

# 针对多尺度目标的轻量级红外目标检测算法

郑璐<sup>1,2</sup>, 彭月平<sup>2</sup>, 周彤彤<sup>1,2</sup>

(1. 武警工程大学 研究生大队, 陕西 西安 710086; 2. 武警工程大学 信息工程学院, 陕西 西安 710086)

**摘要:** 针对现有基于深度学习的红外目标检测算法参数量大、复杂度较高、对多尺度目标检测性能较差等问题, 提出了一种针对多尺度目标的轻量级红外目标检测算法。算法以 YOLOv3 为基础, 采用 MobileNet V2 轻量级骨干网络、设计改进的简化空间金字塔结构 (simSPP)、Anchor Free 机制、解耦头和简化正负样本分配策略 (SimOTA) 分别对 Backbone、Neck 和 Head 进行优化, 最终得到模型大小为 6.25 M, 浮点运算量 2.14 GFLOPs 的 LMD-YOLOv3 轻量级检测算法。在构建的 MTS-UAV 数据集上 mAP 达到 90.5%, 在 RTX2080Ti 显卡上 FPS 达到 99, 与 YOLOv3 相比 mAP 提升了 2.60%, 模型大小为 YOLOv3 的 1/10。

**关键词:** 目标检测; 多尺度; 轻量级算法

中图分类号: TP391.41

文献标志码: A

文章编号: 1001-8891(2023)05-0474-08

## A Lightweight Infrared Target Detection Algorithm for Multi-scale Targets

ZHENG Lu<sup>1,2</sup>, PENG Yueping<sup>2</sup>, ZHOU Tongtong<sup>1,2</sup>

(1. Postgraduate Brigade, Engineering University of PAP, Xi'an 710086, China;

2. College of Information Engineering, Engineering University of PAP, Xi'an 710086, China)

**Abstract:** To solve the problems of large parameters, high complexity, and poor detection performance of multiscale targets in the existing infrared target detection algorithms based on deep learning, a lightweight infrared target detection algorithm for multiscale targets is proposed. Based on YOLOv3, the algorithm uses the MobileNet V2 backbone network, simplified spatial pyramid structure (simSPP), anchor-free mechanism, decoupling head, and simplified positive and negative sample allocation strategies (SimOTA) to optimize the backbone, neck, and head, respectively. Finally, LMD-YOLOv3 with the model size of 6.25 M and floating-point computation of 2.14 GFLOPs was obtained. Based on the MTS-UAV data set, the mAP reached 90.5%, and on the RTX2080Ti dataset, the FPS reached 99. Compared with YOLOv3, mAP increased by 11.7%, and the model size was only 1/10 of YOLOv3.

**Keywords:** object detection, multi-scale, lightweight algorithm

## 0 引言

随着红外成像设备的发展, 红外探测凭借其隐蔽性强、成本低廉和可全天候作业等优势, 在智能监控、精确制导等多个领域得到了广泛应用。作为目标检测中的一个重要分支, 红外目标检测已经成为当前研究热点之一。由于红外图像不具备颜色和纹理等外观信息, 目标检测中的可用特征少, 且红外成像效果还受大气热辐射的不均匀、不稳定性制约, 因此研究适用于复杂背景下的检测算法是红外目标检测研究中的

### 重点和难点。

基于手工固定滑动窗口、步长和固定超参数的传统检测方法对于环境多样性不具有很好的鲁棒性能, 通常只对特定背景有效, 没有办法获取准确的语义信息, 检测的精度不高。基于深度卷积神经网络的方法凭借强大的特征提取和学习能力, 能够从复杂图像中提取特征并进行分层表示, 这些特性使得设计基于深度卷积神经网络的目标检测算法成为当前复杂背景下红外目标检测的有效突破口。近年来, 随着计算机硬件的发展以及红外数据集的扩充, 越来越多的基于

收稿日期: 2022-06-05; 修订日期: 2022-06-23。

作者简介: 郑璐 (1998-), 女, 汉族, 浙江金华人, 硕士研究生。研究方向: 深度学习与目标检测。E-mail: 1095496345@163.com。

通信作者: 彭月平 (1974-), 男, 汉族, 湖北人, 工学博士后, 教授。研究方向: 战场环境建模与仿真。E-mail: 1095496345@qq.com。

基金项目: 装备综合研究项目 (WJ2021A030131); 科研单位自主选题研究项目 (ZZKY20223105)。

深度学习的红外目标检测算法问世。以 Faster RCNN 系列<sup>[1-3]</sup>为代表的两阶段目标检测算法具有较高的检测精度, 以 SSD 系列<sup>[4]</sup>、YOLO 系列<sup>[5-7]</sup>为代表的单阶段目标检测算法在实时目标检测的应用中扮演着重要角色。Liu 等人<sup>[8]</sup>率先将基于深度学习的目标检测算法应用到红外目标检测中, 但由于图像尺寸小、只有灰度信息、特征不明显, 致使机器视觉领域现有的基于深度学习的目标检测算法不适合于红外小目标检测<sup>[9]</sup>。为解决红外目标淹没在背景杂波中时, 虚警率高、检测率低的问题, Zhao 等人<sup>[10]</sup>利用卷积神经网络巧妙地将检测问题转化为模式分类问题, 提出了一种天空复杂云背景下红外小目标检测算法; 文献[11]利用回归型的深度卷积神经网络进行背景成分抑制, 通过阈值分割提取出候选目标区域; Fan 等人<sup>[12]</sup>在分析红外图像特点的基础上, 通过角点检测提取潜在目标区域以保证检测率, 再将潜在目标区域输入基于卷积神经网络的分类器中, 进行非目标区域的剔除, 成功降低了检测的虚警率。对于目标遮挡、形变、目标部位未完全暴露态势下的目标检测问题, Zhang 等人<sup>[13]</sup>在空中飞机目标结构、红外特性分析的基础上, 定义带有约束条件的损失函数, 构建用于空中目标要害部位识别的深度卷积网络模型, 提出一种基于关键点检测卷积网络的空中红外目标要害部位检测算法, 在抗噪性、鲁棒性上表现出优越的性能。

