

面向仓储货物的轻量化目标检测算法

王晨¹, 袁庆霓^{1*}, 白欢¹, 李恒², 宗文泽¹

¹贵州大学现代制造技术教育部重点实验室, 贵州 贵阳 550025;

²贵州大学机械工程学院, 贵州 贵阳 550025

摘要 为了实现对仓储环境下货物的精确检测,提出了一种可用于智能仓储机器人的轻量化仓储货物检测方法(E-YOLOv4-Lite)。该方法以YOLOv4为基础,引入MobileNetv3重构特征提取网络,在PANet中以深度可分离卷积代替标准卷积,减少模型参数量和运算量。融入改进的convolutional block attention module(CBAM),进一步提升网络检测性能,改进的CBAM采用自适应的一维卷积代替通道注意力模块中的全连接层,采用具有膨胀卷积的残差结构扩大空间注意力模块中的感受野。最后,在RPC商品数据集上进行了网络训练和实验测试,其参数量为11.25 MB,检测时间为14.4 ms,每秒传输帧数达到69.2,平均检测精度为95.43%。实验结果表明,改进后的E-YOLOv4-Lite模型具有精度高、实时性好和轻量化的优点,能够更好地满足仓储环境中的货物检测需求。

关键词 图像处理; 目标检测; 注意力机制; 轻量化; 商品数据集

中图分类号 TP391.41; TP183

文献标志码 A

DOI: 10.3788/LOP202259.2410007

Lightweight Object Detection Algorithm for Warehouse Goods

Wang Chen¹, Yuan Qingni^{1*}, Bai Huan¹, Li Heng², Zong Wenzhe¹

¹Key Laboratory of Advanced Manufacturing Technology of the Ministry of Education, Guizhou University, Guiyang 550025, Guizhou, China;

²School of Mechanical Engineering, Guizhou University, Guiyang 550025, Guizhou, China

Abstract To achieve accurate detection of items in the warehousing environment, a lightweight warehousing cargo detection method (E-YOLOv4-Lite) is proposed for use in intelligent warehousing robots. This technique builds on YOLOv4 by introducing the MobileNetv3 network to reconstruct the feature extraction network, replacing standard convolution in PANet with deep separable convolution, and reducing the number of model parameters and processing. The improved convolutional block attention module (CBAM) is integrated to improve network detection performance further. In the channel attention module, the improved CBAM replaces the full connection layer with adaptive one-dimensional convolution, and in the spatial attention module, the residual structure with expansive convolution is used to expand the receptive field. Finally, the network training and experimental tests are conducted through the RPC commodity data set, the number of parameters is 11.25 MB, the detection time is 14.4 ms, the frames per second is 69.2, and the mean average precision is 95.43%. The experimental results reveal that the improved E-YOLOv4-Lite model has the advantages of high accuracy, good real-time performance, and lightweight, allowing it to better meet the needs of cargo detection in storage environments.

Key words image processing; object detection; attention mechanism; lightweight; product dataset

1 引言

2020年8月,国家发改委研究制定了《推动物流业制造业深度融合创新发展实施方案》,以进一步推动物

流业制造业深度融合与创新发展,促进制造业转型升级^[1]。现如今,智慧物流发展迅速,物流分拣逐渐由人力分拣走向智能化,人工智能迎来了越来越多的挑战,机器视觉技术就是其研究热点之一。机器视觉在工业

收稿日期: 2021-09-02; 修回日期: 2021-10-22; 录用日期: 2021-11-01

基金项目: 国家自然科学基金(51865004)、贵州省科技厅(黔科合支撑[2020]4Y140号)、贵州大学研究生创新人才计划项目(2021)、横向课题“基于专利分析的工业机器人技术研究”(K19-0204-001)

