

基于改进 YOLO v3 算法的空中红外目标检测

李彦锴，许媛媛，刘子琪，陈玉清

(上海海事大学 物流工程学院，上海 201306)

摘要：为进一步提升空中作战条件下目标检测的性能，本文通过优化 YOLO v3，提出了一种基于空中红外目标的检测算法 EN-YOLO v3。该算法使用轻量的 EfficientNet 骨干网络作为 YOLO v3 的主干特征提取网络，使模型参数大幅减少，降低模型的训练时间；同时选用 CIoU 作为模型的损失函数，优化模型损失计算方法，提升模型的检测精度。结果表明，优化后的 EN-YOLO v3 目标检测算法与原 YOLO v3 相比模型尺寸减少了 50.03%，精准度提升了 1.17%，能够有效提升红外场景下空中目标的检测效果。

关键词：红外场景；空中目标检测；YOLO v3；模型优化

中图分类号：TP391 文献标识码：A 文章编号：1001-8891(2023)04-0386-08

Aerial Infrared Target Detection Based on Improved YOLO v3 Algorithm

LI Yankai, XU Yuanyuan, LIU Ziqi, CHEN Yuqing

(Department of Logistics Engineering, Shanghai Maritime University, Shanghai 201306, China)

Abstract: To further improve the performance of target detection under air combat conditions, a detection algorithm, namely EN-YOLO v3, based on an air infrared target and the optimization of YOLO v3, is proposed in this paper. The algorithm uses the lightweight EfficientNet backbone network as the backbone feature extraction network of YOLO v3 to reduce the number of model parameters and training time. Additionally, CIoU is used as the loss function of the model to optimize the model loss calculation method and improve its detection accuracy. The results show that compared with the original YOLO v3, the optimized EN-YOLO v3 target detection algorithm reduces the model size by 50.03% and improves the accuracy by 1.17%. This can effectively improve the detection of aerial targets in infrared scenes.

Key words: infrared scene, aerial target detection, YOLO v3, model optimization

0 引言

空中目标检测是对空作战领域的一个重要组成部分。目前很多常用的目标检测算法都是基于可见光图像来完成的，针对黑夜、多云等缺少光源、能见度低的情况，很难使目标有效成像，因此无法获得令人满意的检测效果^[1]。而红外成像技术依靠目标自身的红外辐射工作，无需借助外部环境光，能够在恶劣的观测条件下探测到目标的特征并进行成像，在目标捕获、跟踪等方面有重要应用^[2]。随着国家军队改革的不断深入，作战形式逐渐从机械化作战向智能化、信息化作战的方向发展^[3]。因此通过红外成像所得到的信息精准快速地检测出空中目标是一个非常重要的研究方向。

随着大数据时代的到来，深度学习、人工智能等

收稿日期：2022-05-03；修订日期：2022-06-08。

作者简介：李彦锴（1999-），男，河南南阳人，硕士研究生，研究方向：目标检测、红外图像处理。E-mail:lyk1481392566@163.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:yxu@shmtu.edu.cn。

基金项目：国家自然科学基金（青年项目）（61603246）。

技术展现出了迅猛发展的势头，以深度学习为基础的目标检测算法凭借其优良的检测性能逐渐成为了目标检测领域中人们的首选算法^[4]。当前的目标检测算法主要分为两阶段（two-stage）和单阶段（one-stage）两种。前者的两个阶段具体包括：为区分前景和背景，得到前景区域 RoI（region of interest）的第一阶段（通常用矩形框来表示，被称为 bbox（bounding box）），和利用 RoI 在 CNN 中所提取到的特征进行分类和回归预测的第二阶段（通常该阶段为多分类，需要区分出不同物体的类别）。R-CNN^[5]，Fast R-CNN^[6]，Faster R-CNN^[7]是目前较为常用的两阶段目标检测算法。后者是基于 anchor 直接在输入图像中完成目标的检测分类以及边框调整^[8]，无需其他操作。SSD（single shot multiBox detector）^[9]、YOLO^[10]（You only look once）是目前比较具有代表性的单阶段目标检测算法。

通常情况下,使用两阶段的目标检测算法能够在红外场景下的目标检测中得到更高的准确率,但是这类算法的检测速度难以满足空中作战对于实时性的要求^[11]。而单阶段目标检测算法只需一次特征提取即可完成目标检测,能够更好地适应空中作战中复杂多变的情况^[12]。YOLOv3 目标检测算法在经过前两代的改良之后,凭借其高精度和极快检测速度的优势,得到了广泛的应用^[13]。因此本文以 YOLOv3 算法为主体模型,对其进行改进和优化,使其可以从红外成像技术所得到的空中图像信息里快速精准地检测出目标。

