doi: 10.6041/j.issn.1000-1298.2022.08.031

基于改进 YOLO v4 的自然环境苹果轻量级检测方法

王　卓[1,2]　王　健[1,3]　王枭雄[1,3]　时　佳[1,2]　白晓平[1,2]　赵泳嘉[1,2]

(1. 中国科学院沈阳自动化研究所，沈阳 110016；2. 中国科学院机器人与智能制造创新研究院，沈阳 110169；
3. 中国科学院大学计算机科学与技术学院，北京 100049)

摘要： 针对苹果采摘机器人识别算法包含复杂的网络结构和庞大的参数体量，严重限制检测模型的响应速度问题，本文基于嵌入式平台，以 YOLO v4 作为基础框架提出一种轻量化苹果实时检测方法(YOLO v4-CA)。该方法使用 MobileNet v3 作为特征提取网络，并在特征融合网络中引入深度可分离卷积，降低网络计算复杂度；同时，为弥补模型简化带来的精度损失，在网络关键位置引入坐标注意力机制，强化目标关注以提高密集目标检测以及抗背景干扰能力。在此基础上，针对苹果数据集样本量小的问题，提出一种跨域迁移与域内迁移相结合的学习策略，提高模型泛化能力。试验结果表明，改进后模型的平均检测精度为 92.23%，在嵌入式平台上的检测速度为 15.11 f/s，约为改进前模型的 3 倍。相较于 SSD300 与 Faster R-CNN，平均检测精度分别提高 0.91、2.02 个百分点，在嵌入式平台上的检测速度分别约为 SSD300 和 Faster R-CNN 的 1.75 倍和 12 倍；相较于两种轻量级目标检测算法 DY3TNet 与 YOLO v5s，平均检测精度分别提高 7.33、7.73 个百分点。因此，改进后的模型能够高效实时地对复杂果园环境中的苹果进行检测，适宜在嵌入式系统上部署，可以为苹果采摘机器人的识别系统提供解决思路。

关键词： 采摘机器人；苹果检测；YOLO v4；轻量化；注意力机制；迁移学习

中图分类号： TP391.4；S24　　**文献标识码：** A　　**文章编号：** 1000-1298(2022)08-0294-09　　**OSID:**

Lightweight Real-time Apple Detection Method Based on Improved YOLO v4

WANG Zhuo[1,2]　WANG Jian[1,3]　WANG Xiaoxiong[1,3]　SHI Jia[1,2]　BAI Xiaoping[1,2]　ZHAO Yongjia[1,2]

(1. *Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang* 110016, *China*
2. *Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang* 110169, *China*
3. *School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing* 100049, *China*)

Abstract: Under the picking conditions in unstructured environments, such as overlapping and occlusion, the recognition system based on deep learning in apple picking robot contained complex network structure and large parameter volumes, for which the response speed of detection model was severely limited. In response to this problem, based on the embedded platform, a lightweight apple real-time detection method called YOLO v4-CA, which selected YOLO v4 as the basic framework, was proposed. The proposed method used MobileNet v3 as the feature extraction network, and introduced deep separable convolution in the feature fusion network to reduce network computational complexity. In order to ensure the detection accuracy, coordinate attention was introduced in the key position of the network to strengthen target attention, which can improve the ability to detect dense targets and resist background interference. For the small apple datasets, a combination of cross-domain and in-domain transfer learning strategy was proposed to improve the generalization ability of the model. Experimental results showed that the average precision of the improved model was 92.23%, and the detection speed on the embedded hardware platform was 15.11 frames per second, which was about three times than that of the original YOLO v4 model. Compared with the two representative target detection algorithms of SSD300 and Faster R-CNN, the average precision was increased by 0.91 percentage points and 2.02 percentage points respectively, and the detection speed on the embedded hardware platform was about 1.75 times and 12 times that of the two respectively. Compared with the two lightweight target detection algorithms of

收稿日期： 2021-08-25　**修回日期：** 2021-10-18
基金项目： 国家重点研发计划项目(2020YFB1709603)
作者简介： 王卓(1976—)，男，研究员，博士生导师，主要从事农机精准作业控制技术研究，E-mail：zwang@sia.cn
通信作者： 时佳(1992—)，女，助理研究员，主要从事智能装备研究，E-mail：shijia@sia.cn

DY3TNet and YOLO v5s, the average precision was increased by 7.33 percentage points and 7.73 percentage points respectively. Therefore, the improved model YOLO v4 - CA can efficiently detect apples in a complex orchard environment in real time, and it was suitable for deployment on embedded systems. It can provide solutions for the recognition system of apple picking robots.

Key words: picking robot; apple detection; YOLO v4; lightweight; attention mechanism; transfer learning

0　引言

苹果是我国规模最大的果品之一，苹果园约占全国果园的 18%，年产量约为 4.139×10^7 t[1]。然而由于果园环境复杂，苹果的采摘依旧以人工采摘为主，采摘成本高，效率低，因此，研究苹果采摘机器人代替人工进行自动化采摘具有重要意义。采摘机器人主要由视觉系统和机械臂系统组成[2]，机械臂系统受视觉系统引导完成对果实的采摘，因而对果实进行快速、精准地识别与定位是实现自动采摘的关键[3]。

果园环境较为复杂，枝叶遮挡、果实重叠、光照变化等会影响模型的检测精度，造成误检、漏检等问题; 另外，由于采摘机器人搭载的嵌入式平台算力资源有限，复杂模型的检测速度无法满足任务实时性需求，且难以部署。在保证检测精度的同时提高检测速度成为非结构环境下苹果检测主要的难点问题和研究热点。

