doi:10.6041/j.issn.1000-1298.2025.11.044

基于 CBLP - YOLO 11n 的无人机稻穗轻量化检测方法

王　雪[1]　高　雅[1]　陶桂香[1]　马铁民[1]　张　楠[1]　许善祥[1]　于　庆[2]
(1. 黑龙江八一农垦大学信息与电气工程学院, 大庆 163319; 2. 宁安市东京城镇农业技术推广服务中心, 牡丹江 157421)

摘要: 稻穗准确计数对估算水稻产量至关重要。然而在实际生产场景中,由于稻田环境复杂、水稻品种繁多、穗部形态多异等原因,现有检测方法存在精度不足、模型参数量大的问题。为此本研究提出一种轻量级稻穗检测模型 CBLP - YOLO 11n。首先,在 Backbone 中,选择 C3k2 - CFCGLU 替换原 C3k2 模块,增强模型对稻穗的特征提取和表达能力;其次,使用加权双向特征金字塔 BiFPN 实现多尺度特征信息融合,通过跳跃连接和删除冗余节点的方式,在提高稻穗识别精度的同时,有效降低模型浮点运算量;然后,设计轻量细节增强共享检测头(Lightweight detail-enhanced shared detection head,LDSDH),通过共享卷积降低检测头复杂度;最后采用 Powerful - IoUv2 损失函数替换原有的 CIoU 损失函数,加快模型收敛速度,并优化模型对稻穗的定位准确性。实验结果表明:CBLP - YOLO 11n 模型的检测精确率为 88.2%,召回率为 87.9%,平均精度均值为 93.9%。与 YOLO 11n 相比,CBLP - YOLO 11n 的精确率提高 1.9 个百分点,召回率提高 1.1 个百分点,平均精度均值提高 1.3 个百分点,参数量下降 23.7%,浮点运算量下降 40.6%。相比于其它主流模型,CBLP - YOLO 11n 模型的平均精度均值最高,内存占用量最小,仅为 3.78 MB。所提模型实现了对稻穗的精确识别,可部署在无人机等资源受限设备中,为复杂田间背景下稻穗识别计数提供技术支撑。

关键词: 稻穗计数; 无人机图像; YOLO 11n; 轻量化模型; 多尺度特征融合
中图分类号: TP391.4　　**文献标识码**: A　　**文章编号**: 1000-1298(2025)11-0461-10　　**OSID**:

Lightweight Detection Method of Rice Panicles Based on CBLP - YOLO 11n

WANG Xue[1]　GAO Ya[1]　TAO Guixiang[1]　MA Tiemin[1]　ZHANG Nan[1]　XU Shanxiang[1]　YU Qing[2]
(1. *College of Information Technology*, *Heilongjiang Bayi Agricultural University*, *Daqing* 163319, *China*
2. *Agricultural Technology Extension Service Center*, *Ning'an Dongjing Town*, *Mudanjiang* 157421, *China*)

Abstract: Rice is a globally important food crop and accurate counting of rice panicles is crucial for estimating rice production. However, in actual production scenarios, due to reasons such as complex paddy field environment, diversity of rice varieties, and their panicle morphological features, existing detection methods have insufficient accuracy and a large number of model parameters. To this end, a lightweight rice panicles detection model CBLP - YOLO 11n was proposed. Firstly, in Backbone, the original C3k2 module was replaced with C3k2 - CFCGLU in order to enhance the model's feature extraction and expression ability for rice panicles. Secondly, bidirectional feature pyramid network (BiFPN) was used to achieve multi-scale feature information fusion. By means of skip connections and deletion of redundant nodes, the recognition accuracy of rice panicles was improved while effectively reducing the computational complexity of the model. Then a lightweight detail-enhanced shared detection head (LDSDH) was designed and the complexity of detection head was reduced through shared convolution. Finally, the original CIoU loss function was replaced by Powerful - IoUv2 loss function to accelerate the convergence speed of the model and optimize positioning accuracy of the model for rice panicles. The experimental results showed that the detection precision of the CBLP - YOLO 11n model was 88.2%, the recall was 87.9% and the mean average precision was 93.9%. Compared with YOLO 11n, the CBLP - YOLO 11n model showed improvements in the precision, recall, mAP by 1.9, 1.1 and 1.3 percentage points, respectively. Meanwhile, the CBLP - YOLO 11n reduced the parameter quantity

收稿日期: 2025 - 05 - 05　修回日期: 2025 - 06 - 12
基金项目: 黑龙江省自然科学基金联合引导项目(LH2024E103)、黑龙江八一农垦大学人才引进科研启动项目(XDB202115)和大庆市指导性科技计划项目(zd - 2025 - 046)
作者简介: 王雪(1980—),女,副教授,博士,主要从事图像处理、近红外光谱及其在农业上的应用研究,E-mail: mtmwx@163.com

by 23.7% and the computational quantity by 40.6%. Compared with other mainstream models, the CBLP-YOLO 11n model had the highest average detection accuracy and the smallest memory usage, which was only 3.78 MB. The proposed model realized accurate identification of rice panicles and can be deployed in resource-constrained devices such as drones, providing technical support for identification and counting of rice panicles in complex field backgrounds.

Key words: count of rice panicles; UAV images; YOLO 11n; lightweight model; multi-scale feature fusion

0 引言

水稻是我国的主要粮食作物,提高水稻产量是水稻研究的重要目标之一[1]。开展产量估算有助于及时掌握水稻的生长状况,为农业生产决策提供科学依据,同时也为保障国家粮食安全和宏观调控提供重要支撑。水稻产量和单位面积穗数有着密切关系,因此,快速准确获取单位面积穗数可以提高水稻估产的准确性和时效性。然而传统人工计数稻穗费时费力而且非常容易出错,近年来,由于无人机具有低成本、高灵活性、非接触式和较高分辨率的特点,无人机遥感图像已被广泛应用于农作物监测和计数等任务中[2]。通过选择适当的近地面高度,使用无人机获取大量的农田图像数据,再利用深度学习对图像进行分析处理,可以实现对粮食作物或经济作物的精确计数和估产[3]。

传统的基于机器学习的计数方法曾被广泛应用[4-5]。在大田复杂环境下,传统图像处理方法对纹理、颜色等视觉特征的依赖性较强,其性能易受光照、稻穗遮挡等因素的影响,算法的鲁棒性较差且难以处理大规模田间图像数据。

与传统机器学习方法相比,深度学习在特征提取和复杂数据处理方面具有显著优势。目前,基于深度学习的方法已在农作物检测与计数中获得广泛应用[6-9],主要分为图像分割和目标检测 2 大类[10-12]。图像分割算法的关键在于对目标边缘的精确判断,这依赖于像素级别的图像处理,容易受到大田复杂背景因素的干扰,同时因为计算的复杂度较高,需要更多的计算能力和较长的处理时间[13],对硬件资源要求较高。

