

改进高通道卷积的 YOLOv7-tiny 视觉辅助轻量化算法

欧阳玉旋, 彭垚潘, 张荣芬*, 刘宇红

(贵州大学大数据与信息工程学院, 贵州 贵阳 550025)

摘要:针对市面上大多数视觉辅助系统算法存在参数量大、检测性能低、不便于部署手机移动端等问题,基于 YOLOv7-tiny 设计了一个轻量级的视觉辅助算法。在网络中使用感受野模块(receptive field block, RFB)融合不同尺度的特征信息,提高对不同分辨率大小物体的检测精度;利用激活函数 Silu(sigmoid linear unit)的非线性,增强模型的拟合能力,提升模型的学习速度和检测精度;通过对比实验选择性能更佳的深度卷积(depthwise convolution, DWConv)实现模型的轻量化。实验结果表明,改进后的轻量化模型相比原模型,参数量减少了 52.1%,并获得了最佳的检测性能。与其他主流目标检测算法相比,该算法以 2.90 M 参数量实现了对室内目标更精准的实时检测。

关键词:高通道卷积;感受野模块;激活函数 Silu;深度卷积

中图分类号:TP391.41 **文献标志码:**A

引用格式:欧阳玉旋,彭垚潘,张荣芬,等.改进高通道卷积的 YOLOv7-tiny 视觉辅助轻量化算法[J].山东大学学报(理学版),2025,60(9):62-70,109.

Improvement of the YOLOv7-tiny visual-assisted lightweight algorithm based on high-channel convolution

OUYANG Yuxuan, PENG Yaopan, ZHANG Rongfen*, LIU Yuhong

(College of Big Data and Information Engineering, Guizhou University, Guiyang 50025, Guizhou, China)

Abstract: In response to the issues of large parameter size, low detection performance, and inconvenience for deployment on mobile devices in most existing visual assistance system algorithms, a lightweight visual assistance algorithm is designed based on YOLOv7-tiny. The receptive field block (RFB) is used in the network to fuse the feature information of different scales to improve the detection accuracy of objects of different resolution sizes. The non-linearity of activation function sigmoid linear unit (Silu) is used to enhance the fitting ability of the model, improve the learning speed and detection accuracy of the model. Finally, the depth-wise convolution (DWConv) with better performance is selected by comparison experiment to achieve the lightweight of the model. The experimental results show that the parameters of the improved lightweight model are reduced by 52.1% compared with the original model, and the best detection performance is obtained. Compared with other mainstream object detection algorithms, this algorithm achieves more accurate real-time detection of indoor objects with 2.90 M parameters.

Key words: high-channel convolution; receptive field block; activation function Silu; DWConv

0 引言

根据 2020 年世界卫生组织发布的《世界视觉报告》,截至当年,全球有超过 22 亿人患有视力损伤^[1]。此外,2021 年 5 月中国盲人协会官方网站的数据显示,中国约有 170 万盲人^[1]。如何改善盲人的生活质量已经成为社会关注的问题。对于视障人群来说,视觉的损伤导致其活动范围受限,大多数时间活动在室内,室内复杂的场景又给其生活带来了诸多不便^[2]。为了提升盲人的生活自理能力,各种视觉辅助系统也在不

收稿日期:2024-03-11;网络出版时间:2025-01-10 17:54:29

基金项目:贵州省基础研究(自然科学)资助项目(黔科合基础-ZK[2021]重点 001)

第一作者:欧阳玉旋(1993—),女,硕士研究生,研究方向为深度学习目标检测。E-mail:1621413779@qq.com

*通信作者:张荣芬(1977—),女,教授,硕士生导师,博士,研究方向为机器视觉、智能算法及智能硬件等。E-mail:rfzhang@gzu.edu.cn

断研发中。随着边缘计算的兴起,深度学习模型被推向网络边缘,由于嵌入式平台计算资源和存储资源有限,因此大多数视觉辅助算法停留在研究阶段,无法移植部署到室内机器人、智能监控、智慧安防等嵌入式平台,并应用于实际生活。为了使深度学习模型能够在资源受限的设备和环境中得到广泛应用,满足各种应用场景的需求,对模型进行轻量化研究以便于部署边缘设备具有重要的研究意义。

近年来,目标检测算法主要分为传统目标检测算法和基于深度学习的目标检测算法。基于深度学习的目标检测算法分为双阶段目标检测算法和单阶段目标检测算法。双阶段目标检测算法如 R-CNN (region-based convolutional neural networks)^[3-5] 系列,主要任务是定位和分类。它们首先使用选择性搜索等方法提取感兴趣的候选区域,然后利用卷积神经网络对每个候选区域进行特征提取和分类。然而这类算法存在模型大、泛化性能差、检测速度慢等问题。单阶段目标检测算法如 SSD (single shot multiBox detector)^[6]、YOLO (you only look once)^[7] 系列、RetinaNet^[8] 和 FCOS (fully convolutional one-stage object detection)^[9],不需要显式地找出感兴趣的区域,而是将目标检测问题转化为回归问题,直接对整个图像进行处理得到目标的类别和位置信息。相比双阶段算法,单阶段目标检测算法速度更快,在模型大小方面更具优势,更利于部署边缘端设备。YOLOv3^[10] 采用特征金字塔网络 (feature pyramid network, FPN) 进行多尺度特征融合,将浅层特征和深层特征结合,使得特征图具有较强的语义信息。YOLOv5^[11]、YOLOv6^[12]、YOLOv7^[13] 和 YOLOv8 采用类似的特征金字塔网络,在 FPN 的基础上添加了一条自底向上的路径,进一步融合浅层信息和深层信息,以减少浅层信息的丢失。然而,新增的自底向上的路径增加了模型参数量、运算量和存储空间,不便于部署到嵌入式平台上。

