

基于多任务上下文增强的花椒检测模型

杨前¹,刘兴科¹,罗建桥¹,刘雪垠^{1,2},李柏林¹

(1. 西南交通大学 机械工程学院,四川 成都 610031;

2. 四川省机械研究设计院 研发中心,四川 成都 610063)

摘 要:花椒视觉采摘系统需要检测当前场景中的近景花椒。针对现有检测网络容易混淆相似外观的近景和远景花椒,难以准确检测近景花椒的问题,利用近景花椒与附近枝干的共生现象来区分近/远景花椒,提出基于多任务上下文增强的花椒检测模型。将枝干分割任务被加入目标检测网络中,用于引导全局特征同时表达花椒外观信息和枝干上下文信息;再为每个任务设计注意力模块调整全局特征,避免任务间的干扰。真实花椒数据实验结果表明:所提模型的精度相比 YOLOv4、RetinaNet 和 SSD 分别提高 12.28%、17.23% 和 30.17%。该模型能够大幅度减少远景花椒误检,准确检测近景花椒。

关键词:花椒采摘;目标检测;多任务网络;注意力机制

中图分类号:TP391.41 **文献标志码:**B **文章编号:**1671-5276(2023)01-0113-06

Sichuan Pepper Detection Model Based on Multi-task Context Reinforcement

YANG Qian¹, LIU Xingke¹, LUO Jianqiao¹, LIU Xueyin^{1,2}, LI Bailin¹

(1. School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China;

2. Department of R&D Center, Sichuan Machinery Research & Design Institute, Chengdu 610063, China)

Abstract: In Sichuan pepper visual picking system which needs to detect the close-up pepper in the current scene, the existing object detection network tends to confuse the similar appearances of close-range and long-range pepper, yielding to the difficulty in accurately detecting the close-range pepper. To tackle the problem, the symbiosis of the close-range pepper and the nearby branches and stems is used to distinguish the near/distant pepper, and Sichuan pepper detection model based on multi-task context reinforcement is proposed. The branch and trunk segmentation task are added to the object detection network to guide the global features expressing the appearance information of the peppercorns and the context information of the branches simultaneously. An attention module is designed for each task to adjust the global features to avoid interference between tasks. The experiments on real pepper data show that the accuracy of the proposed model is improved by 12.28%, 17.23% and 30.17% respectively compared with YOLOv4, RetinaNet and SSD, which can drastically reduce the false detection of pepper in the distant view and detect pepper in the close range with accuracy.

Keywords: pepper picking; object detection; multi-task network; attention model

0 引言

我国花椒产量居世界首位,解决花椒自动化采摘问题是推进农业现代化建设的迫切需求^[1]。机器视觉技术的发展,视觉采摘机器人在苹果、番茄等作物采摘中获得了成功应用^[2-3]。花椒视觉采摘系统包括机械臂和安装在机械臂上的相机。首先,相机拍摄当前采摘场景,由视觉算法检测花椒簇位置,如图 1 所示。然后,根据位置检测结果,机械手将花椒簇从枝干上采摘下来。考虑机械臂有限的运动空间以及场景中的复杂障碍物(例如枝条),合理的采摘规划为:采摘明显的近景花椒,忽略远景花椒。因此,视觉系统中花椒检测的目标为:在当前场景中,检出近景花椒簇,剔除远景花椒。



图 1 场景中的近景(实线)和远景花椒簇(虚线)

花椒检测任务可以由流行的目标检测网络完成,例如区域卷积神经网络(region convolutional neural network,

基金项目:四川省科技重点研发项目(2021YFN0020)

第一作者简介:杨前(1994—),男,河南固始人,硕士研究生,研究方向为图像处理和机器视觉。

R-CNN)^[4]、一阶段网络(you only look once, YOLO)^[5]等。目标检测网络一般由特征提取模块和检测模块组成,前者提取目标的外观特征,后者预测目标的位置、尺寸和类别信息。由于特征提取模块能够自动学习目标特征表达,目标检测网络在农作物检测任务中表现了出色性能^[6]。例如,王琪等^[2]提出了基于YOLO深度卷积神经网络的复杂背景下机器人采摘苹果定位;娄甜田等^[7]提出了基于深度卷积网络的葡萄簇检测与分割模型;刘芳等^[8]提出了基于改进型YOLO的复杂环境下番茄果实快速识别方法。对于花椒检测问题,特征提取模块可以学习花椒簇外观特征,克服姿态变化、局部遮挡、光照不均等因素的干扰。

