合的有效性。本文在PAN的基础上,嵌入上下文信息模块[18]和注意力机制模块[19]组成基于多尺度注意力的上下文信息融合结构,如图2所示。

为提升模型对于小目标的检测能力,增加多组卷积核大小为3×3的卷积层扩大模型的感受野,将4条通路拼接后组成多尺度上下文信息层,其结构如图3所示。

模型检测不同尺度目标时,不同特征映射图的贡献程度均不相同。采用注意力层提高模型的特征选择能力,自动调整不同通道维度特征映射图的重要程度,减少冗余信息影响,进而提高其表达能力。注意力层包括全局池化层、全连接层、Sigmoid层、卷积层和Scale层,其结构如图4所示。

2.2.2　分离解耦头结构

目前大多数单阶段目标检测算法采用共享检测头进行目标预测,不可避免会造成训练过程中



注:CBS层指卷积归一化激活层;C3层指具有3组卷积的跨阶段局部层;SPP层指空间金字塔池化层;CSP层指卷积层和跨阶段局部层。

图1　YOLOv5s网络结构

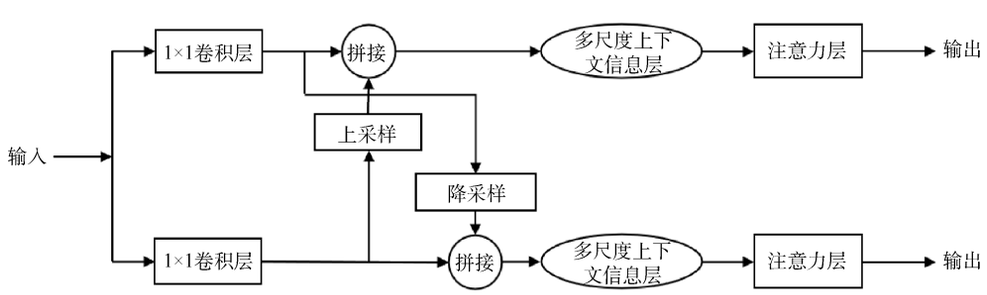


图2　基于多尺度注意力的上下文信息融合结构

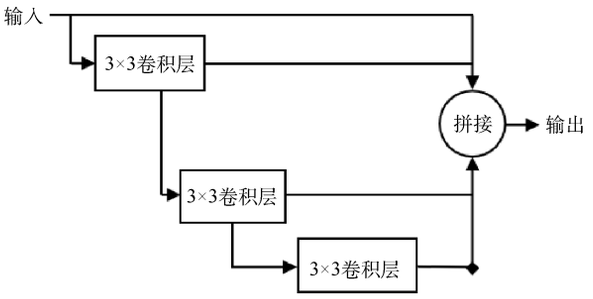


图3　多尺度上下文信息层

检测目标冲突[20],进而影响检测算法性能。为缓解上述问题,本文采用分离解耦头对目标进行分类和回归。与Song等[20]使用分离解耦头的分类输出维度$H \times W \times C$不同,本文设计分离解耦头的分类输出维度为$H \times W \times N$(Anchor)$\times N_C$,其中H和W分别表示特征映射图的高和宽,N(Anchor)表示锚框数量,N_C表示通道数,其结构如图5所示。



图4　注意力层

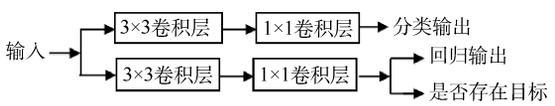


图5　分离解耦头结构

2.3　损失函数

为进一步提高模型的定位精度,使用高效交并损失函数(efficient IOU loss,EIOU)[21]作为回归目标位置的损失函数。YOLOv5s模型中完整交并比损失函数(complete IOU loss,CIOU)将目标框与锚框的宽高比作为优化项,若训练过程的宽高比成比例缩放,则CIOU不会发生变化,从而难以指导模型的训练过程。为缓解上述问题,EIOU直接将预测框和锚框的差异作为优化项

$$L_{\mathrm{EIOU}}=1-S(\mathrm{IOU})+\frac{\rho^2\left(b,b^{gt}\right)}{c^2}+\frac{\rho^2\left(w,w^{gt}\right)}{c_w^2}+\frac{\rho^2\left(h,h^{gt}\right)}{c_h^2},\tag{1}$$

