# 基于改进YOLOv5的目标检测算法研究

邱天衡,王 玲,王 鹏,白燕娥
长春理工大学 计算机科学技术学院,长春 130022

**摘 要:** YOLOv5是目前单阶段目标检测性能较好的算法,但对目标边界回归的精确度不高,难以适用对预测框交并比要求较高的场景。基于YOLOv5算法,提出一种对硬件要求低、模型收敛速度快、目标框准确率高的新模型YOLO-G。改进特征金字塔结构(FPN),采用跨层级联的方式融合更多的特征,一定程度上防止了浅层语义信息的丢失,同时加深金字塔深度,对应增加检测层,使各种锚框的铺设间隔更加合理;其次把并行模式的注意力机制融入到网络结构中,赋予空间注意力模块和通道注意力模块相同的优先级,以加权融合的方式提取注意力信息,使网络可根据对空间和通道注意力的关注程度得到混合域注意力;通过降低网络的参数量和计算量对网络进行轻量化处理,防止因模型复杂度提升造成实时性能的损失。使用PASCAL VOC的2007、2012两个数据集来验证算法的有效性,YOLO-G比YOLOv5s的参数量减少了4.7%,计算量减少了47.9%,而mAP@0.5提高了3.1个百分点,mAP@0.5:0.95提高了5.6个百分点。

**关键词:** YOLOv5算法;特征金字塔(FPN);注意力机制;目标检测
**文献标志码:** A **中图分类号:** :TP391.4 doi:10.3778/j.issn.1002-8331.2202-0093

## Research on Object Detection Algorithm Based on Improved YOLOv5

QIU Tianheng, WANG Ling, WANG Peng, BAI Yan'e
College of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, China

**Abstract:** YOLOv5 is an algorithm with good performance in single-stage target detection at present, but the accuracy of target boundary regression is not too high, so it is difficult to apply to scenarios with high requirements on the intersection ratio of prediction boxes. Based on YOLOv5 algorithm, this paper proposes a new model YOLO-G with low hardware requirements, fast model convergence and high accuracy of target box. Firstly, the feature pyramid network(FPN) is improved, and more features are integrated in the way of cross-level connection, which prevents the loss of shallow semantic information to a certain extent. At the same time, the depth of the pyramid is deepened, corresponding to the increase of detection layer, so that the laying interval of various anchor frames is more reasonable. Secondly, the attention mechanism of parallel mode is integrated into the network structure, which gives the same priority to spatial and channel attention module, then the attention information is extracted by weighted fusion, so that the network can fuse the mixed domain attention according to the attention degree of spatial and channel attention. Finally, in order to prevent the loss of real-time performance due to the increase of model complexity, the network is lightened to reduce the number of parameters and computation of the network. PASCAL VOC datasets of 2007 and 2012 are used to verify the effectiveness of the algorithm. Compared with YOLOv5s, YOLO-G reduces the number of parameters by 4.7% and the amount of computation by 47.9%, while mAP@0.5 and mAP@0.5:0.95 increases by 3.1 and 5.6 percentage points respectively.

**Key words:** YOLOv5 algorithm; feature pyramid network(FPN); attention mechanism; object detection

目标检测是计算机视觉经久不衰的研究方向,被广泛应用在航空航天、交通、医疗、工业、农业、自动驾驶等众多领域,显著地改善着人们的日常生活。

随着大数据时代的到来以及GPU算力的不断增强,深度学习在计算机视觉各领域中逐渐展露其优势,尤其是目标检测任务。目标检测主要分为静态图像目

**基金项目:** 中央引导地方科技发展基金吉林省基础研究专项(202002038JC)。
**作者简介:** 邱天衡(2001—),男,CCF会员,研究方向为目标检测、机器学习;王玲(1979—),通信作者,女,博士,讲师,CCF会员,研究方向为图像处理、机器学习,E-mail:wangling0912@cust.edu.cn;王鹏(1973—),男,博士,教授,研究方向为数据挖掘;白燕娥(1974—),女,硕士,讲师,研究方向为图像处理、机器学习。
**收稿日期:** 2022-02-13 **修回日期:** 2022-04-02 **文章编号:** 1002-8331(2022)13-0063-11

标检测和动态视频目标检测。文献[1]给出了近年来各种图像目标检测算法及其改进方法,视频目标检测主要以图像目标检测为基础,连接循环神经网络提取复杂的时序信息,文献[2]给出了近年来的视频目标检测方法的研究与发展。从2014年开始,基于深度学习的目标检测网络井喷式爆发,先是二阶段网络,如R-CNN、Fast-RCNN[3]、Mask-RCNN[4]等,自2016年文献[5]提出you only look once(YOLOv1)以来,更轻更快的单阶段目标检测网络开始进入学者们的视野,开启了单阶段目标检测网络的新纪元。文献[6-9]均是对单阶段目标检测模型改进的研究,为各研究领域提供了更快、更好的目标检测方法,也为单阶段目标检测算法的实际应用提供了重要理论保障。2020年,YOLOv5问世,以最高140FPS的检测速度震惊世人,使其成为实时条件和移动部署环境上的理想候选者。

为了更好地提取检测目标的特征,许多优秀的卷积神经网络被应用于Backbone中,如VGG[10]、ResNet[11]等,但这些网络训练和预测的代价太大,用于YOLO网络的特征提取会使其失去实时性,无法满足工业应用的要求。随着移动端部署的需求不断增强和模型应用场景的多样化发展,许多轻量化深度神经网络应运而生。MobileNet[12]的基本单元是深度可分离卷积,把标准卷积拆分为深度卷积和点卷积,用较少的计算量获得了几乎无损的精度。ShuffleNet[13]在此基础上,利用组卷积和通道混洗来进一步减少模型参数量。最近,华为诺亚方舟实验室在CVPR 2020上提出了一种新型的端侧神经网络GhostNet[14],利用一些廉价的操作进行变换,在同等参数量的情况下,精度远高于之前的轻量化网络。

文献[15-18]针对特定领域,对YOLOv5进行了轻量化改进,但几种模型均没有对一般数据集,如COCO、PASCAL VOC等进行性能验证。同时,在实际工业应用中,发现YOLOv5s对边界框的回归不够精准,使用更深的YOLOv5m、YOLOv5l等又会受到硬件的制约,均难以满足对实时性和目标框回归准确率要求都很高的场景。为了解决这个问题,本研究基于YOLOv5提出一种针对一般数据集的轻量化和具有更高精度的目标检测模型:

(1)提出跨层加权级联的路径聚合网络(WCAL-PAN)。首先,为了防止浅层目标特征丢失,在PANet[19]结构中加入跨层级联的加权融合结构,将细节信息传递到深层网络;其次,为了获得更加丰富的语义信息,加深金字塔的深度,并对应增加Head部分的检测层,在四种尺度下进行检测,使锚框的铺设间隔更加合理;最后,为了削减上采样过程带来的特征损失,改进了上采样方法。

(2)提出改进CBAM并行注意力模块(P-CBAM)。首先对特征图同时提取空间和通道注意力特征,然后进行加权融合。并行模式的CBAM注意力结构作为一个即插即用的模块,可以插入到Backbone中的每个卷积模块后,用来提高网络的收敛速度、精确度和对目标边界的回归能力。

(3)轻量化网络。以GhostConv作为基本卷积模块,通过廉价的线性变换生成更多的特征图,使用GhostBottleneck替换掉原有的残差块,对整个检测网络进行轻量化处理,以更少的参数量、更快的速度获得更好的检测效果。