1 目标识别算法

1.1 YOLOv3

DarkNet-53 是 YOLOv3 目标检测算法的骨干网络,在算法中主要负责对输入图像进行特征提取。YOLOv3 为了解决深度神经网络中可能出现的冗余、梯度爆炸的问题,在网络中使用残差连接、设置了快捷链路。YOLOv3 的网络结构如图 1 所示:首先将需要检测的图像尺寸调整为 416×416 (本文尺寸均指像素)后输入网络,在 Darknet-53(红框标注的部分)里进行特征提取,提取到的特征被称作特征层,是输

入图片特征的集合。这个过程会对输入的图片进行 5 次压缩,其中在第 3、第 4 和第 5 压缩结束时会得到 L1、L2 和 L3 三个有效特征层。之后借鉴了特征金字塔 (feature pyramid networks, FPN) 的思想,通过上采样和特征连接的方式将这 3 个有效特征层中不同尺度的特征融合在一起,使网络可以实现多尺度的预测。网络输出有 3 个分支 y_1 、 y_2 和 y_3 ,其中 L3 输出的特征是最高层的、最抽象,在经过卷积后会输出给适合较大目标检测的 y_1 ; L2 输出的特征在 L3 的基础上做了一个上采样和一个低层的特征连接 (concat) 的操作,在卷积后会输出给适合检测中等目标的 y_2 ; L1 输出的特征在 L2 的基础上做了一个上采样和一个更低层的特征连接,经过卷积后会输出给适合小目标检测的 y_3 。图中 DBL 结构作为网络的基本组成模块由 3 个操作组成:卷积、批量标准化和激活函数;Resn 结构表示每个 res_block 结构中 res_unit 的数量; concat 结构代表张量拼接。

YOLOv3 会通过 K-Means 聚类的方式得到适合不同尺度大小的先验框。由于这些先验框已经确定好了数据集中常见目标的宽和高,因此能使模型更容易学习。

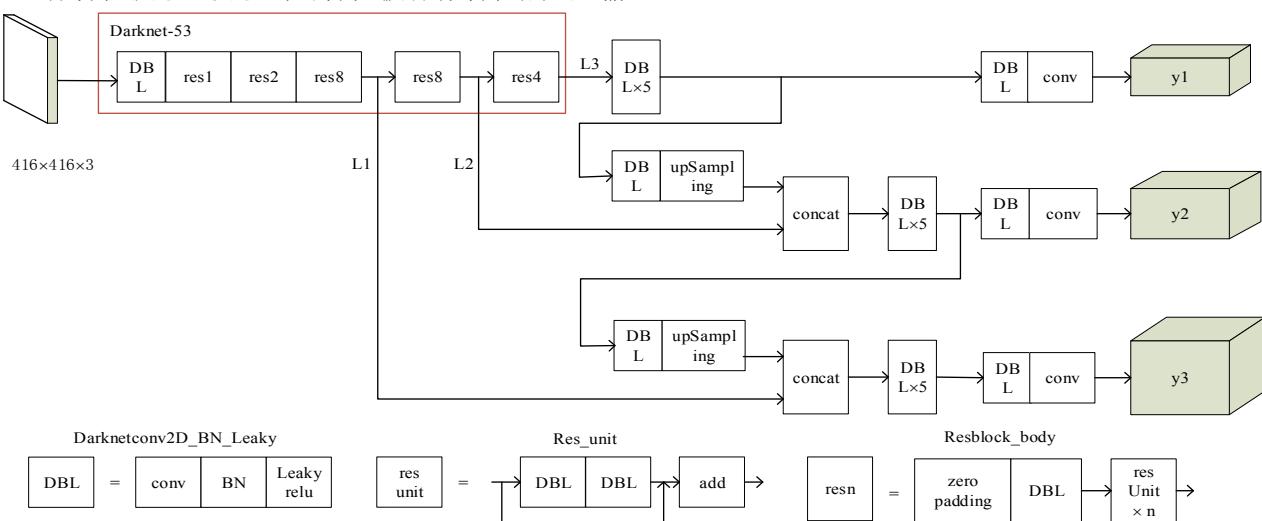


图 1 YOLOv3 网络结构

Fig.1 YOLOv3 network structure

1.2 EfficientNet

EfficientNet 网络是 2019 年 Tan M 等人提出的一组系列网络模型,该系列模型通过平衡缩放网络的深度 (depth)、宽度 (width) 和输入图像分辨率 (resolution) 这 3 个维度,使网络模型在精度、效率和模型大小上得到了最优的效果^[14]。EfficientNet 系列模型包括 8 个不同的网络结构 (EfficientNet-B0~EfficientNet-B7),可以根据实际情况和需求,选择最