式中:S(IOU)表示预测框和锚框交并补的面积,$\frac{\rho^2\left(b,b^{gt}\right)}{c^2}$表示预测框和锚框的中心点损失;$\frac{\rho^2\left(w,w^{gt}\right)}{c_w^2}$表示预测框和锚框的宽损失;$\frac{\rho^2\left(h,h^{gt}\right)}{c_h^2}$表示预测框和锚框的高损失。训练过程中EIOU和CIOU变化情况如表1所示,变化趋势如图6所示。由表1和图6可以看出EIOU有效缓解上述问题。相较于CIOU,使用EIOU的模型收敛效果更好。

表1　训练过程中不同损失函数的变化情况

迭代周期/次	CIOU	EIOU
2	0.175 60	0.178 50
4	0.116 59	0.099 49
6	0.112 10	0.094 57
8	0.106 00	0.088 92
10	0.096 05	0.078 41
12	0.076 32	0.064 58
14	0.065 78	0.055 09
16	0.059 31	0.050 58
18	0.052 08	0.044 62
20	0.050 71	0.043 05

注:CIOU为完整交并损失函数;EIOU为高效交并比损失函数。

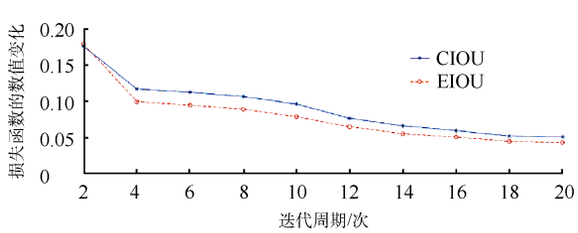


图6　训练过程中不同损失函数的变化趋势

3　实验结果

3.1　实验数据集

数据集是由国防科技大学和中国航天科工二院合作完成,科研人员在外场实际拍摄后,通过标注、滤除错误目标等人工方式构建一套以多架固定翼无人机为检测目标的红外数据集[22]。该数据集包含天空、地面、建筑物和树林等复杂场景,图像分辨率为256像素×256像素,其中训练集包含9 584张图像,测试集为2 400张图像,训练集和测试集属于不同数据分布且相互之间没有重叠,更加符合实际场景的应用需要。由于数据集的标记文件只提供无人机目标中心点位置,笔者重新对红外弱小目标进行标注,从而为评估模型效果提供数据支撑。标注后部分红外弱小目标样本如图7所示。



图7 数据集中部分红外弱小目标样本

3.2 实验环境和参数设置

在深度学习工作站上搭建实验环境,操作系统为Ubuntu18.04,CPU为i5-1135G7,内存为32 G,GPU为NVIDIA GeForce GTX 1050ti。

采用Python语言和Pytorch1.7.1深度学习框架实现算法。训练过程中使用随机梯度下降(stochastic gradient descent,SGD)优化器,学习率(learning rate,LR)设置为0.01,余弦函数衰减策略(cosine learning rate,CLR)控制学习率变化。批处理大小(batch size,BS)为16,训练周期为20,在最初的100次迭代中采用热启动(warm up)策略,权值衰减设为0.000 5,冲量为0.937。由于数据集中大部分图像中包含目标数量较少,训练过程中会出现负样本数量远大于正样本数量的情况,导致模型难以收敛。为缓解上述问题,设置候选目标框与锚框的交并比>0.3时该候选目标框为正样本,从而提高训练过程中正样本数量,进而加快模型收敛速度。