与基于分割的稻穗识别方法相比,基于目标检测的算法能够以较小的计算量实现稻穗的快速准确定位。因此,可采用目标检测网络实现无人机影像中的稻穗识别与计数,该方法避免了像素级分类带来的计算负担,在保证检测精度的同时显著提高了处理效率[14-17]。

基于以上分析,同时考虑到大田背景下稻穗存在小目标、多遮挡、水面倒影和不均匀光照的问题,提高稻穗的识别精度和实现网络模型轻量化是研究的重点。针对上述问题,本文提出 CBLP-YOLO 11n 模型,该模型使用加权双向特征金字塔 BiFPN 实现多尺度特征融合,采用 Powerful-IoUv2 损失函数加快模型收敛速度,设计 C3k2-CFCGLU 和 LDSDH 替换原模型中的 C3k2 模块和检测头,使模型在具有较高准确率的同时实现模型轻量化,在复杂田间环境下具有较高鲁棒性。

1 材料与方法

1.1 无人机水稻稻穗图像采集

无人机遥感图像采集自黑龙江八一农垦大学安达农业科技园(125°20′59″E, 46°24′9″N),研究区域位置如图 1 所示。拍摄时间为 2024 年 8 月 28—29 日 09:00—11:30 和 13:30—17:00,天气晴朗。使用大疆无人机(DJI Air 2S 型)拍摄图像,分辨率为 5 472 像素×3 648 像素,无人机飞行高度设置为 3 m,摄像机角度调整为垂直向下,共采集 116 幅图像。

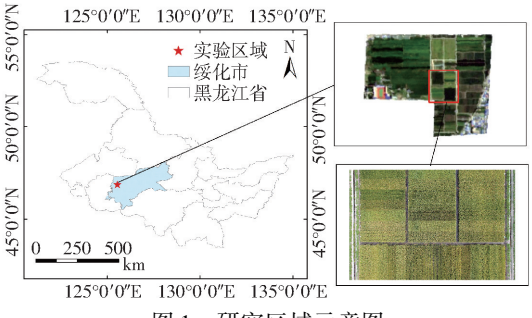


图 1 研究区域示意图

Fig. 1 Schematic of study area

1.2 数据处理

1.2.1 数据裁剪

在训练模型之前,为便于模型训练,同时充分利用高分辨率图像所包含的信息,将无人机拍摄所得图像连续分割为多个尺寸为 512 像素×512 像素的子图像,每个子图像拥有 10% 的重叠区域,然后使用 LabelImg 工具对稻穗进行标记。原始有效图像的数量为 1 281 幅,将数据集按照比例 8∶1∶1 随机分为训练集、验证集和测试集。

1.2.2 数据扩充

由于农田环境多变,无人机图像易受天气、光照和背景等因素影响,稻穗特征信息较为复杂,这些都

为实际田间大规模稻穗图像检测和计数带来一定困难。自然状态下拍摄的稻穗图像主要包括清晰、小目标、弯曲、遮挡、强光和重叠等图像，这些自然场景的差异反映了数据集的复杂性和多样性，能更好地验证检测算法的鲁棒性[18]。同时为了丰富数据集，本研究还对训练集采取了数据增强扩充处理，包括水平垂直翻转、随机旋转 90°、随机裁剪、添加高斯噪声、调节亮度和对比度等操作。原始训练集包含 1 024 幅图像，数据增强后扩充为 1 536 幅图像，这些图像在放大数据集的同时也提高了模型的鲁棒性和泛化能力[19]。

1.2.3　超分辨率重建

考虑到无人机拍摄时受高度、风速、设备和网络传输的影响，部分图像的分辨率不高。为获取图像更多细节特征，本研究采用更适用于真实世界场景的 Real－ESRGAN 模型对图像进行超分辨率重建[20]。Real－ESRGAN 由生成器和判别器 2 部分组成，生成器将低分辨率图像重建为高分辨率图像，而判别器的任务是区分生成器输出的图像与真实高分辨图像，通过对抗训练迫使生成器不断优化，增强生成器的重建能力。

Real－ESRGAN 的生成器核心模块是双重残差密集块（Residual-in-Residual dense block，RRDB），它由多个残差密集块（Residual dense block，RDB）堆叠而成。每个 RDB 通过密集连接和局部跳跃连接来增强特征传播和梯度流动，设输入特征为 x，其公式为

$$F_{\mathrm{RDB}}(x)=x+W_{\mathrm{fusion}}[F_0,F_1,\cdots,F_N] \tag{1}$$

$$F_{\mathrm{RRDB}}=x+\sum_{i=1}^{M}\alpha_i F_{\mathrm{RDB}_i}(x) \tag{2}$$

式中　W_{fusion}——1×1 卷积

F_N——第 N 层输出

α_i——残差缩放因子

$F_{\mathrm{RDB}}(x)$——单个残差密集块 RDB 输出

F_{RRDB}——双重残差密集块 RRDB 输出

RRDB 通过密集连接和双重残差设计，能够有效地提取和融合图像特征，捕捉稻穗图像细节和纹理信息，提高模型对稻穗的细节重建能力。

Real－ESRGAN 的判别器采用 U－Net 结构，如图 2 所示。该结构由编码器和解码器组成，并通过跳跃连接将低层和高层特征融合，既能捕捉图像的全局语义信息又能保留局部细节，但其复杂结构也导致训练的不稳定性增加，因此引入谱归一化（Spectral normalization，SN）技术，通过约束判别器权重矩阵的谱范数，可以防止梯度爆炸或消失，稳定训练过程，同时缓解图像训练导致的过度锐化和伪影问题。

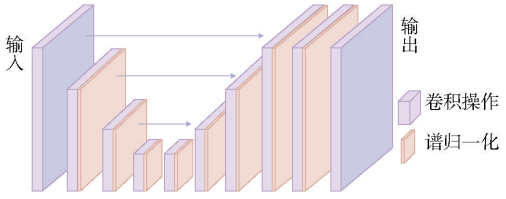
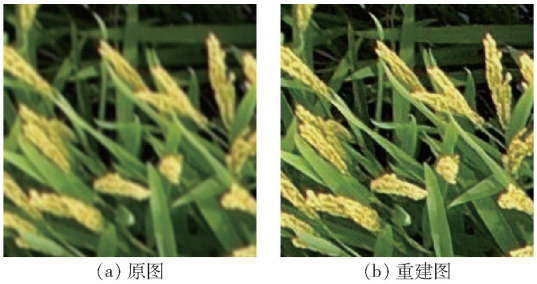


图 2　判别器结构图

Fig. 2　Discriminator structure diagram

对原始训练集的 1 024 幅图像使用 Real－ESRGAN 模型进行超分辨率重建，而验证集和测试集则保持不变。重建前后图像的对比如图 3 所示，可以看出修复后的图像细节更加丰富，边缘更加清晰，模型可学习到更多的特征。



