

基于改进 Oriented R-CNN 的旋转框麦穗检测与计数模型

于俊伟^{1,2,3}, 陈威威^{1,2,4}, 郭园森^{1,2,4}, 母亚双^{1,2,3}, 樊超^{1,2,3}

(1. 河南工业大学粮食信息处理与控制教育部重点实验室, 郑州 450001; 2. 河南工业大学河南省粮食光电探测与控制重点实验室, 郑州 450001; 3. 河南工业大学人工智能与大数据学院, 郑州 450001; 4. 河南工业大学信息科学与工程学院, 郑州 450001)

摘要: 为对干扰、遮挡等复杂的田野环境中麦穗进行精准定位与计数, 该研究提出了一种改进的 Oriented R-CNN 麦穗旋转框检测与计数方法, 首先在主干网络中引入跨阶段局部空间金字塔 (spatial pyramid pooling cross stage partial networks, SPPCSPC) 模块扩大模型感受野, 增强网络感知能力; 其次, 在颈网络中结合路径聚合网络 (PANet, path aggregation network) 和混合注意力机制 (E2CBAM, efficient two convolutional block attention module), 丰富特征图包含的特征信息; 最后采用柔性非极大值抑制算法 (Soft-NMS, soft-non maximum suppression) 优化预测框筛选过程。试验结果显示, 改进的模型对复杂环境中的麦穗检测效果良好。相较原模型, 平均精确度均值 mAP 提高了 2.02 个百分点, 与主流的旋转目标检测模型 Gliding vertex、R3det、Rotated Faster R-CNN、S2anet 和 Rotated Retinanet 相比, mAP 分别提高了 4.99、2.49、3.94、2.25 和 4.12 个百分点。该研究方法利用旋转框准确定位麦穗位置, 使得框内背景区域面积大幅度减少, 为实际观察麦穗生长状况和统计数量提供了一种有效的方法。

关键词: 图像识别; 作物; 注意力机制; 麦穗; Oriented R-CNN

doi: 10.11975/j.issn.1002-6819.202312068

中图分类号: S24

文献标志码: A

文章编号: 1002-6819(2024)-06-0248-10

于俊伟, 陈威威, 郭园森, 等. 基于改进 Oriented R-CNN 的旋转框麦穗检测与计数模型[J]. 农业工程学报, 2024, 40(6):

248-257. doi: 10.11975/j.issn.1002-6819.202312068 <http://www.tcsae.org>

YU Junwei, CHEN Weiwei, GUO Yuansen, et al. Improved Oriented R-CNN-based model for oriented wheat ears detection and counting[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(6): 248-257. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202312068 <http://www.tcsae.org>

0 引言

小麦是重要的粮食农作物之一, 养育了全世界近三分之一的人口。小麦产量关系到国家粮食安全, 小麦产量预测是进行农业生产管理决策的重要参考, 因此, 如何快速且准确地预估麦穗产量具有重要意义^[1-2]。

传统图像处理方法是从麦穗图像中人工提取特征, 如麦穗的形状、颜色及纹理等, 然后通过分类器构建模型从而实现对图像中麦穗识别与计数^[3-4]。FERNANDEZ-GALLEGO 等^[5]利用拉普拉斯频率滤波器去除麦穗图像中的高低频元素并使用中值滤波器去除高噪声, 分割局部峰值完成对图像中麦穗的自动计数。刘哲等^[6]改进 Bayes 抠图算法分割麦穗图像后采用图像滤波、二值化等方法分离出小穗实现麦穗小穗的自动计数。杜颖等^[7]利用颜色特征识别并提取麦穗骨架, 通过计算骨架角点实现麦穗的自动计数。但人工提取麦穗特征依靠经验丰富的专业人员且依赖麦穗图像质量, 所以在复杂的田野环境中传统图像处理技术对麦穗的计数效果并不理想^[8-9]。

目前, 基于卷积神经网络的检测模型在麦穗检测与

计数领域广泛应用。鲍文霞等^[10]利用拥挤场景识别网络 (CSRNet) 生成的麦穗密度, 根据密度图中所有密度值的总和, 精确统计单位面积麦穗数。杨蜀秦等^[11]设计单位面积小麦区域并加入特征重组上采样模块和注意力机制, 提高模型精度。KHAKI 等^[12]提出了麦穗检测模型 WheatNet, 与传统预测框进行麦穗定位相比, 该模型使用点标记麦穗目标并完成计数。黄硕等^[13]在 YOLOv5 模型基础上, 在颈网络输出分支加入混合注意力增强网络模型的感知能力, 提高模型检测能力。这些方法为麦穗定位与计数提供了有效手段, 但基于水平框的检测模型存在固有局限性, 例如对遮挡和重叠目标处理能力弱、难以处理复杂背景环境等。

旋转目标检测通过带有旋转角度的矩形进行目标检测, 能够获得目标的位置和方向信息, 有效提高检测准确性和鲁棒性, 在自动驾驶、遥感影像、工业检测和农业等领域有广泛的应用前景^[14]。李朝辉等^[15]利用多尺度技术检测不同尺度物体并加入了环形圆滑标签解决角度回归问题提高模型检测精度。为提高模型检测精度, 庄文华等^[16]采用瓶颈注意力加强对关键目标的特征提取能力, 有效提高了模型的检测精度。SUN 等^[17]利用无人机拍摄田野麦穗图像, 并提出改进的旋转框小麦头检测与计数模型, 从俯视的角度对田野环境麦穗检测与计数。

