基于改进 YOLOv5 的路面坑洼检测研究

陈佳宁[1] 郝 娟[1] 刘晓群[1]
CHEN Jianing HAO Juan LIU Xiaoqun

摘 要 路面坑洼是常见的路面病害之一，危害极大，因此路面坑洼的识别与检测对人们的日常生活十分重要。基于 YOLOv5 算法，提出了一种改进的路面坑洼检测方法。首先，使用轻量型网络 MobilNetV3 替换原 YOLOv5 主干网络；其次，在模型的 head 部分加入三个无额外参数的 SimAM 注意力模块，以提取更加重要的通道特征，剔除无关通道信息，使模型捕获更多关键信息；最后，将原模型的损失函数更换为精度与收敛能力更加优秀的 Alpha-IoU 损失函数。为验证改进模型的有效性，在公开数据集 Annotated Potholes Image Dataset 上进行对比实验，实验结果表明，改进后的 YOLOv5 模型体积缩短了 48%，平均检测精度与检测速度分别提高了 14.8% 和 6.6%。改进算法能够快速有效地检测出路面坑洼，具有一定的应用价值。

关键词 YOLOv5；路面检测；轻量化；损失函数；注意力模块

doi：10.3969/j.issn.1672-9528.2024.05.032

0 引言

随着城市交通的不断发展，道路的安全性和舒适性成为人们关注的焦点。路面坑洼作为一种常见的道路缺陷，不仅影响了车辆的行驶体验，还可能导致交通事故和车辆损坏 [1]。因此，准确、高效地检测和识别路面坑洼变得至关重要。在过去的几十年里，许多传统的路面坑洼检测方法都依赖于人工巡查，这不仅费时费力，而且容易出现漏检和误检的情况。目标检测是计算机视觉领域的核心问题之一，其任务是在图像中精准地识别出所有被关注的目标物体，并进一步确定它们的类别和位置。鉴于不同类别物体在外观、形状和姿态上呈现出显著差异，同时受到光照条件、物体遮挡等多种因素的影响，目标检测问题一直是计算机视觉领域中最具挑战性和复杂性的问题之一 [2]。随着计算机视觉和深度学习技术的迅猛发展，自动化的路面坑洼检测方法逐渐受到关注。

近年来，深度学习技术在计算机视觉领域取得了举世瞩目的进展与成功。目前的目标检测算法主要分为两大流派：one-stage 算法和 two-stage 算法 [3]。其中，two-stage 算法中 Faster R-CNN[4] 在目标检测任务中表现出色，但其计算复杂度较高，不适合实时应用。one-stage 算法中 SSD[5] 是另一种流行的目标检测算法，但其对小目标的检测效果也不尽如人意。在复杂光照和天气条件下，SSD 的鲁棒性可能受到影响。本文提到的 one-stage 目标检测模型 YOLOv5 因其高效的实时性能和准确性而备受关注。然而，现有的 YOLOv5 模型在路面坑洼检测方面仍存在一些挑战，例如对小尺寸坑洼的检测不够精准，以及在复杂光照和天气条件下的鲁棒性不足，模型参数与计算量提升，导致训练速度变慢、精度减低。

基于以上不足，本文将 YOLOv5 的主干模型换为轻量级网络 MobilNetV3，在损失少量精度的前提下，极大降低参数量、计算量以及推理时间。为了在不引入额外参数的前提下提高模型检测精度，本文提出将 SimAM 注意力模块整合至网络架构中，由于 SimAM 注意力模块的主要操作均基于所设定的能量函数选择机制，从而减少了不必要的结构调整，提高了整体的效率与灵活性。最后将原模型的损失函数替换为更加适合小数据集的 Alpha-EIOU 损失函数来提升检测精度。

