doi:10.12068/j.issn.1005－3026.2024.06.002

基于改进YOLOv4轻量化网络的机械手状态检测算法

郭立新[1]，毕素涛[1,2]，赵明扬[2]
（1.东北大学 机械工程与自动化学院，辽宁 沈阳 110819；2.季华实验室，广东 佛山 528200）

摘 要：YOLOv4网络结构复杂、参数较多、模型较大，因此极大地限制了其在工业上的应用.针对这一问题，提出一种改进YOLOv4的轻量化网络.首先，采用GhostNet代替YOLOv4主干网络，简化网络结构，降低模型参数量；其次，为了弥补网络简化后带来的精度损失，在其余两个输出特征层后加入Spatial Pyramid Pooling结构，加强特征提取；再次，加入Squeeze and Excitation Network通道注意力机制，增强网络重要信息提取能力；最后，将损失函数CIOU替换为SIOU，加快模型收敛，进而产生更好的模型.实验结果表明，在满足工业要求的前提下，改进后的轻量化网络相比于YOLOv4网络，在牺牲较小检测精度的情况下，模型参数量和计算量大幅下降，同时检测速度得到了提升，从而证明了改进算法在光纤插拔任务中机械手夹持状态识别检测的有效性.

关 键 词：YOLOv4；GhostNet；深度可分离卷积；注意力机制；损失函数
中图分类号：TP 391 **文献标志码**：A **文章编号**：1005-3026(2024)06-0769-07

State Detection Algorithm of Manipulator Based on Improved YOLOv4 Lightweight Network

GUO Li-xin[1], *BI Su-tao*[1,2], *ZHAO Ming-yang*[2]
(1.School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110819, China; 2.Jihua Laboratory, Foshan 528200, China. Corresponding author: GUO Li-xin, E-mail: lxguo@mail.neu.edu.cn)

Abstract: The YOLOv4 network is difficult to be widely used in industry due to its complex structure, many parameters, and large model size. In view of this problem, an improved lightweight network based on YOLOv4 is proposed. Firstly, GhostNet is used to replace the YOLOv4 backbone network, simplifying the network structure and reducing the number of model parameters; Secondly, in order to make up for the accuracy loss caused by network simplification, Spatial Pyramid Pooling structure is added after the other two output feature layers to enhance feature extraction. Thirdly, the attention mechanism of channel, which is Squeeze and Excitation Network, is added to improve the network's ability to extract important information. Finally, the loss function CIOU is replaced by SIOU to accelerate the convergence of the model and thus produce a better model. Experimental results show that, on the premise of meeting industrial requirements, compared with YOLOv4 network, the improved lightweight network significantly reduces the number of model parameters and the amount of computation, while improving the detection speed, at the same time, at the expense of less detection accuracy, thus proving the effectiveness of the improved algorithm in the identification and detection of the clamping state of the manipulator in the optical fiber plugging task.

Key words: YOLOv4; GhostNet; deepthwise separable convolution; attention mechanism; loss function

伴随着工业技术的快速发展，对于人工进行 光纤插拔的重复度较高的工作，开始逐渐由光纤

收稿日期：2023-02-02
基金项目：国家自然科学基金资助项目(52275283).
作者简介：郭立新(1968-)，男，辽宁沈阳人，东北大学教授，博士生导师.

插拔机器人进行替代.在光纤插拔作业任务中,最重要的步骤之一是对机械手状态是否夹持光纤进行判别,从而决定机器人下一步的操作.由于在光纤插拔作业过程中,机械手一直处于工作状态,其背景较为复杂,采用传统图像处理的方法来进行判别,已不能满足工作需求.

目前,基于深度学习的目标检测算法已经在工业领域被普遍应用,其主要分为两类:一类以R-CNN[1],Fast R-CNN[2]为代表的两阶段检测算法,先进行区域生成,再通过卷积神经网络进行识别分类;另一类是以SSD[3],YOLO[4-5]为代表的单阶段目标检测算法,将目标检测问题转化为一个回归问题来解决,实现了真正意义上的端到端.虽然单阶段目标检测算法精度略低于两阶段目标检测算法,但其具有较快的检测速度.同时,基于深度学习的目标检测算法也存在缺点,其参数多、模型大,对设备有较高的要求.针对这一问题,本文提出一种改进YOLOv4[5]的轻量化网络,在降低精度较小的情况下,其参数量和计算量大幅下降,同时每秒传输帧数(FPS, frames per second)得到提升.

