

# 基于 YOLO 的钢缆表面损坏检测<sup>①</sup>

刘际驰, 吕后坤, 李 伟

(大连工业大学, 大连 116034)

通信作者: 李 伟, E-mail: [leeway0504@hotmail.com](mailto:leeway0504@hotmail.com)



**摘 要:** 为了解决检测钢缆表面损坏时检测设备资源有限、时间过长等问题, 将深度学习的先进技术以及卷积神经网络 (CNN) 应用于钢缆表面损坏检测. 提出了一种基于 YOLO 的缺陷检测网络模型, 将 GhostNet 融入主干网络, 并基于 ShuffleNet 及注意力机制提出了新的特征提取模块 (ShuffleC3), 再对 Head 部分进行剪枝改进. 实验结果表明, 改进后网络相比基线 YOLOv5s 的平均精度提高 1.1%, 参数量和计算量分别降低了 43.4% 和 31%, 模型大小减少了 42.3%. 可以在降低网络计算成本的同时, 保持较高的识别精确度, 更好地满足了对钢缆材料表面损坏检测的要求.

**关键词:** 深度学习; 卷积神经网络; YOLO; 钢缆; 注意力机制; 表面损坏检测; 目标检测

引用格式: 刘际驰, 吕后坤, 李伟. 基于 YOLO 的钢缆表面损坏检测. 计算机系统应用, 2024, 33(1): 134-140. <http://www.c-s-a.org.cn/1003-3254/9364.html>

## YOLO-based Surface Damage Detection of Steel Cables

LIU Ji-Chi, LYU Hou-Kun, LI Wei

(Dalian Polytechnic University, Dalian 116034, China)

**Abstract:** To solve the limited resources and long time of detection equipment in detecting surface damage of steel cables, this study applies advanced technology of deep learning and convolutional neural networks (CNNs) to surface damage detection of the cables. On this basis, it proposes a YOLO-based defect detection network model to integrate GhostNet into the backbone network, and a new feature extraction module (ShuffleC3) based on ShuffleNet and attention mechanism, and then prunes and improves the Head part. Experimental results show that compared with the baseline YOLOv5s, the average accuracy of the improved network is increased by 1.1%. In addition, the number of parameters and calculations are reduced by 43.4% and 31% respectively, and the model size is reduced by 42.3%. Thus, the proposed model can reduce the network computing cost and maintain higher identification accuracy, which better meets the requirements for surface damage detection of steel cable materials.

**Key words:** deep learning; convolutional neural network (CNN); YOLO; steel cable; attention mechanism; surface damage detection; object detection

钢缆是由多根多股钢丝绳围绕一根纤维芯或钢丝绳芯捻制而成的钢丝束, 在实际使用中会出现损坏等问题造成经济损失或人员伤亡, 因此及时检测识别到钢缆的损坏尤为重要. 目标检测方法分为单阶段和双阶段两种, 各自代表性的算法分别是 YOLO 与 R-CNN,

其中 YOLO 在运行速度与对较小目标检测上拥有更好的性能, 故本文使用 YOLO 算法对钢缆表面的损坏进行检测识别实验.

YOLO 目标检测算法是一种常用的单阶段目标检测算法, 其版本众多, 有 YOLOv3<sup>[1]</sup>、YOLOv5、

① 收稿时间: 2023-07-01; 修改时间: 2023-08-11; 采用时间: 2023-08-24; csa 在线出版时间: 2023-11-24

CNKI 网络首发时间: 2023-11-28

YOLOv7<sup>[2]</sup>等,其中YOLOv5的性能最为优秀.YOLOv5在YOLOv4<sup>[3]</sup>的基础上进行了一部分改进,使性能更佳,且目标检测方面的算法和文献大多都以YOLOv5为基础进行开展,使用五代网络进行改进方便进行对照和比较,近几年推出的版本在实质上与YOLOv5相差不多,因此本文采用YOLOv5进行研究与改进.