近年来，深度学习技术不断发展，基于卷积神经网络的苹果检测算法也因其鲁棒性强、自适应性强以及准确性高而被广泛应用[4-6]。其中，应用于苹果检测任务中的算法主要分为两类，一类是侧重于精度，将检测分为定位和分类两个过程的 two - stage 算法，如 Faster R - CNN[7]、R - FCN[8] 等，GAO 等[9] 针对枝叶遮挡问题，使用改进的 Faster R - CNN 网络对密叶果树中的苹果进行检测，mAP 为 87.9%，单幅图像平均检测时间为 0.241 s。另一类是侧重于速度，将检测过程中的定位和分类转化为回归问题的 one - stage 算法，如 YOLO[10]、SSD[11] 等。张恩宇等[12] 将 SSD 算法与 U 分量阈值分割法相结合识别自然环境中的青苹果，拥有较好的检测效果; 武星等[13] 使用一种轻量化的 YOLO v3 卷积神经网络检测苹果，mAP 为 94.69%，工作站和嵌入式开发板上的检测速度分别为 116.96、7.59 f/s; FU 等[14] 基于 YOLO v3 - tiny 提出了 DY3TNet 模型，对果园中的猕猴桃进行检测，平均检测精度达 90.05%，GPU 下单幅图像检测时间为 34 ms，实现了猕猴桃的快速检测。目前，基于高性能平台开展的苹果检测研究，已取得阶段性进展，而在算力资源有限的嵌入式设备上，检测精度与速度的平衡值得进一步研究。

本文以果园中非结构环境中的苹果作为检测任务，针对算力资源有限的嵌入式平台，提出一种轻量化苹果实时检测方法 YOLO v4 - CA。该方法以 YOLO v4 为基础框架，基于 MobileNet v3 改进网络主干，并使用深度可分离卷积优化特征融合网络，压缩模型，减少模型计算量; 引入坐标注意力机制，弥补因模型轻量化以及非结构化环境所造成的精度损失; 提出一种将跨域迁移与域内迁移相结合的学习策略，提高模型的泛化能力。在台式计算机及嵌入式平台 Jetson AGX Xavier 上分别将本文提出的检测算法与主流目标检测模型进行对比。

1　改进的自然环境苹果检测方法

1.1　YOLO v4 网络结构

YOLO v4[15] 是目前最先进的实时检测模型之一，它在 YOLO v3 的基础上进一步优化，使得总体性能显著提高。其网络结构有 3 大改进: CSPDarkNet53 替换 DarkNet53 作为特征提取网络，促进底层信息融合，增强特征提取能力; 提出空间金字塔池化模块 SPP[16]，在最后一层输出中进行 4 个不同尺度的最大池化操作，有效提高感受野，提取出最显著的上下文特征; 将特征金字塔网络 FPN[17] 结构修改为路径聚合网络 PAN[18]，在 FPN 的自底向上结构中添加一个自顶向下的结构，进一步提取和融合不同尺度特征。

1.2　网络结构轻量化改进

YOLO v4 在多类别检测任务中具有优异的识别精度和速度，而本文所研究的识别任务仅对苹果进行单类识别，原始模型具有参数冗余，存在不必要的计算开销，另外，采摘机器人多搭载嵌入式设备部署识别任务，算力资源有限，冗余的计算会影响模型的检测速度。因此，为使模型部署在嵌入式设备时满足实时性需求，本文基于 YOLO v4 对其特征提取网络和特征融合网络进行轻量化改进。

1.2.1　基于 MobileNet v3 特征提取网络的结构改进

MobileNet 是一种适用于移动端的轻量级神经网络。本文使用 MobileNet v3[19] 轻量级神经网络作为 YOLO v4 - CA 的特征提取网络。MobileNet v3 保留 MobileNet v2[20] 中具有线性瓶颈层的逆残差结

构，并将 SENet[21] 中的轻量级注意模块集成其中作为 bneck 基本块，提高网络对于特征通道的敏感程度，增强网络的特征提取能力；在深层网络中使用 h－swish代替 ReLU，降低运算量，提高模型性能。本文所使用的 MobileNet v3 网络参数如表 1 所示，将特征层 8、14、17 提取到的特征图输出，作为后续特征融合层的输入。

表 1 MobileNet v3 主干

Tab. 1 MobileNet v3 backbone

特征层	输入尺度	基本单元	通道注意力	激活函数
1	416×416×3	Conv2D	不施加	h－swish
2	208×208×16	bneck,3×3	不施加	ReLU
3	208×208×16	bneck,3×3	不施加	ReLU
4	104×104×24	bneck,3×3	不施加	ReLU
5	104×104×24	bneck,5×5	施加	ReLU
6	52×52×40	bneck,5×5	施加	ReLU
7	52×52×40	bneck,5×5	施加	ReLU
8	52×52×40	bneck,3×3	不施加	h－swish
9	26×26×80	bneck,3×3	不施加	h－swish
10	26×26×80	bneck,3×3	不施加	h－swish
11	26×26×80	bneck,3×3	不施加	h－swish
12	26×26×80	bneck,3×3	施加	h－swish
13	26×26×112	bneck,3×3	施加	h－swish
14	26×26×112	bneck,5×5	施加	h－swish
15	13×13×160	bneck,5×5	施加	h－swish
16	13×13×160	bneck,5×5	施加	h－swish
17	13×13×160	Conv2D,1×1	不施加	h－swish

1.2.2 基于深度可分离卷积的特征融合网络结构改进

深度可分离卷积[22] 将卷积过程分解为逐通道卷积和逐点卷积，相较于传统卷积能够大幅减少参数计算量，将 YOLO v4 特征融合部分路径聚合网络(PAN)结构中的普通卷积替换为深度可分离卷积，进一步压缩模型，提高模型计算效率。