通信作者: *qnyuan@gzu.edu.cn

领域的 3 大主要应用是视觉测量、视觉引导和视觉检测^[2]。

目前存在的主流视觉算法大多都是基于深度学习的目标检测算法,主要分为两类:一类是以 YOLO^[3-6]、SSD^[7]算法为代表的 one-stage 算法;另一类是以 RCNN^[8]、Faster-RCNN^[9]为代表的 two-stage 算法。目前在智能化生产过程中多采用 one-stage 算法进行目标检测:文献[10]中提出了一种用于桥梁裂缝检测的深度学习模型 YOLOv4-FPM,该模型对损失函数和网络进行了优化,提高了网络的每秒传输帧数(FPS),但是存在参数量较大难以部署于移动设备的问题;文献[11]提出了一种用于水果检测的轻量级卷积神经网络 YOLOv4-Lite,替换主干特征网络后大幅度减少了网络参数量,但是由于参数量的减少,出现了检测精确率不高的问题;文献[12]提出了一种改进 YOLOv4 模型对果园中的障碍物进行检测,该模型将标准卷积替换为深度可分离卷积,并将 YOLOv4 主干网络改为逆残差结构,网络参数量减少但是对小目标的检测性较差;文献[13]提出了一种基于改进 YOLOv3 算法的变电设备检测方法,该方法加入了多尺度特征融合,但是参数量较大导致检测效率不高;文献[14]提出了一种多尺度特征融合的 DenseNet 模型,引入了软阈值函数改善检测效果,由于采用多个阶段处理图像,提高了检测精度,但是增加了检测时间。

针对以上问题,为了在智能仓储移动机器人上实现实时有效的货物检测,本文提出了一种轻量化网络

模型 E-YOLOv4-Lite,该模型是在 YOLOv4 基础上进行改进的。在特征提取阶段,采用相比于 CSPDarknet-53 更加轻便的 MobileNetv3 来进行搭建,利用深度可分离卷积和逆残差模块减少网络参数量和计算量,并且将 MobileNetv3 根据 YOLOv4 的特征提取网络进行参数改进,使其能够更好地满足 YOLOv4 网络检测层的需求。将特征提取层中 PANet 部分的标准卷积改为深度可分离卷积,并引入改进的 convolutional block attention module(CBAM),使其能够提高目标检测的效率,减少网络预测时间,更好检测有遮挡的货品。在公开的 RPC 多目标商品数据集上进行训练和测试,提高仓储货品目标检测的准确性,并将改进后的算法与目前主流的轻量化算法 YOLOv4-tiny 进行效果对比。

2 所提算法

2.1 MobileNetv3

MobileNet 是一种轻量级深度神经网络模型。MobileNetv3^[15]综合了 MobileNetv1^[16]的深度可分离卷积、MobileNetv2^[17]的逆残差结构和 MnasNet 中的 squeeze and excitation(SE)注意力机制,能够在最大程度减少参数量的情况下保证准确率。MobileNetv3 中的 Block 模块如图 1 所示,输入特征图依次经过 1×1 的标准卷积和 3×3 的深度可分离卷积后,输入 SE 注意力结构块,经过池化层和全连接层激活后与原输入相乘,即可引入 SE 注意力机制,最后与输入特征图相加,得到输出特征图。

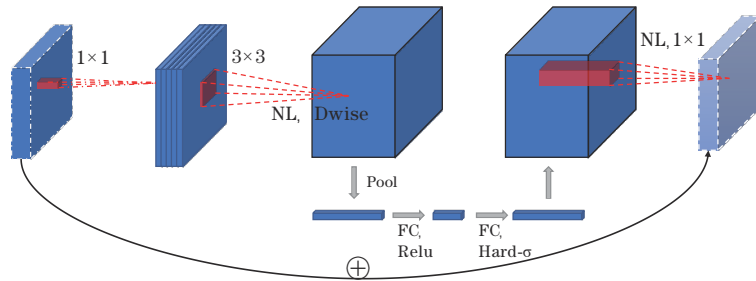


图 1 MobileNetv3 中的 Block 模块

Fig. 1 Block module in MobileNetv3

2.2 CBAM 注意力机制

CBAM 是一种简单有效的卷积神经网络注意力模块,它由通道注意力模块和空间注意力模块两部分组成,将特征图输入 CBAM 后即可得到加入注意力机制后的特征图,对输入特征进行进一步细化。CBAM 中的通道注意力和空间注意力结构如图 2 所示。

给定一个张量 $F \in \mathbf{R}^{C \times H \times W}$ 作为输入送入 CBAM 中:首先对输入 F 在不同通道内进行最大池化和平均池化,并将池化后的两个一维张量输入共享全连接层进行相加,生成一维通道注意力 $M_c \in \mathbf{R}^{C \times 1 \times 1}$,然后将得

