

基于 YOLO v7 - ST 模型的小麦籽粒计数方法研究

王 玲¹ 张 旗¹ 冯天赐¹ 王一博¹ 李雨桐² 陈 度¹

(1. 中国农业大学工学院, 北京 100083; 2. 北大荒农业服务集团黑龙江农机服务有限公司, 哈尔滨 150090)

摘要: 针对小麦考种过程中籽粒堆积、粘连和遮挡现象导致计数准确率低等问题, 本文基于电磁振动原理设计了高通量小麦籽粒振动分离装置, 通过分析受力探讨了籽粒离散分离程度的主要影响因素, 并引入二阶离散系数建立了籽粒离散度等级评价方法。在此基础上, 引入 Swin Transformer 模块构建 YOLO v7 - ST 模型, 对不同离散度等级下小麦籽粒进行计数性能测试。试验结果表明, YOLO v7 - ST 模型在 3 种离散度等级下平均计数准确率、F1 值和平均计数时间的总平均值分别为 99.16%、93% 和 1.19 s, 相较于 YOLO v7、YOLO v5 和 Faster R - CNN 模型, 平均计数准确率分别提高 1.03、2.34、15.44 个百分点, 模型综合评价指标 F1 值分别提高 2、3、16 个百分点, 平均计数时间较 YOLO v5 和 Faster R - CNN 分别减少 0.41 s 和 0.36 s, 仅比 YOLO v7 模型增大 0.09 s。因此, YOLO v7 - ST 模型可实现多种离散度等级下不同程度籽粒遮挡和粘连问题的准确快速检测, 大幅提高小麦考种效率。

关键词: 小麦; 籽粒计数; 目标检测; 离散度等级; 电磁振动; YOLO v7 - ST

中图分类号: TP391.4 文献标识码: A 文章编号: 1000-1298(2023)10-0188-10 **OSID:** 

Wheat Grain Counting Method Based on YOLO v7 - ST Model

WANG Ling¹ ZHANG Qi¹ FENG Tianci¹ WANG Yibo¹ LI Yutong² CHEN Du¹

(1. College of Engineering, China Agricultural University, Beijing 100083, China

2. Heilongjiang Agricultural Machinery Service Co., Ltd., Beidahuang Agricultural Service Group Co., Ltd., Harbin 150090, China)

Abstract: Aiming at the problems of low counting accuracy due to seed accumulation, sticking and shading phenomena in the wheat seed testing process, a high-throughput wheat seed vibration separation device was designed based on the principle of electromagnetic vibration. The main influencing factors of the degree of seed dispersion and separation were discussed by analyzing the forces, and the second-order dispersion coefficient was introduced to establish the seed dispersion grade evaluation method. On this basis, the YOLO v7 - ST model was then built by using the Swin Transformer module and was tested for counting performance under different discrete degree levels. The experimental results showed that the mean counting accuracy, F1 value and mean counting time of the YOLO v7 - ST model were 99.16%, 93% and 1.19 s under the three dispersion levels, respectively. Compared with that of the YOLO v7, YOLO v5 and Faster R - CNN models, the mean counting accuracy was improved by 1.03 percentage points, 2.34 percentage points and 15.44 percentage points, respectively, and the F1 values of the comprehensive evaluation index of the model was increased by 2 percentage points, 3 percentage points and 16 percentage points, respectively. The mean counting time was decreased by 0.41 s and 0.36 s compared with that of YOLO v5 and Faster R - CNN, respectively, and it was only 0.09 s slower than that of the YOLO v7 model. Overall, the YOLO v7 - ST model provided accurate and efficient detection of grains under various discrete degree levels, significantly improved the efficiency of wheat breeding.

Key words: wheat; grain count; object detection; dispersion level; electromagnetic vibration; YOLO v7 - ST

0 引言

小麦是我国重要粮食作物之一, 提高小麦产量

和品质对保障我国粮食安全具有重要意义。考种作为选育高产高质小麦品种的重要手段, 其效率和精度直接影响育种结果。小麦考种涉及植株和籽粒的

多种性状参数,如千粒质量、粒型、产量等,而籽粒个数计算是千粒质量测量、产量预估等的基础数据。因此,快速准确的小麦高通量籽粒计数能加快考种效率,对提升小麦产量具有重要意义^[1-3]。

然而,在复杂农业生产环境下,小麦籽粒计数存在不同程度的重叠遮挡、密集粘连等问题,严重影响计数精度。近年来,机器视觉技术被广泛应用于农作物品质检测、产量预估和籽粒计数的研究,基于此,研究人员提出了基于传统数字图像处理的籽粒计数方法^[4-6]。VISEN 等^[7]使用分类和凹点检测对遮挡籽粒进行分割,通过确定惯性等效椭圆重叠程度,将谷物表征为孤立的内核或一组粘连的内核,再通过凹点检测法和最近邻准则绘制分割线。李立君等^[8]提出基于凸壳理论的遮挡油茶果定位检测算法,该算法通过颜色特征和凹点搜寻对重叠目标进行粗分割,再利用轮廓提取算法提取出了凸壳上的有效轮廓。相比于人工和传感器的逐个检测方法^[9-11],数字图像处理方法大大提高了检测效率和计数精度,然而,该方法高度依赖图像质量和研究人员对目标特征的分析,不同场景的泛化能力及鲁棒性较差。

深度学习方法因其对抽象特征的强大学习能力,在目标物遮挡、粘连情况下仍能够通过关键特征有效识别目标^[12-14]。SUN 等^[15]针对重叠水稻种子计数提出了一种基于轮廓分组预标记的深度学习目标检测方法,该方法以欧几里得距离和散度函数作为综合标准预先标记水稻种子轮廓,并将预标记结果整合到其特征提取层,通过 Faster R-CNN 进行分类计数,平均错误率 1.06%。ZHAO 等^[16]开发了一种基于改进的 YOLO v5 的小麦麦穗检测方法,通过添加微尺度检测层,利用置信度系数对多层特征图中的检测框进行融合,以提高遮挡条件下的检测精度,平均检测准确率 94.1%。XU 等^[17]针对自然环境下柑橘的重叠、遮挡问题,将通道注意力机制和预测框筛选算法 Soft DIoU NMS 引入 YOLO v4 中,有效地提高了检测精度。龙燕等^[18]提出了一种改进全卷积单阶段无锚框网络 FCOS 的苹果目标检测方法,该网络引入联合交并比损失函数,能更好反映预测框和真实框重合度,使网络能够关注到被果实和树叶遮挡的极难识别的目标,并针对不同密集度和不同遮挡程度进行对比试验,结果表明改进的 FCOS 准确率可达 96.0%。因此,深度学习方法应用于多种农作物重叠、遮挡检测与识别的研究取得了一定效果^[19-20],但相关研究多是针对目标与复杂背景之间的遮挡问题,当多个待检测目标相互遮挡且遮挡程度较高,仅显现极小的局部特征时,相关研

究的算法模型不能准确从仅有的未被遮挡的局部特征中识别待检测目标,从而将其与粘连的其他检测目标识别为同一目标,造成漏检。

此外,目前基于深度学习的重叠遮挡的研究场景多是针对少量目标的低通量检测,然而,小麦室内考种具有高通量计数需求,大量籽粒存在不同程度的遮挡、粘连等问题,同时小麦籽粒目标较小,遮挡目标的识别难度大幅提高,相关研究中的算法模型无法满足高通量小目标快速准确的计数需求。因此,本文基于 Swin Transformer 提出改进 YOLO v7-ST 网络模型,该模型利用移位窗口和多头注意力机制可有效识别被遮挡籽粒与其粘连籽粒的关联特征,并提高对遮挡目标局部特征的识别能力。以此为基础,基于电磁振动原理设计小麦籽粒振动分离装置,并引入二阶离散系数对籽粒离散均匀度进行试验研究,从而对不同程度遮挡、粘连的小麦籽粒进行分级检测,以期实现各级离散度下籽粒的准确快速计数。