旋转目标检测虽然在许多场景下取得了较为成功的应用, 但田野环境光照、风速以及麦穗遮挡等因素导致

收稿日期: 2023-12-08 修订日期: 2024-01-14

基金项目: 国家自然科学基金青年基金项目 (62006071); 2021 年度河南省科技攻关计划项目 (212102210152); 河南工业大学粮食信息处理中心开放课题 (KFJJ2023004)

作者简介: 于俊伟, 博士, 副教授, 研究方向为计算机视觉、智能信息处理。Email: yujunwei@126.com

麦穗检测与计数难度较大,因此在麦穗检测与计数任务中需要考虑以下3方面的问题:1)麦穗受趋光特性或风的影响会倾斜生长出现重叠和遮挡,部分麦穗呈现不完整导致难以仅通过外形进行检测与计数;2)麦穗生理性差异,同种麦穗因为光照、湿度等环境因素出现大小及外形等性状差异,影响检测结果;3)田野环境的复杂性,风速以及光照导致采集的麦穗图像清晰度参差不齐,给麦穗检测与计数带来极大困难。为解决这些问题,本文拟采用 Oriented R-CNN^[18] 作为检测模型并针对麦穗检测与计数场景进行研究与优化。

针对麦穗分布密集,外形与大小存在差异,环境复杂等问题,本文在主干网络加入跨阶段局部空间金字塔(SPPCSPC,spatial pyramid pooling cross stage partial networks)^[19] 模块扩大模型感受野,加强对麦穗特征提取能力,增强网络感知能力;同时,在颈网络中引入路径聚合网络(PANet, path aggregation network)^[20] 并在特征融合分支加入混合注意力机制(E2CBAM, efficient two convolutional block attention module)丰富特征图包含的麦穗特征信息;最后使用柔性非极大值抑制算法(Soft-NMS, soft-non maximum suppression)^[21] 优化预测框筛选流程,避免因麦穗遮挡导致漏检,以期满足复杂环境下麦穗的精准快速检测与计数任务需求,为实际观察麦穗长势及数量统计提供一定的技术参考与指导。

1 材料与方法

1.1 试验数据

1.1.1 数据收集与分析

试验材料选取全球小麦数据集^[22](GWHD, <https://www.kaggle.com/competitions/global-wheat-detection/data>)和南京白马镇麦穗图像,其中 arvalis_1、arvalis_2、arvalis_3 以及 inrae_1 种类的麦穗图像拍摄于法国; enthz_1 种类的麦穗图像拍摄于瑞士; rres_1 种类的麦穗图像拍摄于英国; Usask_1 种类的麦穗图像拍摄于加拿大,将南京白马镇采集麦穗图像作为验证集检验模型鲁棒性。所选取样本包含多角度、多位置、环境多样、背景复杂的麦穗图像,如图1所示。共选取分辨率 1 024×1 024 像素 PNG 格式图像共 1 000 张。

由于麦穗形状狭长、分布密集,采用旋转框标注能更准确地表示麦穗区域,避免框内包含更多背景,两种标注的效果对比如图2所示。表1统计了采用水平框(HBB)与旋转框(OBB)标注的麦穗平均像素面积,在7类小麦图像中,旋转框使得麦穗标注的像素面积至少减少30.57%,种类为 arvalis_2 的麦穗标注框面积更是减少了62.63%,因此旋转框标注能更准确地获得麦穗特征,减少背景和相邻目标的干扰。

1.1.2 数据集构建

本试验采用 roLabelImg 标注工具对原始图片中每个麦穗目标所在的区域进行旋转框标注并注明其所属类别,按照 PASCAL VOC 格式生成标注文件(.xml),旋转框标注信息表示为 (x,y,w,h,θ) 分别表示框中心点坐标 (x,y) 、

宽和高 (w,h) 以及相较于水平框的旋转角度 θ 。因模型训练需要,将 PASCAL VOC 标注文件(.xml)转化为 DOTA 标注文件(.txt),旋转框的位置采用八参数即4个顶点坐标 (x_1,y_1) 、 (x_2,y_2) 、 (x_3,y_3) 、 (x_4,y_4) 表示。将构建的旋转框标注全球小麦数据集(RGWHD)按照 7: 1: 2 的比例随机分为训练集、验证集和测试集。

