

DOI: 10.19344/j.cnki.issn1671-5276.2024.06.042

基于跨模态特征融合的 RGB-D 花椒图像显著性检测

李节, 孙成龙, 王逸涵, 杨前, 李柏林

(西南交通大学 机械工程学院, 四川 成都 610031)

摘要:针对现有显著性检测模型无法有效地协同花椒枝干彩色图像和深度图像特征, 建立基于注意力的 RGB-D 图像花椒枝干显著性检测模型。由两个单流卷积网络分别提取彩色和深度图像特征; 设计基于空间和通道注意力机制的跨模态融合模块, 用于融合多尺度的彩色流和深度流特征; 研发多尺度监督机制, 用于缓解由于采用最近邻域上采样的解码方式导致边缘预测不准确的问题。实验结果表明: 该方法的平均精确度、平均召回率、综合评价指标和平均绝对误差均优于对比显著性目标检测方法。

关键词:花椒自动化采摘; 图像处理; RGB-D 显著性目标检测; 跨模态融合; 注意力机制; 多尺寸监督

中图分类号: TP391.41 **文献标志码:** A **文章编号:** 1671-5276(2024)06-0211-07

RGB-D Pepper Image Saliency Detection Based on Cross-modal Feature Fusion

LI Jie, SUN Chenglong, WANG Yihan, YANG Qian, LI Bailin

(School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: To address the inability of existing saliency detection models to utilize the features of pepper branch color images and depth images effectively, an attention-based RGB-D image pepper branch saliency detection model is proposed. Color and depth image features are extracted separately by two single-stream convolutional networks. A cross-modal fusion module based on spatial and channel attention mechanisms is designed to fuse multi-scale color stream and depth stream features. A multi-scale supervision mechanism is developed to alleviate the inaccurate edge prediction caused by the use of nearest-neighbor upsampling decoding. Experimental results show that the average accuracy, average recall rate, comprehensive evaluation index and average absolute error of the proposed method are all superior to the compared salient object detection methods.

Keywords: automated pepper harvesting; picture processing; RGB-D significance target detection; cross-mode fusion; attention mechanism; multi-dimension supervision

0 引言

花椒是四川省重要经济作物, 提升花椒采摘的自动化水平对于我国西部乡村振兴具有重要意义。得益于近些年来计算机软硬件的发展, 基于视觉的采摘机器人被广泛地应用于苹果、柑橘和葡萄的自动化采摘。作为智能采摘机器人^[1]的重要组成部分, 视觉系统通常被设计用于识别并定位果实位置, 从而引导机械部分完成采摘。不同于苹果、柑橘等的采摘, 簇状花椒的采摘点无法直接被观测到, 而采摘点的估计需要利用枝干和花椒簇的交点来确定。

如图 1 所示, 考虑到机械臂的运动空间及复杂采摘场景中的干扰物(枝条、叶子等), 合理的花椒采摘规划应为采摘明显的近景花椒, 忽略远景花椒。由于近景花椒一般位于前景中较粗的主

枝干上, 因此前景主枝干的提取是花椒采摘点估计的重要前提。



图 1 户外花椒图像及深度图像

花椒前景的主枝干提取任务是一种显著性目标检测^[2](salient object detection, SOD)问题, 旨在实现图像场景中感兴趣区域的快速提取并过滤背景噪声的干扰。文献[3]提出了一种 RGB-SOD 算法用于农田中的昆虫检测并取得了良好的效果, 但易受到复杂环境的干扰, 无法有效用于具有低对比度、相似前景与背景、复杂背景等特点的

基金项目:四川省科技计划重点研发项目(2021YFN0020)

第一作者简介:李节(1997—), 男, 四川达州人, 硕士研究生, 研究方向为图像处理、机器视觉, lijie295195@163.com。

花椒主枝干提取。为了实现复杂农业场景中 SOD,文献[4]提出了双流主干网络用于同时提取柑橘图像的彩色和深度特征,提供具有鲁棒性的显著性线索。文献[5]以跳层结构为基础提取跨模态间的多层次互补信息。为了更好地获取跨模态间的互补信息用于显著性推理,文献[6]提出了一种流体金字塔结构用于引导深度图像和彩色图像的信息融合。复杂的农作环境中采集到的彩色图像和深度图像中跨模态信息往往是非耦合的(图1),采集到的花椒深度图像中还包含了与前景相似的噪声,而现有的一些研究表明线性的跨模态融合方法无法有效地抑制相似噪声的干扰,从而影响最终的识别结果。为了抑制深度图像中的噪声对显著推理造成的影响,文献[7]利用边缘一致性、区域不确定性和模型方差来评估深度图质量,并以此指导深度图与彩色图的选择性融合。然而,该方法依赖手动设计的质量评价标准,无法应对花椒采摘场景中的各种复杂背景的干扰。