1 籽粒振动分离装置设计

1.1 系统组成

小麦籽粒振动分离装置如图 1 所示,由工业相机、光源、电磁振动器、振动控制器、输送平板、卸料斗、图像采集处理单元和固定支架组成。相机采用华睿科技 A7500CG20 型彩色 CMOS 相机,分辨率为 500 万像素,搭配 16 mm 焦距镜头,采样视野范围为 350 mm×200 mm。光源采用亮度强、光损失少、成像清晰、亮度均匀的同轴光源,为图像采集装置提供稳定的光照环境。

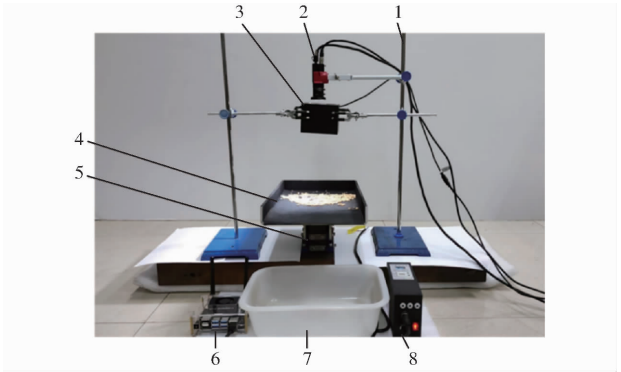


图 1 小麦籽粒振动分离装置

Fig. 1 Wheat seed vibration separation device

1. 支架 2. 工业相机 3. 光源 4. 输送平板 5. 电磁振动器
6. 图像处理单元 7. 卸料斗 8. 振动控制器

由于散落在输送平板上的小麦籽粒呈堆积状态,难以直接进行图像采集与处理,因此,基于电磁振动原理消除籽粒堆积现象,使籽粒离散分布于图像采集范围内,可提高考种测量精度。电磁振动器

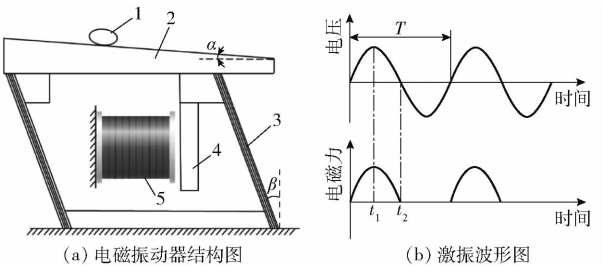


图2 电磁振动器工作原理图

Fig.2 Diagram of working principle of electromagnetic vibrator
1. 小麦籽粒 2. 输送平板 3. 弹簧片 4. 衔铁 5. 电磁线圈

工作原理图如图 2 所示,电流在正半周期时,交流电流过电磁线圈产生电磁力吸附衔铁,衔铁带动弹片产生形变存储势能;当位于负半周期时,线圈内没有电流流过,电磁力消失,弹片释放弹性势能产生激振力,输送平板在激振力的作用下向前运动,其中 β 为振动方向角,取 15° , α 为输送倾角。电流通过电磁线圈时所产生的激振力 $F^{[21-22]}$ 为

$$F=0.004\pi^2f_0^2Hm(1-\xi^2)\tag{1}$$

式中 f_0 ——振动频率,Hz
 H ——振幅,mm
 m ——物料质量,kg
 ξ ——激振频率与固有频率之比

随电流的周期性变化,与输送平板接触的籽粒沿振动方向向前滑动一定距离,而堆积于表层的籽粒滑落至平板后才能向前滑移。当振动器以一定频率和振幅连续振动时,籽料被连续输送出去。适当的激振力有利于籽粒平稳前进,当激振力较大时,籽粒前进速度过快,表层和底层籽粒在激振力作用下同时快速前进,不能及时分离;而激振力过小时,籽粒分离所需时间过长。待籽粒堆积现象消除后进行图像采集与处理,进而籽粒在电磁振动作用下从输送平板末端滑出,完成卸料。

1.2 小麦籽粒离散度分级方法

1.2.1 籽粒离散度影响因素分析

小麦籽粒在电磁振动过程中主要受激振力 F 、重力 G 和摩擦力 F_f (图 3),籽粒在电磁振动作用下随输送平板运动时,所受合力 F' 和运动加速度 a 为

$$\begin{cases} F'=F\cos(\alpha+\beta)+G\sin\alpha-F_f \\ a=0.004\pi^2f_0^2H(1-\xi^2)\cos(\alpha+\beta)+g\sin\alpha-\mu g \end{cases}\tag{2}$$

式中 g ——重力加速度, m/s^2
 μ ——摩擦因数

由式(2)可知,籽粒的运动加速度受振动频率、振幅、输送倾角和摩擦因数的影响,当电磁振动器振动参数与输送平板表面粗糙度不同时,籽粒在激振力、重力和摩擦力的共同作用下向前运动,籽粒的分

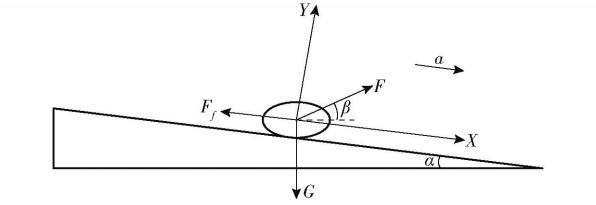


图3 籽粒受力分析图
Fig.3 Force analysis diagram of seed

离效果即籽粒离散度不同。为探究小麦籽粒高通量计数需求中不同离散度下籽粒粘连和遮挡程度对籽粒计数准确度的影响,以小麦籽粒离散度为评价目标,选择振动频率 A 、振幅 B 、摩擦因数 C 及输送倾角 D 为试验因素,通过设计正交试验采集不同离散度的试验样本。

电磁振动器振动过程中,当激振频率等于固有频率时将发生共振,此时振幅达到最大值,但振动装置工作不平稳。为达到高效、稳定的工作状态,应使振动电机工作在亚共振区,即激振频率与固有频率之比 ξ 为 $0.85\sim0.98^{[23]}$ 。试验中,采用SDVC31型数字调频振动送料控制器进行振动频率调节,通过固定控制器输出电压并调整激振频率,发现当激振频率为 73 Hz 时,振幅最大,即振动器固有频率为 73 Hz ,振动频率 f_0 为 $62.5\sim69.4\text{ Hz}$ 。

引入机械指数 K 以计算籽粒振幅。

$$K=4\pi^2f_0^2\frac{H}{g}\tag{3}$$

机械指数是用来衡量振动强弱的物理量,机械指数越大表明振动越剧烈,为使得物料稳定输送,并保证振动冲击对槽体的损坏较小,通常 K 取 $2\sim5^{[23]}$ 。由式(3)计算可知,籽粒振幅为 $0.13\sim0.26\text{ mm}$ 。

探究输送平板摩擦因数对籽粒离散度的影响时,采用不同粗糙度的砂纸作为表面材料(所以摩擦因数以砂纸的粒度为度量,即单位为目),基于已有的研究分析和实际作业效果,砂纸分别选择 30 、 60 、 120 、 240 目。此外,输送平板倾角 α 为正值时,平板向下倾斜,籽粒运动加速度增大,反之,籽粒运动加速度减小;根据试验效果,平板倾角 α 为 $-3^\circ\sim6^\circ$ 。每个因素选择 4 个水平进行试验,试验因素和水平如表 1 所示 $^{[24-25]}$ 。