（a）原图　　（b）重建图

图 3　超分重建前后稻穗图像

Fig. 3　Rice panicle images before and after reconstruction

2　模型构建与训练

2.1　YOLO 11 网络结构

YOLO 11 是 Ultralytics 团队在 2024 年 9 月发布的高效一阶段目标检测算法[21]，其在 YOLO v8 的基础上进行改进，主要由主干网络（Backbone）、颈部网络（Neck）和头部网络（Head）3 部分组成。具体改进为采用 C3k2 替换原 C2f 模块，该结构有助于聚合来自不同尺度的特征，增强特征提取和融合能力；Backbone 中增加一个 C2PSA 模块，是在 C2 中引入多头注意力机制，进一步提高模型对重要特征的捕捉能力；在 Head 部分，在解耦头中的分类分支中加入 2 个深度卷积（Depthwise convolution，DWConv），减少模型的参数量和浮点运算量。

YOLO 11n 是 YOLO 11 系列最小的网络模型，拥有较高的检测精度和较快的检测速度，且模型内存占用量也较小，具有轻量化、高精度和实时性的特点，可以部署在无人机等资源受限的设备中，实现对大田中稻穗的快速检测，本文在 YOLO 11n 的基础上进行改进。

2.2　CBLP－YOLO 11n 模型网络结构

由于大田环境下水稻图像具有背景复杂、稻穗相互遮挡的特点，而稻穗目标检测一般都部署在无人机等边缘设备上，这类设备的计算能力有限，因此需要对模型进行轻量化改进，本研究提出 CBLP－

YOLO 11n 网络结构。

以 YOLO 11n 为基线模型进行下列改进:首先,将 Backbone 中的 C3k2 全都替换为 C3k2 - CFCGLU 模块,通过整合通道信息,突出重要特征并抑制噪声和不重要信息,增强模型的特征表达能力;其次,使用 BiFPN 取代 Neck 中原特征融合方法 PANet,增强不同网络层之间特征信息的传递,实现模型的多尺度特征融合,增强对小目标和相互遮挡目标的识别精度;然后设计了 LDSDH 检测头,采用共享卷积和 DeConv 来减少参数量,提高检测速度;此外,为加快边界框回归速度,引入 Powerful - IoUv2 损失函数弥补原 CIoU 损失函数锚框过度增大和收敛速度慢的问题,改进后的网络模型结构如图 4 所示。

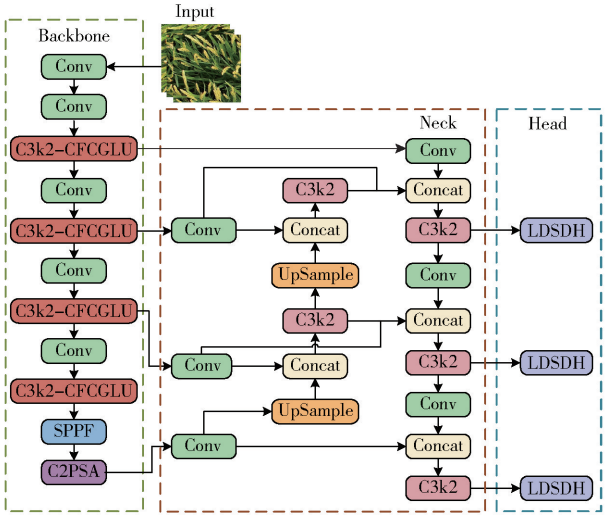


图 4 CBLP - YOLO 11n 模型网络结构图

Fig. 4 CBLP - YOLO 11n model network structure diagram

2.2.1 C3k2 - CFCGLU 模块

在颈部网络中,本研究设计了 C3k2 - CFCGLU 模块,该模块将 C3k2 中的 Bottleneck 替换为 CFCGLU 模块。CFCGLU 由 ConvFormer[22-23] 和卷积门控线性单元[24](Convolutional gated linear unit, CGLU)组成,其结构如图 5 所示。

ConvFormer 是 Transformer 的变形,它将 Transformer 编码器中的 Self - Attention 机制替换为可分离卷积(Separable convolution, SEConv),保留 Transformer 的原始架构,形成了一个不含任何注意力机制的纯 CNN 模型。SEConv 先通过逐点卷积进行通道扩展,提升特征的多样性,然后通过深度卷积进行空间特征提取,最后再通过逐点卷积压缩通道。该方法在降低原模型参数量和浮点运算量的同时也保持了较强的特征提取能力。

CFCGLU 在 ConvFormer 的基础上将 Channel MLP 替换为 CGLU,CGLU 是一种通道混合器,在门控线性单元[25](Gated linear unit, GLU)门控分支的

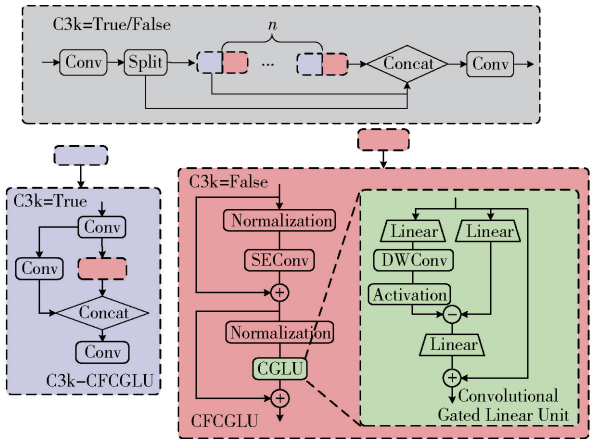


图 5 C3k2 - CFCGLU 网络结构图

Fig. 5 C3k2 - CFCGLU model network structure diagram

激活函数前添加了 3 × 3 的 DWConv,这样可以利用深度卷积提取的局部特征信息调整门控信号,实现门控通道注意力机制,可以根据输入动态地选择重要的特征,抑制不重要的特征,同时还可以作为位置编码满足 CFCGLU 模块对位置信息的需求,有利于稻穗的目标定位和实现更准确的边界框回归预测。

表 1 为 YOLO 11n 使用 C3k2 - CCFCGLU 改进前后的特征提取网络结构,将 Backbone 中的4个

表 1 YOLO 11n 改进前后的 Backbone 网络结构

Tab. 1 Backbone network structure of YOLO 11n and Backbone network structure of improved YOLO 11n

