

[文章编号] 1003—4684(2022)01-0046-08

基于 Mask R-CNN 的槟榔片分割算法研究

舒 军¹, 何俊成², 李振亚²

(1 湖北工业大学电气与电子工程学院, 湖北 武汉 430068

2 湖北工业大学太阳能高效利用及储能运行控制湖北省重点实验室, 湖北 武汉 430068)

[摘 要] 针对槟榔片加工过程中自动化程度低,提出一种基于 Mask R-CNN 的槟榔片自动分级算法:采用改进高分辨率网络作为主干网络提取槟榔片表面信息,减少多次卷积过程中信息损失,提高轮廓分割准确度;采用改进卷积块注意力模型加强内轮廓中有效信息提取,抑制无效信息,提高有核无核籽判断准确率;同时采用深度可分离卷积减少高分辨率主干网络带来的参数量、计算复杂度增加问题。实验结果表明改进后网络相比未改进前分割精度提高 2.5%,同时有核无核籽判断准确率提高 5.93%,总浮点计算量降低 64.9%。

[关键词] Mask R-CNN; 高分辨率网络; 卷积块注意力; 槟榔片

[中图分类号] TP391.4 [文献标识码] A

槟榔片的分级作为槟榔加工中重要一步,主要通过测量槟榔片切面内外轮廓的长度和宽度完成分级,其分级标准见表 1,同时需要检测内轮廓内核是否去除干净。传统槟榔片分级采用人工操作,生产成本高,效率低,而且因人为因素分级标准不一。引入自动化设备进行自动分级是解决问题的良好方式,其关键技术在于槟榔片外轮廓与背景分割和内轮廓与外轮廓进行分割并同时完成内轮廓残留核的判断。

表 1 槟榔分级标准

等级	外观	果径/mm	
		长度	宽度
泡籽	无缺陷	$L>70$	$D>21$
尾期籽	无缺陷	$L\leq 45$	$D\leq 15$
小籽	无缺陷	$45<L\leq 54$	$15<D\leq 19$
中籽	无缺陷	$54<L\leq 58$	$16<D\leq 21$
大籽	无缺陷	$58<L\leq 70$	$16<D\leq 21$
次品	有缺陷	—	—

传统的图像分割方法主要包括基于阈值、边缘、区域、聚类以及图论的图像分割方法。它们主要依靠像素之间的差异完成不同区域分割,然而当像素之间差异较小或者图像内容复杂时会极大影响分割的准确度^[1]。槟榔片内外轮廓颜色差异较小,传送带脏污同样会导致轮廓分割效果差,无法完成残留核信息判断,仍需单独设计分类算法进行处理,增加系统复杂性。

近年来深度学习发展如火如荼,针对图像分割领域深度学习网络模型进展迅速,涌现出许多优秀网络模型如 U-Net、SegNet 等,同传统分割算法相比,其分割精度、鲁棒性大大增强。但是在槟榔片分级项目中,同传统算法一样,无法在完成分割的同时判断残留核的信息。Mask R-CNN 在 Faster R-CNN 基础上发展而来^[2],在检测、分类的基础上增加语义分割网络。能够完成槟榔片内外轮廓语义分割的同时判断核残留信息。

针对槟榔片自动化加工领域中出现问题,本文以能够同时完成分割、分类任务的 Mask R-CNN 为基础,并对其进行有针对性改进,完成槟榔片自动分级任务。

1 Mask R-CNN 网络

Mask R-CNN 是一个两阶段的实例分割网络,能够同时完成目标检测、分类、语义分割等任务。该算法在 Faster R-CNN 的基础上将图像分割和目标检测融合在一起,可以实现实例级的图像分割。整体框架见图 1。

Mask R-CNN 中采用 ResNet-50 作为主干,通过对输入图片进行不同尺度的特征提取获取 C1、C2、C3、C4、C5 层特征。然后使用这些特征层构建特征金字塔,实现多尺度融合^[3]得到有效特征层 P2、P3、P4、P5、P6。在区域提取网络,有效特征层根据不同感受视野的大小设置不同比例、尺寸的候

[收稿日期] 2021—07—30

[第一作者] 舒 军(1973—),男,湖北荆门人,工学博士,湖北工业大学副教授,研究方向为机器视觉,深度学习

[通信作者] 何俊成(1995—),男,湖北恩施人,湖北工业大学硕士研究生,研究方向为机器视觉,深度学习

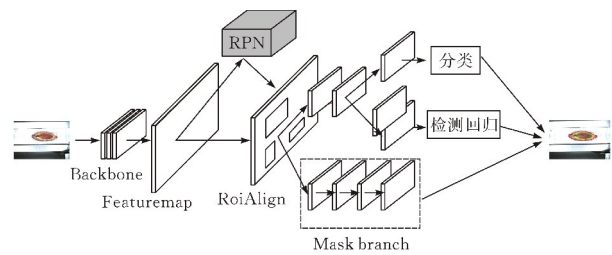


图 1 Mask R-CNN 基本结构

选框,生成可能包含目标的区域候选框。最后通过非极大值抑制(NMS)^[4]对重复候选框进行筛选去除。候选框在有效特征层上经过特征池化 ROI Pooling^[5]操作完成截取,送入目标检测与语义分割网络。在目标检测网络对候选框进行回归定位,再次进行非极大值抑制,获取最终预测框,同时对候选框内目标完成检测分类。将获得的预测框送入语义分割网络,利用此预测框对有效特征层进行截取,并对其像素点进行分割,最终生成分割完整语义信息。

