doi:10.6041/j.issn.1000-1298.2023.12.025

基于改进 DeepLabv3 + 的番茄图像多类别分割方法

顾文娟　魏　金　阴艳超　刘孝保　丁　灿
(昆明理工大学机电工程学院, 昆明 650500)

摘要: 番茄图像中多类别目标的准确识别是实现自动化采摘的技术前提,针对现有网络分割精度低、模型参数多的问题,提出一种基于改进 DeepLabv3 + 的番茄图像多类别分割方法。该方法使用幻象网络(GhostNet)和坐标注意力模块(Coordinate attention,CA)构建 CA - GhostNet 作为 DeepLabv3 + 的主干特征提取网络,减少网络的参数量并提高模型的分割精度,并设计了一种多分支解码结构,用于提高模型对小目标类别的分割能力。在此基础上,基于单、双目小样本数据集使用合成数据集的权值参数进行迁移训练,对果实、主干、侧枝、吊线等 8 个语义类别进行分割。结果表明,改进的 DeepLabv3 + 模型在单目数据集上的平均交并比(MIoU)和平均像素准确率(MPA)分别为 68.64%、78.59%,在双目数据集上的 MIoU 和 MPA 分别达到 73.00%、80.59%。此外,所提模型内存占用量仅为 18.5 MB,单幅图像推理时间为 55 ms,与基线模型相比,在单、双目数据集上的 MIoU 分别提升 6.40、6.98 个百分点,与 HRNet、UNet、PSPNet 相比,内存占用量压缩 82%、79%、88%。该研究可为番茄采摘机器人的智能采摘和安全作业提供参考。

关键词: 番茄; 语义分割; DeepLabv3 + ; GhostNet; CA 注意力模块
中图分类号: TP391.4　　**文献标识码**: A　　**文章编号**: 1000-1298(2023)12-0261-11　　**OSID**:

Multi-category Segmentation Method of Tomato Image Based on Improved DeepLabv3 +

GU Wenjuan　WEI Jin　YIN Yanchao　LIU Xiaobao　DING Can
(*Faculty of Mechanical and Electrical Engineering*, *Kunming University of Science and Technology*, *Kunming* 650500, *China*)

Abstract: Accurate identification of multi-category targets in tomato images is the technical premise for automatic picking. Aiming at the problems of low segmentation accuracy and the large number of model parameters in existing networks, a multi-category segmentation method based on improved DeepLabv3 + was proposed for tomato images. The method used GhostNet and coordinate attention (CA) to construct CA - GhostNet as the backbone feature extraction network of DeepLabv3 +, reducing the number of parameters in the network. And a multi-branch decoding structure was designed to improve the segmentation accuracy of the model for small target categories. Then the weight parameters of the synthesized dataset were used for migration training based on the single and binocular small sample dataset. Eight semantic categories such as fruit, trunk, branch and thin line were segmented. The results showed that mean intersection over union (MIoU) and mean pixel accuracy (MPA) of improved DeepLabv3 + model were 68.64% and 78.59% on the monocular dataset, respectively. The MIoU and MPA were 73.00% and 80.59% on the binocular dataset. In addition, the memory occupation of the proposed model was only 18.5 MB, and the inference time of a single image was 55 ms. Compared with the baseline model, the MIoU on the monocular and binocular datasets was increased by 6.40 percentage points and 6.98 percentage points, respectively. Compared with HRNet, UNet and PSPNet, the memory occupation was reduced by 82%, 79% and 88%, respectively. The research result can provide reference for intelligent picking and safe operation of tomato picking robot.

Key words: tomato; semantic segmentation; DeepLabv3 + ; GhostNet; CA attention module

收稿日期: 2023 - 04 - 23　修回日期: 2023 - 07 - 30
基金项目: 国家自然科学基金项目(52065033)和云南省科技厅重大专项(2022AG050002 - 4)
作者简介: 顾文娟(1985—),女,讲师,博士,主要从事智能制造、机器视觉和图像处理等研究,E-mail: 20110065@kust.edu.cn
通信作者: 阴艳超(1977—),女,教授,博士,主要从事智能制造、知识工程和工业大数据等研究,E-mail: 20090143@kust.edu.cn

0 引言

番茄采摘是其生产链上劳动力投入最大的环节之一,现阶段主要靠人工完成。随着科技的进步,番茄采摘机器人的发展应用有望解决其产业链中劳动力短缺的难题[1]。然而,番茄采摘环境为非结构化环境,光照复杂,茎干交错,吊线纵横,果实与枝叶遮挡严重,无伤采摘以及对不同目标采取的避障策略是制约番茄采摘机器人发展的关键因素[2]。因此,番茄图像中多类别目标的准确识别是实现自动采摘的技术前提,研究具有高识别度的视觉检测系统对实现番茄的自动化采摘具有重要意义。

番茄图像目标识别方法通常可以分为传统方法和深度学习方法两类。传统方法主要采用基于颜色、纹理和形状等特征的图像分割算法,如基于阈值的分割算法、分水岭算法、边缘检测算法等。李寒等[3]提出了一种快速归一化互相关函数(FNCC)的算法,结合霍夫圆检测的方法解决了绿色番茄难以检测的问题。丁永军等[4]融合小波变换和分水岭算法对番茄冠状叶片进行分割,平均误分率为21%。尹建军等[5]使用颜色空间转换的阈值分割方法对番茄果实进行自动分割,该方法分割速度较快,但误分割区域较多。梁喜凤等[6]通过形态学方法和阈值法对番茄串进行分割,该方法对图像拍摄角度敏感,其精度亦有待提升。上述算法通常需要人工设置一些参数和规则,具有一定的局限性和不稳定性。

近年来,随着计算机视觉技术的发展,不少学者将深度学习方法应用到农业领域的对象识别中。此类方法可以自动学习图像特征,无需手动设置参数和规则,具有更好的鲁棒性和泛化能力。龙洁花等[7]使用跨阶段局部网络改进 Mask R-CNN 的主干网络,提升了番茄果实成熟度的平均分割精度,但该方法存在模型参数多、分割速度慢的问题。胡春华等[8]融合杨树苗 RGB 图像和深度数据构建融合数据集,使用 SegNet 对叶片进行分割,茎和叶的交并比分别达到 75.9%、67.9%。冯青春等[9]使用迁移训练的方式,基于 Mask R-CNN 提出了一种番茄植株整枝操作中主茎和侧枝目标分割的方法,由于缺乏茎秆形态立体信息,其定位精度有待提升。任守纲等[10]基于反卷积提出了用于番茄病害识别的 DGVGGNet 模型,但其标注数据有限,模型泛化能力不足。刘文波等[11]将 ResNet-101 与 FPN 结构相结合,改进了 SOLOV2 模型用于番茄病虫害检测,其平均精度为 42.3%,该方法使用的 ResNet-101 主干网络参数较多,对硬件的要求较高,且精度有待提升。熊俊涛等[12]使用 ResNet-34 为主干网络的 DeepLabv3 模型对荔枝花、叶进行分割,其平均交并比为 73.4%。以上算法相较于传统方法,其鲁棒性和适用性虽取得了较大的提升,但存在网络参数多、模型精度低、计算开销大的问题。