## 1　YOLOv5概述

YOLO算法基于整个图片进行预测,一次性给出所有的检测结果。经过不断更新迭代,现已推出了YOLOv5,按照模型大小递增可分为s、m、l、x,各模型仅在网络的深度和宽度上有所不同,均由输入端、Backbone、Neck、Head四部分构成。输入端使用Mosaic数据增强、自适应初始锚框计算、图片缩放等对图像进行预处理;Backbone采用了Focus下采样、改进CSP结构、SPP池化金字塔结构提取图片的特征信息;Neck主要采用FPN+PAN的特征金字塔结构,实现了不同尺寸目标特征信息的传递,解决了多尺度问题;Head采用三种损失函数分别计算分类、定位和置信度损失,并通过NMS提高网络预测的准确度。

Conv模块为复合卷积模块,是许多重要模块的基本组成部分,结构如图1所示。


**图1　Conv结构**

Fig.1　Conv structure

该模块封装了卷积层、BN层以及激活函数层。卷积层通过autopad函数实现自适应padding的效果。

Focus模块结构如图2所示。首先将输入图片按照2倍下采样切分为四部分,然后在通道维度拼接得到12维的特征图,再经过3×3的复合卷积模块进一步提取特征信息,生成32维的特征图。Focus下采样不但信息丢失少,而且通过reshape减少了卷积所带来的FLOPs,提升了网络的速度。

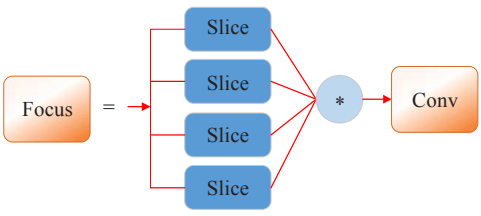



**图2　Focus结构**

Fig.2　Focus structure

Bottleneck为基本残差块,被堆叠嵌入到C3模块中进行特征学习,结构如图3所示。

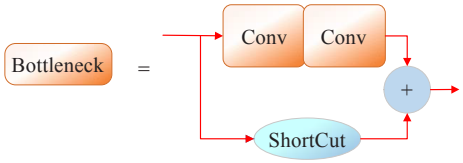



**图3　Bottleneck结构**

Fig.3　Bottleneck structure

利用两个Conv模块将通道数先减小再扩大对齐,以此提取特征信息,并使用ShortCut控制是否进行残差连接。

C3模块是改进后的BottleneckCSP模块,结构如图4所示。

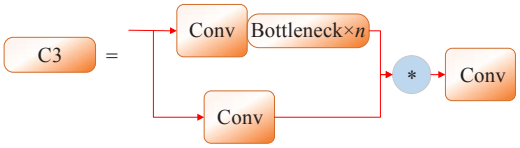



**图4　C3结构**

Fig.4　C3 structure

在C3模块中,输入特征图会通过两个分支,第一个分支先经过一个Conv模块,之后通过堆叠的Bottleneck模块对特征进行学习;另一分支作为残差连接,仅通过一个Conv模块。两分支最终按通道进行拼接后,再通过一个Conv模块进行输出。

SPP模块是空间金字塔池化模块,可以扩大感受野,结构如图5所示。

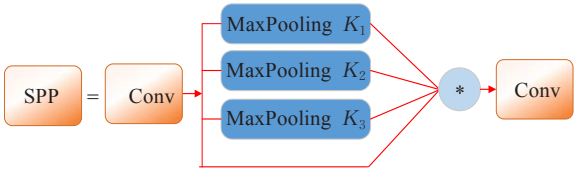



**图5　SPP结构**

Fig.5　SPP structure

先将输入特征图经过一个Conv模块减半通道数,然后分别做三种不同卷积核的最大池化下采样,再将三种池化结果与输入特征图按通道进行拼接,合并后的通道数为原来的两倍,以较小的代价最大限度地提升了感受野。

基于上述介绍,YOLOv5的基本架构如图6所示。

## 2　本文算法

### 2.1　网络整体结构

本研究基于YOLOv5提出了改进网络YOLO-G,使用WCAL-PAN和P-CBAM来提高网络的回归精度和收敛速度,并引入Ghost相关模块降低网络的复杂度。模型的网络结构如表1所示。其中,“from”表示该层模块对应的输入层,-1表示上一层。“Add”表示WCAL-PAN中跨层加权相加模块,“Ghost”表示该层引入了Ghost模

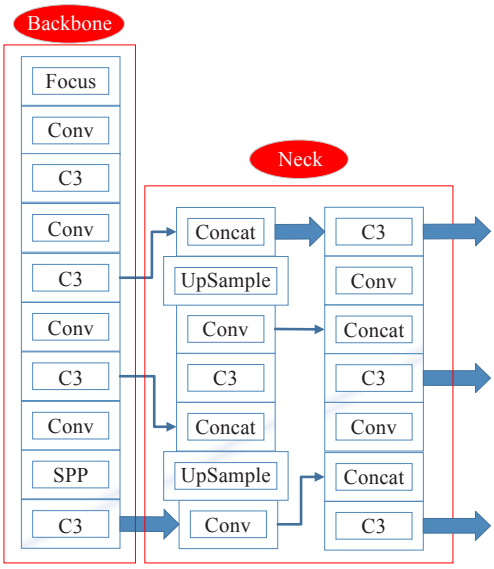



**图6　YOLOv5架构图**

Fig.6　YOLOv5 architecture diagram

**表1　YOLO-G算法架构**

Table 1　Architecture of YOLO-G algorithm

| 层数 | from | 参数量 | 模块名称 |
|---|---|---|---|
| 0 | -1 | 3 520 | Focus |
| 1 | -1 | 10 756 | GhostConv |
| 2 | -1 | 10 016 | GhostC3 |
| 3 | -1 | 40 868 | GhostConv |
| 4 | -1 | 41 840 | GhostC3 |
| 5 | -1 | 159 460 | GhostConv |
| 6 | -1 | 173 304 | GhostC3 |
| 7 | -1 | 466 468 | GhostConv |
| 8 | -1 | 325 016 | GhostC3 |
| 9 | -1 | 925 028 | GhostConv |
| 10 | -1 | 656 896 | SPP |
| 11 | -1 | 575 112 | GhostC3 |
| 12 | -1 | 103 872 | SimpleGhostConv |
| 13 | -1 | 0 | Upsample |
| 14 | [-1,8] | 0 | Concat |
| 15 | -1 | 466 512 | SimpleGhostC3 |
| 16 | -1 | 52 864 | SimpleGhostConv |
| 17 | -1 | 0 | Upsample |
| 18 | [-1,6] | 0 | Concat |
| 19 | -1 | 208 608 | SimpleGhostC3 |
| 20 | -1 | 18 240 | SimpleGhostConv |
| 21 | -1 | 0 | Upsample |
| 22 | [-1,4] | 0 | Concat |
| 23 | -1 | 53 104 | SimpleGhostC3 |
| 24 | -1 | 75 584 | SimpleGhostConv |
| 25 | [-1,20] | 0 | Concat |
| 26 | [-1,6] | 36 482 | Add |
| 27 | -1 | 143 072 | SimpleGhostC3 |
| 28 | -1 | 298 624 | SimpleGhostConv |
| 29 | [-1,16] | 0 | Concat |
| 30 | [-1,8] | 105 730 | Add |
| 31 | -1 | 368 208 | SimpleGhostC3 |
| 32 | -1 | 669 120 | SimpleGhostConv |
| 33 | [-1,12] | 0 | Concat |
| 34 | -1 | 695 744 | SimpleGhostC3 |
| 35 | [23,27,31,34] | 96 300 | Detect |