1 YOLOv4算法简介

YOLOv4目标检测算法是在YOLOv3[4]的基础上进行改进得到的,相比于YOLOv3,其具有更快的检测速度和更高的精度.YOLOv4的网络结构如图1所示.

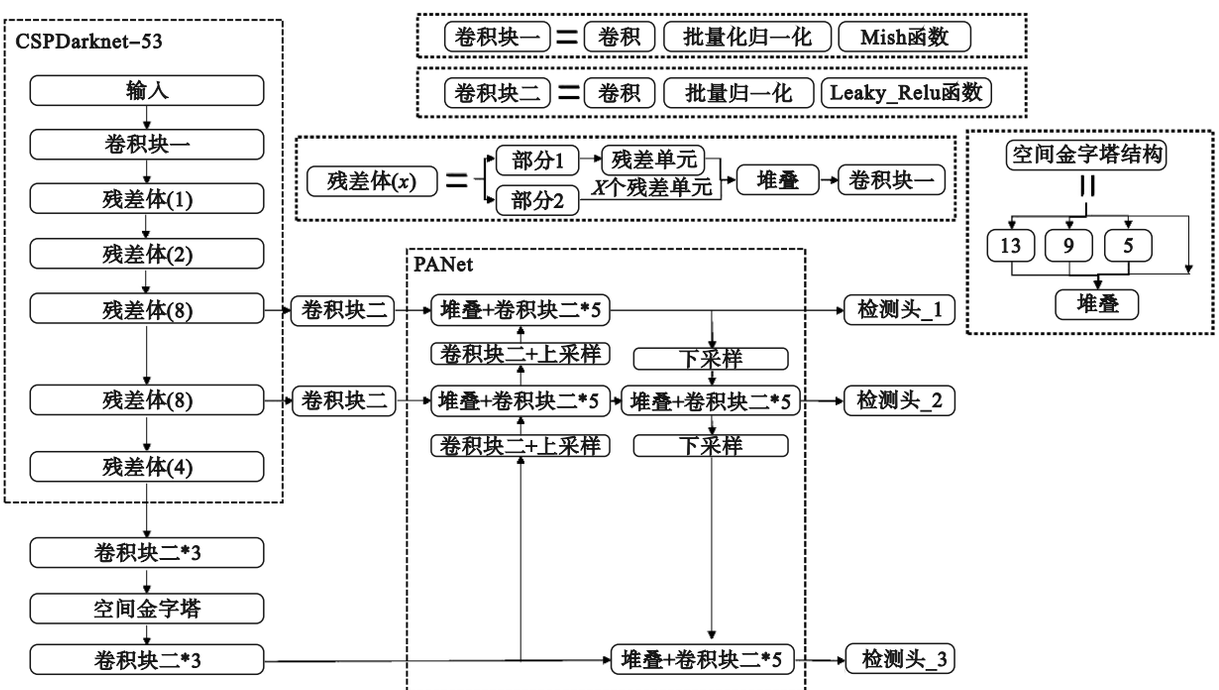

图1 YOLOv4网络结构图

Fig. 1 Structure of YOLOv4 network

YOLOv4其主要由3部分组成:主干网络CSPDarknet-53;加强特征提取网络SPP[6](spatial pyramid pooling)、PANet[7](path aggregation network);预测网络.相比于YOLOv3,YOLOv4在Darknet-53网络结构中引入了CSPNet(cross stage partial network),CSPNet[8]算法的核心思想是将特征图分为两部分:一部分经过瓶颈操作;另一部分直接传递到下一层.二者进行堆叠,从而使整个网络架构能够在减少计算量的同时实现更丰富的梯度组合.主干网络的主要作用是用来对图像进行初步的特征提取,在YOLOv4中,输入图像经过主干网络CSPDarknet-53后,得到3个不同尺度的初步特征图.对得到的特征图进一步使用SPP和PANet加强特征提取,其中特征SPP对输入的特征图分别利用4个不同的尺度进行最大池化处理,池化核大小分别为13×13,9×9,5×5,1×1,最后将特征图进行堆叠,增大网络的感受野.PANet对主干网络提取到的特征图进行采样、卷积和堆叠,融合自上而下和自下而上的特征图,目的是加强网络特征融合,获取更有效的特征.预测网络的作用是利用经过特征加强网络处理过的特征图对目标进行定位与分类.