目前的目标检测算法大多是基于 anchor box 设定, 能够检测到的目标大小取决于 anchor box 的大小, 在 anchor box 数量一定的情况下, 算法能够检测到目标的大小跨度是有限的, 而增加 anchor box 数量会带来更多计算量。在实际场景中, 待检测的目标可能从各个方位以不同尺度出现, 因此研究红外多尺度目标的检测是充满挑战且具有重要意义的工作。针对上述问题, 本文从降低模型复杂度、强化红外特征、提高多尺度目标检测性能 3 个角度出发, 对 YOLOv3 算法进行优化。在特征提取阶段, 采用轻量级的 MobileNet V2 网络代替 Darknet-53 作为骨干网络进行特征提取; 在特征融合阶段, 改进 SPP 空间金字塔结构, 通过实验证明改进后的 SimSPP 能够以更小的参数量达到相同的性能; 在检测阶段取消 anchor 机制, 采用解耦头分别完成分类和回归任务, 采用简化的正负样本分配策略作为样本匹配方案, 提出了针对多尺度目标的轻量级红外目标检测算法。

## 1 LMD-YOLOv3 算法介绍

### 1.1 YOLOv3 算法介绍

YOLOv3(You Only Look Once)是 2018 年 Joseph

Redmon 等人提出的一种端到端的目标检测算法<sup>[14]</sup>, 在 Darknet-19 的基础上, 借鉴了 ResNet 的残差结构, 使用跳跃连接缓解深层网络中出现的梯度消失问题, 形成 Darknet-53 特征提取网络; 特征融合阶段, 采用 FPN 特征金字塔进行加强特征提取, 在提取多特征图的基础上, 将不同大小的特征图进行融合, 用小尺寸特征图检测大尺寸目标, 用大尺寸特征图检测小尺寸目标, 对小目标检测的效果有了明显提升; 采用全卷积结构, 通过  $3 \times 3$  和  $1 \times 1$  卷积进行特征整合和通道数调整, 完成目标预测, 最终形成“Darknet53+FPN+YOLOHead”的模型结构, 网络结构如图 1 所示。

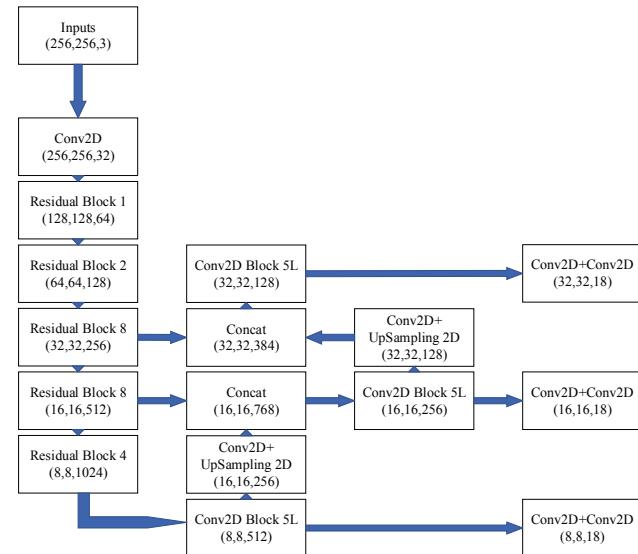


图 1 YOLOv3 网络结构

Fig.1 YOLOv3 network structure

如图 1 所示, 首先在输入端调整图像尺寸为  $256 \times 256$ , 利用 DBL 卷积模块和多个 Residual Block 残差模块组成的 Darknet-53 特征提取网络得到 8 倍、16 倍和 32 倍下采样特征图; 在网络颈部进行浅层特征与高层语义特征的融合, 32 倍下采样特征经过卷积生成小尺寸特征图, 中间尺寸特征图由上采样后的 32 倍特征与输出的 16 倍特征 Concat 后生成, 大尺寸特征图则由 16 倍和 8 倍特征融合生成; 检测头对 3 个不同尺度特征图分别划分不同大小的栅格图, 对边界框的 4 个偏移量、目标置信度和类别进行预测, 产生多个 Bounding Box, 最后通过非极大值抑制和 IoU 交并比计算, 得到预测结果。与其他目标检测算法相比, YOLOv3 模型结构简单, 检测速度快, 且具备一定的多尺度目标检测能力。

### 1.2 LMD-YOLOv3 算法

为提升算法对多尺度红外目标的检测性能, 本文基于 YOLOv3 算法进行改进, 提出了模型大小仅 6.26 M, 浮点运算量仅 2.15 GFLOPs 的 Lightweight