网络结构的轻量化改进能够大幅降低模型的参数量和计算量，但与此同时会带来检测精度上的损失，因此，需要对模型进行进一步优化以提高模型检测精度。

1.3 引入坐标注意力机制的特征融合网络

注意力机制是一种仿生物视觉机制。通过快速扫描全局图像，筛选出感兴趣的区域，投入更多的注意力资源，并抑制其他无用信息，提高视觉信息处理的效率与准确性[23]。

自然环境下的苹果常出现果实重叠和枝叶遮挡的问题，造成模型检测精度的损失，本文使用一种将位置信息与通道信息相结合的坐标注意力机制[24] 施加于网络的关键位置中，增加模型对苹果特征的敏感程度。对于任务中较难识别的重叠、遮挡目标分配高权重以增加关注度，对于不感兴趣的自然背景分配低权重加以抑制，提高自然环境下苹果的识别精度。

如图 1 所示，坐标注意力机制(Coordinate attention，CA)包含信息嵌入以及注意力生成两部分。信息嵌入阶段对特征进行汇聚，对输入特征图的所有通道，分别沿水平坐标和垂直坐标方向进行平均池化，获取到尺寸为 $C\times H\times 1$ 和 $C\times 1\times W$ 的特征图。在注意力生成阶段，将获取到的两幅特征图拼接为 $C\times 1\times(H+W)$ 的特征图，然后采用 1×1 卷积将其通道维数以收缩率 r 从 C 维压缩至 C/r 维，并使用 ReLU 函数进行非线性激活，再将获取到的结果沿空间维分解为 $C/r\times H\times 1$ 的水平注意张量和 $C/r\times 1\times W$ 的垂直注意张量。之后，再使用两组 1×1 的卷积将通道维从 C/r 维升至 C 维，并使用 Sigmoid 函数进行非线性激活。最后，将获取到的两个注意图 $C\times H\times 1$ 和 $C\times 1\times W$ 与输入的特征图相乘，完成坐标注意力的施加。将坐标注意力机制引入至特征融合网络 PAN，如图 2 所示的位置 2 处，位于信息交汇处，使得坐标注意力能够充分获取不同尺度的特征信息，通过两个不同方向注意图的施加，判断目标是否存在于注意图对应的行与列中，提升网络对密集目标的识别效果，缓解枝叶遮挡、果实重叠引起的检测精度损失。另外，图 2 所示的网络中于位置 1 及位置 3 处所施加的 CA 模块仅用于后续对照试验说明用，不作为最终网络结构的一部分。