模型	模块	卷积核大小	输出通道数	浮点运算量	参数量
原 YOLO 11n	Conv	3 × 3	64		
	Conv	3 × 3	128		
	C3k2		256	$3.276\,8 \times 10^8$	6 520
	Conv	3 × 3	256		
	C3k2		512	$3.276\,8 \times 10^8$	25 840
	Conv	3 × 3	512		
	C3k2		512	$2.752\,51 \times 10^8$	86 528
	Conv	3 × 3	1 024		
	C3k2		1 024	$2.752\,51 \times 10^8$	345 088
	SPPF		1 024		
	C2PSA		1 024		
改进后 YOLO 11n	Conv	3 × 3	64		
	Conv	3 × 3	128		
	C3k2 - CCFCGLU		256	$3.716\,1 \times 10^8$	7 400
	Conv	3 × 3	256		
	C3k2 - CCFCGLU		512	$3.305\,09 \times 10^8$	26 108
	Conv	3 × 3	512		
	C3k2 - CCFCGLU		512	$1.874\,85 \times 10^8$	59 068
	Conv	3 × 3	1 024		
	C3k2 - CCFCGLU		1 024	$1.821\,65 \times 10^8$	228 664
	SPPF		1 024		
	C2PSA		1 024		

C3k2 都替换为 C3k2－CCFCGLU 模块，虽然改进后的前 2 个 C3k2－CCFCGLU 的浮点运算量和参数量略有提高，但 Backbone 的总浮点运算量降低 11.1%，参数量降低 30.8%。

2.2.2　BiFPN 结构

YOLO 11n 中的 PANet[26] 采用双向特征融合方法提取图像的低层细节信息和高层语义信息，网络可以在不同尺度上进行特征交互，使得模型在检测不同尺寸目标时都能充分利用多尺度特征，从而提高检测精度。然而自底向上和自顶向下的路径聚合会增加网络的计算量，尤其是在处理高分辨率图像时，会导致推理速度变慢，因此本文将其替换为 BiFPN 结构[27]。

BiFPN 在 PANet 的基础上，通过融合不同层级的特征，删除仅有一个输入且没有特征融合的节点，增加跳跃连接通路和自适应权重分配机制，可以减少信息在传递过程中的丢失，确保高层和低层特征的有效融合，提高模型的检测精度。PANet 和 BiFPN 的网络结构如图 6 所示。

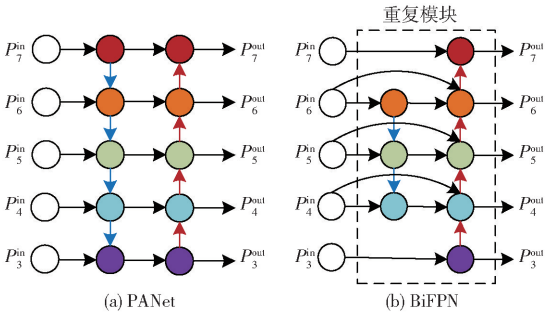


图 6　PANet 和 BiFPN 网络结构图

Fig. 6　PANet and BiFPN network structure diagrams

对于具有不同分辨率的特征，BiFPN 引入自适应权重分配机制，在特征融合过程中为每个输入添加一个额外的权重，让网络能够学习到每个输入的重要特征，从而实现更高级别的特征融合，这有助于识别和区分具有相似特征的目标。相较于 PANet 结构，BiFPN 拥有更少的网络节点，并在处于同一层级的输入和输出节点之间增加一条额外的跳跃连接，在不增加太多计算量前提下融合了更多的特征。将每个双向路径视为一个特征网络层重复多次，从而实现更高级的特征融合。

针对本文所提出的稻穗检测模型而言，引入 BiFPN 能够在不增加额外计算成本的同时显著提高模型性能。通过多尺度特征融合，能更全面地提取稻穗图像的关键信息，增强深层特征对稻穗目标的语义表达能力。相较于 PANet，BiFPN 的跳跃连接与权重动态分配策略，能够更充分地复用不同层级特征信息，从而在保持模型高效性的同时，实现检测精度提升与漏检率降低，更适合实际田间大规模稻穗的检测。

2.2.3　LDSDH 检测头

YOLO 11n 采用的解耦头由多个卷积层构成，各卷积层通过独立的卷积操作处理特征图，并不共享参数，这种设计虽然能在一定程度上提高检测精度，但检测头的参数量和计算量极大。针对这一问题，本研究提出一种轻量细节增强共享检测头 LDSDH，如图 7 所示。该检测头旨在降低计算成本的同时，增强对细节特征的捕捉能力，提高稻穗的检测精度。

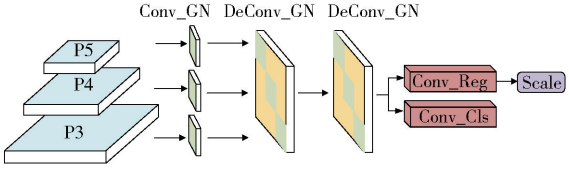


图 7　LDSDH 网络结构图

Fig. 7　LDSDH network structure diagram

首先，将检测头中的基础 Conv（图 8a）改进为 Conv_GN（图 8b）模块，该模块将卷积后提取的特征通过组内归一化的方式在通道上进行分组，提高检测头分类定位的性能。其次，为降低参数量，采用共享卷积来提取 P3、P4 和 P5 层的多尺度特征。此外，针对不同检测头对应目标尺度不一致的问题，引入 Scale 层对特征进行动态缩放，使得模型能够自适应地调整特征表示，这一设计有效降低了检测头的复杂度，并进一步提升了检测精度。

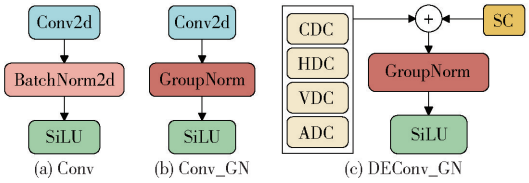


图 8　3 种卷积网络结构图

Fig. 8　Three convolutional network structure diagrams

然而，特征缩放操作可能导致检测精度下降。因此，本研究设计了 DEConv_GN 模块，在 Conv_GN 基础上引入细节增强卷积[28]（Detail enhanced convolution，DEConv）替代原有卷积，在保证模型轻量化的同时，进一步提升特征提取能力，其结构如图 8c 所示。通过并行部署 5 种卷积：标准卷积（Standard convolution，SC）、中心差分卷积（Central difference convolution，CDC）、水平差分卷积（Horizontal difference convolution，HDC）、垂直差分卷积（Vertical difference convolution，VDC）和角度差分卷积（Angular difference convolution，ADC）来提取特征，然后使用分组归一化将通道分组。其中标准卷积用于获得强度级信息，更关注像素的实际亮度

或颜色值,能有效捕捉图像中整体的色彩和亮度特征。而差分卷积用于增强梯度级信息,重点提取稻穗的边缘、纹理和形态特征,从而提升模型在复杂环境下的鲁棒性和泛化能力。

为进一步优化模型性能,避免因并行部署 5 个卷积导致参数量和推理时间增加,采用卷积的重参数化技术将多分支卷积结构简化为单个卷积,计算公式为

$$F_{out} = \mathrm{DEConv}(F_{in}) = \sum_{i=1}^{5} F_{in}K_i = F_{in}\sum_{i=1}^{5} K_i = F_{in}K_{cvt} \tag{3}$$