针对相似前景与背景、复杂背景中的花椒枝干提取问题,本文提出基于注意力机制^[8]和多尺度监督^[9]的花椒主枝干显著性检测模型。模型的编码器采用主流的双分支主干网络来提取彩色图像特征和深度图像特征。此外,在多个尺度上,本文提出特征融合增强模块(feature fusion enhancement module, FFEM)并将其嵌入到两个分支网络中,利用空间与通道注意力模式来实现跨模态特征的判别融合。在多尺度分割标签的监督下,特征融合增强模块能够自动学习空间权值图和通道权值向量。本文方法如下。

1) 针对复杂的农作环境中花椒主枝干的检测

问题,提出一种基于跨模态特征融合的 RGB-D 花椒枝干图像显著性检测模型。

2) 所提模型在多个编码层级采用通道权值向量调整彩色和深度图像的串联特征,并计算调整后特征的空间权值图,提取主枝干的形状特征并抑制背景噪声。

3) 采用多尺度监督的方式来缓解上采样过程中的边缘信息丢失,提高模型对主枝干的分割效果。

1 所提模型

1.1 网络结构

花椒主枝干的精确提取是预测花椒采摘点的重要前提。为了在复杂农作环境下完成花椒主枝干的精确提取,本文提出了一种跨模态的枝干显著性检测模型,如图2所示。模型采用花椒彩色和深度图像两种输入信息,由两个单流卷积网络获取多尺度的编码特征。单流卷积网络采用类似 Unet 编码器架构,通过连续地组合卷积编码层(包括卷积层、批归一化层、线性整流层和最大池化层)来编码彩色图像和深度图像特征。为了有效地利用彩色和深度编码特征,提出特征融合增强模块用于实现跨模态特征的判别融合并剔除特征中相似背景的噪声干扰。该模块借鉴了注意力机制聚焦于感兴趣区域的特性,通过在线性融合过程中嵌入非线性注意力单元来改善融合后特征的显著性表达。在上采样过程中,一个反向的解码网络被用于解码图像特征,非线性注意力单元被嵌入到网络的每一层级来进一步精炼特征表示。最终,通过多级监督的方式,显著性推理模块完成最终的预测输出。