Multiscale infrared target Detection- You Only Look Once (LMD-YOLOv3)，具体结构如图2所示。本文分别从 Backbone、Neck、Head 三个方面对改进进行说明。

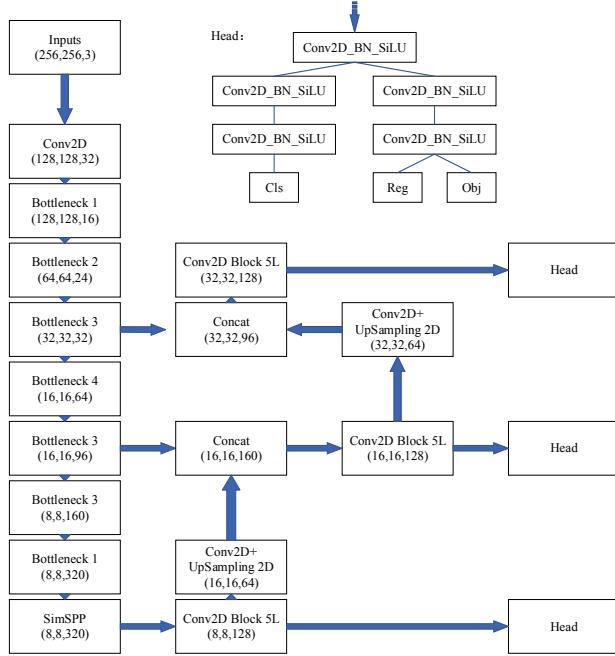


Fig.2 LMD-YOLOv3 network structure

### 1.2.1 Backbone 改进

为尽可能地减少模型运算量，提高算法检测速度，改善 Darknet-53 特征提取网络带来大量参数的问题，节省计算资源，本文对 MobileNet V2<sup>[15]</sup>骨干网络进行优化。MobileNet V2 是 Andrew G. Howard 等于 2018 年在 MobileNet V1<sup>[16]</sup>的基础上提出的改进版本，是目前轻量级神经网络中最具代表性的网络之一。

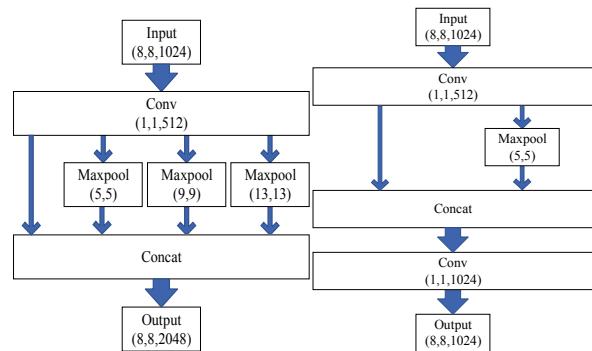
MobileNet 系列网络的主要思想是采用深度可分离卷积来减少运算量及参数量。深度可分离卷积将标准卷积拆分为深度卷积和逐点卷积，首先使用单通道卷积核对每个输入通道进行卷积，然后使用多通道的  $1 \times 1$  卷积核将每个通道的卷积结果进行线性组合，构建新的特征。假设输入为  $D_F \times D_F \times M$  尺寸特征图，输出为  $D_F \times D_F \times N$  尺寸特征图，采用标准卷积核大小为  $D_K \times D_K$ ，则使用标准卷积计算量为  $D_K \times D_K \times M \times D_F \times D_F \times N$ ，第一阶段的深度卷积计算量为  $D_K \times D_K \times M \times D_F \times D_F$ ，第二阶段逐点卷积计算量为  $M \times N \times D_F \times D_F$ ，采用深度可分离卷积的总计算量为  $D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F$ ，即当标准卷积核为  $3 \times 3$  时，深度可分离卷积运算量与标准卷积相比降低了 9 倍。

与 MobileNet V1 相比，MobileNet V2 的核心模块是 Inverted resblock，在  $3 \times 3$  深度卷积之前对通道

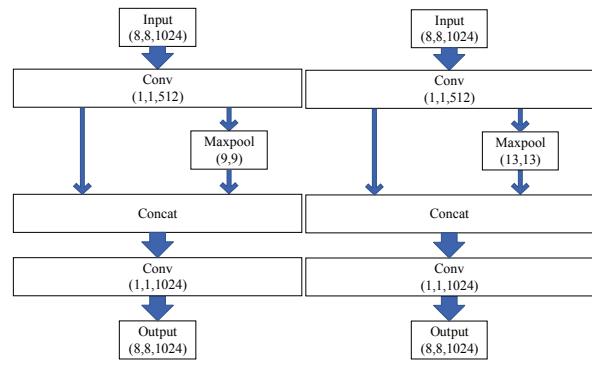
数进行扩张，使网络能够获得更多特征；在  $1 \times 1$  卷积进行通道降维后，用线性变换代替非线性的 ReLU，防止滤除有用信息对特征造成破坏。因此，MobileNet V2 在大幅减少模型参数量的同时，也能一定程度上保证模型的检测性能不受影响。