块。“Simple”标记的模块表示不添加P-CBAM机制。

## 2.2　跨层加权级联的路径聚合网络(WCAL-PAN)

深度学习的浅层网络关注细节信息,如边缘特征,在获取简单特征的基础上,可以帮助网络更准确的回归目标边界;深层网络侧重提取高级语义信息,可以提取到更加复杂的特征,能够帮助网络准确地检测出目标。FPN结构据此使用浅层特征区分简单的目标,深层特征区分复杂的目标,旨在获得鲁棒性更强的检测结果。YOLOv5的FPN结构是基于PAN的,创建了自下而上的路径增强,加速了底层信息的流动,能够很好地融合各层次的语义信息。为了进一步增强模型对浅层语义的关注度,充分融合FPN各层所提取出的语义信息,增强网络对目标边界的回归能力,本研究对YOLOv5的FPN进行改进,称为weighted connections across layers-path aggregation network(WCAL-PAN),具体改进点如下:

(1)在同一尺寸的输入、输出节点间加入跨层加权连接[20]。跨层级联结构能够有效地将浅层的细节、边缘、轮廓等信息融入到深层的网络中,可以在几乎不增加计算量的同时,融合到目标的浅层细节信息,使网络对目标边界的回归更加精准,有效提升预测框与真实框的交并比。同时,考虑到使用跨层级联时浅层特征的融入会对深层语义信息造成一定的影响,所以采用可学习的方式进行融合。以下给出本研究所使用的两种融合方式:

在特征融合过程中,由于顶层和底层的节点信息流动速度较快,经历的卷积数目较少,所以对细节信息的损失不多,为了减小模型的复杂度,所以直接采用concat操作按通道进行特征融合,过程如图7所示。

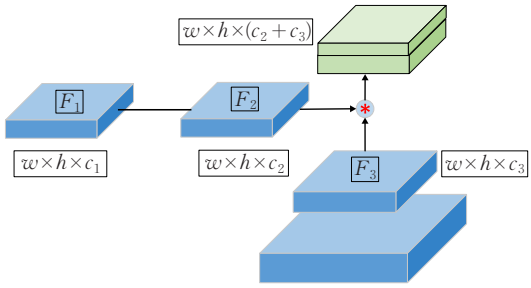



**图7　一般特征融合示意图**

Fig.7　Diagram of general feature fusion

对于其他层的节点,相邻路径上采用concat操作进行特征融合,不相邻路径上采用可学习权重的加权add操作进行特征融合,add操作既可以减少计算量,也可以减少无效浅层信息的融合。计算见公式(1):

$$Out=\sum_i\frac{\mu_i\cdot x_i}{\epsilon+\sum_j\mu_j} \tag{1}$$

式中 $x_i$ 表示每个要进行融合的特征图;$\mu_i$ 是该特征图的权重系数,可通过学习进行更新,初始的权重系数设定为1,表示两层特征图对等融合;$\epsilon$ 是一个很小的数字($\leqslant 10^{-3}$),可有效防止数值不稳定的情况。将权重标准化到0~1之间,提高训练的速度的同时,可以防止训练不稳定的情况发生。依据公式(1),对于某一中间层的特征融合方式如图8所示。

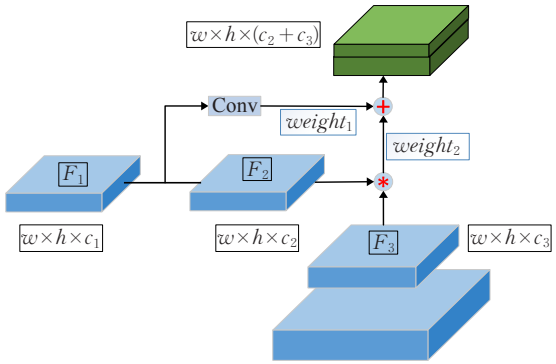



**图8　跨层级联示意图**

Fig.8　Linking diagram of across hierarchies

在图7和图8中,给定某层输入特征图 $F_1\in\mathbb{R}^{w\times h\times c_1}$,自顶向下路径对应层的特征图 $F_2\in\mathbb{R}^{w\times h\times c_2}$,自底向上路径对应层的特征图 $F_3\in\mathbb{R}^{w\times h\times c_3}$,“*”表示concat操作,“+”表示add操作,$weight_1$、$weight_2$ 分别是两条路径上特征图融合的权重值。

从两图中可以看出,顶层和底层输出节点的特征融合采用的是concat操作;而中间层节点的特征融合的过程中首先经历了concat操作,之后与经过通道对齐后的输入层进行加权add操作。最终在输出节点得到的特征图是含有细节、边缘和高级语义信息的复合特征图。为了便于理解,以中间层 $P4$ 为例,各路径上输出的计算如公式(2)、(3)所示:

$$P4^{td}=Concat(P4^{in},Resize(Conv(P5^{in}))) \tag{2}$$

$$P4^{out}=\frac{\mu_1\cdot Concat(P4^{td},Resize(Conv(P3^{out}))+\mu_2\cdot P3^{out}}{\epsilon+\mu_1+\mu_2} \tag{3}$$

式中,$Pk^{in}$ 表示第 $k$ 层的输入,$Pk^{td}$ 表示自顶向下路径中第 $k$ 层中间节点的输出,$Pk^{out}$ 表示自底向上路径中第 $k$ 层输出节点的输出。

(2)向上加深特征金字塔深度。FPN高层感受野大,包含的语义信息更高级,可以增加网络的学习能力,进一步提高检测精度。YOLOv5的FPN为3层,基于改进(1),本研究将其加深为4层,可以充分利用所提跨层级联结构。除此之外,为了匹配FPN的深度,本研究增加Detect部分的检测层,分别命名为tiny、small、medium、large,依次对 $P3$、$P4$、$P5$、$P6$ 输出的特征图进行目标检测,增加检测层之后锚框的铺设间隔变得更加合理,训练的稳定性以及模型的收敛速度和精度都会得到有效提升。

基于改进点(1)、(2),本文采用的FPN结构简化版如图9所示。

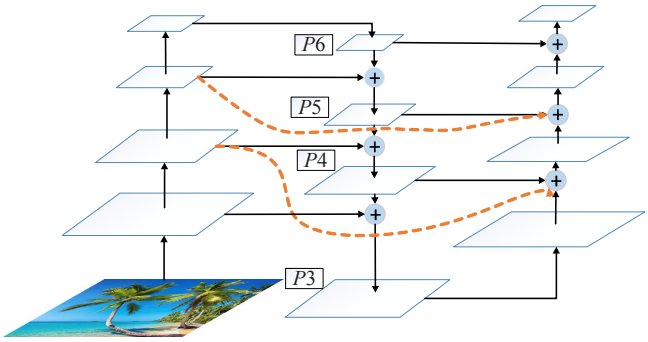



**图9　本文所用FPN结构**

Fig.9　FPN structure used in this paper

图9中橙色虚线即为跨层级联,从图中可以看出,跨层加权融合仅仅用于两个中间层 $P4$ 、$P5$ ,对于顶层 $P6$ 和底层 $P3$ ,由于信息流动造成的损失不多,考虑到模型的运行效率,本研究直接将两部分特征图按通道进行拼接。为了客观给出加深金字塔对网络造成的影响,表2给出了加深金字塔前后YOLOv5s的效果对比,其中WCAL-PAN_1表示加深金字塔后的PAN模块。

