基于 RCR-YOLO 的红外多尺度目标检测算法

陈笑寒，许媛媛
（上海海事大学 物流工程学院，上海 201306）

摘要：红外目标检测一直在军用和民用领域具有广泛的应用，目前针对在复杂背景下的红外多尺度目标检测中存在的漏检及误检问题，本文提出了一种改进的 YOLOv5s 算法 RCR-YOLO。首先将原 YOLOv5s 的骨干网络 CSPDarkNet53 更换为 ResNet50，避免了深层网络产生的梯度消失，增强了网络的特征提取能力，然后在骨干网络末端添加 CA 注意力机制模块，获取不同位置的特征信息，最终在颈部网络中加入 Res2Net 模块，通过引入多分支结构和逐级增加的分辨率来提高网络的表达能力并可以更好地处理多尺度特征信息，进而增强检测性能。实验结果表明，该方法优于 Faster R-CNN、SSD、YOLOv3 这些主流的目标检测算法，相较于 YOLOv5s，在保持 mAP_{50} 为 99.5%的基础上，将 $mAP_{50\text{-}95}$ 提高了 1.1%，拥有更好的检测效果，可以有效地完成复杂背景下的多尺度红外目标检测任务。

关键词：红外目标检测；YOLOv5；深度学习；多尺度
中图分类号：TN219 **文献标识码**：A **文章编号**：1001-8891(2025)04-0459-09

Infrared Multi-Scale Target Detection Algorithm Based on RCR-YOLO

CHEN Xiaohan，XU Yuanyuan
(*Department of Logistics Engineering, Shanghai Maritime University, Shanghai* 201306, *China*)

Abstract: Infrared target detection has been widely used in both military and civilian fields. To address the issues of missed and false detections in infrared multi-scale target detection under complex backgrounds, an improved YOLOv5s algorithm, RCR-YOLO, is proposed in this paper. First, the backbone network CSPDarkNet53 of the original YOLOv5s was replaced with ResNet50 to avoid gradient vanishing caused by the deep network and to enhance the network's feature extraction capability. Subsequently, the CA attention mechanism module was added to the end of the backbone to capture feature information from different locations. Finally, the Res2Net module was added to the neck network to improve the network's representational ability and process multi-scale feature information by introducing a multi-branch structure and progressively increasing resolution, thereby enhancing detection performance. Experimental results show that this method outperforms mainstream target detection algorithms such as Faster R-CNN, SSD, and YOLOv3. Compared to YOLOv5s, mAP50–95 increased by 1.1%, while mAP50 remained at 99.5%, indicating better detection performance. The algorithm effectively performs multi-scale infrared target detection under complex backgrounds.

Keywords: infrared target detection, YOLOv5, deep learning, multi-scale

0 引言

近年来，目标检测是计算机视觉领域研究的热点，其被广泛应用于许多领域，如农业中的害虫检测[1]，工业检测[2-3]，红外图像检测[4-5]，无人机捕获图像检测[6-8]等领域。目标检测算法分为传统目标检测算法和基于深度学习的目标检测算法，其中传统目标检测算法有基于手工特征设计的 HOG[9]（histogram of oriented gradient）检测器，也有基于传统滑动窗口检测方式的 DPM[10]（deformable parts model），但随着深度学习的发展，使用神经网络进行目标检测的算法越来越多，基于深度学习的目标检测算法主要分为

收稿日期：2024-02-29；**修订日期**：2024-04-01.
作者简介：陈笑寒（2000-），男，安徽合肥人，硕士研究生，研究方向：目标检测，红外图像处理。E-mail：2416724731@qq.com。
通信作者：许媛媛（1980-），女，山东莱芜人，副教授，博士，研究方向：复杂系统多尺度建模与优化、深度学习及其应用。E-mail：yyxu@shmtu.edu.cn。

两大类，一类是两阶段目标检测算法，如 R-CNN[11]、Fast R-CNN[12]、Faster R-CNN[13]、SPPNet[14]等，两阶段目标检测算法精度较高但检测速度较慢，另一类是单阶段目标检测算法，如 SSD 系列[15-18]、YOLO 系列[19-22]等，该类算法虽然检测精度一般，但检测速度快，适合完成实时检测任务。

目前基于深度学习的红外场景下的目标检测研究较少，Ding 等人在 SSD 的基础上结合了 Adaptive Pipeline Filter（APF）来提高红外小目标检测的精度[23]，Wei 等人在 YOLOv5 中加入 UNet 结构，可以更好地学习从红外和可见光域到共享特征空间的行人数据映射[24]，Jiang 等人对比了不同的 YOLO 模型在热红外（TIR）遥感多场景图像和视频中的检测效果，并选出了最优模型[6]。

少有研究是关于多尺度红外目标检测的，因此对于复杂背景下的多尺度红外目标检测，本文提出了一种改进的 YOLOv5s 检测算法 RCR-YOLO，相对于 YOLOv6s 和 YOLOv7s，YOLOv5s 具有更小的模型、更快的速度，这让其在训练、部署及实时应用方面更占优势。本文的主要贡献包括：首先将原始 YOLOv5s 的骨干网络更换为 ResNet50，提高了模型对图像中不同尺度特征的提取能力；然后在骨干网络末端添加 CA 注意力机制，增强网络的感受野和定位目标的能力；最后在颈部网络中加入 Res2Net 模块，提高了特征融合能力和网络感知能力。经过与不同的目标检测算法对比之后，表明本文的算法可以很好地完成复杂背景下的多尺度红外目标检测任务。