双阶段算法在准确性方面表现较好但速度较慢,而单阶段算法在速度方面较快但准确性有所不足,不同算法采用不同的策略来解决这些问题,但在模型轻量化和模型部署到嵌入式平台方面依然存在挑战。实现目标检测网络的轻量化常用的方法是使用轻量化注意力模块和替换普通卷积为轻量化卷积,例如,ResNet^[14] 引入 BottleNeck^[15] 结构来减少大尺寸卷积核的通道数。GhostNet^[15] 和 ShuffleNet^[16] 则采用深度可分离卷积^[17] 的设计,提出了轻量化高效特征提取模块,并构建相应的轻量化网络模型。文献[18]通过引入深度可分离卷积实现模型的轻量化。文献[19]通过引入 Ghost 模块和无参注意力模块 SimAM^[20] 对模型进行轻量化改进。文献[21]引入深度可分离卷积和高效通道注意力机制大幅降低模型计算量。但文献[14-21]提出的方法方法都是对各卷积层进行封装集成或是简单替换普通卷积层,未充分认识到网络中存在的大量冗余信息,这将导致模型计算消耗大,检测性能差。本文通过将高通道卷积层替换为深度卷积 (depthwise convolution, DWConv)^[17],剔除密集、冗余信息,保留关键特征信息,在大幅度压缩模型的同时增强了模型对特征信息的理解能力。

为了改善盲人视觉辅助系统算法的检测性能和可部署性,本文以 YOLOv7-tiny 为研究模型,在保证模型高检测性能的前提下对模型进行轻量化处理,使模型更易于部署到嵌入式平台。本文主要工作如下:

1) 采用感受野模块 (receptive field block, RFB)^[22],使用不同大小卷积核融合不同尺度的特征信息,帮助模型更好地理解 and 捕捉输入数据中的关联性和上下文信息,获得更丰富的语义信息,提升对不同尺度目标的识别率;

2) 利用激活函数 Silu (sigmoid linear unit)^[23] 的非线性抽象特征,更好地建立输出数据与预测数据之间的映射关系,加速模型的收敛速度,并进一步提升模型的识别率;

3) 在网络下采样过程中,浅层语义信息逐步丢失,当输出通道为高通道时,会产生大量的噪声信息,利用深度卷积,可剔除网络中无效、冗余的特征信息,降低模型的计算消耗。

1 改进算法

1.1 改进的 YOLOv7-tiny 算法

YOLOv7-tiny 是 YOLOv7 系列中最小的模型,参数量约 6 MB,是面向 GPU 的边缘架构,为算法部署到嵌入式或移动端提供了可行的方案。原模型网络中存在由大量标准卷积密集连接的结构,导致模型下采样过程中存在大量冗余信息,消耗了大量计算资源,不利于网络的特征提取。并且受限于室内环境、光线的影响,模型在刻画目标语义信息方面存在局限,降低了模型的检测性能。

为了提升模型的检测性能并将模型部署到移动设备进行实时精准的检测,本文对 YOLOv7-tiny 进行改进。通过引入 RFB 特征融合模块代替原始的 SPPCSPC 模块,在小幅度降低模型参数量的同时加强模型理解输入数据关联性和上下文信息能力。采用 Silu 激活函数增强模型的拟合能力,适应非线性数据分布和复杂的决策边界,提升模型的识别精度。随着模型下采样的进行,深层高通道卷积处会产生大量冗余的噪声信息。将高通道卷积(通道数为 512 和 256 且卷积核大小为 3)替换为 DWConv,剔除冗余的噪声信息,提升通道有效信息与无效信息之间的比值,实现了对模型高质量的压缩。改进后的网络结构如图 1 所示。