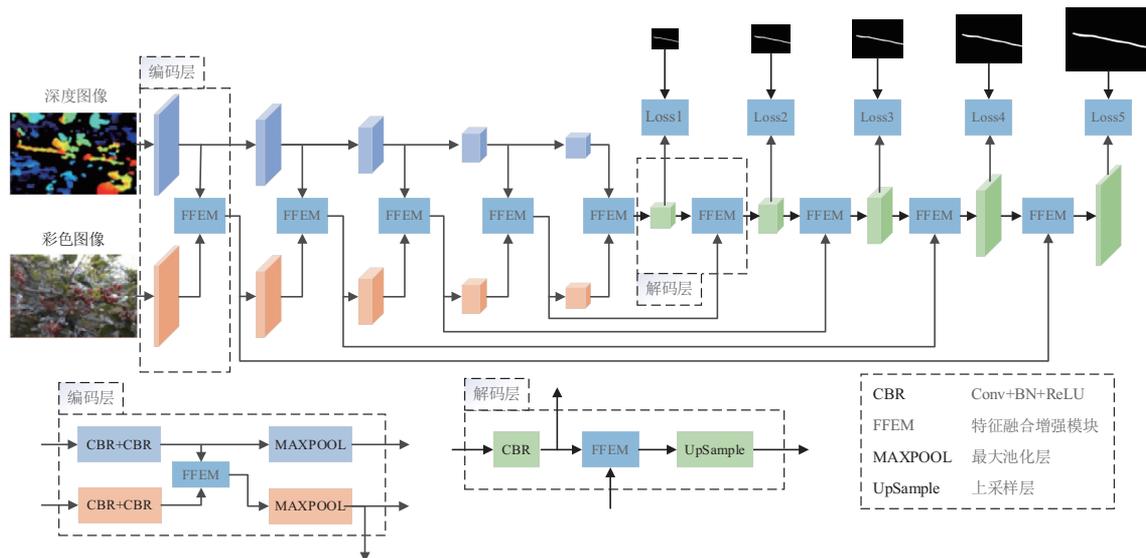


图2 所提模型网络结构

1.2 基于注意力的跨模态融合模块

彩色花椒图像中主枝干的提取会受到相似前景枝干的干扰。因此,为准确地区分主枝干还需要深度图像提供额外的显著性线索。然而,复杂的农作环境易导致深度图的深度线索缺失,使得深度图像中枝干与附近叶子、花椒等对象融为一体。从低质量的深度图像中分辨出目标枝干仍然需要借助颜色、纹理等外观信息。因此,本文设计了跨模态融合模块来同时提取彩色模态和深度模态中包含的与主枝干相关的编码信息。为了减少彩色模态和深度模态中与显著性目标相似的背景信息对主枝干提取的干扰,本文在融合模块中嵌入注意力机制来精炼融合后的编码特征。

单一层级跨模态融合模块的结构如图 3 所示。该模块首先接收来自同层级的彩色模态和深度模态特征 $X_i^{RGB} \in R^{C_i \times (H/2^i) \times (W/2^i)}$ 和 $X_i^{DEP} \in R^{C_i \times (H/2^i) \times (W/2^i)}$,其中参数 C, H, W 和 i 分别表示编码特征的通道数量、尺度和层级系数, R 表示实数空间。

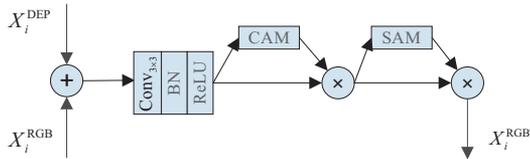


图 3 特征融合增强模块

针对双模态的特征,首先采用拼接操作聚合跨模态特征,并采用卷积操作对聚合特征进行非线性映射:

$$F_{\text{fusion}} = \text{ReLU}(\text{BN}(\text{Conv}_{3 \times 3}(\text{Concat}(X_i^{RGB}, X_i^{DEP})))) \quad (1)$$

式中: $\text{Conv}_{3 \times 3}$ 表示采用 3×3 尺寸的卷积核进行步长为 1 的标准卷积操作; $\text{BN}(\cdot)$ 和 $\text{ReLU}(\cdot)$ 分别代表批归一化和线性整流操作。

对于聚合后的跨模态特征 F_{fusion} ,分别采用通道注意模块和空间激活模块来计算该特征的通道权值图和空间权值图。最终,融合模块的编码输出将表示为输入编码在权值图上的加权映射。具体的计算过程如下:

$$F_{\text{CAM}} = C_{\text{CAM}}(F_{\text{fusion}}) \quad (2)$$

$$F = F_{\text{CAM}} \otimes F_{\text{fusion}} \quad (3)$$

式中: C_{CAM} 表示通道注意力模块; \otimes 表示对应元素相乘;通道权值图 $F_{\text{CAM}} \in R^{C_i \times (H/2^i) \times (W/2^i)}$ 。

$$F_{\text{SAM}} = S_{\text{SAM}}(F) \quad (4)$$

$$X_i^{RGB'} = F_{\text{SAM}} \otimes F \quad (5)$$

式中: S_{SAM} 表示空间注意力模块;空间权值图 $F_{\text{SAM}} \in R^{C_i \times (H/2^i) \times (W/2^i)}$;模块最终输出的编码特征 $X_i^{RGB'} \in R^{C_i \times (H/2^i) \times (W/2^i)}$ 。

通道注意力通过自适应地计算输入特征通道权值图来为判别力强的重要通道赋予较高权值,从高维冗余的特征图中选择对显著性表达更加有利的特征表示。空间注意力机制通过自适应计算来增强显著性区域的特征表示。由于显著性区域被赋予更高的空间权值,多模态特征中与主枝干相似的背景噪声能够被更好地抑制。