1 YOLOv5s 概述

YOLOv5 模型作为一个开源项目于 2020 年被提出[25]，其版本也在不断迭代。YOLOv5 是一种单阶段的目标检测模型，根据网络的深度和性能分为 YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 四个模型，其中 YOLOv5s 的参数最少，结构最小，所以检测速度最快，但是检测精度较低，YOLOv5x 参数最多，有最高的检测精度，但是检测速度最慢，所以 YOLOv5s 是在对实时性需求较高的情况下最好的选择。YOLOv5s 模型主要由 Backbone、Neck 和 Head 三大部分组成，其网络结构如图 1 所示。

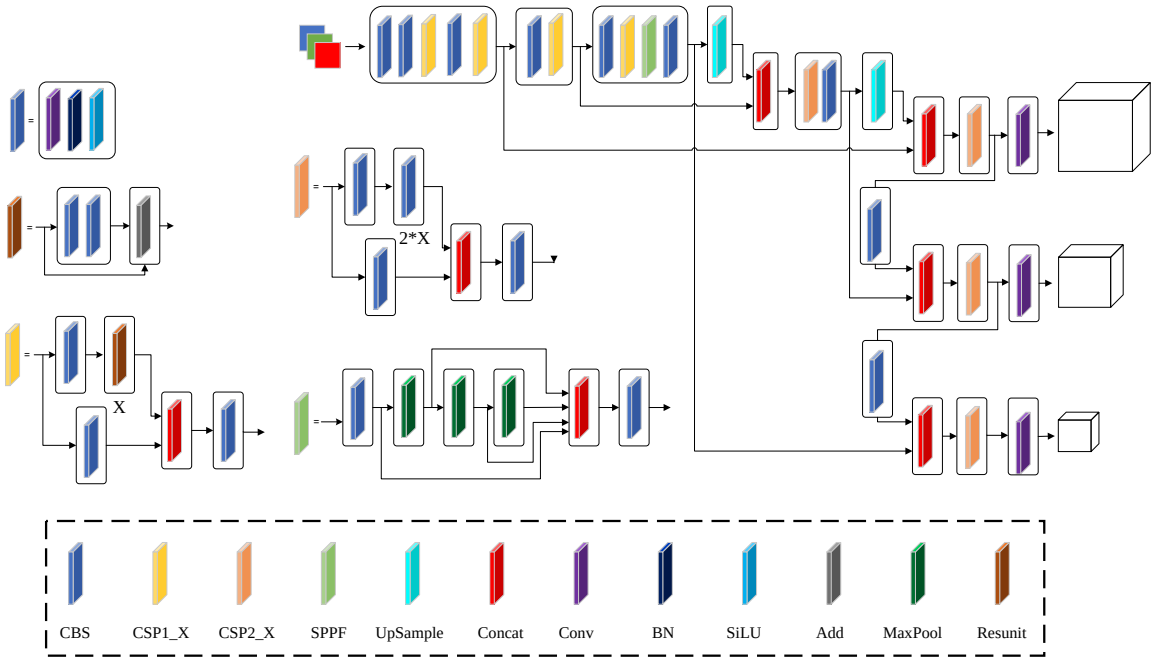

图 1 YOLOv5s 网络结构

Fig.1 YOLOv5s network structure

YOLOv5s 在输入端使用了 Mosaic 数据增强，对图片采用随机缩放、裁剪、排布等方式进行拼接，这大大丰富了检测的数据集并且减少了 GPU 的占用，同时 YOLOv5s 还使用了自适应锚框计算和自适应图片缩放，这些操作不仅提高了推理速度也使得目标检测能力有所提升。

YOLOv5s 的骨干网络主要由 CBS 模块、CSP1 模块和 SPPF 等模块构成，其主要作用是提取特征，并不断缩小特征图。

YOLOv5s 的颈部采用了 FPN（feature pyramid network）及 PAN（path aggregation network）结构，利用 Backbone 部分提取到的信息，加强特征融合。FPN

结构通过自顶向下进行上采样，使得底层特征图包含更强的图像强语义信息；PAN 结构自底向上进行下采样，使顶层特征包含图像位置信息，两个特征最后进行融合，使不同尺寸的特征图都包含图像语义信息和图像特征信息，保证了对不同尺寸的图片的准确预测。

YOLOv5s 的头部是用来对提取到的特征图进行多尺度目标检测的部分，通过多个分支分别预测不同尺度的目标框。

2 算法改进

本文在原 YOLOv5s 模型的基础上进行了一定的改进，首先将原 YOLOv5 中的骨干网络 CSPDarkNet53 更换为 ResNet50，解决了深层模型梯度消失的问题，提高了网络对图像中特征的提取能力；其次在骨干网络的末端加入了坐标注意力（coordinate attention，CA）[27]注意力机制模块，以此来捕获不同位置的特征信息，使得特征图的重点区域得到更多的关注；最后在颈部的 C3 中添加了 Res2Net 模块，通过引入多分支结构和逐级增加的分辨率来提高网络的表达能力，使得网络可以更好地处理多尺度和多分辨率的特征信息，从而增强最终的检测性能。

2.1 Backbone 网络改进

ResNet[26]是一种用于图像分类识别的卷积神经网络，其将残差网络应用于所提取红外图像的分类识别中，能够有效地解决网络深度越来越大时训练困难的问题并能够取得良好的识别结果。原 YOLOv5 的骨干网络采用 CSPDarkNet53 作为特征提取网络，在实际测试中对红外图像中的目标特征信息的提取能力还有待提升，而基于残差连接的 ResNet 能较好地解决深层模型中梯度消失问题，同时加入的部分正则项可以增加模型在训练过程中的收敛速度，所以针对红外图像中存在的细节特征不明显、分辨率较低等问题，本文选择 ResNet50 来替换 CSPDarkNet53 作为模型的特征提取网络。