然而,受远景花椒干扰,现有目标检测网络难以准确检测近景花椒。图1表明近/远景花椒外观相似,现有网络仅学习目标外观特征,容易误检出远景花椒。从图1可以发现,近/远景花椒的主要区别是近景花椒附着在明显的枝干上。鉴于近景花椒和枝干的共生现象,花椒检测模型不仅需要提取目标外观特征,而且应该考虑目标附近的枝干上下文信息,从而区分近/远景花椒。在目标检测领域,上下文信息主要指目标附近的邻域信息或目标之间的共生关系^[9-10]。现有基于上下文增强的目标检测方法主要分为3种:1)上下文窗口^[9],通过人工设置比目标区域更大的窗口来表达目标周围的上下文信息;2)特征金字塔^[11],通过融合不同层次的特征实现在更大感受野上表达目标附近邻域信息;3)目标关系推理^[12],通过定义推理网络表达目标之间的共生关系。上下文窗口和特征金字塔无法准确表达花椒与枝干共生关系,而上下文推理需要专门设计复杂的推理模块,通用性差。因此,现有基于上下文增强的目标检测方法难以有效解决近景花椒识别问题。

针对现有目标检测模型容易混淆近/远景花椒、难以准确检测近景花椒簇的问题,引入枝干分割任务学习枝干上下文信息,本文提出基于多任务上下文增强的花椒检测模型,模型框架如图2所示。首先,增加与检测模块平行的枝干分割模块,花椒检测任务和枝干分割任务共同引导特征提取模块学习全局特征。全局特征同时包含花椒外观特征和枝干上下文信息,达到区分近/远景花椒的目的。然后,构造任务注意力模块来调整全局特征,让不同任务更加灵活地共享全局特征,避免任务之间相互干扰。所提模型与现有相关方法的差异体现在两个方面。

1)不同于现有基于上下文增强的目标检测方法。上下文窗口或特征金字塔仅关注目标本身的邻域信息,忽略了目标间的关系。目标关系推理通过专用网络表达目标关系。相比而言,本文所提方法引入枝干分割任务学习关于目标的上下文信息,既利用了近景花椒与枝干的共生关系,又避免了复杂的推理结构。

2)不同于现有的检测-分割多任务模型^[13]。现有检测-分割多任务模型通过分割目标本身轮廓来增强目标特征表达,无法表达目标间的上下文信息。相比而言,本文所提方法通过分割目标附近的枝干可学习近景花椒的上下文信息。

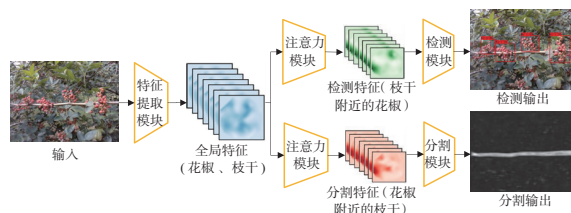


图2 基于多任务上下文增强的花椒检测模型

1 所提方法

1.1 多任务网络

多任务网络通常由一个特征提取模块和若干个相关联的平行任务模块构成,所有相关联任务共享特征提取模块^[14]。多个模块共同引导特征学习过程能够防止特征过拟合到特定任务,从而提高特征泛化能力。本文的近景花椒附着在明显的枝干上,近景花椒与枝干之间有着很强的上下文关联信息。通过引入有关联的枝干分割任务来提升花椒检测任务的能力。花椒检测与枝干分割组成的多任务网络主要有以下几个优点^[15]。

1)隐式数据增广,枝干分割任务带来的额外信息相当于增加了样本大小。特别是训练中存在噪声时,相当于做了数据增强。花椒检测任务和枝干分割任务都有各自的噪声,只学习花椒检测任务可能会使得模型容易过拟合花椒检测任务的噪声,而加入枝干分割任务后,两个任务同时学习通过平均噪声来得到更好的模型泛化能力。

2)归纳偏置和特征关联,花椒检测和枝干分割任务嵌入在相关任务的上下文环境中。在这样的环境中,学习两个任务的归纳偏置可以通过在不同任务上抽样来完成,减轻了单个任务上样本采样的负担。同时,多任务网络可以搜索到所有含花椒和枝干的特征,其中也包含花椒与枝干的关联特征。

3)提高动态优化特性,多任务之间在共同学习过程中可以改变权值更新的动态特性,可能使网络更适合多任务学习。单任务学习时,梯度的反向传播倾向于陷入局部最小值,往往单任务单独学习时有多个局部最小值,通过多个任务交互作用,可以帮助单个任务学习过程规避鞍点和局部最小点。

4)窃听机制,对于枝干分割任务来说很容易学习到枝干相关的特征,而这些特征对于花椒检测任务来说很难学到。这是因为花椒检测任务更倾向学习花椒相关的特征而阻碍了枝干相关特征的学习,通过多任务学习,允许模型窃听,即花椒检测任务使用枝干分割来学习枝干相关的特征。