**表2　加深金字塔前后YOLOv5s的效果对比**

Table 2　Effect comparison of YOLOv5s before and after deepening pyramid

| 采用模块 | 参数量 | GFLOPs | mAP@0.5/% | mAP@0.5:0.95/% | FPS |
|---|---|---|---|---|---|
| PAN | 7 114 785 | 16.5 | 78.4 | 51.5 | 74.6 |
| WCAL-PAN_1 | **12 395 500** | **16.7** | **83.2** | **58.2** | **60.7** |

可见,加深金字塔后,虽然模型精度mAP@0.5和mAP@0.5:0.95获得了大幅提升,但参数量的大幅增加使得加载网络需要更多的显存,加大了模型训练对硬件的要求,同时也影响了模型运行的速度。为了解决这样的问题,本研究引入了Ghost系列模块对网络进行轻量化处理,在一定程度上弥补加深金字塔后网络复杂度上升所带来的负面影响。

(3)改进YOLOv5上采样方法。YOLOv5采用最邻近插值法进行上采样,该方法选用单个参考点像素值进行估计,虽然速度快、开销小,但上采样过程中会造成很严重的特征损失,降低小目标的检测精度。双线性插值法利用4个点估计插值,得到的特征图更加细腻,细节的损失更少,于是本研究将上采样方法改为双线性插值法,二者复杂度仅仅是常数级的差距,相对于精确度的提升,带来的计算开销是可以接受的。

表3为使用PAN结构和两种WCAL-PAN结构的YOLOv5s的实验精度对比,WCAL-PAN表示完全使用本研究所提出的FPN结构。从表中可以看出,使用WCAL-PAN比PAN的mAP@0.5指标提升了4.9个百分点,高交并比要求下的mAP@0.5:0.95指标提高了8.0个百分点,比加深金字塔的WCAL-PAN_1在mAP@0.5:0.95指标上提升了1.6个百分点,证明了跨层级联结构能进一步提高网络对边界的回归精度。总的来说,WCAL-PAN结构使得网络各层次语义信息融合得更加合理充分,WCAL-PAN的引入使网络精度有了大幅上升,尤其是高交并比要求下的精度进一步提高,证明网络从WCAL-PAN结构中融合到了更加有效的特征信息,可以更好地回归目标的边界框,契合高交并比下的工业目标检测任务。

**表3　WCAL-PAN和PAN在YOLOv5s下的效果对比**

Table 3　Effect comparison of WCAL-PAN and PAN in YOLOv5s

%

| 采用模块 | mAP@0.5 | mAP@0.5:0.95 |
|---|---|---|
| PAN | 78.4 | 51.5 |
| WCAL-PAN_1 | 83.0 | 57.9 |
| WCAL-PAN | **83.3** | **59.5** |

## 2.3　并行混合域卷积注意力模块(P-CBAM)

注意力机制通过给不同部分的特征图赋予权重或硬性选择部分特征图,抑制无用信息,以达到选择更优特征的目的。文献[21]结合通道和空间的信息,提出了一种混合域卷积注意力模块(convolutional block attention module,CBAM),该模块首先逐通道提取全局特征,生成通道注意力特征图,并以此作为空间注意力的输入,最终生成混合域特征图,可以有效提高模型的收敛速度和检测精度。事实上,在深度卷积神经网络中,有些层更加关注通道特征,而引入空间特征则会让网络变得敏感,甚至会产生许多非像素信息;有的层更加关注空间特征,而引入通道特征容易对网络造成过拟合的情况。但CBAM空间和通道注意力串行的信息交流方式忽略了上述特点,本研究基于此进行改进。对于不同的特征图数据,由于无法预知特征图对通道和空间的关注程度,首先赋予通道注意力模块和空间注意力模块相同的优先级,然后以可学习的加权融合方式提取混合域特征信息,以此进行特征图空间和通道的信息交流,充分利用通道和空间维度的注意力信息,称为parallel-convolutional block attention module(P-CBAM),结构如图10所示。

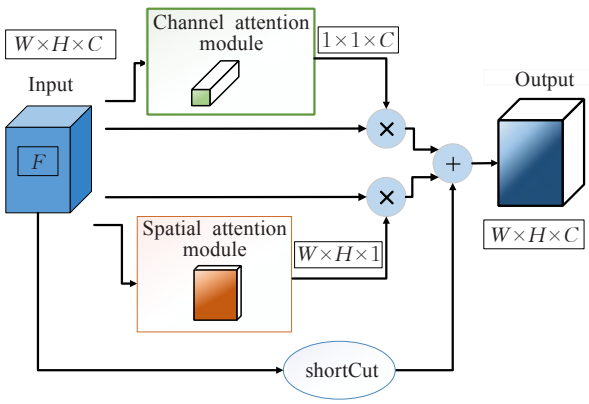



**图10　P-CBAM结构**

Fig.10　P-CBAM structure

对于输入特征图 $F \in \mathbb{R}^{W\times H\times C}$ ,通道注意力模块和空间注意力模块输出的计算如公式(4)、(5)所示:

$$M_C(F)=\sigma\left(MLP\left(AvgPool(F)+MLP\left(MaxPool(F)\right)\right)\right) \quad (4)$$

$$M_S(F)=\sigma\left(f^{7\times7}\left(Concat\left(AvgPool(F),MaxPool(F)\right)\right)\right) \quad (5)$$

式中，$M_C(F)\in\mathbb{R}^{1\times1\times C}$ 为通道注意力模块的输出，$M_S(F)\in\mathbb{R}^{W\times H\times1}$ 为空间注意力模块的输出，$\sigma$ 表示sigmoid函数，MLP表示包含两个全连接层和ReLU激活函数的多层感知机，$f^{7\times7}$ 表示一个卷积核大小为7×7的卷积运算。基于公式(4)、(5)，P-CBAM输出公式如(6)~(8)所示：

$$F_C=M_C(F)\otimes F \quad (6)$$

$$F_S=M_S(F)\otimes F \quad (7)$$

$$F_{\text{out}}=\begin{cases}\dfrac{\omega_1\cdot F_C+\omega_2\cdot F_S}{\omega_1+\omega_2}, & \text{ShortCut is False}\\ \dfrac{\omega_1\cdot F_C+\omega_2\cdot F_S+\omega_3\cdot F}{\omega_1+\omega_2+\omega_3}, & \text{ShortCut is True}\end{cases} \quad (8)$$

式中，$F_C$ 和 $F_S$ 分别为通道和空间注意力特征图，$\otimes$ 表示元素乘法，在该过程中，$M_C$ 和 $M_S$ 被沿着通道和空间维度进行广播。$F_{\text{out}}$ 是对两种类型的特征图进行加权融合的结果。本研究通过ShortCut控制残差连接，使用简便的归一化除法保证训练的稳定性。P-CBAM从对等的角度获取一维通道和二维空间的注意力信息，能够更加有针对性地提取图片特征，提升图像识别效果，以下从定性定量两个角度证明P-CBAM的有效性。

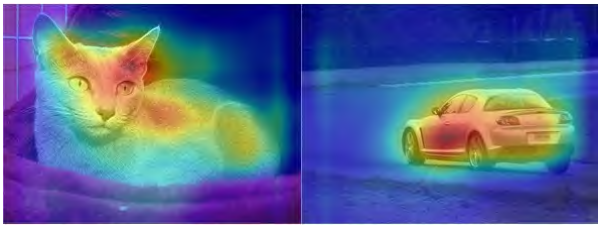
表4为YOLOv5s与加入各种注意力机制后在VOC2007测试集下的精度对比，从表中可以看出加入注意力机制普遍能够提升网络的精度。原始算法在加入CBAM后造成了精度下降，而加入P-CBAM后，在mAP@0.5指标上比最优的ECA机制仅差0.2个百分点，相比原始算法提高0.6个百分点；在mAP@0.5:0.95指标上，P-CBAM在4种注意力机制中获得了最优的效果，相比原始算法提高1.8个百分点，更适用于高交并比下的目标检测任务，证明赋予空间和通道注意力机制相同的优先级并以加权的方式提取注意力信息是对CBAM模块有效的改进方法。