ResNet50 是 ResNet 系列中重要的一员，其包含两个基本残差块，分别为 Conv Block 和 Identity Block，结构如图 2 所示，其中 Conv Block 的输入和输出的维度是不同的，所以不能连续串联，它的作用是改变网络的维度，Identity Block 输入和输出维度相同，可以串联，作用是加深网络，ResNet50 网络结构如图 3 所示。

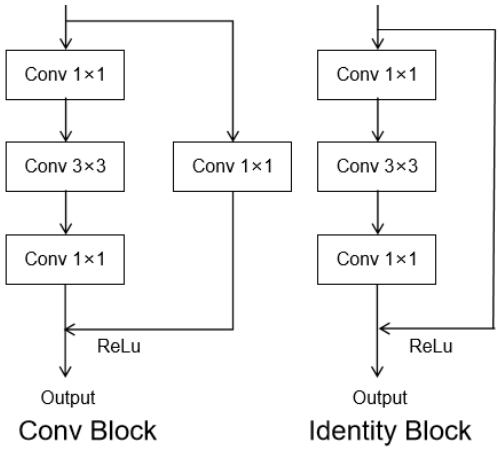

图 2 Conv Block 和 Identity Block 结构

Fig.2 Conv Block and Identity Block structure

2.2 CA 模块

Hou 等在 2021 年提出了坐标注意力（coordinate attention, CA）模块，这种注意力机制在通道注意力内嵌入位置信息，不仅能捕获跨通道信息，还能捕获方向感知与位置信息，CA 模块对特征图中通道信息沿横向和纵向进行编码，并在两个空间方向上对特征进行聚合，这样不仅可以获取空间方向上的长期依赖关系而且可以在扩大网络全局感受野的同时保留准确的位置信息。如图 4 为本文标签位置分布图，可以看出样本检测框的尺度存在不均衡的情况，因此添加 CA 模块来增强网络的全局感受野和准确定位目标的能力，从而应对本文图像标注尺度不均衡的问题。

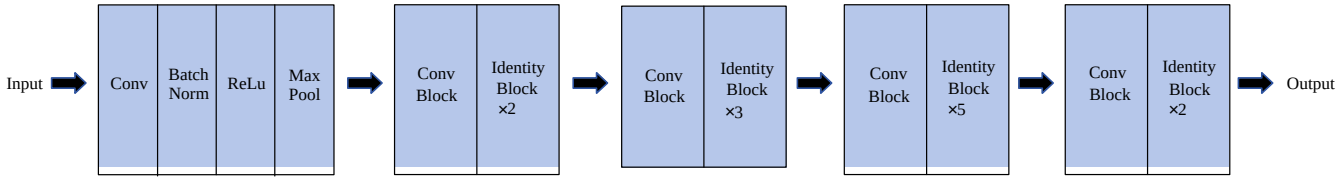

图 3 ResNet50 网络结构

Fig.3 ResNet50 network structure

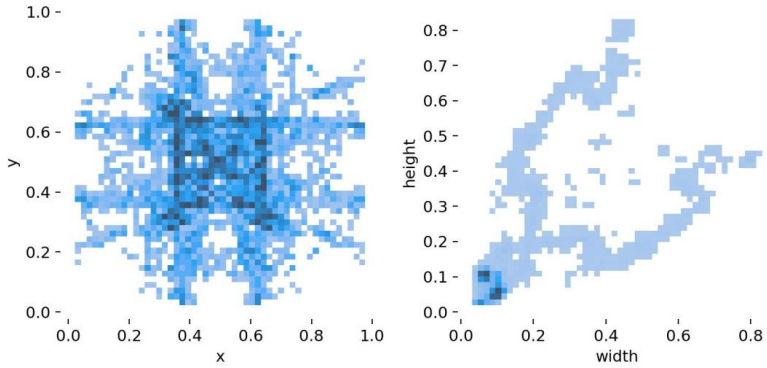

图 4 样本标签位置分布

Fig.4 Sample label location distribution

CA 编码过程如图 5 所示，首先输入通常是一个特征图 $C\times H\times W$，C 是通道数，表示特征图中的不同特征通道，H 是高度，表示特征图的垂直维度，W 是宽度，表示特征图的水平维度，对特征图的水平和垂直两个方向分别使用$(H,1)$和$(1,W)$的池化核进行平均池化，然后生成 z^h 及 z^w 两个水平及垂直独立方向的感知特征图，之后将具有特定方向信息的两幅特征图拼接到第三维度上，再利用 1×1 卷积和非线性激活函数产生过程特征图 $f=R^{(C/r)\times 1\times (H+W)}$，接着将 f 在通道维度拆分为 f^h 和 f^w 两个特征向量，然后分别利用 1×1 卷积进行升维操作，再经过激活函数 Sigmoid(x)得到注意力向量 g^h 和 g^w，最后将输入特征图和两个特征权重相乘从而增强了特征图的表示能力。