3.3 度量指标

3.3.1 精确率、召回率及其所包围的面积

精确率(precision,P)指真实样本在结果中被预测正确的比例,召回率(recall,R)指真实样本在原有数据中被预测正确的比例,其计算式为

$$P=\frac{P_{\mathrm{T}}}{P_{\mathrm{T}}+P_{\mathrm{F}}},\tag{2}$$

$$R=\frac{P_{\mathrm{T}}}{P_{\mathrm{T}}+N_{\mathrm{F}}}。\tag{3}$$

式中:P_{T}表示被正确预测为正样本的数量,P_{F}表示被错误预测为正样本的数量,N_{F}为被错误预测为负样本的数量。

P和R曲线所包围的面积平均精度计算式为

$$P_{\mathrm{A}}=\int_0^1 P(R)\mathrm{d}R。\tag{4}$$

3.3.2 浮点运算次数

浮点运算次数(floating point operations,FLOPs)指模型的计算能力,用于衡量模型的计算复杂度,目前已经成为衡量模型推理性能的参考标准。

3.4 消融实验

为验证不同模块对于算法性能的贡献程度,将其拆分为7组进行消融实验,结果如表2所示。第1组是YOLOv5s模型(5s)的测试结果。第2组(5s+aug)在第1组的基础上增加动态小目标数据增强策略,与第1组相比P_{A}提高2.1%,由于数据增强策略没有增加推理开销,因此参数量和FLOPs与第1组相同。第3组(5s+aug+4-8)使用4和8倍下采样特征映射图,与第2组相比P_{A}提高1.2%,结果表明4和8倍的下采样特征映射图能够更好捕捉小目标的特征,改善模型的特征提取能力。由于骨干网络省去了16和32倍的特征映射图,使得参数量和FLOPs减少。第4组(5s+aug+4-8+mc)添加多尺度上下文信息层,通过融合不同尺度的上下文信息,增大了模型的感受野,与第3组相比P_{A}提高0.7%。第5组(5s+aug+4-8+mc+se)添加注意力层,通过自适应调整不同通道的权重达到特征选择的目的,与第4组相比P_{A}提高0.8%,由于增加注意力层参数量和FLOPs

表2 消融实验结果

方法	参数量/10^6	GFLOPs	精确率/%	召回率/%	平均精度/%
5s	7.06	16.5	65.2	57.1	63.1
5s+aug	7.06	16.5	77.0	57.4	65.2
5s+aug+4-8	1.63	13.4	83.7	53.8	66.4
5s+aug+4-8+mc	1.59	12.6	85.4	51.9	67.1
5s+aug+4-8+mc+se	1.60	12.7	80.5	59.8	67.9
5s+aug+4-8+mc+se+decoupled	1.64	13.5	84.4	55.8	68.8
5s+aug+4-8+mc+se+decoupled+EIOU	1.64	13.5	83.4	58.6	69.2

注:GFLOPs为每秒10亿次的浮点运算次数。

略微增加。第 6 组(5s+aug+4-8+mc+se+decoupled)添加分离解耦头结构缓解目标冲突问题,与第 5 组相比 P_A 提升 1.2%。由于检测头单元分别对目标进行分类和回归,减少卷积层的复用次数,因此参数量和 FLOPs 有所增加,但是仍然小于 YOLOv5s 模型的计算复杂度。第 7 组(5s+aug+4-8+mc+se+decoupled+EIOU)在第 6 组的基础上采用 EIOU 提升模型的收敛效果,与第 6 组相比 P_A 达到 69.2%。由于替换损失函数只增加训练过程的开销,因此参数量和 FLOPs 没有变化。

3.5　对比实验

为进一步评估模型的检测效果,将本文算法与 Faster Rcnn、YOLOv3、YOLOv5s 模型使用相同数据集进行比较,其结果如表 3 所示。实验结果显示该模型在测试集上 P 为 83.4%,R 为 58.6%,P_A 达到 69.2%,优于其他 3 种模型,表明该模型针对红外弱小目标的漏检率和误检率均降低,测试集的部分检测结果如图 8 所示。

表 3　对比实验结果

模型	精确率/%	召回率/%	平均精度/%
Faster R-CNN	59.3	49.1	45.2
YOLOv3	65.6	51.9	58.3
YOLOv5s	65.2	57.1	63.1
本文算法	83.4	58.6	69.2