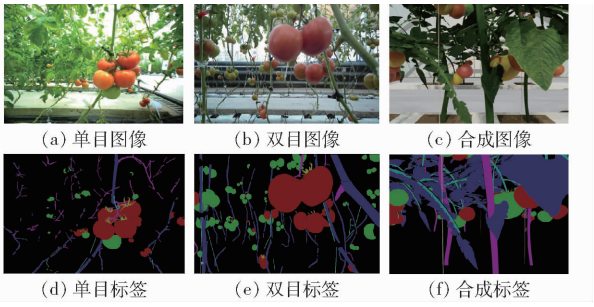
针对上述问题,本文提出一种基于改进 DeepLabv3+的番茄图像多类别分割方法,该方法使用 GhostNet 网络和 CA 注意力模块构建 CA-GhostNet 作为 DeepLabv3+的主干特征提取网络,减少网络的参数量并提高模型的分割精度,设计一种多分支解码结构,用于提高模型对小目标类别的分割能力。在此基础上,基于单、双目小样本数据集使用合成数据集的权值参数进行迁移训练,对果实、主干、侧枝、吊线等 8 个语义类别进行分割。以解决复杂环境下番茄图像多类别分割模型复杂度高和准确率低的问题,为番茄采摘机器人的智能采摘和安全作业提供参考。

1 番茄数据集与多类别分割模型

1.1 数据集

本研究采用国家农业数据中心提供的番茄公共数据集[13],该数据集包括单目数据集、双目数据集和合成数据集 3 个番茄数据集,部分样例如图 1 所示。单、双目数据集为实地采集,图像采集地点为江苏省农业科学院的番茄生产智能温室,该温室育有苏粉 11 号等品种番茄共 18 行,番茄植株的株距为 25 cm,行距为 90 cm。由于图像采集场景存在近距离、高密度特点,因此对相机分辨率和焦距选择有一定要求。单目数据集使用德国 uEye SE 工业相机采样,镜头焦距为 4.16 mm,图像分辨率为 1 600 像素×1 200 像素,双目数据集使用国产 Lena 双目相机采样,定焦镜头焦距为 3.6 mm,左右图分辨率均为 1 280 像素×720 像素。图像采集时间为 2020 年 12 月至 2021 年 2 月,整体天气情况为晴天,温室中的番茄属于成熟期。采集过程中,镜头距离番茄植株 20~40 cm,俯仰角在 -45°到 45°之间变换,相机沿着植株行方向每隔 20 cm 增量进行成像,每行植株都有顺光和逆光两种拍摄情况。该公共数据集中有单目图像 750 幅,双目图像 400 幅,但是多幅连续图像存在高度相似的特点,易导致模型过拟合,故选取具有代表性的图像各 100 幅,使用 Labelme 工具进行标注,语义标签包括成熟果实、绿色果实、花托、主干、侧枝、吊线、剪枝切口和背景 8 个类别。合成数据集使用软件 Plantfactory 和 Blender 对温室番茄及其生长环境进行模拟,生成番茄的合成图像,语义标签多了叶片,共 9 个类别,图像分辨率为 1 600 像素×

1 200 像素，有 3 250 幅。此外，合成数据集存在一定的人工设定和约束，可能无法反映不同品种的番茄在不同生长阶段和不同环境下的特征变化，即无法覆盖真实环境中的所有场景和情况。



(a) 单目图像　(b) 双目图像　(c) 合成图像

(d) 单目标签　(e) 双目标签　(f) 合成标签

图 1　番茄数据集样例

Fig. 1　Samples of tomato datasets

在实际任务中获取大规模的标注数据进行训练是非常困难的，合成数据集的样本特征与真实数据集近似，且具有数据数量可观、制作成本低和标注便捷等优势。因此，利用合成数据集进行预训练，模型可以学习到番茄的形状、颜色、纹理等通用特征，再将这些通用特征迁移到真实场景下的番茄图像分割任务中，可以减少单、双目数据集的标注需求。其次，选取多源番茄图像构成的数据集，可增加数据的多样性，以应对不同相机参数、拍摄角度和光照条件变化下的图像识别问题。但是不同数据集之间的数据分布和标注方式略有差异，强行将所有数据集组合到一起会导致模型的性能下降。因此，按照图像来源将番茄数据集划分为单目数据集、双目数据集和合成数据集，有助于验证模型的性能和泛化能力。

1.2　样本增强

深度学习是一种基于数据驱动的方法，数据集的质量和规模对于模型的准确性和泛化能力具有重要影响。由于单、双目数据集的样本数量较少，而且花托、侧枝、吊线、剪枝切口等语义类别属于难以检测的小目标类别，为防止网络因训练样本不足导致模型精度低、泛化能力差等问题，本研究结合 Imgaug 数据增强库和 OpenCV 库完成样本的数据增强。单、双目数据集使用标注过的 100 幅图像进行数据增强，增强数量各 1 000 幅，按照比例 8∶1∶1 随机划分训练集(800 幅)、验证集(100 幅)和测试集(100 幅)，用于后续模型的训练和测试。采用的数据增强方法包括图像旋转、左右翻转、图像缩放、随机裁剪和投射变换，按照概率 0.8 进行组合增强，数据增强样例如图 2 所示。通过数据增强，可以扩充小目标类别的样本，使训练数据更加丰富，在一定程度上缓解样本不足和类别不均衡带来的问题。

1.3　DeepLabv3 + 模型

DeepLabv3 +[14] 是一种在分割精度和计算效率之间取得平衡的语义分割模型，已经在许多领域中得到了广泛的应用。它通过引入一系列新的特征提取模块和空洞卷积模块，进一步提高模型的分割精度和计算效率。DeepLabv3 + 采用 Xception[15] 模型作为其主干网络，以增加模型的深度和感受野。同时，为了减小分辨率下降对分割精度的影响，DeepLabv3 + 使用了一种空洞空间金字塔池化(Atrous spatial pyramid pooling，ASPP)的多尺度特征提取模块。该模块可以在不同尺度下提取特征，并通过空洞卷积将这些特征进行整合，从而使模型能够更好地理解不同尺寸和形状的物体。除了 ASPP 模块，DeepLabv3 + 还使用了一种解码器(Decoder)的结构，用于将低分辨率的特征图上采样到原始分辨率，并与高分辨率的特征图进行融合，从而提高分割精度。



图 2　数据增强样例

Fig. 2　Sample of data enhancement

1.4　番茄图像多类别分割模型

考虑到番茄采摘环境复杂以及小目标类别难以检测问题，本文设计了一种基于改进 DeepLabv3 + 的番茄图像多类别分割模型，整体结构如图 3 所示。本文从以下几方面进行改进：

(1)在编码器中，使用 GhostNet 和 CA 注意力模块构建 CA - GhostNet，代替 Xception 作为 DeepLabv3 + 的主干特征提取网络，减少网络的参数量并提高模型的分割精度。

(2)在解码器中，设计了一种多分支解码结构，提高对小目标类别的分割能力。此外，在浅层特征分支中嵌入 CA 注意力模块，可以减少高分辨率特征图潜在的噪声干扰，并提高模型精度。

(3)在 CA 模块中，使用 H - Sigmoid 函数取代 Sigmoid 函数，减少指数运算，降低运算成本。在 G - bneck 模块中，使用 Swish 函数和 CA 模块重构 G - bneck 模块，提高模型的运算速度和精度。