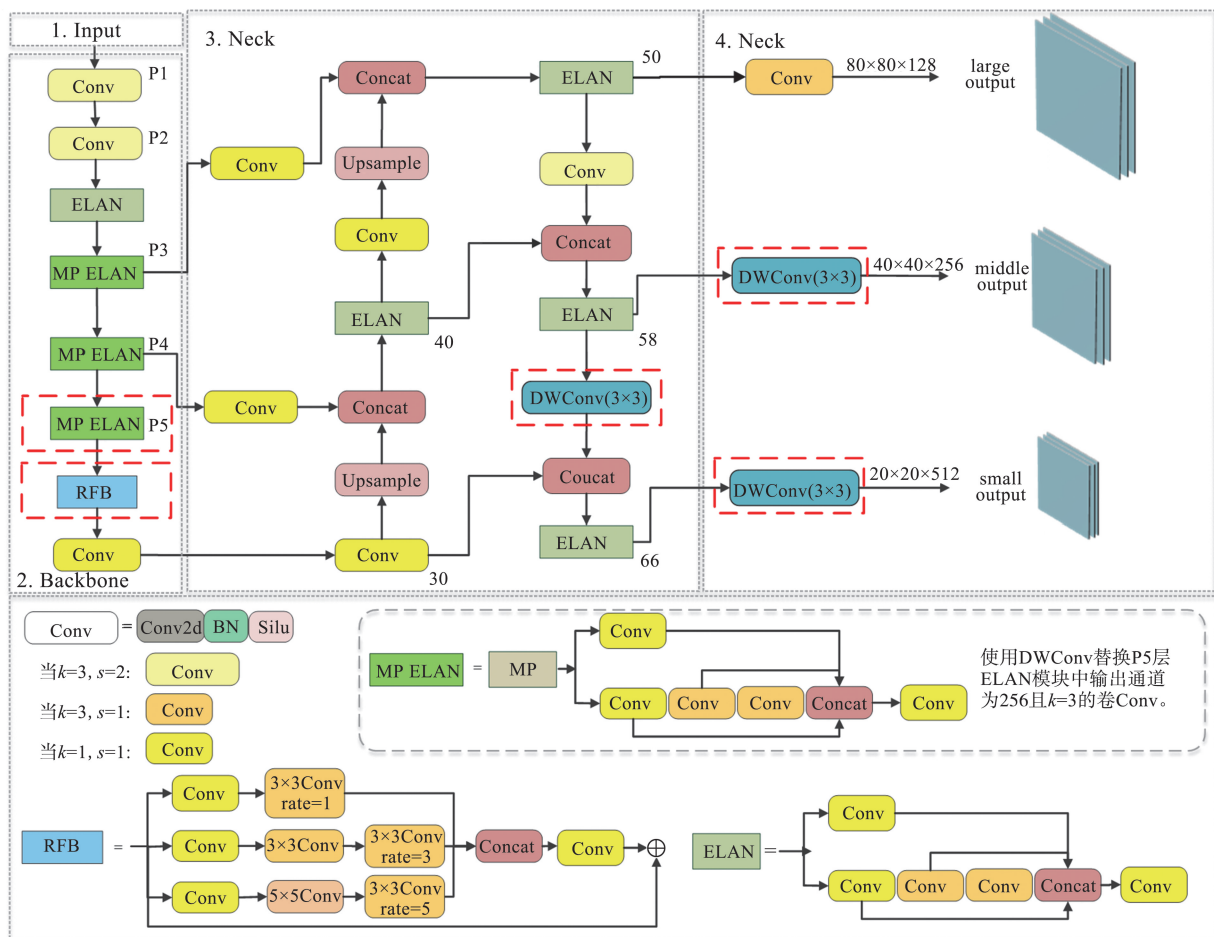


图 1 改进的 YOLOv7-tiny 网络结构图

Fig.1 Diagram of improved YOLOv7-tiny network structure

1.2 感受野模块

RFB 采用多分支的空洞卷积结构设计,通过调整空洞卷积的参数,可以有效扩大卷积层的感受野。这种多分支结构继承了 ResNet 检测模型中的残差结构和 InceptionV2^[24]结构,如图 2 所示。RFB 通过调整膨胀率的大小,可以实现对不同尺度信息的捕捉。较低的膨胀率可以用于捕捉细节信息,而较高的膨胀率可以用于捕捉更宽广的语义信息。通过在网络中使用不同膨胀率的空洞卷积,可以在不增加参数量和计算复杂度的情况下扩大卷积操作的感受野,提高模型对多尺度目标的表示能力。RFB 通过 1×1 卷积有效减少了网络计算开销,同时促进了目标特征信息的融合。模块通过将多个具有不同内核的分支和空洞卷积层相结合,帮助模型更好地捕捉图像中的上下文信息和长距离依赖关系。残差结构通过引入跨层连接,允许信息在网络中直接传递,避免在多次卷积之后特征细节过度丢失,有助于解决梯度消失的问题,并且优化了模型的训练过程。模块最后将各分支拼接重组后与输入特征相加得到了语义丰富的多层次特征表示。

传统的卷积操作只能在相邻的像素点上计算,而空洞卷积通过卷积核与固定间隔的像素点进行乘法运算,可以在不增加参数数量的情况下提高模型的感受野范围^[25]。空洞卷积的实际卷积核大小表示为