花椒检测与枝干分割组成的多任务网络结构图如图3所示。花椒检测任务采用主流的YOLOv4目标检测网络^[16],对应图3中的特征提取模块加上检测模块。特征提取模块:神经网络中的特征分为低维特征与高维特征。低维特征通常是指图像中的一些小的细节信息,例如边缘、角点、颜色、纹理等;高维特征是建立在低维特征之上的,将低维特征高维抽象化、语义化,用于图像中物体检

测和形状识别。通常神经网络中都会使用这两种类型的特征,神经网络的前几层学习低维特征,后几层学习高维特征。检测模块:提取更深的语义信息(位置、尺寸和类别信息)。利用“金字塔”结构,在 52×52 、 26×26 、 13×13 的3个尺度下提取花椒的位置、尺寸和类别信息,这3个特征图相当于将原始图片划分成 52×52 、 26×26 、 13×13 的网格。检测原理为如果物体的中心落在网格中,那么该网格就负责预测这个物体的位置、尺寸和类别,如图4所示,其中每个网络的输出值 y 为 $(p_c, b_x, b_y, b_h, b_w)$, p_c 为网格内有目标物的概率,即置信度; b_x 为预测框的 x 坐标; b_y 为预测框的 y 坐标; b_h 为预测框的高; b_w 为预测框的宽。每个网格预测3组值,故检测网络的输出维度为 $3 \times (p_c, b_x, b_y, b_h, b_w) = 15$ 。通过置信度阈值和非极大抑制得到最终的预测框。

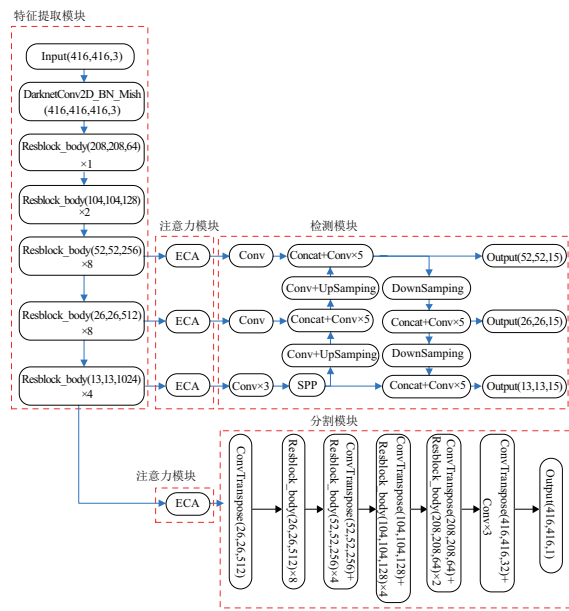


图3 多任务网络结构图

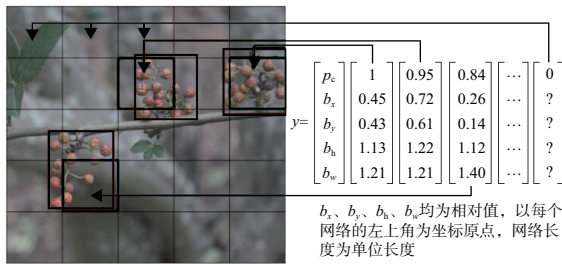


图4 花椒检测网络输出原理图

本文的检测任务之所以采用YOLOv4,因为在复杂的自然环境中花椒的形状和大小各异,而它的特征提取模块在相同的内存消耗和计算复杂度时,具有更强的学习能力,能够适应更加复杂的环境。再加上它的检测模块采用特征金字塔结构,该结构有3个输出,对应在3个尺度上分别输出小、中、大花椒的尺寸与位置,能够适应花椒大小尺度的变化。

枝干分割任务采用经典的“编码器-解码器”结构,对

应图3中的特征提取模块-分割模块。枝干分割属于语义分割(也叫像素级分类,哪些像素点是枝干),文献[17]和文献[18]提出了基于全卷积神经网络的编码-解码思想。这种“编码器-解码器”结构在图像语义分割研究中获得了广泛关注,极大地促进了语义分割的发展。如图5所示[18],语义分割中的“编码器-解码器”本质上为卷积运算,因其卷积过程特征图尺寸先逐渐减小、通道数逐渐增加后,尺寸逐渐增大、通道数逐渐减少,输出与输入保持相同的尺寸。因此得名“编码器-解码器”。