1.4.1　Ghost 模块

较高的模型复杂度会导致番茄图像分割模型过拟合。GhostNet[16] 是一种轻量级深度神经网络架构，将之作为 DeepLabv3 + 的主干特征提取网络，可以降低模型复杂度。普通的卷积操作会产生很多相似的特征图，这些冗余的信息并不会有效提高网络

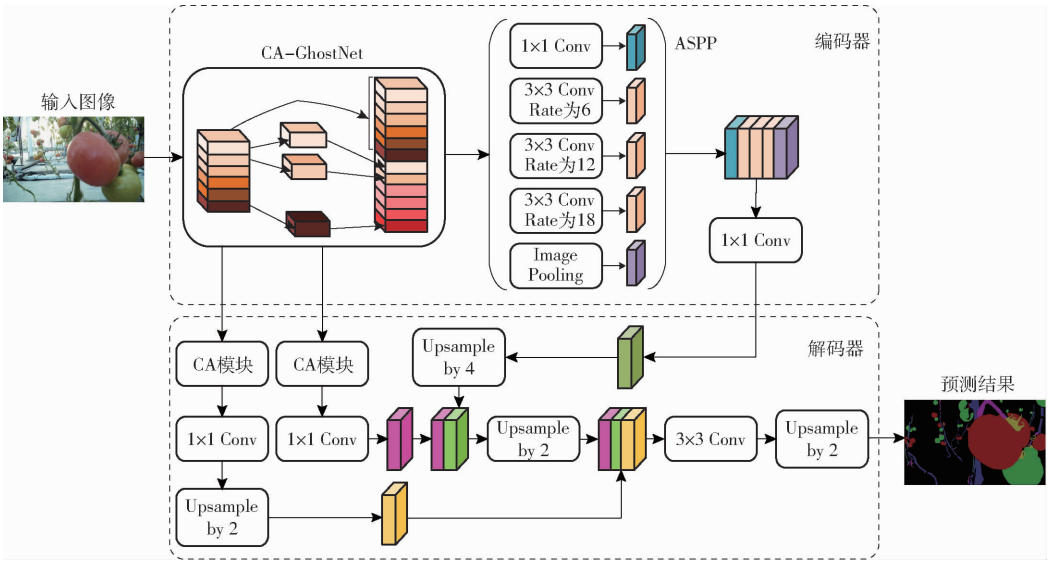


图 3 改进的 DeepLabv3 + 模型结构图

Fig. 3 Improved DeepLabv3 + model structure

的特征提取能力,反而会干扰模型的学习效果。Ghost 模块先通过普通卷积生成部分真实特征层,然后对真实特征层进行线性操作生成幻象特征层,最后将真实特征层和幻象特征层拼接形成新的输出特征层,如图 4b 所示。Ghost 模块利用特殊的分组卷积技术和逐通道卷积技术进行线性操作,使得一个卷积核只需要计算一次就可被多个输出通道共享,因此 Ghost 模块的参数数量通常比普通卷积小得多。

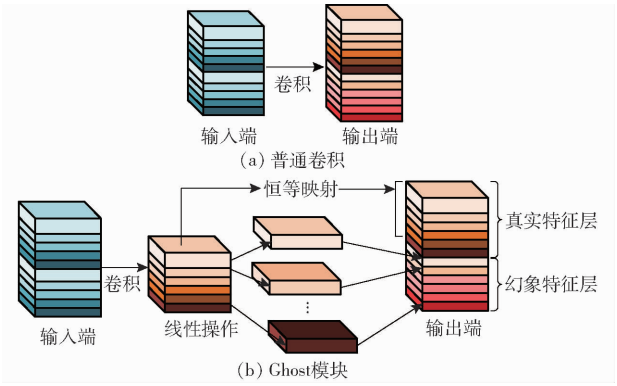


图 4 普通卷积与 Ghost 模块

Fig. 4 General convolution and Ghost module

假设输入特征图的尺寸为 $h \times w \times c$,输出特征图尺寸为 $h' \times w' \times n$,采用的卷积核尺寸为 $k \times k$,线性操作核尺寸为 $d \times d$,Ghost 模块中的分组数为 s,则普通卷积与 Ghost 模块所需的参数量 P_{conv}、P_{ghost} 分别为

$$P_{conv} = h'w'nk^2c \tag{1}$$

$$P_{ghost} = h'w'\frac{n}{s}k^2c + h'w'(s-1)\frac{n}{s}d^2 \tag{2}$$

因此,Ghost 模块与传统卷积的理论加速比 R_p 为

$$R_p = \frac{P_{conv}}{P_{ghost}} = \frac{nc}{\frac{n}{s}c + (s-1)\frac{n}{s}} = \frac{sc}{c+s-1} \approx s \tag{3}$$

其中,$d \times d$ 与 $k \times k$ 相似且 $s \ll c$,则理论参数压缩比 R_c 为

$$R_c = \frac{nck^2}{\frac{n}{s}ck^2 + (s-1)\frac{n}{s}d^2} = \frac{sc}{c+s-1} \approx s \tag{4}$$

Ghost 模块通过对输入通道进行分组,将卷积操作中的参数量减少到原来的 $1/s$,从而大幅降低了模型的计算和存储开销。理论上,分组数越多,Ghost 模块的参数压缩比就越高,但是过高的分组数会对模型的精度造成一定影响。这是因为分组数增加后,每个卷积核只能看到输入张量中的部分信息,导致模型难以学习到全局特征。因此,综合考虑模型的精度、计算和存储开销等因素,在保证模型精度的前提下实现参数压缩,本文分组数为 2。

1.4.2 CA 注意力模块

为了从复杂环境中将番茄图像的多类别语义目标分割出来,模型的特征提取和表示能力需要进一步提升。因此,引入坐标注意力模块[17](CA)提高模型对番茄图像的特征提取能力。GhostNet 原网络中采用的注意力机制为挤压和激励注意力模块(Squeeze and excitation,SE)[18],如图 5a 所示。SE 模块只关注通道之间的权重调整,对局部特征的提取有限,不能很好地捕捉空间细节。为此,CA 模块能够在通道和空间两个方面同时调整注意力,弥补 SE 模块的不足。CA 模块的计算量虽然比 SE 模块略大,但相比于其他注意力模块,仍然相对较小,速度较快。原始 CA 模块中使用 Sigmoid 函数,虽有助于提升网络性能,却降低了检测推理时间。相较于

Sigmoid 函数，H－Sigmoid 函数计算更加简单，不需要进行指数运算，可以提高计算效率[19]。因此，使用 H－Sigmoid 函数代替 CA 模块中的 Sigmoid 函数形成新的注意力模块，提高模型的运行速度，改进后的 CA 模块结构如图 5b 所示。