### 1.2.2 Neck 改进

受 SPPNet 网络<sup>[17]</sup>的启发，我们注意到在 YOLOv3 骨干网后加入 SPP 模块可能有助于提升模型的检测精度。SPP 通过多个不同尺寸内核的最大池化操作，再将所有输出特征图进行 concat 通道拼接实现局部特征和全局特征在特征图级别的融合，丰富特征图的表达能力。因此，我们在参考 YOLOv4<sup>[5]</sup>网络中 SPP 模块结构的基础上对其进行改进，结构如图 3 所示，其中图 3(a)为 YOLOv4 中 SPP 模块结构，图 3(b)为改进的 SPPA 模块，图 3(c)为改进的 SPPB 模块，图 3(d)为改进的 SPPC 模块。



(a) YOLOv4 中 SPP 模块  
(a) SPP in YOLOv4  
(b) 改进的 SPPA 模块  
(b) The improved SPPA



(c) 改进的 SPPB 模块  
(c) The improved SPPB  
(d) 改进的 SPPC 模块  
(d) The improved SPPC

图 3 SPP 模块结构

Fig.3 The structure of SPP

当数据集图片尺寸为  $256 \times 256$  时，32 倍下采样后进入 SPP 模块时宽高为 8，通道数为 1024。由于 SPP 的本质是融合不同尺度感受野信息，空间金字塔池化结构的最大池化核大小决定了该结构能够得到特征图感受野的大小。我们猜想：最大池化核大小越

接近需要池化的特征图尺寸, 空间金字塔池化结构能够得到越接近特征图全局的感受野信息, 特征融合后对于检测性能的提升效果越好。同时, 为减少计算复杂度, 在新的 SPP 模块中, 仅保留最接近需要池化的特征图尺寸的 Maxpool 操作。最后加入卷积模块完成通道数的调整, 保持 SPP 模块输入输出图片的宽高和通道数不变, 使改进后的 SPP 模块能够即插即用到其他检测算法中。

### 1.2.3 Head 改进

受 YOLOX<sup>[7]</sup> 目标检测算法的启发, 我们对 YOLOv3 检测 Head 部分做了 3 点改进:

#### 1) 采用 Anchor Free 机制

为解决目前 Anchor Base 算法检测头复杂度高, 且训练前聚类分析生成的 Anchor 集合对数据集依赖性强, 导致算法泛化性能较差的问题, 我们采用 Anchor Free 的方法, 将 YOLOv3 中对每个位置进行 3 个预测调整为进行一个预测, 每次仅预测网格左上角的宽、高偏移量及预测框的宽、高, 有效降低了检测器参数数量和浮点运算量。另一方面 Anchor Free 的引入使 LMD-YOLOv3 算法能够检测到的目标尺度不再受预设的 Anchor Box 大小限制, 大大提升了算法对多尺度目标的检测性能。

#### 2) 采用解耦头 (Decoupled Head)

由于分类与回归任务之间存在冲突, 采用耦合的检测头会对模型性能造成影响, 而分离的头部又会带来运算复杂度的增加。为平衡模型检测精度和检测速度, 我们首先采用  $1 \times 1$  卷积进行降维, 再通过两个  $3 \times 3$  卷积分别进行分类和回归, 以增加少量参数为代价实现了检测性能的大幅提升。

#### 3) 采用简化的正负样本分配策略

YOLO Head 中仅将采样中心点作为正样本, 导致周围高质量样本被忽略, 正负样本分配 (Optimal Transport Assignment, OTA) 策略考虑全局最优, 具有更强的高质量样本搜索能力。为兼顾算法性能和训练时长, 我们采用简化的正负样本分配策略, 首先计算预测  $g_{gi}$  的成本表示, 然后选择固定区域内成本最低的前  $K$  个预测作为正样本。SimOTA 比 YOLO Head 计算方法更靠近全局最优, 较 OTA 又减少了 25% 的训练耗时。SimOTA 中  $g_{gi}$  和  $p_j$  预测之间成本计算公式如下:

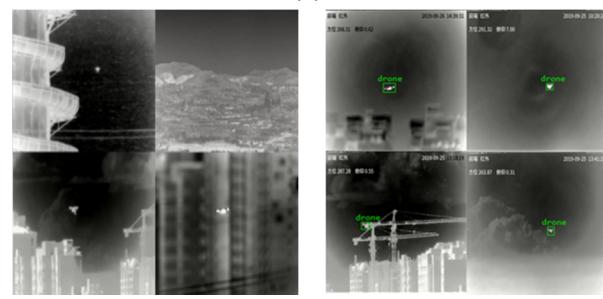
$$C_{ij} = L_{ij}^{\text{cls}} + qL_{ij}^{\text{reg}} \quad (1)$$

式中:  $q$  为平衡系数;  $L_{ij}^{\text{cls}}$  和  $qL_{ij}^{\text{reg}}$  是  $g_{gi}$  和预测  $p_j$  之间的分类损失和回归损失。

## 2 实验结果分析

### 2.1 MTS-UAV 数据集

由于目前缺乏相关公开可用的空中红外多尺度目标数据集, 本实验在 “The 1st Anti-UAV Workshop & Challenge” 比赛提供的红外无人机视频数据集的基础上构建 MTS-UAV 数据集涵盖多种不同复杂场景, 如图 4(a)所示。我们在提供的红外无人机数据集中筛选出符合要求的 36 个红外视频段, 使用 FFmpeg 以  $\text{fps}=5$  的速率将视频提取成帧, 调整图像大小为  $256 \times 256$ , 编写 python 程序进行对应标签的提取, 将标签格式由 json 转化成 xml, 最终得到由 9214 张图片和 9214 个对应的标签文件组成的 VOC 数据集。MTS-UAV 数据集为验证数据集标签的准确性, 对数据集进行可视化处理, 如图 4(b)所示。