表1 正交试验因素水平
Tab.1 Factors and levels of orthogonal test

水平	因素			
	A/Hz	B/mm	$C/\text{目}$	$D/(^\circ)$
1	62.5	0.13	30	-3
2	64.8	0.17	60	0
3	67.1	0.21	120	3
4	69.4	0.26	240	6

将 50 g 小麦籽粒直接导入振动分离装置,采用四因素四水平的 L_{16} 正交试验方案进行试验,图 4 为不同影响因素下小麦籽粒的离散分布情况,摄像头采集的图像尺寸为 2 448 像素 $\times$ 1 604 像素。由图 4 可知,受电磁振动作用高通量籽粒的堆积现象基本消除,籽粒平铺于振动平板上。

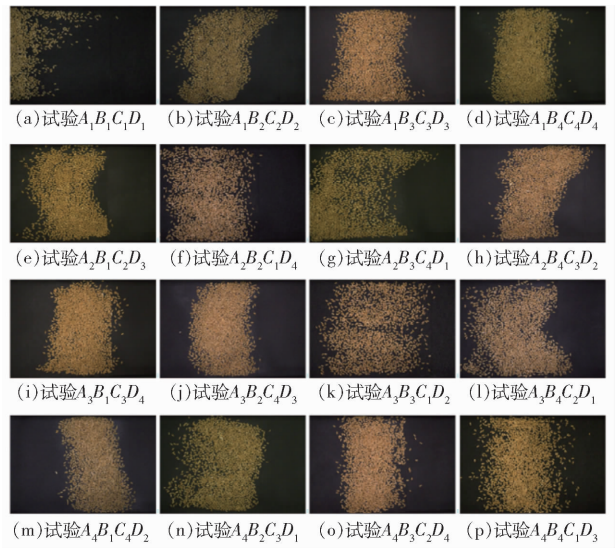


图 4 基于正交试验方案的籽粒样本离散度分布图
Fig. 4 Dispersion distribution of seed samples based on orthogonal test scheme

1.2.2 籽粒离散度评价方法

由图 4 可知,受振动频率、振幅、输送倾角和摩擦因数的影响,籽粒的离散分布情况差异明显,进而产生不同程度的籽粒遮挡和粘连问题。评价数据离散度通常采用离散系数,但在籽粒分布图中,局部区域籽粒极度密集或极度稀疏,导致数据极差大,传统离散系数无法反映籽粒分布的真实情况。二阶离散系数对异常值的敏感度较低,受局部离散度的影响较小,因此,引入二阶离散系数作为研究籽粒离散度的评价指标,计算式为^[26]

$$c_v = \frac{\sigma}{\sqrt{\bar{x}}}$$

(4)

式中 c_v ——二阶离散系数
 σ ——数据标准差
 $\bar{x}$ ——像素占比平均值

为计算原始图像的二阶离散系数,需获取二值图中各局部区域目标像素占比作为基础数据,因此,将对原始图像进行预处理。首先采用 Laplace 算子和锐化方法对原始图像进行组合增强,增强后的图像像素分布出现明显分界线,如图 5 所示,再通过自适应阈值分割将图像二值化。

由于光照强度、砂纸晶粒噪声等因素影响,二值图中存在较多尺寸不一的噪声和孔洞,基于传统十

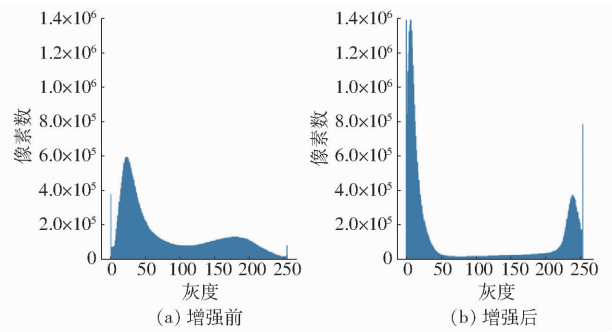


图 5 图像像素分布
Fig. 5 Image pixel distribution

字形和矩形卷积操作的形态学处理方式难以同时消除噪声和填充孔洞。因此,分别对灰度 0 像素和灰度 255 像素连通域面积进行统计分析,通过面积阈值筛选出目标噪声和孔洞,采用四邻域漫水填充算法分别对其进行反向填充。为衡量籽粒分布特征,将处理后的二值图裁剪为 24 个 408 像素 $\times$ 401 像素的局部区域,通过统计各区域籽粒像素百分比作为基础数据,再计算该组数据的总体平均值和标准差,即可由式(4)得到原始图像籽粒分布的二阶离散系数,图像处理过程如图 6 所示。

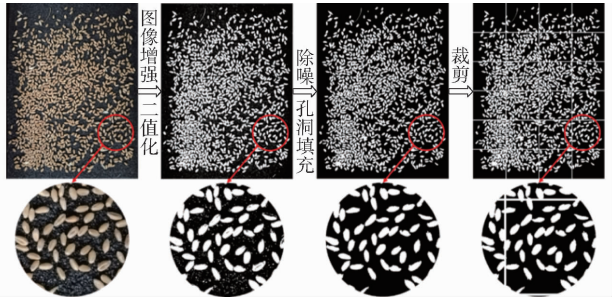


图 6 籽粒原始图像预处理效果图
Fig. 6 Pre-processing effect of original image of seeds

1.2.3 籽粒离散度分级

二阶离散系数是衡量一组数据离散度的相对统计量,其值越小,则数据的离散度越低,籽粒分布越均匀;反之数据离散度高,籽粒分布不均匀,遮挡与粘连现象明显。基于籽粒离散度评价方法,对 16 组试验样本进行离散度系数计算,每组试验重复 3 次,取平均二阶离散系数为试验结果,籽粒离散度评价如表 2 所示。结合图 4 中各组试验样本的籽粒离散分布情况,将籽粒离散度分为 3 个等级,其中,当平均二阶离散系数 $c_v \leq 0.4$ 时,离散度等级为“稀疏”, $0.4 < c_v < 0.5$ 时为“中等密集”, $c_v \geq 0.5$ 时为“密集”。

2 小麦籽粒计数算法设计

2.1 数据集构建

试验所用小麦籽粒于 2021 年 6 月在北京小

表 2 正交试验结果

Tab.2 Results of orthogonal test

试验 序号	因素				c_v	离散度 等级
	A/Hz	B/mm	C/目	D/(°)		
1	62.5	0.13	30	-3	0.501 5	密集
2	62.5	0.17	60	0	0.401 8	中等密集
3	62.5	0.21	120	3	0.455 1	中等密集
4	62.5	0.26	240	6	0.599 2	密集
5	64.8	0.13	60	3	0.448 1	中等密集
6	64.8	0.17	30	6	0.461 9	中等密集
7	64.8	0.21	240	-3	0.388 3	稀疏
8	64.8	0.26	120	0	0.478 4	中等密集
9	67.1	0.13	120	6	0.543 4	密集
10	67.1	0.17	240	3	0.521 6	密集
11	67.1	0.21	30	0	0.361 0	稀疏
12	67.1	0.26	60	-3	0.423 4	中等密集
13	69.4	0.13	240	0	0.520 8	密集
14	69.4	0.17	120	-3	0.422 6	中等密集
15	69.4	0.21	60	6	0.527 3	密集
16	69.4	0.26	30	3	0.477 5	中等密集