合适的模型来完成目标检测。该系列模型能够以较少的模型参数量得到很好的检测结果,是当下目标检测性能最优的网络之一^[15]。

EfficientNet 网络通过采用复合缩放的方式,以一定的比例对卷积神经网络的 3 个维度进行扩展,寻找一组能使网络检测精度最大化的最优参数。模型缩放的关系式如公式(1)所示:

$$\begin{aligned} & \max_{d,w,r} \text{Accuracy}(N(d,w,r)) \\ \text{s.t. } & N(d,w,r) = \bigodot_{i=1, \dots, s}^{\wedge d \cdot L_i} F_i(X_{(r \cdot \hat{H}_{i,r}, w \cdot \hat{W}_{i,w}, \hat{C}_i)}) \quad (1) \end{aligned}$$

Memory(N) \leq target_memory

FLOPS(N) \leq target_flops

式中： d 、 w 和 r 分别代表网络的深度、宽度和分辨率的倍率；Memory 代表记忆存储器；FLOPS 代表每秒浮点运算次数。这组关系式的含义在于如何优化 d 、 w 和 r 三个参数，使模型能在给定条件的约束下获得最佳的精度。由于这 3 个参数存在一定的联系，因此通过一个复合系数 ϕ 对它们进行复合调整，以得到最优的组合参数。复合调整方法如公式(2)所示：

$$\begin{aligned} \text{depth: } & d = \alpha^\phi \\ \text{width: } & w = \beta^\phi \\ \text{resolution: } & r = \gamma^\phi \quad (2) \\ \text{s.t. } & \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \\ & \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \end{aligned}$$

式中： α 、 β 和 γ 分别代表网络深度、宽度和分辨率的分配系数。复合系数 ϕ 是一个可调的复合系数，复合系数的值是根据当前所拥有的资源进行人工调节的。对于基线网络 EfficientNet-B0 来说，通过调整公式(2)中的复合系数 ϕ ，就可扩展出剩余的 Efficientnet-B1 到 Efficientnet-B7 网络。

EfficientNet 是由多个倒置瓶颈卷积模块(Mobile Inverted Bottleneck Convolution, MBCConv)、卷积层、全局平均池化层和全连接层共同构成的^[16]。MBCConv 的网络结构如图 2 所示：利用 1×1 的卷积升高输入特征层的维数，然后进行一次深度可分离卷积，在标准化和激活函数之后施加注意力机制，再利用 1×1 的卷积降低特征层的维数使其与输入保持一致，最后将降维处理后的特征层与左边的残差边相加后输出。其中深度可分离卷积里的特征大小有 3×3 和 5×5 两种；注意力机制包含了最大池化、两次 1×1 的卷积、激活函数以及与原特征图相乘等操作。

Swish 激活函数如公式(3)所示：

$$f(x) = x \cdot \text{sigmoid}(nx) \quad (3)$$

式中： x 为输入； n 为可调参数。Swish 函数没有上边界，不会出现梯度饱和的现象；拥有下边界，可以产生更强的正则化效果；非单调且处处连续可导，可以使模型更容易训练。

2 YOLOv3 算法的改进

为了进一步提高 YOLO v3 目标检测算法对于空中红外目标的检测精度，降低模型的复杂度，拟使用

EfficientNet 的骨干网络代替原 YOLO v3 算法中的 DarkNet-53 完成对输入图像的特征提取，并使用性能优越的 CIoU 作为网络的损失函数，优化模型的损失计算方法，提升模型的预测精准度。

