doi:10.6041/j.issn.1000-1298.2024.05.023

基于频域数据增强与轻量化 YOLO v7 模型的成熟期香梨目标检测方法

郑文轩　杨　瑛

(江苏第二师范学院物理与电子信息学院, 南京 210013)

摘要: 为实现香梨自动化采摘,本文以 YOLO v7 - S 为基础模型,针对果园中香梨果实、果叶和枝干之间相互遮挡,不易精准检测的问题,设计了一种轻量化香梨目标检测 M - YOLO v7 - SCSN + F 模型。该模型采用 MobileNetv3 作为骨干特征提取网络,引入协同注意力机制(Coordinate attention, CA)模块,将 YOLO v7 - S 中的损失函数 CIoU 替换为 SIoU,并联合 Normalized Wasserstein distance (NWD)小目标检测机制,以增强网络特征表达能力和检测精度。基于傅里叶变换(Fourier transform,FT)的数据增强方法,通过分析图像频域信息和重建图像振幅分量生成新的图像数据,从而提高模型泛化能力。实验结果表明,改进的 M - YOLO v7 - SCSN + F 模型在验证集上的平均精度均值(mAP)、精确率和召回率分别达到 97.23%、97.63% 和 93.66%,检测速度为 69.39 f/s,与 Faster R - CNN、SSD、YOLO v3、YOLO v4、YOLO v5s、YOLO v7 - S、YOLO v8n、RT - DETR - R50 模型在验证集上进行性能比较,其平均精度均值(mAP)分别提高 14.50、26.58、3.88、2.40、1.58、0.16、0.07、0.86 个百分点。此外,改进的 M - YOLO v7 - SCSN + F 模型内存占用量与 YOLO v8n 和 RT - DETR - R50 检测模型对比减少 16.47、13.30 MB。本文提出的检测模型对成熟期香梨具有很好的目标检测效果,为背景颜色相近小目标检测提供参考,可为香梨自动化采摘提供有效的技术支持。

关键词: 香梨; 目标检测; YOLO v7; 数据增强; 傅里叶变换; 注意力机制

中图分类号: S24; TP391.4　　**文献标识码**: A　　**文章编号**: 1000-1298(2024)05-0244-10　　**OSID**:

Mature Stage Pear Detection Method Based on Frequency Domain Data Augmentation and Lightweight YOLO v7 Model

ZHENG Wenxuan　YANG Ying

(*School of Physics and Electronic Information Engineering*, *Jiangsu Second Normal University*, *Nanjing* 210013, *China*)

Abstract: In the practice of modern agricultural production, the method of harvesting agricultural products is gradually shifting towards mechanization and intelligence. An increasing number of robots are being introduced into actual production and progressively replacing traditional manual labor. However, in natural environments, factors such as weather, lighting, the similarity in color between fruits and their backgrounds, and mutual occlusion between fruits and branches significantly increased the difficulty of fruit target detection. To accurately detect pears in natural environments, a lightweight pear detection method M - YOLO v7 - SCSN + F was designed based on the YOLO v7 - S foundational model. This model introduced MobileNetv3 into the YOLO v7 - S model as its backbone feature extraction network, thereby reducing the number of parameters in the network. It incorporated a coordinate attention (CA) mechanism in the model's feature fusion layer to enhance the network's feature representation capabilities. The loss function CIoU in YOLO v7 - S was replaced with SIoU, which was used in conjunction with the normalized Wasserstein distance (NWD) mechanism for small target detection, further improving the detection accuracy for fragrant pears. Based on the Fourier transform (FT) data augmentation method, new image data was generated by analyzing the frequency domain information of images and reconstructing the amplitude components, thereby enhancing the model's generalization ability. Experimental results

收稿日期: 2024 - 02 - 04　修回日期: 2024 - 03 - 25

基金项目: 国家自然科学基金项目(51466014)

作者简介: 郑文轩(1980—),男,教授,博士,主要从事图像分析和计算机视觉研究,E-mail: wenxuanzhengdx@163.com

通信作者: 杨瑛(1979—),女,教授,博士,主要从事农业信息技术研究,E-mail: 591908603@qq.com

showed that the improved M－YOLO v7－SCSN＋F model achieved mean average precision (mAP), precision, and recall rates of 97.23%, 97.63% and 93.66%, respectively, on the validation set, with a detection speed of 69.39 f/s. The proposed detection model improved performance compared with Faster R－CNN, SSD, YOLO v3, YOLO v4, YOLO v5s, YOLO v7－S, YOLO v8n and RT－DETR－R50 models on the validation set, with mean average precision (mAP) enhancements of 14.50, 26.58, 3.88, 2.40, 1.58, 0.16, 0.07 and 0.86 percentage points, respectively. Furthermore, the improved M－YOLO v7－SCSN＋F model reduced its parameter count by 16.47 MB and 13.30 MB, respectively, when compared with the advanced YOLO v8n and RT－DETR－R50 detection models. The detection model introduced demonstrated a high degree of effectiveness in target detection for mature pears, offering a reference for detecting small objects with backgrounds of similar color, and provided effective technical support for the automation of pear harvesting.

Key words: pear; object detection; YOLO v7; data augmentation; Fourier transform; attention mechanism

0　引言

2022 年,我国香梨产量突破 1.8×10^7 t,成为果农主要的经济收入来源之一。目前香梨采摘方式主要依靠手工采摘,费时费力,效率低,很难满足作业周期短、采摘任务重、采摘需求急的要求[1]。在现代农业生产实践中,农产品采摘方式正逐渐向机械化和智能化方向发展,更多的机器人开始投入生产实际,并逐渐取代了传统的人工劳作。然而,在自然环境中,由于天气、光线、果实颜色与背景相似,果实和枝干之间相互遮挡等问题的影响,极大地增加了水果目标检测难度。