汤山镇国家精准农业研究示范基地试验田获得,千粒质量为 36 g,每次试验选取籽粒 50 g,约 1 400 粒。基于籽粒振动分离装置采集不同离散度图像 72 幅,图像分辨率为 2 448 像素 × 1 604 像素,并使用 Labellmg 标注软件手工标记小麦籽粒。为保证测试结果能够真实反映模型泛化能力,均匀选取原始数据集中不同密度等级的 32 幅图像作为测试集,剩下 40 幅作为训练集和验证集。由于原始图像分辨率较高,难以满足网络输入尺寸要求,因此,通常采用下采样和裁剪两种方式对原始图像进行处理。由于强制下采样会降低图像分辨率,导致图像特征严重丢失,造成部分籽粒无法识别。因此,本文采用步长为 543 的重叠滑窗策略将原始训练集和验证集中的高分辨率图像裁剪为 640 像素 × 640 像素的低分辨率图像,满足模型输入限制,并对重叠部分的预测结

果进行融合,以解决传统裁剪方式因边缘籽粒误分割而引入的计数误差。

为在原始图像样本量有限的情况下更好地提取图像特征,本文采用光度失真和几何失真对样本量进行扩充,通过调整图像色调及饱和度,添加随机缩放、平移、剪切和旋转等,将裁剪后小尺寸图像扩充至 1 224 幅。为进一步提高模型的泛化能力,采用 Mixup^[27]和 Mosaic^[28]增强方法,通过从训练图像中随机选取 2 个样本及其标签进行加权求和以及使用 4 幅图像分别覆盖不同区域等方法,丰富数据集中图像背景,增加数据的多样性,防止样本扩充所产生的图像特征相似度高而导致的模型过拟合,数据增强方法的实际效果如图 7 所示。

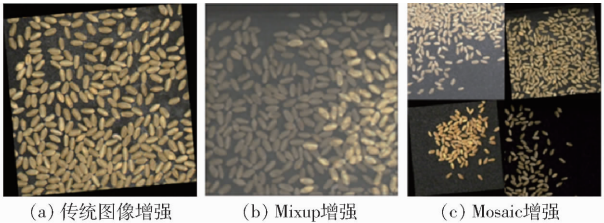


图 7 不同数据增强方法实际效果图

Fig.7 Actual effect of different data enhancement methods

2.2 YOLO v7-ST 网络模型

YOLO v7^[29]算法是基于深度神经网络进行对象的识别与定位,通过引入高效聚合网络等多个学习策略提高模型学习能力,并大幅提高检测速度。而小麦籽粒计数任务中,当籽粒粘连严重甚至形成遮挡时,会严重影响卷积神经网络的输出结果,因此,为提高模型对不同离散度下籽粒计数任务的检测精度,本文以 YOLO v7 作为基线,引入 Swin Transformer^[30]模块,提出了 YOLO v7-ST 模型,模型网络结构如图 8 所示。

YOLO v7 在主干网络中采用的多分支堆叠模块 E-ELAN 可利用分组卷积来扩展计算模块的通道和基数,实现在不改变原有梯度路径的情况下堆叠

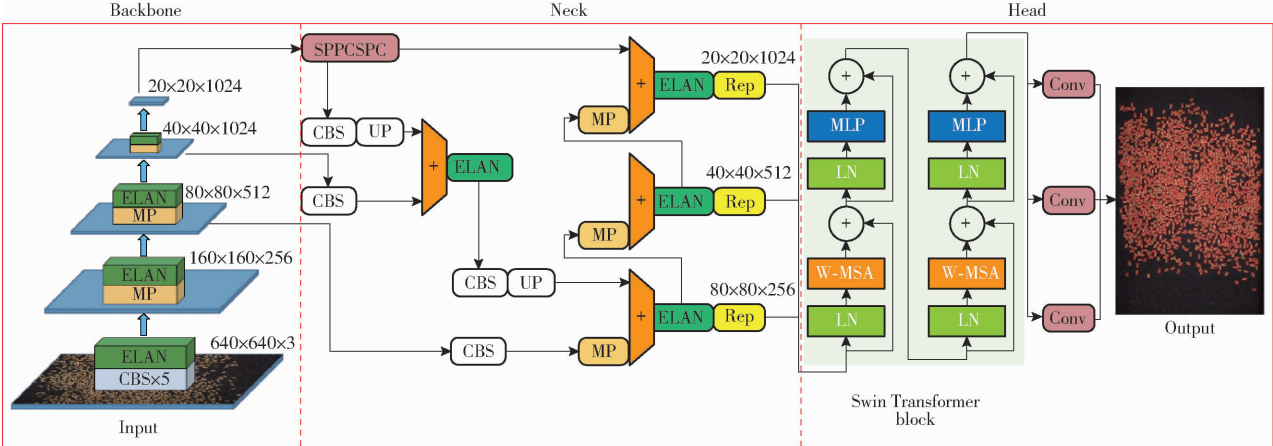


图 8 YOLO v7-ST 网络结构框图

Fig.8 YOLO v7-ST network structure

更多模块,增强网络学习能力。颈部沿用特征金字塔结构对多特征进行融合,头部网络中使用重参数化卷积 RepConv,在训练过程中并联 3×3 卷积、 1×1 卷积和恒等映射,为不同的特征映射提供了更多的梯度多样性。同时添加辅助检测头,计算网络中间层损失值以监督梯度下降方向,最后将辅助头和检测头的结果进行融合,可有效提高模型性能。

然而,在本文“密集”分布等级下,存在较多因重叠遮挡而产生的小目标籽粒。对于这种高度密集情况下的小目标检测场景,经过主干和颈部网络提取到的特征信息中包含大量干扰信息,由传统卷积模块进行的权重分配会导致小目标特征被视为冗余信息而丢失,因此考虑在检测头中添加 Swin Transformer 模块以更好地整合从特征融合网络中输出的特征信息。

Swin Transformer block 首先通过 Patch Embedding 对输入图像进行切块编码,经归一化后送入带移动窗口的多头注意力机制层,再经 Concat、LayerNorm 以及多层感知机 MLP 后输出特征图。根据本文检测类别数使用 1×1 卷积对特征图通道数进行调整,得到 $[20, 20, 18]$ 、 $[40, 40, 18]$ 和 $[80, 80, 18]$ 3 个输出结果,其中 18 代表 3 个先验框的 6 个参数,分别包括每个特征点的回归参数 $[x, y, w, h]$,用于调整获得预测框位置,以及每个特征点是否包含目标和所包含的目标种类。

相较于传统的自注意力机制在对当前位置信息进行编码时过度关注自身位置的缺陷, Swin Transformer 的多头注意力机制不仅可以帮助模型独立地关注不同的特征子空间,还可通过窗口的移位操作加强局部信息之间的联系,捕获丰富的上下文信息。在公开数据集 VisDrone2021 中, Swin Transformer 模块在高密度遮挡对象上的识别准确率提高了 30% ~ 50%^[31],这也更加证实了其对遮挡对象的检测能力。

2.3 模型训练及后处理

本文模型训练通过 NVIDIA RTX3080 GPU 加速,显存 10 GB,采用 Python 3.8 及 PyTorch 1.7.0 完成程序设计,梯度下降采用 Adam 优化器。为加快模型拟合速度,通过迁移学习加载已在 ImageNet 数据集^[32]中训练完成的主干网络权重参数,且采用冻结训练方式将主干网络冻结,只训练头部网络,初始学习率为 0.001,经过 50 次 (Epoch) 训练后解冻,共训练 110 次。冻结训练时占用显存较小,设置 Batch Size 为 8,解冻后设置 Batch Size 为 4。采用余弦退火策略^[33]对学习进行动态调整,训练损失值变化曲线如图 9 所示,模型随迭代次数增加逐步收敛,在