1.3 通道注意机制和空间激活机制

卷积网络输出的中间层特征中包含反映不同内容的通道,例如彩色模态中枝干的主要形状、细节轮廓、语义信息,深度模态中主枝干与背景的深度差异性、目标的深度轮廓等。在主枝干的显著性检测中,特征图中背景的细节轮廓信息会干扰显著性目标的检测,造成分割结果中出现与主枝干结构相似的背景枝干,甚至花椒和叶子等背景对象。因此,对多通道编码特征进行差异化关注,可以增强其中与显著性预测相关的特征表达,起到抑制多模态特征中背景信息的干扰作用,有利于前景主枝干提取。因此在跨模态融合模块每个层级中,通道注意机制被嵌入用于引导网络对融合后的跨模态编码特征中与显著性预测相关的通道,进行重点关注,其结构如图 4 所示。

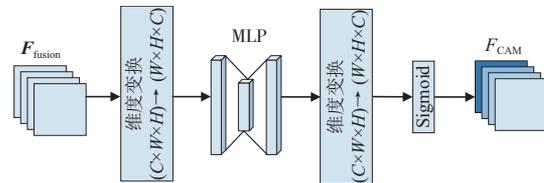


图 4 通道注意力模块

首先,对输入的串联特征 F_{fusion} 进行转置:

$$F_{\text{fusion}}^T = \text{Permute}(F_{\text{fusion}}) \quad (6)$$

式中 $\text{Permute}(\cdot)$ 表示转置操作,转置后的编码特征 $F_{\text{fusion}}^T \in R^{(H/2^i) \times (W/2^i) \times C_i}$ 。

然后,采用包含一个隐含层的多层感知机对转置特征进行非线性映射:

$$F^T = \text{MLP}(F_{\text{fusion}}^T) \quad (7)$$

式中 MLP 由两个全连接层与一个 ReLU 激活函数层组成。与 $\text{CBAM}^{[10]}$ 类似,本文在 MLP 进行特征映射时,采用 reduction 为 r 的调节通道衰减系数。 MLP 输出的编码 $F^T \in R^{(H/2^i) \times (W/2^i) \times C_i}$ 。

随后,采用转置操作还原编码特征的维度,同

时采用 Sigmoid 激活函数将还原后的特征映射到 $[0,1]$ 并获得最终的通道权值图 F_{CAM} :

$$F_{CAM} = \sigma(\text{Permute}(F^T)) \quad (8)$$

式中 σ 表示 Sigmoid 函数。

编码特征中的二维图反映了不同内容的语义激活,二维图特定的权值图能够引导网络增强相应区域的语义响应,从而抑制非感兴趣区域特征的表达。本文进一步嵌入空间激活机制到模态融合模块中。空间注意力机制依据显著性监督来自适应计算有利于枝干预测的二维权值图,从而引导枝干显著性特征的表达。

在空间激活机制中(图 5),对输入的编码特征采用 7×7 的卷积操作,获取更大的感受野,使网络能够更加有效地利用上下文空间信息。同时,为减少较大卷积核带来的计算负担,上述卷积操作采用了 reduction 为 r 的可调节通道衰减系数:

$$F_M = \text{Conv}_{7 \times 7}^r(F_{CAM}) \quad (9)$$

式中: F_M 为卷积后的编码特征; $\text{Conv}_{7 \times 7}^r$ 表示采用 7×7 卷积核和通道衰减系数 r 的卷积操作。

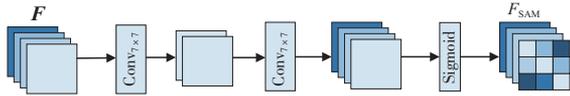


图 5 空间注意力模块

随后采用 7×7 卷积核和通道衰减系数 $1/r$ 的卷积操作将该特征映射到与输入特征相同的特征空间:

$$F = \text{Conv}_{7 \times 7}^{1/r}(F_M) \quad (10)$$

最后采用 Sigmoid 激活函数将编码特征中的特征值映射到 $[0,1]$ 以得到最终的空间权值图 F_{SAM} :

$$F_{SAM} = \sigma(F) \quad (11)$$

F_{SAM} 中每个特征点的取值范围是 $[0,1]$, 某个位置的权值较大则表明此处的特征被增强, 否则被削弱。通过训练阶段优化卷积核参数, F_{SAM} 能够自适应地根据彩色特征 X_i^{RGB} 和深度特征 X_i^{DEP} 为分割目标区域赋予较大特征权值, 增强前景主枝干区域的特征, 抑制远景枝干的干扰。因此, 融合跨模态特征的空间权值图 F_{SAM} 能够引导模型更加关注特征图中的局部重要区域。