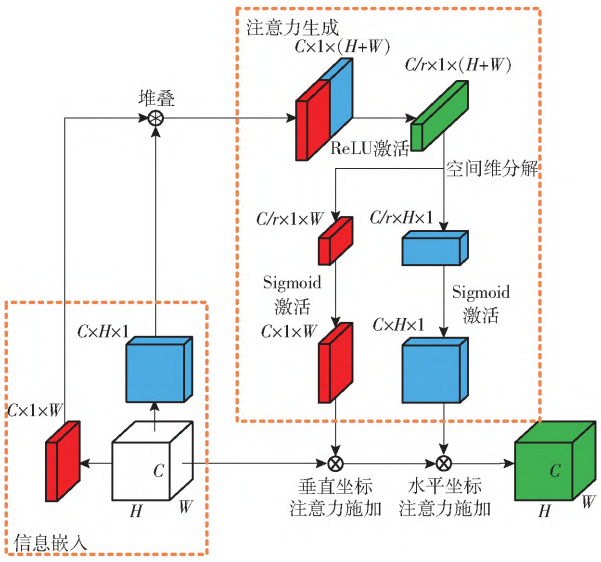

图 1 坐标注意力机制

Fig. 1 Coordinate attention mechanism

基于 YOLO v4 模型进行网络轻量化改进，于特征融合层引入坐标注意力机制后的网络(YOLO v4－CA)结构如图 3 所示。

1.4 跨域迁移与域内迁移相结合的学习策略

模型的训练需要大量数据，大规模的苹果数据

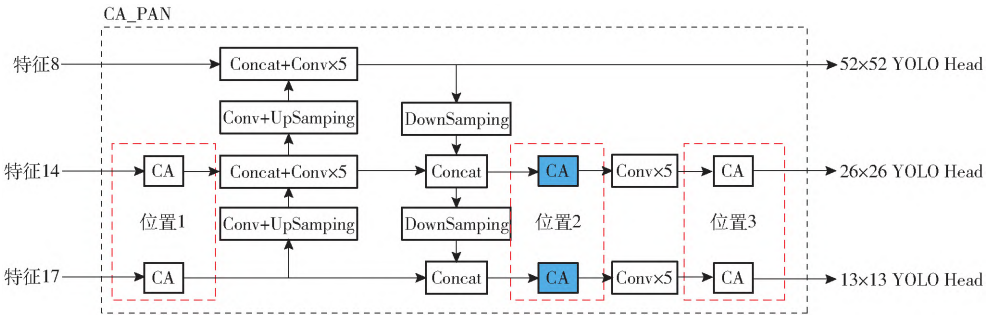

图 2　施加坐标注意力机制的特征融合网络

Fig. 2　Feature fusion network with coordinate attention mechanism

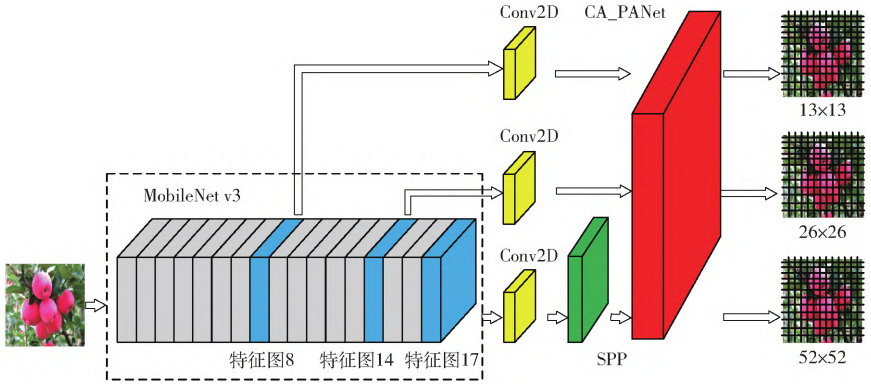

图 3　YOLO v4 − CA 网络结构

Fig. 3　YOLO v4 − CA network structures

集获取比较困难，成本高，而当数据不充足时，通常采用数据增强的方式扩充数据集，并以跨域迁移的方式进行知识迁移，对于苹果检测任务，识别对象为自然环境中的苹果，迁移前后源域与目标域相似度低，通常方法所带来的精度提升有限。因此，本文针对苹果检测任务提出一种将跨域迁移与域内迁移相结合的学习策略[25]，在通常的源域与目标域之间，即通用数据集与自然环境苹果数据集之间添加仅含有苹果特征的数据集作为过渡域，并采用亮度调整以及缩放的方式进行数据扩充，丰富数据集中不同光影及尺度下的苹果特征，减少其与目标域的差异性，提升迁移学习的效果，进而提高模型的检测精度。

具体地，跨域迁移与域内迁移相结合的学习策略分为 2 个阶段: 进行跨域迁移学习，使用通用数据集下训练得到的参数对网络主干部分进行初始化，并利用仅含有苹果特征的数据集对模型进行微调，习得苹果特征; 进行域内迁移学习，利用自然环境苹果数据集在阶段 1 训练好的模型上进一步微调，习得受复杂环境影响的苹果特征。

2　网络训练与检测试验

2.1　数据集准备

本文试验所采用数据集分为两部分。数据集 1 来自于开源的 Fruit − 360 数据集[26]，该数据集包含 120 种不同的水果和蔬菜，每幅图像均取自实验室环境，并在获取后去除目标以外的背景，提取其中 Braeburn、Crimson Snow、Pink Lady、Red 1、Red 2、Red 3 共 6 个品种的苹果图像，共计 3 767 幅图像，其中训练集为 2 804 幅图像，测试集为 963 幅图像。数据集 2 使用自建数据集，图像源自互联网，以苹果、苹果树、自然环境苹果等作为关键词进行检索获得，经过筛选，保留 1 057 幅图像作为数据集，并以 8∶1∶1的比例将其分为训练集 845 幅，验证集 106 幅，测试集 106 幅，根据样本的遮挡情况对数据集进行划分，划分结果如表 2 所示，其中，轻度遮挡样本为平均遮挡程度小于 30% 的样本，重度遮挡样本为平均遮挡程度大于 30% 的样本。

表 2　数据集 2 遮挡情况及其数量

Tab. 2　Occlusion and quantity of datasets 2　幅

数据集	无遮挡样本个数	轻度遮挡样本个数	重度遮挡样本个数	总数量
训练集	2 015	1 569	1 022	4 606
验证集	244	189	82	515
测试集	274	251	109	634

使用 LabelImg 图像标注工具对数据集进行人工标注，在标注过程中忽略图像中遮挡超过 80% 的目标，获得 PASCAL VOC 格式的 XML 文件作为标签文件。将数据集 2 中图像的分辨率调整为网络输入时所需要的 416 像素 × 416 像素，使用 K-means 算法对标签中边界框的尺寸进行聚类，聚类中心设

置为 9，将聚类结果作为网络的先验框，分别为(15，21)、(28，35)、(40，53)、(52，73)、(67，97)、(83，62)、(91，116)、(113，159)、(167，214)。

为提高二阶段域内迁移学习效果，对数据集 1 进行增强以增加源域与目标域的相似性。将图像变换至 HSI 空间，将 I 通道下数据随机调整至原来的 0.8～4.0 倍；根据数据集 2 的先验框尺寸，将图像进行 0.15～2 的随机等比例缩放。进行上述操作，将数据集 1 扩充至原来的 4 倍，共计 11 966 幅图像，处理效果如图 4 所示。

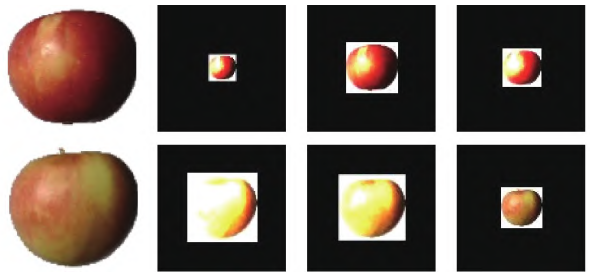

图 4 数据集 1 扩充后图像

Fig. 4 Images of expanded datasets 1

2.2 损失函数

本文模型的损失函数(Loss)由置信度损失、类别损失以及边界框损失 3 部分组成。

2.3 评价标准

本文以准确率 P、召回率 R、平均精度(Average precision，AP)作为模型检测精度的评价指标。

另外，从 3 方面评价模型的性能，使用单位时间图像检测数量(fps)评价模型的检测速度，浮点数计算量(FLOPs)评价模型的计算复杂度，内存占用量评价模型的大小。

2.4 试验平台

本文模型训练平台为台式计算机，配置为 Intel(R) Core(TM) i7－8700 3.20 GHz CPU，内存 16 GB，GPU 为 NVIDIA TITAN V，显存 12 GB，运行环境为 Windows 10 系统，Python 版本为 3.6，Pytorch 版本为 1.2.0，CUDA 版本为 10.0，cuDNN 版本为 7.4.1。模型测试平台除上述高性能台式机外，还有一台 Jetson AGX Xavier 嵌入式平台，搭载 NVIDIA Carmel ARMv8.2 CPU、GPU 为 NVIDIA Volta，能够达到每秒 11 万亿次浮点数计算量，运行环境为 Ubuntu 系统，Python 版本为 3.6，Pytorch 版本为 1.6.0，CUDA 版本为 10.2，cuDNN 版本为 8.0.0。