1 YOLOv5 网络介绍

YOLOv5 是一种一阶段目标检测算法，在检测过程中不产生图像备选框，而是一步到位直接生成检测目标的坐标和类别。它在目标检测领域扮演着重要角色。为了适应不同的使用场景，YOLOv5 共有四个版本的目标检测网络，分别是：YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x，各个网络版本的模型大致相同，区别在于对应网络的深度与特征图宽度依次递增，具体的网络模型如图 1 所示。为了实现轻量化模型，在本文的改进中，选择模型体积最小的 YOLOv5s。

1. 河北建筑工程学院 河北张家口 075000

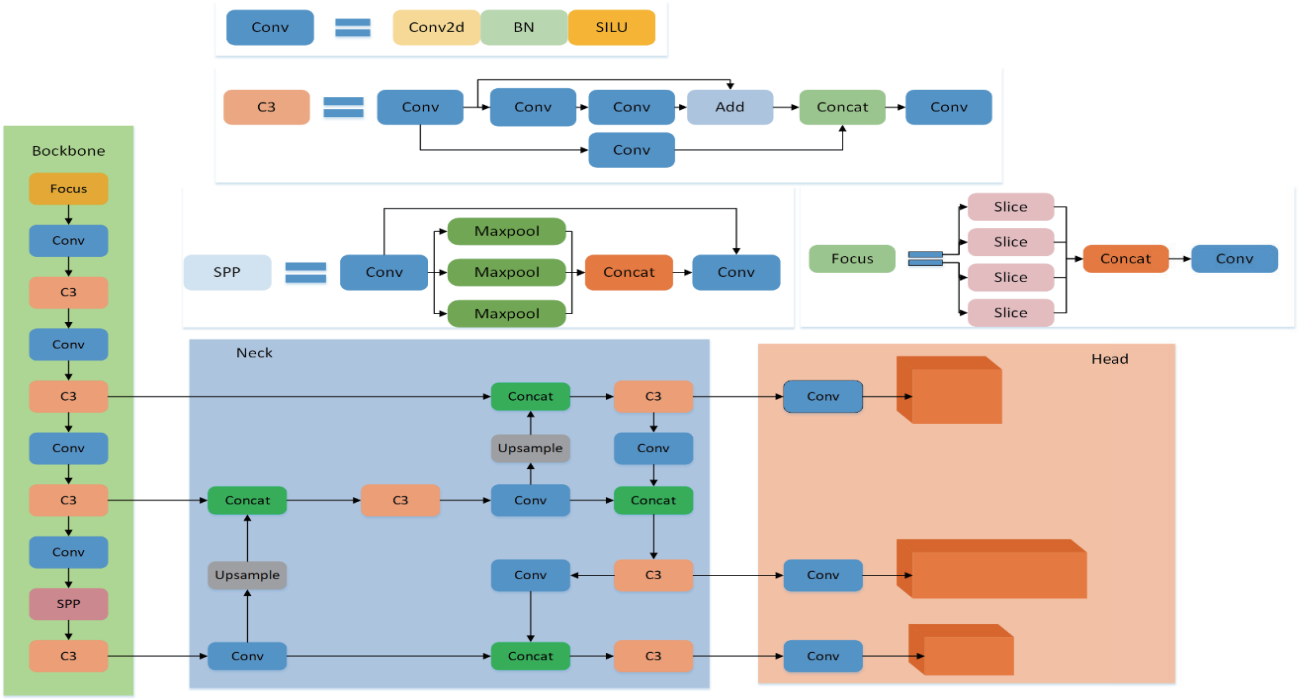

图 1 YOLOv5 网络结构图

YOLOv5 的网络模型可以分为以下部分：Input（输入端）：用于接收并且对数据进行预处理；Backbone（主干网络）：这是网络的主体部分，用于从图像中提取丰富的特征表示；Neck（颈部）：负责对特征图进行多尺度特征融合，连接主干网络和头部；Head（头部）：头部进行最终的回归预测 [6]。