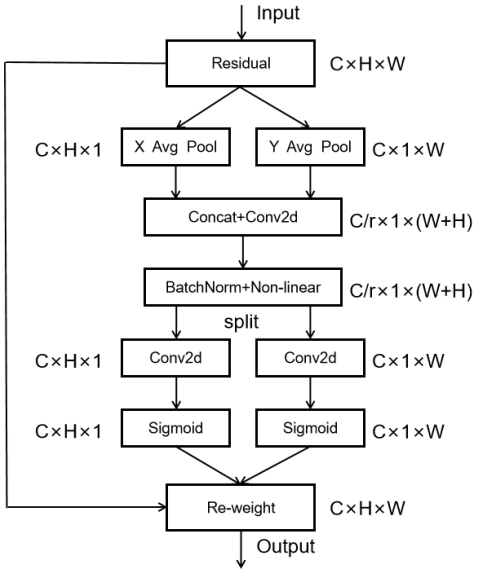

图 5 CA 编码

Fig.5 CA coding

2.3 Neck 网络改进

Res2Net[28-29]是在 2019 年提出的一种用于图像处理的深度卷积神经网络，其通过构造残差块内的分层残差连接可以在更细粒度上表示多尺度特征，同时也增大了每个网络层的感受野，因此本文将 Res2Net 添加到颈部的 C3 结构中，使得网络可以更好地捕捉和处理不同尺度的特征信息。

Res2Net 模块如图 6 所示，首先将输入特征经过一个 1×1 卷积层，然后送入多个卷积子模块，除了第一个子模块其余的子模块都会采用不同的 3×3 卷积核，每个子模块的 3×3 卷积核都可以接收所有其左边的特征信息，这样每一个输出都能有效地扩大感受野的范围，之后多个子模块的输出级联在一起，形成最终的输出。正因为 Res2Net 将多个子模块的信息融合在一个残差块中，所以能够捕捉到更丰富的特征信息，提高了特征融合能力和网络感知能力。

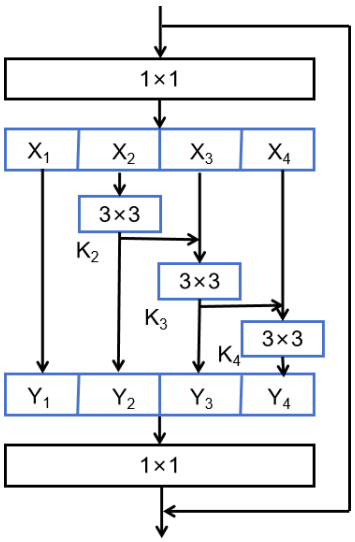

图 6 Res2Net 模块

Fig.6 Res2Net module

3 实验与结果分析

3.1 数据集

为了验证所提出的改进方法对不同尺寸的红外目标（飞机）及干扰的检测效果，选取了航天研究所的部分图片与红外图像弱小飞机目标检测跟踪公开数据集里的部分图片组建了本文的数据集。航天研究所的图片是包含目标（飞机）和干扰的，其中以大目

标和中目标居多，小目标较少，经过对图片的逐一挑选最后获得了 2500 张数据图片，另一个红外图像弱小飞机目标检测跟踪公开数据集则含有 22 个数据段的图片数据，每个数据段里是不同场景的图片，为了满足数据集样本的多样性，基本在每个数据段中都挑选了一些合适的图片数据，共筛选出 2270 张图片，这些图片包含了近距离、远距离、单个目标、多个目标、天空背景、地面背景等各种场景。所以本文最终得到一个包括 4770 张含目标（飞机）和干扰的数据集，然后按照 train:val＝9:1 的比例将图片划分成训练集和验证集，并使用 labelimg 软件对数据集进行了人工标注，共有两个标记类别，分别是飞机和干扰，标注完成后，每一张图片都对应着一个 xml 文件，此文件中包含了图片的主要信息，如图片的名称、路径、大小尺寸和通道以及图片中标注对象的类别和标注框的起点终点等，采用 VOC 数据集格式进行保存。图 7 为数据集中的部分样本图像。

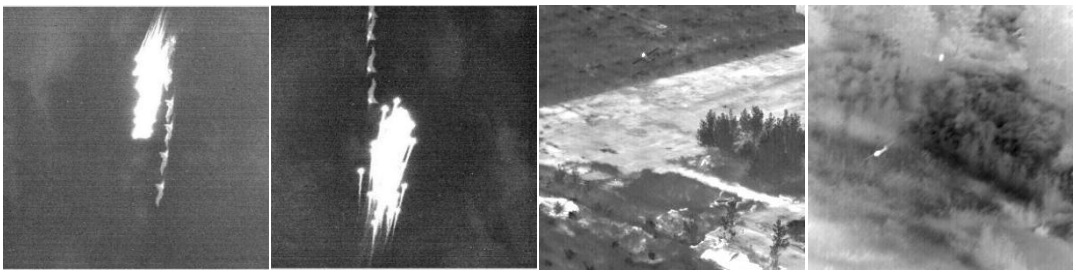

图 7 本文数据集部分样本

Fig.7 Part of the sample diagram of the data set in this paper

3.2 实验环境及评价指标

本文所有的实验均是在 Ubuntu20.04 操作系统上进行的，CPU 是 Intel(R) Xeon(R) Gold 5318Y，GPU 为显存大小是 16 GB 的 NVIDIA A16，开发语言使用的是 Python，其版本为 3.9，采用的深度学习框架是 PyTorch 2.0，并使用了 CUDA 进行 GPU 并行加速，使得训练时间大幅缩短，CUDA 版本为 11.7。

本实验进行训练时的网络参数如表 1 所示，模型在训练时使用 Warmup 进行预热训练，以此来减缓模型在初始阶段对小批量数据的过拟合现象，避免模型振荡以便保证模型深层次的稳定性，之后采用随机梯度下降算法对学习率进行更新。

表 1 实验训练参数

Table 1 Experimental training parameter

Parameters	Value
Epochs	100
Batch-size	16
Optimizer	SGD
Learning rate	0.01
Warmup_epochs	3
Weight_decay	0.0005