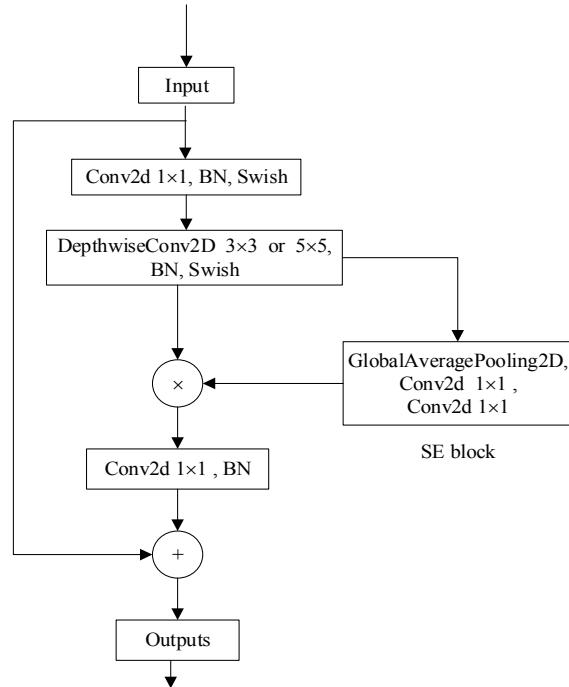


图 2 MBCConv 网络结构
Fig.2 MBCConv network structure

2.1 网络结构的改进

为了能够提升 YOLO v3 目标检测算法的性能，在考虑了实验环境以及数据集大小等实际因素后，设计了一组对比实验，通过对分析这组实验结果后，本文选择使用 EfficientNet-B4 的主干特征提取网络完成特征提取（实验结果分析见 3.3）。EfficientNet-B4 共包含 32 个 MBCConv 模块，为了能够使 EfficientNet-B4 的主干特征提取网络的输出尺寸与 DarkNet-53 的输出尺寸保持一致，使网络可以正常运行，取出第 10 个、第 22 个和第 32 个 MBCConv 模块的输出层作为有效特征层，继续完成不同尺度间的特征融合。同时网络针对不同尺度的目标分别构建检测层，提升模型检测的精准度。

改进后的网络结构图如图 3 所示：将尺寸为 416×416 的图像输入网络模型中，在 Stem、Module2、Module3、Module4 和 Module6 这 5 个模块中对输入的图像进行长和宽的压缩，Module1、Module5 和 Module7 不对图像进行压缩，只进行特征提取。通过主干特征提取网络，可以在 Module 3、Module 5 和 Module 7 处得到 3 种不同尺度的有效特征层 (L1: 52×52 , L2: 26×26 , L3: 13×13)，分别用来检测小、中、大 3 个不同尺度大小的目标。之后通过卷积、

上采样、张量拼接等方式将深层网络与浅层网络融合，增加低层特征的表达能力，提升网络的检测性能，最后将不同尺度的目标分配到不同的预测层 (y_1 、 y_2 、 y_3) 实现目标检测。

2.2 损失函数的改进

IoU (Intersection over Union) 是目标检测领域的常用指标，通过计算“预测框”和“真实框”的交并比来反映目标检测的精准度^[17]。YOLO v3 的损失函

数由 3 部分组成：预测框定位损失，目标置信度损失和目标类别损失^[18]。其中中心点距离损失和预测框的宽高损失组成了预测框损失，因为预测框本身是一个矩形，所以通过 IoU 来反映预测框的精准度是很直观的。IoU 的定义如公式(4)所示：

