doi:10.6041/j.issn.1000-1298.2025.04.034

基于深度增强与特征抗噪的夜间串番茄成熟度识别方法

王　新[1]　唐　灿[1]　朱建新[2]　郭彩平[1]　刘艺豪[1]　王书茂[1]
(1. 中国农业大学工学院, 北京 100083; 2. 山河智能装备股份有限公司, 长沙 410100)

摘要: 机器人自动化采收是温室串番茄收获作业的有效解决方案,其中串番茄成熟度是机器人采摘决策的重要依据。本文利用 Deep White－Balance 与 Zero－DCE 深度神经网络分别实现串番茄图像偏色修正和阴影细节增强,改善补光环境与夜间环境下图像偏色与局部弱光等问题。同时引入深度残差收缩网络,在 YOLO v5s 中构建 RSBottleneck－CW 模块对特征图进行软阈值化处理,抑制图像中的噪声干扰。实验结果表明,在夜间环境下,图像经 Zero－DCE 算法增强处理后,检测模型召回率达到 0.924,捕获到了更多的番茄果实与果串目标。在补光环境下,图像经过 Deep White－Balance 与 Zero－DCE 联合处理后恢复了真实色彩并增强了纹理细节,检测模型平均精度均值(mAP)达到 0.849,相比于处理前提升 0.038。而嵌入 RSBottleneck－CW 模块的 YOLO v5s 对特征图噪声表现出了很强的适应性能,不管是否对图像进行深度增强,其 mAP 与 F1 值始终比原始 YOLO v5s 更高,夜间环境下 mAP 和 F1 值最高分别为 0.902、0.844,补光环境下 mAP 和 F1 值最高分别为 0.868、0.817。检测模型检测出果实与果串后,利用边框匹配算法可以获取到串番茄最终的成熟度。当串番茄成熟度为 90% ~100%时,夜间环境与补光环境下串番茄成熟度识别平均绝对误差分别为 1.837%、1.067%,可为串番茄采摘机器人夜间自动采收作业提供决策依据。

关键词: 串番茄; YOLO v5s; 目标检测; 成熟度; 夜间环境
中图分类号: S641.2; TP391.4　　**文献标识码**: A　　**文章编号**: 1000-1298(2025)04-0363-12　　**OSID**:

Tomato Cluster Ripeness Recognition at Night Based on Depth Enhancement and Feature Noise Reduction

WANG Xin[1]　TANG Can[1]　ZHU Jianxin[2]　GUO Caiping[1]　LIU Yihao[1]　WANG Shumao[1]
(1. *College of Engineering*, *China Agricultural University*, *Beijing* 100083, *China*
2. *Sunward Intelligent Equipment Co.*, *Ltd.*, *Changsha* 410100, *China*)

Abstract: Robotic automated harvesting proves to be an efficient solution for greenhouse tomato cluster harvesting operations. The ripeness of tomato clusters stands as a vital criterion influencing the decision-making process for the harvesting robot. Aiming to employ Deep White－Balance and Zero－DCE deep neural networks for color cast correction and shadow detail enhancement in tomato images to enhance image quality by addressing color cast issues and improving local illumination in both fill light and night environments, the concept of the deep residual shrinkage network was introduced, incorporating the RSBottleneck－CW module into YOLO v5s. This module conducted soft threshold processing on the feature map to effectively suppress noise interference in the image. Experimental results demonstrated that in the night environment, after enhancing the image solely with the Zero－DCE algorithm, the recall of the detection model reached 0.924, capturing more tomato fruits and trusses. In a supplementary light environment, the image underwent joint processing with Deep White－Balance and Zero－DCE to restore authentic colors and enhance texture details. This resulted in the detection model achieving an mAP of 0.849, reflecting a 0.038 increase compared with that of the before processing. The YOLO v5s integrated with the RSBottleneck－CW module exhibited robust adaptability to feature map noise. Irrespective of whether the image underwent depth enhancement, its mAP and F1-Score consistently surpassed those of the original YOLO v5s. In the nighttime environment, the highest recorded mAP and F1-Score values were 0.902 and 0.844, respectively. Similarly, in the supplementary light environment,

收稿日期: 2024－02－29　修回日期: 2024－07－08
基金项目: 湖南省智能农机装备创新研发项目(202404710710436)
作者简介: 王新(1982—),女,副教授,博士,主要从事农业装备智能测控技术研究,E-mail: wangxin117@cau.edu.cn

the peak mAP and F1-Score values reached 0.868 and 0.817, respectively. After the detection model detected the fruits and trusses, the final ripeness level of the tomato clusters were determined by using the bounding boxes aligning algorithm. In the ripeness stage of tomato clusters ranged from 90% to 100%, the average absolute errors in ripeness recognition under nighttime and supplementary light conditions were 1.837% and 1.067%, respectively. These findings can serve as decision-making criteria for night automated harvesting operations of tomato-picking robots.

Key words: tomato cluster; YOLO v5s; target detection; ripeness; night environment

0 引言

串番茄作为高经济价值果蔬,在我国农业生产中占据重要地位。面对串番茄收获这种劳动密集型的作业,智能机器人自动化采收正成为重要发展趋势[1-3]。在机器人自动化收获过程中,判断串番茄成熟度是一个关键环节。串番茄属于呼吸跃变型水果,虽然过早采收后仍会继续成熟,但其风味会大打折扣[4-5]。因此,快速而准确地评估串番茄成熟度对于机器人正确执行采摘任务至关重要。夜间串番茄成熟度识别是机器人执行夜间采收作业的前提条件。目前,设施农艺规范性种植,为机器人自动化收获提供了有利条件。机器人可以全天、无间歇地进行采收作业,来实现其利用率最大化[6-7]。为了满足这种工作方式,夜间环境下果实成熟度自动检测显得尤为重要,但同时也面临诸多挑战。目前已有夜间水果检测方法研究,比如芒果[8]、荔枝[9]、柑橘[10]等。学者们通过添加主动光源来获取夜间图像并利用深度学习模型来检测水果。然而,由于夜间环境下成像对曝光度敏感性高,图像易出现高噪声、颜色失真等问题[11-12]。对于芒果、荔枝、柑橘等水果的检测,由于果实与背景的特征差异明显,检测结果不易受到噪声和颜色失真的干扰。但是,对于同一水果不同成熟度类别,由于各类别之间差异较小,很容易受到果实表面颜色和纹理影响[13-14],导致成熟度分类出现误判。此外,串番茄是一种喜光植物,在设施种植中常需要进行夜间补光以促进果实发育与产量增收[15-16]。由于补光灯的颜色干扰(如红蓝混合补光灯),图像会出现严重色偏,这对于成熟度识别是一个巨大挑战。

为了满足货架期延长和销售链的需求,串番茄通常整串采摘。与单个番茄成熟度识别不同[17-18],需要把整串看作一个整体,判断其成熟度以达到收获标准。已有研究将其视为两阶段任务,即首先检测番茄串,再计算其中成熟果实像素面积比值来获取成熟度[19-20],这种方法存在误差累积,很容易造成成熟度误判。

针对上述问题,本文提出一个结合 Deep White-Balance 和Zero-DCE 的图像预处理方案,以恢复图像真实色彩并增强纹理细节。同时,通过在 YOLO v5s 模型中构建深度残差收缩模块,对特征图进行软阈值化处理,改善图像高噪声对检测结果的干扰。此外,利用 YOLO v5 能同时检测不同尺度目标优势,一次性识别出所有串番茄和果实,然后通过边框匹配算法直接得到串番茄成熟度,以避免两阶段方法误差累积。

