

# 层聚合网络和跨阶段自适应空间特征融合的小目标检测<sup>①</sup>



于龙昆<sup>2</sup>, 占强波<sup>1</sup>, 沈红<sup>2</sup>, 王子昊<sup>1</sup>

<sup>1</sup>(南昌大学 信息工程学院, 南昌 330031)

<sup>2</sup>(南昌大学 先进制造学院, 南昌 330031)

通信作者: 沈红, E-mail: [shenhong@ncu.edu.cn](mailto:shenhong@ncu.edu.cn)

**摘要:** 传统的目标检测算法存在检测效果不佳及检测效率低等问题, 针对这些问题, 提出了一种基于 YOLOv7 网络改进的小目标检测方法. 该方法在原网络的高效层聚合模块 (efficient layer aggregation network, ELAN) 中添加了更多路径, 且将不同路径中的特征信息有效融合后引入 SKNet 网络, 使得模型更加关注网络中不同尺度大小的特征, 提取出更多有效信息; 同时为了加强小目标对空间信息的感知能力, 设计了一个 eSE 模块连接在 ELAN 末端, 以此构建新的高效层聚合网络模块 (enhanced features efficient layer aggregation network, EF-ELAN), 该模块完整地保留了图像特征信息, 提高了网络的泛化能力. 同时设计了一种 CS-ASFF (cross stage-adaptively spatial feature fusion) 模块来应对小目标检测出现的特征尺度不一致问题, 该模块基于 ASFF 网络和 Nest 连接方式进行改进, 对特征金字塔的每一张图片进行卷积、池化等操作提取权重, 将特征信息作用在某一层上, 同时利用其余特征层来加强网络的特征处理能力. 实验结果表明, 本文提出的算法在 DIOR 数据集和 DOTA 数据集上的平均精准率分别提高了 1.5%、2.1%, 实验结果验证了所提出的算法能够有效地提升小目标的检测效果.

**关键词:** 小目标检测; 特征融合; 注意力机制; 深度学习

引用格式: 于龙昆, 占强波, 沈红, 王子昊. 层聚合网络和跨阶段自适应空间特征融合的小目标检测. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9686.html>

## Small Object Detection Based on Layer Aggregation Network and Cross Stage-adaptive Spatial Feature Fusion

YU Long-Kun<sup>2</sup>, ZHAN Qiang-Bo<sup>1</sup>, SHEN Hong<sup>2</sup>, WANG Zi-Hao<sup>1</sup>

<sup>1</sup>(School of Information Engineering, Nanchang University, Nanchang 330031, China)

<sup>2</sup>(School of Advanced Manufacturing, Nanchang University, Nanchang 330031, China)

**Abstract:** Traditional object detection algorithms often face challenges such as poor detection performance and low detection efficiency. To address these problems, this study proposes a method for detecting small objects based on an improved YOLOv7 network. This method adds more paths to the efficient layer aggregation module (ELAN) of the original network and effectively integrates the feature information from different paths before introducing the selective kernel network (SKNet). This allows the model to pay more attention to features of different scales in the network and extract more useful information. To enhance the model's perception of spatial information for small objects, an eSE module is designed and connected to the end of ELAN, thus forming a new efficient layer aggregation network module (EF-ELAN). This module preserves image feature information more completely and improves the generalization ability of the network. Additionally, a cross stage-adaptively spatial feature fusion module (CS-ASFF) is designed to address the

① 基金项目: 国家自然科学基金 (42365006); 江西省自然科学基金 (20232BAB202040)

收稿时间: 2024-05-12; 修改时间: 2024-06-04; 采用时间: 2024-06-17; csa 在线出版时间: 2024-11-15

issue of inconsistent feature scales in small object detection. This module is improved based on the ASFF network and the Nest connection method. It extracts weights through operations such as convolution and pooling on each image of the feature pyramid, applies the feature information to a specific layer, and utilizes other feature layers to enhance the network's feature processing capabilities. Experimental results show that the proposed algorithm improves the average precision rate by 1.5% and 2.1% on the DIOR and DOTA datasets, respectively, validating its effectiveness in enhancing the detection performance of small objects.

**Key words:** small object detection; feature fusion; attention mechanism; deep learning

随着计算机视觉的不断发展,目标检测成为计算机视觉领域最具有挑战性的任务之一,未来10年内,目标检测也将各个领域受到广泛的探索与研究.在计算机视觉领域中,目标检测承担了许多重要的任务,包括对图像中的目标进行跟踪、分类、检测、分割等一系列任务.主要任务是利用大量的算法模型使计算机能够对所需要检测的目标快速地识别定位,并检测出目标在图像中的精确位置,从而完成对目标的感知认识<sup>[1]</sup>.随着深度学习的快速发展,基于深度学习的目标检测在各种检测任务上的准确率和效率上表现出优秀的性能,并广泛应用于自动驾驶、城市交通、智慧医疗、工业检测、环境监测等领域<sup>[2]</sup>.

随着深度学习的发展,YOLO系列模型经历了巨大变革,YOLOv1<sup>[3]</sup>由Redmon等人在2015年的CVPR会议上提出,YOLOv1是一种高效的目标检测算法,直接从整幅图像中预测边界框和置信度.YOLOv2<sup>[4]</sup>是实时目标检测领域的又一重要突破,由Redmon等人在2016年提出.相比于YOLOv1,YOLOv2在准确率和速度方面都有了明显的提升.YOLOv2应用了Darknet19骨干网络.此外YOLOv2还引入了锚框(anchor box)概念,旨在解决不同目标尺度大小及纵横比例失衡等相关问题.每个锚框预测一组边界框和类别概率,使得模型在训练过程中对不同尺度的物体有着更好的匹配效果.YOLOv3<sup>[5]</sup>引入了多种策略,在特征提取器上,使用了Darknet53提升网络检测效率,同时采用了类别感知的锚框方法,加强了网络对物体的位置预测和定位能力.为了弥补这一不足,研究者们进行了诸多尝试,Qin等人提出了YOLOv3-SPP<sup>[6]</sup>.与YOLOv3不同的是,该模型引入了空间金字塔池化(SPP)模块,旨在提高目标检测的精度.通过使用不同大小的池化窗口对特征图进行池化操作,并将这些池化后的特征图进行拼接,SPP模块能够提取出更丰富的特征信息,从而提

高检测的准确性.因此,YOLOv3-SPP的提出是对YOLO系列算法的进一步优化和改进.紧接着YOLOv4<sup>[7]</sup>、YOLOv5<sup>[8]</sup>的提出进一步提升了检测精度和速度.随着YOLOv7<sup>[9]</sup>网络的诞生,针对遥感图像小目标检测算法存在特征丢失及检测效果不佳等问题,在YOLOv7网络基础上进行改进,构建一种EF-YOLOv7-CS网络.主要内容如下:(1)设计了一种强化特征融合模块(enhanced feature-effective layer aggregation network, EF-ELAN),更多的关注网络中不同尺度大小的特征,同时提取出更多有效信息.(2)设计了一种CS-ASFF(cross stage-adaptively spatial feature fusion)网络,将特征信息作用在某一层上,同时利用其余特征层来加强网络的特征处理能力.