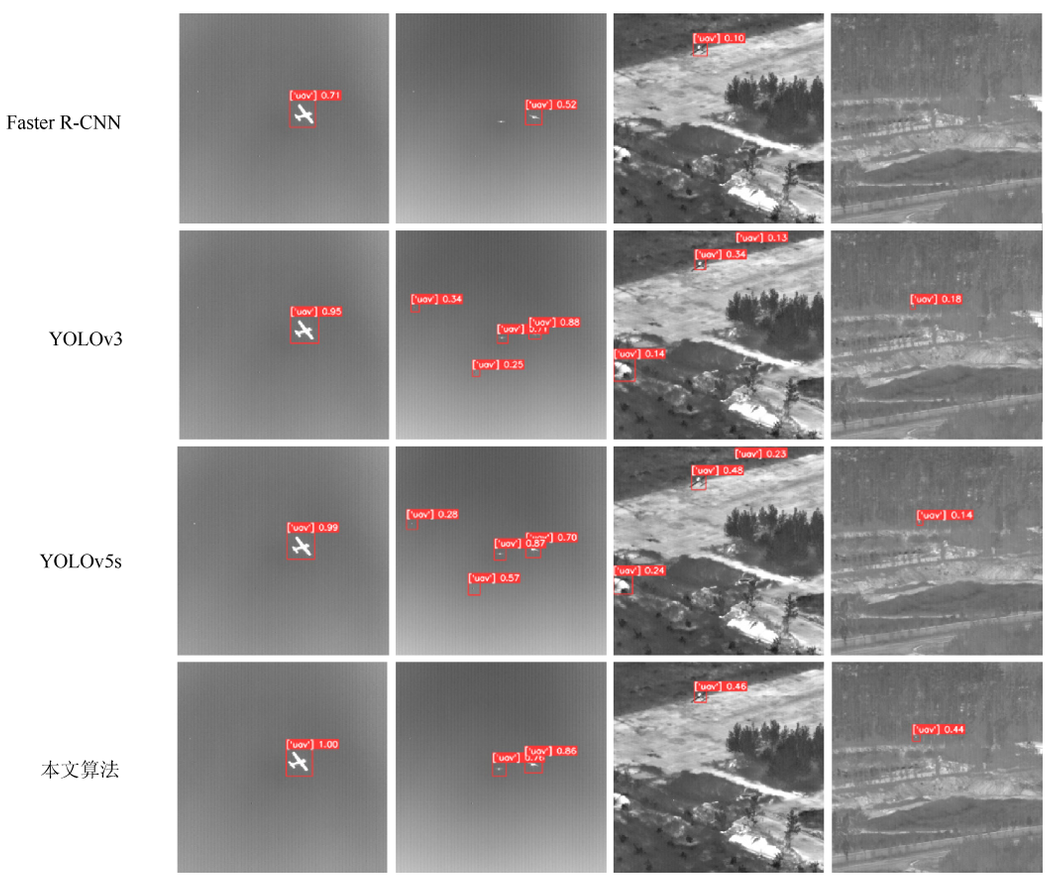


图 8　测试集的部分检测结果

4　讨　论

测试集的检测结果表明,对于第 1 列图像所包含简单背景下的目标,4 种算法均能有效检测。然而第 2、3、4 列图像包含复杂背景下的小目标,Faster R-CNN、YOLOv3、YOLOv5s 采用 16、32 倍特征映射图进行多次下采样,难以对小目标进行有效特征提取,并且训练集和测试集的数据分布不同,存在漏检、误检以及泛化能力较差等问题。由于 YOLOv3、YOLOv5s 采用领先的骨干网络和损失函数,与 Faster R-CNN 相比检测效果有所提升,但是同样存在漏检和误检的情况。本文算法均能检测复杂环境下的红外弱小目标,并且预测目标的置信度相对较高,有效验证模型的泛化能力。本文模型之所以能够提升对于红外弱小目标的检测效果,在于使用动态小目标数据增强策略,增加训练过程中小目标的多样性,提高模型的泛化能力。而后在骨干网络中添加基于多尺度注意力的上下文信息融合结构和

分离解耦头结构,扩大模型感受野的同时,通过数据驱动的方式对特征映射图赋予不同的权重,使其聚焦重要的特征映射图,进而提升模型的判别能力。最后使用EIOU优化目标框的分类置信度和回归位置,同时还提高模型的收敛速度。