(a) MTS-UAV 数据集 (b) 可视化处理后的数据集  
(a) MTS-UAV data set (b) Visualized data set

图 4 MTS-UAV 数据集部分图片

Fig.4 Part of data set MTS-UAV

MTS-UAV 数据集采用 xml 标签文件, 其中包含图片大小 “ $256 \times 256$ ”、标签类别 “drone”, 以及目标位置坐标 “ $xmin$ ”、“ $ymin$ ”、“ $xmax$ ” 和 “ $ymax$ ” 等信息。我们通过提取标签文件中每个目标的上述信息, 对数据集中无人机目标的先验框大小及相关特性进行了统计分析, 得到目标的平均宽高分别为 17.0、11.5, 目标与图片的宽高比分别为 0.0646 和 0.0450。数据集无人机目标大小分布由  $1 \times 1$  到  $64 \times 64$ , 目标尺度差异较大, 适用于复杂场景下对空中多尺度目标检测算法的研究。

### 2.2 实验细节和评估指标

本实验模型训练测试使用硬件平台为 11th Gen Intel ® Core™ i9-11900F CPU, GeForce RTX 2080 Ti 11G GPU; 软件使用 Ubuntu 20.04 系统, python 3.8, Pytorch 1.7.0 深度学习框架; 目标检测算法框架使用商汤科技联合香港中文大学开发的基于 PyTorch 的深度学习目标检测框架 MMDetection。在 MTS-UAV 数据集的基础上, 按照 8:2 的比例划分为训练数据集和测试数据集。实验共进行 epoch 为 100 次的迭代训练,

使用 Adam 梯度下降算法，训练初始化学习率为 0.0001，采用马赛克数据增强。

在检测实时性方面，本实验采用 Parameters、FLOPs、FPS，即模型参数量、浮点运算量和每秒处理帧数 3 项指标评估模型的大小、复杂程度和检测速度。在检测精度方面，采用准确率 (Precision,  $P$ )、召回率 (Recall,  $R$ )、阈值 IOU 为 0.5 的平均精度 (Average Precision, AP) 三项指标评判算法模型的检测性能<sup>[27]</sup>，其中  $P$ 、 $R$  和 AP 值分别表示为：

$$P = \frac{X_{TP}}{X_{TP} + X_{FP}} \quad (2)$$

$$R = \frac{X_{TP}}{X_{TP} + X_{FN}} \quad (3)$$

$$AP = \int_0^1 P(R)dR \quad (4)$$

式(2)和式(3)中： $X_{TP}$  表示正确检测的目标数； $X_{FP}$  表示被检错的目标数； $X_{FN}$  表示未被检出的目标数。

### 2.3 SimSPP 模块实验与结果分析

本节对文章提出的 SimSPP 模块设计的有效性进行验证，实验首先对比了在 YOLOv3 模型特征融合阶段在顶层特征图的处理上加入一个不同的 SPP 模块后，算法对多尺度目标检测性能的变化。我们在 MTS-UAV 数据集上对算法进行训练，并在测试集上进行验证，结果如表 1 所示。

表 1 SimSPP 模块在 YOLOv3 算法上实验结果对比

Table 1 Comparison of experimental results of SimSPP module on YOLOv3 algorithm

Model	Recall	mAP	FPS	FLOPs	Params
YOLOV3	90.70%	87.90%	74	12.41G	61.52M
YOLOV3+	90.90%	88.60%	71	12.57G	64.15M
<b>SPP</b>					
YOLOV3+	90.20%	88.00%	73	12.51G	63.10M
<b>SPPA</b>					
YOLOV3+	90.60%	88.50%	72	12.51G	63.10M
<b>SPPB</b>					
YOLOV3+	90.90%	88.30%	72	12.51G	63.10M
<b>SPPC</b>					

由于分类与回归任务之间存在冲突，采用耦合的检测头会对模型性能造成影响，而分离的头部又会带来运算复杂度的增加。为平衡模型检测精度和检测速度，我们首先采用  $1 \times 1$  卷积进行降维，再通过两个  $3 \times 3$  卷积分别进行分类和回归，以增加少量参数为代价实现了检测性能的大幅提升。表中从目标平均召回率、检测精度、每秒检测帧数、浮点运算量和模型参

数量等 5 个指标对比分析 SimSPP 模块对网络性能的影响。其中，SPP 表示传统具有 3 个不同尺度池化核的特征金字塔模块，SPPA 表示剔除两个较大池化核后，仅保留  $5 \times 5$  池化核的特征金字塔模块，SPPB、SPPC 分别表示仅有  $9 \times 9$  池化核以及  $13 \times 13$  池化核的特征金字塔模块。从表 1 可以看出对于  $8 \times 8$  大小的顶层特征图，采用  $9 \times 9$  池化核的特征金字塔模块效果最佳。与不使用 SPP 模块的 YOLOv3 相比以极小的运算代价，获得了 mAP 提升，与采用传统 3 个池化核的特征金字塔模块算法相比，在保持检测精度的同时，实现了模型参数和运算量的优化，提升了算法检测速度。