## 1 YOLOv7 算法介绍

YOLOv7模型主要由骨干网络(backbone)、颈部网络(neck)和头部网络(head)组成<sup>[10]</sup>,其网络结构如图1所示.

### 1.1 骨干网络(backbone)

Backbone是YOLOv7的主干特征提取网络,在骨干网络中对输入的图片进行特征提取,提取到的特征可以被称作特征层,是输入图片的特征集合,通常特征提取网络是一个经过预先训练的网络结构,可以提取图像大部分有效特征.Backbone依次经过4个CBS模块和若干个ELAN及MP模块,其中CBS模块包含卷积(Conv)、批量归一化(BatchNorm)和激活函数(如SiLU)等操作,这些模块在网络的不同阶段被用来提取和整合特征信息;MP模块通常进行下采样操作,该模块由两个分支组成,其中一条分支经过一次最大池化和CBS操作,另一条分支经过两次CBS操作,该模块结构如图2所示.ELAN模块是一种高效的层聚合网络,ELAN模块由多个CBS构成.在ELAN模块中,输

入输出特征的大小保持不变,而通道数在开始的两个 CBS 中会有所变化. 后续的 CBS 输入通道和输出通道

则保持一致. 经过最后一个 CBS 后, 输出为所需的通道数. 其模块结构如图 3 所示.

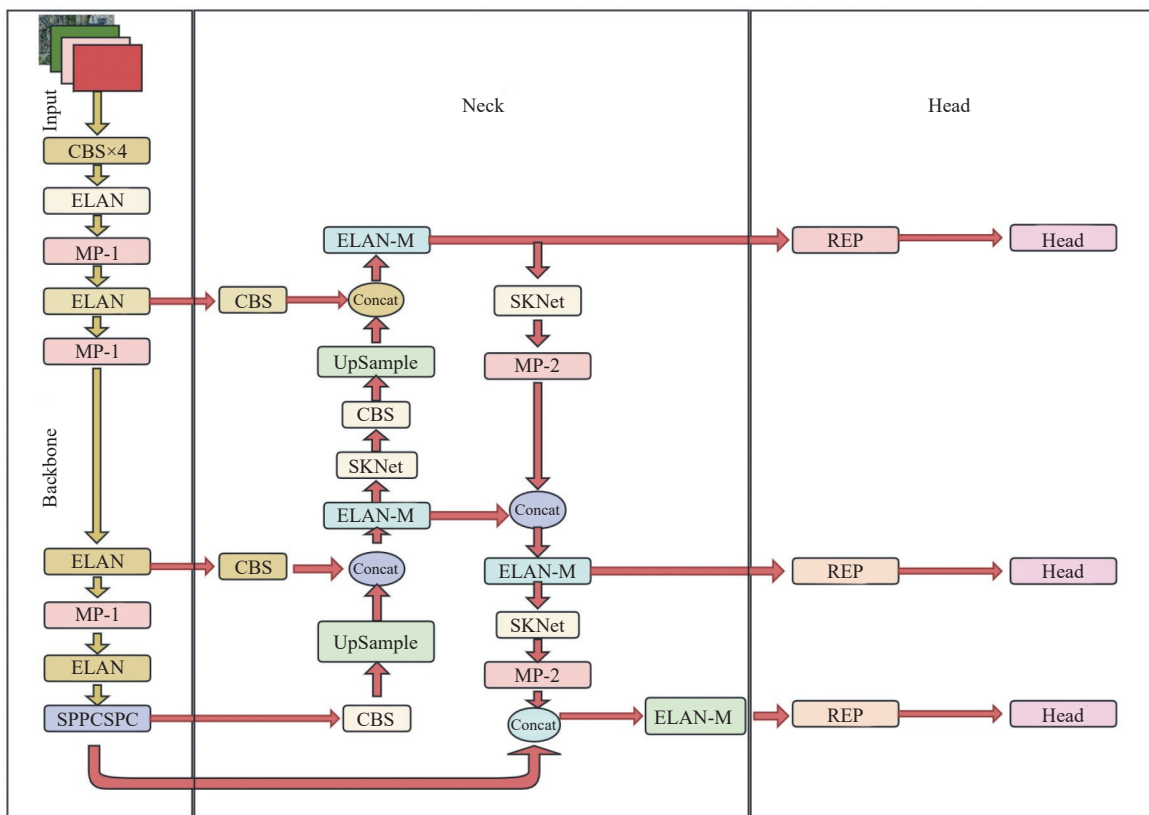


图 1 YOLOv7 结构示意图

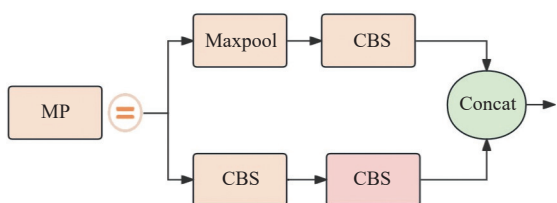


图 2 MP 模块示意图

### 1.2 颈部网络 (neck)

颈部网络负责对特征图进行目标检测, YOLOv7 的颈部网络主要采用了名为 SPP-PAN 的结构, SPP-PAN 结合了空间金字塔池化 (SPP) 和路径聚合网络 (PAN) 的思想. SPP 模块通过引入不同尺度的池化操作, 能够有效地提取和利用多尺度的特征信息, 增强模型对于不同尺寸目标的检测能力<sup>[11]</sup>. 而 PAN 模块则通过自底向上的特征金字塔结构, 将底层特征和高层特征进行融合, 从而进一步提升了特征的丰富性和鲁棒性. Neck 部分主要由 MP-2、SPPCSPC、ELAN-M 模块组成. SPPCSPC 模块能够使网络获得不同的感受野, 区分大

目标和小目标, 并能减少卷积层连续堆叠带来的高计算成本. 其结构如图 4 所示. MP-2 模块仅存在通道数不同的变化; ELAN-M 模块则在 ELAN 基础上改变了拼接方式, 在两组卷积的基础上增加了两组 3×3 卷积.