随着计算机算力的不断增强,深度学习技术在目标检测这一领域展现出了显著的优势[2]。相较于传统的目标检测方法,深度卷积神经网络能够自动地从其训练数据中提取多个层次的特征,使得该模型展现出更高的泛化和特征提取能力[3-4]。YONG 等[5]提出了一种多簇猕猴桃果实识别算法,其识别准确率达到 94.78%。黄小玉等[6]提出桃子检测模型,该方法是以 MTCNN(多任务卷积神经网络算法)为主干的一种网络,对成熟桃子的检测准确率达到 89.3%。HE 等[7]提出 SPP－Net 网络,从之前的单一尺度输入变为了效率更高的多尺度输入。REN 等[8]提出了 Faster R－CNN 网络弥补目标检测速度慢的问题。REDMON 等[9]提出了速度更快的网络模型 YOLO,但这种方法检测精度低。后续研究者提出了 YOLO v2、YOLO v3、YOLO v4 和 YOLO v5 等来改进检测精度低的问题,但这些网络模型在小目标检测上表现欠佳。随着目标检测技术的不断发展,TIAN 等[10]提出了一种改进的 YOLO v3 模型,用于在不同光照条件、复杂背景以及遮挡等情况下检测不同成熟度的苹果。王立舒等[11]提出一种改进 YOLO v4－Tiny 的蓝莓成熟度识别方法,该方法在遮挡与光照不均等复杂场景中平均精度较高。LÜ 等[12]提出了一种轻量级的目标检测 YOLO v5－CS 模型,在 YOLO v5 网络中加入了注意力模块,自然环境中对绿色柑橘识别准确率达 98.23%。YOLO v7 引入了结构重参数化和动态标签分配,使得模型具有更快的速度和更高的精度,同时具备更强的鲁棒性,能很好地应对光线不足和遮挡情况下的目标检测,在智慧农业各领域中得到了广泛的应用[13]。虽然,目标检测在农副产品检测研究领域已取得了良好的效果,但对于水果目标检测而言,其精度与速度仍需提升,当检测距离远且有遮挡时,会产生漏检与误检等现象。

针对现有目标检测模型对自然环境中香梨检测精度不佳的问题,本文提出一种改进的 M－YOLO v7－SCSN＋F 成熟期香梨目标检测模型。该模型将 YOLO v7－5 的骨干网络替换为轻量级的 MobileNetv3 网络,并在特征融合层中引入协同注意力机制(Coordinate attention, CA)模块,以增强网络的特征表达能力,通过实验对比将原 YOLO v7 中的损失函数 CIoU 替换为 SIoU,并与 Normalized Wasserstein distance (NWD)小目标检测机制联合使用,以进一步提高香梨检测精度。同时使用傅里叶变换的数据增强方法[14],分析图像频域信息,重新构建图像的振幅分量来生成新的图像信息。

1　实验数据

1.1　图像采集

所用的香梨数据集拍摄于塔里木大学园艺实验站新梨 7 号实验区,研究对象为成熟期新梨 7 号香梨。实验所用图像使用佳能 EOS 77D 型相机拍摄,针对智能机器人采摘梨果时遇到的采摘距离不同,光照强弱不均的情况,数据采集设计了远距离(距离为 50～100 cm)和近距离(距离为 10～30 cm)2 种图像拍摄模式,拍摄时间选择在 07:00—08:00、13:00—14:00、18:00—19:00 3 个时间段,拍摄时间

选择在 8 月 15—30 日,此时段为香梨成熟期但还未采摘。实验共采集香梨图像 4 539 幅,其中近距图像 2 270 幅,远距图像 2 269 幅,采集图像信息见表 1、图 1。

表 1 采集图像信息

Tab. 1 Details of acquired images

采集距离	采集条件	图像数量/幅
近距离	顺光	676
	逆光	581
	枝叶遮挡	553
	光照不均	460
远距离	顺光	663
	逆光	612
	枝叶遮挡	576
	光照不均	418



(a) 近景 (b) 远景 (c) 顺光 (d) 逆光

图 1 不同场景香梨图像

Fig. 1 Images of pear in different scenes

练 2 种方法进行。冻结骨干网络训练时初始学习率设为 0.01,批量大小(Batchsize)设为 16,动量设为 0.93;解冻后网络训练学习率设为 0.01,Batchsize 设为 8,动量设为 0.93,迭代次数为 300。

2 研究方法

2.1 香梨检测模型构建

为了解决在自然环境下香梨果实受光照和其他环境因素的影响,检测效果不佳的问题,本研究对 YOLO v7 - S 模型进行了改进,模型将 MobileNetv3 引入 YOLO v7 - S 模型中作为其骨干特征提取网络,同时在特征融合层中引入协同注意力机制(Coordinate attention, CA)模块,将 YOLO v7 - S 中的损失函数 CIoU 替换为 SIoU,与 Normalized Wasserstein distance (NWD)小目标检测机制联合使用。实验采用傅里叶变换(Fourier transform,FT)的数据增强方法,进一步提升模型泛化能力。为描述方便,将本改进方法简称为 M - YOLO v7 - SCSN + F。改进的 M - YOLO v7 - SCSN + F 模型如图 2 所示。

2.1.1 YOLO v7 - S 模型主干网络替换

为了使改进的 YOLO v7 - S 模型更符合生产实际,使用 MobileNetv3 网络作为 YOLO v7 模型的骨干网络实现特征提取。MobileNetv3 网络与传统卷积神经网络相比,在准确率小幅降低的前提下大大减少了模型参数量与运算量[17],实现模型轻量化,更有利于部署在移动端,其网络参数设计如表 2 所示。MobileNetv3 在主干网络中加入了注意力机制(Squeezeand excitation networks, SE)模块[18],使用深度可分离卷积(Depthwise separable convolutions)和逆残差结构(Inverted residual with linear bottleneck)。同时使用 H - Swish 激活函数代替 Swish 激活函数,提高了模型性能。

2.1.2 引入 SE 通道注意力机制和 CA 协同注意力机制

在 MobileNetv3 主干网络中加入了 SE 注意力机制模块,SE 注意力机制主要考虑通道注意力(Channel attention),如图 3 所示。在 YOLO v7 - S 特征融合层引入协同注意力机制模块,该注意力机制不仅可以把通道注意力和空间注意力进行有效的结合,同时在通道注意力中融入了位置信息,避免全局 pooling - 2D 操作造成的位置信息丢失,如图 4 所示。

2.1.3 SIoU 损失函数与 NWD 小目标检测机制联用

YOLO v7 - S 模型算法中,预测框回归损失采用 CIoU,但 CIoU 并没有考虑到预测框与真实框方向的不匹配。这种方向的不匹配,可能导致收敛速度减慢和收敛效率低。针对 CIoU 存在的问题,经过实验对比,采用 SIoU 损失函数替换 YOLO v7 - S 使用的 CIoU 损失函数[19]。SIoU 损失函数进一步考虑