同时，为检验 SimSPP 模块在其他算法中的适用性，我们在 MTS-UAV 数据集上对 YOLOX-s 算法进行实验，具体结果如表 2 所示。

YOLOX 算法在特征提取的最后阶段有一个 SPP 模块，我们选取 YOLOX 中较小的 YOLOX-S 模型进行训练，表中 YOLOX-SA 代表特征金字塔模块中仅有  $5 \times 5$  池化核，YOLOX-SB、YOLOX-SC 分别表示特征金字塔模块中仅有  $9 \times 9$  池化核、 $13 \times 13$  池化核。由表 2 可以看出，将 SimSPP 模块应用到 YOLOX-S 算法中，能够在保证检测精度的同时，实现检测速度的提升。对 YOLOX 算法进一步分析发现，该 SPP 模块作用的特征图尺寸也是  $8 \times 8$ ，因此，结合前两组实验，我们发现，在 MTS-UAV 数据集上，特征金字塔结构是有效的，且当池化核大小设置为最接近这一层特征图的大小时 SimSPP 模块性能达到最佳，能够更好地实现检测速度与检测精度的均衡。

表 2 SimSPP 模块在 YOLOX-s 算法上实验结果对比

Table 2 Comparison of experimental results of SimSPP module on YOLOX-s algorithm

	YOLOX-S	YOLOX-SA	YOLOX-SB	YOLOX-SC
mAP	80.90%	80.70%	80.90%	80.60%
FPS	84	90	91	91

### 2.4 LMD-YOLOv3 模型实验与结果分析

#### 2.4.1 消融实验与结果分析

本节通过 LMD-YOLOv3 中不同模块间的消融实验，对比算法中不同模块设计，验证算法中各个模块在空中多尺度红外目标检测问题上的有效性。实验共设 4 组，现将具体实验设置做如下说明：

实验 1：实验使用以 Darknet53 为骨干网 Backbone、以 FPN 为颈部 Neck、以网络第 3 次、第 4 次和第 5 次下采样特征图通过 3 个 YOLO Head 进行预测的 YOLOv3 作为消融实验的基准 Baseline。