$$N = n + (n-1)(r-1), \quad (1)$$

其中, N 是空洞卷积核大小, n 是普通卷积核大小, r 是空洞卷积的膨胀率。

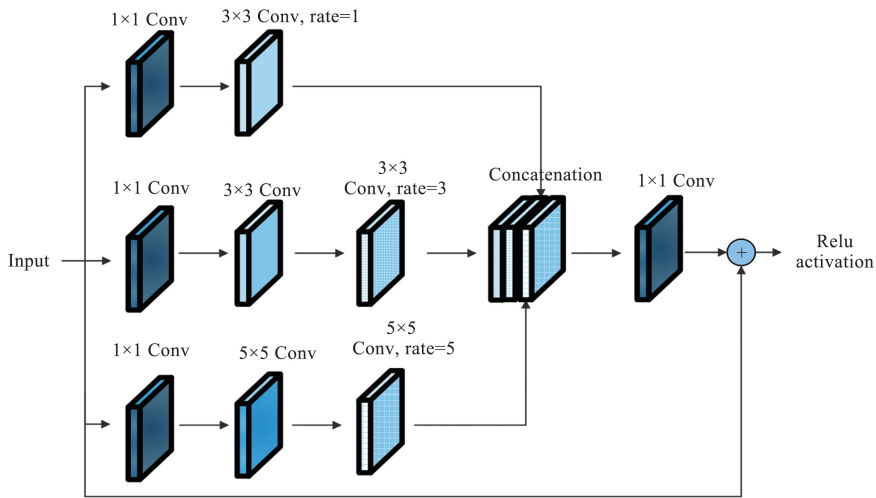


图 2 RFB 结构图

Fig.2 The structure diagram of RFB

1.3 Silu 激活函数

Silu 激活函数是一种非线性函数,通过对神经网络的输出进行非线性变换,增强模型的表达能力和学习能力,定义为

$$\text{Silu}(x) = x \times \text{sigmoid}(x), \tag{2}$$

其中 $\text{sigmoid}(x)$ 是 Sigmoid 函数,定义为

$$\text{sigmoid}(x) = \frac{1}{(1+e^{-x})}. \tag{3}$$

Silu 激活函数在输入范围内具有较大的梯度,可以改善梯度消失问题,促进更有效的反向传播训练过程,在接近 0 和远离 0 的区域都具有较小的梯度,有利于训练过程的稳定性。本文选取 Silu 激活函数,主要是考虑其良好的非线性,能使神经网络更好地适应非线性数据分布和复杂的决策边界,学习到复杂的函数关系,提高神经网络的表达能力和学习能力。

1.4 轻量化卷积

随着深度学习技术的不断发展,卷积神经网络作为模型的核心组件,在提取输入数据特征信息方面发挥了关键作用。为满足不同应用场景的需求,业界已经涌现出多种轻量级卷积结构,这些新型卷积模块在减少参数量和计算复杂度的同时,保持了模型良好的检测性能。为解决大多数视觉辅助系统算法参数量大、检测性能低、不便于部署移动端等问题,本文通过对 DWConv 和 Ghost 2 种轻量级卷积进行实验对比,以选取更优的轻量化改进方案,实现高性能的轻量级模型,使模型更易于部署到边缘设备。

1.4.1 深度卷积 DWConv

深度卷积相对于普通卷积而言,具有计算复杂度低和模型小的优势,更利于部署计算资源和存储资源有限的移动端平台。图 3 为普通卷积结构图,当输入为一个边长 D 、通道为 N 的卷积,通过 M 个边长为 k 、通道为 N 的卷积核,得到一个边长为 D 、通道数为 M 的特征图。仅仅考虑乘积计算,其参数量 $P = k \times k \times N \times M$ 。

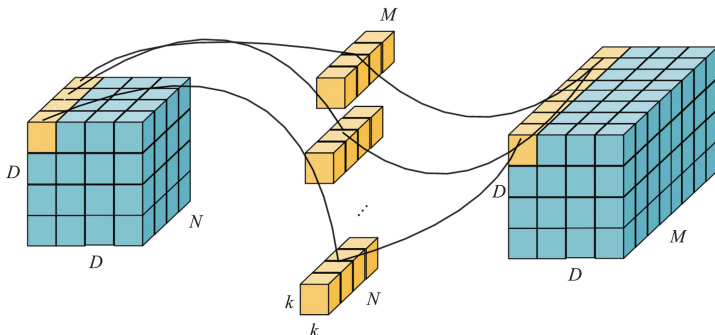


图 3 普通卷积结构图

Fig.3 General convolution structure diagram

图 4 为深度卷积结构,当输入为一个边长为 D 、通道为 N 的卷积,需要通过 N 个边长为 k 、通道为 1 的卷积核,得到一个边长为 D 、通道数为 N 的特征,此时的参数量 $P' = k \times k \times N \times 1$ 。由此可知,如果将普通卷积替换为深度卷积,参数量减少为原本的 $1/M$ 。本文将 YOLOv7-tiny 网络的高通道(通道数为 512 和 256)密集卷积模块,卷积核大小为 3×3 的普通卷积层替换为 DWConv,假设输入通道数为 N ,通过深度卷积操作后,仅单一卷积,得到通道数为 512 的特征图,所需的参数量将从普通卷积的参数量 $3 \times 3 \times N \times 512$ 变成 $3 \times 3 \times N \times 1$,降为原来的 $1/512 \approx 0.20\%$,对于通道数为 256 的参数量降为原来的 $1/256 \approx 0.39\%$ 。可以看出,使用深度卷积代替普通卷积,能大幅度降低模型的参数量。