1 材料和方法

1.1 数据集

1.1.1 数据采集

针对夜间环境和补光环境共进行两轮实验数据采集。首轮实验于 2023 年 4 月 20 日 19:00—23:00 在北京市朝阳区明日农场进行,旨在拍摄夜间环境下串番茄图像。此次采集过程中,实验区域由主动光源照明。所使用设备包括 Realsense L515 型相机、便携式计算机、可调节和移动的自制相机支架。为了获得均匀光照,自制一个 LED 灯板用于实验,其四周对称贴有 LED 日光灯条,如图 1a 所示。第 2 轮实验于 4 月 24 日 19:00—23:00 在明日农场拍摄补光环境下串番茄图像。补光灯为彗控全光谱补光灯(红蓝光混合),光谱波长为 400~750 nm,额定功率为 36 W。同时温室采用冠层补光方式,补光灯均匀布置在番茄植株行间 3 m 高位置处,以提供足够光源来促进果实发育,如图 1b 所示。每次图像采集时,相机垂直于番茄垄面,并距离植株 0.4~0.5 m,高度为 1.2~1.4 m。保存为 1 280 像素×720 像素的 RGB 图,同时包含与之对齐的深度帧。在夜间环境下,共获取 364 幅图像;在补光环境下,共获取 153 个图像对,其中每对图像包含 1 幅无色卡图像和 1 幅有色卡图像,无色卡图像通过有色卡图像来校正颜色,并作为后续深度白平衡(Deep White-Balance)任务的参考。采用国际标准的 datacolor Spyder Checkr 24 色卡。数据示例如图 2a 所示。

1.1.2 数据注释与划分

串番茄以整串中每个果实成熟度为标准来判断其是否应采收。为了降低复杂性同时满足不同销售模式的采摘需求,根据番茄成熟过程的生理特性变化与颜色表征[21],将果实分为 Green、Pink、Red 共 3

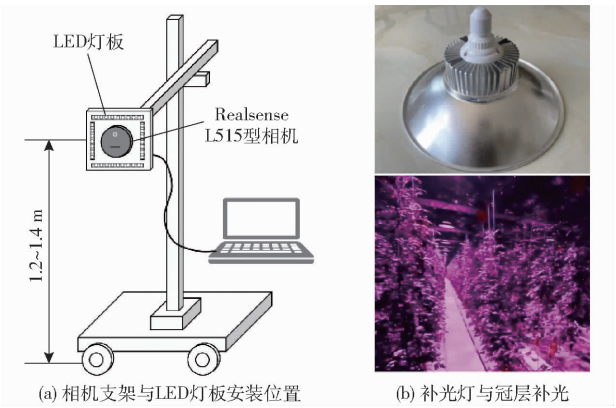

图 1　数据采集设备与补光现场

Fig. 1　Data acquisition equipment and lighting scene

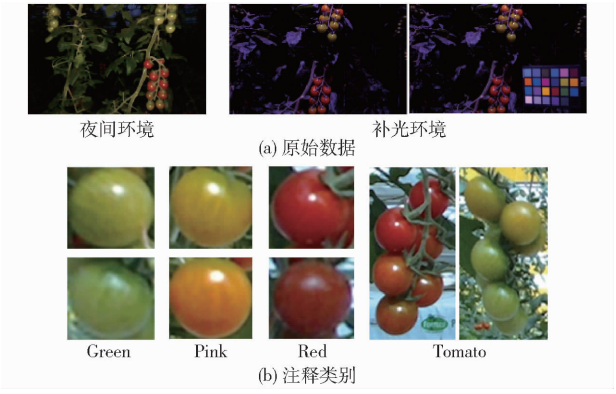

图 2　原始数据与成熟度注释类别

Fig. 2　Original data and ripeness annotation categories

个成熟度类别，如图 2b 所示。Green 阶段番茄表面呈绿色或浅绿色，此时将不再膨胀。Pink 阶段番茄表面呈黄色或浅黄色，此时番茄正处于转色期，糖分开始累积。Red 阶段番茄表面呈红色或暗红色，此时番茄已经成熟且质地开始软化，具有最佳风味。同时增加 1 个 Tomato 类别，用于检测番茄果串。所有数据均以当天日间的番茄果实颜色为参照，利用 Labelimg 工具标注出图像的 Green、Pink、Red、Tomato 边框，用于后续检测任务。将两种环境数据均分为训练集、验证集和测试集，训练集和验证集用于模型训练，测试集用于验证模型泛化性能。最终获得夜间环境下训练、验证、测试的样本边框数量分别为 3 861、669、849，补光环境下分别为1 613、517、593。

1.1.3　深度白平衡模型数据集

在深度白平衡模型训练任务中，需要色彩正确的图像作为训练数据的参考真值，这可以通过手动获取。原始的偏色图像手动色彩校正主要分 2 个阶段：使用色卡中性灰区域作为参考来校正图像的极端色偏；利用色卡配套的 Spyder Checkr 专业工具对校正后图像色彩进一步精细调整。具体步骤是：首先在 Lightroom 中用吸管点击图像中色卡的中性灰区域，Lightroom 自动对图像进行白平衡处理，使得点击的色卡中性区域 RGB 值趋于一致，从而获取到该场景下色彩校正矩阵 $\widetilde{C}_1$，利用该矩阵对图像对中另 1 幅无色卡图像进行白平衡处理来校正颜色，这时色彩校正矩阵基本可以还原图像真实色彩。然后裁剪出初步校正后的色卡区域，用 Spyder Checkr 工具计算出另一个色彩校正矩阵 $\widetilde{C}_2$，利用此矩阵对上一步中初始白平衡处理后的无色卡图像再进行一次颜色校正得到最终微调后的图像，数据预处理流程如图 3 所示。

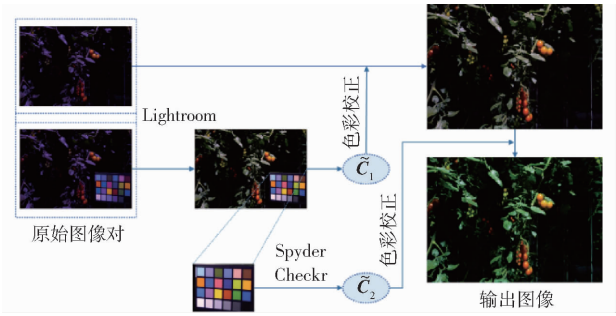

图 3　深度白平衡模型数据集处理流程图

Fig. 3　Dataset processing flowchart of Deep White－Balance model

1.2　方法总体流程

所提方法主要包括图像预处理、目标检测和边框匹配 3 个模块，总体处理流程如图 4 所示。为改善图像质量，补光环境下图像需要经深度白平衡与光照增强预处理，而夜间环境下图像仅通过光照增强预处理即可。随后目标检测网络对增强后高质量图像提取目标特征，检测出果串和不同成熟度果实。最后将检测到的结果进行边框匹配和过滤，计算出果串成熟度。

1.3　图像深度增强方法

1.3.1　深度白平衡

温室番茄种植通常会在夜间采用补光方式来提升果实风味和提高产量。这些补光灯通常具有特定光谱范围，且照射出的光线带有一定颜色，从而可以最大化植株对光线的利用率[22－23]。而相机自带的白平衡功能无法应对这种极端环境，最终相机获取到的图像具有严重色偏。这种由于相机错误白平衡造成的图像色偏会使得图像颜色分布趋于一致性，从而给深度神经网络特征提取带来巨大挑战[24]。特别是番茄成熟度判断这种高度依赖果实颜色特征的任务，模型在这种偏色和低对比度图像中很容易出现错误判断。深度白平衡处理可有效改善图像色偏，提升模型对成熟度识别的准确度。