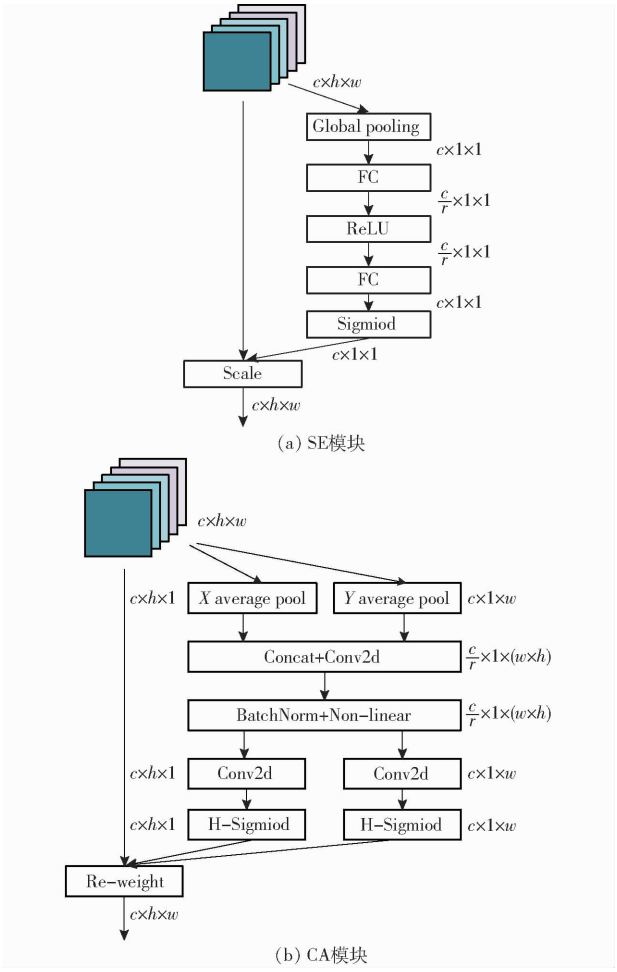


图 5　SE 模块和 CA 模块

Fig. 5　SE block and CA block

假设输入为 x，CA 模块分别沿着 2 个空间方向进行全局平均池化操作，高度为 h 的第 c 通道的输出可以表示为

$$z_c^h(h) = \frac{1}{w}\sum_{0\leqslant i\leqslant w} x_c(h,i) \tag{5}$$

宽度为 w 的第 c 通道的输出表示为

$$z_c^w(w) = \frac{1}{h}\sum_{0\leqslant j\leqslant h} x_c(j,w) \tag{6}$$

紧接着将不同方向感知的特征图融合，使用点卷积将其维度降低为原来的 c/r，然后将经过批量归一化和非线性操作处理得到形如 $1\times(w\times h)\times c/r$ 的特征图 f，即

$$f = \delta(F_1([z^h, z^w])) \tag{7}$$

之后将特征图沿空间维度分解成 2 个单独的张量，即

$$g^h = \sigma(F_h(f^h)) \tag{8}$$

$$g^w = \sigma(F_w(f^w)) \tag{9}$$

最后通过卷积变换以及 H－Sigmoid 激活函数生成 2 个权重，将 2 个权重与原特征图相乘实现加权操作，即

$$y_c(i,j) = x_c(i,j)g_c^h(i)g_c^w(j) \tag{10}$$

1.4.3　CA－GhostNet 网络

GhostNet 使用类似 MobileNetV2 中的瓶颈结构来防止梯度消失并补充原始特征信息，解决随着网络加深出现的网络退化问题。Ghost 瓶颈模块采用先升维再降维的结构形式，能减少特征图在不同维度之间转换时带来的信息损失，同时在特定的网络层中嵌入 CA 模块加强特征提取。为了实现加深网络的作用，Ghost 瓶颈结构中的第 1 个 Ghost 模块用于升维，第 2 个 Ghost 模块用于降维，使其与经过残差连接的通道数保持一致，如图 6a 所示。当步长(stride)为 2 时，网络进行下采样，可以更好地捕捉番茄图像中的全局信息，在第 1 个 Ghost 模块升维后添加深度卷积和 CA 注意力模块，在残差结构中加入深度卷积和点卷积，如图 6b 所示。将原始网络中的 ReLU 激活函数替换为性能更加优异的 Swish 函数，有助于提高模型的泛化能力[20]。G－bneck 可以用更少的参数和计算量建立高效的算法网络，通过堆叠 G－bneck 构建 CA－GhostNet 主干特征提取网络，其网络结构参数如表 1 所示。

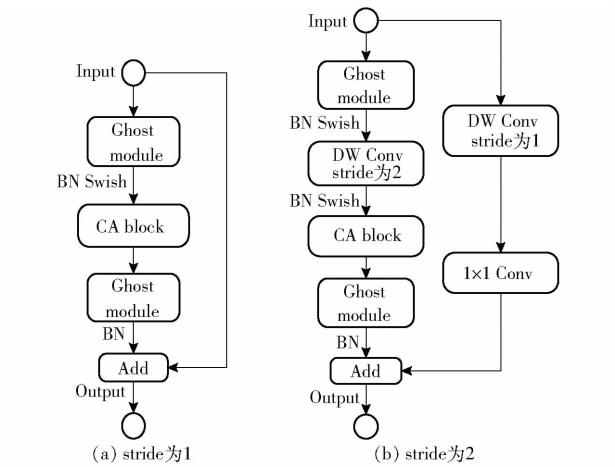


图 6　G－bneck 模块

Fig. 6　G－bneck module

1.4.4　多分支解码结构

DeepLabv3 + 中解码器主要用于融合浅层特征和深层特征，提高分割精度。浅层特征是指网络前几层提取的特征，这些特征注重细节信息的捕捉，包含了更多的颜色、边缘纹理信息，但是由于经过的卷积少，其语义性低，噪声多。深层特征通过网络的层级结构逐渐提取，其感受野较大，拥有更丰富的语义信息。编码器的下采样操作可以减小特征图的尺寸并提取更高级别的语义信息，以便更好地理解图像

表 1 CA - GhostNet 网络结构参数

Tab. 1 CA - GhostNet network parameters

网络层数	操作方法	扩张系数	输出通道	CA模块	步长
1	3×3 Conv		16	否	2
2	G - bneck	16	16	否	1
3	G - bneck	48	24	否	2
4	G - bneck	72	24	否	1
5	G - bneck	72	40	是	2
6	G - bneck	120	40	是	1
7	G - bneck	240	80	否	2
8	G - bneck	200	80	否	1
9	G - bneck	184	80	否	1
10	G - bneck	184	80	否	1
11	G - bneck	480	112	是	1
12	G - bneck	672	112	是	1
13	G - bneck	672	160	是	1
14	G - bneck	960	160	否	1
15	G - bneck	960	160	是	1
16	G - bneck	960	160	否	1
17	G - bneck	960	160	是	1
18	1×1 Conv		160	否	1

的整体结构和语义内容。然而,番茄图像中的花托、细线、切口、侧枝等属于小目标类别,编码器的下采样操作会造成小目标类别的细节信息丢失。在原模型中,使用双线性插值将深层特征直接上采样 16 倍变为与输入图像尺寸相同的图像,这种方法无法还原小目标类别的具体细节。因此,本文借鉴了 FPN[21] 的思想,在解码器中再插入一个浅层特征分支,形成多分支解码结构,用于提高模型对小目标类别的分割能力,如图 7 所示。多分支解码结构通过在编码器中使用深层特征来提取丰富的语义信息,并在解码器中进行适当的上采样和融合操作,可以帮助模型更好地理解番茄图像的细节和语义内容,从而提高分割任务的准确性和鲁棒性。此外,在 2 个浅层特征分支之后嵌入 CA 注意力模块,可以减少噪声干扰并提高模型的精度。