1.2 香梨图像数据集制作

为了改善网络训练效果和提高模型泛化能力,采用随机翻转、剪裁、Mosaic 算法和 Mixup 算法[15]、傅里叶变换频域数据增强算法[16]等数据增广方法增加样本数量。图像增广后数据集共 8 631 幅,以 7∶2∶1划分训练集(6 042 幅)、测试集(1 726 幅)和验证集(863 幅)进行模型训练和测试。

1.3 实验环境配置

使用 Win10 操作系统,GPU 型号为 NVIDIA GeForce RTX - 3090,处理器型号为 Intel(R) Core(TM) i7 - 12700KF 3.60 GHz,深度学习框架为 PyTorch 2.0,编程平台为 PyCharm,编程语言为 Python 3.8,所有对比算法均在相同环境下运行。为提高网络训练效率,本实验采用冻结训练和解冻训

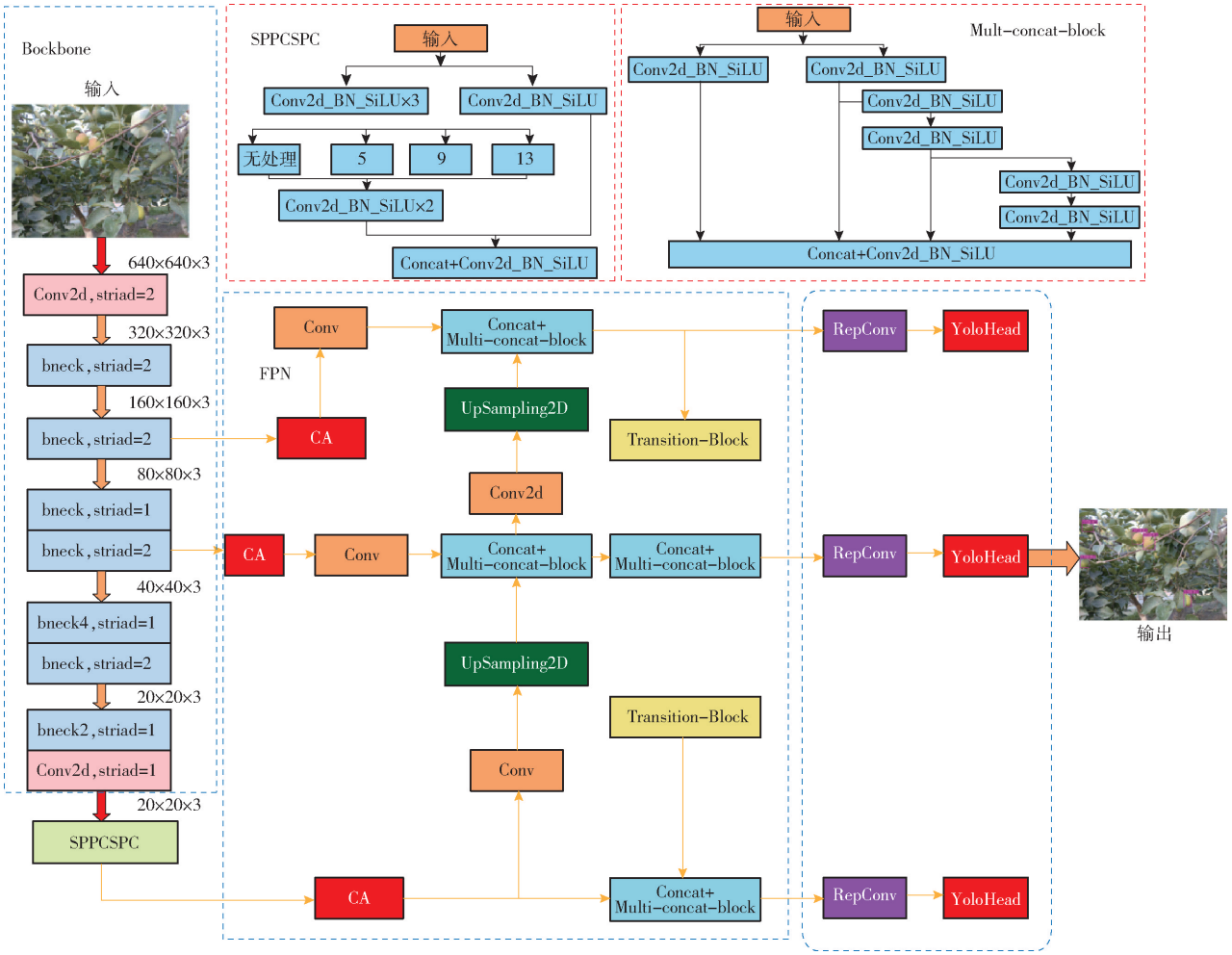


图 2　改进的 M－YOLO v7－SCSN＋F 模型图

Fig. 2　Diagram of improved M－YOLO v7－SCSN＋F network model

表 2　MobileNetv3(small) 网络参数

Tab. 2　MobileNetv3(small) network parameters

输入	操作	升维通道	输出通道	SE	N L	步长
224×224×3	Conv2d,3×3		16		HS	2
112×112×16	bneck,3×3	16	16	√	RE	2
56×56×16	bneck,3×3	72	24		RE	2
28×28×24	bneck,5×5	88	24		RE	1
28×28×24	bneck,5×5	96	40	√	HS	2
14×14×40	bneck,5×5	240	40	√	HS	1
14×14×40	bneck,5×5	240	40	√	HS	1
14×14×40	bneck,5×5	120	48	√	HS	1
14×14×48	bneck,5×5	144	48	√	HS	1
14×14×48	bneck,5×5	244	96	√	HS	2
7×7×96	bneck,5×5	576	96	√	HS	1
7×7×96	bneck,5×5	576	96	√	HS	1
7×7×96	Conv2d,1×1		576	√	HS	1
7×7×576	pool,7×7					1
7×7×576	Conv2d,1×1		1024		HS	1
1×1×1 024	Conv2d,1×1		k			1

注："√"表示本层引入 SE 模块。

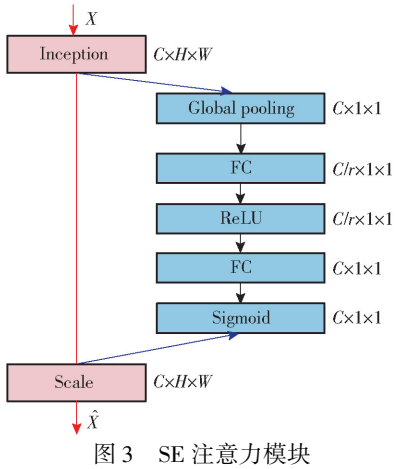


图 3　SE 注意力模块

Fig. 3　SE attention module

了真实框和预测框之间的向量角度，重新定义相关损失函数，SIoU 主要包含：角度损失(Angle cost)、距离损失(Distance cost)、形状损失(Shape cost)、IoU 损失(IoU cost)，图 5 为 SIoU 参数示意图。