5 结束语

针对红外弱小目标成像距离较远难以有效检测的问题,本文提出一种改进算法,通过在YOLOv5s模型的基础上引入动态小目标数据增强策略、改进网络结构和损失函数,从而提升对于红外弱小目标的检测能力。采用消融和对比实验评估该模型在红外弱小目标数据集上的有效性。实验结果表明训练的模型能够检测复杂背景下的红外弱小目标,同时减少漏检和误检的产生。

参 考 文 献

[1] 李俊宏,张萍,王晓玮,等.红外弱小目标检测算法综述[J].中国图象图形学报,2020,25(9):1739-1753.

[2] KIM S. High-speed incoming infrared target detection by fusion of spatial and temporal detectors[J]. Sensors, 2015,15(4):7267-7293.

[3] GAO C,MENG D,YANG Y,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.

[4] FU J,LIU J,TIAN H,et al. Dual attention network for scene segmentation[C]//DAVIS L. IEEE conference on computer vision and pattern recognition. California:IEEE,2019:3141-3149.

[5] DONG L, WANG B, MING Z, et al. Robust infrared maritime target detection based on visual attention and spatiotemporal filtering [J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(5):3037-3050.

[6] DENG H,SUN X,LIU M,et al. Small infrared target detection based on weighted local difference measure [J]. IEEE Transactions on Geoscience & Remote Sensing,2016,54(7):4204-4214.

[7] HARIHARAN B, ARBELAEZ P, GIRSHICK R, et al. Hyper columns for object segmentation and fine-grained localization [C]//BISCHOF H. IEEE conference on computer vision and pattern recognition. Boston:IEEE, 2015:447-456.

[8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once:unified,real-time object detection[C]// BAJCSY R. IEEE conference on computer vision and pattern recognition. Las Vegas:IEEE,2016:779-788.

[9] 鞠默然,罗海波,刘广琦,等.采用空间注意力机制的红外弱小目标检测网络[J].光学精密工程,2021,29(4):843-853.

[10] REDMON J, FARHADI A. Yolov3: an incremental improvement[J/OL].[2022-03-28].arXiv:1804.02767v1.

[11] 蔡伟,徐佩伟,杨志勇,等.复杂背景下红外图像弱小目标检测[J].应用光学,2021,42(4):643-650.

[12] GLENN J. Yolov5 [EB/OL]. [2022-11-18]. https://github.com/ultralytics/yolov5,2021.

[13] 代犇,赵旭,李连鹏,等.基于改进YOLOv5的复杂背景红外弱小目标检测算法[J].红外技术,2022,44(5):504-512.

[14] HE K, ZHANG X, REN S, 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.

[15] WANG C, LIAO H, WU Y, et al. Cspnet: a new backbone that can enhance learning capability of cnn [C]//BOULT T. In Proceedings of the IEEE computer vision and pattern recognition workshops. Seattle:IEEE, 2020:1571-1580.

[16] LIU S,QI L,QIN H,et al. Path aggregation network for instance segmentation[C]//BROWN M. Proceedings of the IEEE international conference on computer vision. Salt Lake City:IEEE,2018:8759-8768.

[17] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//CHELLAPPA R. Proceedings of the IEEE international conference on computer vision. Honolulu Hawaii: IEEE, 2017: 936-944.

[18] NAJIBI M, SAMANGOUEI P, CHELLAPPA R, et al. SSH: single stage headless face detector [C]// CHELLAPPA R. In Proceedings of the IEEE international conference on computer vision. Honolulu Hawaii:IEEE,2017:4885-4894.

[19] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[C]//BROWN M. Proceedings of the IEEE international conference on computer vision. Salt Lake City:IEEE,2018:7132-7141.

[20] SONG G,LIU Y,WANG X. Revisiting the Sibling Head in Object Detector[C]//BOULT T. Proceedings of the IEEE computer society conference on computer vision and pattern recognition. Virtual:IEEE,2020:11560-11569.

[21] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing,2021(506):146-157.

[22] 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(中英文网络版),2020,5(3):291-302.

(责任编辑:马田田)