式中 K_i——CDC、HDC、VDC、ADC 和 VC 卷积核

K_{cvt}——卷积加法操作

$\mathrm{DEConv}(\ )$——细节增强卷积操作

F_{in}、F_{out}——输入特征、输出特征

LDSDH 在降低计算复杂度的同时,实现对全局信息和背景信息高效捕获,可以达到在复杂背景环境下提升稻穗定位识别精度的目的。

2.2.4 Powerful - IoUv2 损失函数

边界框回归对目标检测至关重要,通过对锚框微调使其逼近真实框,从而实现模型的定位性能。YOLO 11n 使用 CIoU[29] 作为边界框回归的损失函数,CIoU 通过锚框与真实框的重叠面积、中心点距离和纵横比来计算损失,但其存在不合理的惩罚因子,导致收敛过程中出现锚框异常增大,收敛速度减慢的现象。因此,本研究引入 Powerful - IoUv2(PIoUv2)损失函数,该函数通过优化损失计算方式,有效解决锚框异常增大的问题,在加快收敛速度的同时,提升了边界框的定位精度,增强模型性能。

Powerful - IoU[30](PIoU)是一种结合尺寸自适应惩罚因子和梯度调整机制的边界框回归损失函数,该函数优化了边界框回归过程,帮助锚框更加直接有效地回归。PIoUv2 则是在 PIoU 的基础上引入了一个由单个超参数控制的非单调注意力层,增强对中质量和高质量锚框的聚焦能力,提升模型的检测精度。PIoUv2 的表达式为

$$P = \left(\frac{\mathrm{d}\omega_1}{\omega_{gt}} + \frac{\mathrm{d}\omega_2}{\omega_{gt}} + \frac{\mathrm{d}h_1}{h_{gt}} + \frac{\mathrm{d}h_2}{h_{gt}}\right)\Big/4 \tag{4}$$

$$L_{PIoU} = L_{IoU} + 1 - \mathrm{e}^{-P^2} \quad (0 \leqslant L_{PIoU} \leqslant 2) \tag{5}$$

$$q = \mathrm{e}^{-P} \quad (q \in (0,1]) \tag{6}$$

$$u(\lambda q) = 3\lambda q \mathrm{e}^{-x^2} \tag{7}$$

$$L_{PIoUv2} = u(\lambda q)L_{PIoU} = 3(\lambda q)\mathrm{e}^{-(\lambda q)^2}L_{PIoU} \tag{8}$$

式中 ω_{gt}、h_{gt}——真实框的宽和高

$\mathrm{d}\omega_1$、$\mathrm{d}\omega_2$、$\mathrm{d}h_1$、$\mathrm{d}h_2$——真实框和锚框交集到最小外接框的距离

P——PIoU 的惩罚因子

q——PIoUv2 的惩罚因子

$u(\lambda q)$——注意力函数

λ——注意力函数的超参数

L_{PIoU}——Powerful - IoU 损失函数

L_{IoU}——IoU 损失函数

L_{PIoUv2}——Powerful - IoUv2 损失函数

公式(4)中所涉及的参数如图 9 所示。

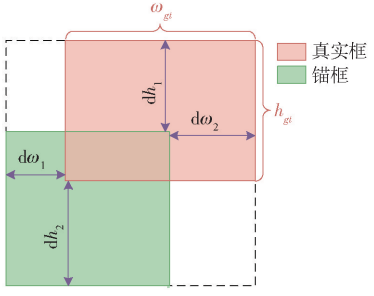


图 9 PIoUv2 参数示意图

Fig. 9 PIoUv2 parameter diagram

PIoUv2 的惩罚因子 q 与最小外接框的尺寸无关,因此扩大锚框不会导致收敛缓慢,这有效解决了锚框异常放大问题。通过锚框和真实框的大小和质量动态调整梯度,加快了模型的收敛速度,有效提高模型对稻穗中重叠和小目标的识别精度和速度。

3 结果与分析

3.1 参数配置及评价指标

实验所用操作系统为 Windows 11,CPU 型号为 12th Gen Intel(R) Core(TM) i9 - 12900H@2.50 GHz,GPU 型号为 NVIDIA GeForce RTX3060,运行内存为 8 GB,编程语言为 Python 3.9.19,深度学习框架为 Pytorch 1.12.1,CUDA 及 GPU 加速库 cuDNN 版本分别为 11.6 和 8.8.1。训练时图像尺寸调整为 640 像素×640 像素,初始学习率设置为 0.01,批量大小设置为 16,工作线程数设置为 4,动量设置为 0.937,迭代次数为 400,采用随机梯度下降 SGD 作为优化器。采用精确率(Precision,P)、召回率(Recall,R)、平均精度均值(Mean average precision,mAP)、浮点运算量、参数量和模型内存占用量评估模型性能。其中,$\mathrm{mAP}_{0.5}$是指 IoU 阈值设为 0.5 时的平均精度。

3.2 数据增强对检测结果的影响

由于无人机拍摄所得稻穗图像受复杂大田环境等因素的影响,为提高检测精度,分别使用 albumentations 库和 Real - ESRGAN 对图像进行增强处理。为了验证图像增强处理对无人机稻穗检测的有效性,使用 YOLO 11n 分别对经过增强处理和未经过增强处理的图像进行训练,训练所得模型再

对测试集进行检测，实验结果如表 2 所示。在未经过增强处理的数据集上模型的平均精确率为 91.8%，而在经过增强处理的数据集上模型的平均精确率为 92.6%，证实了数据增强的必要性和有效性。

表 2　数据增强后检测结果

Tab. 2　Detection results of data enhancement　%

数据处理	P	R	$mAP_{0.5}$
无增强	86.0	85.6	91.8
增强	86.3	86.8	92.6

3.3　损失函数

为验证不同损失函数对模型的优化效果，分别对比 CIoU、GIoU、SIoU、MPDIoU、Wise－IoUv3 和本文所采用的 Powerfur－IoUv2 损失函数在 YOLO 11n 中的性能表现，实验结果如表 3 所示，损失函数收敛速度如图 10 所示。

表 3　不同损失函数模型性能对比

Tab. 3　Performance comparison of different loss function models　%

损失函数	P	R	$mAP_{0.5}$
CIoU	86.3	86.8	92.6
GIoU	87.1	86.9	92.6
SIoU	87.1	86.7	93.0
MPDIoU	87.4	86.4	93.0
Wise－IoUv3	87.1	86.3	92.8
Powerful－IoUv2	87.7	87.2	93.2

由表 3 可知，相较于其它常用损失函数，本文提出的 PIoUv2 在精确率、召回率和 $mAP_{0.5}$ 均最高，与 YOLO 11n 中原 CIoU 损失函数相比，PIoUv2 的精确率、召回率和 $mAP_{0.5}$ 值分别高 1.4、0.4、0.6 个百分点。由图 10 可知，PIoUv2 在模型训练期间的收敛速度快于其它损失函数，并保持与其他算法相同的收敛趋势，其最终收敛值为 0.586 05，是所有损失函数中最低的。结果表明，PIoUv2 作为损失函数，可以加快模型收敛速度，提高模型检测精度。