YOLOv5网络主要可以分为以下4个部分:输入端、Backbone主干、Neck网络、Head检测输出端.

输入端主要对输入图像进行处理,将输入图像统一尺寸、归一化等操作.通过Mosaic操作对输入数据进行增强,将随机的4张输入图片进行随机缩放、拼接、裁剪,这样的操作可以丰富检测数据集,增加了网络的鲁棒性,并减小了Mini-batch值,减小了GPU负担,再通过自适应锚框计算与自适应图片缩放对接下来的目标检测进行性能提升.

Backbone主干部分主要由Conv模块,C3模块和SPPF模块组成,主干网络的目的是提取输入图片中的图像特征,以方便后续的目标检测工作,其中C3模块可以增加网络的深度和感受野,提升特征提取的性能,SPPF模块主要由最大池化部分和常规卷积所组成,其可以从同一特征图中实现不同尺度的特征提取,并生成三尺度特征图,有助于提高检测精度.

Neck网络部分采用了FPN<sup>[4]</sup>特征金字塔与PANet<sup>[5]</sup>相结合.通过进行自顶向下和自下而上的网络进行特征融合,并将Backbone主干与Head检测端相结合,进一步提升检测能力.

Head检测端主要用于预测特征图上不同大小的目标,YOLOv5与YOLOv4相同,使用了多尺度的检测头,对不同尺寸的特征图都有很好的检测效果.

在钢缆目标检测的实际应用中,会出现检测设备资源有限、检测时间过长等问题,故本文对YOLO进行优化,减小模型大小与参数数量,并同时保持较高的检测性能.通过减小模型复杂度和计算开销,达到缓解上诉实际应用问题的目的.

## 1 实验方法及改进

本文网络将Backbone中的卷积模块替换为拥有更高的性能和更少的参数数量的GhostConv模块;在C3模块的基础上,将ShuffleNet和注意力机制与原本的C3模块进行结合,提出了ShuffleC3模块,使网络拥有更高的检测性能和更轻量的结构,最后对Head检测

端进行改进.

由于相比其他五代网络,YOLOv5s网络拥有更为轻量的结构和更低的运算成本,为了追求更好的检测性能度和更低的运算成本之间的平衡,本文在YOLOv5s网络的基础上进行改进.YOLOv5网络结构如图1所示.接下来详细介绍本文提出的GhostConv模块和ShuffleC3模块.

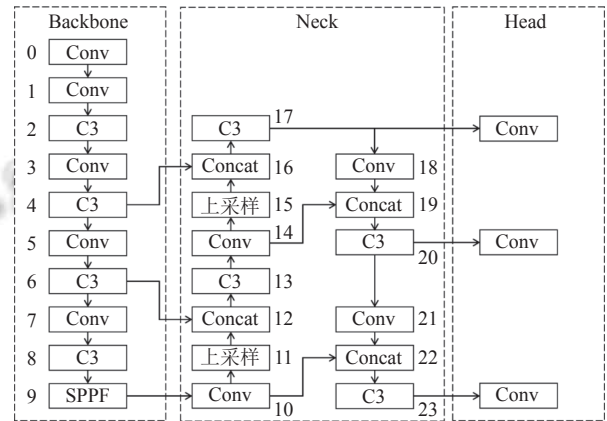


图1 YOLOv5网络结构

### 1.1 GhostConv 卷积模块

神经网络在进行特征提取时,会产生大量相似的冗余特征,它们会占用大量的计算资源.受GhostNet<sup>[6]</sup>启发,引入GhostConv模块,其通过廉价的操作生成更多的特征图,从而减少中间扩展过程中的内存消耗.GhostConv模块结构是在普通卷积模块基础上进行改进,将常规的卷积操作变为了两步,第1步先对输入信息进行常规卷积,获得一些特征图,第2步再对第1步获得的特征图进行线性操作,生成冗余特征图,最后将两步的输出进行Concat连接.其与普通卷积对比如图2所示.