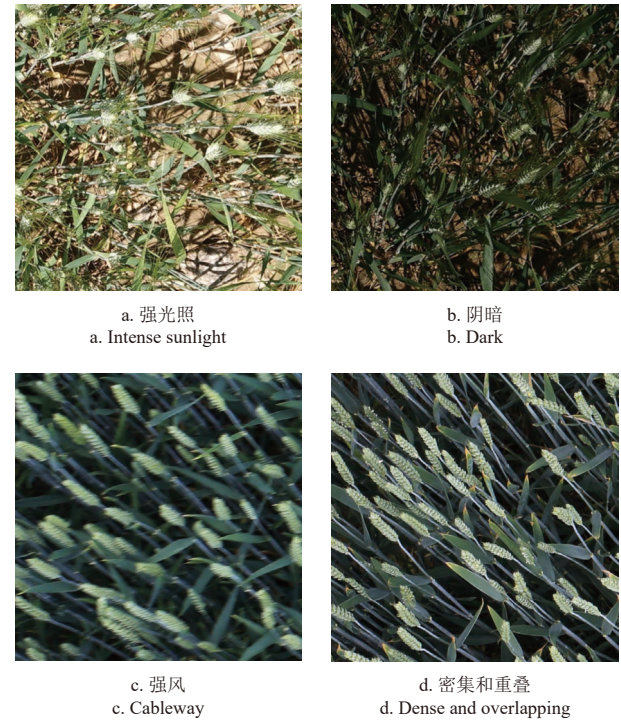


图1 麦穗样本

Fig.1 Sample diagram of wheat ears

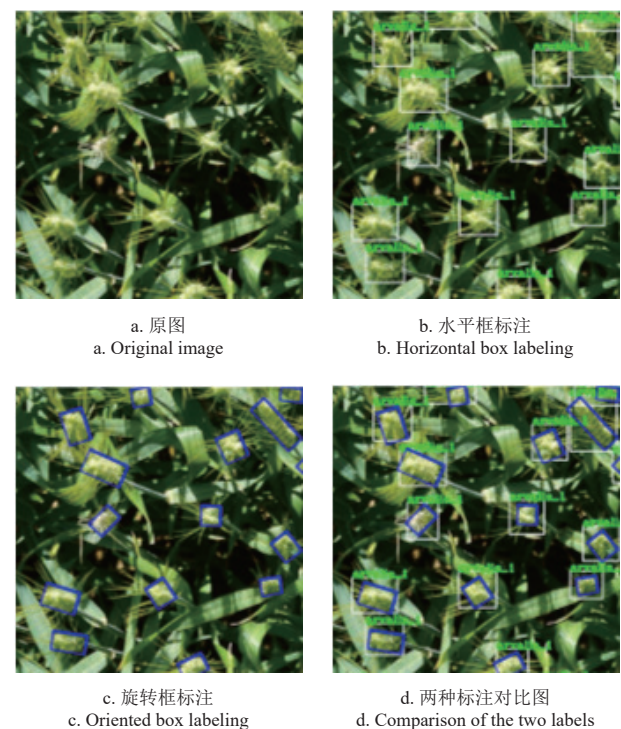


图2 水平框与旋转框视觉效果对比

Fig.2 Comparison of the visual effect of horizontal bounding boxes and oriented bounding boxes

表 1 水平框与旋转框平均面积差异
Table 1 Difference of average area of horizontal and oriented bounding boxes

麦穗种类 Wheat variety	平均面积 Average area		(HBB-OBB)/ HBB/%
	水平框 (HBB) Horizontal bounding boxes	旋转框 (OBB) Orientation bounding boxes	
arvalis_1	4 442.3	1 961.4	55.85
arvalis_2	5 918.8	2 212.1	62.63
arvalis_3	9 851.0	5 568.7	43.47
ethz_1	3 334.5	2 315.0	30.57
inrae_1	14 867.3	6 164.4	58.54
rres_1	6 140.4	3 816.4	37.85
usask_1	12 790.2	5 412.4	57.68

注: arvalis_1~arvalis_3 以及 inrae_1 麦穗拍摄于法国; ethz_1 麦穗拍摄于瑞士; rres_1 麦穗拍摄于英国; Usask_1 麦穗拍摄于加拿大。

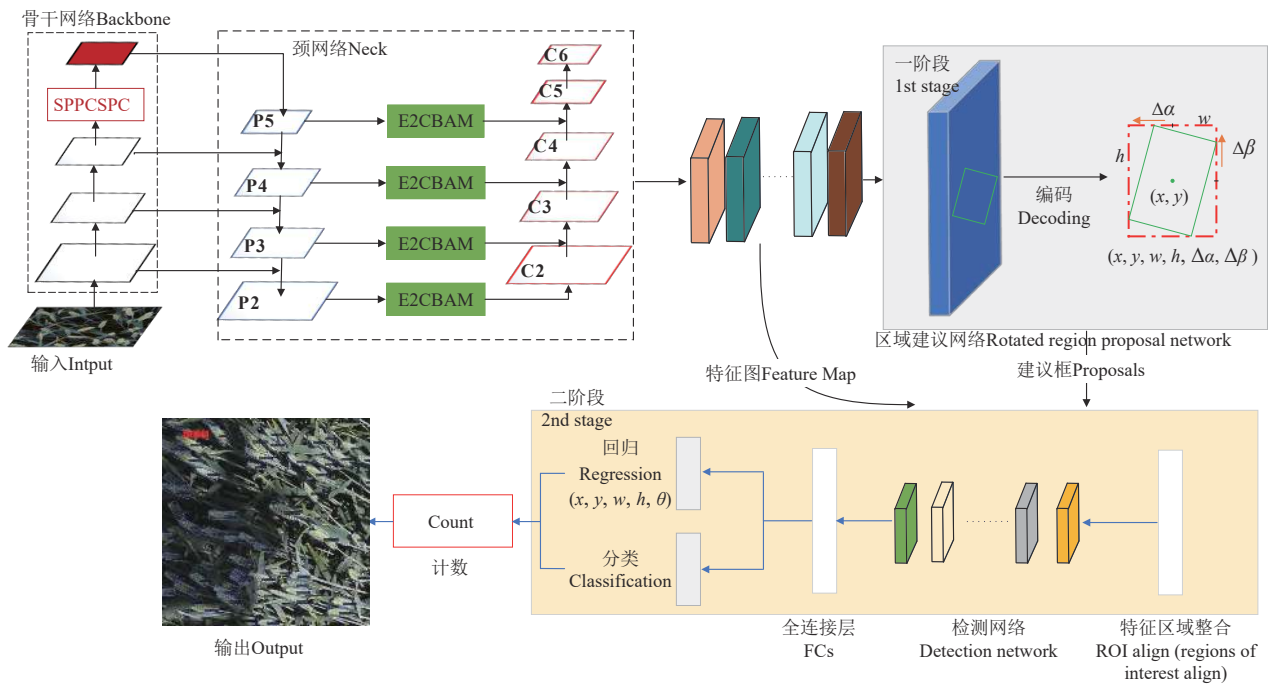
Note: arvalis_1 ~ arvalis_3 and inrae_1 wheat were taken in France; ethz_1 wheat was taken in Switzerland; rres_1 wheat was taken in England; Usask_1 wheat was taken in Canada.