2 Mask R-CNN 网络改进

在槟榔片品质分级中,存在以下几个难点。第一,槟榔内外轮廓之间界限不明,且轮廓上存在着的脏污会使分割受到较大影响。第二,槟榔片内部残留的核与果壁粘贴紧密,且不同槟榔片的残留核大小形态颜色各异,判别比较困难。第三,为保证选片分级的效率,必须要对算法速度进行优化,以满足厂商的生产要求。针对以上问题,本文在 Mask R-CNN 的基础上对其进行改进:首先采用改进的高分辨率特征提取网络(HRNet)^[6],提取槟榔片表面的信息,维持高分辨率特征信息,减少图片卷积计算时损失,提高分割精度。其次引入改进注意力模块(CBAM)^[7],充分考虑空间重要信息,增强有效信息,抑制无效信息,加强有核籽的判断准确率。最后,采用深度可分离卷积(Seperable Convolution)^[8]代替普通常规卷积方式,将一次卷积分为 Depthwise 和 Pointwise 两个过程分别进行计算,理论与实验结果均表明深度可分离卷积能够有效降低参数量,减少浮点运算总量。

2.1 改进高分辨率特征提取网络

采用并行连接不同分辨率并重复进行多尺度融合提取 HRNet 高分辨率特征,维持高分辨率的表示,无需从低分辨率特征图上采样渐渐恢复成原图大小,可直接保持高分辨率的特征图,在高分辨率特征上直接提取深层抽象语义信息,而不是仅仅将其作为低分辨率分支特征信息的补充。不仅能够提取更多丰富信息而且得到的位置信息更加精确。不同分辨率支路之间通过特征融合单元,进行特征图尺寸大小转换,便于不同特征进行融合相加。其基

本网络结构见图 2。

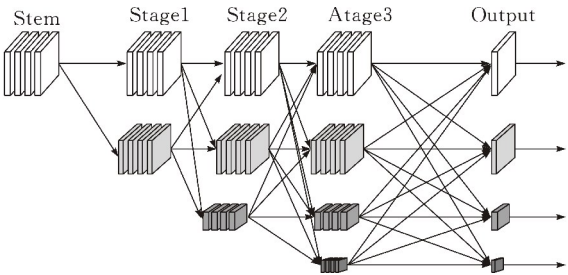


图 2 HRNet 主干特征提取网络结构

HRNet 高分辨率网络虽然能够保持以较高分辨率最终完成语义分割,并减少卷积、池化运算过程图片信息的损失,但是后续不同阶段增加的低分辨率分支深度逐渐变浅,不利于输入样本图片低分辨率抽象特征信息的获取。低分辨率子网络特征图内包含输入图像的全局特征,如果不能对其进行充分学习,对后续区域网络构建不利,最终导致网络训练不充分,对网络最终训练效果有较大影响,不利于最终分割效果。

提高深度学习神经网络获取抽象特征信息能力,主要体现在增加网络深度与扩展网络宽度两方面。增加网络深度通过叠加卷积块实现,多次卷积能够获得更好的非线性表达能力,拟合更加复杂特征。但是网络的加深带来一定的负面影响,深层网络往往伴随梯度不稳定,网络退化等问题,导致网络整体性能出现下降。ResNet 网络中提出的跃层连接结构在一定程度上缓解网络退化问题,但是网络加深带来的负面影响不能被完全消除^[9];扩展网络宽度通过对同一输入特征图进行多维度不同卷积计算实现,从不同维度学习更加丰富特征,最后将不同特征进行融合相加得到包含丰富语义信息特征图。相比于通过叠加密集卷积块增加网络深度,扩展网络宽度不仅能够获得丰富抽象语义信息,同时保证整体网络结构的稀疏性,减少不确定因素的影响。

对于低分辨率分支训练不充分、无法获取丰富特征语义信息。通过扩展网络宽度的方式从空间不同维度提高网络低分辨率分支对抽象信息的获取。相比于通过加深网络深度达到网络训练最优化,采用多种不同尺度卷积核分别对低分辨率分支输入特征图进行卷积计算,不仅能够获取不同尺度下特征图抽象信息,同样能够最大程度避免由于网络过深导致网络退化。改进的低分辨率分支卷积块如图 3 所示。