1.4 显著性推理

特征解码阶段采用邻域插值的上采样操作会丢失目标边缘轮廓的部分特征, 使得分割结果中物体的边缘较为粗糙^[11]。然而花椒采摘点的估

计需要精确的枝干边缘作为依据。因此, 采用多尺度监督的方式来减少上采样过程中边缘信息的丢失, 如图 6 所示。在特征解码阶段, 本文采用与 Skip-Unet 类似的解码网络, 通过采用阶梯式的最近邻域上采样层和卷积激活操作来对融合后的跨模态特征编码进行最终的解码映射。在上采样过程中, 跳跃连接被用于从编码器中获取部分特征映射来提升解码器的特征丰度, 缓解因注意力机制而丢失的部分编码特征。

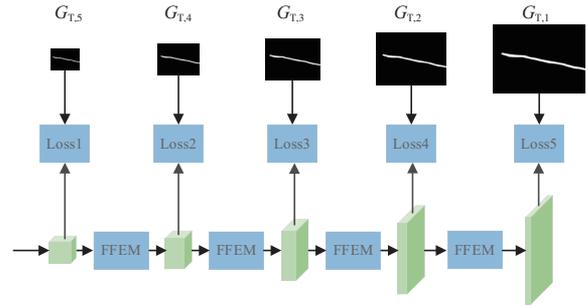


图 6 显著性推理流程

在上采样过程中的每一个层级, 对应尺度的真实标签值 $G_{T,i}$ 被依次用于监督特征解码过程, 以损失计算的方法来引导枝干特征的显著表达。其中, 不同尺度的真实标签值通过下采样真实标签图获得。

2 实验结果与分析

2.1 实验设置

1) 数据集

实验采用的花椒图像采集于四川省冕宁县, 品种为红花椒, 处于盛果采摘期(2021年7月2日—7月6日), 果实的颜色以红色为主。采用 Intel RealSense D435i 深度相机采集花椒的彩色和深度图像, 并将二者尺寸进行对齐。表 1 详细地列出了图像数据集的信息。为了保证数据的多样性, 采集的图像涵盖了 3 种天气条件(晴天、多云、雨后)和一天中的两个时间段(上午和下午), 如图 7 所示。从 5 棵不同大小的花椒树上总共收集了 1 725 张彩色图像和对应的深度图, 其中 1 042 张彩色图像被标记用于识别算法的训练和验证。其中, 70%(721) 的标记图像被用作训练数据, 剩余 30%(321) 的标记图像被用于实验验证, 以测试识别算法的拟合性能。此外, 剩余 683 张未标记的图像则被用来测试算法的识别效果。在数据标注方面, LabelMe 软件被用于手动标注主枝干的分割掩码。

表 1 数据集分布

项目	训练集	验证集	测试集	总和
标注	721	321	0	1 042
未标注	0	0	683	683
分辨率	640(pixels)×480(pixels)			

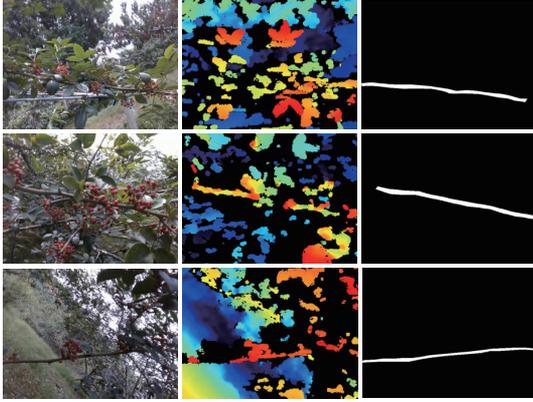


图 7 晴天、多云及雨后数据集图像

2) 模型参数

实验框架基于 PyTorch1.2 框架搭建,训练阶段采用 Adam 优化器来训练网络,选取学习率、批量、迭代次数和通道衰减系数 r 分别设置为 0.001、6、800 和 16。测试阶段,将模型预测概率大于 0.7 的像素作为分割目标。所有实验环境均采用 Ubuntu 18.04 的设备环境,显卡为 NVIDIA GeForce RTX 3090。