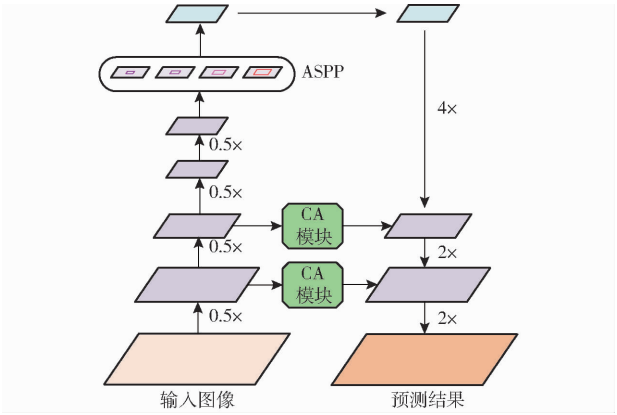


图 7 多分支解码结构

Fig. 7 Multi-branch decoding structure

1.4.5 损失函数

交叉熵损失函数[22](Cross Entropy Loss, CE Loss)是一种常用的图像分割任务损失函数,通常用于衡量模型预测结果与真实标签之间的差异,并作为优化目标进行模型训练,其计算公式为

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\mathrm{lb}(p_{ij}) \tag{11}$$

式中 N——像素点个数 M——类别数

y_{ij}——符号函数,若样本 i 的真实类别等于 j 则取 1,否则取 0

p_{ij}——像素点 i 属于 j 的预测概率

CE Loss 主要关注像素分类结果,计算每个像素点预测值与真实值之间的差异。但是在类别不均衡的情况下,模型可能会倾向于预测更多的像素点为负样本,进而导致模型性能下降。在单、双目番茄数据集中,由于花托、侧枝、吊线和剪枝切口的像素数量远小于背景类别的像素数量,类别不均衡会导致模型偏向于预测背景类别,引用戴斯损失函数[23](Dice Loss)作为辅助损失函数解决这一问题。Dice Loss 在计算损失时,通过计算模型预测结果与真实标签之间的相似度,将其转换为一个可优化的目标函数。相较于 CE Loss, Dice Loss 对于类别不均衡的问题具有更好的鲁棒性,可以有效地缓解类别不均衡问题带来的影响,其计算公式为

$$L_{\mathrm{Dice}} = 1 - \frac{2\sum_{i=1}^{N} p_i t_i + \varepsilon}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} t_i^2 + \varepsilon} \tag{12}$$

式中 p_i——模型对 i 的预测值

t_i——对应的真实标签

ε——很小的常数,防止分母为 0

然而,Dice Loss 在反向传播的过程中可能导致梯度剧烈变化,造成模型训练不稳定。基于以上考虑,本文使用 CE Loss 与 Dice Loss 构成复合损失函数 L_{loss},其计算公式为

$$L_{\mathrm{loss}} = L_{\mathrm{CE}} + L_{\mathrm{Dice}} \tag{13}$$

该复合损失函数同时考虑了 CE Loss 和 Dice Loss 的优点,可以降低类别不均衡问题对模型造成的影响,提高模型的训练效率。此外,复合损失函数还可以避免梯度剧烈变化,从而更加稳定地训练模型。

2 试验结果与分析

2.1 试验环境

本试验主机中央处理器为 i5 - 12490f,运行内存为 32 GB,显卡为 GeForce RTX3060 12 GB,操作系

统为 Ubuntu 22.04LTS。使用 Anaconda 3 配置深度学习虚拟环境，Python 版本为 3.8，Cuda 版本为 11.3，torch 版本为 1.12.0。通过多次试验得出，使用 SGD 优化器，初始学习率设置为 0.007，学习率下降方式为余弦退火，权值下降因子为 0.000 1，动量因子为 0.9，可以较快达到局部收敛。对于单目数据集和合成数据集，输入网络的图像分辨率设置为 800 像素 ×600 像素，批量大小设置为 4。对于双目数据集，输入网络的图像分辨率为 1 280 像素 × 720 像素，批量大小设置为 2。

2.2　评价指标

本研究中用于测量分割模型准确度的评估指标为准确率（Accuracy）、平均查准率（Mean precision，MP）、交并比（Intersection over union，IoU）、平均交并比（Mean intersection over union，MIoU）、平均像素准确率（Mean pixel accuracy，MPA）。

2.3　试验

2.3.1　迁移训练试验

本文的训练策略为先在合成数据集上进行预训练，然后在单、双目数据集上加载合成数据集的权值参数进行迁移训练，最终在单、双目数据集的测试集上验证结果。通过在合成数据集上进行预训练，可以使模型保留番茄图像大部分的特征，在训练单、双目小样本数据集时只需要微调参数就可以得到更好的初始表现和收敛速度，从而加快模型的训练过程。本文基于 DeepLabv3 + 模型设计试验，模型迭代次数为 50，结果如表 2 所示。从表 2 可以看出，模型使用迁移训练后，其 MIoU 和 MPA 均有不同程度的提升，在单目数据集上的 MIoU 和 MPA 分别提高 3.68、0.74 个百分点，在双目数据集上的 MIoU 和 MPA 分别提高 4.82、2.94 个百分点。因此，使用合成数据集进行迁移训练可以解决标记数据不足的问题，并提高模型的性能和泛化能力。

表 2　迁移训练试验结果

Tab.2　Experiment results of transfer training　%

数据集	迁移训练	MIoU	MPA
单目数据集	否	56.50	69.82
	是	60.18	70.56
双目数据集	否	56.45	66.67
	是	61.27	69.61

2.3.2　轻量化主干网络对比试验

将 CA − GhostNet、MobileNetV2[24] 和 MobileNetV3[25] 替换 DeepLabv3 + 模型的原主干网络 Xception，验证 CA − GhostNet 网络对番茄图像的分割性能，在两类测试集上的分割结果如表 3 所示。由表 3 可知，CA − GhostNet 在单目数据集上的 MIoU、MPA、MP 和 Accuracy 分别为 62.24%、69.75%、82.48%、95.23%，除 MPA 比 MobileNetV3 低 1.26 个百分点外，MIoU、MP 和 Accuracy 指标均高于其他轻量化主干网络。在双目数据集上，CA − GhostNet 的 MIoU、MPA、MP 和 Accuracy 分别为 66.02%、76.55%、83.05%、94.68%，所有精度指标均高于其他轻量化网络。除此之外，与 Xception、MobileNetV2 和 MobileNetV3 相比，CA − GhostNet 的参数量最少，仅为 4.58×10^6。几种轻量化模型都是借助深度可分离卷积替换普通卷积达到减少计算参数的目的，但是 Xception 并没有在残差结构上做调整，导致计算过程较为零散，参数较多，训练速度慢。MobileNetV2 使用了线性瓶颈和反向残差结构，有效降低了计算量，同时网络层次更深，模型更小，但精度却不及其他模型。MobileNetV3 使用神经网络架构搜索技术进行网络结构构建，使得模型结构更加合理化，达到了更佳性能，但其精度有待提升。CA − GhostNet 在 MobileNetV3 的基础上，使用 Ghost 模块替换深度可分离卷积，去除了许多重复冗余的特征，有效降低了参数量，引用 CA 注意力模块使得模型的特征提取能力进一步提高。因此，使用 CA − GhostNet作为 DeepLabv3 + 的主干特征提取网络，保证了模型的轻量化，且拥有较高的特征提取能力。