角度损失(Angle cost)计算公式为

$$\Lambda = 1 - 2\sin^2\left(\arcsin\frac{c_h}{\sigma} - \frac{\pi}{4}\right) = \cos\left(2\left(\arcsin\frac{c_h}{\sigma} - \frac{\pi}{4}\right)\right) \tag{1}$$

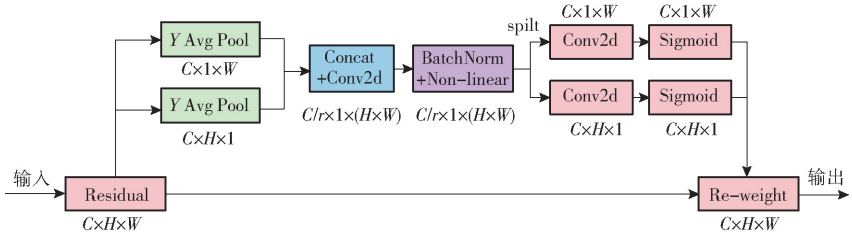


图 4 CA 注意力模块

Fig. 4 CA attention module

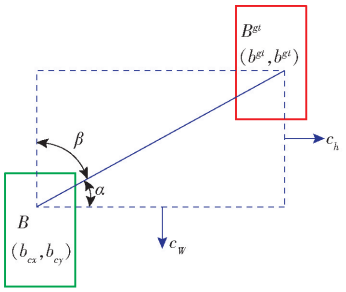


图 5 SIoU 参数示意图

Fig. 5 Schematic of SIoU parameter

其中

$$\frac{c_h}{\sigma}=\sin\alpha \tag{2}$$

根据角度损失 Λ 公式,定义距离损失 Δ 计算公式为

$$\Delta=\sum_{t=x,y}(1-\mathrm{e}^{-\gamma\rho_t})=2-\mathrm{e}^{-\gamma\rho_x}-\mathrm{e}^{-\gamma\rho_y} \tag{3}$$

其中

$$\begin{cases}\rho_x=\left(\dfrac{b^{gt}_{cx}-b_{cx}}{c_w}\right)^2\\ \rho_y=\left(\dfrac{b^{gt}_{cy}-b_{cy}}{c_h}\right)^2\\ \gamma=2-\Lambda\end{cases} \tag{4}$$

形状损失 Ω 定义为

$$\Omega=\sum_{t=w,h}(1-\mathrm{e}^{-w_t})^{\theta}=(1-\mathrm{e}^{-w_w})^{\theta}+(1-\mathrm{e}^{-w_h})^{\theta} \tag{5}$$

其中

$$\begin{cases}w_w=\dfrac{|w-w^{gt}|}{\max(w,w^{gt})}\\ w_h=\dfrac{|h-h^{gt}|}{\max(h,h^{gt})}\end{cases} \tag{6}$$

式中 w、h、w^{gt}、h^{gt} 分别为预测框和真实框的宽和高。θ 控制对形状损失的关注程度,为避免过于关注形状损失而降低对预测框的移动,θ 范围为 2 ~ 6。

SIoU 损失定义为

$$Loss_{\mathrm{SIoU}}=1-IoU+\frac{\Delta+\Omega}{2} \tag{7}$$

实验中将 NWD[20] 小目标检测方法引入香梨的目标检测中,将 NWD 和 SIoU 联合使用,实验选择其各占比 50% 的策略进行,即:Loss_loc = (1 - IoU_ratio)(1.0 - NWD) + IoU_ratio(1.0 - SIoU),这个策略既可以提升目标检测精度,同时又保证模型检测速度。

2.2 基于傅里叶变换的频域数据增强

YOLO v7 - S 模型中使用 Mosaic 算法和 Mixup 算法等数据增强方法。这些算法更注重样本的空域表现,但忽略了频域信息。图像频域信息可体现灰度变化的速度,保留图像轮廓信息,对图像进行频域数据增强处理能有效降低外部环境对检测精度的影响。为此,本文引入基于傅里叶变换的频域数据增强算法,以增强 YOLO v7 模型鲁棒性。

信号在时域和频域之间的变换依赖离散傅里叶变换(Discrete Fourier transform, DFT),二维离散傅里叶变换广泛应用于图像变换,它能将空间域的图像转换到频域上进行研究。对于尺寸为 $M\times N$ 的图像 $f(x,y)$,傅里叶变换可以表示为

$$F(\mu,\upsilon)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\mathrm{e}^{-\mathrm{j}2\pi\left(\frac{\mu x}{M}+\frac{\upsilon y}{N}\right)}$$
$$(\mu=0,1,\cdots,M-1;\upsilon=0,1,\cdots,N-1) \tag{8}$$

其中 μ、υ 为频率变量,μ 对应 x 轴,υ 对应 y 轴,$F(\mu,\upsilon)$ 是图像 $f(x,y)$ 傅里叶变换后的频谱。傅里叶反变换能将频域系数变换回空间域,获得重建图像,相应的二维傅里叶反变换为

$$F(x,y)=\frac{1}{MN}\sum_{\mu=0}^{M-1}\sum_{\upsilon=0}^{N-1}f(\mu,\upsilon)\mathrm{e}^{-\mathrm{j}2\pi\left(\frac{\mu x}{M}+\frac{\upsilon y}{N}\right)}$$
$$(x=0,1,\cdots,M-1;y=0,1,\cdots,N-1) \tag{9}$$

式(8)可分离为

$$F(\mu,\upsilon)=\sum_{x=0}^{M-1}\mathrm{e}^{\frac{-\mathrm{j}2\pi\mu x}{M}}\sum_{y=0}^{N-1}f(x,y)\mathrm{e}^{\frac{-\mathrm{j}2\pi\upsilon y}{N}} \tag{10}$$

式(9)可分离为

$$F(x,y)=\frac{1}{MN}\sum_{\mu=0}^{M-1}\mathrm{e}^{\frac{\mathrm{j}2\pi\mu x}{M}}\sum_{y=0}^{N-1}F(\mu,\upsilon)\mathrm{e}^{\frac{\mathrm{j}2\pi\upsilon y}{N}} \tag{11}$$