到的注意力 $M_c(F)$ 与原输入 F 相乘后生成 F' 传入空间注意力模型;其次,将输入 F' 按空间进行全局最大值池化和均值池化,并将池化后的两个二维张量拼接后进行卷积,最终得到一个二维的空间注意力 $M_s(F') \in \mathbf{R}^{1 \times H \times W}$,最后将空间注意力 $M_s(F')$ 与 F' 相乘即可引入 CBAM 注意力。整体引入注意力的过程可以描述为

$$F' = M_c(F) \otimes F, \quad (1)$$

$$F'' = M_s(F') \otimes F'. \quad (2)$$

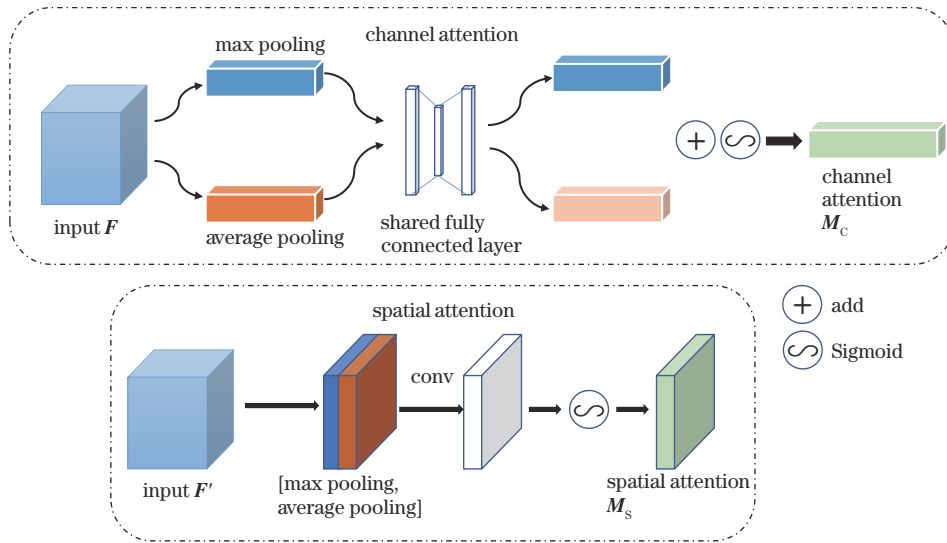


图2 CBAM中的通道注意力和空间注意力结构

Fig. 2 Channel attention and spatial attention structure in CBAM

3 算法改进

3.1 改进的YOLOv4-Lite网络模型

本研究借鉴文献[11]中的方案,将YOLOv4的主干特征提取网络CSPDarknet-53改为MobileNetv3,并在原来的基础上将原PANet中的 3×3 的标准卷积换成深度可分离卷积、激活函数改为Hard-swish来减少计算量,构成了全新的YOLOv4-Lite网络。将标准的 3×3 卷积替换为深度可分离的 1×1 和 3×3 卷积后,能够在不损失精度的情况下,减少参数量。由于在移动设备中计算Sigmoid函数的代价太大,而Hard-swish在深度网络中效果更好,因此采用Hard-swish作为激活函数。

以一个 416×416 的输入图像为例,首先在主干特征提取网络中经过5次Bottleneck块结构后,输出 52×52 、 26×26 、 13×13 的3个有效特征层。将 13×13 的特征层输入空间金字塔池化(SPP)网络,利用不同尺寸的池化层进行特征融合,提高感受野并分离出有效特征。随后将3组特征层输入PANet进行融合, PANet中自下而上的特征融合路径能够有效融合更丰富的特征信息。最后,经过特征融合后的3组特征层预测每个位置的3个边界框。若数据集中含有 k 个类别,每个边界框将给出 $3 \times (5+k)$ 个预测值,其中前4个值为边界框的位置信息,即横坐标、纵坐标、边界框宽度、边界框高度,第5个值表示边界框中存在该目标的置信度。整体网络结构如图3所示。