1.2 基于改进 Oriented R-CNN 的麦穗检测模型构建

Oriented R-CNN 是目前旋转目标检测算法领域中检测精度相对较高的两阶段检测模型, 其主要结构包括输入端数据预处理、主干网络 (Backbone)、颈网络 (Neck)、区域候选网络 (Rotated RPN, rotated region proposal network)、头部网络 (Head)。在模型预处理阶段采用图像旋转等策略增强数据集样本多样性; 主干

网络 Backbone 使用 ResNet-50^[23] 网络对图像进行特征提取, 其核心思想是残差学习, 在残差结构中添加跨层链接, 通过堆叠残差模块构建深度网络, 解决梯度不稳定性问题; 颈网络 Neck 采用特征金字塔网络 (FPN, feature pyramid network), 通过融合不同尺度的特征图, 丰富了特征图所包含的语义和目标位置信息; 区域候选网络 (Rotated RPN) 在特征图上生成尺度不同的建议框; 头部网络 (Head) 将特征图信息与建议框信息整合输入至 ROI 层后经过全连接层得到模型的输出信息。

在田野环境中, 针对麦穗目标小、遮挡导致模型精度不理想的问题, 提出一种改进 Oriented R-CNN 的检测模型, 结构如图 3 所示, 在特征提取网络 ResNet-50 中引入跨阶段局部空间金字塔模块 SPPCSPC 增大感受野, 提升模型对小目标的检测能力; 此外, 为了进一步提高模型特征学习能力, 在颈网络 FPN 加入路径聚合网络 PANet 和混合注意力机制 E2CBAM 丰富特征图所包含的特征信息; 最后, 针对小目标检测, 采用柔性非极大值抑制算法 Soft-NMS 优化预测框的筛选过程, 避免部分预测框因筛选算法被误删导致小目标漏检。



注: P2、P3、P4、P5、C2、C3、C4、C5 和 C6 表示特征图; 变量 (x, y, w, h) 表示框的中心点坐标以及宽高; $\Delta\alpha$ 和 $\Delta\beta$ 表示边中点偏移量; θ 表示边界框回归角度; SPPCSPC 代表跨阶段局部特征金字塔模块; E2CBAM 代表混合注意力模块。

Note: P2, P3, P4, P5, C2, C3, C4, C5 and C6 represent feature maps; The variables (x, y, w, h) represent the center point coordinates of the frame and its width and height; $\Delta\alpha$ and $\Delta\beta$ represent the offset of the edge midpoint; θ represents the boundary box regression angle; SPPCSPC stands for spatial pyramid pooling cross stage partial networks; E2CBAM stands for hybrid attention mechanism module.

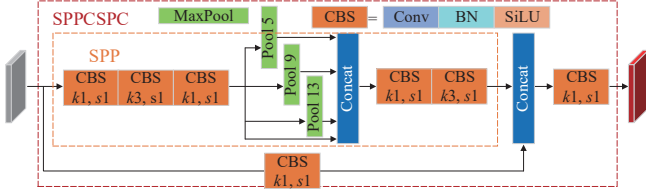
图 3 改进后的 Oriented R-CNN 结构
Fig.3 Improved Oriented R-CNN structure

1.2.1 主干网络中引入跨阶段局部空间金字塔模块

跨阶段局部空间金字塔 (SPPCSPC, spatial pyramid pooling cross stage partial networks) 模块是以空间金字塔池化模块 (SPP, Spatial Pyramid Pooling) 为基础, 结合跨阶段局部网络 (CSP, cross stage partial networks)^[24]

组合而成, 如图 4 所示。CSP 模块将主干网络提取的特征输入至卷积分支和 SPP 模块分支, 最后将两个分支输出的特征图合并进行 Concat 操作进行特征融合, 减少了模型一半计算量的同时提高了模型的速度及精度。

该操作过程细化为:



注： $k1$ 表示卷积核大小等于 1×1 ； $k3$ 表示卷积核大小等于 3×3 ； $s1$ 表示步长为 1；Pool 5、Pool 9、Pool 13 分别表示最大池化的池化核大小为 5、9、13；CBS 模块中 Conv 表示卷积操作；BN 表示批量归一化；SiLU 是激活函数。
Note: $k1$ represents the convolution kernel_size=1; $k3$ represents the convolution kernel_size=3; $s1$ represents a step size of 1; Pool 5, Pool 9, and Pool 13 indicate that the max pool core size is 5, 9, and 13, respectively; In the CBS module, Conv represents the convolution operation; BN represents batch normalization; SiLU is the activation function.