本文采用目标检测模型常用的平均精度（average precision, AP）、平均精度均值（mean AP, mAP）、精确率 *P*、召回率 *R* 以及 FPS 作为模型的评价指标来评估算法的性能，AP 是对于单个类别，通过计算不同置信度阈值下的精确率-召回率曲线围成的面积来评估模型性能的一个指标，AP 的取值范围在 0～1 之间，越接近 1 表示模型性能越好，mAP 是多个类别的 AP 的平均值，其综合考虑了所有类别的检测性能，是一个更全面的评估指标，*P* 是指模型预测为正类别的样本中，真正是正类别的样本所占的比例，是评价模型检测准确率的指标，*R* 是指模型正确检测到的正类别样本占所有正类别样本的比例，是评价模型能否将目标全部准确检测出来的指标，FPS 是一种用于评估模型推理速度的重要指标，表示模型每秒处理的图像帧数，FPS 越高，说明模型推理速度越快，上述衡量指标的计算公式分别为：

$$\mathrm{AP}=\int_0^1 P\mathrm{d}R \tag{1}$$

$$\mathrm{mAP}=\frac{\sum_{i=1}^{N}\mathrm{AP}_i}{N} \tag{2}$$

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}\times 100\% \tag{3}$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}\times 100\% \tag{4}$$

式中：AP 的值是 *P*-*R* 曲线与坐标轴所围成的面积值；mAP 的值是通过所有类别的 AP 值求均值所得到的；*N* 表示检测类别的总数；mAP 值越大表示模型的检测效果越好，识别精度越高；TP 表示模型预测为正的正样本；FP 表示模型预测为正的负样本；FN 表示模型

预测为负的正样本。

3.3 消融实验

为了验证所提方法的有效性，本文进行了消融实验来评估不同模块在相同实验条件下对目标检测算法性能的影响，消融实验中选择 YOLOv5s 作为基准模型，并采用 AP、mAP、P、R 和 FPS 作为实验评估指标，训练了 100epoch 后的结果如表 2 所示。

从表中可以看出，相较于原始 YOLOv5s 算法，模型 B 将骨干网络更换为 ResNet50，在保持 mAP_{50} 不变的情况下，$mAP_{50\text{-}95}$ 有 0.6%的提升，相对飞机来说，干扰的 $AP_{50\text{-}95}$ 提升得较为明显，有 1%的涨幅，模型 C 在模型 B 的基础上又增加了 CA 模块，通过增强网络的全局感受野和准确定位目标的能力，解决了本文图像标注尺度不均衡的问题，使得飞机与干扰的 $AP_{50\text{-}95}$ 都有了少量的提高，最终的 $mAP_{50\text{-}95}$ 也有 0.7%的提升，模型 D 在模型 C 基础上在颈部加入 Res2Net 模块，从而可以捕捉更多有用的特征信息，增强了特征融合和网络感知能力，使得飞机的 $AP_{50\text{-}95}$ 达到 69.8%，干扰的 $AP_{50\text{-}95}$ 达到 88.8%，最终的 $mAP_{50\text{-}95}$ 达到 79.3%，比原始 YOLOv5s 提高了 1.1%，虽然改进后的模型在 FPS 上有一定的下降，但是其在推理速度和检测精度的综合性能上会比 YOLOv5s 更占优势。训练及验证时的损失变化分别如图 8、9 所示。对于检测精度来说，$mAP_{50\text{-}95}$ 是比 mAP_{50} 更加严格的衡量指标，所以结果证明本文改进的方法在原 YOLOv5s 的基础上对检测精度有了一定的提高，可以完成对多尺度红外目标的检测任务。

3.4 对比实验

为了验证本文改进后的算法相比于其他目标检测算法的优越性，本文使用当前主流的目标检测算法 Faster-RCNN、SSD 及 YOLOv3 进行对比实验，结果如表 3 所示。

由表可知，Faster-RCNN 的 FPS 很低，说明其推理速度很慢，并且准确度较低，这可能会导致检测结果中有误检的情况，SSD 的 FPS 较高，速度较快，但是召回率较低，可能会造成部分目标的漏检，RCR-YOLO 的飞机与干扰的 AP_{50} 均达到 99.5%，性能较为均衡。总而言之，无论是准确度、召回率还是 mAP 值，本文改进后的算法均优于其他几个主流的目标检测算法，并且在精度和速度的综合性能上也表现得更加优越，在复杂背景下的多尺度红外目标检测中展现了高效的性能。

3.5 实验分析

为了更加直观地显示改进模型的效果，图 10～13 展示了几种不同目标检测算法的检测结果以及与之对应的 RCR-YOLO 的检测结果。

表 2 消融实验结果

Table 2 Ablation results

Model	Algorithm	AP_{50}/(%)		$AP_{50\text{-}95}$/(%)		*P* /%	*R* /%	mAP_{50} /%	$mAP_{50\text{-}95}$ /(%)	FPS
		Aero-plane	Interference	Aero-plane	Interference					
A	YOLOv5s	99.5	99.5	69.1	87.2	99.4	99.7	99.5	78.2	81.3
B	YOLOv5s+ResNet50	99.4	99.5	69.3	88.2	99.7	99.6	99.5	78.8	27
C	YOLOv5s+ResNet50+CA	99.5	99.5	69.4	88.4	99.5	99.8	99.5	78.9	28.2
D	YOLOv5s+ResNet50+CA+Res2Net(RCR-YOLO)	99.5	99.5	69.8	88.8	99.6	99.6	99.5	79.3	28.2