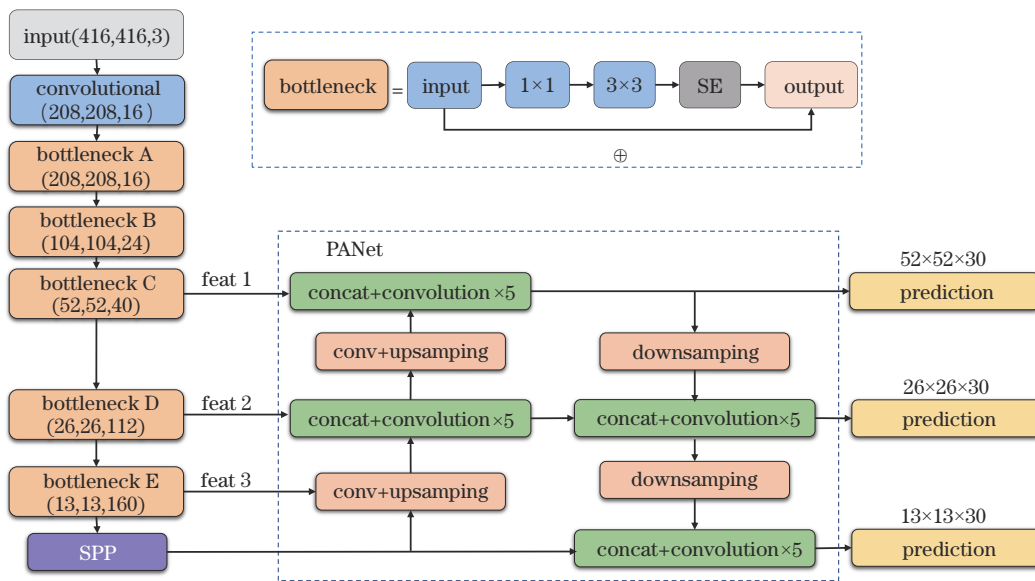


图3 YOLOv4主干提取网络改进

Fig. 3 Improvement of YOLOv4 backbone extraction network

3.2 改进的 E-CBAM 注意力机制

CBAM 注意力机制在计算生成通道注意力模块时,会使用全连接层进行特征映射。但由于全连接层的参数量巨大,即使先对共享全连接层的通道特征进行倍数压缩,其参数量仍然很大。另外在空间注意力模块中,为了聚合更广泛的特征,CBAM 机制使用一个 7×7 的卷积核来提取空间特征,相比于 3×3 的卷积核,虽然增大了感受野,但同时模块的参数量也随之增

大。因此,本研究借鉴文献[18]中的思想,采用自适应维度的一维卷积代替全连接层,有效减少了参数计算量。针对遮挡货品检测精度不高的情况,利用具有膨胀卷积的残差块对空间注意力模型中的参数量进行缩减,在增大感受野的同时提高对遮挡货品的检测精度。本研究将其称为 efficient-convolutional block attention module (E-CBAM), E-CBAM 的整体结构如图 4 所示。

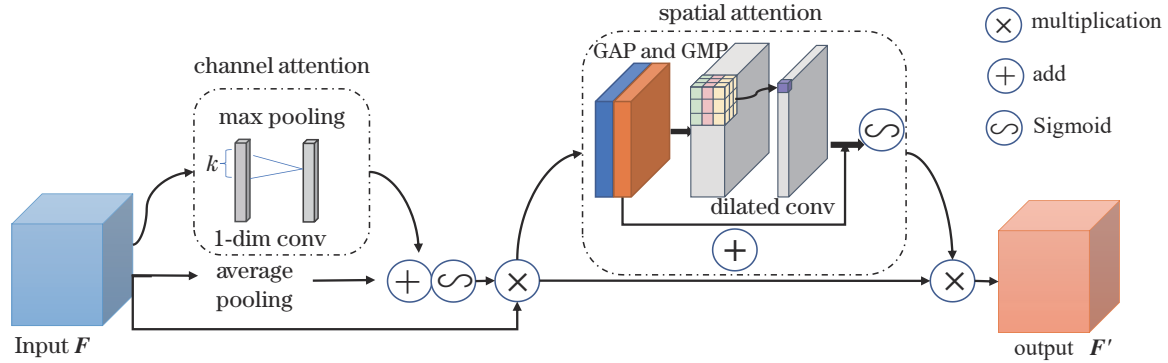


图 4 E-CBAM 整体结构

Fig. 4 E-CBAM overall structure

3.2.1 通道注意力改进

与 CBAM 类似,通道注意力的关注点在于输入特征图。首先利用全局最大值池化和均值池化来映射特征信息,形成两个不同的通道描述信息,用 F_{avg}^C 表示 F 经过平均池化的通道描述信息,用 F_{max}^C 来表示 F 经过最大池化的通道描述信息。采用自适应维度为 k 的一维卷积来聚合 k 个邻域通道内的特征信息,然后将两个通道的信息元素相加,并通过 Sigmoid 函数激活,生成通道注意力 $M_c(F) \in \mathbb{R}^{C \times 1 \times 1}$,再与原来的输入特征相乘从而注入通道注意力机制。具体的计算过程可描述为

$$M_c(F) = \sigma \left\{ f_{\text{ld}}^k \left[\text{AvgPool}(F) + f_{\text{ld}}^k (\text{MaxPool}(F)) \right] \right\} = \sigma \left[f_{\text{ld}}^k (F_{\text{avg}}^C) + f_{\text{ld}}^k (F_{\text{max}}^C) \right], \quad (3)$$

$$k = \left\lfloor \frac{\log_2 C}{2} + \frac{1}{2} \right\rfloor_{\text{odd}}, \quad (4)$$