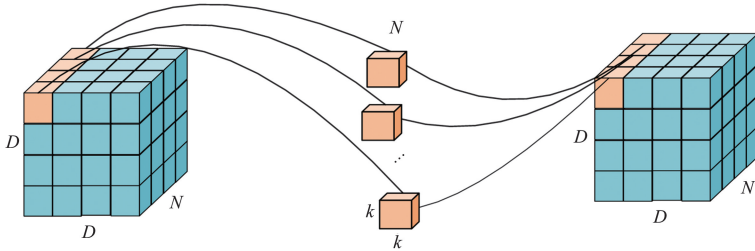


图 4 深度卷积结构图
Fig.4 Depthwise convolution structure diagram

1.4.2 幽灵卷积 GhostConv

在普通卷积中,一些特征图中的特征信息可能非常相似,通过消除冗余的特征信息能显著减少计算特征信息的操作量。GhostConv 是一种特殊的卷积结构,主要由卷积、线性变换 Φ 运算和特征拼接 3 个步骤组成,如图 5 所示。它首先通过常规卷积提取特征信息,得到一般特征 I 。然后,对一般特征图中的每个通道的特征进行线性运算,得到相同通道数的 Ghost 特征图,这些 Ghost 特征图是通过简单的线性变换得到,能够在几乎不影响模型检测精度的情况下,有效减少普通卷积的冗余性,降低模型计算复杂度。最后将一般特征 I 和 Ghost 特征图叠加起来作为输出。与普通的卷积相比,Ghost 模块能够在保持识别性能的同时,减少通用卷积层的计算消耗和模型参数量。

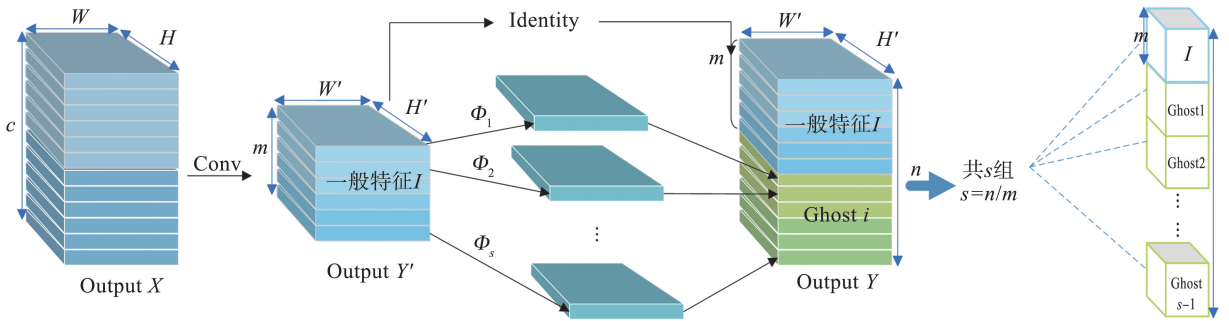


图 5 Ghost 模块
Fig.5 Ghost module

一个输入特征为 $X, X \in \mathbf{R}^{H \times W \times c}$ ($H \times W \times c$ 分别表示输入特征图的高、宽和通道数),输出特征 $Y, Y \in \mathbf{R}^{H \times W \times n}$,卷积核为 $k \times k$ 。通过普通卷积需要的参数量为

$$n \times c \times k \times k. \tag{4}$$

当输入特征 X 经过 Ghost 卷积模块,要想获得与普通卷积相似的识别性,需要经过普通卷积和线性变换 2 个步骤,获得输出特征 Y 。普通卷积生成 m 个特征图 $Y', Y' \in \mathbf{R}^{H \times W \times m}$,其中 $m \leq n$,对普通卷积输出的特征图进行线性变换 Φ ,结合拼接操作即可得到与普通卷积类似的 n 个特征图,且 $n = m \times s$ 。Ghost 卷积模块的浮点运算量包含普通卷积参数量和为线性变换参数量,分别表示为

$$m \times c \times k \times k, \tag{5}$$

$$(s-1) \times \frac{n}{s} \times k \times k. \tag{6}$$

普通卷积参数量与 Ghost 卷积模块参数量之间的关系为

$$r_c = \frac{n \times c \times k \times k}{m \times c \times k \times k + (s-1) \times \frac{n}{s} \times k \times k}, \tag{7}$$

当 $s \ll c$ 时,有

$$r_c \approx \frac{s \times c}{s + c - 1} \approx s, \tag{8}$$

由式(8)可知,当使用 Ghost 模块代替普通卷积,参数量将减少至原本的 $1/s$ 。

2 实验结果与分析

2.1 实验数据集

根据室内场景目标检测需要,对数据集^[18]进行增补,构建符合视障人群需求的室内目标检测数据集。数据集共 12 507 张图片,包含 sofa、table、chair、bed 等 14 个类别。图 6 为目标实例及其标签的分布情况。