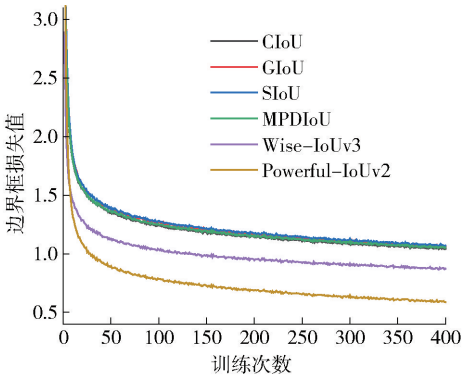


图 10　损失函数收敛曲线

Fig. 10　Convergence curves of loss function

3.4　消融实验

为验证本文提出的改进模块的优化效果，在 YOLO 11n 的基础上进行了消融实验，实验结果如表 4 所示。相较于基线模型，模型 1 主干网络采用 C3k2－CFCGLU 模块，相比于原始模型的召回率提高 0.8 个百分点，说明该模块能够更好地捕捉复杂背景下稻穗的细节特征，减少稻穗漏检的可能性；模型 2 的颈部网络加入 BiFPN 模块后，相比于原始模型的平均精度均值提高 0.6 个百分点，其参数量下降 25.5%，说明 BiFPN 可以将不同层级的特征图进行更全面的融合，并且模块中的跳跃连接可以显著减少模型的参数量。模型 3 对检测头 LDSDH 的改进使得模型的平均精度均值相比于原模型提高 0.6 个百分点，参数量和浮点运算量分别降低 24.3% 和 15.3%，但推理速度略有降低，说明该模块在牺牲一定速度的情况下提高了模型的检测精度并实现轻量化。模型 5 是 BiFPN 和 LDSDH 在基线网络上的组合，其平均精度均值没有单独加这 2 个模块的模型 2 和模型 3 高，但其参数量有了明显优化，相较于基线模型下降 35.1%。模型 6 在基线模型上分别加入了 C3k2－CFCGLU、BiFPN 和 LDSDH 模块，其平

表 4　消融实验结果

Tab. 4　Ablation experiment results

模型	C3k2－CFCGLU	BiFPN	LDSDH	PIoUv2	P/%	R/%	$mAP_{0.5}$/%	浮点运算量	参数量	推理速度/(f·s^{-1})
YOLO 11n					86.3	86.8	92.6	6.313×10^9	2.582×10^6	122.44
模型 1	√				86.9	87.6	92.9	6.179×10^9	2.440×10^6	96.21
模型 2		√			88.3	86.9	93.2	6.277×10^9	1.923×10^6	94.29
模型 3			√		88.2	86.4	93.2	4.777×10^9	2.186×10^6	66.10
模型 4	√	√			86.8	86.5	93.1	6.143×10^9	1.780×10^6	95.83
模型 5		√	√		88.4	86.3	93.0	4.950×10^9	1.676×10^6	71.01
模型 6	√	√	√		87.2	86.8	93.4	4.816×10^9	1.534×10^6	71.80
模型 7		√	√	√	88.6	86.7	93.5	4.950×10^9	1.676×10^6	113.45
CBLP－YOLO 11n	√	√	√	√	88.2	87.9	93.9	4.816×10^9	1.534×10^6	103.81

注：“√”表示使用此模块。

均精度均值较基线网络提高 0.8 个百分点,参数量和浮点运算量也有了显著降低。模型 7 融合了 BiFPN、LDSDH 和 PIoUv2 损失函数,其平均精度均值相较于基线模型提高 0.9 个百分点,虽然其推理速度略有下降,但相比于仅引入 BiFPN 和 LDSDH 改进的模型 5,模型 7 在推理速度和检测精度方面均有所提升。

综合以上改进,本研究提出的模型 CBLP-YOLO 11n 取得了较好的检测性能。其精确率、召回率和平均精度均值分别提升 1.9、1.1、1.3 个百分点,分别达到 88.2%、87.9%、93.9%。参数量和浮点运算量也分别下降 23.7% 和 40.6%,实现了模型的轻量化,同时推理速度提高到 103.81 f/s,可部署在无人机上实现对稻穗的快速准确检测。

3.5 不同主流模型检测效果对比

为验证改进后的模型性能,将改进后的模型与目前主流模型 SSD、Faster R-CNN、YOLO 系列[31-34]等基于 CNN 的目标检测模型和基于 Transformer 的 RT-DETR[35]模型进行对比,结果如表 5 所示。

表 5 不同模型对比实验结果

Tab. 5 Comparison of test results with different models

模型	$mAP_{0.5}$/%	浮点运算量	参数量	推理速度/(f·s^{-1})	模型内存占用量/MB
SSD	89.4	$3.042\,8\times10^{10}$	$2.374\,6\times10^{7}$	40.30	91.67
Faster R-CNN	87.1	$1.372\,16\times10^{11}$	$4.134\,8\times10^{7}$	17.60	315.79
YOLO v5n	91.5	5.973×10^{9}	2.216×10^{6}	82.60	4.47
YOLO v7-tiny	90.6	$1.302\,2\times10^{10}$	6.007×10^{6}	74.60	11.72
YOLO v8n	92.4	8.085×10^{9}	3.006×10^{6}	88.90	5.99
RT-DETR	86.3	$1.259\,76\times10^{11}$	$4.013\,4\times10^{7}$	25.80	63.11
YOLO v9t	89.3	$1.069\,3\times10^{10}$	2.617×10^{6}	25.30	5.83
YOLO v10n	91.2	8.568×10^{9}	2.763×10^{6}	85.50	5.55
YOLO 11n	92.6	6.313×10^{9}	2.582×10^{6}	122.44	5.27
YOLO 12n	92.3	5.819×10^{9}	2.509×10^{6}	81.40	5.23
本文模型	93.9	4.816×10^{9}	1.534×10^{6}	103.81	3.78

从表 5 可知,YOLO 11n 与 YOLO 系列其它轻量级模型相比,平均精度均值比 YOLO v5n、YOLO v7-tiny、YOLO v8n、YOLO v9t、YOLO v10n、YOLO 12n 分别高 1.1、2.0、0.2、3.3、1.4、0.3 个百分点,其浮点运算量、参数量和模型内存占用量也是除 YOLO 12n 以外最小的,推理速度达到 122.44 f/s,可以为后期部署在计算资源有限的硬件设备上奠定基础。改进后的 CBLP-YOLO 11n 模型,平均精度均值达到 93.9%,比原模型提高 1.3 个百分点,其模型内存占用量最小,比原模型减少 28.3%,尽管推理速度略有下降,但仍可满足实时检测的需求。