1.1 Input

YOLOv5 的输入端对输入数据进行了一些处理，首先会对数据样本进行 Mosaic 数据增强，Mosaic 通过随机缩放、随机裁剪和随机排布的方式将多张图像拼接在一起，这有助于模型更好地适应不同目标的大小和位置。除此之外，输入端还会进行自适应锚框计算，针对不同的数据集，YOLOv5 会根据初始设定的锚框长宽进行自适应计算，以更好地适应不同尺度的目标。最后还支持自适应图片缩放：YOLOv5 在输入图像上进行自适应缩放，以适应不同大小的输入图像，提高检测速度。

1.2 Backbone

Backbone 部分是网络的主体部分，用于从图像中提取丰富的特征表示，主要包括 Focus、C3 和 SPP 三个模块。Focus 用于在输入图像进入主干网络之前进行特征处理，类似于一种特殊的下采样方法，旨在有效提取图像中的信息。Focus 模块将输入特征图切分成多个低分辨率的子特征图，然后将这些子特征图拼接在一起。这样，它将图像的宽度和高度信息转换到通道维度，从而减少了下采样带来的信息损失。

Backbone 主干网络采用了 C3（CSPDarknet53）结构，这是对 YOLOv3 版本中的 Darknet53 进行的改进。CSPDarknet53 是一种交叉阶段部分（CSP）结构，将 Darknet53 与 CSP 融合。它在特征提取过程中引入了跨层连接，从而更好地融合不同尺度的特征。通过 CSP 策略，CSPDarknet53 将特征图分成两部分，一部分直接用于预测，另一部分经过跨层连接后再与原始特征图融合。这种结构有助于提高特征的表达能力，使模型更好地适应不同目标的大小和位置。

SPP 模块是一种空间金字塔池化方法，用于多尺度目标检测的特征提取。SPP 模块将输入特征图分成不同大小的网格，并在每个网格中进行池化操作，得到固定大小的特征向量。在 YOLOv5 中，SPP 模块的结构包括多个不同大小的池化操作，通过将这些池化结果拼接在一起，实现局部特征和全局特征的融合，提高检测精度。

1.3 Neck

Neck 模块是连接主干网络（Backbone）和头部网络（Head）的部分，负责特征融合和处理，以提高检测的准确性和效率。在 YOLOv5 中，Neck 模块采用一种名为 PANet（path aggregation network）[7] 的特征融合模块。PANet 是一种用于处理多尺度目标检测的技术，它可以通过在骨干网络上添加不同尺度的特征层来实现。PANet 通过自顶向下和自下向上两个部分来实现特征融合。自顶向下通过上采样和与更粗粒度的特征图融合来实现不同层次特征的融合，自下向上通过使用一个卷积层来融合来自不同层次的特征图。

1.4 Head

Head 模块是 YOLOv5 网络的最后一部分，它负责将特征图转换为目标检测的最终输出，包括目标类别、位置和置

信度。其中预测层是此部分的核心，通过卷积操作将特征图转换为目标检测的预测结果。预测层输出的通道数取决于目标类别数和每个目标的预测信息（例如边界框坐标、置信度等）。其中，边界框损失通过损失函数 GIoU（generalized intersection over union）[8] 来度量真实值与预测值之间的距离，通过反向传播来训练网络。

2 YOLOv5 网络改进

原 YOLOv5 网络需要大量的训练数据才能训练出较好的模型。如果训练数据不足，会导致模型的准确率下降。在处理小目标时，YOLOv5 容易出现漏检或误检的情况。经过测试，YOLOv5 对路面坑洼的检测存在漏检错检，并且在阴雨天或对积水坑洼的检测效果不尽如人意，检测效率不足。为了改善上述问题，本文对网络的三部分进行了改进。

2.1 MobilNetV3 轻量级网络

相对于重量级网络而言，轻量级网络的特点是参数少、计算量小、推理时间短。MobileNetV3 是一种典型的轻量级卷积神经网络，它在保持高性能的同时，具有较小的模型大小和低计算复杂 [9]。具体的网络结构如图 2 所示。