深度白平衡基于相机白平衡原理，将输入的

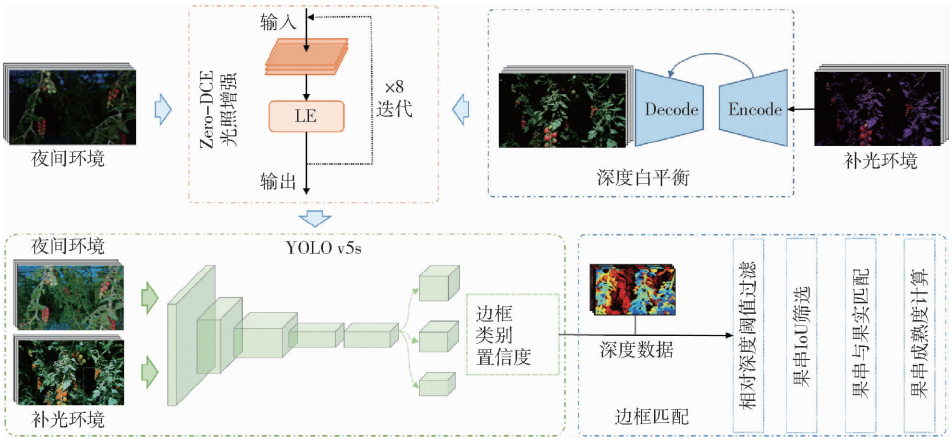

图 4 数据总体处理流程图

Fig. 4 Overall data processing flowchart

s-RGB图像编码为 raw-RGB 图像的潜在表示，然后用解码器将 raw-RGB 潜在表示解码为目标图像[25]。流程可表示为

$$\hat{I}_{WB^{(t)}} = g_t(f(I_{WB}^{(in)})) \tag{1}$$

其中 $f(\cdot)$ 为将输入图像 $I_{WB}^{(in)}$ 编码为 raw-RGB 图像的潜在表示，$g_t(\cdot)$ 为将 raw-RGB 图像潜在表示解码为目标图像。解码器设置为 g_A、g_T、g_S 3 种，分别对应生成正确 WB(A)图、白炽灯 WB(T)图、阴影 WB(S)图。因本文仅需要正确的 WB(A)图结果，裁剪掉其他 2 个解码器。为了在满足白平衡需求下进一步减少模型参数量，网络第 4 级的编码和解码层被舍弃。调整后的模型结构如图 5 所示。该网络采用 U-Net 结构[26]，由 1 个 3 级编码器和 1 个 3 级解码器组成。同时在编码器与解码器不同尺度上增加了跳跃连接，以促进特征复用，减少细节丢失。

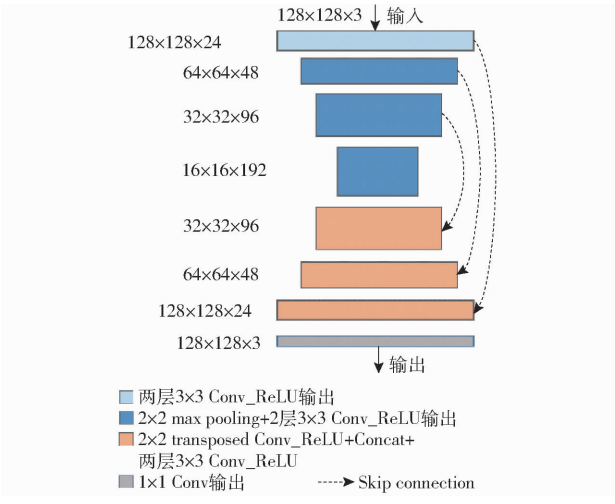

图 5 深度白平衡网络结构图

Fig. 5 Schematic of Deep White-Balance network structure

1.3.2 Zero-DCE 光照增强

由于夜晚温室环境下相机曝光不佳、植株叶面遮挡等，串番茄图像易出现局部弱光或阴影问题。这种图像会引起检测网络难以捕捉到目标，出现一些漏检现象。Zero-DCE 是一种无监督学习的弱光增强算法[27]，其核心思想是利用一组高阶曲线对低光图像进行不断自我迭代来实现图像增强。图像每个通道的每个像素对应一个独立的高阶曲线，迭代方式为

$$LE_n(x) = LE_{n-1}(x) + A_n(x)LE_{n-1}(x)(1 - LE_{n-1}(x)) \tag{2}$$

其中，n 为迭代次数，$LE_n(x)$ 为图像应用高阶曲线增强后的结果，$A_n(x)$ 为与图像尺寸相同的可学习参数图。参数 $A_n(x)$ 通过一个轻量的卷积神经网络 DCE-Net 学习获得。DCE-Net 结构和图像迭代流程如图 6 所示。网络由 6 个相同 3×3 卷积层组成且卷积操作后不改变特征图尺寸。最后一层特征图还需经过一次 3×3 卷积与 Tanh 激活，进而生成 8 组最终的参数图。每组参数图对输入 RGB 图像进行依次迭代。迭代完成的图像中值滤波后作为最终输出，以抑制增强带来的噪声扩大。

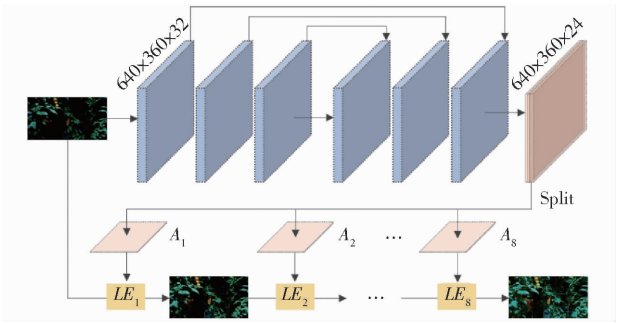

图 6 Zero-DCE 处理图像流程图

Fig. 6 Zero-DCE image processing flowchart

DCE-Net 可以在没有配对或不配对的数据集上进行训练，损失函数为

$$L = L_{spa} + L_{exp} + \alpha L_{col} + \beta L_{tvA} \tag{3}$$

其中 L_{spa} 为空间一致性损失值，用于保持增强图像空间一致性。L_{exp} 为曝光损失值，用于控制曝光水平。L_{col} 为颜色恒常性损失值，用于防止图像增强后

出现色偏。L_{tvA}为光照平滑损失值，用于保持增强图像像素单调关系。α、β为权重。该模型具有参数量小、运行速度快等优点，并且能够在保持正常亮度区域同时增强阴影区域亮度、提高纹理细节。这对改善夜晚番茄图像质量，提升下游番茄视觉检测任务效果具有重要作用。

1.4　引入深度残差收缩模块的 YOLO v5s

夜晚拍摄的串番茄图像中存在非常多噪声，这会使得网络提取到的特征图出现很多无用信息干扰。利用深度残差收缩模块的软阈值化思想，在 YOLO v5s 中嵌入类似模块来促进网络对噪声图像无用信息的抑制和有用信息的保留。

深度残差收缩网络是解决一维高噪声信号故障诊断问题的深度卷积神经网络[28]。深度残差收缩模块结构如图 7a 所示。其主要思想是在卷积层中嵌入一个类似 SE 注意力模块[29]分支来学习一组阈值，然后对主路径上的特征值进行软阈值化处理。软阈值化处理为

$$y=\begin{cases}x-\tau & (x>\tau)\\ 0 & (-\tau\leqslant x\leqslant\tau)\\ x+\tau & (x<-\tau)\end{cases} \tag{4}$$

其中 τ 为设定的阈值。通过将绝对值小于某个阈值的特征值置为零来实现噪声抑制，同时保留信号的有用部分。

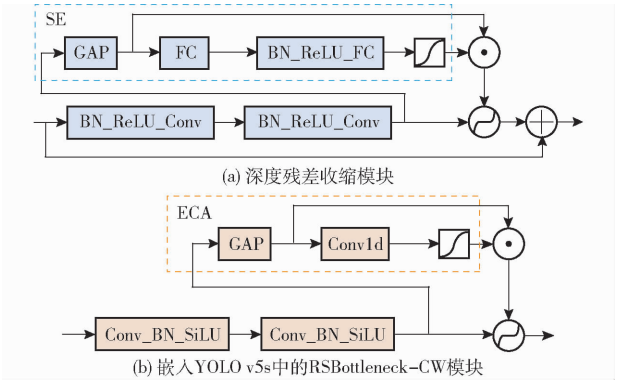

图 7　不同深度残差收缩模块结构图

Fig. 7　Different depth residual shrinkage module structure diagram

参考深度残差收缩网络思想设计出具有通道独立阈值的瓶颈模块 RSBottleneck－CW，结构如图 7b 所示。该模块利用参数量更少、特征提取能力更强的 ECA 注意力模块[30]代替 SE 模块来学习阈值。输入特征图先经过 2 个卷积层，再通过 ECA 分支学习一组通道独立阈值。主路径上特征图的每个通道都对应有一个阈值，特征图再经过软阈值化处理后得到 RSBottleneck－CW 模块的最终输出。RSBottleneck－CW 模块被嵌入到 YOLO v5s 的颈部来提升网络特征提取能力。