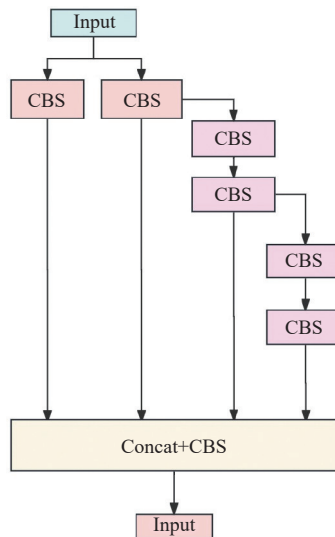


图 3 ELAN 模块示意图

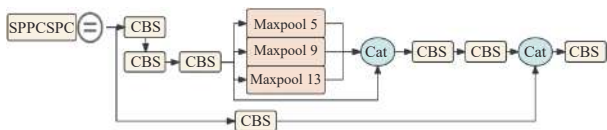


图4 SPPCSPC 模块示意图

### 1.3 头部网络 (head)

Head 网络主要用于预测目标位置和类别. 此外, YOLOv7 的 head 网络还使用了对称卷积来增加网络的感受野, 以提高检测性能. 感受野的扩大有助于模型更好地捕捉图像中的上下文信息, 从而更准确地判断

目标的类别和位置<sup>[12]</sup>.

## 2 EF-YOLOv7-CS 算法介绍

EF-YOLOv7-CS 算法主要由 YOLOv7 改进而来, 本文分别从 YOLOv7 主干网络中的 ELAN 模块和颈部网络中的特征融合金字塔网络结构进行改进, 通过设计强化聚合网络 (EF-ELAN)、跨阶段自适应空间特征融合网络 (CS-ASFF) 对其改进, 得到的 EF-YOLOv7-CS 结构如图 5 所示. 下面分别介绍这两个创新点.

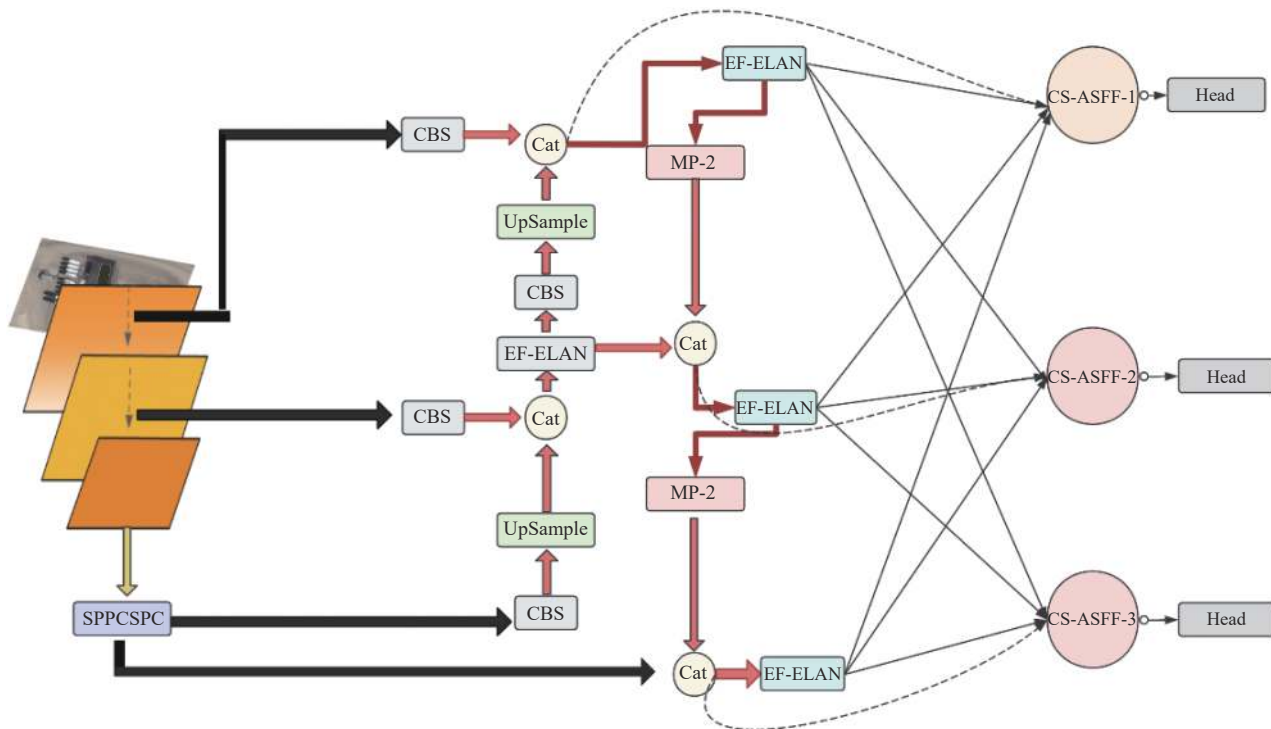


图5 EF-YOLOv7-CS 模块示意图

### 2.1 强化特征高效层聚合网络 (EF-ELAN)

针对处理目标特征及信息表征能力不足等问题, 本文设计了一种 EF-ELAN 结构替换 YOLOv7 模型 neck 中的 ELAN 模块, 以强化金字塔网络对目标的细节感知能力, 该结构使用了多分支堆叠模块, 不同于原 ELAN 模块, EF-ELAN 结构取消了跳跃连接, 旨在处理分支上更多有效信息, 与原 ELAN 结构存在相似之处的是, EF-ELAN 包含了更多的分支模块, 与特征聚合块组合使用. 该结构如图 6 所示, 首先, 第 1 个分支进行一次 CBS 操作输出特征信息, 与第 3 个分支第 1 次 CBS 输出特征进行一次拼接处理, 同时进行 SKNet 操作, 接着第 4 个分支的第 1 次 CBS 输出特征