2.5 模型训练

模型的训练策略分为两阶段，阶段 1 进行跨域迁移学习，利用大规模数据集 VOC 训练好的 MobileNet v3 网络预训练权重对网络参数进行初始化，并利用 Fruit－360 数据集对网络进行微调；阶段 2 进行域内迁移学习，利用自建的苹果数据集对网络进行进一步微调。

训练过程分为两步，首先，冻结网络骨架部分，批量大小为 64，初始学习率为 1×10^{-3}，训练轮次为 50；接着，解冻训练，批量大小为 16，初始学习率为 1×10^{-4}，训练轮次为 50。训练中所使用到的优化器均为 Adam，参数为默认值，每训练一轮学习率衰减为原来的 0.9。

训练过程中使用 Tensorboard 记录数据，每进行一次迭代，写入训练集损失；每训练一个轮次，写入验证集损失，并保存模型权重。损失值变化曲线如图 5 所示，共训练 100 个轮次，将后 50 轮次中验证集损失最低的模型作为训练结果以进行后续分析。

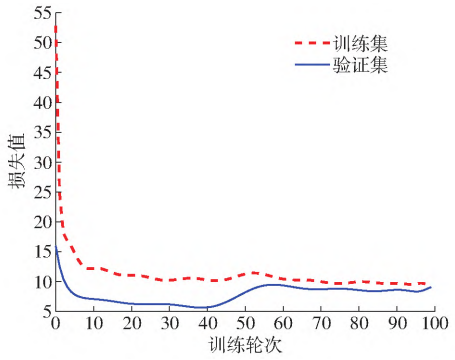

图 5 损失值变化曲线

Fig. 5 Change curves of loss value

3 试验与结果分析

为验证本文针对苹果检测任务所设计方法的有效性，比较不同策略施加前后对模型性能的影响，在相同测试集下与改进前模型、两种常用目标检测模型以及两种轻量级目标检测模型进行综合对比。

3.1 网络轻量化对模型检测效果的影响

由表 3 可知，将 YOLO v4 的原特征提取网络 CSPDarkNet53 替换为 MobileNet v3，并将特征融合网络中的普通卷积替换为深度可分离卷积后，模型的浮点数计算量降低 88.38%，模型内存占用量降低 78.03%，AP 降低 6.71 个百分点。说明网络轻量化能够有效降低模型计算量、内存占用量，但同时会损失一定检测精度，因此，需要对模型进一步优化，提高综合能力。

表 3 网络轻量化对模型的影响

Tab. 3 Effect of network lightweighting on model

模型	AP/%	FLOPs	模型内存占用量/MB
YOLO v4	93.23	5.95×10^{10}	244.0
网络轻量化模型	86.52	6.92×10^{9}	53.6

3.2　坐标注意力机制施加不同位置对模型检测效果的影响

在网络轻量化模型的基础上，将坐标注意力机制 CA 施加在图 2 所示的特征融合网络中的不同位置，对比施加位置不同对模型检测能力的影响，结果如表 4 所示，在位置 2 处施加使模型 AP 提高了 1.21 个百分点，而在位置 1 和位置 3 处施加分别使模型 AP 降低了 6.82、1.33 个百分点，说明 CA 在特征融合网络中不同位置的施加并不一定会带来模型检测性能的提升，而由于位置 2 处于特征提取网络中不同尺度信息的交汇处，相较于位置 1 和位置 3 能够使注意力机制进行信息嵌入阶段获取更加丰富的特征信息，进而提升模型的检测效果。另外，由表 4 可知，CA 于不同位置施加时模型额外内存占用量较低。

表 4　施加注意力机制至不同位置检测能力对比

Tab. 4　Comparison of detection capabilities with attention mechanism at different locations

施加位置	AP/%	模型内存占用量/MB
无	86.52	53.6
位置 1	79.70	53.7
位置 2	87.73	54.1
位置 3	85.19	53.7

3.3　不同注意力机制对模型检测效果的影响

在网络轻量化模型的基础上，在图 2 所示的位置 2 处施加不同的注意力机制，对比不同注意力机制对模型检测能力的影响，由表 5 可知，施加 SE 在收缩率为 32 时模型 AP 最高，为 86.74%；施加 CBAM 在收缩率为 8 时模型 AP 最高，为 86.26%；施加 CA 在收缩率为 32 时模型 AP 最高，为 87.53%，相较于施加前，SE、CA 分别提高 0.22、1.01 个百分点，CBAM 降低 0.26 个百分点，说明在模型特征融合网络中施加注意力机制并不一定能够带来检测精度的提升，需根据特定任务加以选择。本文所引入的 CA 模块使用两个一维注意图进行特征编码，通过嵌入不同尺度的特征信息，以一种近似于坐标的形式决定图像中目标的关注程度，能够有效提高模型对于密集目标的敏感程度，进而改善苹果检测任务中果实重叠、枝叶遮挡对检测精度带来的负面影响。另外，由表 5 可知，于特征融合网络中施加不同注意力机制带来的额外内存占用量较低，结合表 4 得出以下结论: 对于内存空间及算力受到约束的任务中，可通过在网络中施加合适的注意力机制改善模型的检测性能。

表 5　施加不同注意力机制的检测能力对比

Tab. 5　Comparison of detection capabilities with different attention mechanisms

注意力机制	收缩率	AP/%	模型内存占用量/MB
无		86.52	53.6
SE	8	84.74	54.9
	16	85.73	54.2
	32	86.74	53.9
CBAM	8	86.26	54.9
	16	85.23	54.2
	32	80.30	53.9
CA	8	86.16	55.5
	16	84.61	54.6
	32	87.53	54.1