YOLO v5 是一种基于锚框的单阶段目标检测器，得益于其颈部特征融合网络和 3 个不同尺度检测头的设计，YOLO v5 可以很轻松地实现大、小目标的同时检测。基于这个优点，采用 YOLO v5 来同时检测出串番茄图像中目标尺寸差异较大的果实和果串。YOLO v5s 网络中包含很多 C3 模块，这些 C3 模块由 n 个 Bottleneck_False/True 组成，这里将所有的 Bottleneck_False 模块替换为带软阈值化处理的 RSBottleneck－CW 模块，用于提升网络对噪声数据的适应性能。

1.5　边框匹配与过滤

串番茄整串采收要求将 1 串视为整体来判断其成熟度。通过 YOLO v5 检测出果串和果实边框后，需要进一步找到每个果串边框中包含的各成熟阶段的果实。此外，由于串番茄并行种植模式，采集的图像中具有背景果实干扰。在边框匹配同时，融入目标深度信息来进行相对深度阈值筛选，从而过滤掉背景误检边框。边框匹配整体算法流程为

Algorithm 1：Bounding boxes aligning

Input：$B=\{b_1,\cdots,b_N\}$，$C=\{c_1,\cdots,c_N\}$，$D=\{d_1,\cdots,d_N\}$，d，u

B is the list of detection boxes,

C contains corresponding detection categories,

D contains depth values of corresponding detection boxes,

d is the depth threshold,

u is the IoU threshold

Output：All truss boxes, each element contains multiple fruit boxes

$T\leftarrow\{\}$　# trusses
$F\leftarrow\{\}$　# fruits
$m\leftarrow\operatorname{argmin} D$
for b_i in B do
　if $(d_i-m)\leqslant d$ then
　　if c_i is Tomato then
　　$T\leftarrow T+b_i$
　　else
　　$F\leftarrow F+b_i$
　　end
　end
end
$T_f\leftarrow\{\}$　# filtered trusses
for t_i in T do
　$t\leftarrow\{t_{i+1},\cdots,t_m\}$，$m\leftarrow\mathrm{Len}(T)$
　　for t_j in t do
　　$iou=\mathrm{IoU}(t_i,t_j)$

```
        if iou ⩾ u then
            T_f + S_min(t_i, t_j)   # adding small area box
        end
    end
end
T ← T − T_f   # trusses updating
O ←{ }
for t_i in T do
    T_F ←{t_i}
    for f_i in F do
        c ← CenterPoint f_i
        if c ∈ Area t_i then
            T_F ← T_F + f_i   # adding fruits
        end
    end
    O ← O + T_F
end
return O
```

先遍历所有检测结果,对每个结果进行相对深度阈值筛选。对相对深度小于某阈值的检测框进行保留,相对阈值筛选为

$$\begin{cases} d_i = D_i - D_{\min} \\ d_i \leqslant d_{thre} \end{cases} \tag{5}$$

式中 d_i——相对深度

D_i——每个框深度

$D_{\min}$——所有框中最低深度

d_{thre}——筛选阈值

通过这种方法过滤可以不受镜头与垄面距离变化的影响。然后对相对深度阈值筛选后的边框结果分类别统计,且计算各果串边框的 IoU 值,对 IoU 值大于某阈值的边框进行剔除,以消除同一果串的多个检测结果。最后遍历所有果串,统计中心落在果串边框中的所有果实边框,完成果实果串匹配。最终计算出串番茄成熟度,即

$$R = \frac{N_R + 0.5N_P}{N_R + N_P + N_G} \times 100\% \tag{6}$$

其中 N_R、N_P、N_G 分别表示 Red、Pink、Green 果实数量。为了避免在果实重叠时成熟度计算出现较大误差,没有采用计算像素面积比值来获取成熟度[19-20],而利用果实数量比值作为串番茄成熟度的方法,在面对不同成串姿态时具有更稳健的结果。

1.6 实验设计

针对夜间和补光两种环境情况来讨论串番茄成熟度识别效果:对夜间照明环境仅采用光照增强预处理,补光环境则采用深度白平衡和光照增强预处理。所有情况均考虑使用深度残差收缩模块提升目标检测网络性能。

1.6.1 模型训练方法

对于深度白平衡模型,预训练模型是在 12 000 幅规模数据集上训练而成。这里需要在串番茄夜间补光环境数据集上进行迁移学习,以获得满足当前环境需求的白平衡模型。然后对原始图像进行九角切片来丰富模型采样范围和扩充数据。训练时从每幅输入图像上进行随机区域裁剪得到 8 幅 128 像素 × 128 像素的图像作为模型训练输入数据,详细流程如图 8 所示。训练批次大小为 32,训练轮次为 50,优化器为 Adam,学习率为 0.000 1。

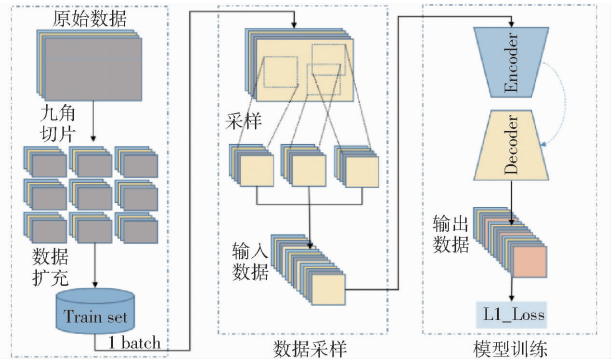

图 8 深度白平衡模型训练方法

Fig. 8 Training method of Deep White - Balance model

利用白天数据训练出的模型进行迁移学习可以有效提升夜间模型检测的性能[31]。所以本文夜间环境和全光谱环境下的目标检测模型均采用了迁移学习方法。预训练模型为温室白天不同天气和光照条件下的串番茄成熟度识别模型。迁移学习训练轮次均为 200,在前 20 轮次对骨干网络进行冻结训练,学习率为 0.01。同时在所有训练阶段采用 mosaic 数据增强,提升模型鲁棒性。最终保存在验证集上表现最好的模型进行测试评估。

所有模型训练和测试均在 1 台 Intel(R) Core(TM) i7 - 10700KF CPU @ 3.80 GHz, 10 GB NVIDIA GeForce RTX 3080 GPU,运行内存 32 GB 的计算机上完成。

1.6.2 结果评价方法

针对深度白平衡任务,对比迁移学习前、后以及使用图像九角切片后结果。以色差(ΔE)、平均角度误差(Mean angular error, MAE)、均方误差(Mean squared error, MSE)等指标来定量判断结果。

针对目标检测任务,分情况讨论夜间条件和补光条件下的模型检测效果。夜间条件下对比分析应用光照增强、RSBottleneck - CW 模块前后的结果。补光条件下对比分析应用深度白平衡、光照增强、以及 RSBottleneck - CW 模块前后效果。目标检测时

常用评价指标召回率（Recall，R）、精确率（Precision，P）、平均精度均值（mean Average Precision，mAP）、平均精度（AP）以及 F1 值定量评价检测任务。采用平均绝对误差评价最终的串番茄成熟度识别准确率。