与第 1 个分支进行拼接后进行一次 SKNet 操作, 每一次连接之后的卷积操作中都使用了相同的填充方式 (same-padding)<sup>[13]</sup> 以保证卷积后的感受野不变, 保留网络的多尺度检测能力, 对于越深层的特征所采用的连续卷积越少. 将 4 个分支进行连接处理后, 引入 eSE<sup>[14]</sup> 模块增强特征, 网络可以充分利用不同层的特征信息, 实现更有效的特征融合, 这种结构有助于提升网络的特征表示能力, 从而提高小目标检测等任务的准确性和鲁棒性. eSE 结构如图 7 所示. eSE 结构由 3 个模块组成, eSE 根据 SENet 中 SE 模块的设计理念得到, 首先将输入特征经过一次全局池化操作, 不同于 SE 模块中使用两次全连接层进行处理, 该模块仅通过一次全

连接层, 主要为了减少通道信息丢失, 最后通过 Sigmoid 层输出检测结果, 其中 SKNet 模块如图 8 所示.

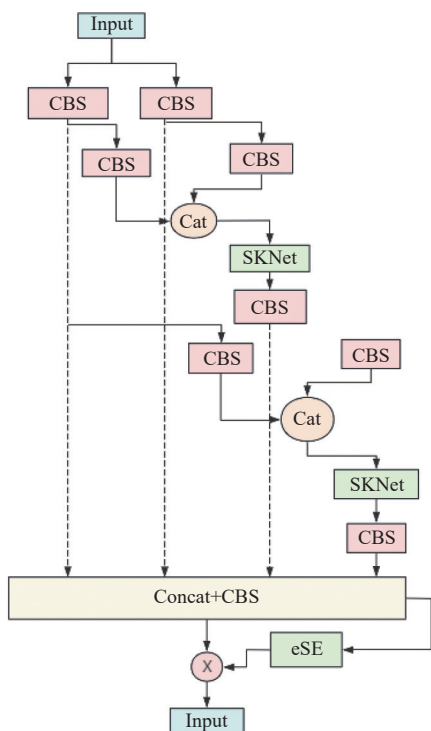


图 6 EF-ELAN 模块示意图

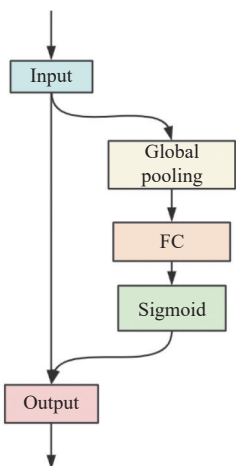


图 7 eSE 模块示意图

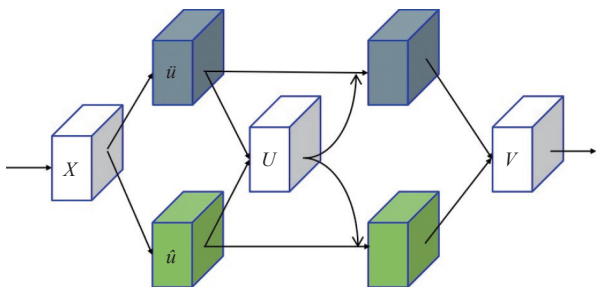


图 8 SKNet 模块示意图

### 2.2 跨阶段自适应空间特征融合 (CS-ASFF)

特征融合是一种有效解决目标检测尺度变化挑战的方法. 但是当目标尺度变化范围过大时, 其性能可能会受到影响. 因此提出了一种新的数据驱动的金字塔特征融合方式, 自适应空间特征融合 (ASFF)<sup>[15]</sup>. 它学习了在空间上过滤冲突信息以抑制梯度反向传播的时不一致的方法, 从而改善了特征的比例不变性, 并且推理开销降低. ASFF 自适应空间特征融合机制对于每个级别的深度特征都有很好的利用, 所有其他级别的特征都将调整为相同的形状并且根据所学习的权重图进行空间融合, ASFF 模型如图 9 所示, 图中展示了来自不同尺度大小特征层相互融合的过程. 在此之前有一种基于 Nest<sup>[16]</sup>连接和空间模型, 在提出的融合策略中, 开发了空间注意模型, 它们描述了每个空间位置以及每个具有深层特征的通道的重要性. Nest 连接体系结构充分利用了深层特征, 并保留了编码器网络提取于来自不同比例特征的有效信息, 对于多尺度深度特征的融合, 用于保留来自先前层的更多信息. 但是, 当在网络体系结构中使用跳跃连接时, 语义间隙会导致意外结果. Nest 连接的框架如图 10 所示, 通过 Nest 连接, 可以约束语义间隙的影响, 并保留更多的信息以获得更好的结果. Nest 连接策略并非针对深度功能而设计, 并且在小目标中检测性能较低, 一是不能提取多尺度特征, 深度特征没有得到充分利用, 二是融合策略未精心设计融合深度特征. 因此本文提出了一种跨阶段自适应空间特征融合机制 CS-ASFF, 通过利用 Nest 连接方式中的跨阶段连接方式结合 ASFF 理念设计得到. CS-ASFF 模型如图 11 所示.

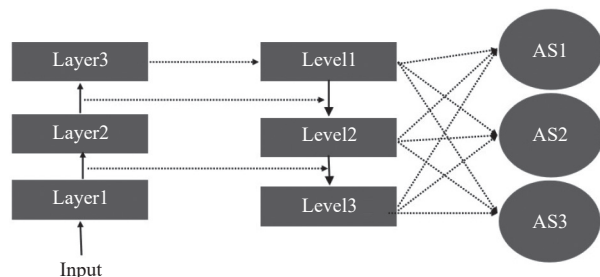


图 9 ASFF 模块示意图

以图 11 中 CS-ASFF-2 模块为例, 连接了来自 C1、C2、C3 这 3 个不同层和一个 C2'层的不同特征, 其中 C2'是输出 C 经过跳跃连接直接输出得到, 将 4 个不同层的特征信息乘上相应的权重并求和, 得到了 CS-ASFF-2 模块, 其中 C1 通过一次 1×1 卷积将通道数调整到与