3) 评价指标

为更好地评估模型的综合性能,采用平均精确度 P 、平均召回率 R 、 F_{measure} 和平均绝对误差 (mean absolute error, MAE) 作为评价指标。如表 2 所示,模型检测结果的定义主要分为真阳性 (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。

表 2 模型不同检测结果的定义

分类	真实为枝干	真实为背景
分割为枝干	TP	FP
分割为背景	FN	TN

P 和 R 计算公式如下:

$$P = \frac{T_p}{T_p + F_p} \quad (12)$$

$$R = \frac{T_p}{T_p + F_n} \quad (13)$$

F_{measure} 是对精确度和召回率的整体表现评估,计算公式如下:

$$F_{\text{measure}} = \frac{(1 + \beta^2) \times P \times R}{\beta^2 \times P + R} \quad (14)$$

式中 β^2 是一个超参数,通常取 0.3。 P 、 R 、 F_{measure} 数值越大,显著性目标检测效果越好。

MAE 值用于评估显著预测图和真值图之间的平均绝对差值,代表显著性检测的整体效果:

$$M_{\text{AE}} = \frac{1}{N \times M} \sum_{x=1}^N \sum_{y=1}^M |S(x, y) - G(x, y)| \quad (15)$$

式中: N 和 M 分别为图像的长和宽; $S(x, y)$ 和 $G(x, y)$ 分别为 (x, y) 处的显著预测值和真值。MAE 值越小,表明模型预测的结果与真实标注图间差异越小,因此枝干的分割性能越好。

2.2 对比实验

为测试本文模型的性能,本文采用主流的显著性模型进行对比,包括全局上下文感知渐进聚合显著性目标检测网络 (global context-aware progressive aggregation network for salient object detection, GCPANet)^[12]、金字塔特征注意力显著性检测网络 (pyramid feature attention network for saliency detection, PFAN)^[13]、基于 CNN 的跨视图转移和多视图融合 RGB-D 显著性检测网络 (CNNs-based RGB-D saliency detection via cross-view transfer and multiview fusion, MV-CNN)^[14]、基于分层动态滤波 RGB-D 显著性检测网络 (hierarchical dynamic filtering network for RGB-D salient object detection, HDFNet)^[15]。指标结果定量比较如表 3 所示。

表 3 不同模型的花椒检测性能

模型	P	R	F_{measure}	M_{AE}
PFAN	0.817 2	0.823 2	0.818 6	0.062 7
GCPANet	0.837 0	0.825 6	0.834 3	0.052 8
HDFNet	0.832 7	0.843 6	0.835 2	0.053 8
MV-CNN	0.842 8	0.844 6	0.843 2	0.042 6
本文模型	0.852 3	0.862 5	0.854 6	0.041 2

从表 3 中可以看出:本文模型取得了最佳的显著性检测效果,其中 P 、 R 、 F_{measure} 和 M_{AE} 指标分别为 0.852 3、0.862 5、0.854 6 和 0.041 2。与 MV-CNN 网络相比,指标性能分别提升了 0.95、1.79、1.14 和 0.14 个百分点。相比于其他的 3 种网络,

指标性能分别有 1.53~3.51、1.89~3.93、1.94~3.61 和 1.16~2.15 个百分点的提升。此外,为了清晰地观察实验结果,本文绘制了各个模型的 $P-R$ 曲线。如图 8 所示, $P-R$ 曲线表明本文模型优于 MV-CNN 模型,同时明显优于 HDFNet、GCPANet 和 PFAN 模型。