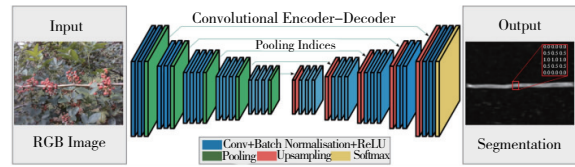


图5 语义分割中的编码-解码结构

本文中的编码器(特征提取模块)用于提取枝干的特征,对输入图片进行多次的卷积运算,图片的尺寸逐渐减小而通道数逐渐增加(由原来的 $416 \times 416 \times 3 \rightarrow 13 \times 13 \times 1024$),将低维的枝干特征进行高维语义化。解码器(分割模块)收集这些语义信息,通过上采样的形式“解码”,特征图的通道数减小而尺寸增加($13 \times 13 \times 1024 \rightarrow 416 \times 416 \times 1$),最终将枝干区域对应到原图相应的像素点上。解码器输出的是一张与输入图片尺寸一致,通道为1的特征图,同花椒检测任务类似,预测每个像素是枝干的概率,像素值也为概率值,最终将概率大于0.5的像素可视化得到枝干区域。

由于花椒检测任务和枝干分割任务共享特征提取模块,因此特征提取模块提取的全局特征包含花椒和枝干的语义信息。模型训练阶段,全局特征被两个任务同时引导来表达花椒外观信息和近景花椒附近的枝干上下文信息,实现了抑制远景花椒的目的。

1.2 任务注意力模块

全局特征虽然同时包含花椒外观和枝干信息,增强了区分近/远景花椒的能力,但是容易导致任务之间相互干扰,影响每个任务的学习能力。例如,对于花椒检测任务,全局特征中的枝干信息(非重要的部分)将影响花椒检测效果。因此,直接将全局特征送入检测或分割模块不利于任务间协同,需要针对不同任务调整全局特征。在特征调整方面,最近出现的通道注意力机制[19-21]利用相关特征图来学习权重分布,再用学出来的权重施加在原特征图之上最后进行加权求和。通过对卷积特征不同通道进行加权来改善特征表达,能够增强特征中的有效信息,抑制噪声。本文利用注意力机制学习特征图权重分布的特性,分别在检测和分割模块前加入注意力模块,调整输入到花椒检测任务和枝干分割任务的特征分布,两个任务按需筛选适合的特征,实现了从特征注意力到任务注意力的转变。

为调整全局特征,分别在检测和分割模块前加入任务注意力模块,调整全局特征形成更加关注特定任务的特征表达。具体来说,注意力模块采用高效通道注意力(efficient channel attention, ECA)网络[21],如图6所示。

相比其他注意力机制^[19-20],ECA 不仅通过去除降维过程增强了对有效通道的关注能力,而且采用一维卷积代替全连接提高了计算效率。ECA 的计算过程如下:

首先,采用全局平均池化(global average pooling, GAP)将全局特征 $X_0 \in \mathbf{R}^{C \times H \times W}$ 转化为一维向量 $X_1 \in \mathbf{R}^C$, C 、 H 、 W 分别为全局特征的通道数、高度、宽度。式(1)定义了上述 GAP 操作:

$$X_{1,c} = \text{GAP}(X_0(c, :, :)) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W X_0(c, i, j) \quad (1)$$

然后,采用长度为 k 的一维卷积核 X_1 进行卷积操作,经过激活函数处理后获得通道权值 $w \in \mathbf{R}^C$:

$$w = \delta(\text{Conv1D}_k(X_1)) \quad (2)$$

$$k = \text{odd}\left(\frac{\log_2 C + b}{\gamma}\right) \quad (3)$$

式中: $\text{Conv1D}_k(\cdot)$ 表示一维卷积操作; $\delta(\cdot)$ 表示激活函数,采用 Sigmoid 激活函数;卷积核长度 k 根据超参数 γ 和 b 计算; $\text{odd}(\cdot)$ 表示取最靠近的奇数。

最后,由通道权值调整全局特征获得特定的任务特征 $X_t \in \mathbf{R}^{C \times H \times W}$:

$$X_t(c, :, :) = w_c X_0(c, :, :) \quad (4)$$

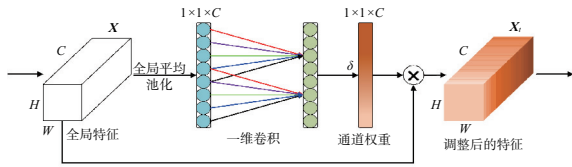


图6 任务注意力模块(ECA网络)

所提模型共包含4个结构相同的注意力模块,记为A1-A4。由图3可知,检测模块的输入为3个不同尺度的特征图,A1-A3分别用于调整这3个尺度的特征,形成针对检测任务的特征;分割模块只有1个输入,A4用于调整特征提取模块最后输出的特征,形成针对分割任务的特征。

在训练集标注信息的引导下,注意力模块能够自动学习全局特征中不同通道的权重,改善检测和分割任务对所需特征的表达能力。对于花椒检测任务 t ,权重 w_c 较大,表明通道 c 储存的特征对花椒检测任务的贡献强,全局特征 X_t 的第 c 维被增强。相比一般多任务网络中各任务直接共享全局特征^[14-15],注意力模块实现了更加灵活的特征共享,能够避免不同任务间的干扰,提高每个任务的学习能力。