式中: σ 表示 Sigmoid 激活函数; f_{ld}^k 表示卷积核大小为 k 的一维卷积操作; C 表示输入特征图的通道数; $\lfloor \cdot \rfloor_{\text{odd}}$ 表示与 \cdot 最接近的奇数。

3.2.2 空间注意力模块的改进

在通道注意力模块后加入空间注意力模块能够有效提高网络模型的检测能力,使网络关注于图像有用的信息,为了解决深层网络中加入卷积层造成网络退化的问题,本研究将原网络中的卷积结构替换为具有膨胀卷积的残差结构,如图 4 所示。在空间注意力模块中首先将已经引入通道注意力的特征信息进行全局平均池化和全局最大池化,生成两个不同的通道描述信息 $F_{\text{avg}}^{1 \times 1}$ 和 $F_{\text{max}}^{1 \times 1}$,并将两个通道信息进行拼接,生成一

个更加有效的空间特征层。随后利用具有膨胀卷积的残差结构进一步聚合上下层空间中的信息,提高感受野。经 Sigmoid 函数激活后生成空间注意力模型 $M_s(F') \in \mathbb{R}^{1 \times H \times W}$ 。最后,将空间注意力模型 $M_s(F')$ 与输入特征对应元素相乘即可注入空间注意力机制。具体的计算过程可描述为

$$M_s(F') = \sigma \left\{ (1 + f_{\text{dilation}}^{3 \times 3} f_{2d}^{1 \times 1}) [\text{GAP}(F') + \text{GMP}(F')] \right\} = \sigma \left[(1 + f_{\text{dilation}}^{3 \times 3} f_{2d}^{1 \times 1}) (F_{\text{avg}}^{1 \times 1} + F_{\text{max}}^{1 \times 1}) \right], \quad (5)$$

式中: $f_{\text{dilation}}^{3 \times 3}$ 表示卷积核大小为 3 的膨胀卷积; $f_{2d}^{1 \times 1}$ 表示卷积核大小为 1 的标准卷积。本研究采用的是膨胀率为 2 的膨胀卷积。

3.3 E-yolov4-Lite 网络模型

所提 E-CBAM 模型可以加入卷积神经网络的任意位置,为了进一步优化模型,将改进的注意力机制引入 PANet 层,主要原因如下: 1) 在不改变主干网络的情况下引入注意力机制,可以利用公开数据集中预训练权重进行迁移学习,大大缩减网络的训练时间,而且主干网络 MobileNetv3 中已经存在 SE 注意力模块,不需要再引入新的注意力机制; 2) 在深层的神经网络中引入注意力机制,能够有效获取更加深刻的特征信息和丰富的语义信息。

综合以上两点,本研究在改进网络的基础上将 E-CBAM 注意力机制添加到 PANet 中,在同样大小通道数的特征层进行卷积操作后引入 E-CBAM,由于 PANet 中 26×26 的特征层需要经过两次拼接和卷积操作,所以总共需要引入 7 个注意力模块,具体过程如图 5 所示。