实验2：在实验1的基础上，在Neck部分顶层添加simSPP模块。

实验3：在实验2的基础上，在Head部分采用Anchor Free机制、Decoupled Head以及简化的正负样本分配策略。

实验4：在实验3的基础上，在Backbone部分用MobileNet V2代替Darknet53为骨干网进行特征提取。

将完成100个epoch训练后的模型在测试数据集上进行评估测试，损失函数如图5，具体结果如表3所示。

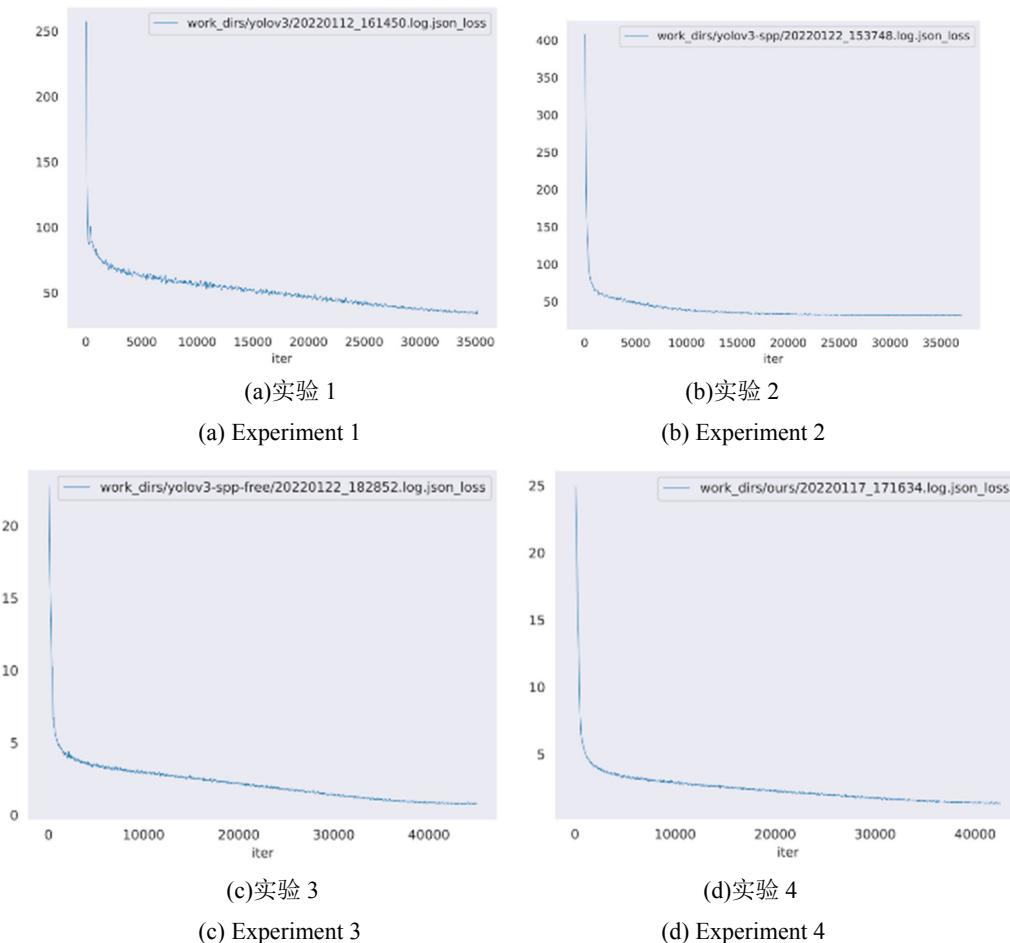


图5 损失函数曲线图

Fig.5 Loss function graph

表3 LMD-YOLOv3消融实验结果对比

Table 3 LMD-YOLOv3 comparison of ablation experiment results

	Recall	mAP	FPS	FLOPs	Params
Experiment 1	90.70%	87.90%	74	12.41G	61.52M
Experiment 2	90.60%	88.50%	72	12.51G	63.10M
Experiment 3	91.20%	90.40%	74	11.19G	46.20M
Experiment 4	91.20%	90.50%	99	2.14G	6.25M

如表3所示，随着组件的增加，算法性能逐步提升。Baseline的实验结果如实验1所示，YOLOv3算法在数据集上达到了87.90%的检测精度和每秒74帧的检测速度，在目前主流检测算法中占据优势地位。表中实验2在FPN的基础上添加了SimSPP模块，对比实验1和实验2，我们发现改进的简化SPP结构仅用0.1G的运算量为代价，增强了特征图的表达能力，

使算法精度提升了0.6%。对比实验2和实验3，证明了Head部分的改进使算法检测精度和速度都有所提升，有助于提高算法对多尺度目标的检测性能。此外，由图5中(b)、(c)对比发现，Decoupled Head将分类与回归任务分离，使其能够加速模型训练收敛。实验4在实验3的基础上采用MobileNet V2轻量级网络，将模型压缩为YOLOv3的1/10，使算法的检测速度有了大幅提升。实验P-R(precision-recall)曲线如图6。

#### 2.4.2 横向对比实验与结果分析

本节在MTS-UAV数据集上对LMD-YOLOv3算法与其他常规目标检测算法的综合性能进行比较，主要以召回率、检测精度、每秒检测帧数、浮点运算量和模型参数量等5个指标进行衡量，结果如表4所示。

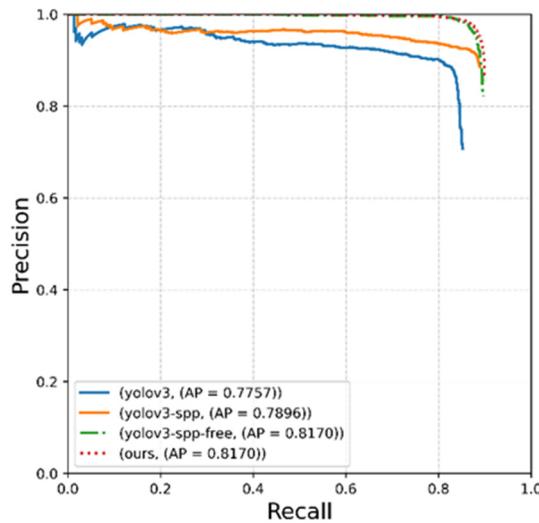


图 6 P-R 曲线图

Fig.6 P-R graph

从实验结果可见,本文所提的算法凭借 90.50% 的检测精度,与其他常规目标检测算法相比具有更好的表现。与 YOLOv3 相比, LMD-YOLOv3 算法在提升了 2.60% 目标检测精度和 99 帧/s 检测速度的同时,将召回率由 90.70% 提升至 91.20%。在目标检测算法中,准确率越高则产生虚警概率越低,召回率越高则发生漏检概率越低。对于空中红外目标的检测而言,召回率指标与准确率同样重要,尤其是在检测入侵目标

时,我们可以接受一定范围内的虚警,但是不希望产生过多的漏检。而许多算法研究中只关注检测精度、速度的提升,忽略了召回率的大幅下降,这是不符合实际应用需求的。

表 4 横向实验结果对比

Table 4 Comparison of horizontal experimental results

	Recall/%	mAP/%	FPS	FLOPs/G	Params/M
YOLOv3	90.70	87.90	74	12.41	61.52
YOLOv4	85.57	86.99	70	11.30	63.9
YOLOX-s	88.20	80.90	84	2.13	8.94
Faster-RCNN	81.60	81.30	49	26.24	41.12
LMD-YOLOv3	91.20	90.50	99	2.14	6.25
(Ours)					

## 2.5 检测结果可视化

为验证 LMD-YOLOv3 算法在实际检测场景中的有效性,我们抽取测试集中部分图像,对算法检测结果进行评估与可视化分析。图 7 展示了 YOLOv3 算法和改进后的 LMD-YOLOv3 算法在 MTS-UAV 数据集的部分测试集上的检测结果以及 LMD-YOLOv3 算法各阶段特征图。