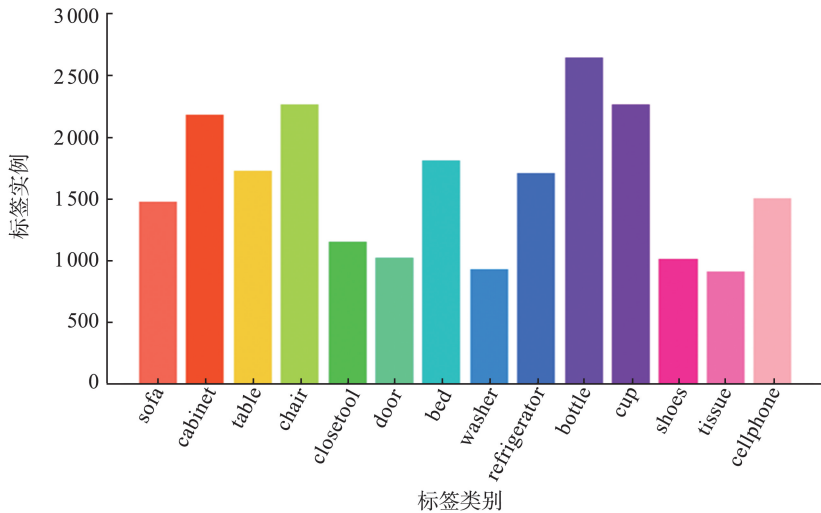


图 6 数据标签分布
Fig.6 Data label distribution

2.2 实验平台与训练设置

实验使用 Ubuntu 20.04 操作系统,CPU 为 AMD 3900X,内存 32GB,GPU 选用 NVIDIA GTX3090,实验环境为 Python3.8、Pytorch1.8.2、CUDA11.1。模型训练 200 次,batchsize 为 16,采用 Adam 优化器进行优化,初始学习率为 0.001。实验以参数量、平均精度 mAP@ .5 和 mAP@ .5:.95 和推理速度 FPS 为主要的性能指标对模型性能进行评估。

2.3 实验情况

2.3.1 特征融合模块实验

本文通过不同卷积核获得不同的感受野,提取到不同尺度的特征信息,将来自不同层或不同分支的特征进行整合,帮助模型更好地理解 and 捕捉输入数据中的关联性和上下文信息,获得更丰富的语义信息,以提升模型检测性能。本节在原模型基础上将 RFB 与 SPP、SPPF 和原模型中的 SPPCSPC 模块进行特征融合实验对比,验证 RFB 在模型中发挥的重要作用。实验结果如表 1 所示。

表 1 特征融合模块对比实验
Table 1 Comparative experiments of feature fusion module

方案	特征融合模块	参数量/M	mAP@ .5/%	mAP@ .5:.95/%	FPS
1	SPP	5.85	67.6	44.4	313
2	SPPF	5.85	67.5	44.5	313
3	SPPCSPC	6.05	67.7	44.4	244
4	RFB	5.88	68.5	46.0	250

与 SPP 和 SPPF 模块相比,虽然本文选取的 RFB 模块参数量略高,但是检测精度明显提升;与原模型的 SPPCSPC 模块相比,RFB 模块不仅提升了推理速度,同时具有更少的参数量,并且检测精度最高,在更利于部署移动端的情况下,能实现对室内目标更精准的实时检测。

2.3.2 激活函数改进实验

考虑到模型的检测性能和泛化性能,本文选取 Silu 激活函数,它可以更有效地捕获非线性抽象特征,从而提高模型的检测性能。在原模型(激活函数为 Leaky ReLU^[26])基础上对几种常见的激活函数 Mish^[27]、Elu^[28]、PRELU^[26]和原模型的 Leaky ReLU^[26]进行实验对比。具体实验情况见表 2,通过实验证明使用 Silu 激活函数,本文模型拥有最好的检测精度和的检测速度。

表 2 激活函数对比实验
Table 2 Comparison experiment of activation function

方案	激活函数	参数量/M	mAP@ .5/%	mAP@ .5:.95/%	FPS
1	baseline	6.05	67.7	44.4	244
2	Mish	6.05	68.7	46.1	130
3	Elu	6.05	68.6	45.4	227
4	PRELU	6.05	67.4	44.3	233
5	Silu	6.05	69.4	46.3	243

2.3.3 模型轻量化实验

由于移动端设备的计算资源和存储空间有限,单纯提升模型检测的精确性,如 YOLOv7-tiny+RFB+Silu 模型,并不能满足本文实际应用的需要。为了更利于算法部署到移动端设备,本文进一步对模型进行压缩,设计了模型轻量化实验方案,通过将深度卷积 DWConv 和幽灵卷积 GhostConv 分别加入 YOLOv7-tiny 改进模型(YOLOv7-tiny+RFB+Silu)对模型进行轻量化实验,验证了深度卷积对模型压缩的改进优势。具体实验情况见表 3。表中“-512-”表示逐步替换卷积的输出通道数;“-Head”表示网络的 3 个检测头;“(512-256) * 3”表示输出通道为 512 和 256 且卷积核大小为 3×3 的卷积。

表 3 轻量化实验
Table 3 Lightweight experiments

方案	模型	参数量/M	mAP@ .5/%	mAP@ .5:.95/%	FPS
1	baseline	6.05	67.7	44.4	244
2	YOLOv7-tiny+RFB+Silu	5.88	69.5	46.9	132
3	GhostConv-512	4.99	69.2	46.5	294
4	GhostConv-512-256	3.88	68.6	46.1	270
5	GhostConv-512-256-128	3.42	67.6	45.3	250
6	DWConv-512-256	1.85	65.0	42.3	250
7	DWConv-Head	4.29	68.4	45.5	326
8	DWConv-(512-256) * 3	2.90	68.3	45.4	333