2　YOLOv4算法改进

2.1　Ghostnet替换主干网络

Ghostnet[9]是由华为诺亚方舟实验室提出,旨在通过廉价的操作获得更多特征图. Ghostnet轻量级网络主要通过堆叠Ghost模块得出Ghost瓶颈形成的. Ghost模块的主要作用是代替普通卷积,首先采用普通的1×1卷积对输入的特征图进行通道数的压缩,然后对压缩后的特征图分别进行深度可分离卷积(逐通道卷积)和恒等映射,最后将二者堆叠生成更多的特征图. Ghost模块的结构如图2所示.

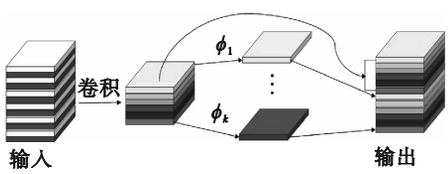

图2　Ghost模块结构

Fig. 2　Structure of Ghost module

Ghost瓶颈是由Ghost模块组成的瓶颈结构,如图3所示.图中步长为1的结构中其主干部分由两个Ghost模块串联组成,其中第1个模块扩大通道数,第2个模块将通道数进行压缩使其与输入通道数一致,最后将输入特征图与经过处理后的特征图进行相加.由于步长为1,因此不会对输入特征图的高和宽进行压缩,其主要目的为加深网络深度.图中步长为2的结构中其主干部分在两个模块之间加入了步长为2的深度可分离卷积,可以将输入的特征图的高和宽进行压缩,使其大小为原来的1/2,同时在残差边部分也有一个步长为2的深度可分离卷积,以保证在相加操作时,二者特征图长和宽相等.

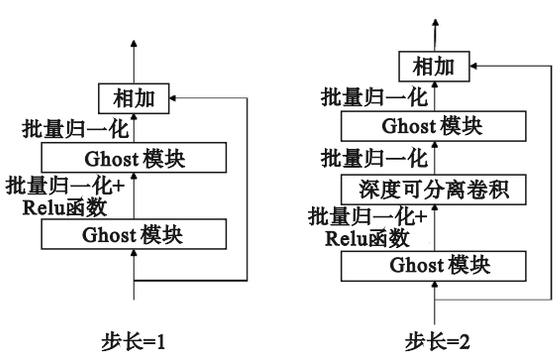

图3　Ghost瓶颈结构

Fig. 3　Structure of Ghost bottleneck

Ghostnet是由多个Ghost瓶颈组成,Ghost模块相比于普通卷积具有较少的参数与计算量,因此使用Ghostnet代替CSPDarknet-53作为YOLOv4的主干网络可以简化网络,减少网络的参数量与计算量,加快网络的推理速度. Ghostnet作为YOLOv4主干网络结构如图4所示.

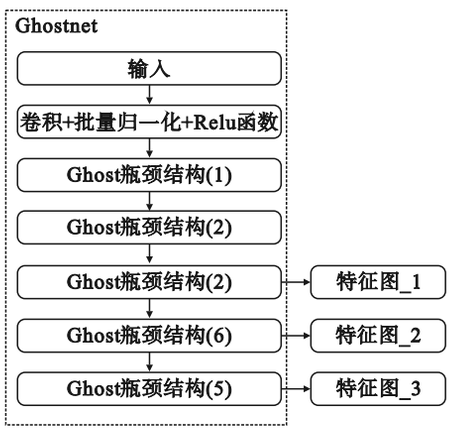

图4　改进YOLOv4的主干网络

Fig. 4　Backbone network of improved YOLOv4

2.2　深度可分离卷积代替普通卷积

深度可分离卷积[10]是将普通卷积拆分为逐通道卷积与逐点卷积,相比于普通卷积,其具有更少的参数与计算量.普通卷积如图5a所示,深度可分离卷积如图5b所示.