低分辨率分支输入特征经过不同支路,采用不同感受视野的卷积提取输入特征图像抽象信息。每次卷积计算完成后,使用 Relu 函数进行激活,增加网络支路特征的非线性。同时模块中引入残差连

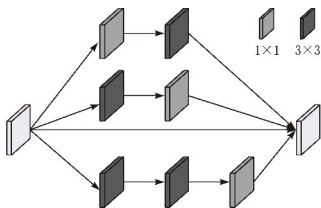


图 3 改进后低分辨率支路

接,加速网络的收敛,不仅能够避免模型出现过拟合导致退化,还能够进一步利用输入特征层的原始特征参数,增强输入原始特征提取能力。最终将各支路分别得到的特征图进行融合相加,使用 BN 进行参数归一化,同时采用 Relu 激活函数进行激活操作,增加网络的非线性。最后将输出特征层结果与其余分辨率支路进行多特征融合后,输入至后续模块。

将改进后低分辨率分支卷积块应用于整个网络进行重新训练,实验条件设置相同。相比于改进前,改进后网络语义分割 Loss 值经过多次迭代训练后更好更快地达到收敛状态见图 4b。

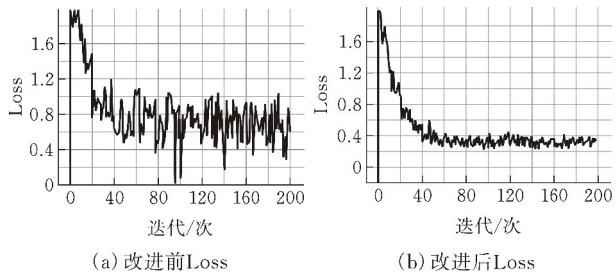


图 4 改进前后 Loss

最终 Loss 值稳定在 0.3 左右。改进前特征提取网络语义分割 Loss 走势见图 4a,其分割 Loss 值基本在 0.4~1.0 之间震荡,没有完全收敛至一确定数值。实验结果表明,改进前网络训练并不充分,仍在一定程度上丢失一部分信息特征。改进后网络低分辨率卷积块分别经过不同分支进行特征提取,最终融合相加,充分利用输入特征层信息,整体网络更快地达到收敛状态,训练达到最优状态。

2.2 改进卷积块注意力模型

注意力机制的本质是使网络模型聚焦到显著特征上,而抑制或者忽略无关特征。在槟榔片的分级系统中,增加网络模型对残核的关注,以便更好地判断是否存在残留核。

卷积块注意力模型由 Woo 等人于 2018 年提出,其将通道注意力和空间注意力结合,能够嵌入到任意网络模型中^[10]。空间注意力模型结构见图 5。

首先输入特征层 C_i 分别进行全局平均池化和全局最大池化,将所有通道特征图转换为单通道特征图。将池化之后的特征图进行叠加,随后进行卷积核为 7×7 的卷积操作后再利用 sigmoid 进行激

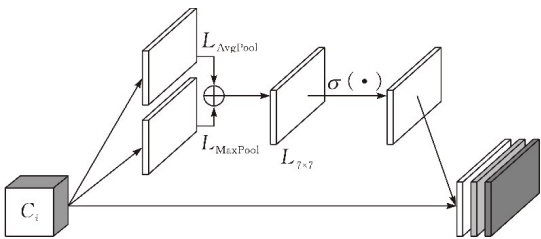


图 5 空间注意力模型结构

活,最后将激活后的特征图与输入特征图相乘,输出最终的加权特征。如:

$$M_S(C_i) = \sigma(f^{7 \times 7}([AvgPool(C_i); MaxPool(C_i)])) = \sigma(f^{7 \times 7}([F_{avg}^S; F_{max}^S]))$$

通道注意力机制将特征层各通道分别进行全局平均池化和全局最大池化,空间结构见图 6。

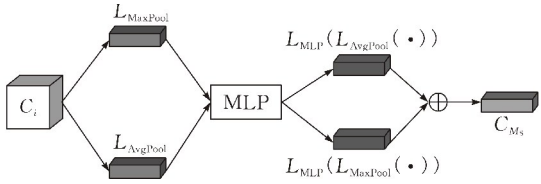


图 6 通道注意力机制

分别进行池化后,将输出特征层经过多层感知机(MLP)进行参数共享,输出特征进行对应相加操作,最后经过 sigmoid 函数进行激活,生成最终通道注意力特征图。如下:

$$M_C(C_i) = \sigma(MLP(AvgPool(C_i)) + MLP(MaxPool(C_i))) = \sigma(W_1(W_0(F_{avg}^C)) + W_1(W_0(F_{max}^C)))$$

在传统注意力网络结构中,针对槟榔核检测,使用全局最大池化与全局平均池化能够一定程度上加强有效核特征信息的提取,但是对无效干扰信息不能进行有效抑制。对槟榔片训练过程进行不同阶段特征图可视化分析(图 7),图 7b 为没有添加卷积块注意力模型网络,可以看出特征图中存在大量无关特征信息干扰网络判断。