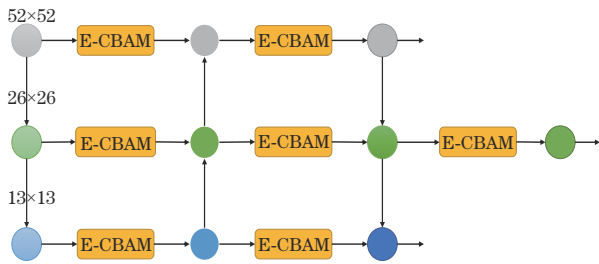


图 5 PANet中加入E-CBAM机制

Fig. 5 E-CBAM mechanism added to PANet

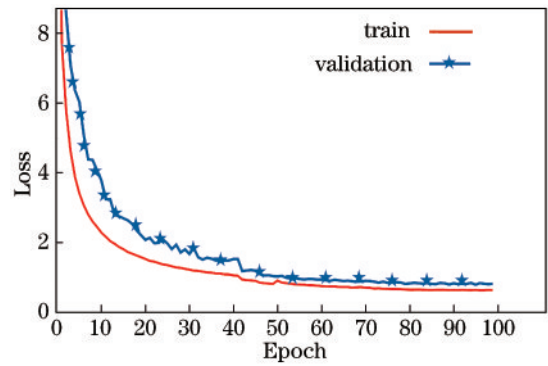


图 6 Tensorboard中的损失曲线

Fig. 6 Loss curve in Tensorboard

4 实验结果与分析

4.1 实验环境

实验过程中训练环境和测试环境保持一致,在 Windows 10 操作系统上进行训练和测试,采用 NVIDIA® GeForce™ RTX1080, 8 GB 的显卡,并运用 CUDA 10.2 加速训练。开发环境为 Anaconda 3, 采用 TensorFlow 2.2.0 作为深度学习框架,训练过程中 batch_size 设置为 16。首先冻结主干网络参数训练 50 个周期,解冻后对整个网络再训练 50 个周期,学习率设置为 10^{-4} 。图 6 为 E-YOLOv4-Lite 模型采用 Tensorboard 得到的训练过程损失曲线。

4.2 数据集

采用公开的 RPC 商品数据集^[19]进行训练和测试。RPC 数据集包含 200 个小类和 17 个大类,数据集中的图片充分考虑了日常生活条件下灯光及物体遮挡等问题,能够进一步提高算法的精确度。选择 12000 张图片作为基础数据集,按照 9:1 的比例进行划分,选择 10800 张图片进行训练,剩余 1200 张图片进行测试。为了进一步提高训练精度,采用 Mosaic 进行数据增强,从而增加训练数据量,提升单个 GPU 的训练能力。图 7 是部分 RPC 数据集示例。



图 7 部分数据集示例

Fig. 7 Some examples of datasets

4.3 实验

4.3.1 改进注意力机制消融实验

为了验证注意力机制的有效性,采用消融实验的方法对改进进行验证。表 1 展示了 MobileNetv3-YOLOv4、CBAM+YOLOv4-Lite 和 E-YOLOv4-Lite 在 RPC 数据集上的表现,其中参数增量以 MobileNetv3-YOLOv4 为基础进行对比,mAP 表示平均检测精度,实验中将 CBAM 的共享全连接层的通道压缩率设置为 16。

从表 1 可以看出:改进的注意力机制使 mAP 提高了 2 个百分点,且参数增量很少;而 CBAM 使参数增

表 1 改进注意力机制消融实验

Table 1 Improved attention mechanism ablation experiment

Model	mAP / %	Parameter
MobileNetv3-YOLOv4	93.25	0
CBAM+YOLOv4-Lite	94.75	94894
E-YOLOv4-Lite	95.43	196

量巨大,mAP 却只提升了 1 个百分点。这是由于 CBAM 中采用的部分池化难以获取全部特征,而改进的注意力采用全局池化,尽量获取到图片的全部特征。即在 YOLOv4-Lite 网络中引入改进的 E-CBAM 机制能够在有效提高 mAP 的同时,不会引入大量的参数,对轻量化系统开发具有很好的应用意义。

4.3.2 检测效果对比实验分析

为了验证算法的性能,将 E-YOLOv4-Lite、文献 [11] 中的 YOLOv4-Lite、目前主流的轻量化算法 YOLOv4-tiny 及经典的单阶段检测算法 SSD 等 4 种算法进行了对比实验。参数量用来描述目标检测算法的复杂程度,各模型在商品数据集中训练的参数量如表 2 所示。由表 2 可以看出,E-YOLOv4-Lite 相较于 YOLOv4-Lite 参数量基本持平,而 SSD 拥有比较大的参数量。虽然 E-YOLOv4-Lite 在引入了 E-CBAM 后参数量微量上升,但网络模型的性能提高显著。

表 3 为使用 4 种模型分别对物品进行检测的结果,其中精确率 (Precision) 计算正确检测目标数与整体检

表 2 各模型参数量对比

Table 2 Comparison of parameters of each model

Model	Parameter
YOLOv4-tiny	6062814
YOLOv4-Lite	11791741
E-YOLOv4-Lite	11791937
SSD	26285486

表 3 4 种模型进行物品检测的综合性能对比

Table 3 Comparison of comprehensive performance of 4 models for item detection

Model	Precision / %	Recall / %	F1	FPS	Time / s	mAP / %
YOLOv4-tiny	98.95	94.89	0.9677	70.3	0.0141	94.53
YOLOv4-Lite	97.73	93.47	0.9493	71.1	0.0140	93.25
E-YOLOv4-Lite	99.03	95.51	0.9718	69.2	0.0144	95.43
SSD	96.02	91.64	0.9378	29.1	0.0344	91.59