表 3　不同轻量化主干网络的对比试验结果

Tab.3　Comparative experiment results of different lightweight backbone networks

数据集	主干网络	MIoU/%	MPA/%	MP/%	Accuracy/%	参数量
单目数据集	Xception	57.83	64.61	79.31	95.15	5.69×10^7
	MobileNetV2	56.45	66.67	71.72	92.84	6.02×10^6
	MobileNetV3	60.76	71.01	75.83	93.82	5.04×10^6
	CA − GhostNet	62.24	69.75	82.48	95.23	4.58×10^6
双目数据集	Xception	59.41	67.88	76.45	92.34	
	MobileNetV2	51.03	59.65	66.39	92.81	
	MobileNetV3	59.46	66.98	79.25	93.63	
	CA − GhostNet	66.02	76.55	83.05	94.68	

不同轻量化主干网络的分割结果如图 8 所示，由于经过数据增强后的图像只突出局部细节的特点，并不能反映各模型的整体分割效果，因此使用标注过的 100 幅原始图像进行可视化分析。在图像 1 中，Xception、MobileNetV2、MobileNetV3 对右上角吊线的检测不敏感，MobileNetV2 对远景的果实存在漏分割现象；在图像 2 中，MobileNetV2 和 MobileNetV3 对左侧堆叠的绿色果实存在明显的误分割，CA - GhostNet 在吊线上的分割效果优于其他模型；在图像 3 中，Xception 对吊线的分割不完整，MobileNetV2 将部分主干识别成侧枝，MobileNetV3 在花托部分存在误分割且没有识别到远景的果实；图像 4 中，Xception 和 MobileNetV3 没有检测到主干上的剪枝切口，MobileNetV3 将右方的背景识别成侧枝。以 CA - GhostNet 为主干的 DeepLabv3 + 模型总体分割效果优于 Xception、MobileNetV2 和 MobileNetV3，但其模型性能仍有提升空间，在图像 2 中 CA - GhostNet 将下方的绿色果实分割成红色果实，在图像 3 中没有识别到远景的绿色果实，在图像 4 中对细线的分割不完整。

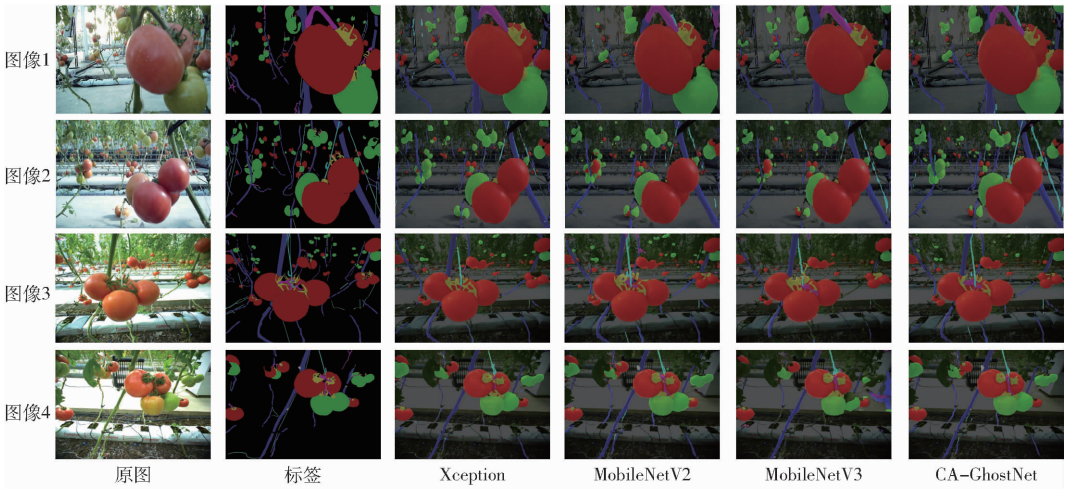


图 8 不同轻量化主干网络的分割结果

Fig. 8 Segmentation results of different lightweight backbone networks

2.3.3 消融试验

为了验证在解码器中引入 CA 注意力模块和多分支解码结构的贡献，使用 CA - GhostNet 为主干特征提取网络的模型为基线模型，在此模型的基础上进行消融试验。试验结果如表 4 和表 5 所示，引入 CA 模块，单、双目数据集的 MIoU 在基线模型的基础上分别提升 2.70、1.24 个百分点；引入多分支解码结构，单、双目数据集的 MIoU 在基线模型的基础上分别提升 3.59、5.62 个百分点。单独引用 CA 模块和多分支解码结构，每个类别的交并比均有不同程度上的提升，且引入多分支解码结构的效果优于 CA 模块。

同时引入 CA 模块和多分支解码结构，加强了模型对小目标的分割能力。在单目数据集上，成熟

表 4 单目数据集消融试验结果

Tab. 4 Results of monocular dataset ablation experiment %

试验方法	IoU								MIoU
	背景	成熟果实	绿色果实	花托	主干	侧枝	吊线	剪枝切口	
基线模型	94	90	79	48	67	34	40	51	62.24
+ CA 模块	94	91	81	50	68	39	41	55	64.94
+ 多分支	95	91	80	51	70	38	47	56	65.83
+ 多分支 + CA 模块	96	91	82	60	73	44	49	54	68.64

表 5 双目数据集消融试验结果

Tab. 5 Results of binocular dataset ablation experiment %

试验方法	IoU								MIoU
	背景	成熟果实	绿色果实	花托	主干	侧枝	吊线	剪枝切口	
基线模型	91	95	66	59	70	48	42	57	66.02
+ CA 模块	91	95	71	60	72	50	41	59	67.26
+ 多分支	94	95	75	66	72	49	42	81	71.64
+ 多分支 + CA 模块	95	96	77	66	75	45	50	80	73.00

果实、绿色果实、花托、主干、侧枝、吊线和剪枝切口的交并比在基线模型的基础上分别提升 1、3、12、6、10、9、3 个百分点，MIoU 整体提升 6.40 个百分点。在双目数据集上，成熟果实、绿色果实、花托、主干、吊线和剪枝切口的交并比在基线模型的基础上分别提升 1、11、7、5、8、23 个百分点，侧枝的交并比下降 3 个百分点，MIoU 整体提升 6.98 个百分点。多分支解码结构在小目标的识别上起到了重要作用，改善了下采样带来的小目标细节轮廓信息丢失的问题，CA 注意力模块不仅获取了通道间信息，还考虑了方向相关的位置信息，在解码器中嵌入 CA 模块有利于模型减少浅层特征分支中潜在的噪声干扰。因此，本文提出的 CA 模块和多分支解码结构有助于提高模型的精度，使模型更好地定位和识别番茄图像中的小目标类别。

2.3.4　不同模型对比试验

为了验证本文提出的番茄图像分割模型的有效性，选取 HRNet[26]、UNet[27]、PSPNet[28] 与改进的 DeepLabv3 + 模型进行对比验证，试验结果如表 6 所示。改进的 DeepLabv3 + 模型内存占用量为 18.5 MB，与 HRNet、UNet、PSPNet 相比，分别减少 84%、89%、90%，输入图像分辨率为 800 像素 ×600 像素时，单幅图像分割时间为 55 ms，满足了实时性要求。分割精度方面，本文模型在单目数据集上的 MIoU 和 MPA 分别为 68.64%、78.59%，MPA 比 UNet 低 0.03 个百分点。在双目数据集上，本文模型的 MIoU 和 MPA 分别达到 73.00%、80.59%，优于其他模型。合成数据集虽然只作为预训练数据集使用，除 MIoU 比 UNet 低 0.42 个百分点外，MPA 也优于其他模型。因此，本文所提模型具有内存占用量少、精度高的特点。