80 轮训练后达到稳定,将此时的训练结果作为最终权重参数。

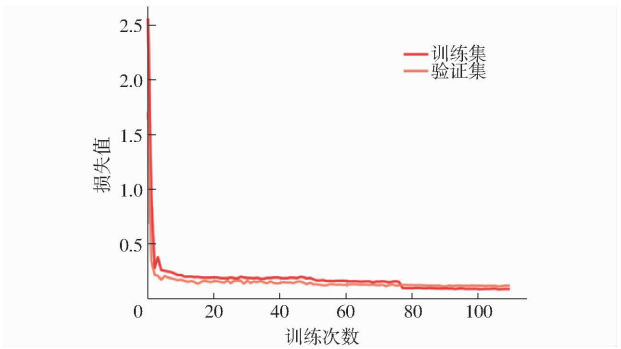


图 9 训练损失变化曲线

Fig. 9 Training loss variation curves

YOLO v7-ST 网络模型输入图像是经过重叠滑动窗裁剪过后的 640 像素 × 640 像素的小尺寸图像,需将小尺寸图像的预测结果拼接成原始图像尺寸。由于图像拼接的重叠部分存在多个预测框,如图 10a 所示,因此,采用非极大抑制算法 (NMS) 和 Soft-NMS 算法^[34]对重叠预测框进行筛选。如图 10b 所示,最右侧籽粒与中间籽粒紧密贴合造成遮挡,两者交并比较大,采用 NMS 算法导致置信度较低的右侧籽粒预测框被“抑制”,仅留下置信度较高的中间籽粒预测框。而 Soft-NMS 算法基于交并比的高斯衰减函数降低重叠框的置信度,衰减函数公式为

$$S'_i = S_i e^{-\frac{\text{IoU}(M, b_i)^2}{\sigma}} \tag{5}$$

式中 i ——样本索引
 S_i ——置信度分数
 b_i ——第 i 个预测框
 S'_i ——衰减后的置信度分数
 M ——置信度最大的预测框
 $\text{IoU}()$ ——交并比,两预测框重叠面积与两预测框并集面积之比

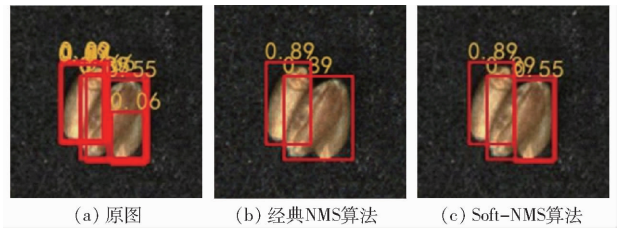


图 10 后处理效果对比

Fig. 10 Post-processing effect comparisons

由式(5)可知,与预测框 M 重叠度越高,即交并比越大的预测框,其置信度衰减越大。重复上述过程更新所有预测框置信度,并设置置信度阈值,删除小于该阈值的预测框,其筛选结果如图 10c 所示, Soft-NMS 算法有效保留了置信度 0.55 的右侧预测框。

3 结果与分析

为量化分析网络模型性能,采用平均计数准确率 (Mean counting accuracy, MCA)、精 确 率 (Precisoin, P)、召回率 (Recall, R)、F1 值和平均计数时间 (Mean counting time, MCT) 作为评价指标。其中,平均计数准确率表示模型预测籽粒数与实际籽粒数的比值,用于表征计数结果的准确性;精确率反映模型的查准能力;召回率反映模型的查全能力;F1 值是评价精确率和召回率的综合指标;平均计数时间即单幅图像的检测时间。

3.1 预测结果分析

采用本文构建的 YOLO v7 - ST 模型对不同离散度等级的 32 幅测试集图像进行预测,单幅图像籽粒个数约 1 400 粒,不同离散度等级下籽粒分布情况如图 11 所示。采用预测框的红色中心点对籽粒进行标记,由表 3 可知,YOLO v7 - ST 模型在各离散度等级下的平均计数准确率平均值为 99.16%,F1 值平均值为 93%,平均计数时间的平均值为 1.19 s,其中籽粒离散度等级处于“稀疏”时,计数准确率为 99.93%;当籽粒处于“中等密集”等级时,籽粒粘连程度增加,计数准确率达 99.80%,表明模型对严重粘连籽粒有较好的识别效果;当籽粒处于“密集”等级时,计数准确率下降 2.06 个百分点,且精确率普遍高于召回率,说明模型具有更强的查准能力,查全能力稍弱,因此模型计数准确率下降的主要原因在于遮挡程度增加造成的漏检,但计数准确率仍可达到 97.74%,模型综合评价指标 F1 值均在 90% 以上。检测速度方面,由于模型采用重叠滑窗将单幅图像裁剪 15 次,即需检测 15 次,当籽粒分布密集度高时,部分小尺寸图像中不存在籽粒,即模型跳过该区域的检测,因此单幅图像检测时间和籽粒离散度呈负相关;单幅图像模型平均计数时间为 1.19 s,能够满足高通量快速准确检测的需求。

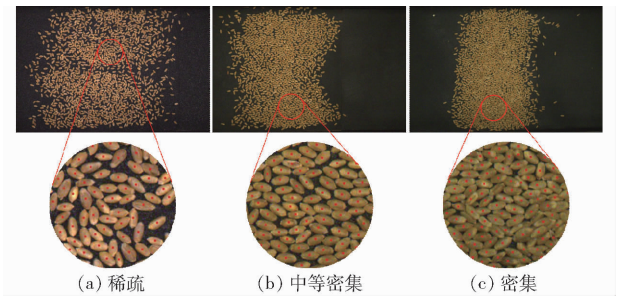


图 11 各离散度等级计数效果

Fig. 11 Counting effects of each dispersion level

为验证 Soft - NMS 对拼接后的预测框进行后处理的有效性,在表 3 中增加了 YOLO v7 - ST 模型不进行后处理的对照组试验结果。结果表明,不进行

后处理时模型的计数准确率大幅下降,3 个离散度等级下的平均 MCA 仅有 38.74%,下降了 60.42 个百分点。精确率同样下降明显,但召回率有明显提升,平均召回率 95.37%,提升 5.28 个百分点,这是因为未进行后处理的模型保留了拼接后的所有预测框,因此模型查全能力更强,但其中包含大量裁剪处被误分割的“破损籽粒”检测出的低置信度预测框,导致模型查准能力较差,即精确率低,而添加 Soft - NMS 可以保留“破损籽粒”完整部分预测出的高置信度预测框,并剔除仅靠破损部分识别出的低置信度预测框。经后处理的 YOLO v7 - ST 模型计数准确率和 F1 值大幅提升,充分证明了 Soft - NMS 后处理的重要性。

表 3 各离散度等级计数结果评价
Tab.3 Evaluation of counting results of each dispersion level

模型	离散度等级	MCA/ %	P/ %	R/ %	F1 值/ %	MCT/ s
YOLO v7 - ST	稀疏	99.93	99.32	91.60	95	1.23
	中等密集	99.80	96.75	90.76	94	1.19
	密集	97.74	91.40	87.91	90	1.15
	平均值	99.16	95.82	90.09	93	1.19
YOLO v7 - ST - 无后处理	稀疏	45.38	63.63	99.12	78	1.21
	中等密集	36.80	60.72	97.17	75	1.19
	密集	34.03	55.60	89.81	69	1.10
	平均值	38.74	59.98	95.37	74	1.17