1.3 损失函数

损失函数(loss function)是用来估量模型预测值与真实值不一致的程度,它是一个非负实值函数,模型训练是损失函数最小化的过程,损失函数越小,模型的鲁棒性就越好。

由图4可知,花椒检测任务的损失函数包括目标置信度损失 l_{obj} 和预测框回归损失 l_{box} 。目标置信度损失 l_{obj} 采用交叉熵,分为有目标和无目标,表达式如下:

$$l_{\text{obj}} = \sum_{i=0}^{S^2} \sum_{j=0}^B I_{i,j}^{\text{obj}} [\hat{C}_i \ln(C_i) + (1 - \hat{C}_i) \ln(1 - C_i)] + \sum_{i=0}^{S^2} \sum_{j=0}^B I_{i,j}^{\text{noobj}} [\hat{C}_i \ln(C_i) + (1 - \hat{C}_i) \ln(1 - C_i)] \quad (5)$$

式中: S^2 表示网格大小; B 代表每个网格产生的候选框(3个); $I_{i,j}^{\text{obj}}$ 表示如果在 i,j 处的网格内有目标,其值为1,否则为0; $I_{i,j}^{\text{noobj}}$ 表示如果 i,j 处的网格内没有目标,其值为1,否则为0; $\hat{C}_i$ 表示预测的置信度; C_i 表示真实的置信度。

预测框回归损失 l_{box} 采用C-IoU(complete-intersection over union)损失,C-IoU损失不仅考虑了预测框与真实框的重合度,还增加了中心距离和宽高比的惩罚项。惩罚项的作用是控制预测框能够尽可能快速地与真实框接近。若真实框和预测框的中心距离越远,宽高相差越大,相应的惩罚值越大(图7)。其数学表达式如式(6)、式(7)、式(8)所示。

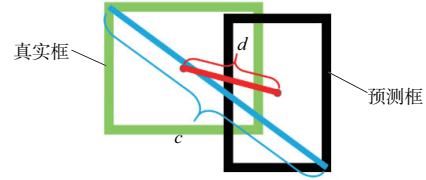


图7 基于C-IoU的预测框回归损失示意图

$$l_{\text{box}} = 1 - I_{\text{ou}} + \frac{d^2}{c^2} + \alpha v \quad (6)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2 \quad (7)$$

$$\alpha = \frac{v}{(1 - I_{\text{ou}}) + v} \quad (8)$$

式中: I_{ou} 为预测框与真实框面积的交集与并集比值; d 为预测框与真实框中心点的距离; c 为预测框与真实框对角线距离; w^{gt} 和 h^{gt} 为真实框的宽和高; w 和 h 为预测框的宽和高; v 为衡量长宽比一致性; α 为长宽比一致性的权重。

由上文介绍的枝干分割任务就是将枝干区域对应到原图相应的像素点上,即枝干区域的像素值为1,其他区域像素值为0。类别数为2个,即枝干区域和其他区域;枝干分割任务采用广义相似度损失函数^[22],表达式如下:

$$l_{\text{GDL}} = 1 - 2 \frac{\sum_{l=1}^2 \omega_l \sum_n r_{ln} p_{ln}}{\sum_{l=1}^2 \omega_l \sum_n (r_{ln} + p_{ln})} \quad (9)$$

$$\omega_l = \frac{1}{\left(\sum_{n=1}^N r_{ln} \right)^2} \quad (10)$$

式中: r_{ln} 为类别 l 在第 n 个像素的真实值; p_{ln} 为对应的预测值; ω_l 为每个类别的权重。

所提模型总的损失函数为

$$l_{\text{oss}} = l_{\text{box}} + l_{\text{obj}} + l_{\text{GDL}} \quad (11)$$

2 实验与结果分析

2.1 实验设置

样本图像来自成都市金堂县某花椒种植园,采用英特尔D435i深度相机,并将相机固定在机械手末端,距离最近花椒簇200 mm~500 mm范围内拍摄,图像像素为640×480,格式为24位PNG。选择1 800张图像构成数据集,对所有图像进行人工标注,采用“Labellmg”标注工具标注

近景花椒簇真实框,采用“Labelme”标注工具对应的枝干区域进行蒙版标注。训练集、验证集和测试集的划分比例为 7:2:1。

训练阶段,采用随机梯度下降优化器对网络进行优化,动量设置为 0.9,衰变值为 0.000 5,批尺寸为 64,迭代轮次为 1 000,初始学习率为 0.001。模型由 Python3.7 结合深度学习库 PyTorch1.7 实现,在配备 GeForce RTX 3090 的服务器上运行。