傅里叶变换幅度谱和相位谱分别为

$$|F(\mu,\upsilon)|=\sqrt{R^2(\mu,\upsilon)+I^2(\mu,\upsilon)} \tag{12}$$

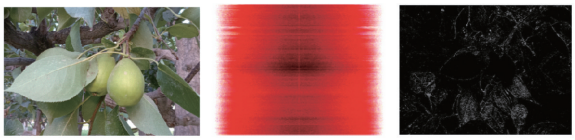
$$\phi(\mu,\upsilon)=\arctan(I(\mu,\upsilon)/R(\mu,\upsilon)) \tag{13}$$

其中,R 和 I 表示傅里叶变换系数实部和虚部。

由式(10)、(11)可知,傅里叶变换及傅里叶反变换的核函数具有分离性,且指数部分具有周期性。快速傅里叶变换基于逐次加倍的思想消除了傅里叶变换中的多余计算,提高了计算速率。

为分析快速傅里叶变换的效果及可行性,从香

梨数据集的训练集中随机选择 1 幅图像。对图像进行快速傅里叶变换，分离图像中的振幅和相位信息，分别进行可视化。振幅、相位分量可视化结果如图 6 所示。



(a) 原图　(b) 振幅分量信息　(c) 相位分量信息

图 6　傅里叶变换振幅分量、相位分量信息图

Fig. 6　Diagrams of Fourier transform amplitude and phase components

从图 6 可以看出，振幅分量给出了图像对比度信息，相位分量给出了图像轮廓信息。在目标检测任务中，相位分量相较于振幅分量更为重要，因为图像轮廓信息能帮助模型更好地提取特征，提升检测精度。

通过分离图像的相位分量，并构建不同的振幅信息来生成新图像。新图像与原图像轮廓相同，但幅值不同，以此达到数据增强的目的，从而增加训练样本多样性，提高所训练模型泛化性。需要注意的是，新的振幅信息来自训练样本中的其他图像，并非随机产生，具体过程描述为：

将 2 幅图像 $f(x_1,y_1)$ 和 $f(x_2,y_2)$ 分别进行快速傅里叶变换

$$F(\mu_1,v_1)=\mathrm{FFT}(f(x_1,y_1)) \tag{14}$$

$$F(\mu_2,v_2)=\mathrm{FFT}(f(x_2,y_2)) \tag{15}$$

取对应的振幅 $|F(\mu_1,v_1)|$ 和 $|F(\mu_2,v_2)|$，随机产生一个 $[0,1]$ 内的随机因子 α，以这个随机因子 α 和 $(1-\alpha)$ 为权重，组合得到两个新的振幅为

$$|F'(\mu_1,v_1)|=\alpha|F(\mu_2,v_2)|+(1-\alpha)|F(\mu_1,v_1)| \tag{16}$$

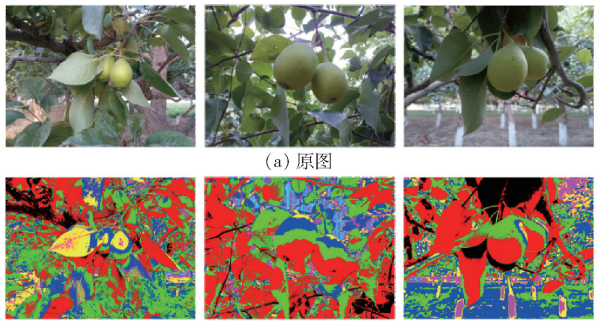
$$|F'(\mu_2,v_2)|=\alpha|F(\mu_1,v_1)|+(1-\alpha)|F(\mu_2,v_2)| \tag{17}$$

重新组合相位分量和新的振幅分量，然后进行反傅里叶变换，还原为时域图像。从香梨数据集中随机选取 3 幅图像，当执行上述数据增强算法并取随机因子 α 为 0.8 后，得到的增强样本如图 7 所示。

由图 7 可以看出，经过傅里叶变换数据增强的图像保留了原有的轮廓信息，但交换了各自的色彩和对比度信息。这些图像由于色彩特征的改变成为新的图像样本，从而达到数据增强效果。

2.3　模型评价指标

在模型检测精度方面，选用精确率（Precision）、召回率（Recall）和平均精度均值（Mean average precision，mAP）作为评价指标；在模型检测性能方面，选用检测速度和模型内存占用量作为评价指标，



(a) 原图

(b) 傅里叶变换数据增强

图 7　傅里叶变换数据增强效果

Fig. 7　Effectiveness diagrams of Fourier transform data augmentation

这些评价指标的具体计算方法见文献[21－22]。

3　实验结果与分析

3.1　YOLO v7 模型性能对比

对比 YOLO v7－S、YOLO v7－M、YOLO v7－L 3 个模型的精确率、召回率、平均精度均值、模型内存占用量和检测速度，见表 3，并综合考虑选择模型内存占用量最小、检测速度最快的 YOLO v7－S 作为基础模型进行改进。

表 3　不同 YOLO v7 检测模型在验证集中的性能结果

Tab. 3　Performance comparison of different YOLO v7 detection models in validation set

模型	图像尺寸/（像素×像素）	平均精度均值/%	精确率/%	召回率/%	模型内存占用量/MB	检测速度/（f·s^{-1}）
YOLO v7－S	640×640	95.65	96.21	93.53	37.27	75.61
YOLO v7－M	640×640	94.83	96.55	91.89	38.53	73.41
YOLO v7－L	640×640	94.33	96.16	92.11	39.21	73.97

3.2　YOLO v7 模型消融实验

设计 5 组消融实验，分别为 YOLO v7－S、YOLO v7－S＋CA 模型（特征融合层中加入 CA 注意力模块）、MobileNetv3－YOLO v7－S 模型（替换 YOLO v7－S 骨干网络为 MobileNetv3）、改进后的 M－YOLO v7－SCSN 模型（MobileNetv3－YOLO v7－S＋CA＋SIoU＋NWD）和 M－YOLO v7－SCSN＋F 模型（M－YOLO v7－SCSN＋傅里叶变换数据增强）的检测性能，采用模型平均精度均值（mAP）和检测速度两个评价指标，在验证集对模型进行评价，评价结果如表 4 所示。从表 4 可知，使用 MobileNetv3 替换 YOLO v7－S 骨干网络，虽其平均精度均值（mAP）和检测速度略有下降，但其模型内存占用量比 YOLO v7－S 减少 27.8%。在检测精度方面，本实验设计的 M－YOLO v7－SCSN＋F 模型平均精度均值高于其他 4 组网络模型。