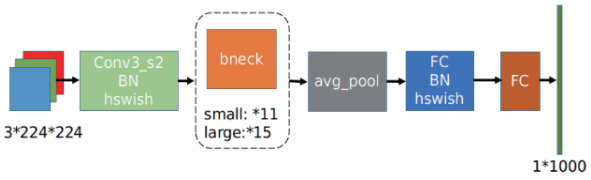

图 2 MobileNetV3 网络结构图

MobileNetV3 继承了 MobileNet 系列的主要特点，其中通道可分离卷积是其发挥轻量级作用的关键因素。通道可分离卷积将卷积操作分为两个步骤：首先进行通道压缩，然后再进行正常的 1*1 卷积，这样做能显著减少计算和参数数量。同时，网络相较于前代网络将激活函数更改为 Hard-swish 激活函数。这是一个快速轻量级的激活函数，它是在 ReLU 激活函数的基础上进行改进的，通过增加一定的非线性特性和使用更少的计算代价来提高模型的表现。具体的激活函数公式为：

$$H-swish[x]=x\frac{RELU6(x+3)}{6} \tag{1}$$

$$\text{RELU6}=\text{MIN}(6,\text{MAX}(0,x)) \tag{2}$$

2.2 SimAM 注意力模块

注意力机制是一种模仿人类视觉和认知系统的方法，允许神经网络在处理输入数据时集中注意力于相关的部分。通过引入注意力机制，神经网络能够自动地学习并选择性地关注输入中的重要信息，提高模型的性能和泛化能力。现有的注意力模块通常关注信道或空间维度，通过额外的子网络生成注意力权值，但这两种注意机制与人脑中的特征和空间注意不完全对应。如图 3 所示，为了更好地模拟人脑的注意机制，SimAM 设计出一种通过能量函数来估计三维权重的方法，通过评估每个神经元的重要性，度量神经元之间的线性可分性，同时保持轻量级属性 [10]。

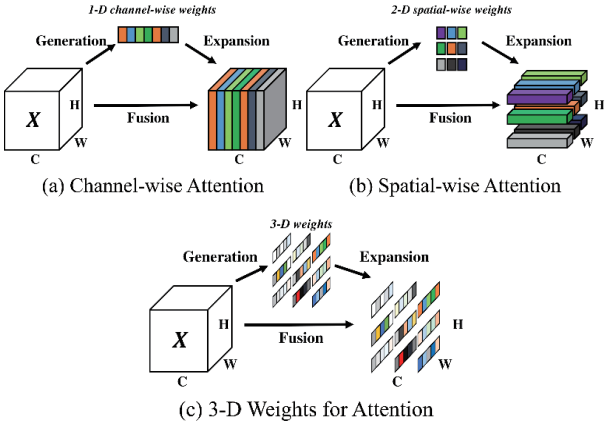

图 3 SimAM 注意力机制实现三维权重

能量函数公式为：

$$e_t^*=\frac{4\left(\hat{\sigma}^2+\lambda\right)}{\left(t-\hat{u}\right)^2+2\hat{\sigma}^2+2\lambda} \tag{3}$$

式中：μ 和 σ 是除 t 以外的所有神经元的平均值和方差。公式表明，能量越低，神经元与周围神经元越不同，对视觉处理越重要，所以注意力模块根据 e_t^* 值的大小来反向对不同神经元赋予不同的权重，提取更加重要的通道特征，剔除多余特征，使模型精度提高。