不同模型的分割结果如图 9 所示。HRNet 模型在图像 2 和图像 4 中对成熟果实和绿色果实的边缘处理不到位，存在误分割现象，在图像 5 中对番茄主干部分的分割效果不够细致（边缘带有锯齿状）。

表 6　不同模型在番茄数据集上的试验结果

Tab. 6　Experimental results of different models on tomato datasets

模型	单目数据集		双目数据集		合成数据集		内存占用量/	推理时间/
	MIoU/%	MPA/%	MIoU/%	MPA/%	MIoU/%	MPA/%	MB	ms
HRNet	64.12	72.91	65.47	74.64	72.98	82.35	119.1	50
UNet	67.95	78.62	70.89	80.52	76.89	84.37	176.1	111
PSPNet	50.36	55.92	53.06	59.18	60.83	71.87	187.2	91
本文模型	68.64	78.59	73.00	80.59	76.47	85.29	18.5	55

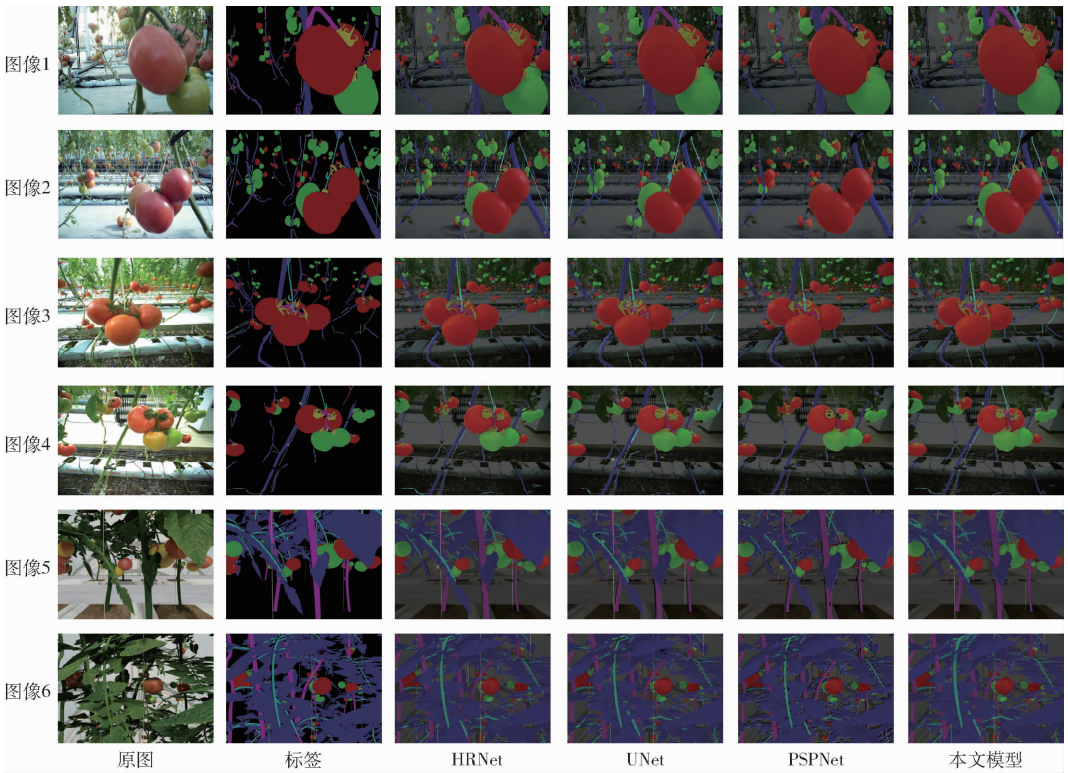


图 9　不同模型对番茄图像的分割结果

Fig. 9　Segmentation results of tomato images by different models

UNet 模型采用编-解码器结构,提高了网络的分割精度,但是对小目标的分割效果略差,例如在图像 3 中将花托分割成侧枝,以及将左下角的成熟果实误分割为绿色果实,在图像 4 中没有检测到主干上的剪枝切口,还将部分背景识别成吊线。PSPNet 模型由于在不同网络尺寸上使用了池化操作,导致高分辨率特征图丢失了小目标类别的细节信息,图像漏分割和误分割现象严重,例如在图像 2、图像 4 和图像 5 中,对吊线的识别率不高,在图像 2 中将下方的绿色果实识别为成熟果实,在图像 3 中将部分花托分割成侧枝,与其他模型相比,分割效果较差。本文基于 CA - GhostNet 主干网络改进的 DeepLabv3 + 模型,引入了多分支解码结构和 CA 注意力模块,不仅模型体量小,且能够提取多尺度信息,细化分割结果,保持番茄图像中小目标类别提取的完整性,达到了较为理想的分割效果。

3 结论

(1)以番茄图像多类别目标分割为研究对象,针对现有模型分割精度低、速度慢、参数多等问题,提出了一种改进的 DeepLabv3 + 轻量化模型。该模型内存占用量仅为 18.5 MB,单幅图像推理时间 55 ms,在 3 个番茄数据集上表现出了较强的泛化能力。

(2)使用 GhostNet 和 CA 注意力模块构建 CA - GhostNet 作为 DeepLabv3 + 模型的主干特征提取网络,提高了模型的特征提取能力,并有效降低了参数量。在解码器中引入 CA 模块和多分支解码结构,加强模型对小目标类别的分割能力。改进的 DeepLabv3 + 模型在单、双目数据集上的 MIoU 分别为 68.64%、73.00%,与基线模型相比,MIoU 分别提高 6.40、6.98 个百分点。

(3)基于单、双目数据集使用了数据增强和迁移学习的训练策略,对果实、主干、侧枝、吊线等 8 个语义类别进行分割,在有限的数据下得到了更好的分割结果。与 HRNet、UNet、PSPNet 相比,本文所提模型精度最高,内存占用量最少,且拥有较好的分割效果。

参 考 文 献

[1] 张勤, 刘丰溥, 蒋先平, 等. 番茄串收机械臂运动规划方法与试验[J]. 农业工程学报, 2021, 37(9): 149 - 156.
ZHANG Qin, LIU Fengpu, JIANG Xianping, et al. Motion planning method and experiments of tomato bunch harvesting manipulator[J]. Transactions of the CSAE, 2021, 37(9): 149 - 156. (in Chinese)

[2] 刘继展. 温室采摘机器人技术研究进展分析[J]. 农业机械学报, 2017, 48(12): 1 - 18.
LIU Jizhan. Research progress analysis of robotic harvesting technologies in greenhouse[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(12): 1 - 18. (in Chinese)

[3] 李寒, 张漫, 高宇, 等. 温室绿熟番茄机器视觉检测方法[J]. 农业工程学报, 2017, 33(增刊 1): 328 - 334,388.
LI Han, ZHANG Man, GAO Yu, et al. Green ripe tomato detection method based on machine vision in greenhouse[J]. Transactions of the CSAE, 2017, 33(Supp. 1): 328 - 334,388. (in Chinese)