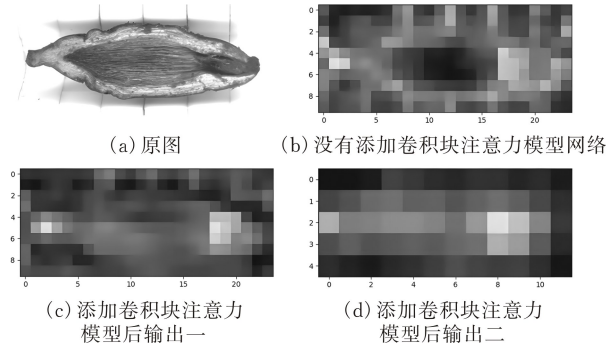


图 7 传统注意力模块特征图可视化对比

图 7c、7d 为添加卷积块注意力模型后不同阶段输出的特征图。由图中可以看出,较多无关干扰信息经过注意力模型的卷积池化加权后被抑制,同时有效信息得到一定加强。但是对无效信息不能够完

全抑制,如图 7c 左侧区域。由此可见,常规注意力结构在槟榔片分级研究案例中有一定效果,同时也存在一定弊端,无效干扰信息错误提取对最终结果的判断有较大的影响。

为了能够进一步抑制无效干扰信息,提高残留核的判断能力。对传统注意力网络进行补充:增加步长为 1,卷积核 5×5 最大池化层与 3×3 平均池化层,分别对空间注意力与通道注意力进行调整,以通道注意力结构为例,改进见图 8。平均池化操作即对一点邻域内像素值求平均,并使用求得平均值代替该特征点像素值。改进后注意力模块通过不同尺度全局最大池化与全局平均池化对含有抽象语义信息的输入特征图进行平滑操作,降低无关信息干扰。为防止最大池化对有效抽象语义信息造成过多损失,同时尺度为平均池化操作对损失抽象语义信息进行一定补充。不同尺度下输入特征图进行池化操作并完成加权计算,最终送入 MLP 完成参数共享。

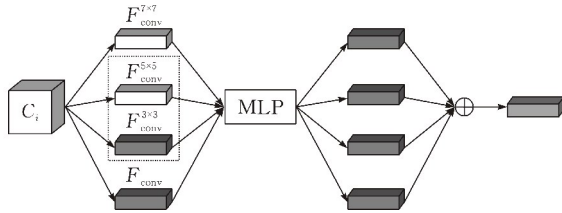


图 8 改进后通道注意力模块

在特征层各通道进行卷积池化计算时,不同大小卷积核尺寸会有不同大小的输出特征图。计算公式如下:

$$C_o = \frac{C_i - K + 2P}{S} + 1$$

其中 C_o 为输出特征图大小, C_i 为输入特征图大小, K 为卷积核尺寸, P 为填充像素个数, S 为卷积步长。为了使经过不同池化卷积核大小计算得到的特征图能够进行相加融合,对不同卷积核计算时,当卷积核尺寸为 7×7 ,填充像素 Padding 设为 3,当卷积核尺寸为 5×5 ,填充像素 Padding 设为 2,当卷积核尺寸为 3×3 ,填充像素 Padding 设为 1。改进后通道注意力结构,如下:

$$M'_C(C_i) = \sigma(W_1(W_0(L_{7 \times 7})) + W_1(W_0(L_{5 \times 5})) + W_1(W_0(L_{3 \times 3})) + W_1(W_0(L_{avg})))$$

其中: C_i 表示输入特征层, $L_{7 \times 7}$ 、 $L_{5 \times 5}$ 、 $L_{3 \times 3}$ 和 L_{avg} 分别表示 7×7 的卷积、 5×5 的卷积、 3×3 的卷积、平均池化层, $W_0 \in R_{C/r \times C}$ 和 $W_1 \in R_{C \times C/r}$ 分别为 MLP 的权重。改进后注意力模块不同阶段特征见图 9,可以看出干扰无效信息通过两种不同尺度最大池化与平均池化后能够得到有效抑制,同时有

效核特征信息并没有受到较大影响,依然能够得到有效加强。实验效果对比可以看出,对卷积块注意力模块的改进达到预期目的,能够加强语义分割过程中有效特征,同时抑制无效干扰特征,提高后续分类的准确率。

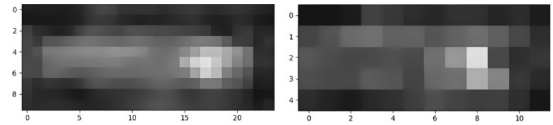


图 9 改进后特征图可视化对比

在高分辨率网络特征融合阶段引入改进后卷积注意力模块,增强有效特征,抑制无效特征。计算公式如下:

$$P_i = C_i \otimes M'_C(C_i) \oplus L_{\text{upsampling}}(P_{i-1})$$

其中: C_i 代表第 i 阶段输出得到的特征图, P_i 代表融合后生成的特征图, M'_C 表示改进后通道注意力模型, $L_{\text{upsampling}}$ 表示上采样。计算示意图见图 10。