2.3 Alpha-IoU 损失函数

YOLOv5 原损失函数中的 GIoU 预测框损失函数，其计算基于边界框回归指标的聚合效果：GIoU 计算目标检测的损失时，决定于预测框和真实框之间的距离、重叠区域和纵横比等指标的聚合。当真实框包含预测框，且真实框与预测框的大小固定时，其 IoU 恒定为预测框与真实框面积的比值，则无论预测框在真实框中的哪个位置，损失都不变，造成梯度消失。这导致收敛速度较慢且效率较低，导致检测器不准确。后续的有关改进研究中提出了 DIoU[11] 与 CIoU[12] 等损失函数在 IoU 损失中引入惩罚项以缓解梯度消失问题。本文使用的 Alpha-IoU 损失函数是在现有损失函数的基础上引入 power 变换构建的一个新的 IoU 损失函数。首先，将 Box-Cox 变换应用于 IoU 损失；然后推广为 power IoU loss，其中 α 是一个权重参数。除此之外，还添加了一个附加的 power 正则项，使得 Alpha-IoU 能够概括现有的基于 IoU 的损失，包括 GIoU、DIoU 和 CIoU 等 [13]。Alpha-IoU 不引入额外的参数，直接估计三维权重。针对不同的模型要求，适当选择 α（即 α>1）有助于提高 IoU 目标的损失和梯度自适应加权的边界框回归精度。经过多次实验，Alpha-IoU 在小数据集和噪声的边界框提供更强的鲁棒性。在这里，只依次列出 Alpha-

CIoU、Alpha-GIoU 与 Alpha-DIoU 的公式（4）～（6）以供参考。

$$L_{\alpha-CIoU}=1-IoU^{\alpha}+\left(\frac{|C/(B\cup B^{gt})|}{|C|}\right)^{\alpha} \tag{4}$$

$$L_{\alpha-DIoU}=1-IoU^{\alpha}+\frac{\rho^{2\alpha}(b,b^{g_t})}{C^{2\alpha}} \tag{5}$$

$$L_{\alpha-DIoU}=1-IoU^{\alpha}+\frac{\rho^{2\alpha}(b,b^{g_t})}{C^{2\alpha}}+(\beta\mu) \tag{6}$$

3 实验与结果分析

3.1 实验平台及参数配置

为验证改进措施的有效性，在公开数据集 Annotated Potholes Image Dataset 上进行对应实验，具体实验环境及其配置如表 1 所示。在模型中选择的图片大小为 640×640，由于数据集偏小，考虑到过拟合问题，将迭代次数设置为 100 轮，批次大小选择 16，初始学习率为 0.01，循环学习率为 0.2，动量设置为 0.917，权重衰减 0.000 5，为了改善模型泛化能力，最小化损失锐化，采用 SAM 优化器，实验均是在不加载预训练权重的情况下进行。

表 1 实验环境配置

参数	配置
操作系统	Windows 11
CPU	AMD Ryzen 7745HX 3.6GHZ
GPU	NVIDIA GeForce RTX 4060 8GB
内存	16GB
编程语言	Python 3.11.5
深度学习	Pytorch 2.2.0
并行计算架构	CUDA version 11.8

3.2 数据集准备

本次实验采用的数据集为 kaggle 上公开的路面坑洼图像数据集 Annotated Potholes Image Dataset，数据集共有 665 张图片，带有对应的标注信息。同时，通过网络爬虫和实地拍摄对原数据集扩充，经过 labelimg 标注工具标注检测目标，最终将数据集扩充至 820 张数据。数据集中，小型坑洼和中型坑洼共占数据集的 80%，余下的 20% 数据为大型坑洼。最后，将数据集按照 8:1:1 的比率划分为训练集、验证集和测试集。

本文采用的数据集格式为 PASCOL VOC，但是 YOLOv5 只支持 txt 的数据集文件，所以要将 PASCOL VOL 格式下的 xml 文件转换成 txt 文件。经过上述的数据集准备，每个文件都标注出了目标框的左上角 $(x_{\min},y_{\max})$ 与右下角 $(x_{\max},y_{\min})$ 顶点坐标。YOLOv5 所需要的文件要求中心点坐标 (x,y) 以及目标框的宽 w 与高 h，转换公式如式（7）～（9）所示，其中 i_w 与 i_h 表示输入图像的宽与高：