选择精度 (precision)、召回率 (recall) 和平均准确率 (average precision, AP) 等指标评价花椒检测性能。当模型输出的预测框与人工标注的真实花椒簇之间的交并比大于 0.5 时,认为花椒检测结果正确。模型检测结果分为真阳性 (true positive, TP)、假阳性 (false positive, FP)、真阴性 (true negative, TN)、假阴性 (false negative, FN) 4 种情况,如表 1 所示。例如 TP 表示检测为近景花椒且真实为近景花椒的检测框数量。

表 1 模型不同检测结果的定义

真实情况	检测为花椒簇	预测为背景
花椒簇	TP	FN
背景	FP	TN

精度表示检测结果中真实近景花椒簇的占比,反映了模型区分近/远景花椒的能力,计算公式如式 (12) 所示。

$$P_{\text{recision}} = \frac{T_p}{T_p + F_p} \quad (12)$$

召回率表示真实近景花椒簇被检出的比例,反映了模型识别目标的能力,计算公式如式 (13) 所示。

$$R_{\text{ecall}} = \frac{T_p}{T_p + F_N} \quad (13)$$

平均精度表示 Precision-Recall 曲线的面积,是对模型性能的综合评价,计算公式如式 (14) 所示, p 代表 Precision, r 代表 Recall。

$$A_p = \int_0^1 p(r) dr \quad (14)$$

2.2 实验一:花椒检测性能

为测试所提模型的花椒检测性能,设置下列对比方法:

1) 现有目标检测网络,分别采用近年来流行的目标检测网络进行花椒检测,包括 SSD 网络^[23]、RetinaNet 网络^[24]以及 YOLOv4 网络;2) 多任务网络,去除所提模型中的任务注意力模块,仅在 YOLOv4 基础上增加与检测模块平行的枝干分割模块;3) 多任务+注意力,即所提模型。实验结果如表 2 所示。

表 2 不同方法的花椒检测性能

方法	P_{recision}	R_{ecall}	A_p
SSD	0.727 0	0.815 6	0.743 2
RetinaNet	0.807 2	0.833 2	0.819 1
YOLOv4	0.842 8	0.834 6	0.838 5
多任务网络	0.917 4	0.886 9	0.893 8
多任务+注意力	0.946 3	0.881 4	0.915 9

从表 2 中可以观察到以下 3 点现象。

1) YOLOv4 的花椒检测性能优于 SSD 和 RetinaNet,说明所提模型选择 YOLOv4 作为基础的目标检测网络是合理的。

2) 相比 YOLOv4,多任务网络的性能大幅提高。花椒检测精度从 0.842 8 提高到 0.917 4,意味着检测结果中误检的近景花椒大幅减少。精度提高的原因是枝干分割任务引导模型学习近景花椒附近明显的枝干信息,增强了全局区分近/远景花椒的能力,能够防止误检出近景花椒。实验结果说明,引入枝干分割任务能够增强花椒的上下文表达,提高检测性能。

3) 相比多任务网络。所提模型进一步提高了精度 (0.917 4、0.946 3) 和 AP 值 (0.893 8、0.915 9),召回率基本不变。对于花椒检测任务,注意力模块是能够调整全局特征,抑制全局特征中枝干信息的,增强花椒信息的表达。实验结果说明,在多任务网络基础上增加任务注意力机制能够促进任务协同,进一步提高花椒检测性能。

2.3 实验二:枝干分割任务的作用

为分析多任务网络增强目标上下文信息的作用,对比 YOLOv4 和多任务网络中特征提取模块提取的全局特征。特征提取模块最终输出一个 1 024 维的特征图,图 8 展示了 1 024 维中的前 8 维特征图以及原始输入图像。对比图 8(a) 和图 8(b) 发现,多任务网络提取的全局特征在近景花椒和枝干对应区域具有较高响应,远景花椒区域响应较低。实验结果说明,增加任务能够学习近景花椒附近的枝干上下文信息,增强特征区分近景和远景花椒的能力。

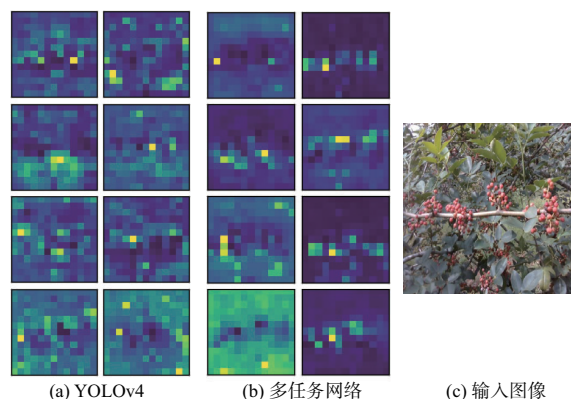


图 8 YOLOv4 和多任务网络提取的全局特征

进一步地,图 9 对比了 YOLOv4 和多任务网络的花椒检测效果。相比 YOLOv4 模型,多任务网络剔除了 YOLOv4 误检的远景花椒簇,补全了枝干附近 YOLOv4 漏检的近景花椒簇。实验结果说明,多任务网络能够利用枝干上下文信息提高检测近景花椒簇的能力。