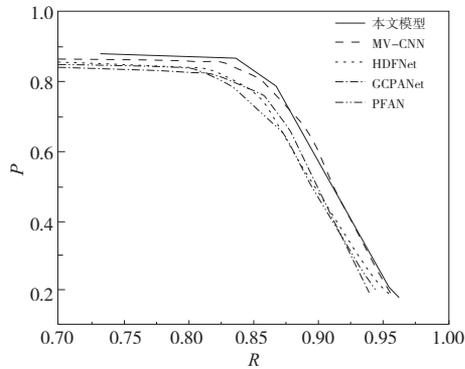


图 8 不同模型 $P-R$ 曲线对比

为了更直观地分析结果,本文进一步展示了各个模型检测的可视化结果,如图 9 所示。

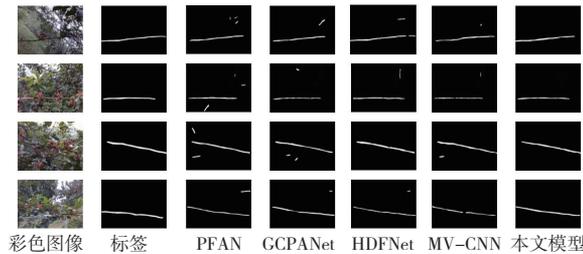


图 9 不同模型的花椒识别效果

从图 9 中可以看出,本文模型能够在杂乱背景、前景和不易区分背景、多个对象等复杂场景中,准确检测到显著花椒枝干区域。如第一、第二行图片中存在较多与前景主枝干类似的背景枝干,但本文模型能够充分抑制背景枝干噪声,将前景主枝干识别出来。第三行图片中前景区域存在多个枝干对象,本文模型仍能够将前景主枝干识别出来。这表明本文模型能够有效过滤冗余信息,准确地输出识别结果。

2.3 消融实验

为测试本文所提不同模块对模型显著性检测的影响,设置以下模型进行对比实验:A,基础模型,经典的 U-Net 单模态骨干网络模型;B,双编码器模型,在 U-Net 模型的基础上,增加另一编码器通道提取深度图像特征,两通道在编码阶段结束后特征直接拼接进入解码阶段;C,跨模态多尺

度特征融合模型,该模型同样采用双通道提取图像特征,并在编码阶段的 5 个层次分别加入特征融合模块进行特征融合,融合后的特征进入彩色图像通道继续编码;D,多尺度监督模型,在 U-Net 模型的基础上,在解码阶段进行多尺度监督;E,本文模型,同时采用跨模态多尺度特征融合模块和多尺度监督模块。上述模型的检测性能如表 4 所示。

表 4 不同模型的花椒检测性能

模型	P	R	F_{measure}	M_{AE}
A	0.816 5	0.812 9	0.815 7	0.060 7
B	0.800 6	0.807 3	0.802 1	0.065 3
C	0.842 7	0.840 3	0.842 1	0.048 2
D	0.828 5	0.830 6	0.829 0	0.053 6
E	0.852 3	0.862 5	0.854 6	0.041 2

从表 4 中可以看出:模型 B 由于未采用特征融合模块对深度图像和彩色图像进行特征融合,导致被冗余的深度信息干扰,评价指标反而差于基础模型;模型 C 在引入多尺度特征融合增强模块后,评价指标得到了明显的提升;模型 D 的多尺度监督模块也促进了模型的评估性能。多尺度特征融合增强模块和多尺度监督模块的联合使用和仅使用单个模块相比,指标性能分别有 0.96~2.38、2.22~3.19、1.25~2.56 和 0.70~1.24 个百分点的提升。此外,本文绘制了 5 种模型检测的 $P-R$ 曲线,如图 10 所示。从中可以看出,使用多尺度特征融合增强模块和多尺度监督模块后模型的检测曲线能够将其他曲线完全包住,这证明了其性能优于其他几种检测模型。

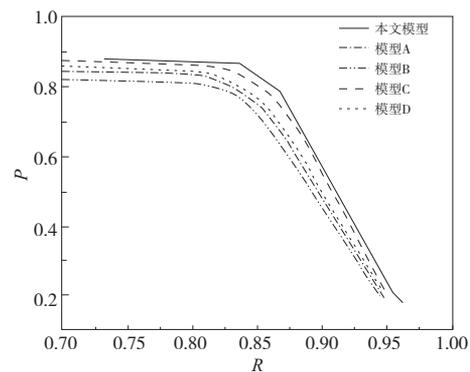


图 10 消融实验 $P-R$ 曲线对比

为了更加直观地反映出多尺度特征融合增强模块和多尺度监督模块对显著性检测的影响,本文对这几个模型的部分检测结果进行了可视化,

如图 11 所示。从图 11 中可以看出:模型 B 增加了预测图的噪声数量,带来了负面效果;模型 C 又极大程度上抑制了噪声,这定性证明了多尺度特征融合增强模块的作用;使用模型 D 的多尺度监督模块后,模型提取边缘信息的能力更强;同时本文模型使用了多尺度特征融合增强模块和多尺度监督模块,可以得到轮廓清晰且无噪声的前景主枝干预测图。