根据表 3 数据,方案 3—5 使用幽灵卷积 GhostConv 循序渐进地替换网络中的高通道卷积(通道数为 512、256、128),发现通过使用 GhostConv 进行逐步替换高通道卷积块,模型得到了不同程度的压缩,同时,还从不同的指标上提升了模型的性能,由于 GhostConv 分为一般特征图和 Ghost 特征图 2 个部分,虽然减少了模型参数量但增加了网络的深度,使得模型推理速度 FPS 提升有限;方案 6 当使用深度卷积 DWConv 替换网络中的高通道卷积(通道数为 512 和 256)时,虽然模型得到了极致的压缩,但模型检测精度损失严重,达不到设计小模型、高精度的目的。方案 7 当仅将检测头端的 3 个高通道卷积块(通道数为 512、256、128)替换为深度卷积 DWConv,虽然提升了模型的检测精度和检测速度,但模型参数量并没有得到明显的改善。通过数次实验使用深度卷积 DWConv 替换网络中的各卷积块发现,与使用幽灵卷积 GhostConv 的方案 3—5 相比,方案 8 使用深度卷积 DWConv 替换网络中通道数为 512 和 256 且卷积核大小为 3 的高通道卷积,模型的参数量也获得了大幅度的缩减,虽然模型检测略低于方案 3 和 4,但模型获得了最好的推理速度。相比原模型,方案 8 模型 mAP@ .5 和 mAP@ .5:.95 分别提高了 0.6% 和 1.0%,模型参数量减少了 52.1%,较好地实现了模型检测精度和模型大小之间平衡,同时具有比原模型更优的推理速度,这充分证实深度卷积不仅可实现对网络中冗余信息的剔除,提升模型对特征的理解能力,还能降低模型的计算消耗,提升模型的检测速度,

达到了本文设计高性能轻量化模型的初衷。

2.3.4 与其他算法的实验对比

为确保模型对比的公平性,在保证各项参数指标不变的情况下,将本文模型与 Faster R-CNN-resnet50^[5]、SSD-vgg^[6]、Retinanet^[8]、Fcos^[9]、YOLO 系列等主流的目标检测算法在同一数据集上进行实验对比。同时增加模型大小指标以评估各模型部署嵌入式平台的可靠性。具体实验结果见表 4。

表 4 本文模型与其他目标检测算法对比
Table 4 Compare with other object detection algorithms

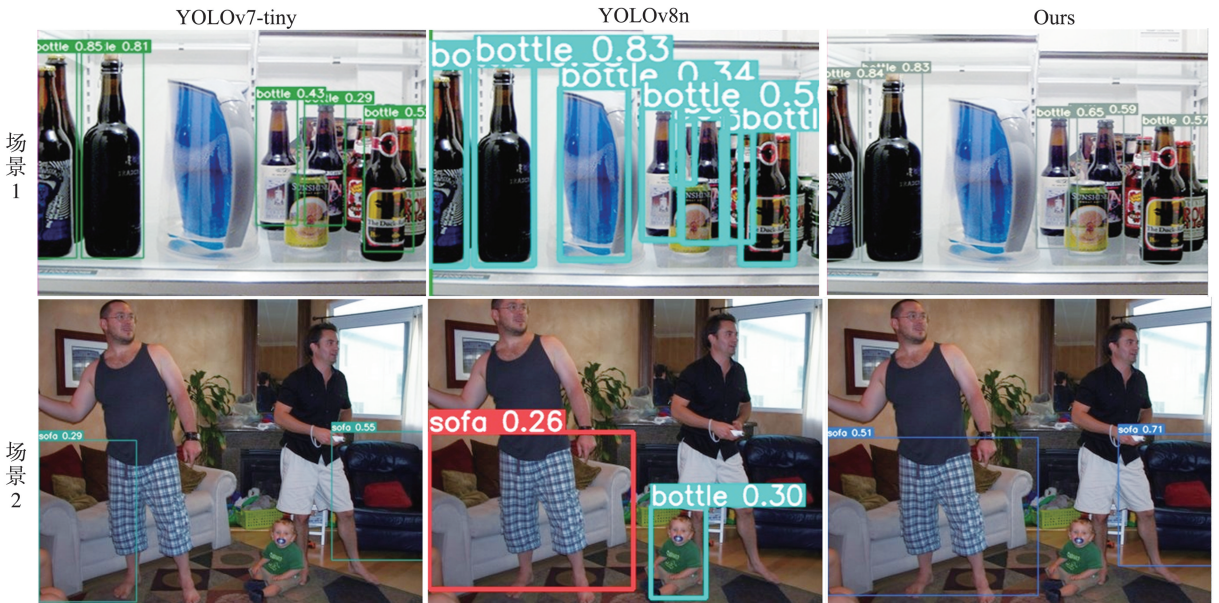
方案	模型	参数/M	mAP@.5/%	mAP@.5:.95/%	模型大小/MB
1	Faster R-CNN-resnet50	136.96	47.93	—	114.0
2	SSD-vgg	25.35	49.33	—	102.0
3	Retinanet	36.60	62.25	—	146.7
4	Fcos	31.14	66.22	—	123.0
5	YOLOv3- tiny	8.70	50.10	26.2	17.5
6	YOLOv5s	7.06	66.50	43.7	14.4
7	YOLOv8n	3.01	66.30	44.6	6.2
8	YOLOv7-tiny	6.05	67.70	44.4	12.3
9	Ours	2.90	68.30	45.4	6.0