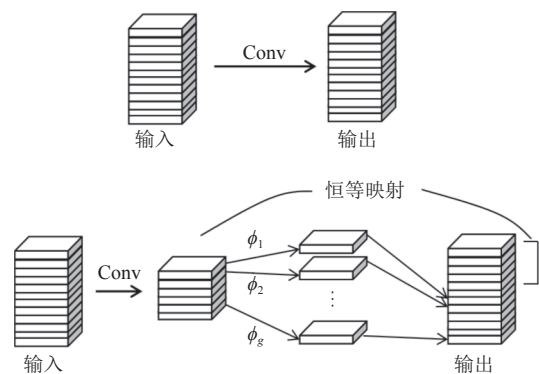


图2 常规卷积与GhostConv对比

GhostConv 模块通过使用分组卷积进行线性变换, 假设输入的特征图大小为 $[C_1, H, W]$ , 输出特征图大小 $[C_2, H', W']$ , 其中  $C$  为通道数,  $H$  与  $W$  是特征图高与宽, 经过分组卷积分为  $g$  个组后, 输入特征图每组尺寸变为  $W \times H \times C_1/g$ , 单个卷积核每组尺寸变为  $K \times K \times C_2/g$ ,  $K$  为卷积核大小, 由此可计算参数量从普通卷积的  $K \times K \times C_1 \times C_2$  变为  $K \times K \times C_1 \times C_2/g$ , 可以明显降低网络的参数量与计算量. 同样假设输入的特征图大小为 $[C_1, H, W]$ , 输出特征图 $[C_2, H', W']$ , 其中常规卷积核为  $K$ , 线性变换卷积核为  $D$ ,  $g$  表示每个通道产生的总映射数, 假设  $K=D$ ,  $g$  远小于  $C_1$ , 如下公式代表了常规卷积与 GhostConv 模块的计算量和参数量之比:

计算量之比:

$$r_s = \frac{C_2 \times H' \times W' \times C_1 \times K \times K}{\frac{C_2}{g} \times H' \times W' \times C_1 \times K \times K + (g-1) \times \frac{C_2}{g} \times H' \times W' \times D \times D}$$

$$= \frac{C_1 \times K \times K}{\frac{1}{g} \times C_1 \times K \times K + \frac{(g-1)}{g} \times D \times D} \approx \frac{C_1 \times g}{C_1 + g - 1} \approx g$$

参数量之比:

$$r_c = \frac{C_1 \times C_2 \times K \times K}{\frac{C_2}{g} \times C_1 \times K \times K + (g-1) \times \frac{C_2}{g} \times D \times D}$$

$$\approx \frac{C_1 \times g}{C_1 + g - 1} \approx g$$

因此, 理论上使用 GhostConv 模块可以节约  $g$  倍计算成本, 减少  $g$  倍参数. 通过将 GhostConv 模块引入 Backbone 主干并替换常规卷积, 将  $g$  设定为 2, 可以大大减少计算成本, 并保证网络的稳定性.

## 1.2 ShuffleC3 模块

### 1.2.1 ShuffleNetV2

ShuffleNetV1<sup>[7]</sup>中, 使用了分组卷积来进行降低参数量, ShuffleNetV2<sup>[8]</sup>中在其基础上提出了 channel split 方法. channel split 代替分组卷积, 减少分组卷积所需要的计算量, 将输入的  $C$  维通道一分为二, 得到左右两个分支, 由于网络中设计的分支数量越多、碎片化程度越高, 会导致网络的速度越慢, 故其中一个分支保持恒等映射, 另一个分支通过 3 个卷积进行特征提取, 再对输出进行 Concat 连接操作. 在连接之后, 由于每次常规卷积操作都是针对同一分支进行卷积处理, 导致每个组内没有进行信息交流, 于是提出 channel shuffle 的思想, 如图 3 所示, 通过 channel split 所得的  $g$  组特

征矩阵, 对其再次进行  $g$  组划分, 将每个通道进行位置变换, 拼接出新的特征矩阵, 达到特征信息融合的目的, 以此来提高特征的重用性和特征提取性能.