C2 相同, 并利用插值法将图像分辨率调整到与 C2 相同大小, C3 通过一个步长为 2 的 3×3 最大池化层, 接着进行一次步长为 2 的 3×3 卷积将通道数调整至与 C2 通道数相同, C2' 通过一次 1×1 卷积将通道数调整至与 C2 相同. 学习权重流程如图 12 所示, 同时 4 个参数需要满足以下两个条件: 一是不同层到 L 层级的权重参数和为 1, 二是每个权重参数对应的权重范围在 0-1 之间, 如式 (1) 所示:

$$\alpha'_{ij} + \beta'_{ij} + \gamma'_{ij} + \mu'_{ij} = 1, \alpha'_{ij}, \beta'_{ij}, \gamma'_{ij}, \mu'_{ij} \in [0, 1] \quad (1)$$

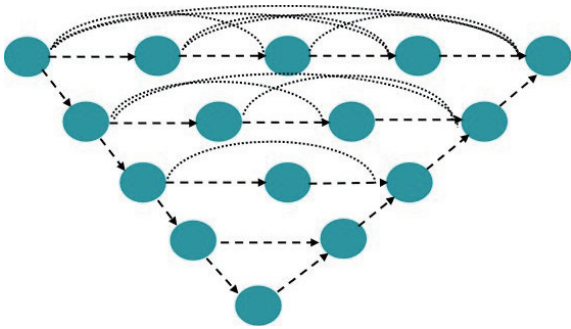


图 10 Nest 连接结构示意图

文中以 C2 层的 feature map 上的点 (i, j) 为例, 其梯度如式 (2) 所示:

$$\frac{\partial \mathcal{L}}{\partial C_{ij}^1} = \alpha_{ij}^1 \cdot \frac{\partial \mathcal{L}}{\partial y_{ij}^1} + \alpha_{ij}^2 \cdot \frac{\partial \mathcal{L}}{\partial y_{ij}^2} + \alpha_{ij}^3 \cdot \frac{\partial \mathcal{L}}{\partial y_{ij}^3} + \alpha_{ij}^4 \cdot \frac{\partial \mathcal{L}}{\partial y_{ij}^4} \quad (2)$$

在本文改进的 YOLOv7 模型 head 网络中, 添加 CS-ASFF 跨阶段自适应空间融合网络, 该模型对不同特征层输出特征进行筛选, 通过学习权重来选择不同层级特征进行融合, 将融合后的特征进行尺度调整, 以适应目标的不同尺度, 这种结构在处理小目标尺度大小差异上效果显著<sup>[17]</sup>.

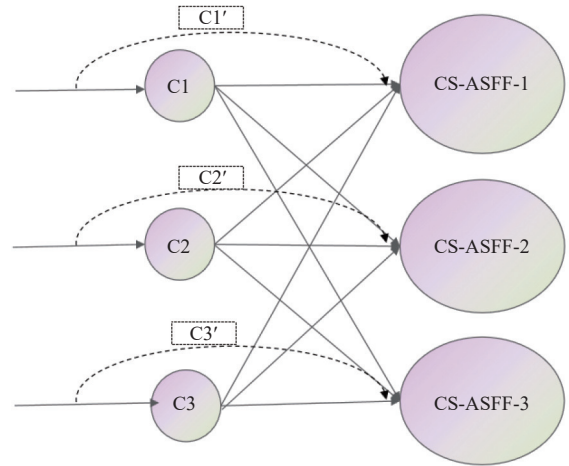


图 11 CS-ASFF 连接结构示意图

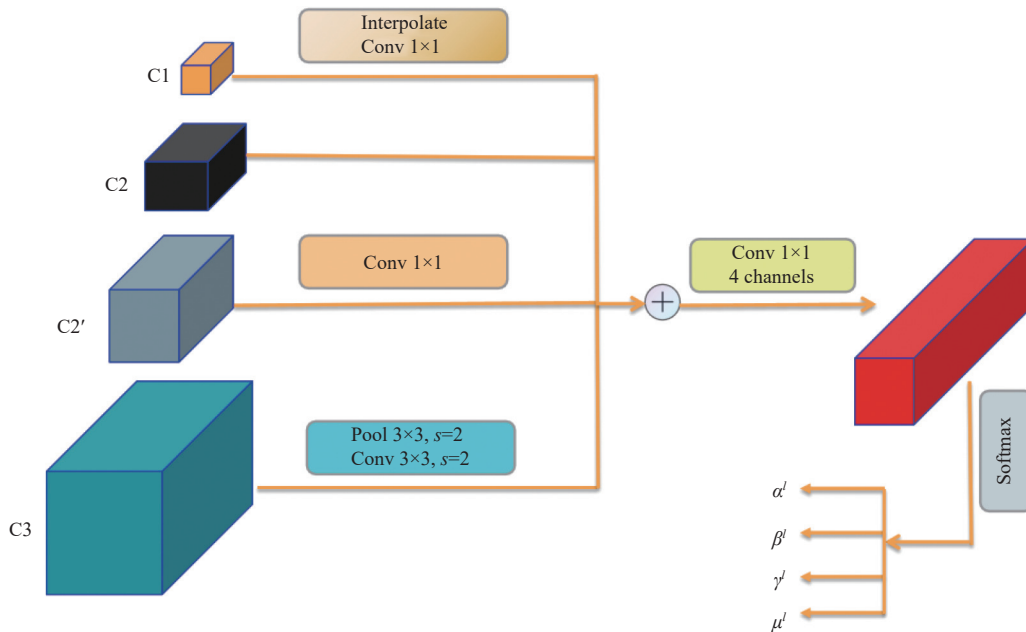


图 12 CS-ASFF 学习权重流程示意图

### 3 实验结果与分析

#### 3.1 实验环境配置

实验以 PyTorch 作为深度学习框架, 显卡为 NVIDIA

GeForce GTX 4090; 运行内存为 24 GB; Python 版本为 3.8. 通过使用 PyTorch 1.11.0 来实现所需训练模型, 初始 Lr 设置为 0.001, 输入图像的大小设置为 640×640,

Batchsize 设置为 16.

### 3.2 实验数据集

本文所使用的数据集是 DIOR 和 DOTA 数据集, DIOR 由 23 463 张遥感图像组成, DIOR 是一个用于光学遥感图像目标检测的大规模基准数据集. 涵盖 20 个对象类. 其中训练集包含 21 117 张图片, 验证集包含 2346 张图片. 其中 DOTA 是一个用于目标检测的大规模数据集, 其包含 2 806 张无人机图像, 15 个对象类, 其中训练集包含 2 586 张图片, 验证集包含 220 张图片.