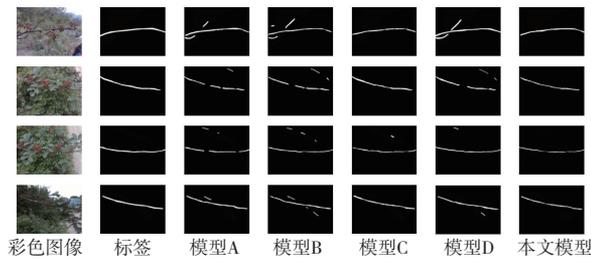


图 11 消融实验可视化对比

3 结语

针对现有显著性目标检测模型难以准确定位复杂场景下花椒枝干的问题,本文提出跨模态特征融合的 RGB-D 花椒图像显著性检测模型。本文方法首先采用双分支主干网络来提取彩色图像特征和深度图像特征,并利用注意力机制来引导这两个模态特征的融合;最后引入多尺度监督方法用于提升模型对于显著性枝干边缘的检测性能。实验结果表明:本文模型的各项评估指标均优于其他显著性目标检测模型,能够得到更加精确的枝干检测结果。

参考文献:

[1] 杨前,刘兴科,罗建桥,等. 基于多任务上下文增强的花椒检测模型[J]. 机械制造与自动化,2023,52(1): 113-118,149.

[2] LIU J J, HOU Q B, LIU Z A, et al. PoolNet⁺: exploring the potential of pooling for salient object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 887-904.

[3] 黄世国,洪铭淋,张飞萍,等. 基于 F~3Net 显著性目标检测的蝴蝶图像前背景自动分割[J]. 昆虫学报, 2021, 64(5): 611-617.

[4] SUN Q X, CHAI X J, ZENG Z K, et al. Noise-tolerant RGB-D feature fusion network for outdoor fruit detection[J]. Computers and Electronics in Agriculture, 2022, 198: 107034.

[5] 陈曦涛,訾玲玲,张雪曼. 采用跳层卷积神经网络的 RGB-D 图像显著性检测[J]. 计算机工程与应用, 2022, 58(2): 252-258.

[6] LIU Z Y, LIU J W, ZUO X, et al. Multi-scale iterative refinement network for RGB-D salient object detection[J]. Engineering Applications of Artificial Intelligence, 2021, 106: 104473.

[7] WANG X H, LI S, CHEN C, et al. Depth quality-aware selective saliency fusion for RGB-D image salient object detection[J]. Neurocomputing, 2021, 432: 44-56.

[8] NIU Z Y, ZHONG G Q, YU H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.

[9] WANG N, CUI Z G, SU Y Z, et al. Multiscale supervision-guided context aggregation network for single image dehazing[J]. IEEE Signal Processing Letters, 2021, 29: 70-74.

[10] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//European Conference on Computer Vision. Cham: Springer, 2018: 3-19.

[11] YU J, YAO J H, ZHANG J, et al. SPRNet: single-pixel reconstruction for one-stage instance segmentation[J]. IEEE Transactions on Cybernetics, 2021, 51(4): 1731-1742.

[12] CHEN Z Y, XU Q Q, CONG R M, et al. Global context-aware progressive aggregation network for salient object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 10599-10606.

[13] ZHAO T, WU X Q. Pyramid feature attention network for saliency detection [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 3080-3089.

[14] HAN J W, CHEN H, LIU N, et al. CNNs-based RGB-D saliency detection via cross-view transfer and multiview fusion [J]. IEEE Transactions on Cybernetics, 2018, 48(11): 3171-3183.

[15] PANG Y W, ZHANG L H, ZHAO X Q, et al. Hierarchical dynamic filtering network for RGB-D salient object detection [C]//Vedaldi A, Bischof H, Brox T, et al. European Conference on Computer Vision. Cham: Springer, 2020: 235-252.

收稿日期:2023-04-07