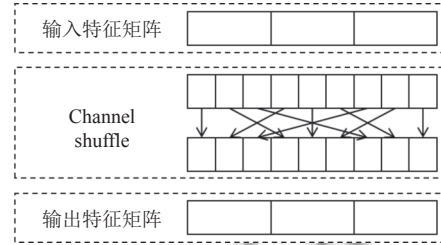


图 3 Channel shuffle 思想结构

C3 模块是 Backbone 中的重要组成部分, 由 3 个卷积模块和 1 个 Bottleneck 模块组成, 本文将 ShuffleNetV2 与 C3 模块相结合, 将 Bottleneck 结构替换为 ShuffleNet Bottleneck 结构. ShuffleNet Bottleneck 结构如图 4 所示, 其中卷积模块包括 Conv 模块、BN 归一化模块、激活函数模.

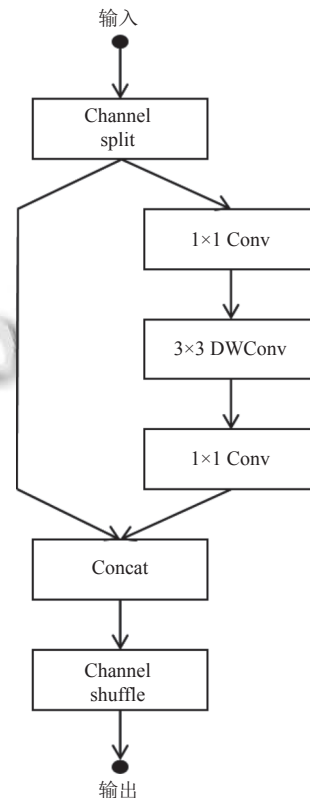


图 4 ShuffleNet Bottleneck 结构

### 1.2.2 CBAM 注意力

CBAM<sup>[9]</sup>是一种轻量级的卷积注意力模块, 它与常见的注意力模块 SENet<sup>[10]</sup>和 ECANet<sup>[11]</sup>不同, 其在通道

维度上进行注意力操作之外,还在空间维度进行了注意力操作.其结构如图5.

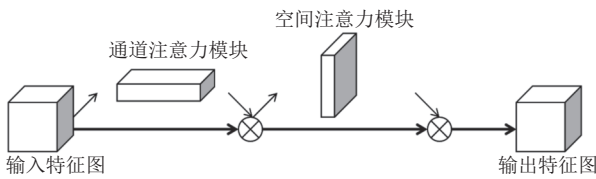


图5 CBAM 结构

通道注意力模块主要关注输入图片中有意义的信息,例如在分类任务中关注为什么分成不同的类别.其在通道上保持维度不变,在空间上压缩维度.假设输入的特征图大小为 $[C, H, W]$ ,将其分别通过最大池化层与平均池化层后,得到两个大小均为 $[C, 1, 1]$ 的特征图,把这两个特征图送入全连接层后,将输出结果相加,再通过 Sigmoid 激活函数的计算,获得代表特征重要程度的通道权重系数,最后再将原输入的特征图与通道权重系数相乘,得到大小为 $[C, H, W]$ 的输出特征图.

空间注意力模块主要关注目标的位置信息,其在空间上保持维度不变,在通道上压缩维度,其原理与通道注意力模块相同.

本文添加 CBAM 注意力到 ShuffleNet Bottleneck 结构中. CBAM 注意力机制可以使网络关注更需要关注的地方,通过并行的最大池化层和平均池化层,使网络在特征提取的过程中,提取到更加丰富、全面的高层特征,提升特征提取的性能.