图 4 SPPCSPC 模块结构

Fig.4 Structure of the SPPCSPC module

1) 将主干网络所提取的特征图分别输入至两分支，其中，SPP 模块分支首先对输入特征图进行 1×1 以及 3×3 的卷积处理提取特征，然后输入至 SPP 模块进行池化操作处理，最后再通过 3×3 的卷积提取特征；卷积分支只对输入特征图进行 1×1 的卷积处理。

2) 将两分支的输出特征图进行 Concat 操作后再经过 1×1 的卷积模块处理调整输出通道，输出包含更多特征信息的不同尺度特征图。

在上述过程中， 1×1 的卷积模块主要目的是为了调整特征图的通道数， 3×3 的卷积模块用于减少因池化处理造成的特征信息丢失，丰富特征图语义信息。卷积分

支主要作用是避免提取过多重复梯度信息同时减少网络传播过程中的梯度信息损失。

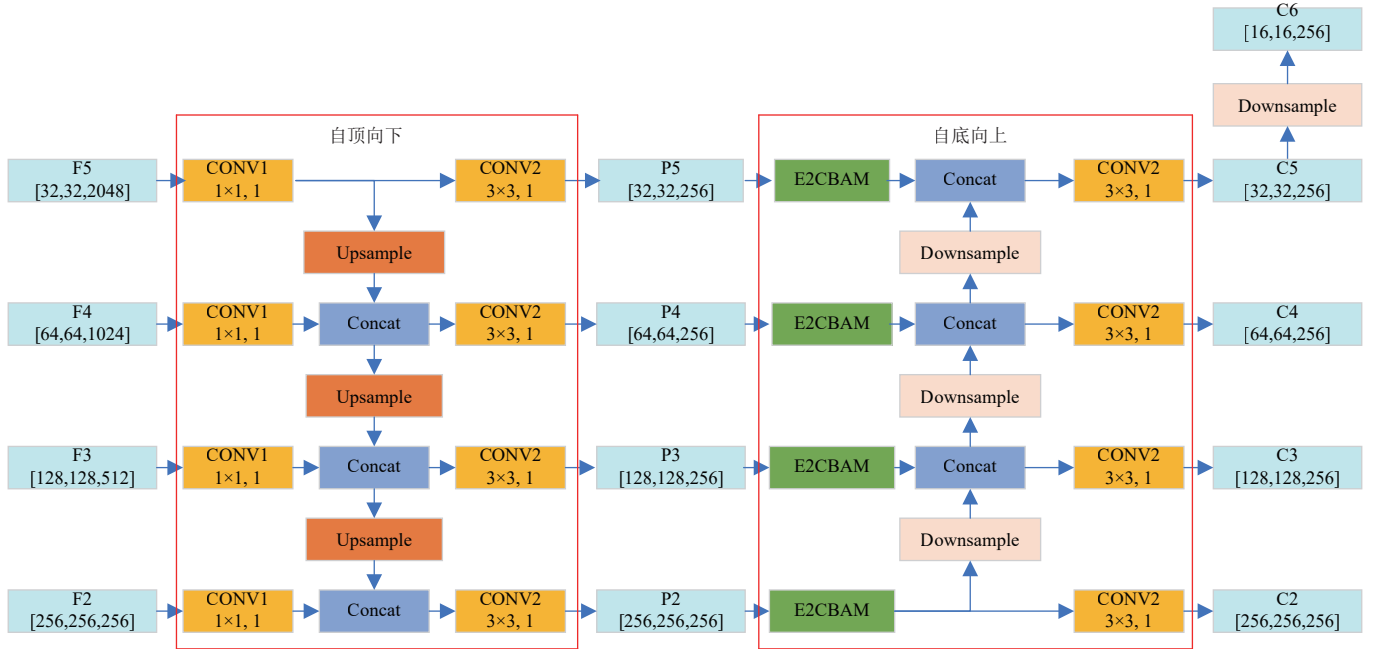
SPP 模块由 4 个最大池化层分支组成，其池化核大小分别为 5、9、13 和 1，4 种池化核大小不同的最大池化分支能扩大感受野，使其能够处理不同尺度的感受野对象，能够更好的区分样本中大目标与小目标。

1.2.2 颈网络中引入路径聚合网络及混合注意力模块

传统特征金字塔网络 FPN 建立了自顶向下的特征融合通道，高层特征图所包含的语义信息和目标位置信息得以传递给低层特征图，使得模型可以学习到更多的目标特征从而提高检测精度。

为了应对田野环境中小麦头部分布不均匀且重叠导致检测精度不理想的情况，引入了路径聚合网络结构 (PANet)，该网络在 FPN 的基础之上搭建了自底向上的信息传递通道，有效解决了 FPN 网络单向信息传递导致低层特征图所包含的信息无法传递给高层特征图的难题，进一步丰富了特征图所包含的语义信息和目标位置信息，能够更好地满足实际麦穗检测任务对精度的需求。