### 3.3 实验评价指标

当正确地预测了一个目标, 即检测出的目标框与真实目标框有足够的重叠, 这个预测被认为是正确的, 即为 (true positive,  $TP$ ), 当错误地预测了一个目标, 即检测出的目标框实际上并没有对应任何真实的目标, 或者与真实目标的重叠度不够, 这个预测被认为是错误的, 即为 (false positive,  $FP$ ). 精度 (precision,  $P$ ) 指正确分类的正样本数与预测为正样本的总数比值, 其公式为:

$$P = \frac{TP}{TP + FP} \quad (3)$$

召回率 (recall,  $R$ ) 是正确分类的正样本数与实际正样本数的比值, 定义如下:

$$R = \frac{TP}{TP + FN} \quad (4)$$

平均精度  $AP$ , 使用积分的方式来计算  $PR$  曲线与坐标轴围成的面积, 定义如下:

$$AP = \int_0^1 P(R) dR \quad (5)$$

其中,  $AP \in (0, 1)$ ,  $AP$  值越大说明其目标检测性能越好.

$mAP$  就是对所有类的  $AP$  值求平均, 其定义如下:

$$mAP = \frac{1}{n} \sum_{u=1}^n AP_u \quad (6)$$

其中,  $AP_u$  表示第  $u$  类目标的  $AP$ .

### 3.4 实验结果与分析

为了验证本文所提出的 EF-ELAN 层聚合网络设计的有效性, 构建 YOLOv7-EF-ELAN 模型, 与原 YOLOv7 模型进行比对, 其实验结果如表 1 所示.

可以看出本文所提出的 YOLOv7-EF-ELAN 与原 YOLOv7 模型相比网络性能均有小幅度提升, 改进后的模型精准率提高了 1.2%, 召回率提升了 2.1%,  $mAP@.5$

提升了 0.9%. 与基线网络相比, YOLOv7-EF-ELAN 网络在模型参数量上减少了 4%, 浮点运算数降低了 4.8%, 其中每秒传输帧数提升了 7.6%, 改进后的网络模型在轻量化上做出了提升, 这使得实验模型在小目标检测的结果上各项数据指标得到了提升, 其中网络模型参数量减少使得算法模型检测速率得到提升; FPS 得到小幅度提升, 模型在实时性方面得到了优化, 算法能够以更快的速度处理数据, 提高了模型的实时响应能力. 实验结果证明, 设计的 EF-ELAN 层聚合网络在检测小目标图像上, 增强了对图像中重要区域的关注, 实现了对重要特征的增强和对不相关特征的抑制.

表 1 实验对比结果

Method	P R $mAP@.5$			Parameters	FLOPs	FPS
	(%)	(%)	(%)			
YOLOv7	91.7	83.5	87.2	36.9M	104.7G	52
YOLOv7-EF-ELAN	92.9	85.6	88.1	35.4M	99.6G	56
Improvement	+1.2	+2.1	+0.9	-4%	-4.8%	+7.6%

为了验证本文所提出的 CS-ASFF 跨阶段自适应空间特征融合网络的有效性, 构建 YOLOv7-CS-ASFF 网络模型, 与原 YOLOv7 模型进行比对, 参数设置相同, 实验使用 DIOR 测试集验证, 其实验结果如表 2 所示.

表 2 模型改进前后的实验对比结果

Method	P R $mAP@.5$			Parameters	FLOPs	FPS
	(%)	(%)	(%)			
YOLOv7	90.7	83.4	87.1	36.9M	104.7G	52
YOLOv7-CS-ASFF	92.9	86.6	89.4	34.2M	101.2G	60
Improvements	+2.2	+3.2	+2.3	-7.3%	-3.3%	+15.4%

从表 2 中可见, 本文所提出的 YOLOv7-CS-ASFF 与原 YOLOv7 模型相比精准率提高了 2.2%, 召回率提升了 3.2%,  $mAP@.5$  提升了 2.3%, 这是因为设计的 CS-ASFF 模型使网络直接学习如何在其他层次对特征进行空间过滤, 只保留有用的信息, 这种自适应的融合方式可以在每个空间位置将不同层次的特征进行融合, 从而提高小目标检测的准确性和稳定性. 设计的算法模型在模型参数量上和 FPS 帧率上有较大优化, 分别提升了 7.3% 和 15.4%, 实验结果验证了 CS-ASFF 模型进一步优化了 YOLOv7 网络, 特征金字塔网络模块更加轻量化.

为了验证改进后模型针对不同数据集的泛化能力, 在 DOTA 数据集上做进一步验证, 其中实验参数环境保持不变, 实验结果如表 3 所示.

表3 模型在DOTA数据集上实验对比结果

Method	Parameters	FLOPs	P (%)	R (%)	$mAP@.5$ (%)	FPS
YOLOv7	36.9M	104.7G	48.3	40.1	55.6	35
Ours	35.2M	99.3G	51.6	44.5	47.7	41
Improvements	-4.6%	-5.2%	+3.3	+4.4	+2.1	+13%

从表3中可以得到,模型在各项评价指标上都有着明显的提升,改进后的算法模型在帧数上得到了巨大提升,提升了13个百分点,该模型在针对不同数据集仍然达到良好的性能提升,验证了EF-ELAN模块和CS-ASFF模块融合对小目标检测速率上的有效提升,融合模块更多关注地对不同尺度特征表达的泛化能力,选择性地关注小目标需要关注的区域,其中模型参数量减少了4.6%,模型处理图像能力更加迅速.在检测精度方面,模型的精准率和召回率分别提升了3.3%和4.4%,这说明模型在精准率和召回率上取得了更好的平衡,在检测能力上能够更广泛的覆盖真正例,同时模型取得了更好的泛化能力,能够在新的数据集上保持稳定的性能,实验结果充分验证了该模型优化策略的有效性.