$$x=\frac{x_{max}+x_{min}}{2i_w} \tag{7}$$

$$y=\frac{y_{max}+y_{min}}{2i_h} \tag{8}$$

$$w=\frac{x_{max}-y_{min}}{i_w} \tag{9}$$

$$h=\frac{y_{max}-y_{min}}{i_h} \tag{10}$$

3.3 性能评价指标

实验的好坏要通过训练结果的指标来评价，本文适用的指标有 Precision（精确率）、Recall（召回率）和 mAP（mean average precision，平均检测精度）。其中，精准率表示模型预测为正样本的结果中，实际为正样本的比例；召回率表示实际为正样本的结果中，模型成功预测为正样本的比例；mAP 则表示数据集中所有类别的平均精度的均值，综合体现了模型的优劣。上述评价指标的计算方法为：

$$P=\frac{T_P}{T_P+F_P} \tag{11}$$

$$R=\frac{T_P}{T_P+F_N} \tag{12}$$

$$F=\frac{\sum_{i=1}^{C}P_i}{C} \tag{13}$$

$$T=\int_0^1 P(r)\mathrm{d}r \tag{14}$$

式中：T_P 表示正样本被正确筛选的数量；F_P 表示负样本被当作正样本筛选的数量；F_N 表示正样本被当作负样本筛选的数量；C 表示类别数目，F 表示 mAP，P_i 表示第 i 个类别的精确度；T 为精度与召回率曲线与坐标轴所围面积。

3.4 消融实验

消融实验是科学研究中常用的一种方法，通过消融实验可以知道每一种改进之间的差异性。本文通过消融实验，依次更改主干模型、添加注意力机制以及修改损失函数，实验结果如表 2 所示。可以看出，在 YOLOv5 基础上改进主干网络为 MobileNetV3 的情况下，在损失少量精度的前提下参数量和计算量大幅减少，检测速率提高，模型具有轻量化特点。在添加注意力机制 SimAM 与改进损失函数这两种情况下，没有带来额外计算量，并且精度和召回率均有所提升。将上述所有改进模块加入到 YOLOv5 模型中，参数量下降 52%，计算量下降 39%，使得模型轻量化，而且精度提高 12%，召回率提高 10%，平均检测精度提高 15%，达到 0.85，检测速率也有小幅度提高，达到 51.8 帧 /s。上述消融实验表明，改进后的 YOLOv5 网络模型满足快速精准进行路面坑洞检测的要求，并且具有轻量化模型的优点。

表 2 消融实验结果

模型	参数/M	计算量/GFLOPs	P	R	mAP	速率/(帧/s)
YOLOv5	7.01	15.8	0.77	0.69	0.74	48.6
YOLOv5+MobileNetV3	3.63	6.2	0.73	0.64	0.70	52.3
YOLOv5+SimAM	7.01	15.8	0.81	0.74	0.78	48.6
YOLOv5+Alpha-IoU	7.01	15.8	0.84	0.73	0.83	48.2
YOLOv5+MobileNetV3+SimAM+AlphaIoU	3.63	6.2	0.86	0.76	0.85	51.8

3.5 检测效果对比

为了直观体现本文改进模型对于路面坑洞检测的有效性，随机挑选几张路面坑洞数据进行对比测试，测试结果如图 4 所示。图 4 上侧为原 YOLOv5 模型的测试结果，下侧为改进 YOLOv5 模型的测试结果。检测对比图可以看出，原模型存在漏检现象，特别是在积水坑洼的条件下；而改进模型对此有所改进，减少了漏检现象，提高了检测表现。