为了充分利用主干网络提取的目标特征信息，在特征图融合阶段进一步增强上下文信息的交互，从而提高模型准确率。针对田野环境中麦穗目标小且分布密集的特点，在 PANet 网络结构特征图融合连接中添加 E2CBAM 混合注意力模块，使得整体网络模型获得更加充分的目标信息，避免小目标漏检，改进后的颈网络结构如图 5 所示。



注：CONV1 模块表示卷积核大小等于 1×1 ，步长为 1 的卷积操作；CONV2 模块表示卷积核大小等于 3×3 ，步长为 1 的卷积操作；Upsample 为上采样模块；Downsample 为下采样模块；E2CBAM 是引入的混合注意力机制；Concat 是拼接操作。
Note: CONV1 module means a convolution operation with kernel_size=1 and stride is 1; CONV2 module means a convolution operation with kernel_size=3 and stride is 1; Upsample is an upsampling module; Downsample is a downsampling module; E2CBAM is the hybrid attention mechanism introduced; Concat is a concatenation operation.

图 5 改进后颈网络结构

Fig.5 Structure of the improved neck network

CBAM 是 WOO 等^[25]提出的一种混合注意力机制模块，从通道和空间两个作用域入手，对输入特征图先

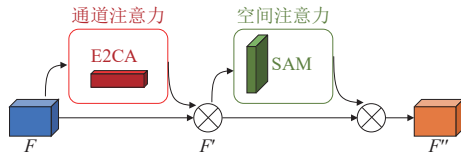
进行通道模块 (CAM) 处理后再通过空间注意力模块 (SAM)，实现端到端的混合注意力机制，其中 CAM

模块注重对特征图通道的分配关系；SAM 模块使得模型更关注对目标类别判断的关键像素区域，减少对非重点像素区域的关注。

如图 6 所示，为了减少添加 CBAM 模块导致的增加模型参数量，在轻量化通道注意力模块（ECA）^[26] 基础上，使用双分支轻量化通道注意力模块（E2CA）代替 CAM 提高模型对特征图通道的关注，最后与 SAM 组成轻量化混合注意力机制模块（E2CBAM），其模块结构表示为

$$\begin{aligned} F' &= M_c(F) \otimes F, \\ F'' &= M_s(F') \otimes F', \end{aligned} \quad (1)$$

式中 F 表示原特征图； F' 表示经过 E2CA 机制模块处理后的特征图； F'' 表示 F' 经过 SAM 模块处理后的特征图； $M_c(F)$ 为通道注意力图； $M_s(F)$ 为空间注意力图。



注：F 表示原特征图； F' 表示经过 E2CA 机制模块处理后的特征图； F'' 表示经过 SAM 模块处理后的特征图。

Note: F represents the original feature map; F' represents the feature map after being processed by the E2CA mechanism module; F'' represents the feature map after being processed by the SAM module.

图 6 混合注意力机制 E2CBAM 结构

Fig.6 Structure of hybrid attention mechanism E2CBAM

ECA 使用 1×1 的卷积核代替全连接层，使得模型参数变小更加轻量。如图 7 所示，E2CA 是由两个并行的 ECA 分支结构组成，首先将输入的特征图进行平均池化（最大池化），特征图从 $[H, W, C]$ 的矩阵转变为 $[1, 1, C]$ 的特征向量；接着对特征图通道数进行计算得到自适应的一维卷积核的大小；然后利用计算得到的自适应卷积核进行一维卷积运算得到特征图每个通道对应的权重；最后使用归一化的权重值与原特征图在通道维度相乘得到通道加权后的特征图，实现流程表示为

$$M_c(F) = \sigma(F_{conv1D}(\text{AvgPool}(x) + \text{MaxPool}(x))) \quad (2)$$

式中 σ 表示 Sigmoid 激活函数； F_{conv1D} 表示自适应一维卷积变换；AvgPool 和 MaxPool 分别表示平均池化和最大池化。自适应卷积核的大小计算式如下：

$$k = \varphi(c) = \left\lfloor \frac{\log_2^c}{\gamma} + \frac{b}{\gamma} \right\rfloor \text{odd} \quad (3)$$

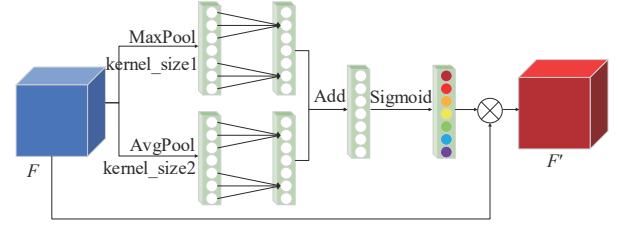
式中 k 表示卷积核大小， c 表示特征图通道数， $\lfloor \cdot \rfloor \text{odd}$ 表示取表达式最近的奇数值。将 γ 和 b 分别设置为 2 和 1 用于改变特征图通道数 c 和自适应卷积核大小之间的比例关系。

如图 8 所示，SAM 首先将输入的特征图分别经过最大池化和平均池化层在通道维度生成两个 $[1, H, W]$ 的权重向量；接着对两个特征向量进行 Concat 拼接得到 $[2, H, W]$ 的特征空间权重；然后经过卷积核大小为 7×7 的卷积层将特征图维度变为 $[1, H, W]$ ，得到通道维度