**表4　各种注意力机制与P-CBAM在YOLOv5s下的对比**

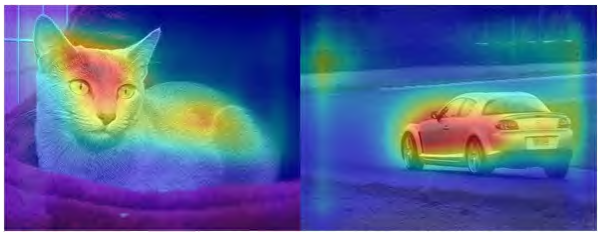
Table 4　Comparison of various attentional mechanisms and P-CBAM in YOLOv5s　%

| 算法 | mAP@0.5 | mAP@0.5:0.95 |
|---|---|---|
| YOLOv5s | 78.4 | 51.5 |
| YOLOv5s+SENet | 78.6 | 52.0 |
| YOLOv5s+CBAM | 78.3 | 51.2 |
| YOLOv5s+ECA | 79.2 | 52.4 |
| YOLOv5s+P-CBAM | **79.0** | **53.3** |

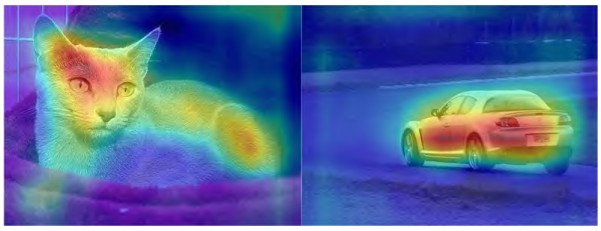
如图11为特征图经过各种注意力机制处理后的加权热力图。从图中可以看出，和其他3种主流注意力机制相比，加入P-CBAM后，网络对检测目标区域的覆盖度和关注程度都获得了提升，证明P-CBAM能够帮助深度卷积网络提取到更加关键的特征信息。



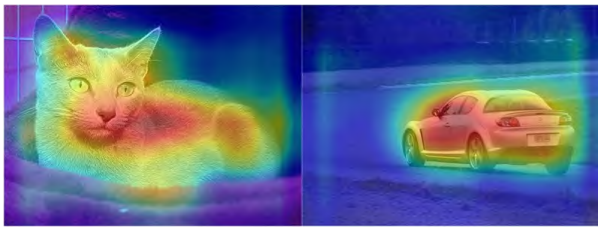
(a)SE热力图



(b)CBAM热力图



(c)ECA热力图



(d)P-CBAM热力图

**图11　各种注意力机制与P-CBAM热力对比**

Fig.11　Thermal contrast of various attentional mechanisms and P-CBAM

此外，为了验证P-CBAM注意力机制在不同类型目标检测问题上的普适性，表5给出了加入P-CBAM前后YOLOv5s在VOC数据集、SKU-110K数据集、Argoverse数据集、VisDrone2019数据集上的检测效果，实验结果格式为mAP@0.5/mAP@0.5:0.95。4种数据集涵盖了各种目标分布类型，除了比较容易检测的大目标外，还包括了一些目标检测领域的重难点问题，如小目标、模糊目标、密集目标、形态多样目标等。

**表5　P-CBAM在多个数据集下的检测效果**

Table 5　Detection effects of P-CBAM in multiple datasets　%

| 数据集 | YOLOv5s | YOLOv5s+P-CBAM |
|---|---|---|
| VOC | 78.4/51.5 | 79.0/53.3 |
| SKU-110K | 87.2/52.9 | 87.5/53.3 |
| Argoverse | 34.6/19.1 | 35.9/19.7 |
| VisDrone | 30.2/15.6 | 30.6/15.9 |

从表中可以看出，加入P-CBAM注意力机制后的YOLOv5s相比原始算法在所有数据集上的精度都获得了一定提升，证明P-CBAM模块对于各种目标检测任务

的普适性。

## 2.4　网络结构轻量化

GhostNet的基本思想是根据特征图之间的联系，把一般卷积拆分为两步，图12给出一般卷积和Ghost卷积的对比示意图。

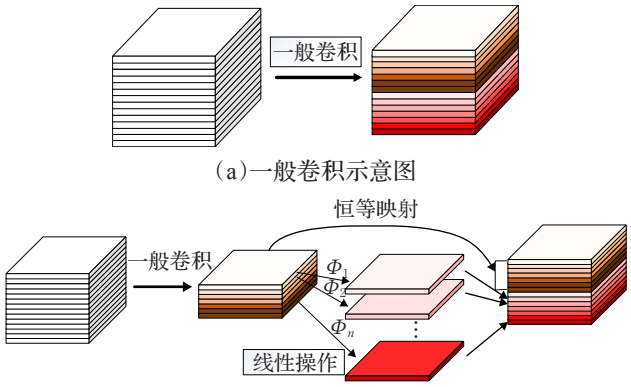



**图12　一般卷积与Ghost卷积对比图**

Fig.12　Contrast diagram of general and Ghost convolution

Ghost模块包含一个少量卷积、一个总体恒等映射和 $m\times(s-1)$ 个线性运算。首先通过一般卷积生成少量特征图，然后将第一步得到的特征图进行廉价线性操作生成Ghost特征图，最后将两组特征图按通道拼接，生成足够多的特征图以匹配给定的输出通道数。

对于输入 $X\in\mathbb{R}^{W\times H\times C}$，一般卷积的输出 $Y\in\mathbb{R}^{W'\times H'\times n}$ 可以表示为 $Y=X\cdot f+b$，其中 $f\in\mathbb{R}^{k\times k\times C\times n}$ 表示卷积核大小为 $k\times k$ 的 $C\times n$ 个卷积运算，$b$ 表示偏置项。一般卷积的FLOPs可表示为 $W'\cdot H'\cdot n\cdot k\cdot k\cdot C$。Ghost卷积采用分步策略，计算如公式(9)、(10)所示：

$$Y'=X\cdot f' \tag{9}$$

$$Y_{ij}=\Phi_{i,j}\cdot Y'_i,\ \ i\in[1,m],j\in[1,s] \tag{10}$$

其中少量卷积结果 $Y'\in\mathbb{R}^{W'\times H'\times m}$ 表示对输入 $X$ 经过一般卷积 $f'\in\mathbb{R}^{k\times k\times C\times m}$ 后生成的 $m$ 个特征图 $(m\ll n)$；之后将 $m$ 个特征图逐个进行线性操作，每个特征图均生成 $s$ 个特征图，共生成 $n=m\times s$ 个特征图。$\Phi_{i,j}$ 表示对第一步卷积中生成的第 $i$ 个特征图 $Y'_i$ 进行第 $j$ 个线性操作，$\Phi_{i,s}$ 表示一个直接的特征恒等映射。为了保证CPU或GPU的高效性和实用性，设每个线性操作的卷积核大小均为 $d\times d$，则一般卷积和Ghost卷积的速度比可用公式(11)进行计算：

$$Rate_S=\frac{W'\cdot H'\cdot n\cdot k\cdot k\cdot C}{W'\cdot H'\cdot m\cdot k\cdot k\cdot C+W'\cdot H'\cdot(n-m)\cdot d\cdot d}=\frac{W'\cdot H'\cdot n\cdot k\cdot k\cdot C}{W'\cdot H'\cdot\frac{n}{s}\cdot k\cdot k\cdot C+W'\cdot H'\cdot(s-1)\cdot\frac{n}{s}\cdot d\cdot d}=\frac{k\cdot k\cdot C}{\frac{1}{s}\cdot k\cdot k\cdot C+\frac{(s-1)}{s}\cdot d\cdot d}\approx\frac{s\cdot C}{s+C-1}\approx s \tag{11}$$