3.2 不同网络模型预测结果对比

为验证本文网络模型的检测效果,在相同测试集下,分别与目标检测领域的主流模型 YOLO v7、YOLO v5、Faster R - CNN 进行对比,结果如表 4 所示。

表 4 不同模型检测结果对比
Tab.4 Comparison of detection results for different models

模型	离散度等级	MCA/ %	P/ %	R/ %	F1 值/ %	MCT/ s
YOLO v7 - ST	稀疏	99.93	99.32	91.60	95	1.23
	中等密集	99.80	96.75	90.76	94	1.19
	密集	97.74	91.40	87.91	90	1.15
YOLO v7	稀疏	99.95	99.29	91.97	95	1.19
	中等密集	99.04	95.43	89.82	93	1.12
	密集	95.39	87.14	84.83	86	1.01
YOLO v5	稀疏	99.90	99.15	91.20	95	1.7
	中等密集	98.77	95.35	88.61	92	1.57
	密集	91.80	85.85	79.24	82	1.54
Faster R - CNN	稀疏	96.30	89.42	85.17	87	1.64
	中等密集	87.13	83.47	77.69	80	1.54
	密集	67.73	63.34	60.92	62	1.47

在检测准确率方面,籽粒离散度等级为“稀疏”时,YOLO 系列模型均有极高的识别准确率,准确率达到 99.90%,Faster R-CNN 准确率较低,仅有 96.30%;当籽粒离散度等级为“中等密集”时,各模型平均计数准确率相较“稀疏”等级,YOLO v7-ST、YOLO v7、YOLO v5 和 Faster R-CNN 分别下滑 0.13、0.91、1.13、9.17 个百分点,F1 值分别下降 1、2、3、7 个百分点,由于 YOLO v7、YOLO v5 和 Faster R-CNN 模型缺少注意力机制对特征的高效提取,易将多个粘连籽粒识别为同一籽粒而造成漏检,因此 YOLO v7-ST 模型对于粘连籽粒的检测效果更优;当籽粒离散度等级为“密集”时,籽粒之间产生严重遮挡,各模型均存在不同程度的漏检,平均计数准确率分别为 97.74%、95.39%、91.80% 和 67.73%,F1 值分别为 90%、86%、82% 和 62%,相较于另外 3 种模型,YOLO v7-ST 模型的检测准确率分别提高 2.35、5.94、30.01 个百分点,F1 值分别提高 4、8、28 个百分点,可见 Swin Transformer 检测头可有效提高遮挡目标的检测效果。检测时间方面,以 YOLO v7 为基线的模型检测速度更快,由于 YOLO v7-ST 增加了模型复杂度,但 Swin Transformer 基于局部移位窗口方法有效降低了参数量,检测时间与 YOLO v7 检测时间相近。

3.3 不同注意力机制预测结果对比

为验证本文网络模型注意力机制的优越性,在相同测试集下,比较了 Swin Transformer 与通道注意力机制 SENet、双重注意力机制 CBAM、自注意力机制 Self Attention 的效果差异,结果如表 5 所示。

对比几种注意力机制的检测结果可知,基于自

注意力机制的网络模型在“密集”等级下的检测精度更好,这是由于自注意力机制依赖于目标之间的相似度,尽管被遮挡的目标存在一定的特征丢失,但可以通过周围的籽粒特征近似表达被遮挡籽粒,具有较强的可解释性。而 SENet 和 CBAM 依赖于通道或空间上的特征重要性,通过卷积的权重学习调整特征的强弱,缺少相互之间的联系。此外,对比 Self Attention 和 Swin Transformer 的多头注意力机制,其本质都是通过非线性变换将输入映射在新的空间中,而本文的多头注意力可以将输入映射在 6 个子空间,再通过特征聚合捕捉所有空间特征信息,信息更丰富,效果更好。

表 5 不同注意力机制检测结果对比
Tab.5 Comparison of detection results for different attention algorithms

模型	离散度等级	MCA/ %	P/ %	R/ %	F1 值/ %	MCT/ s
YOLO v7-ST	稀疏	99.93	99.32	91.60	95	1.23
	中等密集	99.80	96.75	90.76	94	1.19
	密集	97.74	91.40	87.91	90	1.15
YOLO v7-SENet	稀疏	98.67	98.07	90.45	94	1.35
	中等密集	98.01	94.44	87.69	91	1.2
	密集	94.23	86.08	84.75	85	1.18
YOLO v7-CBAM	稀疏	99.91	99.16	91.21	95	1.75
	中等密集	99.15	95.72	88.95	92	1.64
	密集	95.73	89.53	85.13	87	1.61
YOLO v7-Self Attention	稀疏	99.87	99.12	91.17	95	1.22
	中等密集	98.90	95.48	88.73	92	1.13
	密集	96.16	89.93	85.51	88	1.07

图 12 为“稀疏”、“中等密集”和“密集”3 种籽粒离散度等级下模型检测结果热力图,由图 12 可

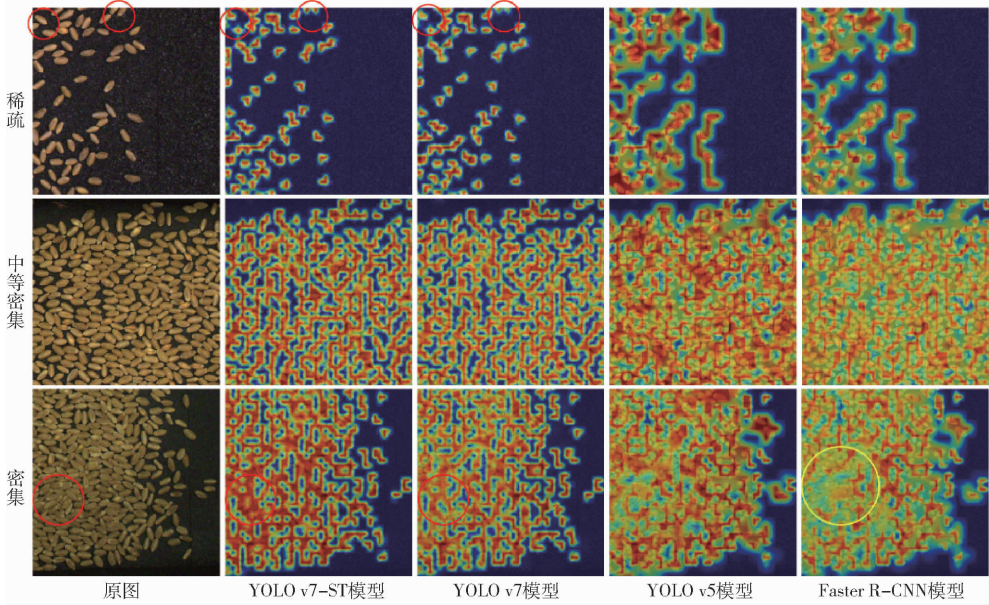


图 12 不同模型检测结果热力图

Fig. 12 Heat map of test results for different models

知,Faster R-CNN 模型识别效果最差,由于其仅关注籽粒中心局部特征,对大部分籽粒轮廓识别不完整,且在图中黄色圆圈标注位置,存在大量密集堆积籽粒的漏检情况。YOLO v5 模型对籽粒识别注意力范围较广,对于密集籽粒注意力低,导致该模型在不同离散度等级下的识别泛化性能弱。YOLO v7 模型对图像背景和目标的区分度更高,识别效果明显优于 YOLO v5;相较于 YOLO v7,YOLO v7-ST 的 Swin Transformer 检测头对全局特征以及被遮挡目标具有更强的检测能力,如图中红色圆圈标注位置,该部分热力图颜色更深,表明 YOLO v7-ST 更能关注到边缘籽粒和遮挡籽粒等检测难点,在 4 种模型中有最高的识别准确率。