表 4 YOLO v7 模型消融实验结果

Tab. 4 Ablation experiment results for YOLO v7 model

改进策略	图像尺寸/（像素×像素）	平均精度均值/%	模型内存占用量/MB	检测速度/（f·s^{-1}）
YOLO v7－S	640×640	95.65	37.27	75.61
YOLO v7－S＋CA	640×640	96.69	37.28	75.11
MobileNetv3－YOLO v7－S	640×640	96.23	26.91	72.93
M－YOLO v7－SCSN	640×640	96.91	28.11	71.65
M－YOLO v7－SCSN＋F	640×640	97.77	28.13	69.39

3.3 不同损失函数性能对比

为了分析不同损失函数的性能表现，将 YOLO v7－S 中原有的损失函数进行替换，表 5 为 CIoU、DIoU[23]、GIoU[24] 和 SIoU[25] 4 种损失函数应用 YOLO v7－S 中的性能表现。图 8 为使用 4 种损失函数模型损失值曲线。由图 8 可以看出，使用 CIoU、DIoU、GIoU、SIoU 损失函数训练模型损失值均可快速收敛并维持在较低损失值。

表 5 不同损失函数性能比较

Tab. 5 Performance comparison of different loss functions

损失函数	图像尺寸/（像素×像素）	平均精度均值/%	精确率/%	召回率/%	训练时长/h
CIoU	640×640	95.26	95.85	91.93	6.27
DIoU	640×640	93.15	93.98	89.56	7.31
GIoU	640×640	93.49	93.73	89.42	6.69
SIoU	640×640	95.54	96.23	91.83	5.46

由表 5 可知，与 CIoU 相比使用 SIoU 损失函数模型的平均精度均值（mAP）和精确率分别高 0.28、0.38 个百分点，召回率低 0.1 个百分点，训练时间短 0.81 h。与 DIoU 相比，使用 SIoU 损失函数的平均精度均值（mAP）、精确率和召回率分别高 2.39、2.25、2.27 个百分点，训练时长缩短 1.85 h；与 GIoU 相比，使用 SIoU 损失函数的平均精度均值（mAP）、精确率和召回率分别提高 2.05、2.5、2.41 个百分点，训练时长缩短 1.23 h；根据上述的分析可知，使用 SIoU 损失函数进行模型训练，综合优势较为明显，检测精度最高。

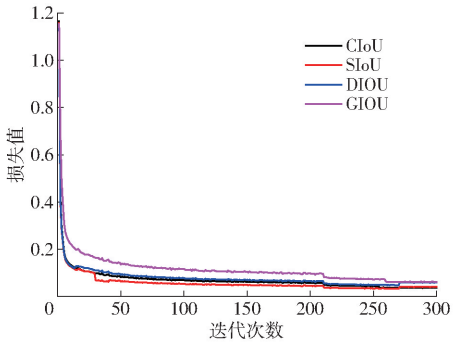


图 8 不同损失函数损失值随迭代次数变化曲线

Fig. 8 Curves of loss values with iteration epochs for different loss functions

3.4 香梨果实图像检测结果

图 9 为不同距离、有无遮挡、不同光照条件下 YOLO v7－S 和改进 M－YOLO v7－SCSN＋F 模型香梨检测结果。

由图 9 可知，无论是在近距离还是远距离无遮挡顺光、逆光的情况下，YOLO v7－S 模型以及 M－YOLO v7－SCSN＋F 模型均可准确识别香梨果实，没有漏检和误检的情况。在有枝叶遮挡的情况下，YOLO v7－S 的检测不够准确，图 9a～9f 中均出现了漏检和误检现象。M－YOLO v7－SCSN＋F 模型仅在光照不均匀且有相邻香梨遮挡（图 9c）和远距离有遮挡（图 9d～9f）中有个别香梨被漏检现象。

从表 6 可以看出，近景拍摄数据集测试时，M－YOLO v7－SCSN＋F 模型均值平均精度（mAP）、精确度和召回率分别比 YOLO v7－S 高 0.70、1.17、1.13 个百分点；而在远景数据集中，M－YOLO v7－SCSN＋F 模型平均精度均值（mAP）、精确率和召回率分别比 YOLO v7－S 模型高 1.47、1.60、2.11 个百分点。



图 9 香梨果实图像检测

Fig. 9 Images detection of pear fruit

表 6　YOLO v7－S 和 M－YOLO v7－SCSN＋F 检测结果对比

Tab. 6　YOLO v7－S and M－YOLO v7－SCSN＋F in different shooting scene　%

模型	场景	平均精度均值	精确率	召回率
YOLO v7－S	远景	93. 69	94. 27	91. 51
	近景	95. 63	95. 74	93. 29
M－YOLO v7－SCSN＋F	远景	95. 16	95. 87	93. 62
	近景	97. 33	96. 91	94. 42

3.5　不同目标检测模型检测性能对比

实验使用 Faster R－CNN[26]、SSD、YOLO v3[27]、YOLO v4[28]、YOLO v5s[29]、YOLO v7－S[30]、YOLO v8n、RT－DETR－R50 模型与 M－YOLO v7－SCSN＋F 模型在验证集上进行性能对比。从表 7 可以看出，改进后的 M－YOLO v7－SCSN＋F 模型在验证集的平均精度均值（mAP）、精确率和召回率分别达到 97. 23%、97. 63% 和 93. 66%，检测速度为 69. 39 f/s。其平均精度均值（mAP）与相应模型对比分别提高 14. 50、26. 58、3. 88、2. 40、1. 58、0. 16、0. 07、0. 86 个百分点。其模型内存占用量分别减少 74. 26、58. 28、33. 82、36. 23、8. 95、9. 14、16. 47、13. 30 MB，改进后的 M－YOLO v7－SCSN＋F 模型更有利于部署在生产设备上。

表 7　不同类型检测模型在验证集上性能比较

Tab. 7　Performance comparison of different types of detection models on validation set