综上所述，将 YOLO v4 的原特征提取网络 CSPDarkNet53 替换为 MobileNet v3，并将特征融合网络中的普通卷积替换为深度可分离卷积，同时于图 2 所示位置 2 处施加收缩率为 32 的 CA 模块能够使改进后模型检测精度达到最佳，因此，后续对比试验的讨论基于该网络结构展开。

3.4　不同迁移学习方式对模型检测效果的影响

比较不同迁移学习方式对模型检测精度的影响。使用 VOC 预训练模型在数据集 2 上进行训练，作为跨域迁移学习; 对模型进行随机初始化并先后在数据集 1 和数据集 2 上进行训练，作为域内迁移学习; 使用 VOC 预训练模型先后在数据集 1 和数据集 2 上进行训练，作为跨域迁移与域内迁移相结合的学习方式。由表 6 可知，进行跨域迁移与域内迁移相结合的学习方式使模型精度达到最优，相较于单独进行跨域迁移和域内迁移的 AP 分别提高 4.7、19.87 个百分点，这是由于两者相结合的学习方式分两阶段进行，在模型掌握通用特征后学习苹果特征，进而再学习自然环境下的苹果特征，相较于跨域迁移，添加过渡域以缓解因源域与目标域相似性低所带来的负面影响; 而相较于域内迁移，通过通用特征对模型进行初始化，弥补因数据集 1 中不具备背景所造成的信息损失，因而能够获得 3 种学习方式中最佳的模型检测精度，具有最强的泛化能力。

表 6　不同迁移学习方式检测能力对比

Tab. 6　Comparison of detection capabilities with different transfer learning methods

学习方式	数据集	AP/%
跨域	VOC 数据集 + 数据集 2	87.53
域内	数据集 1 + 数据集 2	72.36
跨域与域内结合	VOC 数据集 + 数据集 1 + 数据集 2	92.23

3.5　不同检测模型对比试验

为验证本文模型的效果，在相同测试集下，分别与 YOLO v4、SSD300、Faster R - CNN、DY3TNet 以及

YOLO v5s 进行对比，对比结果如表 7、8 所示。对无遮挡、轻度遮挡以及重度遮挡 3 种情况下模型检测效果的对比如图 6 所示，其中红色矩形框为预测结果，橙色圆形框为误检目标，黄色圆形框为漏检目标。由表 7 可知，本文模型的 AP 为 92.23%，相比于 YOLO v4 降低了 1.00 个百分点，相比于 SSD300 及 Faster R－CNN 提升 0.91、2.02 个百分点，相比于 DY3TNet 及 YOLO v5s 提升 7.33、7.73 个百分点。由图 6 可知，YOLO v4－CA 对于无遮挡以及轻度遮挡情况下的样本检测效果优异，而对于遮挡情况较为严重的样本，与 YOLO v4、SSD300 以及 Faster R－CNN 的检测效果相近，但依然存在漏检现象，这是由于遮挡超过 80% 的目标默认不做标注。另外，YOLO v4－CA 在 6 种检测模型中拥有最高的识别准确率，即拥有最低的误检率，避免了在苹果采摘过程中出现误检现象而造成机械臂的误操作，提高了机器人的整体采摘效率，因此，拥有高识别准确率的 YOLO v4－CA 更适合于苹果采摘任务。

由表 8 可知，本文模型内存占用量为 54.1 MB，约为 YOLO v4 的 1/4，SSD300 及 Faster R－CNN 的 1/2，DY3TNet 和 YOLO v5s 的 2 倍；计算量相较于 YOLO v4 减少 87%、相较于 SSD300 及 Faster R－CNN 分别减少 89% 和 93%，相较于 DY3TNet 及 YOLO v5s 减少 44% 和 59%；在高性能台式机平台上单幅图像的检测速度与 YOLO v4 相近，为 15.34 f/s，约为 SSD300 的 1/5，Faster R－CNN 的 2 倍，约为 DY3TNet 及 YOLO v5s 的 1/8 和 1/4；在嵌入式平台 Jetson AGX Xavier 上检测速度为 15.11 f/s，约为 YOLO v4 的 3 倍，SSD300 及 Faster R－CNN 的 1.75 倍和 12 倍，约为 DY3TNet 及 YOLO v5s 的 1/4 和 2/3。对比可以发现，6 种模型部署于高性能台式计算机平台上时普遍拥有不错的检测速度，而当移植到算力有限的嵌入式平台上时，模型检测速度均会产生不同程度的衰减。另外，YOLO v4－CA 的检测速度不如两种轻量级模型，但相较于 YOLO v4 及两种常用的目标检测模型 SSD300、Faster R－CNN，YOLO v4－CA 在嵌入式平台上的检测速度具有明显优势。

综合考虑模型的检测精度与性能，相比于改进前模型以及两种常用的目标检测模型，YOLO v4－CA 更易于在嵌入式平台上部署，同时能够在保证精度的前提下拥有较高的检测速度；相较于两种轻量级模型，YOLO v4－CA 在检测速度上不具有竞争力，但拥有更高的检测精度以及识别准

表 7 不同模型检测精度比较

Tab. 7 Comparison of detection accuracy with different models %

模型	AP	*P*	*R*
SSD300	91.32	91.73	82.07
Faster R－CNN	90.21	73.56	92.43
YOLO v4	93.23	92.78	82.40
DY3TNet	84.90	81.70	81.90
YOLO v5s	84.50	71.50	86.90
YOLO v4－CA	92.23	94.29	78.78

表 8 不同模型检测性能比较

Tab. 8 Comparison of detection performance with different models

模型	模型内存占用量/MB	FLOPs	台式计算机检测速度/(f·s^{-1})	嵌入式平台检测速度/(f·s^{-1})
SSD300	100	6.09×10^{10}	76.77	8.60
Faster R－CNN	108	1.05×10^{11}	7.28	1.24
YOLO v4	244	5.95×10^{10}	15.88	5.02
DY3TNet	26.8	1.24×10^{10}	126.58	52.63
YOLO v5s	28.0	1.70×10^{10}	55.87	22.78
YOLO v4－CA	54.1	6.92×10^{9}	15.34	15.11