4 结论

(1)基于电磁振动原理设计了小麦籽粒振动分离装置,通过籽粒受力分析,研究了振动频率、振幅、输送倾角和摩擦因数对籽粒离散程度的影响,并基于多因素正交试验方案建立了不同籽粒离散度的试验样本集。引入二阶离散系数评价方法,结合图像处理结果和籽粒分布特征将籽粒离散度划分为“稀疏”、“中等密集”和“密集”等级。

(2)提出了一种 YOLO v7-ST 网络模型的高通量小麦籽粒计数方法,并对不同离散度等级试验样

本进行了测试,试验结果表明,YOLO v7-ST 模型对“稀疏”、“中等密集”和“密集”籽粒的平均计数准确率分别为 99.93%、99.80% 和 97.74%,总平均值为 99.16%;模型综合评价指标 F1 值分别为 95%、94% 和 90%,总平均值为 93%;平均计数时间分别为 1.23、1.19、1.15 s,总平均值为 1.19 s,模型检测精度和效率能满足小麦室内考种的检测需求。

(3)为验证本文模型的优越性,与常用的目标检测模型以及其他注意力机制算法进行了对比。籽粒“稀疏”等级时,本文模型的 MCA 与 YOLO v7 和 YOLO v5 结果相近,相比 Faster R-CNN 提高 3.63 个百分点;“中等密集”等级时,本文模型的 MCA 相较于 YOLO v7、YOLO v5 和 Faster R-CNN 分别提高 0.76、1.03、12.67 个百分点,随籽粒离散度等级为“密集”时,MCA 进一步提高 2.35、5.94、30.01 个百分点,此外,本文模型的 F1 值分别提高 4、8、28 个百分点。由此可见,随离散度等级变化,YOLO v7-ST 模型的检测精度更高,针对不同离散度等级的泛化能力更强。YOLO v7 结合不同的注意力机制时,Swin Transformer 的多头注意力机制也比其他通道注意力、空间注意力和自注意力的检测准确率更高。本文模型的 MCT 与 YOLO v7 相近,比 YOLO v5 和 Faster R-CNN 模型快 0.4 s 左右。

参 考 文 献

- [1] 蒋赞,张丽丽,薛平,等. 我国小麦产业发展情况及国际经验借鉴[J]. 中国农业科技导报, 2021, 23(7): 1-10.
JIANG Yun, ZHANG Lili, XUE Ping, et al. Development status of wheat industry in China and international experience for reference[J]. Journal of Agricultural Science and Technology, 2021, 23(7): 1-10. (in Chinese)
- [2] 宋鹏,张晗,王成,等. 玉米高通量自动考种装置设计与试验[J]. 农业工程学报, 2017, 33(16): 41-47.
SONG Peng, ZHANG Han, WANG Cheng, et al. Design and experiment of high throughput automatic measuring device for corn [J]. Transactions of the CSAE, 2017, 33(16): 41-47. (in Chinese)
- [3] 王华. 做好优质专用小麦品种区域试验的措施[J]. 现代农业科技, 2011(1): 117, 119.
WANG Hua. Measures for regional test of high quality wheat varieties[J]. Modern Agricultural Science and Technology, 2011 (1): 117, 119. (in Chinese)
- [4] 林耀海,吕钟亮,杨长才,等. 自然场景图像中的重叠蜜柚识别及试验[J]. 农业工程学报, 2021, 37(24): 158-167.
LIN Yaohai, LÜ Zhongliang, YANG Changcai, et al. Recognition of the overlapped honey pomelo images in natural scene and experiment[J]. Transactions of the CSAE, 2021, 37(24): 158-167. (in Chinese)
- [5] MEBATSION H K, PALIWAL J. A Fourier analysis based algorithm to separate touching kernels in digital images[J]. Biosystems Engineering, 2011, 108(1): 66-74.
- [6] 荀一,鲍官军,杨庆华,等. 粘连玉米籽粒图像的自动分割方法[J]. 农业机械学报, 2010, 41(4): 163-167.
XUN Yi, BAO Guanjun, YANG Qinghua, et al. Automatic segmentation method of adhesive corn kernel image [J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(4): 163-167. (in Chinese)
- [7] VISEN N S, SHASHIDHAR N S, PALIWAL J, et al. AE—automation and emerging technologies[J]. Journal of Agricultural Engineering Research, 2001, 79(2): 159-166.
- [8] 李立君,阳涵疆. 基于改进凸壳理论的遮挡油茶果定位检测算法[J]. 农业机械学报, 2016, 47(12): 285-292, 346.
LI Lijun, YANG Hanjiang. Revised detection and localization algorithm for *Camellia oleifera* fruits based on convex hull theory [J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(12): 285-292, 346. (in Chinese)
- [9] 张霖,赵祚喜,可欣荣,等. 压电式种子计数系统[J]. 农业机械学报, 2011, 42(8): 41-45.
ZHANG Lin, ZHAO Zuoxi, KE Xinrong, et al. Seed-counting system design using piezoelectric sensor[J]. Transactions of the