为了进一步验证结构的有效性,引入不同主流算法模型进行对比实验,探究改进后的融合模型是否对小目标检测具有可靠的检测性能,在DIOR测试集上进行训练,实验结果如表4所示.由表4结果可知,展示了不同算法在数据集上的检测结果.本节提出的算法与主流目标检测算法相比,在DIOR测试集上的检测精度均有提升,与单阶段检测器YOLOv5-n, YOLOv7-

E6相比,改进后的模型 $mAP@.5$ 分别提升了8.7%和4.2%, $mAP@.5:.95$ 均提升了5%以上.该算法在提升检测精度的同时,检测速度并没有下降多少,检测速度FPS比YOLOv7-E6减少了8,但是仍达到了34.与传统的SSD、DSSD、Faster R-CNN检测器相比,其检测精度大幅度提升, $mAP@.5$ 和 $mAP@.5:.95$ 均提升了20%以上,检测速度比较DSSD有着明显的提高,与MobileNet224和EfficientDet-D0检测器相比其检测速度明显下降,主要是网络中参数计算量大,网络深度增加造成,但与之相比其检测精度都存在提升.

表4 主流算法在DIOR数据集上的检测结果

Method	Size	Parameters (M)	FLOPs (G)	FPS	$mAP@.5$ (%)	$mAP@.5:.95$ (%)
SSD <sup>[18]</sup>	512×512	13.2	31.5	40	68.7	42.8
DSSD <sup>[19]</sup>	321×321	11.5	21.4	25	69.8	41.7
Faster R-CNN <sup>[20]</sup>	600×1000	16.2	29.2	35	65.1	40.2
YOLOv5-n	640×640	21.2	40.6	38	82.4	67.3
MobileNet224 <sup>[21]</sup>	224×224	4.2	12.6	78	82.3	69.5
EfficientDet-D0 <sup>[22]</sup>	512×512	3.9	2.5	69	80.3	57.6
YOLOv7-E6	640×640	37.2	125.2	42	86.9	68.6
Ours	640×640	33.4	110.5	34	91.1	73.6

为了验证该算法对遥感图像小目标检测效果的有效性,在DIOR测试集和DOTA测试集上分别进行实验验证,比较算法改进前后遥感图像小目标检测的可视化结果对比.图13和图14分别展示了遥感图像小目标检测在DIOR和DOTA数据集上的可视化结果示意图.

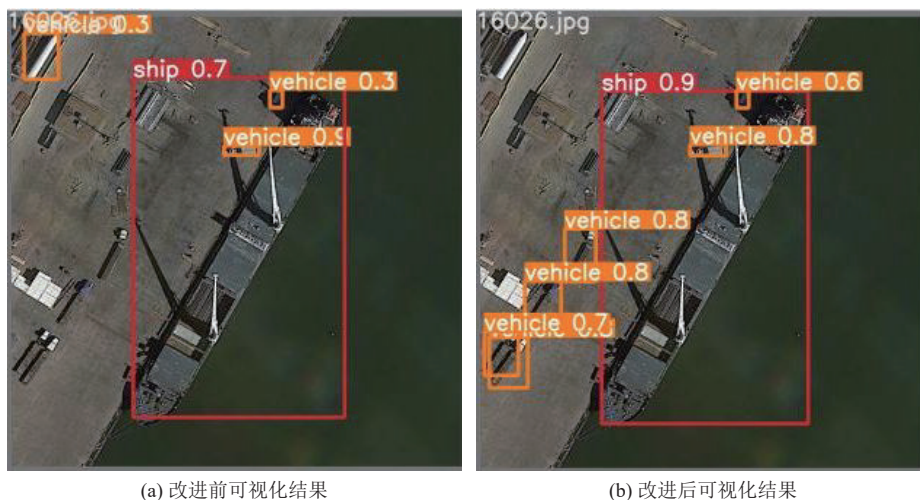


图13 DIOR数据集上的可视化结果比较



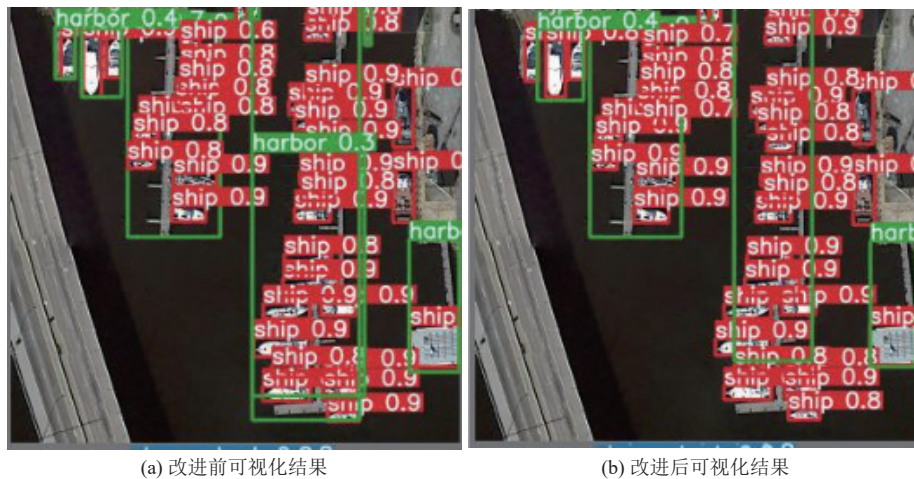


图 14 DOTA 数据集上的可视化结果比较

如图 13 所示, (a) 是原算法检测结果, (b) 是改进后融合算法检测结果. 从图 13 中可以看出改进后的算法检测效果有着显著提升, 不仅体现在目标检测准确率上, 其漏检、误检率也明显下降, 图 13(a) 中存在明显的漏检现象, 原算法未能将图中左下段汽车检测出来, 且左上角存在将仓库错检为汽车的情况, 而改进后的算法, 不仅能够检测出图像中的汽车, 其错检、漏检现象大幅度下降, 对小目标的准确率高, 目标识别定位也较为精确. 如图 14 所示, 针对不同的数据集, 改进后的模型误检率降低, 针对物体重叠, 背景噪声较大区域置信度有着小幅提升.