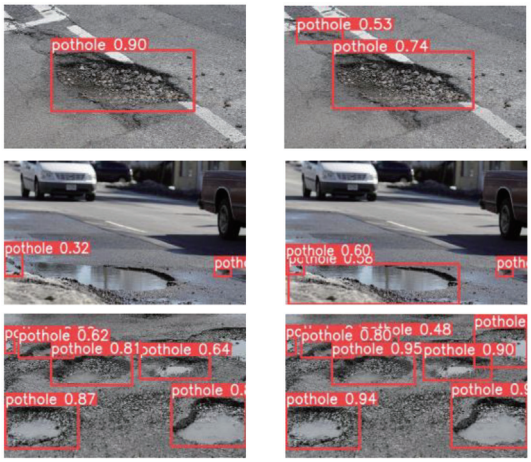

图 4 YOLOv5 与改进 YOLOv5 检测效果对比

4 结语

本文旨在研究 YOLOv5 改进模型在路面坑洼检测中的应用，经过测试研究，首先将原 YOLOv5 的主干网络改为轻量级网络 MobileNetV3，使模型体量大幅缩减，提高检测速度；然后引入无额外参数的 SimAM 注意力机制实现三维权重，对不同的神经元赋予不同的权重，使网络通道专注于重要特征信息，剔除无关通道信息，使模型捕获更多关键信息；最后将损失函数替换为 Alpha-IoU 损失函数，弥补原损失函数的缺陷，提高检测精度，并且经过实验证明了上述改进的准确性。但是本文改进方法仍然对远处小坑洼的检测不是特别敏感，将在后期的工作中进一步改进，提高模型的检测精度与鲁棒性。

参考文献：

[1] 赵潇 . 基于深度学习的路面坑洼检测系统研究与实现 [D]. 宁夏 : 宁夏大学 ,2020.

[2] 方路平 , 何杭江 , 周国民 . 目标检测算法研究综述 [J]. 计算机工程与应用 ,2018,54(13):11-18+33.

[3] 许德刚 , 王露 , 李凡 . 深度学习的典型目标检测算法研究综述 [J]. 计算机工程与应用 ,2021,57(8):10-25.

[4]REN S, KAIMING HE, ROSS B, et al.Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6):1137-1149.

[5]WEI L, ANGUELOV D, ERHAN D, et al.SSD:single shot multibox detector[C]//Computer Vision–ECCV 2016.Cham: Springer, 2016:21-37.

[6] 赵婉月 . 基于 YOLOv5 的目标检测算法研究 [D]. 西安 : 西安电子科技大学 , 2021.

[7]LIU S, QI L, QIN H, et al.Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Piscataway: IEEE, 2018: 8759-8768.

[8]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/CVP Conference on Computer Vision and Pattern Recognition,[v.1].Piscataway: IEEE, 2019: 658-666.

[9]HOWARD A, SANDLER M, CHU G, et al.Searching for mobilenetv3[C]//2019 IEEE/CVF International Conference on Computer Vision.Piscataway:IEEE,2019:1314-1324.

[10]YANG L, ZHANG R, LI L, et al.Simam:a simple, parameter-free attention module for convolutional neural networks[C]//International Conference on Machine Learning,Part 15 of 16.New York:Curran Associates,Inc,2022:11853-11864.

[11]ZHENG Z, WANG P, LIU W, et al.Distance-IoU loss:faster and better learning for bounding box regression[C]//34th AAAI Conference on Artificial Intelligence,v.7,p.4.AAAI Technical Track:Vision. Palo Alto:Association for the Advancement of Artificial Intelligence, 2020:12993-13000.

[12]ZHENG Z, WANG P, REN D, et al.Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE transactions on cybernetics, 2021, 52(8): 8574-8586.

[13]HE J, ERFANI S, MA X, et al.Alpha-IoU:a family of power intersection over union losses for bounding box regression[J]. Advances in neural information processing systems, 2021, 34: 20230-20242.

【作者简介】

陈佳宁（2000—），男，河南许昌人，硕士研究生，研究方向：目标检测。

刘晓群（1968—），通信作者（email: lfyq2@163.com），男，河北张家口人，硕士，教授，研究方向：计算机网络与信息安全。

郝娟（1989—），女，河北保定人，硕士，副教授，研究方向：图像处理。

（收稿日期：2024-03-15）