(a) YOLOv3 检测结果可视化  
(a) Visualization of YOLOv3 test results



(b) LMD-YOLOv3 检测结果可视化  
(b) Visualization of LMD-YOLOv3 test results

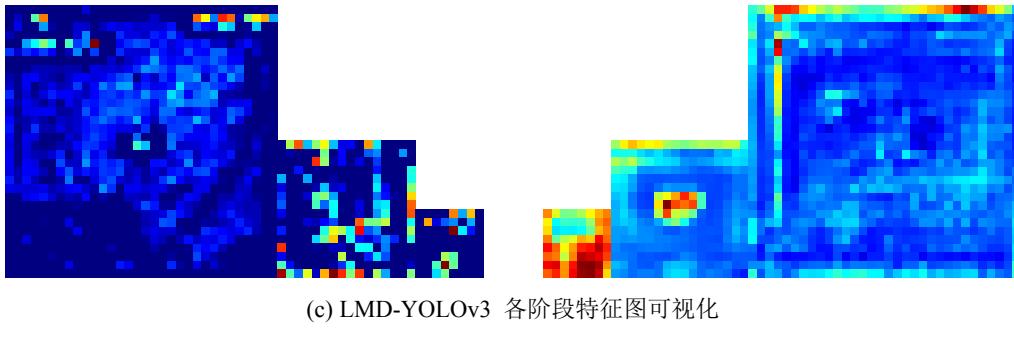


图 7 检测结果可视化

Fig.7 Visualization of test results

由图 7 中(a)、(b)对比可以直观地看出 LMD-YOLOv3 算法与 YOLOv3 算法相比具有更好的检测性能。图 7(c)分别为 3 层 backbone 和 3 层 neck 的输出，可以发现第二层 neck 的输出最能反映出真实目标位置，在下一步研究中我们将重点利用第二层特征进行目标检测。

### 3 结语

本文对红外多尺度目标检测问题进行研究，提出了一种针对多尺度目标的轻量级红外目标检测算法。在完成数据集构建的基础上，对传统 SPP 结构进行改进，仅保留一个 Maxpool 操作，将改进后的 SimSPP 模块添加至模型颈部位置，通过调整池化核大小进行多组实验并发现：当池化核大小设置为最接近这一层特征图的大小时 SimSPP 模块性能达到最佳。在此基础上，结合 Anchor Free 机制、解耦头和简化正负样本分配策略等最新算法优化方法，采用轻量级提取网络完成空中红外目标检测算法设计。通过实验表明，在 RTX2080Ti 设备上 LMD-YOLOv3 检测速度达到 99 帧/s，mAP 达到 90.5%，其他算法相比，本文所提算法对空中红外多尺度目标的检测能力有明显的提升。

### 参考文献：

- [1] 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.
- [2] Girshick R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [3] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. Advances in neural Information Processing Systems, 2015, **28**: 91-99.
- [4] LIU W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//European conference on computer vision., 2016: 21-37.
- [5] Bochkovskiy A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J/OL]. arXiv preprint arXiv, 2020, <https://arxiv.org/abs/2004.10934>
- [6] WANG C Y, Yeh I H, LIAO H Y M. You Only Learn One Representation: Unified Network for Multiple Tasks[J/OL]. arXiv preprint arXiv, 2021, <https://arxiv.org/abs/2105.04206>.
- [7] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021[J/OL]. arXiv preprint arXiv, 2021, <https://arxiv.org/abs/2107.08430>.
- [8] LIU M, DU H, ZHAO Y, et al. Image small target detection based on deep learning with SNR controlled sample generation[M]//Current Trends in Computer Science and Mechanical Automation, 2018: 211-220.
- [9] LIN Liangkui, WANG Shaoyou, TANG Zhongxing. Using deep learning to detect small targets in infrared oversampling images[J]. Journal of Systems Engineering and Electronics, 2018, **29**(5): 947-952.
- [10] ZHAO D, ZHOU H, RANG S, et al. An adaptation of CNN for small target detection in the infrared[C]//IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, 2018: 669-672.
- [11] 谢江荣. 基于深度学习的空中红外目标检测关键技术研究[D]. 上海: 中国科学院大学 (中国科学院上海技术物理研究所), 2019.
- [12] XIE Jiangrong. Research on Key Technologies of Air Infrared Target Detection Based on Deep Learning[D] Shanghai: University of Chinese Academy of Sciences (Shanghai Institute of Technical Physics, Chinese Academy of Sciences), 2019.
- [13] FAN M, TIAN S, LIU K, et al. Infrared small target detection based on region proposal and CNN classifier[J]. Signal, Image and Video Processing, 2021, **15**: 1927-1936.
- [14] 张凯, 刘昊, 杨曦, 等. 基于关键点检测网络的空中红外目标要害部位识别算法[J]. 西北工业大学学报, 2020, **38**(6): 1154-1162.
- [15] ZHANG K, LIU H, YANG X, et al. Key position recognition algorithm of aerial infrared target based on key point detection net-work [J]. Journal of Northwest University of Technology, 2020, **38**(6): 1154-1162
- [16] Redmon J, Farhadi A. Yolov3: An incremental improvement[J/OL]. arXiv preprint arXiv, 2018, <https://arxiv.org/abs/1804.02767>.
- [17] Howard A, Zhmoginov A, CHEN L C, et al. Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation[J/OL]. Computer Science, 2018, <https://arxiv.org/abs/1801.04381v2>.
- [18] Howard A G, ZHUM B, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J/OL]. arXiv preprint arXiv, 2017, <https://arxiv.org/abs/1704.04861>.
- [19] 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.