表 3 对比实验结果

Table 3 Comparative experimental results

Algorithm	AP_{50}/(%)		*P*/(%)	*R*/(%)	mAP_{50}/(%)	FPS
	Aeroplane	Interference				
Faster-RCNN	85.5	97.9	73.3	93.1	91.7	6.3
SSD	97.7	97.9	98.5	85.9	97.8	56.4
YOLOv3	98.7	97.5	97.1	92.8	98.1	18.4
RCR-YOLO	99.5	99.5	99.6	99.6	99.5	28.2

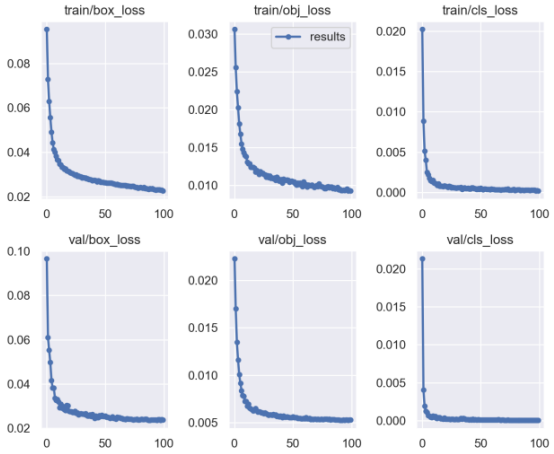

图 8 YOLOv5s 损失变化

Fig.8 YOLOv5s loss changes

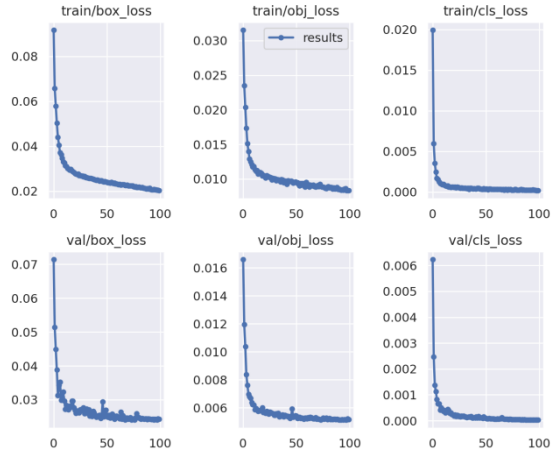

图 9 RCR-YOLO 损失变化

Fig.9 RCR-YOLO loss changes

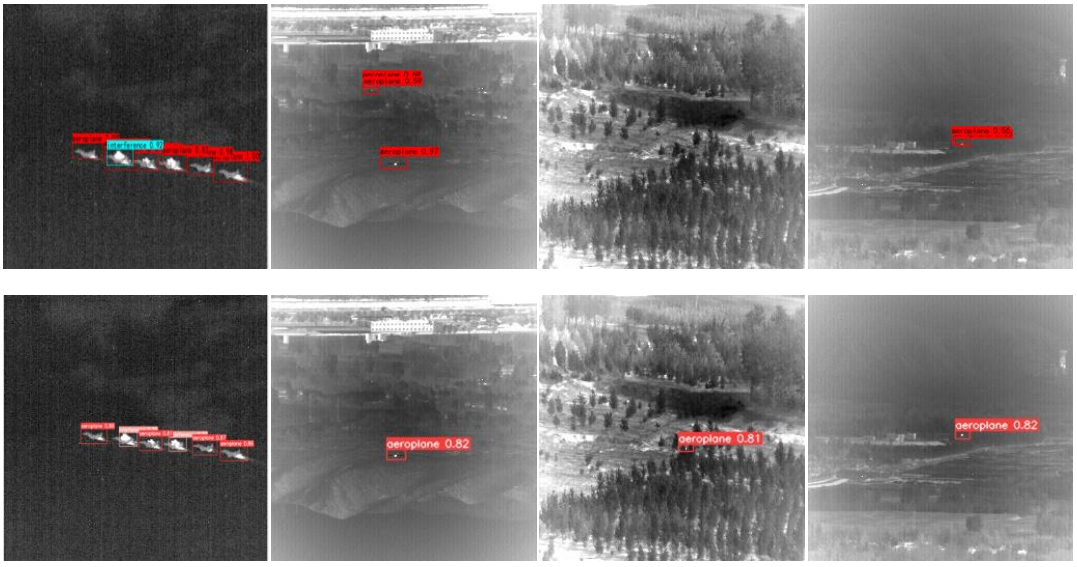

图 10 Faster-RCNN（上）与 RCR-YOLO（下）的检测结果对比

Fig.10 Comparison of detection results between Faster-RCNN(upper) and RCR-YOLO(down)