的空间权重；最后利用 Sigmoid 对空间权重归一化后与原图相乘，对空间位置特征进行加权得到新的特征图 $[C, H, W]$ ，以上操作表示为

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (4)$$

式中 $f^{7 \times 7}$ 表示卷积核大小为 7×7 的卷积运算。

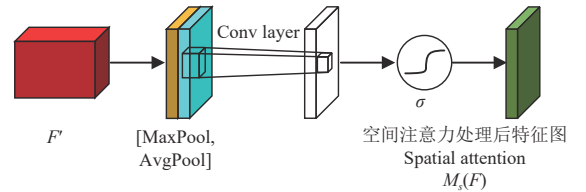


注：kernel_size1、kernel_size2 表示自适应卷积核；Sigmoid 是激活函数；Add 表示拼接操作。

Note: kernel_size1, kernel_size2 represent adaptive convolution kernels; Sigmoid is the activation function; Add represents the concat operation.

图 7 E2CA 模块结构

Fig.7 Structure of the E2CA module



注：Conv layer 表示 7×7 的卷积层； σ 表示 Sigmoid 激活函数。

Note: Conv layer represents a 7×7 convolutional layer; σ represents a Sigmoid activation function.

图 8 SAM 模块结构

Fig.8 Structure of the SAM (Spatial attention module)

1.2.3 柔性非极大值抑制算法

在田野环境中，麦穗分布密集且互相遮挡导致漏检情况比较多。传统的非极大值抑制算法 NMS 将预测框之间交并比大于所设阈值的其他预测框全部过滤，例如当两个麦穗之间重叠程度较高时，非极大值抑制算法 NMS 则会去除其中一个预测框从而导致另一个目标被漏检，NMS 计算式如下：

$$s_i = \begin{cases} s_i, & \text{iou}(M, b_i) < N_t \\ 0, & \text{iou}(M, b_i) \geq N_t \end{cases} \quad (5)$$

式中 s_i 表示第 i 个预测框的得分， M 为得分最高的预测框， b_i 表示第 i 个预测框， N_t 是所设预测框交并比阈值，iou 表示 M 与 b_i 的交并比值。

柔性非极大值抑制算法（Soft-NMS, soft-non maximum suppression）采用更加柔和的边界框筛选算法，在对目标预测框重叠程度较高且超过所设阈值时，通过对目标预测框使用高斯加权修改其置信度值来保留该预测框而不是直接将其过滤，有效减少了目标重叠导致边界框被过滤导致的漏检问题，进而提高模型针对密集小目标的检测能力，Soft-NMS 计算式如下：

$$s_i = \begin{cases} s_i, & \text{iou}(M, b_i) < N_t \\ s_i(1 - \text{iou}(M, b_i)), & \text{iou}(M, b_i) \geq N_t \end{cases} \quad (6)$$

在 Soft-NMS 算法中，置信度高斯加权计算式如下：

$$s_i = s_i e^{-\frac{\text{iou}(M_i, D)^2}{\delta}}, \forall b_i \notin D$$

(7)

式中 δ 为高斯权重。Soft-NMS 对超过阈值的重叠预测框采用“惩罚”方式对其重新赋值置信度值而不是将其过滤。

1.3 模型的训练与试验

1.3.1 评价指标

本试验使用平均精度（average precision, AP）、平均精度均值（mean average precision, mAP）、模型参数量（parameters）以及计算量对模型性能进行评价，其中计算值需要先计算模型训练的准确率（precision, P ）及召回率（recall, R ）。

1.3.2 试验环境及训练参数

试验输入图片尺寸为 1 024×1 024 像素，模型训练超参数具体设置如下：优化器选择使用随机梯度下降法 SGD 并采用 Momentum 优化算法，其中动量衰减项设置为 0.9，权重衰减因子默认为 0.000 1，初始学习率设置为 0.005，24 轮次之后降为 0.000 5，32 轮次之后降为 0.000 05，38 轮次后降为 0.000 001，共训练 40 轮次。

试验的硬件设备主要配置为 i7-12700F 和显存 16GB 的 NVIDIA GeForce RTX 4080 GPU。所使用的操作系统为 Windows 11 且装有 Python3.7 和 Pytorch1.12.1 的环境。

2 结果与分析

2.1 颈网络改进模块试验

本试验以 Oriented R-CNN 模型为基础，在不改变模型其他参数及结构的情况下，训练模型至收敛，模型性能的试验结果见表 2。

表 2 嵌入不同模块后模型检测精度对比

Table 2 Comparison of model detection accuracy with different modules

PANet	CBAM	E2CBAM	平均精度均值 Mean average precision mAP/%	参数量 Parameters P_s/M	浮点计算量 Floating-point operations FLOPs/G
×	×	×	85.77	41.13	198.54
√	×	×	85.96	44.47	223.92
√	√	×	86.12	44.71	223.97
√	×	√	86.27	44.67	223.97

注：“×”表示未采用该模块；“√”表示采用该模块，下同。
Note: "x" indicates that the module has is not used; "√" indicates that the module is used, the same as below.