### 1.2.3 ShuffleC3 模块设计

本文网络使用分组结构的 ShuffleNet Bottleneck 来替换原本 YOLOv5 网络中 C3 模块中的 Bottleneck 结构,分组结构可以降低一部分特征提取过程中所占用的参数量和计算量,但为了维持特征提取准确性与轻量型结构之间的平衡,本文网络在 ShuffleNet Bottleneck 中添加 CBAM 注意力机制、用  $3 \times 3$  的卷积结构来替换原本  $1 \times 1$  的卷积结构以加深网络深度、增加 C3 模块中 Bottleneck 的数量,在原本的基础上增加一次 Bottleneck 结构.由于 SiLU 函数在接近 0 时具有更平滑的曲线,且可以使网络的输出范围在 0 和 1 之间,实验得知其在本网络中获得了优于原 ReLU<sup>[12]</sup>函数的性能,故在 ShuffleNet Bottleneck 的卷积模块中使用 SiLU 函数替换 ReLU 函数.通过上述方法,使得网络获得更低的参数量和计算量,并拥有更高的特征提取性能. ShuffleC3 结构如图 6 所示,  $n$  表示 Bottleneck 结构重复次数.

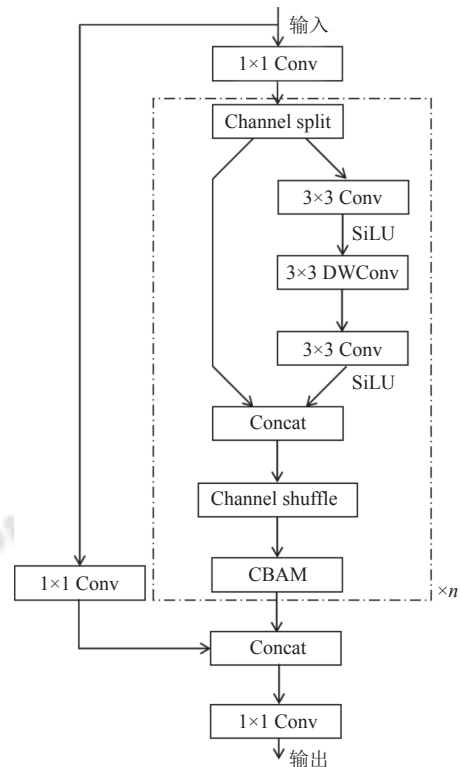


图6 ShuffleC3 结构

### 1.3 Head 部分改进

Head 目标检测头部分主要由卷积层、池化层、全连接层等构成,其用来对 Backbone 主干部分提取出的特征图进行大、中、小目标的多尺度目标检测.锚框 (Anchor) 部分是预定义的一组边界框,用于在特征图上生成候选框,由于本次使用的钢缆数据集的目标多为中型目标和小型目标,故只定义 $[10, 13, 16, 30, 33, 23]$ 小目标检测框和 $[30, 61, 62, 45, 59, 119]$ 中目标检测框.由实验得知,使用预定的锚框所得到的检测性能要优于使用 K-means 聚类所获得的锚框,故本文网络使用预定锚框代替 K-means 聚类所获得的锚框.再在 Head 网络中剪掉大目标检测部分和相对应的与 Backbone 主干相连接的 Neck 网络部分,大幅降低网络的计算复杂度和参数量.改进后的 YOLOv5 网络结构图如图 7 所示.

## 2 实验结果及分析

为了证明改进网络在钢缆表面损坏上识别的优越性,本节给出实验结果与分析.在本节中,首先介绍数据集、实验配置及评估指标,再进行消融实验验证 GhostConv 模块、ShuffleC3 模块、Head 部分改进的