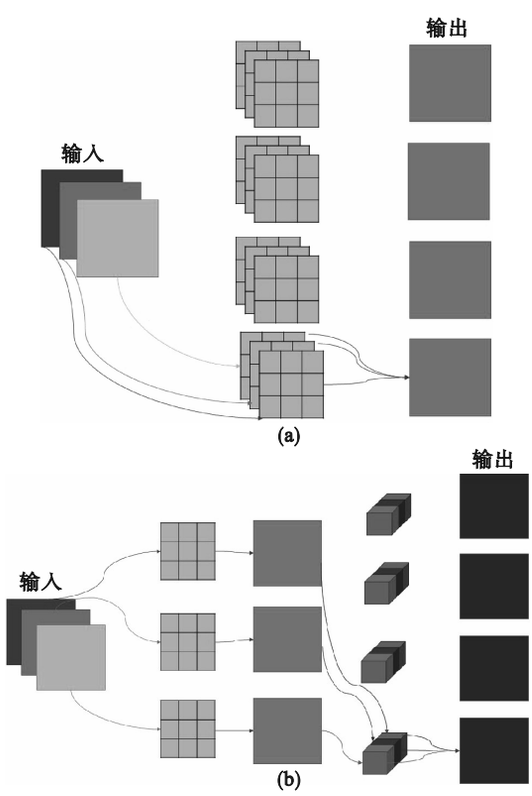

图5　普通卷积与深度可分离卷积

Fig. 5　Traditional convolution and depthwise separable convolution

(a)—普通卷积;(b)—深度可分离卷积.

假设输入特征图大小为$I_W \times I_H \times I_C$,输出特征图大小为$O_W \times O_H \times O_C$,卷积核大小为$K \times K$.其中$I_W$,$O_W$分别表示输入、输出特征图的宽;$I_H$,$O_H$分别表示输入、输出特征图的高;$I_C$,$O_C$分别表示输入、输出特征图的通道数.

$$P_{conv} = K \times K \times I_C \times O_C, \tag{1}$$

$$C_{conv} = K \times K \times I_C \times O_W \times O_H \times O_C, \tag{2}$$

$$P_{dw} = K \times K \times I_C + I_C \times O_C, \tag{3}$$

$$C_{dw} = K \times K \times O_W \times O_H \times I_C + I_C \times 1 \times 1 \times O_W \times O_H \times O_C. \tag{4}$$

式中:P_{conv},C_{conv}分别为普通卷积的参数量和计算量;P_{dw},C_{dw}分别为深度可分离卷积的参数量和计算量.

通过式(1)~式(4)可以得到,在同等条件下,普通卷积和深度可分离卷积的参数量与计算量.记R为深度可分离卷积与普通卷积的计算量之比.

$$R = \frac{C_{dw}}{C_{conv}} = \frac{1}{O_C} + \frac{1}{K^2}. \tag{5}$$

通常情况下,输出特征图的通道数远大于卷积核大小的乘积.假设卷积核的大小为3×3,那么普通卷积的计算量约是深度可分离卷积的计算量的9倍.由此可知,采用深度可分离卷积代替普通卷积可大幅减少计算量与参数量.

2.3 加入空间金字塔池化结构

采用轻量化网络Ghostnet代替CSPDarknet-53,虽然简化了网络结构,减少了参数量与计算量,但同时也导致了网络模型的表达能力不足,从而使网络的检测精度降低.为了解决这一问题,在从主干网络中得到的第1层输出特征图和第2层输出特征图之后分别加入SPP结构,增强网络的感受野和加强网络的特征融合能力.

2.4 融入注意力机制

SENet[11](squeeze-and-excitation networks)是典型的通道注意力网络.SE(squeeze - and - excitation)注意力机制的作用是加强重要特征,减弱一般特征,其通过网络去学习特征权重,获取每一个特征图的重要程度,从而让网络去重点关注重要的特征图,使模型达到更好的效果.SENet网络结构如图6所示.对输入的$C \times H \times W$特征图进行平均池化得到$C \times 1 \times 1$的特征图,对1×1的特征图进行非线性变换并通过Sigmoid函数使其值位于0和1之间,该值表示从每个通道提取出来的权重;最后将输入特征图的每个通道各自乘各自的权重,得到新的特征图.每个通道的权重大小表示每个通道的重要程度,从而使重要信息所在通道加强,强化模型的重要特征提取能力.