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

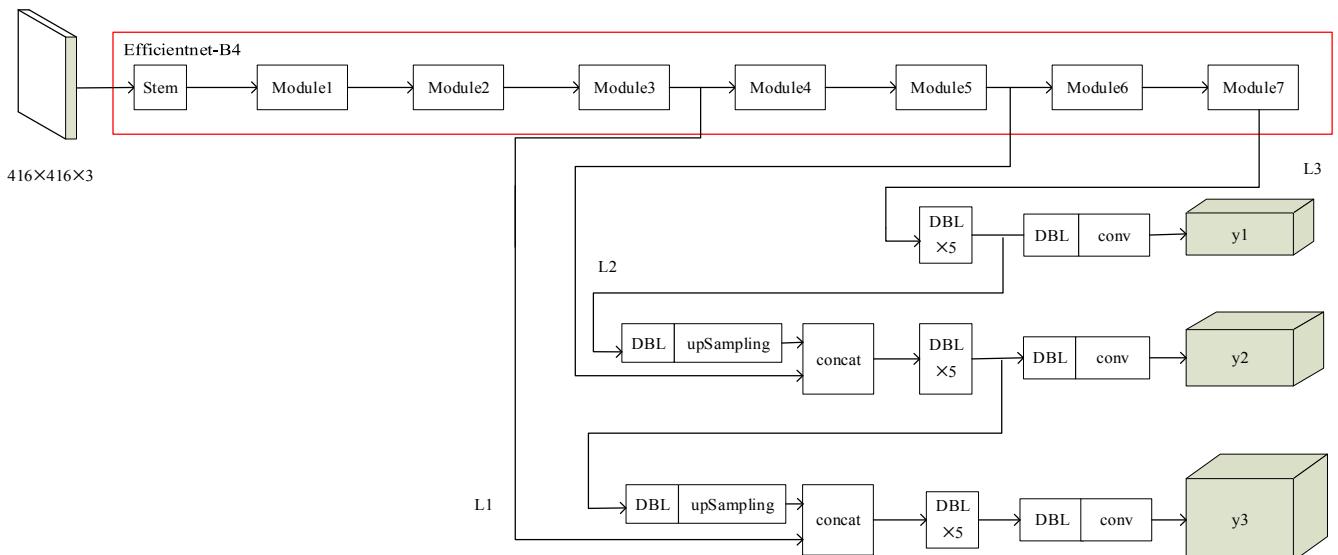


图 3 EN-YOLOv3 网络结构
Fig.3 EN-YOLOv3 network structure

但 IoU 也有其自身的缺陷：对于没有重叠的预测框与真实框来说， $|A \cap B|$ 的值为 0，导致得到的 IoU 也为 0，无法反映出两者之间的距离大小，且由于没有梯度回传，无法继续进行学习训练。为了解决这个问题，后来的研究者们分别提出了 GIoU、DIoU 和 CIoU 作为损失函数^[19]。GIoU 在 IoU 的基础上加强了对于非重合区域的关注，能够更好地反映出预测框与真实框的重合度；但是当目标框与预测框处于同一平面时，GIoU 就会退化为 IoU，会造成收敛变慢、回归不够准确等问题。DIoU 和 CIoU 的作者 Zheng 等人表示一个优秀的回归定位损失应该包括重叠的面积、中心点距离和长宽比这 3 种几何参数^[20]。因此 CIoU 在计算损失时考虑了这些因素，在损失函数的迭代过程中使预测框不断向真实框移动，尽可能保证了预测框和真实框的宽高纵横比更为接近，提高了检测精度，而且加快了预测框的回归收敛速度。CIoU 的损失函数如公式(5)所示：

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + mv \quad (5)$$

式中： c 为能够将预测框和真实框同时包含在一起的最小闭包区域的对角线距离； $\rho(b, b^{\text{gt}})$ 为真实框和预测

框中心点的欧氏距离； m 为权重函数； v 为度量长宽比相似性的函数。 m 和 v 的定义如公式(6)、公式(7)所示：

$$m = \frac{v}{1 - \text{IoU} + v} \quad (6)$$

$$v = \frac{4}{\pi^2} (\arctan \frac{w_1^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w_1}{h})^2 \quad (7)$$

式中： w_1 、 h 和 w_1^{gt} 、 h^{gt} 分别代表预测框的宽高和真实框的宽高。

3 实验结果与分析

本文的实验条件如表 1 所示。

表 1 实验环境

Table 1 Lab environment

Parameters	Configuration
Operating system	Linux
Video memory	16 G
Training framework	TensorFlow-GPU 2.3.0
Programming language	Python
GPU	NVIDIA RTX2080Ti

3.1 实验数据与评价指标

本实验的数据来自某航天研究所的红外成像视频文件,包括单飞行器目标干扰投放和多飞行器目标干扰投放。单目标视频文件共有 2627 帧图片,每帧图片的大小是 512×640 ; 多目标视频文件共有 589 帧图片,每帧图片的大小是 512×640 。实验的目标是:投放干扰源前的飞行器目标检测,投放干扰源后飞行器与干扰的目标检测。

本文从两份视频文件中选取部分合适的图片组成自建数据集。由于该视频文件的红外图像中,部分目标细节模糊、整体的对比度较低,直接利用从视频文件中选取的原图进行模型训练会导致检测精度较低、误差较大,所以需要对图像进行预处理,增强目标清晰度;针对自建数据集的样本较少,采用几何变换(镜像、旋转等)、像素调整(调整对比度、亮度等)方式增加图像样本,扩充自建数据集。目前自建数据集含有 2500 张红外场景图像,按照比例随机取 2025 张图像、225 张图像、250 张图像分别作为模型的训练集、验证集和测试集。

本文使用平均准确率(mean average precision, mAP)对目标检测算法进行性能评价,以了解模型正确分类的能力。计算 mAP 需要两个参数:查准率(Precision)和召回率(Recall)。查准率是指预测正确的正样本占全部正样本的比例,定义如公式(8)所示:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

召回率是指预测为正样本占全部正样本的比例,定义如公式(9)所示:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

式中: TP 表示预测和实际都为正样本; FP 表示实际的负样本被预测为正样本; FN 表示实际的正样本被预测为负样本。

基于召回率和查准率,可计算得到每个种类的 AP(average precision),其定义如公式(10)所示:

$$\text{AP} = \int_0^1 P(R) dR \quad (10)$$

mAP 就是对所有种类的 AP 求均值,其定义如公式(11)所示:

$$\text{mAP} = \frac{\sum_{i=1}^k \text{AP}_i}{k} \quad (11)$$

式中: k 为要检测的目标类别总数; AP_i 和 mAP 的值在 0 和 1 之间,越接近于 1,说明目标检测算法的性能越好。

3.2 训练过程

由于网络内部的参数过多,需要大量训练样本才能拟合出一个性能良好的模型。而本文数据集的样本有限,若直接将其用于模型训练,会产生训练速度慢以及准确率不高的问题。因此决定采用迁移学习的思想,先使用 PASCAL VOC 数据集对网络模型进行训练获得预训练模型,针对空中红外目标的检测,在预训练模型的基础上使用自建数据集进行迁移学习。