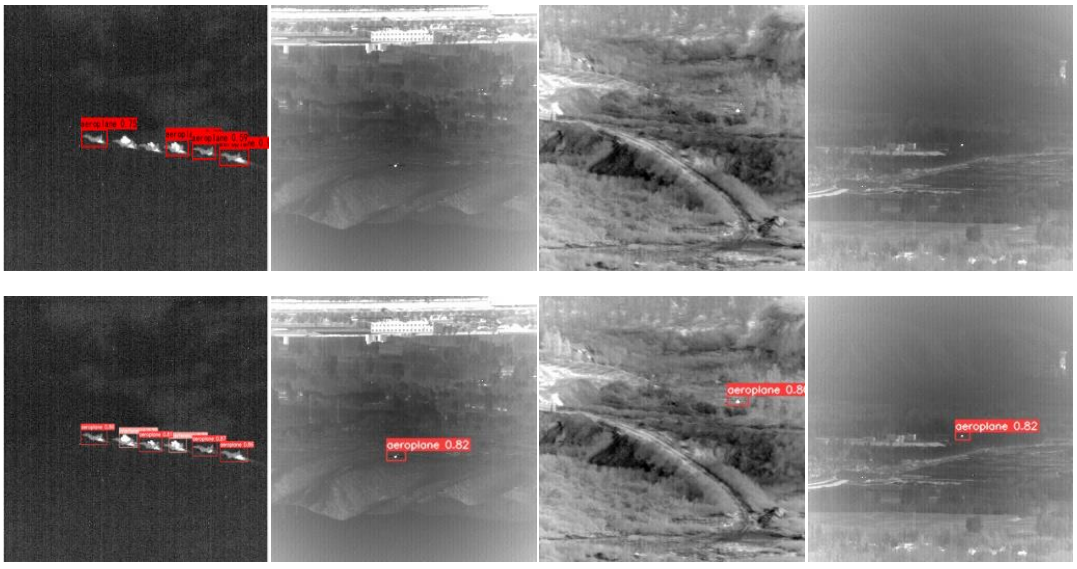

图 11 SSD（上）与 RCR-YOLO（下）的检测结果对比

Fig.11 Comparison of detection results between SSD(upper) and RCR-YOLO(down)

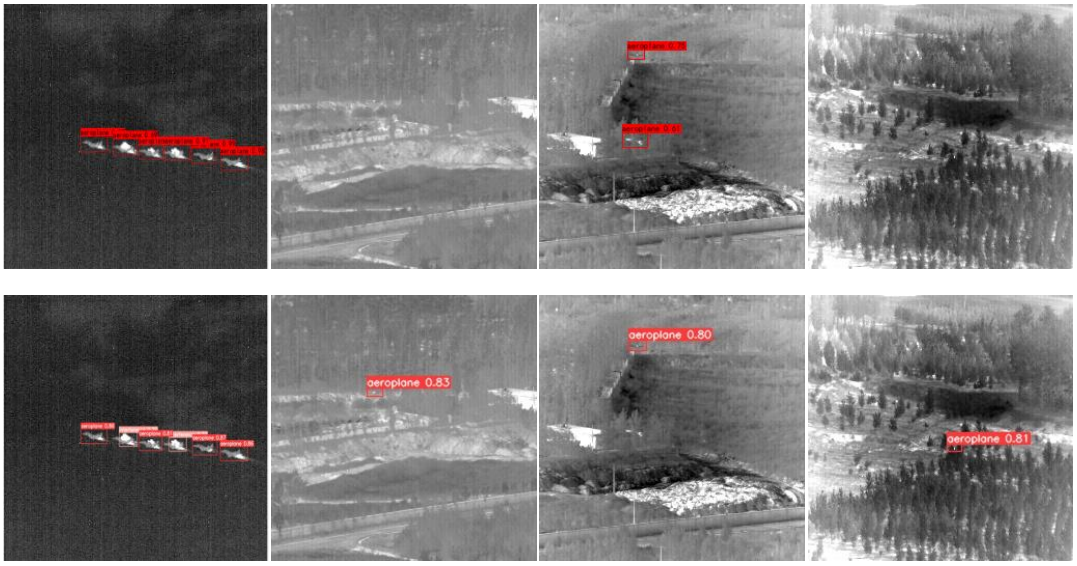

图 12 YOLOv3（上）与 RCR-YOLO（下）的检测结果对比

Fig.12 Comparison of detection results between YOLOv3(upper) and RCR-YOLO(down)

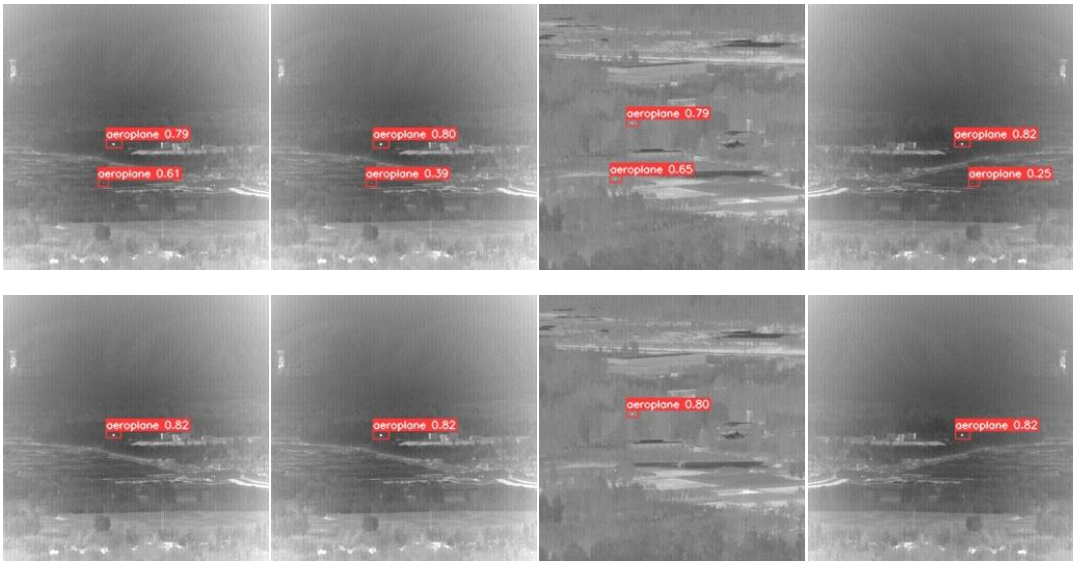

图 13 YOLOv5s（上）与 RCR-YOLO（下）的检测结果对比

Fig.13 Comparison of detection results between YOLOv5s(upper) and RCR-YOLO(down)