2.4 实验三:任务注意力模块的作用

为分析任务注意力模块调整全局特征的作用,对比不同任务注意力模块输出的通道权重。所提模型包括 4 个注意力模块 A1-A4:用于检测任务的 A1-A3 和用于分割任务的 A4。将图 1 输入到所提模型,图 10 展示了 A1-A4 输出的通道权值向量 $w_1 \sim w_4$,图 11 展示了每个权值向量

中最大权值对应通道的特征。



图9 YOLOv4 和多任务网络的花椒检测效果

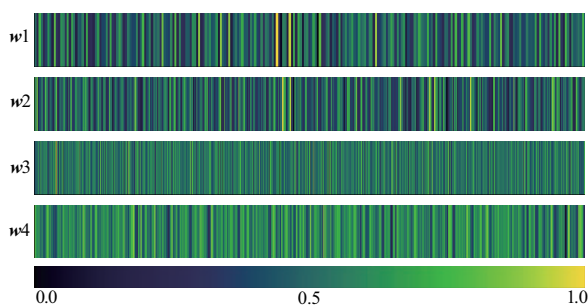


图10 注意力模块中输出的权值向量

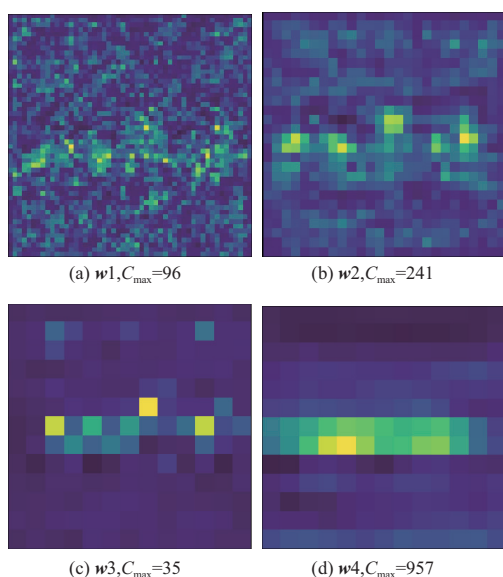


图11 最大权值对应通道的特征

图10显示,不同任务注意力模块输出的通道权重具

有不同的分布。这是因为全局特征同时包含花椒外观信息和枝干信息,不同的任务需要关注全局特征中不同信息成分。图11显示,在用于检测任务的注意力权重 $w1 \sim w3$ 中,最大权值对应的通道特征包含了明显花椒信息,这些通道上的特征被增强后送入检测模块。在用于分割任务的注意力权重 $w4$ 中,最大权值对应的通道特征有着明显的枝干信息,这些通道上的特征被增强后送入分割模块。实验结果说明,任务注意力模型能够调整全局特征,增强特征中的有效信息,促进任务间协同。

3 结语

针对目标检测网络难以检测近景花椒的问题,提出基于多任务上下文增强的花椒检测模型。首先,向 YOLOv4 目标检测网络中加入枝干分割模块,形成检测-分割多任务网络。然后,设计任务注意力模块调整全局特征,增强特征对特定任务的表达能力。实验表明:加入分割任务能够捕捉近景花椒附近的枝干上下文信息,提高区分近/远景花椒的能力;任务注意力模块能够增加特征共享灵活性,促进多任务相互协同。下一步研究方向是结合花椒检测和枝干分割的结果估计采摘点。

参考文献:

- [1] 刘安成,尉倩,崔新爱,等. 花椒采收现状及研究进展[J]. 中国农机化学报,2019,40(3):84-87.
- [2] 王琪,丁柏文,陈萍. 苹果采摘机器人控制系统设计[J]. 机械制造与自动化,2019,48(5):150-152,161.
- [3] 郭凯敏,崔天时,张桢,等. 西红柿采摘机器人视觉系统的研究[J]. 农机化研究,2016,38(12):217-221,226.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2016,38(1):142-158.
- [5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE,2016:779-788.
- [6] 黄豪杰,段先华,黄欣辰. 基于深度学习水果检测的研究与改进[J]. 计算机工程与应用,2020,56(3):127-133.
- [7] 娄甜田,杨华,胡志伟. 基于深度卷积网络的葡萄簇检测与分割[J]. 山西农业大学学报(自然科学版),2020,40(5):109-119.
- [8] 刘芳,刘玉坤,林森,等. 基于改进型YOLO的复杂环境下番茄果实快速识别方法[J]. 农业机械学报,2020,51(6):229-237.
- [9] LENG J X, REN Y H, JIANG W, et al. Realize your surroundings: exploiting context information for small object detection[J]. Neurocomputing,2021,433:287-299.
- [10] QIU H Q, LI H L, WU Q B, et al. Hierarchical context features embedding for object detection[J]. IEEE Transactions on Multimedia,2020,22(12):3039-3050.
- [11] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE,2017:936-944.