迁移学习的过程分为两个阶段:第一个阶段是冻结训练的阶段,这个阶段会冻结模型的主干网络,不改变特征提取网络的参数,只会对网络的部分结构进行微调,这个阶段将学习率设置为 0.001, batch_size 设置为 16, epoch 设置为 50, 冻结的层数为前 468 层;第二个阶段是解冻阶段的训练,这个阶段模型的主干网络不再冻结,特征提取网络开始发生改变,其中学习率设置为 0.0001, batch_size 设置为 8, epoch 设置为 50。

3.3 同系列算法实验对比

为了能够在 EfficientNet 系列网络中选择出最合适的模型作为本文算法的主干特征提取网络,本文设计了一组实验进行对比分析,实验中分别使用 EfficientNet B1-B6 替换 YOLOv3 的主干网络 Darknet-53, 经过模型的训练测试后得到各个模型的 mAP 和 Model size, 实验结果如表 2 所示。从表 2 中可以看出 EfficientNet 系列网络的 B1、B2、B3、B5 的 mAP 都低于 YOLOv3, B4 和 B6 的 mAP 高于 YOLOv3, 而 B4 以不到 B6 一半的模型大小得到了和 B6 相似的检测精度,因此最终选择 EfficientNet-B4 为本文算法的主干特征提取网络。

3.4 实验结果定性分析

在自建数据集中,随机抽取 3 类测试图像,测试 3.2 中通过迁移学习训练所得到的模型,单目标飞行器的检测效果如图 4 所示,多目标飞行器的检测效果如图 5 所示,干扰条件下多目标飞行器的检测效果如图 6 所示。

从图中可以看出,无论是不同飞行姿态下的单目标飞行器,还是有无干扰下的多目标飞行器,均可通过本模型得到正确标定,且标定矩形框的大小与目标区域大小相符。实验结果可以定性表明,改进后的 EN-YOLOv3 算法能够从红外图像中准确检测和定位空中红外目标。

3.5 实验结果定量分析

为定量验证改进算法的有效性,本文采用消融研究的思想设立了 3 组模型并分别进行训练,3 组模型结构见表 3,表格中的“√”和“×”分别代表了使用和不使用该模块。其中,Model 1 是原始的 YOLOv3

模型, Model 2 和 Model 3 是对主干特征提取网络和损失函数分别进行改进验证的模型, Model 3 为本文所构建的模型。

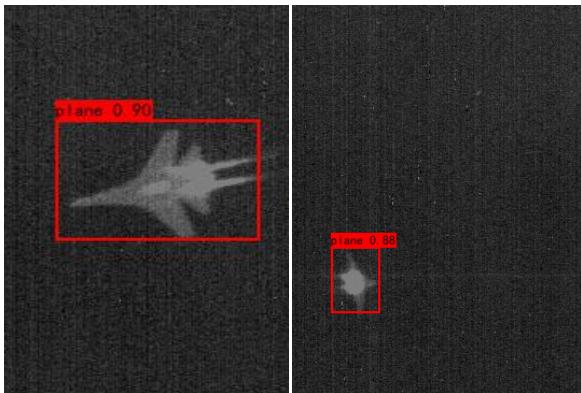


图 4 基于 EN-YOLOv3 模型的单目标检测

Fig.4 Single target detection based on EN-YOLOv3 model

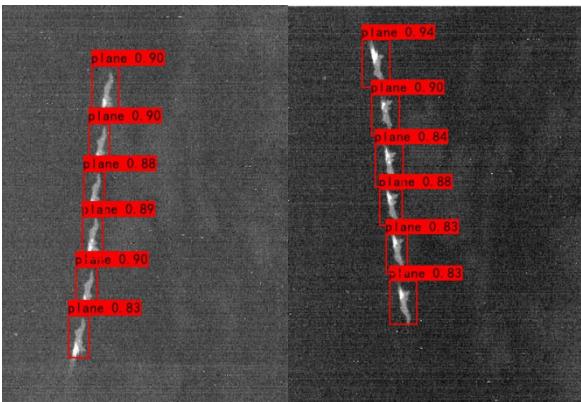


图 5 基于 EN-YOLOv3 模型的多目标检测

Fig.5 Multi-target detection based on EN-YOLOv3 model

不同模型下的 mAP 和 FPS 如表 4 所示, 从表 4 中可以看出, Model 2 使用 EfficientNet-B4 作为网络的主干特征提取网络时, 得到的 mAP 与 YOLO v3 (Model 1) 相比上升了 0.38 个百分点, Model 3 在

Model 2 的基础上使用 CIoU 作为损失函数后, 得到的 mAP 与 YOLOv3 (Model 1) 相比上升了 1.17 个百分点。另外 Model 2 和 Model 3 在目标检测速度方面略低于 Model 1, 但是差距很小, 能够满足红外场景下空中目标检测对于算法实时性的要求。由此可见经过主干网络和损失函数改进后的 Model 3 可以在满足目标检测速度要求的同时更加精准地检测出空中红外目标。