从表 3 可以看出:所提 E-YOLOv4-Lite 模型在准确率上达到 99.03%,具有优秀的精确度;在召回率上达到 95.51%,具有良好的查全性;在 F1 得分为 0.972,具有良好的检测性能;FPS 与 YOLOv4-Lite 相差不多,但相对于经典的 SSD 有明显提升;检测时间略有上升,mAP 相较于 YOLOv4-Lite 提高了 2.18 个百分点,相较于 YOLOv4-tiny 提高了 0.9 个百分点。

测目标数的比例,用于衡量模型分类准确能力,召回率(Recall)通过计算正确检测目标数与验证集总数的比例得到,用于衡量模型查全的能力,平衡分数(F1)是精确率和召回率的调和平均数,用于衡量算法的性能,FPS 为每秒动态视频流的信息数量,数值越高,画面越流畅,mAP 是在不同的召回率下的平均检测精度,用来描述目标检测算法的性能。

图 8 展示了 YOLOv4-tiny 与 E-YOLOv4-Lite 模型对于商品的检测效果:E-YOLOv4-Lite 模型在有遮挡的条件下具有更高的检测精度,无漏检或者误检情况;而 YOLOv4-tiny 在检测图 8 第 2 张图片时,存在明显的小商品漏检情况,说明 E-YOLOv4-Lite 能够有效检测遮挡商品。因此,所提算法具有良好的检测精度,检测效果基本满足仓储环境下的精度要求。

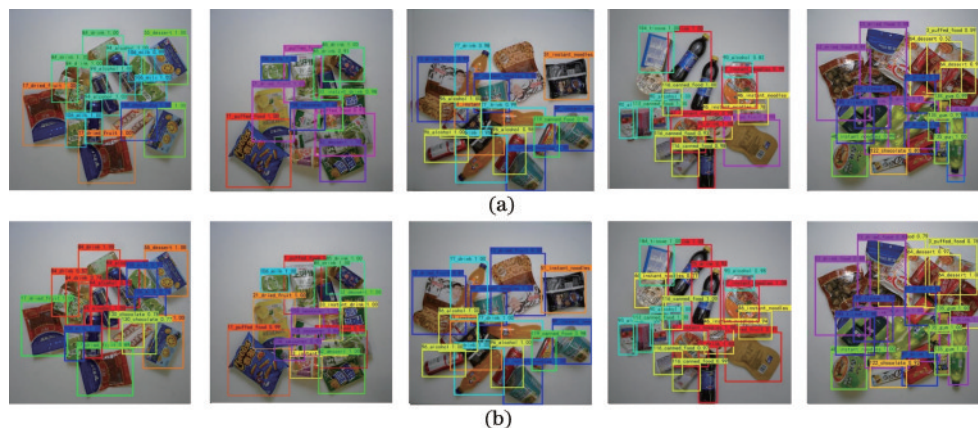


图 8 两种模型检测效果对比。(a) YOLOv4-tiny; (b) E-YOLOv4-Lite

Fig. 8 Comparison of detection effects of two models. (a) YOLOv4-tiny; (b) E-YOLOv4-Lite

5 结 论

提出了一种用于仓储货品检测的轻量化目标检测模型 E-YOLOv4-Lite。将 YOLOv4 的主干特征提取网络替换为参数量更少的 MobileNetv3,在 PANet 中以深度可分离卷积代替标准卷积,激活函数改用 Hard-swish,使得模型的参数量仅为 YOLOv4 的 1/5,更轻量化,适合于智能仓储环境中的嵌入式移动设备的部署。同时,在 PANet 中引入改进的 CBAM 注意力模型,进一步提高了对遮挡物品的检测精度,更好满足仓储环境中的需求。

选用公开 RPC 商品数据集中的 12000 张多目标图片作为基础数据集,对改进注意力机制的效果进行消

融验证,然后将所提方法与 YOLOv4-tiny、文献[11]中的 YOLOv4-Lite 进行对比实验。实验结果表明,改进后的 E-YOLOv4-Lite 模型具有更好的精确度和实时性,mAP 为 95.43%,FPS 达到了 69.2,检测时间仅为 14.4 ms,模型参数量仅为 11.25 MB。E-YOLOv4-Lite 模型在提高了精度的同时,引入了极少参数量,增强了实时性,能够更好满足仓储环境下的嵌入式开发场景。未来将继续探索 E-YOLOv4-Lite 模型在其他领域的应用,如图像分割、抓取定位等。