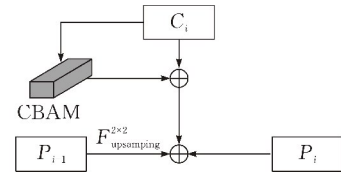


图 10 CBAM 计算示意图

2.3 深度可分离卷积

工业生产中除了对生产的品质有要求,对生产的效率同样重视。随着深度学习网络的不断深化,模型的结构也变得越来越复杂,降低网络参数量提高运算速度也愈发重要。

常规卷积采用“空间-跨通道”的特征学习方式,在进行空间特征提取的同时完成通道合并。深度可分离卷积将常规的一次卷积计算分为 Depthwise convolution 和 Pointwise convolution 两步完成。对于一幅三通道的输入,首先对三个通道分别进行卷积计算,输出三个通道的特征图。其次,使用卷积核对三个通道再次进行卷积计算,得到的输出与一次正常卷积得到的输出大小维度相同,但是一次卷积的参数数量大大降低。假设输入大小为 $h \times w \times c$ 的图像,卷积核尺寸为 $h_1 \times w_1$,输出特征图尺寸为 $H \times W \times N$,则标准卷积参数数量为 $h_1 \times w_1 \times c \times H \times W \times N$ 。对于可分离卷积,在 Depthwise convolution 阶段参数数量为 $H \times W \times c \times h_1 \times w_1 \times 1$,Pointwise convolution 阶段参数数量为 $H \times W \times N \times 1 \times 1 \times c$,所以深度可分离卷积与标准卷积相比如下:

$$\frac{H \times W \times c \times h_1 \times w_1 \times 1 + H \times W \times N \times 1 \times 1 \times c}{h_1 \times w_1 \times c \times H \times W \times N} = \frac{1}{N} + \frac{1}{h_1 \times w_1}$$

针对不同的输入输出通道数及卷积核大小,卷积计算的参数量的缩减幅度并不相同,但是从上述公式可以明显看出,相比于传统卷积,深度可分离卷积参数量在一定程度上会减少,从而加快网络模型的运行速度。

3 实验

本文根据实际生产的要求,对 Mask R-CNN 做出相应的改进。能够更加精确地对内外轮廓进行分割,对翻面、是否存在残留核进行判断。同时对网络运行速度做出一定优化,降低参数数量,提高运行效率。现通过对比试验分析,验证上文的改进算法在槟榔片分级过程中的有效性。

3.1 实验环境及评价指标

3.1.1 实验环境 本文中研究槟榔片分级实验操作系统为 Windows 10;CPU 为 Intel Core i5-4590, GPU 为 Nvidia Quadro P5000,基础软件环境 CUDA 10.0 cuDNN 7.0,gcc/g++ 5.0 OpenCV 3.4.0;深度学习框架为 Keras/TensorFlow。

3.1.2 评价指标 训练完成后的模型文件,采用 COCO 竞赛官方评价指标 AP(average precision)衡量算法的分割效果。采用分类的正确率(accuracy)评价网络模型对翻面籽、有核、无核籽的分类能力。对网络模型运算速度及模型复杂度采用参数量,以及浮点计算量(floating point operations, FLOPs)作为评价标准完成评价。对于一次卷积运算,其浮点计算量公式为

$$FLOPs = (2 \times C_i \times K \times K - 1) \times H \times W \times C_o$$

其中 C_i , C_o 为输入输出特征图通道数, K 为卷积核的大小, $H \times W$ 为输出特征图大小。针对模型激活函数(activation function)以及偏置(bias)其他方面计算的不同考虑,对 FLOPs 的计算方式不尽相同。为便于对比分析,本文采用 Github 中开源文件 count_flops.py 对模型参数量、浮点计算量进行统一计算。

3.2 实验结果分析

实验数据集内共包含槟榔片样本 6726 张,其中用作测试集 1350 张。训练 epoch 为 200 次,step_per_epoch 为 100,初始动量设置为 0.9,batch size 为 2,学习率为 0.001,权重衰减系数 0.0001。选取 (0.50:0.05:1)等不同 IoU 阈值,通过计算 AP 值衡量算法性能。

3.2.1 轮廓分割结果对比 只采用高分辨率的 HRNet 模型作为主干网络、改进 HRNet 模型网络和未改进的 Mask R-CNN 模型训练,数据结果对比见表 2。

表 2 不同主干网络分割 AP 值

主干网络	mAP / %	AP ₅₀ / %	AP ₇₅ / %
ResNet50	61.3	90.2	73.7
HRNet	65.7	91.3	75.58
改进后 HRNet	70.4	92.7	76.6