由化简结果可得一般卷积的计算量大致为Ghost卷积的 $s$ 倍，同理可计算出参数量也近似为 $s$ 倍。Ghost卷积是一个更轻、更快的模块，本研究以此为基础，使用Ghost卷积替换了YOLOv5中的部分一般卷积，替换后的Conv、Bottleneck和C3这3种主要模块结构如图13所示。

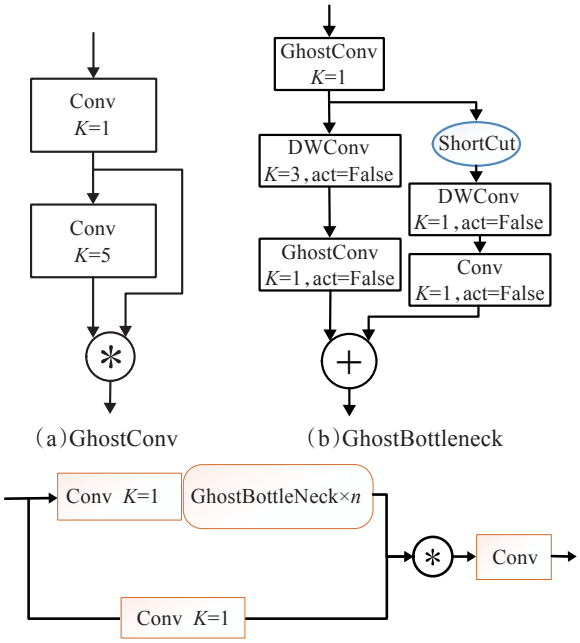



**图13　Ghost系列模块**

Fig.13　Ghost series modules

图13中，$K$ 表示卷积核大小，act表示是否有非线性激活函数层，act=False表示未含有非线性激活函数，DWConv为逐通道卷积。轻量化后的模型在保证准确率降低最少的条件下，大大减少了参数量和计算量，提升了网络的运行速度，表6为输入尺寸为640×640的图像在全部使用Ghost系列模块替换后的网络与YOLOv5s模型的对比结果。

**表6　YOLOv5s和YOLO-G的结果对比**

Table 6　Results comparison of YOLOv5s and YOLO-G

| 算法 | 参数量 | GFLOPs | GPU-speed/ms | 模型大小/$10^6$ |
|---|---|---|---|---|
| YOLOv5s | 7 114 785 | 16.5 | 1.7 | 14.2 |
| 本文算法 | **3 419 817** | **7.8** | **1.4** | **6.8** |

由表6可以看出，替换后网络计算量减少了52.7%，参数量减少了51.9%，模型大小减少了52%，目标检测推理速度提升了18%。实验结果证明了使用Ghost模块对网络进行轻量化的有效性，而参数量和计算量的大幅降低能够有效减小模型训练和预测对硬件的要求，使模型更适配于实际的工业应用。Ghost结构使得网络的复杂度得以降低，能够弥补P-CBAM和WCAL-PAN引入后所带来的计算量和参数量的上升。

## 3　实验结果及分析

### 3.1　实验环境与参数设置

具体实验环境如表7所示。

**表7　实验环境**

Table 7　Experimental environment

| 参数 | 实验环境 |
|---|---|
| 操作系统 | Windows10 64位 |
| CPU | 11th Gen Intel® Core™ i7-11700F CPU @ 2.50 GHz |
| GPU | GeForce RTX 3090(24.0 GB) |
| 内存 | 64 GB |
| Python | 3.8 |
| 深度学习框架 | Pytorch1.8.0、CUDA11.1 |

所有YOLO模型初始学习率为0.01,动量参数为0.92,最终学习率为0.2,BatchSize为64,并采用epoch为3,动量参数为0.8的warm-up方法预热学习率。在warm-up阶段,采用一维线性插值更新学习率,预热结束后采用余弦退火对学习率进行更新。

### 3.2　数据集及预处理

本研究使用PASCAL VOC 2007和2012数据集进行实验,该数据集包含4个大类和20个小类。训练集选用VOC 2007和2012数据集中的train+val部分,共16 551张图片;测试集选用VOC 2007数据集中的test部分,共4 952张图片。

首先对数据集标签格式进行转化,将PASCAL VOC格式转化为YOLO的文本格式,生成id、x、y、w、h,并进行归一化,然后把转换结果按训练集和验证集的方式进行存储。

其次,对640×640大小的图片,采用$k$-means算法重新对数据集中的目标框进行聚类,并对聚类产生的所有anchors使用genetic algorithm算法进行2 000轮变异,保留得到最佳召回率最大的结果,得到的最终anchors如表8所示。

**表8　聚类后的anchors大小**

Table 8　Anchors size after clustering

| 对应检测层 | 类型1 | 类型2 | 类型3 |
|---|---|---|---|
| tiny | 32,39 | 35,90 | 75,57 |
| small | 66,130 | 124,105 | 114,193 |
| medium | 217,164 | 154,319 | 259,295 |
| large | 473,226 | 332,443 | 536,402 |

本研究使用随机增减亮度、对比度、饱和度、色调,以及随机缩放、裁减、翻转、擦除等基本方法和Mosaic方法进行数据增强。Mosaic数据增强选取4张经过进行上述基本方法操作后的图片进行随机编排和拼接,一方面变相地增大了BatchSize,降低显存占用率,另一方面扩充了原始数据集,防止过拟合的发生,提高了模型的整体鲁棒性,数据增强后效果如图14所示。


**图14　数据增强后的效果图**

Fig.14　Rendering after enhanced data

### 3.3　评价指标

本研究使用mAP@0.5、mAP@0.5:0.95、参数量、计算量(GFLOPs)、模型大小、FPS作为目标检测模型的衡量标准。其中,mAP@0.5代表在IOU阈值为0.5时的平均AP;mAP@0.5:0.95代表在IOU阈值为从0.5到0.95,步长为0.05时各个mAP的平均值。

参数量和计算量指标分别从时间和空间上衡量模型的复杂度,mAP@0.5主要用于体现模型的识别能力,mAP@0.5:0.95由于要求的IOU阈值更高,主要用于体现定位效果以及边界回归能力,前5个评价指标的值均与检测效果呈正相关,FPS表示每秒检测图像的数目,值越大检测速度越快。

mAP计算中的AP即P-R曲线下方面积,具体计算基于公式(12)~(14):

$$Precision=\frac{TP}{TP+FP} \tag{12}$$

$$Recall=\frac{TP}{TP+FN} \tag{13}$$

$$AP=\int_0^1 P(R)\mathrm{d}R \tag{14}$$

式中,TP代表IOU大于等于设定阈值的检测框数目,FP代表代表IOU小于设定阈值的检测框数目,FN代表漏检的目标数目。

### 3.4　定量评价

为了验证所提算法YOLO-G的有效性,本研究将其与一些轻量化检测模型和主流的单阶段、二阶段目标检测模型进行了比较,实验结果如表9所示。可见,YOLO-G相比大部分网络不仅复杂度更小,而且精度上也有着显著优势,尤其是mAP@0.5:0.95指标,几乎是所有对比模型中最优的;相比于相同Baseline的轻量化的网络YOLOv5-mobileNet、YOLOv5-ShuffleNet,虽然模型复杂度稍大,但是检测精度方面有着明显的优势;改进后的网络在GPU下的FPS达到51.3,由于WCAL-PAN模块、P-CBAM模块的加入,以及轻量化depth-wise卷积的大量引入,使得YOLO-G的检测速度未达到最优,但也可以满足实时检测任务的需求。而网络参数量和计算量的进一步优化,使得模型时间、空间复杂度大大减少,让模型训练和检测时对硬件的要求更低,可以被广