作用,最后将本文网络方法与其他常用的缺陷识别方法进行了比较。

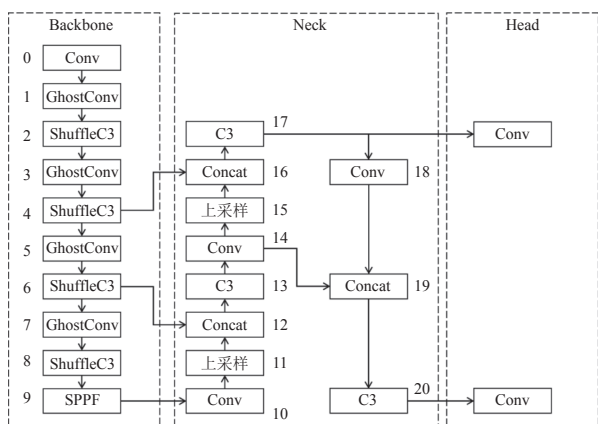


图7 改进后网络结构

### 2.1 数据集

为了证明本文所提出改进方法的有效性,引用了数据集 Cable Damage Computer Vision Project 和数据集 DBTT Computer Vision Project,并对它们进行部分整合与数据增强,获得钢缆损坏数据集 Cable Damage.数据集中为包含两种损坏类型:断裂、烧毁的问题钢缆,共包括 6590 张图像,并按 70%、20%、10% 的比例划分训练集、验证集、测试集.其中,训练集中样本个数为 4614,验证集中样本个数为 1314.训练集在网络参数训练中有着重要作用,可以获得最小的损失函数.而验证集用来评估训练网络识别表面损坏的准确性.数据集中两种损坏类型如图 8 所示。

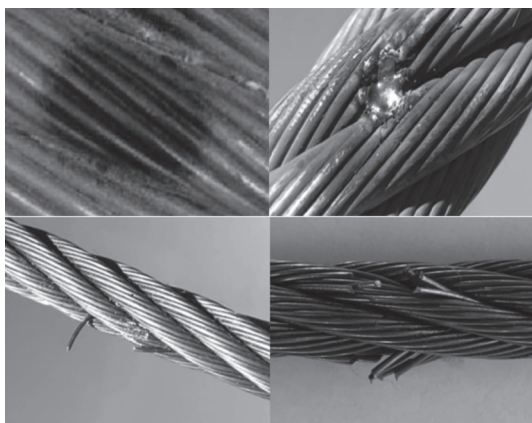


图8 数据集中两种损坏类型

### 2.2 实验配置

本次实验是在 PyTorch 框架下进行的.本文实验使用具有 10 GB 显存的 NVIDIA GeForce RTX 3080

显卡和 Intel(R) Xeon(R) 2.50 GHz Platinum 8255C CPU.网络训练过程进行 200 次迭代,每批次输入数据量大小为 32,将输入图像统一变换为 640×640 大小并归一化,采用随机梯度 (SGD) 下降优化器,采用线性学习率调度策略,初始学习率为 0.01,最终学习率为 0.0001.动量参数和权重衰减系数分别为 0.937 和 0.0005,由于网络结构改变后预权重会对结果产生影响,故本文所有实验均不使用预权重。

改进后网络的测试结果如图 9 所示,可以看出改进后的网络能够准确识别出钢缆表面的损坏。

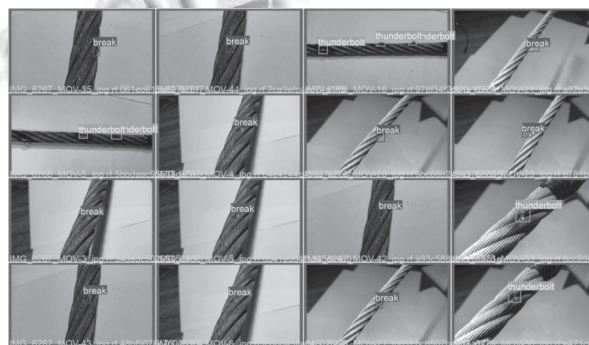


图9 测试结果

### 2.3 评估指标

本文采用平均精度 ( $mAP$ )、精确率 ( $Precision$ )、召回率 ( $Recall$ )、浮点运算量 (GFLOPs)、参数量 (Params)、模型大小和每秒帧数 (FPS) 来综合评价所提出的网络.在目标检测任务中,精确率和召回率是判断网络识别效果的重要指标,计算方法如下。