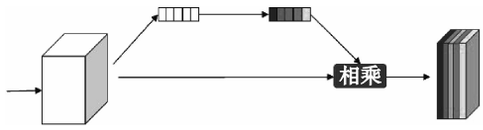

图6 SENet结构

Fig. 6 Structure of SENet

2.5 损失函数SIOU替代CIOU

目标检测算法中损失函数包括分类损失和置信度损失以及边界框回归损失.YOLOv4采用CIOU[12](complete intersection over union)损失函数作为边界框回归损失函数.CIOU将真实框与预测框之间的中心距离、重叠率、尺度以及惩罚项进行综合考虑,计算公式为

$$\mathrm{CIOU} = \mathrm{IOU} - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v. \tag{6}$$

其中:IOU为真实框与观测框的交集与并集的比值;$\rho^2(b, b^{gt})$代表真实框b与预测框b^{gt}之间的欧氏距离;c代表的是能够同时包含与预测框与真实框的最小闭包区域的对角线距离;α和v的计算公式为

$$\alpha = \frac{v}{1 - \mathrm{IOU} + v}, \tag{7}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2. \tag{8}$$

CIOU回归时的损失计算公式:

$$L_{\mathrm{CIOU}} = 1 - \mathrm{CIOU}. \tag{9}$$

从以上公式可以看出,CIOU损失函数并没有考虑到真实框与预测框之间的方向,导致收敛稳定性差.SIOU[13]损失函数引入真实框与预测框之间的角度回归,从而使得目标框的回归变得更加稳定.SIOU主要包含以下部分:

1)角度损失,公式为

$$\Lambda = 1 - 2 \times \sin^2\left(\arcsin\left(\frac{c_h}{\sigma}\right) - \frac{\pi}{4}\right). \tag{10}$$

其中:c_h为真实框和预测框中心点的高度差;σ为真实框和预测框中心点之间的距离.

2)距离损失,计算公式为

$$\Delta = \sum_{t=x,y}\left(1 - \mathrm{e}^{-\gamma\rho_t}\right). \tag{11}$$

其中:$\gamma = 2 - \Lambda$;$\rho_x = \left(\frac{b_{cx}^{gt} - b_{cx}}{c_w}\right)^2$;$\rho_y = \left(\frac{b_{cy}^{gt} - b_{cy}}{c_h}\right)^2$;$c_w$,$c_h$为真实框和预测框最小外接矩形的宽和高.

3)形状损失,计算公式为

$$\Omega = \sum_{t=w,h}\left(1 - \mathrm{e}^{-w_t}\right)^{\theta}. \tag{12}$$

其中:$w_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}$;$w_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}$.

综上所述,SIOU损失函数为

$$L_{\mathrm{SIOU}}=1-\mathrm{IOU}+\frac{\Lambda+\Omega}{2}. \tag{13}$$

2.6　改进的YOLOv4轻量化模型

改进模型主要对以下部分进行了改进:①Ghostnet替代CSPDarknet-53成为YOLOv4网络主干;②使用深度可分离卷积代替普通卷积;③在主干网络输出的第一层和第二层特征图中添加SPP结构;④融入SENet通道注意力机制;⑤将边框回归损失函数由CIOU替换为SIOU.改进后的模型如图7所示.