(下转第149页)

3.4 试验验证

制造 60、70、80 邵尔硬度的丁腈橡胶双面阀板,装配到双塔干燥器换向阀进行充排气保压试验,试验结果如表 4 所示。试验表明:不同硬度丁腈橡胶开始泄漏压力与仿

真计算的结果接近,验证了密封仿真计算的正确性,验证了这类产品的密封机理主要靠橡胶与阀口接触形变产生的接触应力来密封,并且由于上下阀口限位作用,这类产品属于定应变条件的密封,橡胶硬度增加有利于密封。

表 4 不同硬度丁腈橡胶双面阀板密封保压试验结果

单位:kPa

充气 压力	60 邵尔硬度			70 邵尔硬度			80 邵尔硬度		
	开始压力	结束压力	漏气量	开始压力	结束压力	漏气量	开始压力	结束压力	漏气量
300	303.8	302.4	-1.4	304.5	302.8	-1.7	295.6	294.1	-1.5
600	600.0	557.2	-42.8	605.6	599.3	-6.3	599.7	598.5	-1.2
900	900.2	568.7	-331.5	901.7	801.6	-100.1	900.3	898.1	-2.2

4 结语

通过有限元方法研究了橡胶材料硬度、橡胶厚度以及橡胶与金属黏接对双塔干燥器换向阀密封性能的影响。发现橡胶材料硬度对于这类高工作气压的气动密封件至关重要。同时该类金属与橡胶黏接的密封产品,其橡胶黏接强度设计要 ≥ 3 MPa,才能避免工作时高气压对密封件橡胶材料造成撕裂破坏,导致产品失效。

参考文献:

- [1] 张成彦,谭跃进,李金禄,等. 吸附式压缩空气干燥器用组合换向阀[J]. 中国设备工程,2018(5):174-175.

- [2] 张付英,李腾. 基于 ANSYS 的 Y 形圈初始建模[J]. 天津工业大学学报,2013,32(6):85-88.
 [3] RIVIN E I,LEE B S. Experimental study of load-deflection and creep characteristics of compressed rubber components for vibration control devices[J]. Journal of Mechanical Design, 1994,116(2):539-549.
 [4] 杨晓翔. 非线性橡胶材料的有限单元法[M]. 北京:石油工业出版社,1999
 [5] 刘博. 双塔式空气干燥器故障频发分析及改进研究[J]. 华东科技(综合),2019(5):299.

收稿日期:2022-04-08

(上接第 118 页)

- [12] 刘大有,刘春辰,王生生. 环境智能中上下文推理方法研究综述[J]. 模式识别与人工智能,2011,24(5):673-679.
 [13] BOLYA D,ZHOU C,XIAO F Y, et al. YOLACT better real-time instance segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2022,44(2):1108-1121.
 [14] 邵蔚元,郭跃飞. 多任务学习及卷积神经网络在人脸识别中的应用[J]. 计算机工程与应用,2016,52(13):32-37,88.
 [15] 徐培超,陈雯柏,陈祥凤,等. 多任务及 Resnet 网络在人脸多属性识别中的应用[J]. 小型微型计算机系统,2018,39(12):2720-2724.
 [16] BOCHKOVSKIY A,WANG C Y,LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [D/OL]. (2020-05-20)[2021-08-10] https://blog.51cto.com/u_15228753/2819798.
 [17] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
 [18] BADRINARAYANAN V,KENDALL A,CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(12):2481-2495.
 [19] 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.
 [20] WOO S,PARK J,LEE J Y, et al. CBAM:convolutional block attention module [C]//Computer Vision - ECCV 2018, [S.I.: s.n.],2018.
 [21] 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). Seattle, WA, USA: IEEE,2020:11531-11539.
 [22] SUDRE C H,LI W Q,VERCAUTEREN T, et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations [C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, [S.I.:s.n.],2017.
 [23] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [C]//Computer Vision - ECCV 2016, [S.I.:s.n.], 2016.
 [24] LIN T Y,GOYAL P,GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(2):318-327.

收稿日期:2021-09-13