2　实验结果与分析

2.1　深度白平衡结果

深度白平衡测试结果如表 1 所示。ΔE、MAE 与 MSE 的最小值均出现在使用了切片增强和迁移学习的模型结果中。直接采用文献[27]的预训练模型效果较差，这可能是因为预训练模型的数据集中包含较少的类似极端情况的偏色图像。经过对模型在特定数据集上进行微调后，各指标均出现显著改善，但是在部分果实区域仍存在一定偏色，如图 9 中的迁移学习模型结果所示（每行的第 1 列上下分别表示原始图像和参考图像）。这是由于模型训练时输入数据随机取样不足。通过对图像进行九角切片后，图像尺寸缩小，促进了随机取样的均匀性和丰富性。所以最终经过切片和迁移学习微调出来的模型可以恢复更真实的图像色彩。

表 1　补光环境下深度白平衡结果

Tab. 1　Deep White－Balance results under supplementary light environment

参数	预训练模型	迁移学习模型	切片＋迁移学习模型
ΔE	12.09	5.41	4.89
MAE/(°)	27.48	12.26	9.14
MSE	629.85	113.11	89.90

2.2　果实与果串检测结果

优化前后 YOLO v5s 模型在夜间环境与补光环境下的检测结果如表 2 所示。从总体上来看，不管是在夜间环境还是补光环境下，优化后 YOLO v5s－RSBottleneck－CW 始终较原始 YOLO v5s 表现更佳。在两种环境下的最高 P、R、mAP 和 F1 值均来自 YOLO v5s－RSBottleneck－CW，其中夜间环境与补光环境下最高 mAP 分别为 0.902、0.868，均为边框过滤后的结果，较原始 YOLO v5s 分别提高 0.009 和 0.017。

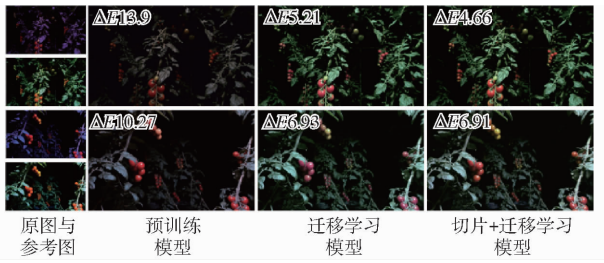

图 9　部分深度白平衡处理结果示例图

Fig. 9　Some examples of Deep White－Balance processing results

图像预处理在不同环境下的应用同样改善了模型的检测结果，如表 2 所示。在夜间环境下，Zero－DCE 增强预处理增强了局部弱光下的纹理细节，使得模型能检测到更多目标，YOLO v5s 的 R 提高 0.013。但是同时也会点亮背景信息，使得模型误检到背景番茄，降低了 P。而通过对结果边框过滤处理后，有效降低背景误检率而不影响召回率。在补光环境下，图像经过深度白平衡颜色校正后，同样使得模型可以检测到更多的目标。在深度白平衡基础上进行光照增强后，YOLO v5s 捕获到更多阴影处的目标，R 得以进一步提高，较未进行任何预处理的结果提升 0.021。图像预处理方法 Zero－DCE 与 Deep White－Balance 对检测模型 YOLO v5s 与 YOLO v5s－RSBottleneck－CW 的 mAP 结果均有提升，特别是在补光环境下改善较明显，如图 10 所示。

表 2　夜间环境与补光环境下不同模型检测结果

Tab. 2　Detection results of different models under night environment and supplementary light environment

环境	预处理方法	精确率	召回率	平均精度均值	F1 值
夜间环境	Original	0.757/0.765	0.911/0.929	0.882/0.900	0.827/0.839
	En	0.727/0.754	0.924/0.931	0.892/0.901	0.813/0.833
	En & filtering	0.744/0.773	0.924/0.931	0.893/0.902	0.824/0.844
补光环境	Original	0.744/0.764	0.859/0.849	0.811/0.811	0.797/0.804
	WB	0.713/0.746	0.872/0.885	0.835/0.845	0.783/0.808
	WB & En	0.709/0.736	0.880/0.899	0.849/0.866	0.784/0.808
	WB & En & filtering	0.735/0.750	0.880/0.899	0.851/0.868	0.799/0.817

注："/"符号左右分别表示 YOLO v5s 与 YOLO v5s－RSBottleneck－CW 结果，Original 表示未预处理，En 表示 Zero－DCE 增强处理，WB 表示 Deep White－Balance 处理，filtering 表示边框过滤处理。下同。

表 3 为两种环境下每个类别的检测结果细节，其均由 YOLO v5s－RSBottleneck－CW 模型获得，并经过了边框过滤处理，其中夜间环境下图像进行了 Zero－DCE 增强处理，补光环境下图像进行了 Deep White－Balance 和 Zero－DCE 增强处理。从表 3 可以看到 Tomato 类平均精度总是最高，这是因为果串

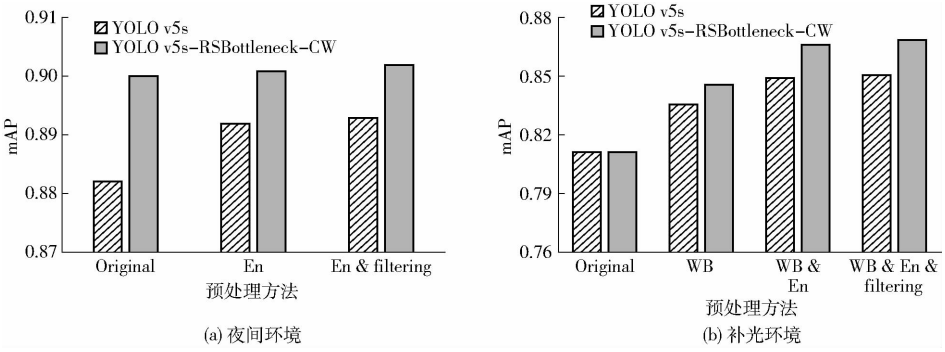

图 10 YOLO v5s 与 YOLO v5s－RSBottleneck－CW 的 mAP 结果

Fig. 10 mAP results of YOLO v5s and YOLO v5s－RSBottleneck－CW

表 3 不同环境下每个类别检测结果

Tab. 3 Detection results for each class in different environments

环境	类别	精确率	召回率	平均精度	平均精度均值
夜间环境	Green	0.787	0.942	0.920	0.902
	Pink	0.703	0.902	0.847	
	Red	0.788	0.919	0.890	
	Tomato	0.813	0.962	0.953	
补光环境	Green	0.699	0.915	0.877	0.868
	Pink	0.682	0.863	0.804	
	Red	0.785	0.874	0.861	
	Tomato	0.835	0.942	0.929	

特征较明显，模型更容易检测到这些目标；而 Pink 类平均精度总是最低，这可能是因为 Pink 类中存在一些处于两种成熟度之间的果实，其颜色纹理等比较相似，某些情况下肉眼也很难分辨出来，而网络难以学习到这种非常细小的差异特征，导致类别判断失败。

YOLO v5s 夜间环境下结果如图 11a 所示。经过 Zero－DCE 增强处理的图像使得模型能够更多地捕捉到图中左下角弱光下的目标，但同时会带来一定的噪声扩大。噪声扩大引起模型将叶片错误检测为番茄果实。引入深度残差收缩模块的 YOLO v5s－RSBottleneck－CW 模型并未出现类似误检，其在扩大的噪声面前表现出更强的适应性。最后经过边框过滤，背景中误检结果和前景中重复检测的果串被消除。补光环境下结果如图 11b 所示。由于图像背景与果实颜色一致性和阴影问题，在原图结果中并未检测出图中左下角目标。Deep White－Balance 处理后 YOLO v5s 和 YOLO v5s－RSBottleneck－CW 均检测到了图中左下角目标，但均出现更多的误检。再经过 Zero－DCE 增强处理