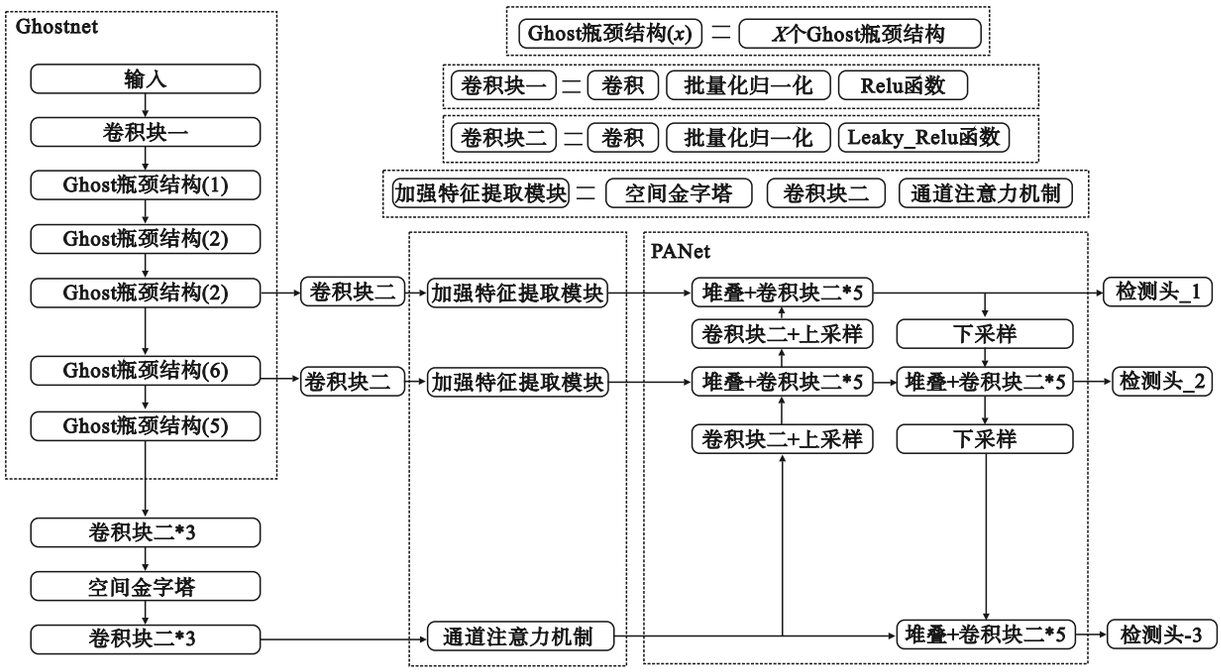

图7　改进YOLOv4轻量化网络结构

Fig. 7　Structure of improved YOLOv4 lightweight network

3　实验与结果分析

3.1　数据集制作与实验设置

目前未有关于机械手抓取光纤的公开数据集,本文所用数据集来源为光纤插拔作业任务时的实地场景采集.共采集1 110张图片,将数据集图片按8∶1∶1的比例划分训练集、验证集和测试集.

本实验采用的硬件配置为Intel(R) Xeon(R) W-2295 CPU @3.0GHz,GPU为RTX5000,内存16GB.环境为Python 3.9,Pytorch 1.10.

实验中,Bacthsizes设置为8,采用Adam优化器,初始学习率为1×10^{-4},采用余弦退火法进行学习率衰减,动量参数为0.94,同时在训练过程中采用了标签平滑来提高模型的泛化能力.

本文主要使用平均检测精度mAP(mean average precision)、每秒的传输帧数(frame per second,FPS)以及模型参数量和计算量作为评价指标对模型进行比较分析.mAP为所有类别的AP(average precision)的平均值,AP与准确率P和召回率R有关,计算公式为

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}},R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{14}$$

$$\mathrm{AP}=\int_0^1 P(R)\mathrm{d}R. \tag{15}$$

其中:TP为正确预测为正样本的数量;FP为错误的预测为正样本的数量;FN为错误的预测为负样本的数量.

3.2　实验结果分析

将改进的YOLOv4轻量化网络对机械手状态进行检测,检测结果如图8所示,准确率变化曲线如图9所示,召回率变化曲线如图10所示.实验结果见表1,由表中数据可以看出,改进的轻量化模型在光纤插拔任务中,机械手夹持状态判别精度高达0.980 3,FPS达到54.在具有较高检测精度的同时,还具有较快的检测速度.改进后的轻量化模型相比于YOLOv4算法,在mAP(平均精度)仅下降0.36个百分点的情况下,FPS得到了提升,同时相比于YOLOv4,参数量下降了81.68%,计算量下降了86.71%.