表 5 为经过训练后各组模型的大小和参数情况。表 5 的数据表明, 就模型大小而言, 改进后算法 (Model 3) 相比于 YOLO v3 (Model 1) 降低了 50.03%, 模型的参数量也大大减少, 表明使用 EfficientNet 作为目标检测模型的主干特征提取网络相比 DarkNet-53 更加轻量高效。对比分析表明, 本文所提出的 EN-YOLO v3 目标检测算法与原 YOLOv3 算法相比有了较大的提升, 说明本文提出的目标检测算法更加适合检测空中的红外目标。

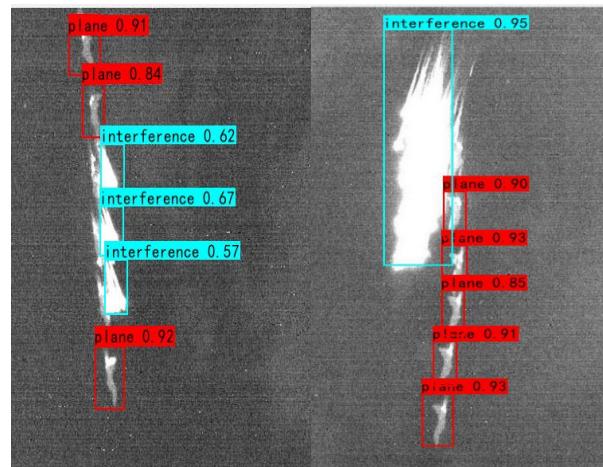


图 6 基于 EN-YOLOv3 模型的干扰条件下的多目标检测

Fig.6 Multi-target detection under interference conditions based on EN-YOLOv3 model

表 2 同系列算法实验对比

Table 2 Experiment comparison of the same series of algorithms

Backbone feature extraction network	mAP@0.5	Model size /MB
Darknet-53	0.8251	235.5
Efficientnet-B1	0.7822	50.9
Efficientnet-B2	0.8216	62.1
Efficientnet-B3	0.8185	78.1
Efficientnet-B4	0.8289	117.7
Efficientnet-B5	0.8203	173.7
Efficientnet-B6	0.8292	238.4

表3 基于消融实验所设立的不同模型结构

Table 3 Different model structures established based on ablation experiments

Model	Backbone feature extraction network		CIoU
	Darknet-53	Efficientnet-B4	
Model 1	√	✗	✗
Model 2	✗	√	✗
Model 3	✗	√	√

表4 不同模型下的 mAP 和 FPS

Table 4 mAP and FPS under different models

Model	Model 1	Model 2	Model 3
mAP@0.5	0.8251	0.8289	0.8368
FPS	45	43	43

表5 不同模型结构下的模型尺寸和参数数量

Table 5 Model size and parameter quantity under different model structures

Model	General parameters	Trainable parameters	Non-trainable parameters	Model size /MB
Model 1	61581727	61529119	52608	235.5
Model 2	30562887	30428887	134000	117.7
Model 3	30562887	30428887	134000	117.7

4 结语

空中红外目标检测是现代空战中的一个重要组成部分，目标检测的能力在实战当中有很大的影响。为了能够进一步提升目标检测算法的性能，本文对于 YOLOv3 算法进行了优化。首先摒弃参数过多的 DarkNet-53，使用 EfficientNet 的骨干网络完成特征的提取，降低模型的复杂度，减少模型的参数量，提升了训练速度；使用 CIoU 作为网络模型的损失函数，在计算损失时考虑中心点距离、重叠面积和长宽比这 3 个要素，使模型的预测框更加符合真实框，提升了模型预测的精准度。实验的测试结果表明，本文提出的 EN-YOLO v3 算法在检测速度略低于 YOLO v3 的情况下，模型大小比 YOLO v3 降低了 50.03%，精准度比 YOLO v3 提升了 1.17%。在下一步的研究工作中，将会以本文为基础，继续扩展数据集，实现红外场景下不同机型的检测。

参考文献：

- [1] 秦鹏, 唐川明, 刘云峰, 等. 基于改进 YOLO v3 的红外目标检测方法[J]. 计算机工程, 2022, 48(3): 211-219.
QIN Peng, TANG Chuanming, LIU Yunfeng, et al. Infrared target detection method based on improved YOLO v3[J]. Computer Engineering, 2022, 48(3): 211-219.