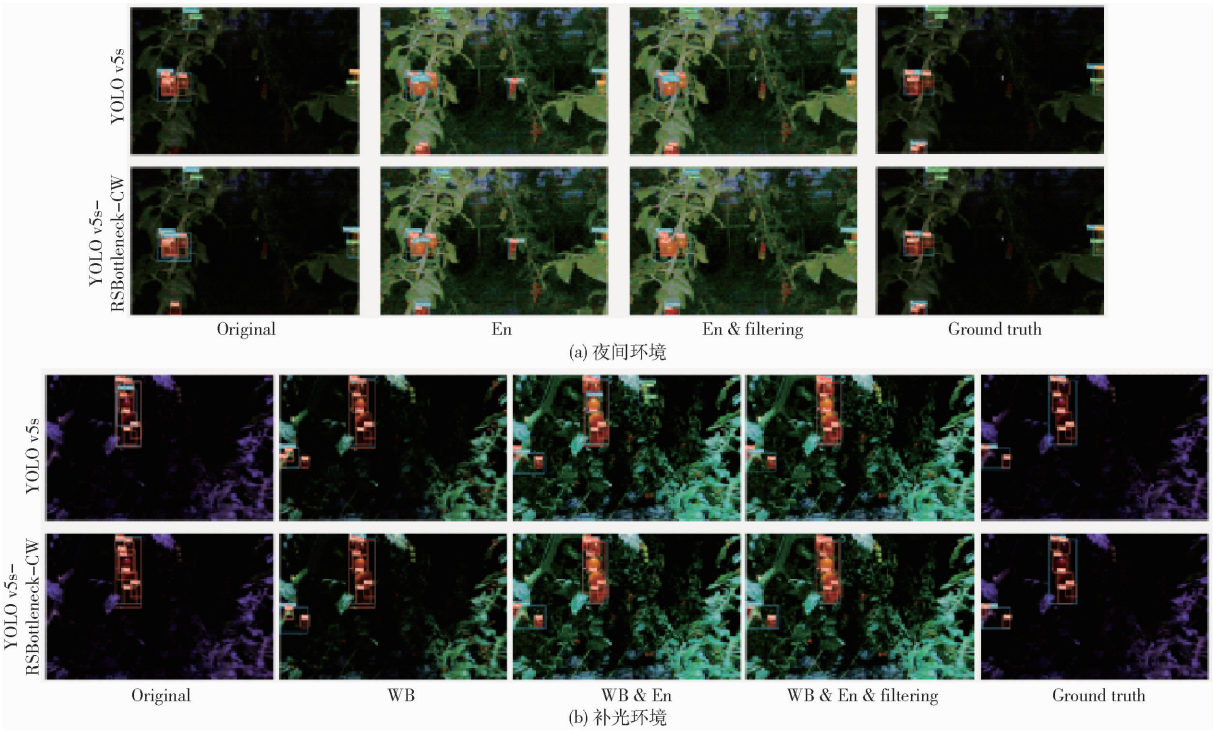

图 11 不同环境下 YOLO v5s 和 YOLO v5s－RSBottleneck－CW 结果示例图

Fig. 11 Examples of YOLO v5s and YOLO v5s－RSBottleneck－CW results in different environments

后，植株茎秆与果实纹理细节增强，左下角目标捕获更加精准。同样的边框过滤进一步降低了背景误检。

表 4 为所提方法与不同 YOLO 系列模型的检测结果对比，所用数据均为深度增强后的图像。夜间环境下 YOLO v8s 具有最高的 mAP 值(0.919)，但其参数量较 YOLO v5s－RSBottleneck－CW 也有所增加，计算成本会更高。补光环境下提出的 YOLO v5s－RSBottleneck－CW 模型在各方面均表现出最好的结果，这是由于补光环境下串番茄图像质量更差，经过偏色校正与光照增强处理后图像中噪声会更加显著，而 YOLO v5s－RSBottleneck－CW 能够很好抑制特征图噪声信息并保留有用信息，相较 YOLO v6s 和 YOLO v8s 表现出更强的噪声适应性能。

表 4　不同 YOLO 模型结果对比

Tab. 4　Comparison of different YOLO model results

环境	模型	精确率	召回率	平均精度均值	参数量
夜间环境	YOLO v6s	0.769	0.918	0.893	1.850×10^7
	YOLO v8s	0.778	0.928	0.919	1.113×10^7
	YOLO v5s－RSBottleneck－CW	0.773	0.931	0.902	7.280×10^6
补光环境	YOLO v6s	0.723	0.857	0.801	1.850×10^7
	YOLO v8s	0.739	0.859	0.861	1.113×10^7
	YOLO v5s－RSBottleneck－CW	0.750	0.899	0.868	7.280×10^6

总的来说，Deep White－Balance 校正了图像色偏，提升了目标检测模型召回率。Zero－DCE 光照增强改善了图像弱光或阴影带来的对比度降低和细节丢失，促进了目标检测模型对目标的特征捕获。光照增强前后 YOLO v5s 提取的特征图如图 12 所示(其中，红色虚线强调的特征图中表现出输入图像黄色虚线区域的特征提取结果差异)。在黄色虚线标注区域中，增强后的图像使得模型更加有效地提取到了该区域特征。结合深度残差收缩模块的 YOLO v5s－RSBottleneck－CW 具有更强的特征提取能力，在结果中也表现得更加出色。特别是当光照增强时带来的扩大噪声，YOLO v5s－RSBottleneck－CW 表现出了更强的鲁棒性。即使在背景点亮情况下，也能在一定程度上避免出现 YOLO v5s 类似的背景误检情况，如图 11b 中的 WB & En 结果所示。

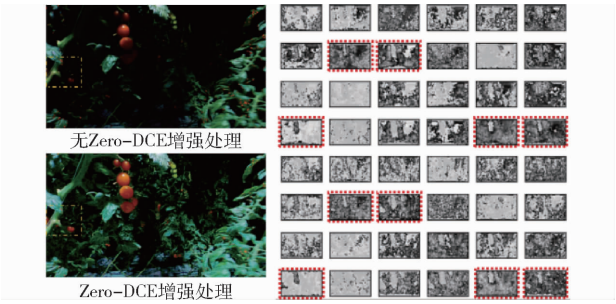

图 12　YOLO v5s 网络第 13 层 C3 模块部分特征图

Fig. 12　Partial feature diagram of C3 module on Layer 13 of YOLO v5s network

利用 GradCAM[32] 观察 YOLO v5s 优化前后对每个类别的关注情况，其中选择 YOLO v5s 的检测层进行激活可视化，以便更好地理解网络，结果如图 13 所示(其中，17、20、23 分别表示 YOLO v5 第 17、20、23 检测层的输出结果，黄色虚线区域表现出模型的结果差异，Green 与 Pink 中 YOLO v5s 出现误检，Red 中出现漏检)。为了更好地分析 YOLO v5s 受噪声信息的影响，图 13 采用的是经过 Deep White－Balance 与 Zero－DCE 处理后的图像，因为光照增强会带来一定的噪声扩大。在 Green 类别中，可以明显看到 YOLO v5s 错误地关注到了叶片和背景果实，而 YOLO v5s－RSBottleneck－CW 则仅关注到了前景中的果实，表现出更强的适应性。在 Pink 类别中 YOLO v5s 部分错误地关注到了 Red 果实，而在 Red 类中又丢失了部分 Red 果实区域。这在 YOLO v5s－RSBottleneck－CW 中均得到了一定改善。最后在 Tomato 类中，模型大目标检测层表现出来的结果并无太大区别。总体来说，通过 GradCAM 热力图可以看到，改进 YOLO v5s 抗干扰性更强，关注区域更加集中，同时对于中小目标的类别判断上也更加