参 考 文 献

- [1] 中华人民共和国国家发展和改革委员会. 关于印发《推动物流业制造业深度融合创新发展实施方案》的通知

- [EB/OL]. (2020-08-22)[2021-04-05]. https://www.ndrc.gov.cn/xxgk/zcfb/tz/202009/t20200909_1237841.html. National Development and Reform Commission. Notice on printing and distributing the implementation plan for promoting the deep integration, innovation and development of logistics industry and manufacturing industry[EB/OL]. (2020-08-22) [2021-04-05]. https://www.ndrc.gov.cn/xxgk/zcfb/tz/202009/t20200909_1237841.html.
- [2] 尹仕斌, 任永杰, 刘涛, 等. 机器视觉技术在现代汽车制造中的应用综述[J]. 光学学报, 2018, 38(8): 0815001. Yin S B, Ren Y J, Liu T, et al. Review on application of machine vision in modern automobile manufacturing[J]. Acta Optica Sinica, 2018, 38(8): 0815001.
- [3] 来文豪, 周孟然, 胡锋, 等. 基于多光谱成像和改进 YOLO v4 的煤矸石检测[J]. 光学学报, 2020, 40(24): 2411001. Lai W H, Zhou M R, Hu F, et al. Coal gangue detection based on multi-spectral imaging and improved YOLO v4 [J]. Acta Optica Sinica, 2020, 40(24): 2411001.
- [4] 孙迎春, 潘树国, 赵涛, 等. 基于优化 YOLOv3 算法的交通灯检测[J]. 光学学报, 2020, 40(12): 1215001. Sun Y C, Pan S G, Zhao T, et al. Traffic light detection based on optimized YOLOv3 algorithm[J]. Acta Optica Sinica, 2020, 40(12): 1215001.
- [5] 来文豪, 周孟然, 王锦国, 等. 多光谱波段筛选的煤矸石快速定位[J]. 中国激光, 2021, 48(16): 1611001. Lai W H, Zhou M R, Wang J G, et al. Fast location of coal gangue based on multispectral band selection[J]. Chinese Journal of Lasers, 2021, 48(16): 1611001.
- [6] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-04-05]. <https://arxiv.org/abs/2004.10934>.
- [7] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [8] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [9] Ren S Q, He K M, 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.
- [10] Yu Z W, Shen Y G, Shen C K. A real-time detection approach for bridge cracks based on YOLOv4-FPM[J]. Automation in Construction, 2021, 122: 103514.
- [11] 王金鹏, 高凯, 姜洪喆, 等. 基于改进的轻量化卷积神经网络火龙果检测方法(英文)[J]. 农业工程学报, 2020, 36(20): 218-225. Wang J P, Gao K, Jiang H Z, et al. method for detecting dragon fruit based on improved lightweight convolutional neural network[J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(20): 218-225.
- [12] 蔡舒平, 孙仲鸣, 刘慧, 等. 基于改进型 YOLOv4 的果园障碍物实时检测方法[J]. 农业工程学报, 2021, 37(2): 36-43. Cai S P, Sun Z M, Liu H, et al. Real-time detection methodology for obstacles in orchards using improved YOLOv4[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(2): 36-43.
- [13] 朱惠玲, 牛哲文, 黄克灿, 等. 基于单阶段目标检测算法的变电设备红外图像目标识别及定位[J]. 电力自动化设备, 2021, 41(8): 217-224. Zhu H L, Niu Z W, Huang K C, et al. Identification and location of infrared image for substation equipment based on single-stage object detection algorithm[J]. Electric Power Automation Equipment, 2021, 41(8): 217-224.
- [14] 徐利锋, 黄海帆, 丁维龙, 等. 基于改进 DenseNet 的水果小目标检测[J]. 浙江大学学报(工学版), 2021, 55(2): 377-385. Xu L F, Huang H F, Ding W L, et al. Detection of small fruit target based on improved DenseNet[J]. Journal of Zhejiang University (Engineering Science), 2021, 55(2): 377-385.
- [15] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 1314-1324.
- [16] Howard A G, Zhu M L, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2021-04-05]. <https://arxiv.org/abs/1704.04861>.
- [17] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.
- [18] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.
- [19] Wei X S, Cui Q, Yang L, et al. RPC: a large-scale retail product checkout dataset[EB/OL]. (2019-01-22) [2021-04-05]. <https://arxiv.org/abs/1901.07249>.