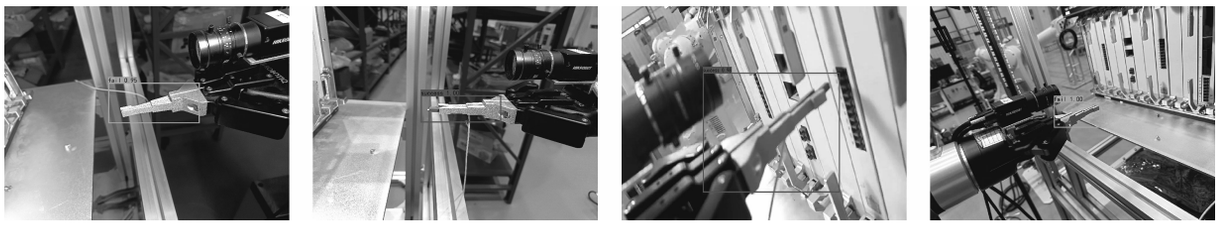

图 8 改进 YOLOv4 轻量化网络检测结果

Fig. 8 Detection result of improved lightweight YOLOv4

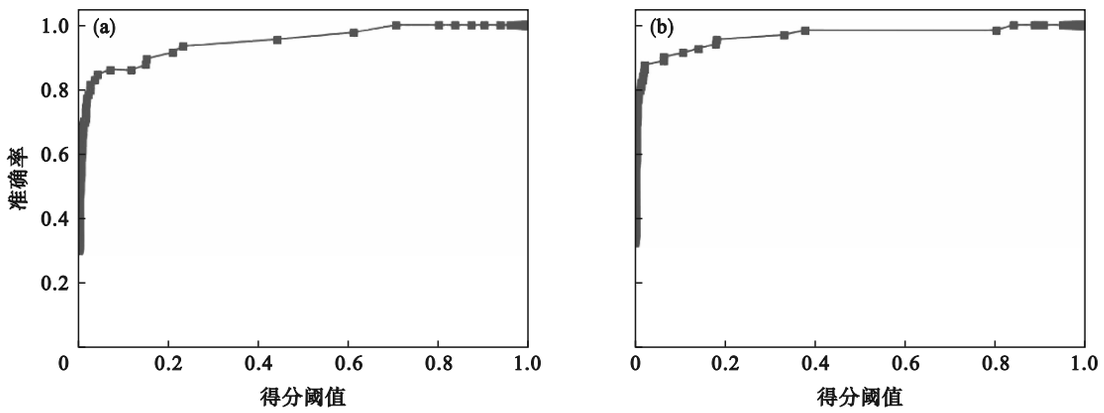

图 9 改进 YOLOv4 轻量化网络准确率变化曲线

Fig. 9 Accuracy change curves of improved lightweight YOLOv4

(a)—类别 fail;(b)—类别 success.

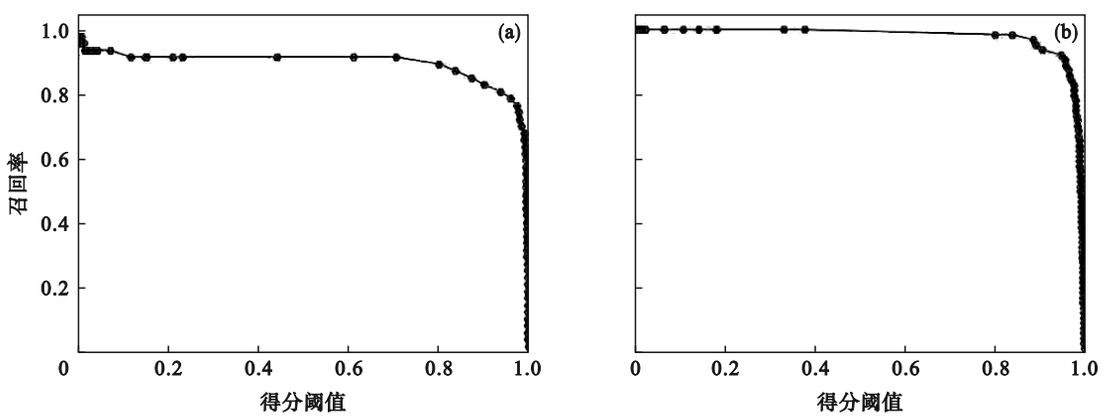

图 10 改进 YOLOv4 轻量化网络召回率变化曲线

Fig. 10 Recall change curves of improved lightweight YOLOv4

(a)—类别 fail;(b)—类别 success.

表 1 改进算法与 YOLOv4 指标对比

Table 1 Comparison between improved algorithm and YOLOv4

模型	参数量×10^{-6}	计算量×10^{-9}/FLOPs	FPS	mAP
YOLOv4	63.94	29.88	42	0.983 9
改进算法	11.71	3.97	54	0.980 3

注:FLOPs(floating point of operations)为浮点运算次数.