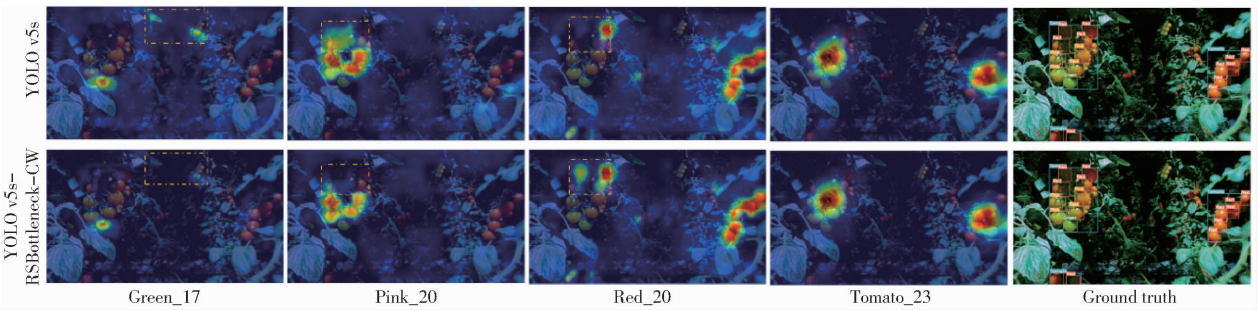

图 13　GradCAM 结果示例

Fig. 13　Example of GradCAM results

准确。这表明嵌入深度残差收缩模块的 YOLO v5s - RSBottleneck - CW 面对具有挑战性的夜间环境可以表现出更佳的检测性能。

为了进一步验证深度残差收缩模块嵌入方法在其他模型中的通用性,将其融合到其他流行的目标检测器中来测试其处理夜间高噪声图像的性能,结果如表 5 所示。表 5 中的对比模型采用以 ResNet50 作为骨干网络的 RetinaNet 与 Faster R - CNN 目标检测器,其中 ResNet50 的基本残差模块用本文所提出的深度残差收缩模块进行替换,对应分别生成了 RetinaNet - RSBottleneck - CW 与 Faster R - CNN - RSBottleneck - CW 模型。为了尽可能保证实验的一致性,对比模型均通过 MMDetection 开源工具库来实现网络的搭建与训练,训练数据采用了深度增强后的图像,同时训练的所有超参数保持相同配置,最后模型性能测试时均考虑使用边框深度过滤算法。由表 5 可知,不管是夜间环境还是补光环境,深度残差收缩模块的嵌入分别使得 RetinaNet 与 Faster R - CNN 出现了不同的性能提升,其中 mAP 与 F1 值结果均出现较为明显改善,但由于参数量增加,模型推理速度有一定下降。相较而言,YOLO v5s - RSBottleneck - CW 依然保持着最好的检测性能,同时模型帧率达到101 f/s,具有很好的实时性能。总的来说,深度残差收缩模块对高噪声图像处理是有效的,其可以很好地拓展到其他模型中以提升高噪声环境下的目标检测性能。

表 5 不同模型检测结果对比

Tab. 5 Comparison of test results from different models

环境	模型	精确率	召回率	平均精度均值	F1 值	帧率/(f·s^{-1})
夜间环境	RetinaNet	0.731	0.908	0.841	0.810	30
	Faster R - CNN	0.741	0.906	0.829	0.815	27
	RetinaNet - RSBottleneck - CW	0.765	0.923	0.862	0.837	24
	Faster R - CNN - RSBottleneck - CW	0.743	0.922	0.856	0.823	23
	YOLO v5s - RSBottleneck - CW	0.773	0.931	0.902	0.844	101
补光环境	RetinaNet	0.694	0.815	0.758	0.750	30
	Faster R - CNN	0.698	0.838	0.773	0.762	27
	RetinaNet - RSBottleneck - CW	0.706	0.837	0.783	0.766	24
	Faster R - CNN - RSBottleneck - CW	0.713	0.866	0.802	0.782	23
	YOLO v5s - RSBottleneck - CW	0.750	0.899	0.868	0.817	101

串番茄不同成熟度对应的最终成熟度识别平均绝对误差 E_R 如表 6 所示,表头中的各成熟度阶段由式(6)结合单果类别标签计算获得。不管是在夜间环境还是补光环境下,当串番茄成熟度为 90% ~100% 时 E_R 总是最小,所提出的方法在该阶段成熟度识别准确率最高。成熟度 90% ~100% 的串番茄基本上为已经成熟或即将成熟,这个阶段正值串番茄采收的最佳时期,而该阶段准确的成熟度识别更有利于机器人对待采摘目标的获取。图 14 为串番茄成熟识别部分结果示例,不管夜间环境或是补光环境下,基本上均能正确检测到果串与不同成熟度果实目标。特别是在果实出现重叠时,也能将果实检测出来并作为成熟度计算依据。由于串番茄成熟度计算时 Green、Pink、Red 3 种果实均参与了计算,在个别果实类别判断错误情况下,计算出来的串番茄成熟度不会发生较大偏差。总的来说在准确检测出果串与果实目标情况下,通过计算成熟果实类别比例的串番茄成熟度表征方法能很好地体现串番茄的生长阶段,同时为机器人采摘决策提供参考。

表 6 不同成熟度阶段的串番茄成熟度识别平均绝对误差 E_R

Tab. 6 Mean absolute error E_R for ripeness recognition of tomato trusses at different ripeness levels

环境	成熟度		
	0% ~25%	25% ~90%	90% ~100%
夜间环境	2.871	8.363	1.837
补光环境	3.325	9.479	1.067

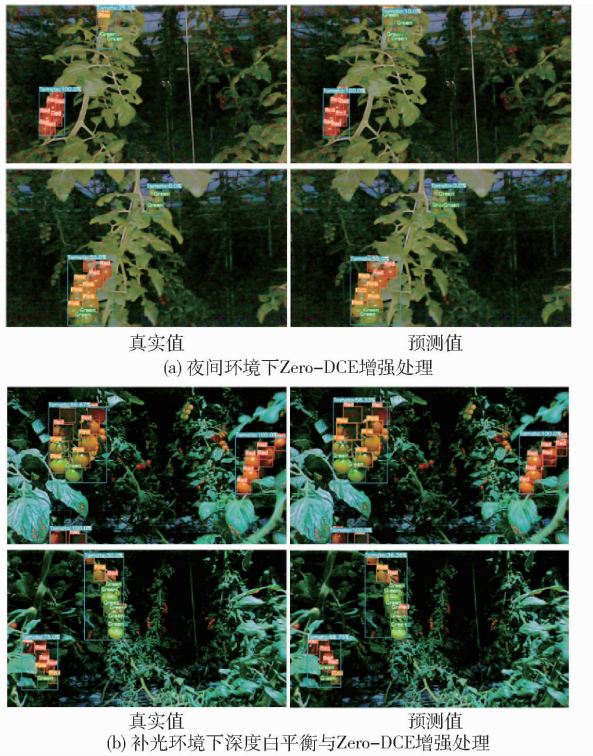

图 14 串番茄成熟度识别结果示例

Fig. 14 Example of tomato trusses ripeness recognition results

3　结论

(1)围绕夜间环境与补光环境两种情况进行讨论,针对图像阴影、偏色以及高噪声问题,提出了一种结合 Deep White－Balance、Zero－DCE 图像深度增强和 RSBottleneck－CW 模块特征抗噪的夜间串番茄成熟度识别方法,为机器人夜间作业提供了视觉技术支持。

(2)对于补光环境下图像严重色偏, Deep White－Balance 实现了图像颜色恢复,测试集平均色差(ΔE)仅为 4.89。对于夜间环境和补光环境下均存在的图像阴影或局部弱光现象,Zero－DCE 光照增强提高了图像对比度,同时增强了弱光与阴影部分的纹理细节。增强后高质量图像给 YOLO v5s 带来了显著的性能增益,在夜间环境和补光环境下 mAP 分别达到 0.892 和 0.849。

(3)对于图像高噪声问题,深度残差收缩模块的嵌入和软阈值化处理使得 YOLO v5s－RSBottleneck－CW 可以有效捕获特征图有用信息并抑制无用信息,从而提升网络检测效果。YOLO v5s－RSBottleneck－CW 结合深度增强处理后,在夜间环境和补光环境下 mAP 分别为 0.901 和 0.866,相对于 YOLO v5s 在未增强处理图像上的 mAP 分别提升 0.019 和 0.055。