精确率:

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

召回率:

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

其中,  $TP$  代表正类被预测为正类的个数,  $FP$  代表正类被预测为负类的个数,  $FN$  代表负类被预测为负类的个数.平均精度 ( $mAP$ ) 是 mean average precision 的缩写,它等于所有类别的平均精度求和除以所有类别.此外,为了比较不同网络的计算复杂度,本文选择了计算量 (GFLOPs) 和参数量 (Params) 来表示不同网络之间的性能差异.另外,本文用每秒帧数 (FPS) 表示推理速度, FPS 的结果为 663 张测试图像的平均值。

## 2.4 对照实验结果及分析

本节通过消融实验验证 GhostConv 模块、ShuffleC3

模块、Head 改进在钢缆损坏识别上的优势, 实验结果如表 1 所示。

表 1 各项改进后实验结果

实验	GhostConv	ShuffleC3	Head	<i>mAP</i> (%)	<i>Precision</i> (%)	<i>Recall</i> (%)	Params (M)	GFLOPs	模型大小 (MB)	FPS
1	—	—	—	83.6	87.5	80.3	7.02	15.8	14.4	106.3
2	√	—	—	83.7	88.3	80.2	6.24	14.0	12.8	112.3
3	—	√	—	84	90.9	78.6	6.52	14.1	13.5	61
4	—	—	√	83.9	91.3	78.2	5.23	14.3	10.8	137
5	√	√	—	84.1	88.7	81.6	5.75	12.3	11.9	57.8
6	√	—	√	84.3	88.9	80.4	4.46	12.5	9.2	126.6
7	√	√	√	84.7	88.9	82.1	3.97	10.9	8.3	64.5

ShuffleC3 模块中引入了注意力机制, 目前注意力机制最常用的包括 SE-Net、ECA-Net、CA-Net<sup>[13]</sup>、

CBAM-Net 这 4 种. 分别加入 4 种注意力进行对比实验, 如表 2 所示。

表 2 ShuffleC3 中添加各注意力机制识别结果

实验	SE	ECA	CA	CBAM	<i>mAP</i> (%)	<i>Precision</i> (%)	<i>Recall</i> (%)	Params (M)	GFLOPs	模型大小 (MB)
1	√	—	—	—	83.9	87.8	82.1	3.97	10.9	8.3
2	—	√	—	—	83.5	88.5	82.7	3.94	10.9	8.2
3	—	—	√	—	84.1	88.1	81.9	3.97	10.9	8.4
4	—	—	—	√	84.7	88.9	82.1	3.97	10.9	8.3

为了验证本文改进网络在钢缆损坏识别上的性能, 将本文网络与其他先进的轻量级网络模型进行对比, 包括 YOLOv3-Tiny、YOLOv6<sup>[14]</sup>、YOLOv7-Tiny, 再进行了与常见的两阶段目标检测网络 Faster-RCNN<sup>[15]</sup> 性能对比. 此外, 将常用的轻量级骨干网络 MobileNetV3<sup>[16]</sup>、ShuffleNetV2、GhostNet 取代 YOLOv5 默认的 Backbone 主干, 进行对比. 结果如表 3 所示。

表 3 常见目标检测网络性能对比

实验	模型	<i>mAP</i> (%)	Params (M)	GFLOPs	模型大小 (MB)
1	YOLOv5s	83.6	7.02	15.8	14.4
2	YOLOv3-Tiny	84.1	8.67	13	16.6
3	YOLOv6s	81.9	18.5	45.3	181
4	YOLOv7-Tiny	83.4	6.01	13.2	12.3
5	Faster-RCNN	61.3	137.1	370.2	108
6	MobileNetV3-YOLOv5	82.5	5.02	11.3	10.4
7	ShuffleNetV2-YOLOv5	81.7	3.79	8.0	8.0
8	GhostNet-YOLOv5	84.2	4.76	7.9	10.0
9	改进YOLOv5	84.7	3.97	10.9	8.3