模型	图像尺寸/（像素×像素）	平均精度均值/%	精确率/%	召回率/%	模型内存占用量/MB	检测速度/（f·s^{-1}）
Faster R－CNN	640×640	82. 73	87. 14	76. 26	102. 39	9. 93
SSD	640×640	70. 65	83. 54	68. 32	86. 41	49. 75
YOLO v3	640×640	93. 35	95. 33	89. 67	61. 95	48. 15
YOLO v4	640×640	94. 83	95. 26	90. 19	64. 36	50. 32
YOLO v5s	640×640	95. 65	96. 43	91. 33	37. 08	53. 66
YOLO v7－S	640×640	97. 07	97. 21	93. 74	37. 27	75. 52
YOLO v8n	640×640	97. 16	97. 29	96. 34	44. 60	81. 83
RT－DETR－R50	640×640	96. 37	96. 72	94. 96	41. 43	83. 17
M－YOLO v7－SCSN＋F	640×640	97. 23	97. 63	93. 66	28. 13	69. 39

3.6　检测热力图

利用 Grad－CAM 产生目标检测热力图[31]，如图 10 所示。通过香梨图像检测热力图，可以看出 M－YOLO v7－SCSN＋F 模型的目标检测热力图更贴近真实梨果区域。通过图 10 中圆形圈出的区域可以看出，原 YOLO v7－S 目标检测热力值相对 M－YOLO v7－SCSN＋F 模型的目标检测热力值更低，表现出漏检现象。通过图 10 可以看出，M－YOLO v7－SCSN＋F 模型可从弱语义的目标中提取更强的目标特征，有着较好的鲁棒性和泛化能力。

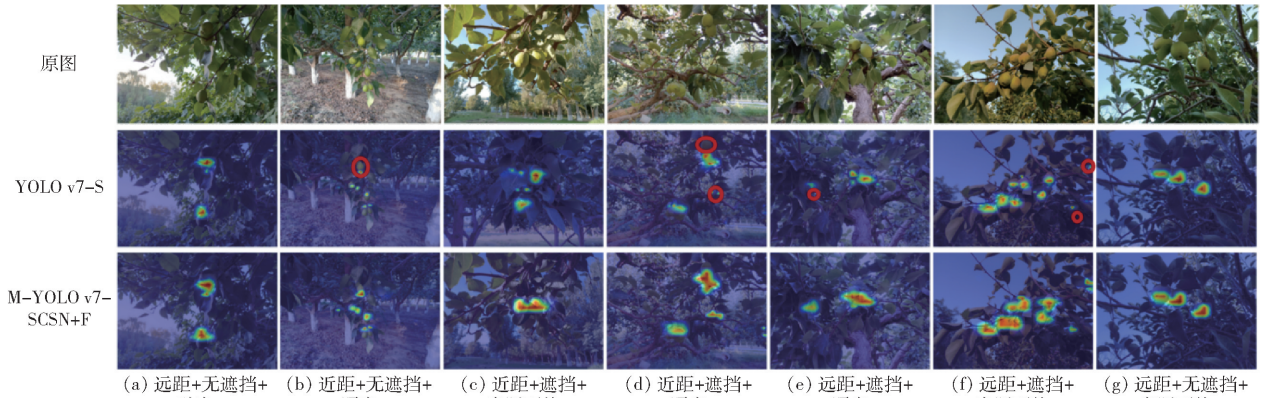


图 10　香梨果实图像检测热力图

Fig. 10　Heat maps of pear fruit image detection

4　结论

（1）以 YOLO v7－S 为基础网络，针对果园中香梨果实、果叶和枝干之间相互遮挡，不易精准定位的问题，设计了一种轻量化香梨目标检测方法 M－YOLO v7－SCSN＋F。该模型在验证集的平均精度均值（mAP）、精确率和召回率分别达到 97. 23%、97. 63% 和 93. 66%，检测速度为 69. 39 f/s。

（2）设计了 YOLO v7－S、YOLO v7－S＋CA、MobileNetv3－YOLO v7、M－YOLO v7－SCSN 和 M－

YOLO v7 - SCSN + F 模型的 5 组消融试验,采用模型检测平均精度均值(mAP)和检测速度两个评价指标,在验证集对模型进行评价,结果表明,M - YOLO v7 - SCSN + F网络模型平均精度均值高于其他 4 组网络模型。

(3)将改进的 M - YOLO v7 - SCSN + F 模型与 Faster R - CNN、SSD、YOLO v3、YOLO v4、YOLO v5s、YOLO v7 - S、YOLO v8n、RT - DETR - R50 模型进行性能对比。结果表明,本研究设计的模型平均精度均值(mAP)与相应模型对比提高 14.50、26.58、3.88、2.40、1.58、0.16、0.07、0.86 个百分点。此外,改进的 M - YOLO v7 - SCSN + F 模型内存占用量比先进的 YOLO v8n 和 RT - DETR - R50 检测模型减少 16.47、13.30 MB,更有利于部署在生产设备上。

参 考 文 献

[1] 邵园园,王永贤,玄冠涛,等. 基于高光谱成像的肥城桃品质可视化分析与成熟度检测[J]. 农业机械学报,2020,51(8):344 - 350.
SHAO Yuanyuan, WANG Yongxian, XUAN Guantao, et al. Visual detection of SSC and firmness and maturity prediction for Feicheng peach by using hyperspectral imaging[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(8): 344 - 350. (in Chinese)

[2] 宋怀波,尚钰莹,何东健. 果实目标深度学习识别技术研究进展[J]. 农业机械学报,2023,54(1):1 - 19.
SONG Huaibo, SHANG Yuying, HE Dongjian. Review on deep learning technology for fruit target recognition[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(1):1 - 19. (in Chinese)

[3] 林景栋,吴欣怡,柴毅,等. 卷积神经网络结构优化综述[J]. 自动化学报,2020,46(1):24 - 37.
LIN Jingdong, WU Xinyi, CHAI Yi, et al. Structure optimization of convolutional neural networks: a survey[J]. Acta Automatica Sinica, 2020, 46(1): 24 - 37. (in Chinese)

[4] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Conference on Computer Vision and Pattern Recognition, 2016: 770 - 778.

[5] YONG H, ZHANG C. LeNet - 5 improvement based on FPGA acceleration[J]. The Journal of Engineering, 2020, 13(7): 526 - 528.

[6] 黄小玉,李光林. 基于改进判别区域特征融合算法的近色背景绿色桃子识别[J]. 农业工程学报,2018, 34(23):23 - 38.
HUANG Xiaoyu, LI Guanglin. Near-color background green peach recognition based on improved discriminative region feature fusion algorithm[J]. Transactions of the CSAE, 2018, 34(23):23 - 38. (in Chinese)

[7] HE K, ZHANG X, REN S. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904 - 1916.