- Chinese Society for Agricultural Machinery, 2011, 42(8): 41–45. (in Chinese)
- [10] 赵博, 樊学谦, 周利明, 等. 气流输送播种机压电式流量传感器设计与试验[J]. 农业机械学报, 2020, 51(8): 55–61. ZHAO Bo, FAN Xueqian, ZHOU Liming, et al. Design and test of piezoelectric flow sensor for pneumatic seeder[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(8): 55–61. (in Chinese)
- [11] 应玉明. 基于单片机控制的工件自动计数系统设计[J]. 机电工程技术, 2011, 40(7): 47–49. YING Yuming. Design of workpiece automatic counting system based on single chip microcomputer control[J]. Electromechanical Engineering Technology, 2011, 40(7): 47–49. (in Chinese)
- [12] 朱学岩, 张新伟, 才嘉伟, 等. 基于无人机图像和贝叶斯 CSRNet 模型的粘连云杉计数[J]. 农业工程学报, 2022, 38(14): 43–50. ZHU Xueyan, ZHANG Xinwei, CAI Jiawei, et al. Adhesion spruce counting based on UAV images and Bayesian CSRNet model[J]. Transactions of the CSAE, 2022, 38(14): 43–50. (in Chinese)
- [13] 宋怀波, 王云飞, 段援朝, 等. 基于 YOLO v5-MDC 的重度粘连小麦籽粒检测方法[J]. 农业机械学报, 2022, 53(4): 245–253. SONG Huaibo, WANG Yunfei, DUAN Yuanchao, et al. Detection method of severe adhesive wheat grain based on YOLO v5-MDC model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(4): 245–253. (in Chinese)
- [14] BRAGUY J, RAMAZANOVA M, GIANCOLA S, et al. SeedQuant: a deep learning-based tool for assessing stimulant and inhibitor activity on root parasitic seeds[J]. Plant Physiology, 2021, 186(3): 1632–1644.
- [15] SUN J, ZHANG Y, ZHU X, et al. Deep learning optimization method for counting overlapping rice seeds[J]. Journal of Food Process Engineering, 2021, 44(9): e13787.
- [16] ZHAO J, ZHANG X, YAN J, et al. A wheat spike detection method in UAV images based on improved YOLOv5[J]. Remote Sensing, 2021, 13(16): 3095.
- [17] XU L, WANG Y, SHI X, et al. Real-time and accurate detection of citrus in complex scenes based on HPL-YOLOv4[J]. Computers and Electronics in Agriculture, 2023, 205: 107590.
- [18] 龙燕, 李南南, 高研, 等. 基于改进 FCOS 网络的自然环境下苹果检测[J]. 农业工程学报, 2021, 37(12): 307–313. LONG Yan, LI Nannan, GAO Yan, et al. Apple fruit detection under natural condition using improved FCOS network[J]. Transactions of the CSAE, 2021, 37(12): 307–313. (in Chinese)
- [19] 黄成龙, 张忠福, 华向东, 等. 基于改进 Faster R-CNN 和 Deep Sort 的棉铃跟踪计数[J]. 农业机械学报, 2023, 54(6): 205–213. HUANG Chenglong, ZHANG Zhongfu, HUA Xiangdong, et al. Cotton boll tracking and counting based on improved Faster R-CNN and Deep Sort[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(6): 205–213. (in Chinese)
- [20] 闫彬, 樊攀, 王美茸, 等. 基于改进 YOLOv5m 的采摘机器人苹果采摘方式实时识别[J]. 农业机械学报, 2022, 53(9): 28–38, 59. YAN Bin, FAN Pan, WANG Meirong, et al. Real-time apple picking pattern recognition for picking robot based on improved YOLOv5m[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(9): 28–38, 59. (in Chinese)
- [21] 宋世国, 王其超. 直线式电磁振动上料器的设计[J]. 机械设计, 1987(5): 30–35. SONG Shiguo, WANG Qichao. Design of linear electromagnetic vibrating feeder[J]. Mechanical Design, 1987(5): 30–35. (in Chinese)
- [22] 王应彪, 赵学观, 徐丽明, 等. 基于电磁振动的玉米种子定向排序输送技术[J]. 农业机械学报, 2015, 46(1): 79–88. WANG Yingbiao, ZHAO Xueguan, XU Liming, et al. Experiment and directional movement technology of corn seed based on electromagnetic vibration[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(1): 79–88. (in Chinese)
- [23] 王艳, 周辉. 电磁振动给料机参数的确定原则[J]. 橡塑技术与装备, 2010, 36(2): 33–38. WANG Yan, ZHOU Hui. Determination principle of electromagnetic vibration feeder parameters[J]. Rubber Plastic Technology and Equipment, 2010, 36(2): 33–38. (in Chinese)
- [24] 俞亚新, 赵匀, 张斌, 等. 基于稻种胚胎定向排列的排种器及参数优化[J]. 江苏大学学报(自然科学版), 2008, 29(3): 194–197. YU Yaxin, ZHAO Yun, ZHANG Bin, et al. Seed metering device based on orienting ordered arrangement in embryo and parameter optimization[J]. Journal of Jiangsu University (Natural Science Edition), 2008, 29(3): 194–197. (in Chinese)
- [25] 赵树行. 电磁振动给料机最佳运动学参数的探讨[J]. 起重运输机械, 1982(7): 28–32. ZHAO Shuxing. Discussion on the optimal kinematics parameters of electromagnetic vibration feeder[J]. Hoisting and Transportation Machinery, 1982(7): 28–32. (in Chinese)
- [26] KVALSETH T O. Coefficient of variation: the second-order alternative[J]. Journal of Applied Statistics, 2017, 44(3): 402–415.
- [27] ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup: beyond empirical risk minimization[J]. arXiv Preprint, arXiv:1710.09412, 2017.

39(11): 1344 – 1350.

LI Jie, LI Shangke, JIANG Liwen, et al. A nondestructive method identifying varieties of green tea based on near infrared spectroscopy and chemometrics[J]. Journal of Instrumental Analysis, 2020, 39(11): 1344 – 1350. (in Chinese)

[16] ZHAO A, TANG X, ZHANG Z, et al. Optimizing Savitzky-Golay parameters and its smoothing pretreatment for FTIR gas spectra[J]. Journal of Instrumental Analysis, 2016, 36(5): 1340 – 1344.

[17] WU Y, PENG S, XIE Q, et al. An improved weighted multiplicative scatter correction algorithm with the use of variable selection; application to near-infrared spectra[J]. Chemometrics and Intelligent Laboratory Systems, 2019, 185(1): 114 – 121.

[18] USTUNDAG O, DINC E. Continuous wavelet transforms and ultra performance liquid chromatography applied to the simultaneous quantitative determination of candesartan cilexetil and hydrochlorothiazide in tablets[J]. Monatshefte fur Chemie, 2021, 52: 1097 – 1106.

[19] XU Y, ZHANG D, SONG F, et al. A method for speeding up feature extraction based on KPCA[J]. Neurocomputing, 2007, 70(4): 1056 – 1061.

[20] CENTNER V, MASSART D L. Elimination of uninformative variables for multivariate calibration[J]. Anal. Chem., 1996, 68: 3851 – 3858.

[21] 叶荣珂,孔庆辰,李道亮,等. 宽度学习的虾新鲜度检测方法[J]. 光谱学与光谱分析, 2022, 42(1): 164 – 169.

YE Rongke, KONG Qingchen, LI Daoliang, et al. Shrimp freshness detection method based on broad learning system[J]. Spectroscopy and Spectral Analysis, 2022, 42(1): 164 – 169. (in Chinese)

[22] 褚小立. 化学计量学方法与分子光谱分析技术[M]. 北京: 化学工业出版社, 2011.

[23] CHEN J, SUN D. Partial least squares regression (PLSR) applied to NIR and HSI spectral data modeling to predict chemical properties of fish muscle[J]. Food Engineering Reviews, 2017, 9(1): 36 – 49.

[24] CERVANTES F, GARCIA-LAMONT L, RODRIGUEZ-MAZAHUA A. A comprehensive survey on support vector machine classification: applications, challenges and trends[J]. Neurocomputing, 2020, 408: 189 – 215.

[25] JIA J, YUAN S, SHI Y, et al. Improved sparrow search algorithm optimization deep extreme learning machine for lithium-ion battery state-of-health prediction[J]. iScience, 2022, 25(4): 103988 – 104005.

[26] JIN J, LIU Z, CHEN C. Discriminative graph regularized broad learning system for image recognition[J]. Science China Information Sciences, 2018, 61(11): 1 – 14.

[27] 李国强,徐立庄. 局部感受野的宽度学习算法及其应用[J]. 计算机工程与应用, 2020, 56(9): 162 – 167.

LI Guoqiang, XU Lizhuang. Application of local receptive field based broad learning system[J]. Computer Engineering and Applications, 2020, 56(9): 162 – 167. (in Chinese)

[28] DU J, VONG C M, CHEN C P. Novel efficient RNN and LSTM-like architectures: recurrent and gated broad learning systems and their applications for text classification[J]. IEEE Transactions on Cybernetics, 2020, 51(3): 1586 – 1597.

(上接第 197 页)

[28] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. arXiv Preprint, arXiv:2004.10934, 2020.

[29] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. arXiv Preprint, arXiv:2207.02696, 2022.

[30] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012 – 10022.

[31] ZHU X, LV S, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 2778 – 2788.

[32] DENG J, DONG W, SOCHER R, et al. Imagenet: a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 248 – 255.

[33] LOSHCHILOV I, HUTTER F. SGDR: stochastic gradient descent with warm restarts[J]. arXiv Preprint, arXiv:1608.03983, 2016.

[34] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS—improving object detection with one line of code[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 5561 – 5569.