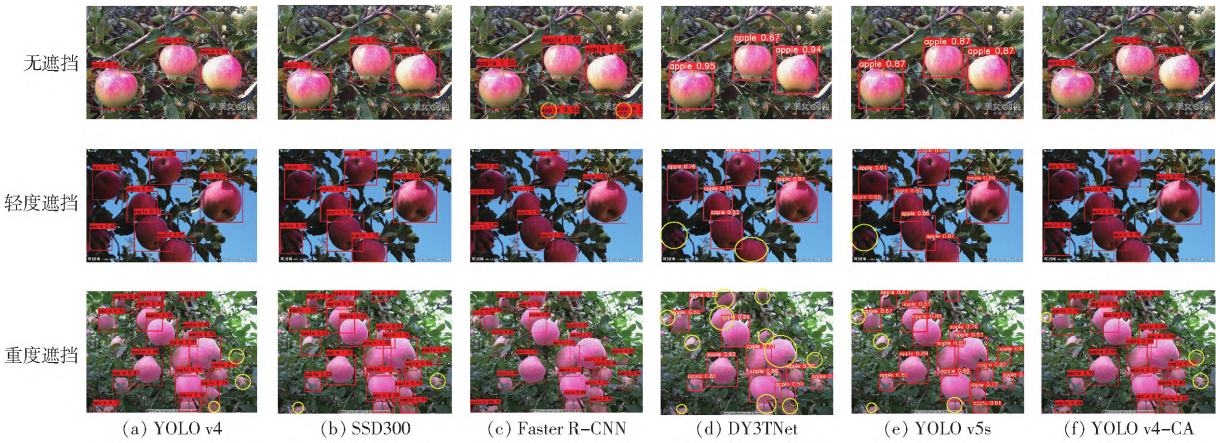

图 6 不同检测方法处理效果对比

Fig. 6 Comparison of detection results with different detection methods

确率。因此,综合以上分析可知,YOLO v4 - CA 实现了检测速度和检测精度的平衡,在保证苹果采摘过程中低误检率的同时提高了检测速度,更适用于苹果采摘任务。

4　结论

(1) 提出了一种改进 YOLO v4 轻量化实时苹果检测方法(YOLO v4 - CA),试验结果表明,YOLO v4 - CA 的平均检测精度达到了 92.23%,内存占用量为 54.1 MB,浮点数计算量为 6.92×10^9,在台式计算机及嵌入式平台 Jetson AGX Xavier 上的检测速度分别达到 15.34 f/s 和 15.11 f/s。模型能够在保证检测精度的同时,满足采摘机器人实时性需求。

(2) 将 CA 注意力机制引入特征融合网络,提升网络对密集目标的识别效果,改善枝叶遮挡、果实重叠对苹果检测带来的精度损失,在仅增加少量内存占用量的前提下 AP 提高 1.01 个百分点。

(3) 针对自然环境中的苹果检测,提出了一种将跨域迁移与域内迁移相结合的学习方法,有效提高了模型的泛化能力,相较于传统的跨域迁移学习 AP 提高 4.7 个百分点。

(4) 为验证本文模型的优越性,与两种常用的目标检测模型以及两种轻量级目标模型进行对比。本文模型的 AP 相较于 SSD300 与 Faster R - CNN 分别提高 0.91、2.02 个百分点,相较于 DY3TNet 与 YOLO v5s 分别提高 7.33、7.73 个百分点,在嵌入式平台上的检测速度分别约为 SSD300 与 Faster R - CNN 的 1.75 倍和 12 倍,约为 DY3TNet 及 YOLO v5s 的 1/4 和 2/3。

参 考 文 献

[1] 李会宾,史云. 果园采摘机器人研究综述[J]. 中国农业信息,2019,31(6):1 - 9.
LI Huibin, SHI Yun. Review on orchard harvesting robots[J]. China Agricultural Informatics, 2019, 31(6): 1 - 9. (in Chinese)

[2] LEHNERT C, SA I, MCCOOL C, et al. Sweet pepper pose detection and grasping for automated crop harvesting[C]// 2016 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2016.

[3] 王丹丹,宋怀波,何东健. 苹果采摘机器人视觉系统研究进展[J]. 农业工程学报,2017,33(10):59 - 69.
WANG Dandan, SONG Huaibo, HE Dongjian. Research advance on vision system of apple picking robot[J]. Transactions of the CSAE, 2017, 33(10): 59 - 69. (in Chinese)

[4] 景亮,王瑞,刘慧,等. 基于双目相机与改进 YOLO v3 算法的果园行人检测与定位[J]. 农业机械学报,2020,51(9):34 - 39,25.
JING Liang, WANG Rui, LIU Hui, et al. Orchard pedestrian detection and location based on binocular camera and improved YOLO v3 algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(9): 34 - 39, 25. (in Chinese)

[5] 何进荣,石延新,刘斌,等. 基于 DXNet 模型的富士苹果外部品质分级方法研究[J]. 农业机械学报,2021,52(7):379 - 385.
HE Jinrong, SHI Yanxin, LIU Bin, et al. External quality grading method of Fuji apple based on deep learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(7): 379 - 385. (in Chinese)

[6] 薛勇,王立扬,张瑜,等. 基于 GoogLeNet 深度迁移学习的苹果缺陷检测方法[J]. 农业机械学报,2020,51(7):30 - 35.
XUE Yong, WANG Liyang, ZHANG Yu, et al. Defect detection method of apples based on GoogLeNet deep transfer learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(7): 30 - 35. (in Chinese)

[7] REN S, HE K, GIRSHICK R, et al. Faster R - CNN: towards real-time object detection with region proposal networks[J]. arXiv preprint arXiv: 1506.01497, 2015.