表9 对比实验

Table 9 Contrast experiments

| 算法 | 输入尺寸 | 参数量/$10^6$ | GFLOPs | 模型大小/$10^6$ | mAP@0.5/% | mAP@0.5:0.95/% | FPS |
|---|---|---|---|---|---|---|---|
| YOLOv5s | 640×640 | 7.100 | 16.50 | 14.2 | 78.4 | 51.5 | 74.6 |
| YOLOv4-tiny | 640×640 | 5.920 | 16.22 | 22.6 | 58.9 | 26.2 | 99.8 |
| YOLOv4-mobileNetv2 | 640×640 | 12.170 | 18.26 | 46.9 | 82.0 | 46.8 | 50.9 |
| YOLOv4-mobileNetv3 | 640×640 | 14.070 | 16.95 | 54.1 | 81.5 | 46.3 | 47.2 |
| YOLOv4-ghostNet | 640×640 | 11.110 | 15.69 | 42.8 | 80.8 | 45.4 | 39.3 |
| YOLOv5-ShuffleNet | 640×640 | 0.455 | 6.20 | 1.3 | 43.5 | 20.3 | 78.7 |
| YOLOv5-mobileNet | 640×640 | 2.790 | 5.60 | 7.4 | 61.2 | 29.0 | 73.5 |
| SSD[22] | 618×618 | 41.180 | 387.90 | 200.0 | 76.7 | — | 48.5 |
| YOLOv3[23] | 640×640 | 61.950 | 156.40 | 117.0 | 82.4 | 57.4 | 55.7 |
| Faster[24] | 1 000×600 | 60.170 | 523.80 | 460.0 | 79.7 | — | 11.6 |
| Cascade[25] | 1 000×600 | 87.980 | 543.80 | 672.0 | 79.1 | — | 10.4 |
| Grid[26] | 1 000×600 | 83.250 | 766.70 | 636.0 | 80.2 | — | 7.5 |
| YOLO-G | 640×640 | 6.780 | 8.60 | 12.8 | 81.5 | 57.1 | 51.3 |

泛应用于一些低成本的工业检测问题。在13组对比实验中,YOLO-G在模型复杂度升序中位列第3,在mAP@0.5指标降序排列中位列第3,与top1仅仅相差0.9个百分点,在mAP@0.5:0.95指标降序排列中位列第2,与top1仅仅相差0.3个百分点。结合模型复杂度和实际应用效果,从总体上看,在高交并比需求的工业任务中,YOLO-G在众多模型中的表现更加出色。

### 3.5 定性评价

本研究还使用了3组场景的图片对YOLOv5和YOLO-G的检测效果进行定性评价,所有实验输入图片大小均为640,置信度阈值为0.25,NMS阈值为0.45,实验结果如图15所示。

第1组实验图片的先验目标数量较少,此时YOLOv5出现了大量漏检的情况,而YOLO-G检测出了更多正确的目标,证明YOLO-G能提取出更丰富的特征;第2组实验图片中的目标较密集、遮挡较严重,YOLOv5漏检的数量进一步增多,且出现了一定的误检框,而YOLO-G的检测精度并未下降;第3组图片中部分目标的特征较为模糊,识别很困难,但YOLO-G的检测效果依然比较出色。总体来说,YOLO-G对正确预测框的置信度和交并比都普遍高于YOLOv5,证明网络提取到了更加丰富的语义信息,表现出了更好的性能。

### 3.6 消融实验

为了进一步验证本研究所提算法的检测性能,探究各个改进方法的有效性,在YOLOv5s的基础上设计了8组消融实验,每组实验使用相同的超参数以及训练技巧,实验结果如表10所示。

其中,WCAL-PAN代表所提FPN结构,P-CBAM代表本研究所提注意力机制,Ghost代表引入Ghost系列模块,“√”代表引入模块,组别7中WCAL-PAN下“√”代

(a)YOLOv5实验效果图

(b)YOLO-G实验效果图

图15 YOLOv5和YOLO-G效果对比

Fig.15 Comparison of YOLOv5 and YOLO-G effects

表10 消融实验

Table 10 Ablation experiments

| 组别 | P-CBAM | WACL-PAN | Ghost | 参数量 | GFLOPs | mAP@0.5/% | mAP@0.5:0.95/% |
|---|---|---|---|---|---|---|---|
| 1 | × | × | × | 7 114 785 | 16.5 | 78.4 | 51.5 |
| 2 | √ | × | × | 7 300 173 | 16.8 | 79.0 | 53.3 |
| 3 | × | √ | × | 12 659 184 | 16.9 | 83.3 | 59.5 |
| 4 | × | × | √ | 3 419 817 | 6.8 | — | — |
| 5 | √ | × | √ | 3 485 337 | 7.8 | 77.6 | 52.0 |
| 6 | × | √ | √ | 6 371 768 | 8.2 | 80.4 | 55.5 |
| 7 | √ | √ | √ | 6 413 156 | 8.2 | 80.7 | 56.3 |
| 8 | √ | √ | √ | 6 780 348 | 8.6 | 81.5 | 57.1 |

表11 小目标消融实验

Table 11 Small target ablation experiments

%

| 算法 | bird | bottle | plant | chair | boat | 总体 |
|---|---|---|---|---|---|---|
| YOLOv5s | 74.8/44.1 | 69.1/43.5 | 50.8/24.1 | 62.8/38.4 | 69.4/38.5 | 65.4/37.7 |
| YOLOv5s+P-CBAM | 72.9/44.0 | 69.1/42.7 | 52.1/25.5 | 62.9/38.4 | 70.8/44.0 | 65.6/39.0 |
| YOLOv5s+WCAL-PAN | 81.1/53.4 | 70.6/45.4 | 59.6/29.9 | 66.7/42.5 | 73.0/42.9 | 70.2/42.8 |
| YOLO-G | 78.8/51.0 | 66.1/40.1 | 60.1/30.4 | 62.8/37.9 | 72.1/42.7 | 68.0/40.4 |