- [2] 董小虎, 傅瑞罡, 高颖慧, 等. 复杂背景下红外小目标自适应检测跟踪[J]. 航空兵器, 2019, 26(6): 22-28.
DONG Xiaohu, FU Ruigang, GAO Yinghui, et al. Detecting and tracking of small infrared targets in complex backgrounds[J]. Aero Weaponry, 2019, 26(6): 22-28.
- [3] 侯毅苇, 李林汉, 王彦. 结合红外显著性目标导引的改进 YOLO 网络的智能装备目标识别研究[J]. 红外技术, 2020, 42(7): 644-650.
HOU Yiwei, LI Linhan, WANG Yan. Research on intelligent equipment target recognition based on improved YOLO network combined with infrared saliency target guidance[J]. Infrared Technology, 2020, 42(7): 644-650.
- [4] 宋忠浩, 谷雨, 陈旭, 等. 基于加权策略的高分辨率遥感图像目标检测[J]. 计算机工程与应用, 2021, 57(13): 199-206.
SONG Zhonghao, GU Yu, CHEN Xu, et al. Object detection in high resolution remote sensing image based on weighting strategy[J]. Computer Engineering and Applications, 2021, 57(13): 199-206.
- [5] Girshick R, Donahue J, Darrell T, et al. Rich feature Hierarchies for accurate object detection and semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [6] Girshick R. Fast R-CNN[C]//IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [7] REN S, HE K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6): 1137-1149.

- [8] 杨涛, 戴军, 吴钟建, 等. 基于深度学习的红外舰船目标识别[J]. 红外技术, 2020, 42(5): 426-433.
YANG Tao, DAI Jun, WU Zhangjian, et al. Infrared ship target recognition based on deep learning[J]. *Infrared Technology*, 2020, 42(5): 426-433.
- [9] LIU W, Anguelov D, Erhan D, et al. SSD: Single Shot Multi Box Detector[C]//European Conference on Computer Vision, 2016: 21-37.
- [10] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]// IEEE Conference on Computer Vision & Pattern Recognition, 2016: 779-788.
- [11] 杨鹏博. 基于深度学习的红外弱小目标检测与主动跟踪研究[D]. 上海: 中国科学院大学(中国科学院上海技术物理研究所), 2021.
YANG Pengbo. Research on Infrared Weak and Small Target Detection and Active Tracking Based on Deep Learning[D]. Shanghai: University of Chinese Academy of Sciences(Shanghai Institute of Technical Physics Chinese Academy of Sciences), 2021.
- [12] 李慕锴, 张涛, 崔文楠. 基于 YOLO v3 的红外行人小目标检测技术研究[J]. 红外技术, 2020, 42(2): 176-181.
LI Mukai, ZHANG Tao, CUI Wennan. Research on infrared pedestrian small target detection technology based on YOLO v3[J]. *Infrared Technology*, 2020, 42(2): 176-181.
- [13] Redmon J, Farhadi A. YOLO V3: an incremental improvement [EB/OL].[2023-04].<https://doi.org/10.48550/arXiv.1804.02767>.
- [14] TAN M, LE Q V. Efficientnet: rethinking model scaling for convolutional neural networks[EB/OL].[2023-04]. <https://doi.org/10.48550/arXiv.1905.11946>.
- [15] 王振亚, 赵继红, 王艳鹏, 等. 基于改进 EfficientNet 网络的药片检测分类的研究[J]. 现代计算机, 2021, 27(28): 27-32.
WANG Zhenya, ZHAO Jihong, WANG Yanpeng, et al. Research on detection and classification of pills based on improved EfficientNet network[J]. *Modern Computer*, 2021, 27(28): 27-32.
- [16] 赵鹏菲, 黄丽佳. 一种基于 EfficientNet 与 BiGRU 的多角度 SAR 图像目标识别方法[J]. 雷达学报, 2021, 10(6): 895-904.
ZHAO Pengfei, HUANG Lijia. A multi-angle SAR image target recognition method based on EfficientNet and BiGRU[J]. *Journal of Radars*, 2021, 10(6): 895-904.
- [17] 许伟, 熊卫华, 姚杰, 等. 基于改进 YOLO v3 算法在垃圾检测上的应用[J]. 光电子·激光, 2020, 31(9): 928-938.
XU Wei, XIONG Weihua, YAO Jie, et al. Application of garbage detection based on improved YOLO v3 algorithm[J]. *Journal of Optoelectronics-Laser*, 2020, 31(9): 928-938.
- [18] 刘芳, 刘玉坤, 张白. 基于 D-YOLOv3 检测网络的温室叶菜幼苗图像检测[J]. 江苏农业学报, 2021, 37(5): 1262-1269.
LIU Fang, LIU Yukun, ZHANG Bai. Image detection of greenhouse leafy vegetable seedlings based on D-YOLOv3 detection network[J]. *Jiangsu Journal of Agricultural Sciences*, 2021, 37(5): 1262-1269.
- [19] Rezatofighi H, Tsoi N, Gwak J Y , 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), 2019: 658-666.
- [20] 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: 12993-13000.