如表 1 所示,改进后采用高分辨率网络进行特征的提取,能够充分训练并利用不同分辨率之间特征的差别进行信息的补充,实现更高精度的分割。以 AP_{50} 为例进行分析,未改进 Mask R-CNN 主干网络采用 ResNet50 进行特征提取,搭建特征金字塔输出特征层,进行后续语义分割。相比于改进前高分辨率特征提取网络 HRNet,其分割准确度低 1.1%,数据表明,高分辨率网络中不同分支之间信息相互补充对特征提取有一定帮助。对高分辨率主干特征提取网络低分辨率进行扩宽改进后,分割准确率提高 1.4%,相比原版 Mask R-CNN 提高 2.5%。实验结果证明对低分辨率分支进行的改进能够进一步促进网络训练,提高抽象特征提取能力,增加语义分割精准度。

针对实验的具体效果,选取具有代表性的样本分割效果图(图 11),图 11a 为训练标注文件,图 11b 为原始 Mask R-CNN 模型分割结果,可以看出原始网络对槟榔片内外轮廓有较大的误分割,受到阴影与传送带脏污影响较大,不能准确识别内轮廓边缘造成误分割。图 11c 主干特征网络为改进前 HRNet 分割结果,相比于原始 Mask R-CNN 网络,其对外轮廓边缘处理相对较好,受到阴影影响较小,但是对内外轮廓处仍不能进行准确分割,不能完整正确识别边缘。图 11d 为改进后 HRNet 作为主干网络分割效果图,可以看出其对外轮廓边缘控制更加稳定,能够有效分割阴影与脏污点边缘。同时相比于前两种算法模型,对内外轮廓分割较为精准。通过样本的分割实际效果与实验量化指标对比,能够证明改进的分割算法性能优越。改进后采用高分辨率网络进行特征提取,能够很好地利用不同分辨率之间特征的差别进行信息补充,从而实现更高精度的分割。

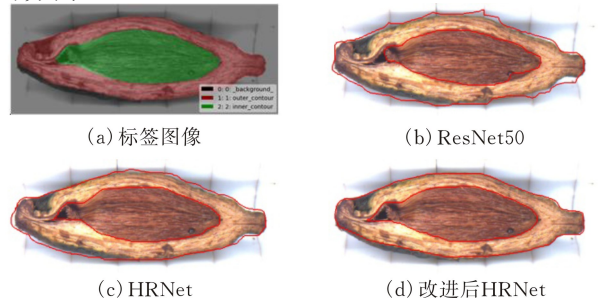


图 11 不同网络分割效果图

3.2.2 卷积块注意力改进效果对比 槟榔片内轮廓残留核的判别对槟榔片的分级至关重要,将其作为分类问题进行处理。在主干特征提取网络特征融合阶段加入改进空间注意力模块,通过增加 5×5 最大池化与 3×3 平均池化,强化内轮廓中残留核信息的提取,同时抑制其余无关信息的干扰。

通过类别激活热图(CAM)^[11]可视化,通过卷积层输出特征图,根据输出的类别对不同通道之间梯度加权映射,能够在模型进行类别判断时,定位图像中哪部分像素对模型最终判断输出结果有更大贡献。通过具体样本进行分析,改进后卷积块注意力模型与改进前模型对残留核有效信息关注对比见图 12。

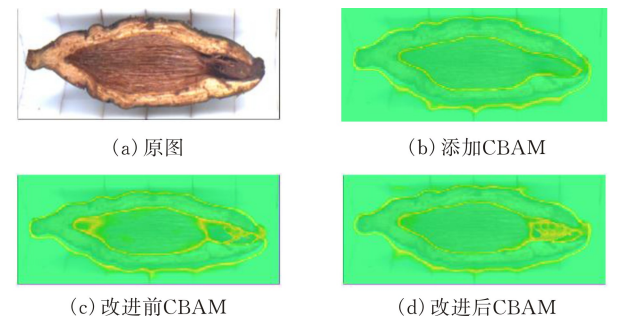


图 12 不同模型类别激活热图

图 12a 为带核槟榔片样本。图 12b 为未添加卷积块注意力结构,主干网络为改进后 HRNet 网络模型,可以看出网络对槟榔片内部残留核的关注较少,不能有效对残留核的有效信息进行加强。仍通过整个内轮廓整体信息对残留进行判断。图 12c 为改进前卷积块注意力结构模型,可以看出改进前卷积块注意力模型对残留核的有效信息有一定的加强作用,同时对无效干扰信息有一定抑制作用。但是并不能完全清除无效信息干扰,对槟榔片类别判断有一定负面影响。图 12d 为改进后卷积块注意力模型结构模型作用后结果,采用多层次不同卷积核分别对输入特征进行池化,将不同特征信息加权后可以看出,槟榔片内部无效干扰信息被抑制,残留的核有效信息得到加强,对残留核信息有较为精准的判断。从图 12 可以看出:相比于改进前的卷积块注意力模型,改进后模型能够加强有效信息,同时抑制无效信息的作用更加明显,能有效提高对槟榔片是否存在残留核的判断准确率。