与其他非轻量级模型相比,改进后模型的平均精度均值相比 SSD、Faster R-CNN 和 RT-DETR 分别高 4.5、6.8、7.6 个百分点,除此之外,本文提出的模型的参数量、浮点运算量和模型内存占用量相比于这 3 种网络低的多,部署在无人机上有较大优势。

综上,本文提出的 CBLP-YOLO 11n 模型可以在稻穗目标检测中平衡检测精确度和检测速度,综合性能较佳。

3.6 检测结果

为了进一步验证 CBLP-YOLO 11n 模型的可靠性和泛化性,本文在测试集上进行了验证。图 11 展示了多种场景下的原图、原模型检测图和改进后模

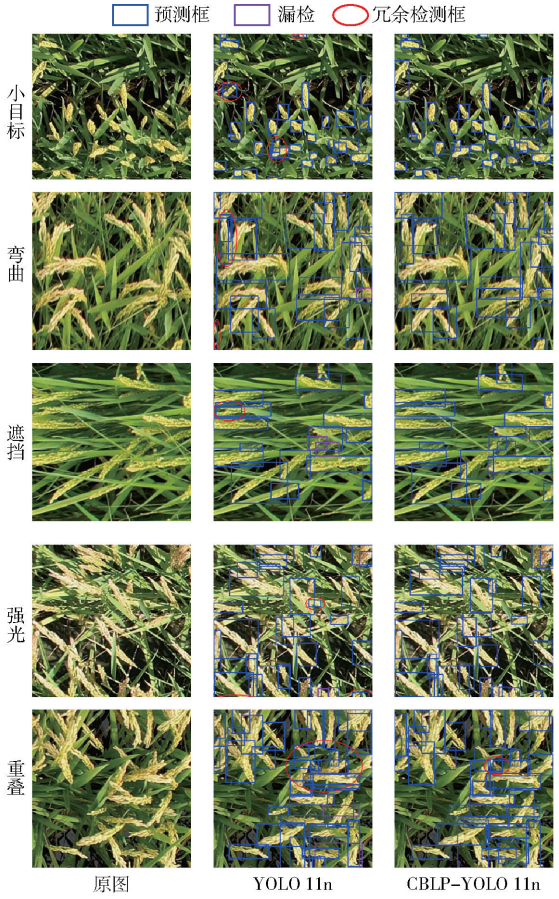


图 11 不同模型稻穗检测结果

Fig. 11 Detection results of panicles with different models

型检测图。

通过观察图像可以看出，在弯曲重叠和遮挡的情况下，由于重叠稻穗之间的边界不清晰，原模型难以准确提取稻穗的位置信息，容易产生漏检的情况。小目标稻穗由于稻穗目标较小，特征不够明显，而强光会使黄绿色的稻叶偏向黄色，稻穗本身也是黄色，导致稻穗与背景颜色区分度太低，因此这 2 种场景下模型都出现重复检测框。

图 12 为图 11 在误检和漏检的局部放大示意图，通过图 12 可以更明显看出，本研究模型基本可以在多种复杂场景和背景下正确识别稻穗，有效降低了稻穗的漏检率和误检率。

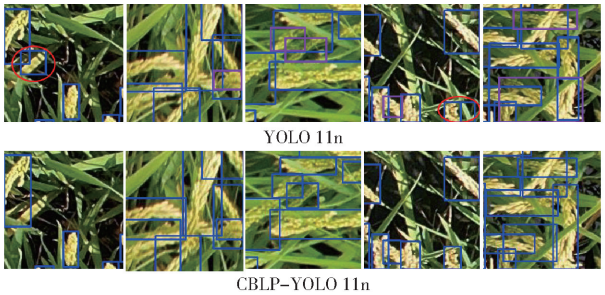


图 12　部分检测结果局部放大图

Fig. 12　Partial enlarged view of test results

3.7　计数相关性研究

为验证本研究提出的检测模型对稻穗的计数效果，本文对测试集中 129 幅稻穗图像进行计数实验，将模型改进前后的预测值和人工标记值进行线性拟合，结果如图 13 所示。从拟合结果来看，基础模型对稻穗的真实值和预测值拟合的 R^2 为 0.80，而改进后的模型的 R^2 提高到 0.86，说明改进后的模型可以更准确地对稻穗进行计数。综上，本文提出的轻量化模型 CBLP－YOLO 11n 可以准确地对稻穗进行检测和计数，有助于实现精准的稻穗估产。

4　结论

（1）为实现复杂田间背景下稻穗的快速检测计数，本研究通过无人机拍摄、图像分割、数据增强和超分辨率重建的方式建立自然田间水稻图像数据集，用于模型训练。

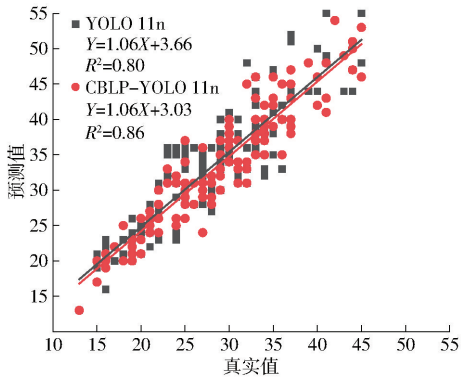


图 13　稻穗真实值与预测值拟合结果

Fig. 13　Fitting result between true value and predicted value of panicles

（2）基于 YOLO 11n 设计了一种用于稻穗检测的轻量化模型 CBLP－YOLO 11n。首先，在主干网络中，引入 C3k2－CFCGLU 模块替换原 C3k2，提高了模型对稻穗的目标定位能力；其次，使用 BiFPN 融合多尺度特征信息，提高检测性能，并通过跳跃连接和删除多余节点的方式有效降低了模型的参数量和浮点运算量；然后，设计轻量细节增强共享检测头 LDSDH，降低了检测头复杂度；最后采用 Powerful－IoUv2 损失函数替换原有的 CIoU 损失函数，实现了对稻穗快速精准实时检测的目的。

（3）消融实验结果表明，改进后模型的精确率、召回率和平均精度均值分别提升 1.9、1.1、1.3 个百分点，参数量和浮点运算量也分别下降 23.7% 和 40.6%，实现了模型的轻量化。CBLP－YOLO 11n 模型的平均精度均值为 93.9%，参数量为 1.534×10^6，浮点运算量为 4.816×10^9，模型内存占用量为 3.78 MB。相比于其它主流模型，改进后的模型检测精度最高，内存占用量最小。

（4）CBLP－YOLO 11n 模型可对输入稻穗图像进行自动计数，具有较高鲁棒性，可以有效降低人工计数成本并提高稻穗计数准确性，为复杂田间背景下无人机稻穗估产提供技术支撑。

参　考　文　献

[1] 冯向前，王爱冬，洪卫源，等. 基于低空无人机遥感的水稻产量估测方法研究进展[J]. 中国水稻科学，2024，38(6)：604－616.
FENG Xiangqian，WANG Aidong，HONG Weiyuan，et al. Research progress in rice yield estimation method based on low-altitude unmanned aerial vehicle remote sensing[J]. Chinese Journal of Rice Science，2024，38(6)：604－616.（in Chinese）

[2] GAO R，JIN Y，TIAN X，et al. YOLOv5－T：a precise real-time detection method for maize tassels based on UAV low altitude remote sensing images[J]. Computers Electronics in Agriculture，2024，221：108991.