### 3.5 消融实验

为了探究目标检测中的小目标不敏感问题. 本节将对设计的 CS-ASFF 跨阶段自适应空间融合特征网络、EF-ELAN 层聚合网络模型融合实验分析, 本节在 DIOR 数据集上进行了一系列的对比实验, 用于分析单个 CS-ASFF 网络、EF-ELAN 层聚合网络以及两个模型融合设计后的网络对遥感图像小目标检测精度的影响. 分别在基线模型中进行消融实验, 结果如表 5 所示. 通过引入 EF-ELAN、CS-ASFF 结构, 与基线网络相比, 其检测精度均有提升. 在基线网络中添加 EF-ELAN 聚合结构,  $mAP@.5$  提升了 0.9%, 添加 CS-ASFF 方法后, 网络  $mAP@.5$  提升了 1.3%; 将 EF-ELAN 和 CS-ASFF 同时融合在基线网络中, 改进后的 EF-YOLOv7-CS 模型相比基线模型其  $mAP@.5$  提升了 1.9%, 其中 EF-YOLOv7-CS 网络在检测性能上略微高于 EF-YOLOv7 网络,  $mAP@.5$  提升了 0.4%, 这说明与单一网络 EF-YOLOv7 相比, 通过添加 CS-ASFF 可以自适应的融合

多尺度特征, 通过学习权重来自动调整不同层级特征之间的信息, 提升检测性能. 其中 EF-YOLOv7-CS 网络在检测性能上也略微高于 CS-YOLOv7 网络,  $mAP@.5$  提升了 0.6%, 这说明 EF-ELAN 网络可以充分利用不同层的特征信息, 实现更有效的特征融合, 这种结构有助于提升网络的特征表示能力, 该实验证明, EF-ELAN 和 CS-ASFF 可以有效融合, 且其检测性能比添加单一网络均有小幅提升证明了 EF-ELAN 模块和 CS-ASFF 模块在遥感小目标检测上的优异性, 可以有效提升遥感图像小目标的检测性能.

表 5 探究不同改进点的消融实验结果

YOLOv7 (Baseline)	EF-ELAN	CS-ASFF	$mAP@.5$ (%)
√	—	—	89.2
√	√	—	90.1
√	—	√	90.5
√	√	√	91.1

## 4 结语

为了解决遥感图像小目标检测任务中可利用特征较少、定位精度偏差、特征易丢失等难点, 本文提出了一种基于层聚合网络 EF-ELAN 和跨阶段自适应空间特征融合 CS-ASFF 的算法模型, 首先将设计的 EF-ELAN 替换原 neck 网络中特征融合部分的 ELAN 模块, 该模块捕捉不同尺度特征层语义信息, 强化金字塔网络对目标的细节感知能力, 更好地应对遥感图像中小目标特征及形状信息表征能力不足的问题, 接着, 设计了一种 CS-ASFF 跨阶段自适应空间特征融合算法, 该算法通过加强特征提取, 增强局部特征的代表能力来融合更多有效信息, 该结构不仅自适应地融合多尺

度特征,而且通过学习权重来自动调整不同层级特征之间的融合信息,有效加强了网络对局部信息的感知能力.最后通过实验验证设计算法的有效性,同时通过可视化分析,得到算法改进前后的遥感图像小目标检测结果,通过对设计算法的消融实验,算法在 DIOR 数据集和 DOTA 数据集上有着显著的检测速度和检测精确率,进一步说明本文提出的改进算法在遥感图像小目标检测上优于目前主流的目标检测算法.

### 参考文献

- 1 卜荟力.复杂场景下基于改进 YOLO 算法的遥感图像目标检测.现代信息科技,2022,6(10):87–90.[doi:10.19850/j.cnki.2096-4706.2022.10.021]
- 2 刘小波,刘鹏,蔡之华,等.基于深度学习的光学遥感图像目标检测研究进展.自动化学报,2021,47(9):2078–2089.[doi:10.16383/j.aas.c190455]
- 3 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788.
- 4 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6517–6525.
- 5 Redmon J, Farhadi A. YOLOv3: An incremental improvement. Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: CVPR, 2018. 7794–7803.
- 6 Qin XY, Li N, Weng C, *et al.* Simple attention module based speaker verification with iterative noisy label detection. Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing. Singapore: IEEE, 2022. 6722–6726.
- 7 Roy AM, Bose R, Bhaduri J. A fast accurate fine-grain object detection model based on YOLOv4 deep neural network. Neural Computing and Applications, 2022, 34(5): 3895–3921. [doi:10.1007/s00521-021-06651-x]
- 8 Zhu XK, Lyu SC, Wang X, *et al.* TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal: IEEE, 2021. 2778–2788.
- 9 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 7464–7475.
- 10 李建东,王岩,曲海成.基于 Transformer 的跨尺度交互学习伪装目标检测.计算机系统应用,2024,33(2):115–124.[doi:10.15888/j.cnki.csa.009395]
- 11 尹宏鹏,陈波,柴毅,等.基于视觉的目标检测与跟踪综述.自动化学报,2016,42(10):1466–1489.[doi:10.16383/j.aas.2016.c150823]
- 12 白明雄,张超,陈棋,等.基于无人机可见光遥感的单木树高提取方法研究.林业资源管理,2021(1):164–172.[doi:10.13466/j.cnki.lyzygl.2021.01.021]
- 13 谢江荣.基于深度学习的空中红外目标检测关键技术研究[博士学位论文].上海:中国科学院大学(中国科学院上海技术物理研究所),2019.
- 14 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.
- 15 陆德彪,郭子明,蔡伯根,等.基于深度数据的车辆目标检测与跟踪方法.交通运输系统工程与信息,2018,18(3):55–62.[doi:10.16097/j.cnki.1009-6744.2018.03.009]
- 16 Li H, Wu XJ, Durrani T. NestFuse: An infrared and visible image fusion architecture based on nest connection and spatial/channel attention models. IEEE Transactions on Instrumentation and Measurement, 2020, 69(12): 9645–9656. [doi:10.1109/TIM.2020.3005230]
- 17 戚玲珑,高建瓴.基于改进 YOLOv7 的小目标检测.计算机工程,2023,49(1):41–48.[doi:10.19678/j.issn.1000-3428.0065942]
- 18 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.
- 19 Huang G, Liu Z, van der Maaten L, *et al.* Densely connected convolutional networks. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 2261–2269.
- 20 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2015. 91–99.
- 21 Ran XQ, Yan TF, Cai TL. MobileNet for differential constellation trace figure. Proceedings of the 13th International Conference on Communication Software and Networks. Chongqing: IEEE, 2021. 168–172.
- 22 Tan MX, Le QV. EfficientNet: Rethinking model scaling for convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning. Long Beach: ICML, 2019. 6105–6114.

(校对责编:孙君艳)