[8] REN S, HE K, GIRSHICK R. Faster R - CNN: towards real-time object detection with region proposal networks[C]// Advances in Neural Information Processing Systems, 2015: 91 - 99.

[9] REDMON J, FARHADI A. Yolo v3: an incremental improvement[J]. arXiv preprint arXiv: 1804.02767, 2018.

[10] TIAN Y, YANG G, WANG Z, et al. Apple detection during different growth stages in orchards using the improved YOLO - v3 model[J]. Computers and Electronics in Agriculture, 2019, 157: 417 - 426.

[11] 王立舒,秦铭霞,雷洁雅,等. 基于改进 YOLO v4 - Tiny 的蓝莓成熟度识别方法[J]. 农业工程学报,2021,37(18):170 - 178.
WANG Lishu, QIN Mingxia, LEI Jieya, et al. Blueberry maturity recognition method based on improved YOLO v4 - Tiny[J]. Transactions of the CSAE, 2021, 37(18):170 - 178. (in Chinese)

[12] LÜ S, LI R, ZHAO Y, et al. Green citrus detection and counting in orchards based on YOLO v5 - CS and AI edge system [J]. Sensors, 2022, 22(2): 576.

[13] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLO v7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]// Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2023: 7464 - 7475.

[14] 吕珊. 面对复杂场景的鲁棒目标检测算法研究[D]. 北京:北京交通大学,2022:10 - 18.
LÜ Shan. Research on robust object detection algorithms for complex scenes[D]. Beijing: Beijing Jiaotong University, 2022: 10 - 18. (in Chinese)

[15] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLO v4: optimal speed and accuracy of object detection[EB/OL][2022 - 11 - 13]. https://arxiv.org/pdf/2004.10934.

[16] 伊建峰,黎思成,吕珊,等. 基于频域数据增强及 YOLO v7 的动火作业检测模型[J]. 计算机应用,2023,43(增刊): 285 - 290.
YI Jianfeng, LI Sicheng, LÜ Shan, et al. Hot work detection model based on frequency domain data enhancement and YOLO v7[J]. Journal of Computer Applications, 2023, 43(Supp): 285 - 290. (in Chinese)

[17] YANG F, LI X, SHEN J. MSB - FCN: multi-scale bidirectional FCN for object skeleton extraction[J]. IEEE Transactions on Image Processing, 2020, 30: 2301 - 2312.

[18] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781 - 10790.

[19] HAN K, WANG Y, CHEN H, et al. A survey on vision transformer[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2022,

45(1)：87－110.
[20] HOU Q,ZHOU D,FENG J. Coordinate attention for efficient mobile network design[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021:13713－13722.
[21] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Computer Vision and Pattern Recognition. IEEE,2016: 779－788.
[22] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Computer Vision and Pattern Recognition. IEEE, 2017: 6517－6525.
[23] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression [C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 1－7.
[24] REZATOFIGHI S H, TSO I N. Generalized intersection over union: a metric and a loss for bounding box regression[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
[25] GEVORGYAN Z. SIoU loss: more powerful learning for bounding box regression[J]. Computer Vision and Pattern Recognition,2022(5):1－12.
[26] 朱红春,李旭,孟炀,等. 基于 Faster R－CNN 网络的茶叶嫩芽检测[J]. 农业机械学报,2022,53(5):217－224.
ZHU Hongchun, LI Xu, MENG Yang, et al. Tea bud detection based on Faster R－CNN network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(5):217－224. (in Chinese)
[27] 武星,齐泽宇,王龙军,等. 基于轻量化 YOLO v3 卷积神经网络的苹果检测方法[J]. 农业机械学报,2020,51(8):17－25.
WU Xing, QI Zeyu, WANG Longjun, et al. Apple detection method based on light-YOLO v3 convolutional neural network [J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(8):17－25. (in Chinese)
[28] 郭建军,何国煌,徐龙琴,等. 基于改进 YOLO v4 的肉鸽行为检测模型研究[J]. 农业机械学报,2023,54(4):347－355.
GUO Jianjun, HE Guohuang, XU Longqin, et al. Pigeon behavior detection model based on improved YOLO v4[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(4):347－355. (in Chinese)
[29] 李健源,柳春娜,卢晓春,等. 基于改进 YOLO v5s 和 TensorRT 部署的鱼道过鱼监测[J]. 农业机械学报,2022,53(12):314－322.
LI Jiangyuan, LIU Chunna, LU Xiaochun, et al. Fish passage monitoring based on improved YOLO v5s and TensorRT deployment[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(12):314－322. (in Chinese)
[30] 宋怀波,马宝玲,尚钰莹,等. 基于 YOLO v7－ECA 模型的苹果幼果检测[J]. 农业机械学报,2023,54(6):233－242.
SONG Huaibo,MA Baoling,SHANG Yuying,et al. Detection of young apple fruits based on YOLO v7－ECA model[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(6):233－242. (in Chinese)
[31] MARTIN D J,FAHAD S K,MICHAEL F,et al. Adaptive color attributes for real-time visual tracking[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2014, 143(1): 1090－1097.

(上接第 243 页)

[18] ZHANG QL, YANG Y. SA－Net: shuffle attention for deep convolutional neural networks[C]//ICASSP－IEEE International Conference on Acoustics, 2021: 2235－2239.
[19] BHAT S, ALHASHIM I, WONKA P. LocalBins: improving depth estimation by learning local distributions[J]. ArXiv, 2022, abs/2203.15132.
[20] LEE J H, HAN M K, KO D W, et al. From big to small: multi-scale local planar guidance for monocular depth estimation [J]. ArXiv, 2019, abs/1907.10326.
[21] SUN L, BIAN J, ZHAN H, et al. SC－DepthV3: robust self-supervised monocular depth estimation for dynamic scenes[J]. ArXiv, 2022, abs/2211.03660.
[22] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]// IEEE/CVF Conference on Computer Vision Pattern Recognition, 2021: 13708－13717.
[23] WANG Q, WU B, ZHU P F, et al. ECA－Net: efficient channel attention for deep convolutional neural networks[C]//IEEE/CVF International Conference on Computer Vision, 2019: 11531－11539.
[24] MEHTA S, RASTEGARI M. Separable self-attention for mobile vision transformers[J]. ArXiv, 2022, abs/2206.02680.
[25] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[J]. ArXiv, 2018, abs/1807.06521.
[26] LI X, HU X, YANG J. Spatial group-wise enhance: improving semantic feature learning in convolutional networks[J]. ArXiv, 2019, abs/1905.09646.