[3] 翟肇裕，张梓涵，徐焕良，等. YOLO 算法在动植物表型研究中应用综述[J]. 农业机械学报，2024，55(11)：1－20.
ZHAI Zhaoyu，ZHANG Zihan，XU Huanliang，et al. Review of applying YOLO family algorithms to analyze animal and plant phenotype[J]. Transactions of the Chinese Society for Agricultural Machinery，2024，55(11)：1－20.（in Chinese）

[4] BAI X，CAO Z，ZHAO L，et al. Rice heading stage automatic observation by multi-classifier cascade based rice spike detection method[J]. Agricultural Forest Meteorology，2018，259：260－270.

[5] ZHOU C，LIANG D，YANG X，et al. Recognition of wheat spike from field based phenotype platform using multi-sensor fusion

and improved maximum entropy segmentation algorithms[J]. Remote Sensing, 2018, 10(2): 246.
[6] DONMEZ C, VILLI O, BERBEROGLU S, et al. Computer vision-based citrus tree detection in a cultivated environment using UAV imagery[J]. Computers Electronics in Agriculture, 2021, 187: 106273.
[7] LI Z, ZHU Y, SUI S, et al. Real-time detection and counting of wheat ears based on improved YOLOv7[J]. Computers Electronics in Agriculture, 2024, 218: 108670.
[8] 李兆东, 章艳芳, 汪蕴红, 等. 基于 YC-YOLO v7 模型的油菜幼苗株数识别方法[J]. 农业机械学报,2024,55(12): 322-332.
LI Zhaodong, ZHANG Yanfang, WANG Yunhong, et al. Identification of rapeseed seedling number based on YC-YOLO v7 model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(12): 322-332. (in Chinese)
[9] 王金鹏, 何萌, 甄乾广, 等. 基于改进 COF-YOLO v8n 的油茶果静态与动态检测计数方法[J]. 农业机械学报,2024, 55(4):193-203.
WANG Jinpeng, HE Meng, ZHEN Qianguang, et al. Camellia oleifera fruit static and dynamic detection counting based on improved COF-YOLO v8n[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(4): 193-203. (in Chinese)
[10] XIAO L, PAN Z, DU X, et al. Weighted skip-connection feature fusion: a method for augmenting UAV oriented rice panicle image segmentation[J]. Computers Electronics in Agriculture, 2023, 207: 107754.
[11] 曹英丽, 赵雨薇, 杨璐璐, 等. 基于改进 DeepLabv3+的水稻田间杂草识别方法[J]. 农业机械学报,2023,54(12): 242-252.
CAO Yingli, ZHAO Yuwei, YANG Lulu, et al. Weed identification method in rice field based on improved DeepLabv3+[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(12): 242-252. (in Chinese)
[12] HONG S, JIANG Z, LIU L, et al. Improved Mask R-CNN combined with Otsu preprocessing for rice panicle detection and segmentation[J]. Applied Sciences, 2022, 12(22): 11701.
[13] WANG Y, AHSAN U, LI H, et al. A comprehensive review of modern object segmentation approaches[J]. Foundations Trends® in Computer Graphics Vision, 2022, 13(2-3): 111-283.
[14] TAN S, LU H, YU J, et al. In-field rice panicles detection and growth stages recognition based on RiceRes2Net[J]. Computers electronics in agriculture, 2023, 206: 107704.
[15] 黄硕, 周亚男, 王起帆, 等. 改进 YOLOv5 测量田间小麦单位面积穗数[J]. 农业工程学报,2022,38(16):235-242.
HUANG Shuo, ZHOU Yanan, WANG Qifan, et al. Measuring the number of wheat spikes per unit area in fields using an improved YOLOv5[J]. Transactions of the CSAE, 2022, 38(16): 235-242. (in Chinese)
[16] WEI J, TIAN X, REN W, et al. A precise plot-level rice yield prediction method based on panicle detection[J]. Agronomy, 2024, 14(8): 1618.
[17] LIANG Y, LI H, WU H, et al. A rotated rice spike detection model and a crop yield estimation application based on UAV images[J]. Computers Electronics in Agriculture, 2024, 224: 109188.
[18] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1): 1-48.
[19] YUAN G, YE T, FU H, et al. Clustering based detection of small target pedestrians for smart cities[J]. Sustainable Energy Technologies Assessments, 2022, 52: 102300.
[20] WANG X, XIE L, DONG C, et al. Real-ESRGAN: training real-world blind super-resolution with pure synthetic data[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021:1905-1914.
[21] KHANAM R, HUSSAIN M. Yolov11: an overview of the key architectural enhancements[J]. arXiv Preprint, arXiv:2410.17725, 2024.
[22] YU W, LUO M, ZHOU P, et al. Metaformer is actually what you need for vision[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022:10819-10829.
[23] YU W, SI C, ZHOU P, et al. Metaformer baselines for vision[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2023, 46(2): 896-912.
[24] SHI D. Transnext: robust foveal visual perception for vision transformers[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024:17773-17783.
[25] SHAZEER N. Glu variants improve transformer[J]. arXiv Preprint, arXiv:2002.05202, 2020.
[26] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:8759-8768.
[27] 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.
[28] CHEN Z, HE Z, LU Z-M. DEA-Net: single image dehazing based on detail-enhanced convolution and content-guided attention[J]. IEEE Transactions on Image Processing, 2024, 33: 1002-1015.
[29] 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, 2020:12993-13000.
[30] LIU C, WANG K, LI Q, et al. Powerful-IoU: more straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism[J]. Neural Networks, 2024, 170: 276-284.
[31] TIAN Y, YE Q, DOERMANN D. Yolov12: attention-centric real-time object detectors[J]. arXiv Preprint, arXiv:2502.12524, 2025.
[32] WANG A, CHEN H, LIU L, et al. Yolov10: real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.
[33] WANG C-Y, YEH I-H, MARK LIAO H-Y. Yolov9: learning what you want to learn using programmable gradient information[C]// European Conference on Computer Vision: Springer, 2024:1-21.
[34] REIS D, KUPEC J, HONG J, et al. Real-time flying object detection with YOLOv8[J]. arXiv Preprint, arXiv:2305.09972, 2023.
[35] ZHAO Y, LV W, XU S, et al. Detrs beat yolos on real-time object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024:16965-16974.