针对槟榔片残留的特征对 Mask R-CNN 网络模型进行相应的改进,最终的分类准确率见图 13。

由表 3 看出:未添加卷积块注意力网络模型对有核槟榔片分类准确率仅为 89.70%,在将主干网络中加入未改进卷积块注意力结构后,其分类准确率达到 92.52%,能够证明卷积块注意力模块,对有效

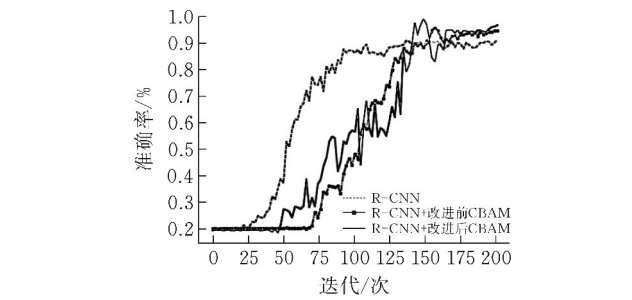


图 13 不同网络分类准确率

特征有一定的加强作用,能够提高类别判断准确率。同时,将改进后卷积块注意力结构加入主干网络模块,对有效特征进行强化后,其准确率达到 95.63%,相比于其余两种算法模型,准确率分别提高 5.93%, 3.11%。实验结果表明,改进后的卷积块注意力模型能够进一步对无效干扰信息进行抑制,同时加强有效特征信息,提高特征信息获取能力,提高槟榔片有核无核籽的判断准确率。

表 3 不同网络模型具体分类准确率

主干网络	准确率/%
改进后 HRNet	89.70
改进后 HRNet+改进前 CBAM	92.52
改进后 HRNet+改进后 CBAM	95.63

3.2.3 参数量与计算量对比结果 高分辨率分支与低分辨率分支特征信息的相互补充,进行多次多尺度融合与低分辨率分支的扩宽,能够使网络进行更加充分训练,带来精度准确率上的提升,但是过多特征融合与低分辨率支路的扩宽,势必会导致参数量与计算量一定程度的增加,为能够兼顾精度与速度,采取深度可分离卷积代替常规普通卷积计算,将一次卷积过程分为两次运算,降低参数数量,减小运算量,提高运算速度。对三种模型进行实验对比,不同算法模型网络参数量及总浮点运算量 FLOPs 量化结果见表 4。

表 4 不同模型参数量与计算量

主干网络	参数量	FLOPs
模型一		
改进前 HRNet	64.18M	60.8GFLOPs
模型二		
改进前 HRNet+深度可分离	34.6M	17.52GFLOPs
模型三		
改进后 HRNet+深度可分离	36.1M	21.35GFLOPs

从表 4 可以看出:模型二使用深度可分离卷积进行计算,参数量相比于改进前未采用深度可分离卷积,主干网络为改进前高分辨率的模型一降低 41.4%,总浮点运算量降低 71.2%。同时由于高分辨率网络的重复多尺度融合与对低分辨率支路扩展的改进,模型三参数量及总浮点运算量相比模型二

增加 4.15%和 17.9%。但是相比于未使用深度可分离卷积计算的模型一参数量降低 43.7%，总浮点运算量降低 64.9%。

实验具体量化指标可以看出：采用深度可分离卷积能够降低参数量与总浮点计算量，从而加快网络运算速度，提高检测能力。对单张图像进行检测耗时实验对比，结果见表 5。

表 5 单张图像检测耗时结果

网络类型	最少时间/ms	最大时间/ms	平均时间/ms
模型一	64.29	77.93	71.11
模型二	42.32	51.65	46.98
模型三	45.24	59.37	52.31

相比于模型一，采用深度可分离卷积后，模型二计算耗时分别降低 33.9%，平均检测时间减少 24.13 ms。模型三网络结构相比模型二更加复杂，相比于模型一其对单张图像耗时降低 26.4%，平均检测时间减少 18.8 ms。

3.2.4 总体网络训练损失 在改进 Mask R-CNN 网络模型中，Loss 值的计算分为三部分，分别为分类误差、检测误差和分割误差， $L = L_{cls} + L_{box} + L_{mask}$ 。在对最优模型进行训练时，随着迭代次数的增加，损失值 Loss 不断下降，在进行 200 次迭代后 Loss 值稳定在 0.1 左右，同时测试集损失值 val_Loss 随 Loss 下降不断下降，最终趋于稳定，说明训练模型的性能良好(图 14。)