注:“—”表示无法计算。

由表 4 对比实验可以看出,本文改进算法相比 Faster R-CNN-resnet50、SSD-vgg 等 8 个算法,在模型轻量化上做出了显著改进,在模型参数量仅 2.90 M 的前提下,平均精度获得了较大的提升,mAP@.5 和 mAP@.5:.95 分别为 68.3% 和 45.4%,比原模型 YOLOv7-tiny 分别提高了 0.6% 和 1.0%,比 SOTA 算法 YOLOv8n 分别提高了 2.0% 和 0.8%,模型大小仅 6.0 MB,相比其他算法,本文改进模型更易于部署到移动终端设备,实现对室内常见目标的精准识别。

2.3.5 实验结果可视化分析

选取 2 组场景图进行可视化对比分析,以评测改进模型的检测性能。如图 7 所示,对于像 bottle 这类的小目标和 sofa 这类大目标,改进算法相比原始算法 YOLOv7-tiny 和 YOLOv8n 有着更高的置信度,检测精度更高。说明本文算法能同时捕捉到高层和低层语义信息,得到的语义信息更加丰富,提高了模型检测精度。



注:图片来自文献[18]。

图 7 改进模型检测结果可视化

Fig.7 Visualization of test results of improved model

3 结论

为了设计一个能对室内常见物体进行精准识别的高性能轻量化模型,以便能够部署到资源受限的设备中,本文对 YOLOv7-tiny 网络模型进行了改进。通过引入 RFB 对多尺度特征进行高效融合,获得丰富的语义表达,并采用 Silu 激活函数的非线性拟合能力,对模型进行优化,增强模型的学习能力和识别性能;使用深度卷积 DWConv 对模型做大幅度的压缩,降低了模型的参数量,同时提高了模型的检测性能。对比实验表明,改进模型大幅度降低了模型空间的消耗,并且具有最好的检测性能和最快的推理速度,为部署在移动端设备提供了可行性,具有潜在的应用价值。在未来的工作中,将对模型做进一步的优化,提升模型在复杂场景下的检测性能,同时保证模型部署到资源有限的边缘端设备或手机移动端能实现对室内常见物体精准快速的检测。

参考文献:

- [1] World Health Organization. World report on vision[R/OL]. (2019-10-08)[2024-03-11]. <https://www.who.int/publications/item/9789241516570>.
- [2] 徐波. 盲人视觉辅助导航系统设计[D]. 北京:中国科学院大学, 2020.
XU Bo. Visual aided navigation system for the blind[D]. Beijing: University of Chinese Academy of Sciences, 2020.
- [3] 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 (CVPR). Columbus: IEEE, 2014:580-587.
- [4] GIRSHICK R. Fast R-CNN[C]//IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015: 1440-1448.
- [5] 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 and Machine Intelligence, 2017, 39(6):1137-1149.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single Shot MultiBox Detector[C]//Computer Vision-ECCV. Cham: Springer International Publishing, 2016: 21-37.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016:779-788.
- [8] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2):318-327.
- [9] TIAN Z, SHEN C, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C]//IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019:9627-9636.
- [10] LAVAL M O. Tomato detection based on modified YOLOv3 framework[J]. Scientific Reports, 2021, 11(1):1-11.
- [11] 郭君斌,于琳,于传强. 改进 YOLOv5s 算法在交通标志检测识别中的应用[J]. 国防科技大学学报, 2024, 46(6):123-130.
GUO Junbing, YU Lin, YU Chuanqiang. Application of improved YOLOv5s algorithm in traffic sign detection and recognition[J]. Journal of National University of Defense Technology, 2024, 46(6):123-130.
- [12] BIST R B, SUBEDI S, YANG X, et al. A novel YOLOv6 object detector for monitoring piling behavior of cage-free laying hens[J]. AgriEngineering, 2023, 5(2):905-923.
- [13] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver: IEEE, 2023:7464-7475.
- [14] SHAFIQ M, GU Z. Deep residual learning for image recognition: a survey[J]. Applied Sciences, 2022, 12(18):8972.
- [15] HAN K, WANG Y, TIAN Q, et al. GhostNet: more features from cheap operations[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle: IEEE, 2020:1577-1586.
- [16] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City: IEEE, 2018:6848-6856.
- [17] JANG J G, QUAN C, LEE H D, et al. Falcon: lightweight and accurate convolution based on depthwise separable convolution[J]. Knowledge and Information Systems, 2023, 65(5):2225-2249.