3.3 消融实验

为验证改进的 YOLOv4 轻量化网络模型每个改进策略对机械手状态检测的有效性,本文设计了消融实验.将 YOLOv4 采用 Ghostnet 作为主干网络且使用深度可分离卷积代替普通卷积的模型记为 YOLOv4_L.本文以 YOLOv4_L 为基准,分别采用不同改进模块进行实验分析.消融实验结果如表 2 所示,表中数据表明,仅对 YOLOv4 进行轻量化改进,即 YOLOv4_L 模型,该模型相比于 YOLOv4 网络模型,参数量大幅下降,检测速度得到提升,但精度却急剧下降.该模型检测精度并不能满足实际作业场景中的需求,需要在其基础上进行改进,提高网络检测精度.

本文在YOLOv4_L模型的基础上分别加入通道注意力网络、空间金字塔结构、损失函数的替换以及综合以上所有改进模块进行实验分析,结果表明不同的改进策略均能使模型的检测精度增加,其中作用较大的是加入SPP结构,相比于YOLOv4_L,mAP增加了1.37%.不同模型对应的检测速度与平均精度如图11所示,图11表明改进算法在具有较快的检测速度下,且有较高的检测精度.

表2　消融试验结果

Table 2　Result of ablation test

模块	SE	SPP	SIOU	参数量 $\times10^{-6}$	FPS	mAP
YOLOv4	—	—	—	63.94	42	0.983 9
YOLOv4_L	—	—	—	11.01	57	0.951 4
YOLOv4_L	√	—	—	11.05	56	0.960 7
YOLOv4_L	—	√	—	11.67	55	0.965 1
YOLOv4_L	—	—	√	11.01	57	0.959 3
YOLOv4_L	√	√	√	11.71	54	0.980 3

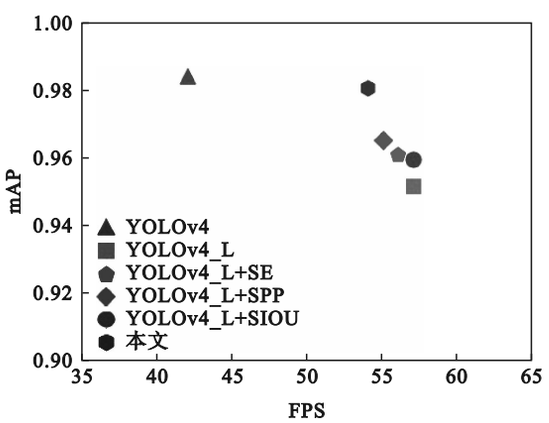

图11　不同模型检测速度与平均精度

Fig. 11　FPS and mAP of different models

4　结　　语

良好的网络模型,不仅要求有较高的检测精度,同时还要求模型具有较小的参数量和计算量.本文针对YOLOv4算法参数多和计算量大的问题,提出了一种改进YOLOv4的轻量化模型算法,使其能够在满足检测精度的情况下,具有较快的检测精度.实验结果表明,在满足工业要求的前提下,改进的算法相比于YOLOv4,在精度仅下降0.36个百分点的情况下,FPS得到了提升,同时参数量下降了81.68%,计算量下降了86.71%,证明了本文改进算法在光纤插拔作业任务中的有效性.

参考文献:

[1] 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.Columbus:IEEE,2014:580-587.

[2] Girshick R. Fast R-CNN [EB/OL]. (2015-04-30) [2022-12-25].https://arxiv.org/abs/1504.08083

[3] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector [C]//European Conference on Computer Vision.Cham:Springer,2016:21-37.

[4] Redmon J, Farhadi A. Yolov3: an incremental improvement [EB/OL]. (2018-04-08) [2022-12-25]. https://arxiv. org/pdf/1804.02767.pdf.

[5] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. (2020-04-23) [2022-12-25].https://arxiv.org/abs/2004.10934.

[6] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.

[7] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 8759-8768.

[8] Wang C Y, Liao H Y, Wu Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, 2020: 1571-1580.

[9] Han K, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 1577-1586.

[10] Chollet F. Xception: deep learning with depthwise separable convolutions [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017: 1800-1807.

[11] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011-2023.

[12] Zheng Z H, Wang P, Liu W, et al. Distance-IOU loss: faster and better learning for bounding box regression [EB/OL]. (2019-11-19) [2022-12-25]. https://arxiv. org/pdf/1911.08287.pdf.

[13] Gevorgyan Z, SIOU loss: more powerful learning for bounding box regression [EB/OL]. (2022-05-25) [2022-12-25].https://arxiv.org/abs/2205.12740.