由检测结果可以看出，Faster-RCNN 由于精确率 *P* 很低，所以会出现较多的误检，从而导致其 mAP50 是最低的，SSD 的召回率 *R* 较低，对于在复杂背景下的小目标检测会出现较多的漏检情况，YOLOv3 相对于 Faster-RCNN 和 SSD 在精确率 *P* 和召回率 *R* 上取得了一定的平衡，但是仍存在部分的漏检和误检，YOLOv5s 对漏检的改善较大，但对部分小目标的检测还会出现误检结果，而本文提出的 RCR-YOLO 不仅避免了检测时的漏检情况还降低了复杂背景下小目标检测的误检率，相对于上述的其他几种目标检测算法，其检测效果更加优越。

4 总结

针对复杂背景下的多尺度红外目标检测中的漏检及误检问题，本文提出了一种改进的 YOLOv5s 算法 RCR-YOLO，首先将模型的骨干网络替换为 ResNet50，提高了网络的特征提取能力，然后在骨干网络的末端添加了 CA 注意力机制，使得特征图的重点区域得到更多的关注，最终在颈部网络中加入 Res2Net 模块，提高了网络处理多尺度和多分辨率特征信息的能力。实验结果表明，改进后的 YOLOv5s 算法在准确度和召回率上均有提升，并将 $mAP_{50\text{-}95}$ 从 78.2%提高到 79.3%，对复杂背景下的多尺度红外目标检测有较好的效果。

参考文献：

[1] LI K, WANG J, Jalil H, et al. A fast and lightweight detection algorithm for passion fruit pests based on improved YOLOv5[J]. *Computers and Electronics in Agriculture*, 2023, **204**: 107534.

[2] ZHANG Y, GUO K. Power plant indicator light detection system based on improved YOLOv5[J]. *Journal of Beijing Institute of Technology*, 2022, **31**(6): 605-612.

[3] YANG H, FANG Y, LIU L, et al. Improved YOLOv5 based on feature fusion and attention mechanism and its application in continuous casting slab detection[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023.

[4] ZHONG S, ZHOU H, MA Z, et al. Multiscale contrast enhancement method for small infrared target detection[J]. *Optik*, 2022, **271**: 170134.

[5] 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. IHBF-based enhanced local contrast measure method for infrared small target detection[J]. *Infrared Technology*, 2022, **44**(11): 1132-1138.

[6] JIANG C, REN H, YE X, et al. Object detection from UAV thermal infrared images and videos using YOLO models[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2022, **112**: 102912.

[7] CAO S, WANG T, LI T, et al. UAV small target detection algorithm based on an improved YOLOv5s model[J]. *Journal of Visual Communication and Image Representation*, 2023, **97**: 103936.

[8] LIU Z, GAO X, WAN Y, et al. An improved YOLOv5 method for small object detection in UAV capture scenes[J]. *IEEE Access*, 2023, **11**: 14365-14374.

[9] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR'05), 2005, **1**: 886-893.

[10] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]//2008 *IEEE Conference on Computer Vision and Pattern Recognition*, 2008: 1-8.

[11] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 580-587.

[12] Girshick R. Fast R-CNN[C]//*Proceedings* of the *IEEE International Conference on Computer Vision*, 2015: 1440-1448.

[13] REN Shaoqing, HE Kaiming, Ross Girshick, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **39**(6): 1137-1149.

[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] LIU W, Anguelov D, Erhan D, et al. Ssd: single shot multibox detector[C]//*Computer Vision–ECCV* 2016: 14*th European Conference*, 2016: 21-37.

[16] FU C Y, LIU W, Ranga A, et al. Dssd: deconvolutional single shot detector[J]. arXiv preprint arXiv:1701.06659, 2017.

[17] Jeong J, Park H, Kwak N. Enhancement of SSD by concatenating feature maps for object detection[J]. arXiv preprint arXiv:1705.09587, 2017.

[18] LI Z, ZHOU F. FSSD: feature fusion single shot multibox detector[J]. arXiv preprint arXiv:1712.00960, 2017.

[19] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 779-788.

[20] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 7263-7271.

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

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

[23] DING L, XU X, CAO Y, et al. Detection and tracking of infrared small target by jointly using SSD and pipeline filter[J]. *Digital Signal Processing*, 2021, **110**: 102949.

[24] WEI J, SU S, ZHAO Z, et al. Infrared pedestrian detection using improved UNet and YOLO through sharing visible light domain information[J]. *Measurement*, 2023, **221**: 113442.

[25] Terven Juan, Diana-Margarita Córdova-Esparza,et al. A comprehensive review of yolo architectures in computer vision: from yolov1 to yolov8 and yolo-nas[J]. *Machine Learning and Knowledge Extraction*, 2023, **5**(4): 1680-1716.

[26] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.

[27] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 13713-13722.

[28] GAO S H, CHENG M M, ZHAO K, et al. Res2net: a new multi-scale backbone architecture[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, **43**(2): 652-662.

[29] 袁志安, 谷雨, 马淦. 面向多类别舰船多目标跟踪的改进 CSTrack 算法[J]. **光电工程**, 2023, **50**(12): 16-31.
YUAN Zhian, GU Yu, MA Gan. Improved CSTrack algorithm for multi-class ship multi-object tracking[J]. *Opto-Electronic Engineering*, 2023, **50**(12): 16-31.