[4] 丁永军, 张晶晶, LEE Won Suk, 等. 小波变换与分水岭算法融合的番茄冠层叶片图像分割[J]. 农业机械学报, 2017, 48(9): 32 - 37.
DING Yongjun, ZHANG Jingjing, LEE Won Suk, et al. Segmentation of tomato leaves from canopy images by combination of wavelet transform and watershed algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(9): 32 - 37. (in Chinese)

[5] 尹建军, 毛罕平, 王新忠, 等. 不同生长状态下多目标番茄图像的自动分割方法[J]. 农业工程学报, 2006, 22(10): 149 - 153.
YIN Jianjun, MAO Hanping, WANG Xinzhong, et al. Automatic segmentation method for multi-tomato images under various growth conditions[J]. Transactions of the CSAE, 2006, 22(10): 149 - 153. (in Chinese)

[6] 梁喜凤, 金超杞, 倪梅娣, 等. 番茄果实串采摘点位置信息获取与试验[J]. 农业工程学报, 2018, 34(16): 163 - 169.
LIANG Xifeng, JIN Chaoqi, NI Meidi, et al. Acquisition and experiment on location information of picking point of tomato fruit clusters[J]. Transactions of the CSAE, 2018, 34(16): 163 - 169. (in Chinese)

[7] 龙洁花, 赵春江, 林森, 等. 改进 Mask R - CNN 的温室环境下不同成熟度番茄果实分割方法[J]. 农业工程学报, 2021, 37(18): 100 - 108.
LONG Jiehua, ZHAO Chunjiang, LIN Sen, et al. Segmentation method of the tomato fruits with different maturities under greenhouse environment based on improved Mask R - CNN[J]. Transactions of the CSAE, 2021, 37(18): 100 - 108. (in Chinese)

[8] 胡春华, 刘炫, 计铭杰, 等. 基于 SegNet 与三维点云聚类的大田杨树苗叶片分割方法[J]. 农业机械学报, 2022, 53(6): 259 - 264.
HU Chunhua, LIU Xuan, JI Mingjie, et al. Single poplar leaf segmentation method based on SegNet and 3D point cloud clustering in field[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(6): 259 - 264. (in Chinese)

[9] 冯青春，成伟，李亚军，等. 基于 Mask R－CNN 的番茄植株整枝操作点定位方法[J]. 农业工程学报，2022，38(3)：128－135.
FENG Qingchun，CHENG Wei，LI Yajun，et al. Method for identifying tomato plants pruning point using Mask R－CNN[J]. Transactions of the CSAE，2022，38(3)：128－135.(in Chinese)

[10] 任守纲，贾馥玮，顾兴健，等. 反卷积引导的番茄叶部病害识别及病斑分割模型[J]. 农业工程学报，2020，36(12)：186－195.
REN Shougang，JIA Fuwei，GU Xingjian，et al. Recognition and segmentation model of tomato leaf diseases based on deconvolution-guiding[J]. Transactions of the CSAE，2020，36(12)：186－195.(in Chinese)

[11] 刘文波，叶涛，李颀. 基于改进 SOLO v2 的番茄叶部病害检测方法[J]. 农业机械学报，2021，52(8)：213－220.
LIU Wenbo，YE Tao，LI Qi. Tomato leaf disease detection method based on improved SOLO v2[J]. Transactions of the Chinese Society for Agricultural Machinery，2021，52(8)：213－220.(in Chinese)

[12] 熊俊涛，刘柏林，钟灼，等. 基于深度语义分割网络的荔枝花叶分割与识别[J]. 农业机械学报，2021，52(6)：252－258.
XIONG Juntao，LIU Bolin，ZHONG Zhuo，et al. Litchi flower and leaf segmentation and recognition based on deep semantic segmentation[J]. Transactions of the Chinese Society for Agricultural Machinery，2021，52(6)：252－258.(in Chinese)

[13] 周玲莉，任妮，张文翔，等. 应用于农业场景视觉解析任务的番茄数据集[J]. 农业大数据学报，2021，3(4)：70－76.
ZHOU Lingli，REN Ni，ZHANG Wenxiang，et al. Tomato dataset for agricultural scene visual parsing tasks[J]. Journal of Agricultural Big Data，2021，3(4)：70－76.(in Chinese)

[14] CHEN Liang-Chieh，ZHU Yukun，PAPANDREOU G，et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]// Proceedings of the European Conference on Computer Vision (ECCV)，2018：801－818.

[15] CHOLLET F. Xception：deep learning with depthwise separable convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2017：1251－1258.

[16] HAN Kai，WANG Yunhe，TIAN Qi，et al. Ghostnet：more features from cheap operations[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2020：1580－1589.

[17] HOU Qibin，ZHOU Daquan，FENG Jiashi. Coordinate attention for efficient mobile network design[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2021：13713－13722.

[18] HU Jie，SHEN Li，SUN Gang. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2018：7132－7141.

[19] 徐浩，杨德刚，蒋倩倩，等. 基于 SSD 的轻量级车辆检测网络改进[J]. 计算机工程与应用，2022，58(12)：209－217.
XU Hao，YANG Degang，JIANG Qianqian，et al. Improvement of lightweight vehicle detection network based on SSD[J]. Computer Engineering and Applications，2022，58(12)：209－217.(in Chinese)

[20] 范晶晶，薛皓玮，吴欣鸿，等. 引入重影特征映射和通道注意力机制的手势识别算法[J]. 计算机辅助设计与图形学学报，2022，34(3)：403－414.
FAN Jingjing，XUE Haowei，WU Xinhong，et al. Gesture recognition algorithm introducing ghost feature mapping and channel attention mechanism[J]. Journal of Computer-Aided Design & Computer Graphics，2022，34(3)：403－414.(in Chinese)

[21] GHIASI G，LIN Tsung-Yi，LE Q V. NAS-FPN：learning scalable feature pyramid architecture for object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2019：7036－7045.

[22] 邓建国，张素兰，张继福，等. 监督学习中的损失函数及应用研究[J]. 大数据，2020，6(1)：60－80.
DENG Jianguo，ZHANG Sulan，ZHANG Jifu，et al. Loss function and application research in supervised learning[J]. Big Data Research，2020，6(1)：60－80.(in Chinese)

[23] 颜宇，陈帆，和红杰. 基于边界增强损失的海马体分割算法[J]. 光电子·激光，2020，31(3)：299－309.
YAN Yu，CHEN Fan，HE Hongjie. Hippocampus segmentation with boundary enhanced loss[J]. Journal of Optoelectronics·Laser，2020，31(3)：299－309.(in Chinese)

[24] SANDLER M，HOWARD A，ZHU Menglong，et al. Mobilenetv2：inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2018：4510－4520.

[25] 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.

[26] SUN Ke，XIAO Bin，LIU Dong，et al. Deep high-resolution representation learning for human pose estimation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2019：5693－5703.

[27] RONNEBERGER O，FISCHER P，BROX T. U-Net：convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015：18th International Conference，Munich，2015：234－241.

[28] ZHAO Hengshuang，SHI Jianping，QI Xiaojuan，et al. Pyramid scene parsing network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2017：2881－2890.