[8] DAI J, LI Y, HE K, et al. R - FCN: object detection via region-based fully convolutional networks[C]// Advances in Neural Information Processing Systems, 2016: 379 - 387.

[9] GAO F, FU L, ZHANG X, et al. Multi-class fruit-on-plant detection for apple in SNAP system using Faster R - CNN[J]. Computers and Electronics in Agriculture, 2020, 176: 105634.

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

[11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]// European Conference on Computer Vision. Springer, Cham., 2016: 21 - 37.

[12] 张恩宇,成云玲,胡广锐,等. 基于 SSD 算法的自然条件下青苹果识别[J]. 中国科技论文,2020,15(3):274 - 281.
ZHANG Enyu, CHENG Yunling, HU Guangrui, et al. Recognition of green apple in natural scenes based on SSD algorithm[J]. China Science Paper, 2020, 15(3): 274 - 281. (in Chinese)

[13] 武星,齐泽宇,王龙军,等. 基于轻量化 YOLO v3 卷积神经网络的苹果检测方法[J]. 农业机械学报,2020,51(8):17 - 25.
WU Xing, QI Zeyu, WANG Longjun, et al. Apple detection method based on Light - YOLO v3 convolutional neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(8): 17 - 25. (in Chinese)

[14] FU L, FENG Y, WU J, et al. Fast and accurate detection of kiwifruit in orchard using improved YOLO v3 - tiny model[J]. Precision Agriculture, 2021, 22(3): 754 - 776.

[15] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLO v4: optimal speed and accuracy of object detection [J]. arXiv preprint arXiv: 2004.10934, 2020.
[16] HE K, ZHANG X, REN S, 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.
[17] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117 - 2125.
[18] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759 - 8768.
[19] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1314 - 1324.
[20] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: inverted residuals and linear bottlenecks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510 - 4520.
[21] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132 - 7141.
[22] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251 - 1258.
[23] 李文涛,张岩,莫锦秋,等. 基于改进 YOLO v3 - tiny 的田间行人与农机障碍物检测 [J]. 农业机械学报,2020,51(增刊1): 1 - 8,33.
LI Wentao, ZHANG Yan, MO Jinqiu, et al. Detection of pedestrian and agricultural vehicles in field based on improved YOLO v3 - tiny [J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(Supp. 1): 1 - 8, 33. (in Chinese)
[24] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [J]. arXiv preprint arXiv: 2103.02907, 2021.
[25] BUKHSH Z A, JANSEN N, SAEED A. Damage detection using in-domain and cross-domain transfer learning [J]. arXiv preprint arXiv: 2102.03858, 2021.
[26] MUREŞAN H, OLTEAN M. Fruit recognition from images using deep learning [J]. arXiv preprint arXiv: 1712.00580, 2017.

(上接第 274 页)

[14] 周建平,林韩,温步瀛. 改进量子遗传算法在输电网规划中的应用 [J]. 电力系统保护与控制,2012,40(19): 90 - 95.
ZHOU Jianping, LIN Han, WEN Buying. Application of improved quantum genetic algorithm in transmission network expansion planning [J]. Power System Protection and Control, 2012, 40(19): 90 - 95. (in Chinese)
[15] 周国华,陈炉,唐承丽,等. 长株潭城市群研究进展与展望 [J]. 经济地理,2018,38(6): 52 - 61.
ZHOU Guohua, CHEN Lu, TANG Chengli, et al. Research progress and prospect on Changsha - Zhuzhou - Xiangtan urban agglomeration [J]. Economic Geography, 2018, 38(6): 52 - 61. (in Chinese)
[16] 匡丽花,叶英聪,赵小敏,等. 基于改进 TOPSIS 方法的耕地系统安全评价及障碍因子诊断 [J]. 自然资源学报,2018,33(9): 1627 - 1641.
KUANG Lihua, YE Yingcong, ZHAO Xiaomin, et al. Evaluation and obstacle factor diagnosis of cultivated land system security in Yingtan City based on the improved TOPSIS method [J]. Journal of Natural Resources, 2018, 33(9): 1627 - 1641. (in Chinese)
[17] 石淑芹,陈佑启,姚艳敏,等. 东北地区耕地自然质量和利用质量评价 [J]. 资源科学,2008(3): 378 - 384.
SHI Shuqin, CHEN Youqi, YAO Yanmin, et al. Assessing natural quality and utilization quality of cultivated land in Northeast China [J]. Resource Science, 2008(3): 378 - 384. (in Chinese)
[18] 吴利,柳德江. 基于 GA - BP 神经网络的玉溪市耕地生态安全评价 [J]. 云南农业大学学报(自然科学),2019,34(5): 874 - 883.
WU Li, LIU Dejiang. Ecological security evaluation of cultivated land in Yuxi City based on GA - BP neural network [J]. Journal of Yunnan Agricultural University (Natural Science), 2019, 34(5): 874 - 883. (in Chinese)
[19] 赵宏波,马延吉. 东北粮食主产区耕地生态安全的时空格局及障碍因子——以吉林省为例 [J]. 应用生态学报,2014,25(2): 515 - 524.
ZHAO Hongbo, MA Yanji. Spatial-temporal pattern and obstacle factors of cultivated land ecological security in major grain producing areas of Northeast China: a case study in Jilin Province [J]. Chinese Journal of Applied Ecology, 2014, 25(2): 515 - 524. (in Chinese)
[20] 郁磊,史峰,王辉,等. MATLAB 智能算法 30 个案例分析 [M]. 北京: 北京航空航天大学出版社, 2015.
[21] 付强,赵小勇. 投影寻踪模型原理及其应用 [M]. 北京: 科学出版社,2006.
[22] 楼文高. 基于群智能最优化算法的投影寻踪理论——新进展、应用及软件 [M]. 上海: 复旦大学出版社,2021.