根据消融实验可以看出在分别单独引入 GhostConv 模块、ShuffleC3 模块、Head 改进后, 网络的计算量和参数量都有明显程度的降低. 在实验 5-实验 7 中分别对以上 3 种改进进行了不同方法的融合, 从数据可以看出实验 7 的融合方法效果最好, 提高了平均精度 1.1%, 参数量和计算量分别降低了 43.4% 和 31%, 模型大小减少了 42.3%. 由于 ShuffleC3 模块中引入了 CBAM 注意力机制, 其会对网络运算速度产生一些影

响. FPS 指标会受到实验配置影响, 在不同的实验设备上会有不同的结果, 因此本文更注重平均精度和网络的计算成本, 故实验 7 的网络模型更加符合要求。

由注意力对比实验可以看出 4 种注意力机制对参数量、计算量和模型大小的影响相差不大, 使用 CBAM 可以获得更高的平均精度, 因此在本文网络中使用 CBAM 添加到 ShuffleC3 模块中。

由不同模型对比实验可知, YOLOv3-Tiny 虽然平均精度略高于 YOLOv5s, 但由于其主干使用 DarkNet-53, 导致有了更高的计算成本. YOLOv6s、YOLOv7-Tiny 在钢缆损坏识别上的表现均不如改进前的 YOLOv5s, 本文网络在保证平均精度的同时, 在计算成本和模型大小上均有优于其他模型的表现, 其中替换骨干的 ShuffleNetV2-v5、GhostNet-v5 网络由于其使用大量轻量级卷积且网络深度较浅, 分别在参数量和计算量上略低于改进后网络. ShuffleNetV2-v5 的参数量低了 0.18M, 计算量低了 2.9G, GhostNet-v5 的计算量低了 3G, 但是它们在减轻运算成本的同时, 因为其较浅的网络无法保证识别的准确度, 所以平均精度均低于改进后网络, 因此改进后网络更加符合本文的要求, 拥有最好的性能。

## 3 结论与展望

本文在基于 YOLOv5 的基础上, 设计了一种针对

钢缆表面损坏识别的轻量级网络,为了更进一步的降低网络参数数量和计算复杂度,节约计算成本,并保证检测的平均精度.本文采用了GhostConv模块、ShuffleC3模块并对Head部分进行改进,在Cable Damage数据集上测试,改进后网络对比YOLOv5s提高1.1%的平均精度,参数数量和计算量分别降低了43.4%和31%,模型大小减少了42.3%.未来,将重点对算法进行进一步优化,以达到更高的准确率、更快的检测速度和更低的模型复杂度.

### 参考文献

- 1 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.
- 2 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.
- 3 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.
- 4 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 936–944.
- 5 Liu S, Qi L, Qin HF, *et al.* Path aggregation network for instance segmentation. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8759–8768.
- 6 Han K, Wang YH, Tian Q, *et al.* GhostNet: More features from cheap operations. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 1577–1586.
- 7 Zhang XY, Zhou XY, Lin MX, *et al.* ShuffleNet: An extremely efficient convolutional neural network for mobile devices. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 6848–6856.
- 8 Ma NN, Zhang XY, Zheng HT, *et al.* ShuffleNet v2: Practical guidelines for efficient CNN architecture design. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 122–138.
- 9 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. Proceedings of the 15th European conference on Computer Vision. Munich: Springer, 2018. 3–19.
- 10 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.
- 11 Wang QB, Wu BG, Zhu PF, *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 11531–11539.
- 12 Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. Fort Lauderdale: JMLR.org, 2011. 315–323.
- 13 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13708–13717.
- 14 Li CY, Li LL, Jiang HL, *et al.* YOLOv6: A single-stage object detection framework for industrial applications. arXiv:2209.02976, 2022.
- 15 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.
- 16 Howard A, Sandler M, Chen B, *et al.* Searching for MobileNetV3. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 1314–1324.

(校对责编:牛欣悦)