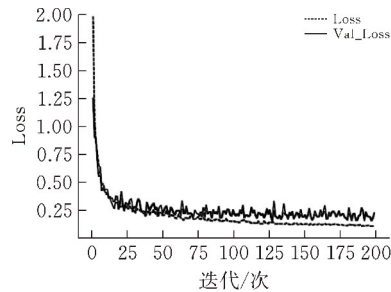


图 14 训练总体 Loss

4 结论

本文对 Mask R-CNN 原网络进行分析介绍，并针对槟榔片自动分级系统难点进行相应的改进：采用高分辨率特征提取网络作为主干网络进行特征提取，加强包含空间位置信息的高分辨率特征与包含丰富语义信息的低分辨率特征融合，扩宽低分辨率子网络，充分提取低分辨率分支进行特征语义信息，提高分割准确度；同时针对是否残留核判别难度大的问题，采用改进卷积块注意力机制，加强有核籽不同特征图中有效特征，抑制无效干扰特征；高分辨率特征的重复利用，带来参数量、计算量的增加。为了在保证准确率、精度的基础上提高运算速度，采用深

度可分离卷积代替传统卷积进行计算，达到降低参数量，减少运算复杂度的要求。实验表明：改进后的 Mask R-CNN 网络在分割精确度、分类判断准确率、网络计算量方面具有一定的优势。

[参 考 文 献]

[1] 黄鹏, 郑淇, 梁超. 图像分割方法综述[J].武汉大学学报(理学版), 2020, 66(6):519-531.

[2] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C] // 2017 IEEE International Conference on Computer Vision (ICCV), 2017,2980-2988. 10.1109/ICCV.2017.322.

[3] LIN T, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C].//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.936-944. 10.1109/CVPR.2017.106.

[4] NEUBECK A, GOOL L V. Efficient non-maximum suppression[C] // 18th International Conference on Pattern Recognition (ICPR'06), 2006.850-855. 10.1109/ICPR.2006.479.

[5] GIRSHICK R. Fast R-CNN[C]// 2015 IEEE International Conference on Computer Vision (ICCV), 2015. 1440-1448. 10.1109/ICCV.2015.169.

[6] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019. 5686-5696. 10.1109/CVPR.2019.00584.

[7] WOO S, PARK J, LEE J-Y, et al. CBAM: convolutional block attention module[C]// Cham:Springer International Publishing, 2018,3-19.

[8] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.1800-1807. 10.1109/CVPR.2017.195.

[9] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning[C]// AAAI, 2017.

[10] 袁嘉杰, 张灵, 陈云华. 基于注意力卷积模块的深度神经网络图像识别[J]. 计算机工程与应用, 2019, 55: 9-16.

[11] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2):336-359.

Research on Grading Algorithm of Betel Nut Slices Based on Mask RCNN

SHU Jun¹, HE Juncheng², LI Zhenya¹

(1 School of Electrical and Electronic Engineering, Hubei Univ. of Tech., Wuhan, 430000, China;
2 Hubei Key Laboratory for High-efficiency Utilization of Solar Energy and Operation Control of Energy Storage System, Hubei Univ. of Tech., Wuhan, 430000, China)

Abstract: Aiming at the low degree of automation in the processing of betel nut slices, an automatic grading algorithm for betel nut slices based on Mask R-CNN is proposed. An improved high-resolution network is used as the backbone network to extract the surface information of betel nut slices, reduce information loss during multiple convolutions, and improve the accuracy of contour segmentation. The improved convolution block attention model is applied to strengthen the extraction of effective information from the inner contour, suppress invalid information, and improve the accuracy of nucleus and seedless judgment. At the same time, deep separable convolution is utilized to reduce high-resolution backbone network bands and the problem of the increasing amount of parameters and computational complexity. The experimental results show that the improved network segmentation accuracy is increased by 2.5% compared with the unimproved network, and the accuracy of the core and seedless judgment is increased by 5.93%, and the total floating-point calculation is reduced by 64.9%.

Keywords: Mask R-CNN; High-resolution network; Convolution block attention; Betel nut slice

[责任编辑：张岩芳]

(上接第 28 页)

Research on Forecasting Method of Sea Wind Speed Based on Improved SARIMA-LSTM

YU Congcong, XIONG Caiquan, XU Shiqiang, GU Xiaohui

(School of Computer Science, Hubei Univ. of Tech., Wuhan 430068, China)

Abstract: In order to improve the accuracy of offshore wind speed prediction, a model of sea wind speed prediction based on STL improved Seasonal Autoregressive Integrated Moving Average Model (SARIMA) and Long Short-term Memory (LSTM) neural network is proposed in this paper. First, the accuracy of seasonal difference step size based on SARIMA model through STL decomposition is improved and the SARIMA model is used to predict the observed wind speed series data to obtain the predicted value and the residual between the predicted value and the observed value; the residual sample is then used set to train the LSTM neural network and predict the residual; finally, the two predicted values are summed to obtain the predicted value of the wind speed series. Three different locations are selected for experiments and comparison with the method before the improvement. The experimental results show that the prediction accuracy of the improved method is higher, and the error is smaller.

Keywords: maritime wind speed; predictive analysis; LSTM; SARIMA

[责任编辑：张岩芳]