(4)边框匹配算法完成了果串与果实检测结果的配对,从而通过不同成熟度果实比例计算出了每串番茄串成熟度水平。同时结合深度信息融合策略,降低了背景误检率,进一步提高了检测结果准确性。在两种环境下经过图像深度增强和边框过滤的 YOLO v5s－RSBottleneck－CW 结果均获得了各自的最高 mAP 和 F1 值,其中夜间环境下最高 mAP 和 F1 值分别为 0.902 和 0.844,补光环境下最高 mAP 和 F1 值分别为 0.868 和 0.817。

参　考　文　献

[1] 刘继展,江应星.农业采摘机器人产业化进程分析与多臂高速化技术走向[J]. 农业机械学报,2024,55(10):1－17.
LIU Jizhan, JIANG Yingxing. Industrialization trends and multi-arm technology direction of harvesting robots[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(10):1－17. (in Chinese)

[2] 郑太雄, 江明哲, 冯明驰. 基于视觉的采摘机器人目标识别与定位方法研究综述[J]. 仪器仪表学报, 2021, 41(9): 28－51.
ZHENG Taixiong, JIANG Mingzhe, FENG Mingchi. Vision based target recognition and location for picking robot:a review[J]. Chinese Journal of Scientific Instrument, 2021, 41(9): 28－51. (in Chinese)

[3] WANG Z, XUN Y, WANG Y, et al. Review of smart robots for fruit and vegetable picking in agriculture[J]. International Journal of Agricultural and Biological Engineering, 2022, 15(1): 33－54.

[4] KARAPANOS I C, CHANDRA M, AKOUMIANAKIS K A, et al. The ripening and quality characteristics of cherry tomato fruit in relation to the time of harvest[C]//Proceedings of the 5th International Conference on Postharvest Unlimited, 2014: 495－500.

[5] TOBARUELA E D, GOMES B L, BONATO V C D, et al. Ethylene and Auxin: hormonal regulation of volatile compound production during tomato (*Solanum lycopersicum* L.) fruit ripening[J]. Frontiers in Plant Science, 2021, 12: 765897.

[6] 姬伟, 吕兴琴, 赵德安, 等. 苹果采摘机器人夜间图像边缘保持的 Retinex 增强算法[J]. 农业工程学报, 2016, 32(6): 189－196.
JI Wei, LÜ Xingqin, ZHAO Dean, et al. Edge-preserving Retinex enhancement algorithm of night vision image for apple harvesting robot[J]. Transactions of CSAE, 2016, 32(6): 189－196. (in Chinese)

[7] 熊俊涛, 郑镇辉, 梁嘉恩, 等. 基于改进 YOLO v3 网络的夜间环境柑橘识别方法[J]. 农业机械学报, 2020, 51(4): 199－206.
XIONG Juntao, ZHENG Zhenhui, LIANG Jiaen, et al. Citrus detection method in night environment based on improved YOLO v3 network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(4): 199－206. (in Chinese)

[8] KOIRALA A, WALSH K B, WANG Z, et al. Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO'[J]. Precision Agriculture, 2019, 20(6): 1107－1135.

[9] LIANG C, XIONG J, ZHENG Z, et al. A visual detection method for nighttime litchi fruits and fruiting stems[J]. Computers and Electronics in Agriculture, 2020, 169 : 105192.

[10] MIRHAJI H, SOLEYMANI M, ASAKEREH A, et al. Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions[J]. Computers and Electronics in Agriculture, 2021, 191: 106533.

[11] 徐新, 穆楠, 张晓龙. 一种鲁棒的夜间图像显著性对象检测模型[J]. 软件学报, 2018, 29(9): 2616－2631.
XU Xin, MU Nan, ZHANG Xiaolong. Robust salient object detection model for nighttime images[J]. Journal of Software, 2018, 29(9): 2616－2631. (in Chinese)

[12] MURUGAN R A, SATHYABAMA B. Object detection for night surveillance using Ssan dataset based modified Yolo algorithm in wireless communication[J]. Wireless Personal Communications, 2023, 128(3): 1813－1826.

[13] TU S, XUE Y, ZHENG C, et al. Detection of passion fruits and maturity classification using Red－Green－Blue depth images [J]. Biosystems Engineering, 2018, 175: 156－167.

[14] TANG C, CHEN D, WANG X, et al. A fine recognition method of strawberry ripeness combining Mask R - CNN and region segmentation[J]. Frontiers in Plant Science, 2023, 14: 1211830.
[15] TEWOLDE F T, NA L, KOUTA S, et al. Nighttime supplemental LED inter-lighting improves growth and yield of single-truss tomatoes by enhancing photosynthesis in both winter and summer[J]. Frontiers in Plant Science, 2016, 7: 448.
[16] PALMITESSA O D, LEONI B, MONTESANO F F, et al. Supplementary Far-Red light did not affect tomato plant growth or yield under mediterranean greenhouse conditions[J]. Agronomy, 2020, 10(12): 1849.
[17] HUANG Y P, WANG T H, BASANTA H. Using Fuzzy Mask R - CNN model to automatically identify tomato ripeness[J]. IEEE Access, 2020, 8: 207672 - 207682.
[18] SU F, ZHAO Y, WANG G, et al. Tomato maturity classification based on SE - YOLOv3 - MobileNetV1 network under nature greenhouse environment[J]. Agronomy, 2022, 12(7): 1638.
[19] TENORIO G L, CAARLS W. Automatic visual estimation of tomato cluster maturity in plant rows[J]. Machine Vision and Applications, 2021, 32(4): 78.
[20] MIAO Z, YU X, LI N, et al. Efficient tomato harvesting robot based on image processing and deep learning[J]. Precision Agriculture, 2023, 24(1): 254 - 287.
[21] RADZEVIČIUS A, KARKLELIENĖ R, VIŠKELIS P, et al. Tomato (*Lycopersicon esculentum* Mill.) fruit quality and physiological parameters at different ripening stages of Lithuanian cultivars[J]. Agronomy Research, 2009, 7(2): 712 - 718.
[22] 李军, 刘凤军, 徐君. 不同发光二极管(LED)光源补光对大棚秋番茄植株生长及果实产量和品质的影响[J]. 江苏农业学报, 2011, 27(6): 1339 - 1343.
LI Jun, LIU Fengjun, XU Jun. Effects of light emitting diode supplement on growth, fruit quality and yield of autumn tomato in greenhouse[J]. Jiangsu Journal of Agricultural Sciences, 2011, 27(6): 1339 - 1343. (in Chinese)
[23] 魏守辉, 肖雪梅, 钟源, 等. 日光温室不同时段补光对番茄果实品质及挥发性物质影响[J]. 农业工程学报, 2020, 36(8): 188 - 196.
WEI Shouhui, XIAO Xuemei, ZHONG Yuan, et al. Effects of supplemental illumination in different periods on the quality and volatile compounds of tomato fruit in solar greenhouse[J]. Transactions of CSAE, 2020, 36(8): 188 - 196. (in Chinese)
[24] AFIFI M, BROWN M S. What else can fool deep learning? Addressing color constancy errors on deep neural network performance[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 243 - 252.
[25] AFIFI M, BROWN M S. Deep white-balance editing[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1397 - 1406.
[26] RONNEBERGER O, FISCHER P, BROX T. U - net: convolutional networks for biomedical image segmentation[C]// Proceedings of the Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, 2015: 234 - 241.
[27] GUO C, LI C, GUO J, et al. Zero-reference deep curve estimation for low-light image enhancement[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1780 - 1789.
[28] ZHAO M, ZHONG S, FU X, et al. Deep residual shrinkage networks for fault diagnosis[J]. IEEE Transactions on Industrial Informatics, 2019, 16(7): 4681 - 4690.
[29] 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.
[30] WANG Q, WU B, ZHU P, et al. ECA - Net: efficient channel attention for deep convolutional neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534 - 11542.
[31] XIAO Y, JIANG A, YE J, et al. Making of night vision: object detection under low-illumination[J]. IEEE Access, 2020, 8: 123075 - 123086.
[32] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 618 - 626.