表不引入跨层加权连接结构。由于引入Ghost模块是为了进行网络轻量化,所以不对单独引入Ghost模块后的网络计算mAP。

从表10中可以看出,Ghost系列模块的引入,使网络计算量减少了52.7%,参数量减少了51.9%,是有效的轻量化手段;加入P-CBAM模块后mAP@0.5提升0.6个百分点,mAP@0.5:0.95提升1.8个百分点,虽然对网络精度提升不多,但其几乎不增加网络的参数量和计算量。同时在引入Ghost模块降低模型复杂度后,虽然mAP@0.5指标下降了0.8个百分点,但是mAP@0.5:0.95依然比之前提升了0.5个百分点,证明P-CBAM模块的引入,提高了模型对目标边界的回归能力,使预测出的目标框更加贴合物体的轮廓,对于一些对IOU要求较高、需要更准确定位物体的任务来说,加入P-CBAM是非常有效的;其次,WCAL-PAN对网络精度的提升是最多的,mAP@0.5提升4.9个百分点,mAP@0.5:0.95提升8个百分点,但网络结构比原始YOLOv5s复杂,对硬件的要求较高,并且牺牲了一些实时性,引入Ghost模块后,在mAP@0.5提升2.0个百分点,mAP@0.5:0.95提升4.0个百分点的情况下,不仅使网络参数量有所降低,并且使计算量减少为原来的50.3%;无论是P-CBAM或者WCAL-PAN都会增加网络复杂度,尤其是使用WCAL-PAN后,网络的参数量增加了44.9%,而引入Ghost可以有效降低网络复杂度,大大地减少计算量,达到速度和精度两方的平衡,最终改进后的模型相比YOLOv5s,参数量减少了4.7%,计算量减少了47.9%,而mAP@0.5提高了3.1个百分点,mAP@0.5:0.95提高了5.6个百分点,模型对目标框的拟合能力进一步加强,且网络运行时对硬件要求更小,可以被广泛应用于一些对目标框IOU要求较高,需要定位得更加准确的工业任务。

另外,针对比较难分辨的小目标,设计了4组消融实验,结果如表11所示,表中实验结果格式为mAP@0.5/mAP@0.5:0.95。从表中可以看到,WCAL-PAN对小目标提升效果较大,mAP@0.5提高了3.3个百分点,mAP@0.5:0.95提高了5.1个百分点,而P-CBAM的作用更多的体现在边框回归精度的上,mAP@0.5:0.95提升了1.3个百分点。最终本文所提模型YOLO-G在小模型检测上,mAP@0.5提高了2.6个百分点,mAP@0.5:0.95提高了2.7个百分点,使模型对小目标检测效果也得到了有效提升。

## 4 结语

本文研究基于YOLOv5提出了一种改进的目标检测算法YOLO-G。采用WCAL-PAN、P-CBAM结构对网络的精度和目标框边界的回归能力进行提升;使用Ghost模块对网络进行轻量化处理,填补WCAL-PAN和P-CBAM模块引入后对网络实时性能造成的损失。YOLO-G和YOLOv5s相比,参数量减少了7.9%,计算量减少了49%,而mAP@0.5提高了3.1个百分点,mAP@0.5:0.95提高了5.6个百分点。但为了减少模型复杂度,实现移动端目标检测,接下来将对网络进行剪枝、蒸馏等处理,进一步对模型进行轻量化;另外,Head阶段中上采样部分仍可继续改进,如果使用一些复杂度较低的图像超分算法,可以更好地检测出小目标。

## 参考文献:

[1] 罗会兰,陈鸿坤.基于深度学习的目标检测研究综述[J].电子学报,2020,48(6):1230-1239.
LUO H L,CHEN H K.Survey of object detection based on deep learning[J].Acta Electronica Sinica,2020,48(6):1230-1239.

[2] 王迪聪,白晨帅,邬开俊.基于深度学习的视频目标检测综

述[J].计算机科学与探索,2021,15(9):1563-1577.
WANG D C,BAI C S,WU K J.Survey of video object detection based on deep learning[J].Journal of Frontiers of Computer Science and Technology,2021,15(9):1563-1577.

[3] GIRSHICK R.Fast R-CNN[C]//Proceedings of IEEE International Conference on Computer Vision(ICCV),2015:1440-1448.

[4] HE K,GKIOXARI G,DOLLÁR P,et al.Mask R-CNN[C]//Proceedings of IEEE International Conference on Computer Vision(ICCV),2017:2980-2988.

[5] REDMON J,DIVVALA S,GIRSHICK R,et al.You only look once:Unified,real-time object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2016:779-788.

[6] 王燕妮,余丽仙.注意力与多尺度有效融合的SSD目标检测算法[J].计算机科学与探索,2022,16(2):438-447.
WANG Y N,YU L X.SSD object detection algorithm with effective fusion of attention and multiscale[J].Journal of Frontiers of Computer Science and Technology,2022,16(2):438-447.

[7] 沈震宇,朱昌明,王喆.基于MAML算法的YOLOv3目标检测模型[J].华东理工大学学报(自然科学版),2022,48(1):112-119.
SHEN Z Y,ZHU C M,WANG Z.YOLOv3 object detection model based on MAML algorithm[J].Journal of East China University of Science and Technology,2022,48(1):112-119.

[8] 谭显东,彭辉.改进YOLOv5的SAR图像舰船目标检测[J].计算机工程与应用,2022,58(4):247-254.
TAN X D,PENG H.Improved YOLOv5 ship target detection in SAR image[J].Computer Engineering and Applications,2022,58(4):247-254.

[9] 王兵,乐红霞,李文璟,等.改进YOLO轻量化网络的口罩检测算法[J].计算机工程与应用,2021,57(8):62-69.
WANG B,LE H X,LI W J,et al.Mask detection algorithm based on improved YOLO lightweight network[J].Computer Engineering and Applications,2021,57(8):62-69.

[10] SIMONYAN K,ZISSERMAN A.Very deep convolutional networks for large-scale image recognition[J].arXiv:1409.1556,2014.

[11] HE K,ZHANG X,SUN S R A J.Deep residual learning for image recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2016:770-778.

[12] HOWARD A G.MobileNets:Efficient convolutional neural networks for mobile vision applications[J].arXiv:1704.04861,2017.

[13] ZHANG X,ZHOU X,SUN M L.ShuffleNet:An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition,2018:6848-6856.

[14] HAN K,WANG Y,TIAN Q,et al.GhostNet:More features from cheap operations[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR),2020:1577-1586.

[15] 杨小冈,高凡,卢瑞涛,等.基于改进YOLOv5的轻量化航空目标检测方法[J/OL].信息与控制:1-7(2021-09-27)[2022-01-22].http://kns.cnki.net/kcms/detail/21.1138.TP.20210927.1729.002.htm.
YANG X G,GAO F,LU R T,et al.Lightweight aerial object detection method based on improved YOLOv5[J/OL].Information and Control:1-7(2021-09-27)[2022-01-22].http://kns.cnki.net/kcms/detail/21.1138.TP.20210927.1729.002.htm.

[16] 林森,刘美怡,陶志勇.采用注意力机制与改进YOLOv5的水下珍品检测[J].农业工程学报,2021,37(18):307-314.
LIN S,LIU M Y,TAO Z Y.Detection of underwater treasures using attention mechanism and improved YOLOv5[J].Transactions of the Chinese Society of Agricultural Engineering,2021,37(18):307-314.

[17] 彭成,张乔虹,唐朝晖,等.基于YOLOv5增强模型的口罩佩戴检测方法研究[J].计算机工程,2022(4):39-49.
PENG C,ZHANG Q H,TANG Z H,et al.A face mask wearing detection method based on YOLOv5 enhancement model[J].Computer Engineering,2022(4):39-49.

[18] 钱坤,李晨瑄,陈美杉,等.基于YOLOv5的舰船目标及关键部位检测算法[J].系统工程与电子技术,2022(6):1823-1832.
QIAN K,LI C X,CHEN M S,et al.Ship target and key parts detection algorithm based on YOLOv5[J].Systems Engineering and Electronics,2022(6):1823-1832.

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

[20] TAN M X,PANG R.EfficientDet:Scalable and efficient object detection[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR),2020:10778-10787.