在颈网络中引入路径聚合网络 PANet，mAP 提高了 0.19 个百分点，在此基础之上分别嵌入了 CBAM 以及 E2CBAM，mAP 分别提高了 0.16 和 0.31 个百分点，参数量分别上涨了 0.24 和 0.20 M，模型浮点数计算量基本不发生变化。相较于 CBAM，E2CBAM 在计算量基本不变的前提下参数量不仅减少了 0.04 M，并且 mAP 提高了 0.15 个百分点。

为了更直观地验证在颈网络中嵌入 E2CBAM 的优越性，在引入 PANet 的基础之上，利用热力图比较颈网络嵌入 CBAM 以及 E2CBAM 对麦穗目标特征的捕获能力，将颈网络输出特征图的 c2 层作为样本进行对比，试验结果如图 9 所示。

由图 9 可知，加入 PANet 的颈网络在嵌入不同注意力模块后，网络对密集麦穗目标的特征信息捕获能力均有明显的提升，尤其在区分麦穗目标与背景上有了显著的提升。对比 CBAM 和 E2CBAM，两种混合注意力机制能够有效将网络感兴趣目标范围控制在麦穗目标区域，但 CBAM 对于背景区域关注较少但对于目标区域的关注程度同样不高，而 E2CBAM 减少对背景区域的关注程度的同时，对目标区域的定位更加准确，证明了 E2CBAM 的优越性，所以研究采用 E2CBAM 模块增强模型对麦穗目标位置区域的关注程度。

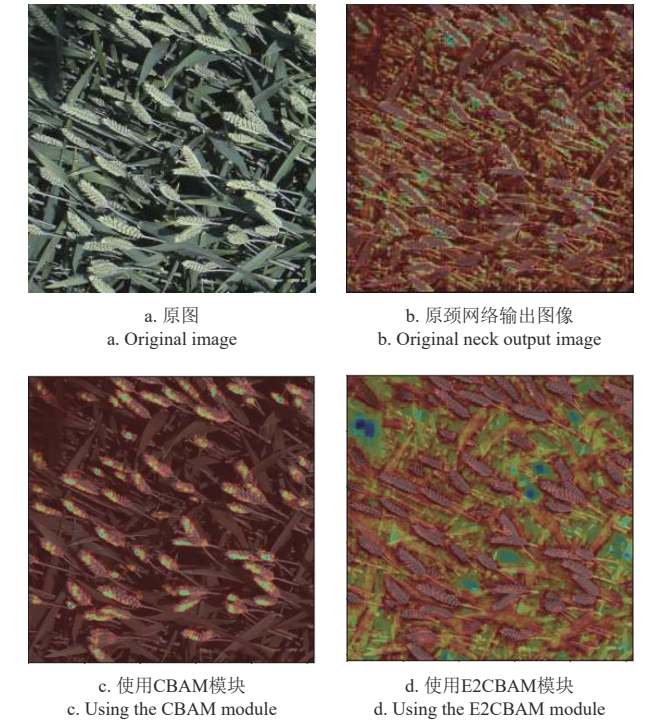


图 9 颈网络中嵌入不同注意力机制可视化结果
Fig.9 Visualization results of different attention mechanisms embedded in the neck network

2.2 改进模型消融试验

在 Oriented R-CNN 基础上，在主干网络 Backbone 中引入 SPPCSPC 模块增大感受野，增强模型对麦穗小目标的检测能力；在颈网络中添加 PANet 融合不同尺度的特征图，并嵌入 E2CBAM 模块使得模型更关注特征图中小目标位置，对麦穗目标定位更加准确；在筛选预测框时，采用 Soft-NMS 避免重叠的麦穗目标漏检。为验证 3 种改进策略对模型检测性能的作用，进行了消融试验，试验结果如表 3 所示。

表 3 改进模型消融试验结果

Table 3 Ablation test results of the improved model

ResNet+ SPPCSPC	PANet+ E2CBAM	Soft-NMS	精确率		mAP/%
			Precision P /%	Recall R /%	
×	×	×	65.69	90.60	85.77
√	×	×	79.06	90.81	86.21
×	√	×	82.09	90.29	86.27
×	×	√	64.60	92.09	86.87
√	√	×	81.79	90.80	86.45
√	×	√	72.70	92.14	87.40
×	√	√	64.97	92.07	87.59
√	√	√	72.50	92.54	87.79

由表 3 数据可知, 采用原 Oriented R-CNN 模型, 该模型在测试集上 mAP 为 85.77%; 在基准模型上在主干网络中使用 SPPCSPC 生成输出特征层 P5, 扩大了感受野使模型能够处理不同尺度的麦穗目标, mAP 提高了 0.44 个百分点; 在基准模型的基础上在颈网络中引入 PANet, 通过融合不同尺度的特征图丰富其所包含的特征信息, 并在特征融合分支中加入 E2CBAM 模块, 使得模型更关注麦穗目标位置区域, 该改进策略使得模型 mAP 提高了 0.50 个百分点; 在基准模型的基础上采用柔性非极大值抑制算法 Soft-NMS, 采用更友好的预测框筛选算法, 有效避免了重叠麦穗漏检, mAP 提高了 1.10 个百分点; 分别将 3 种改进模块进行两两组合验证其组合有效性, 相较于基准模型分别提高了 0.68、1.63 以及 1.82 个百分点; 在基准模型的基础上嵌入 3 种改进模块, mAP 最高, 较原模型提高了 2.02 个百分点。通过消融试验证明 3 种改进模块均能够有效提高模型检测精度, 并在共同使用时能够使得模型 mAP 达到最高值 87.79%。

2.3 Oriented R-CNN 改进前后试验结果

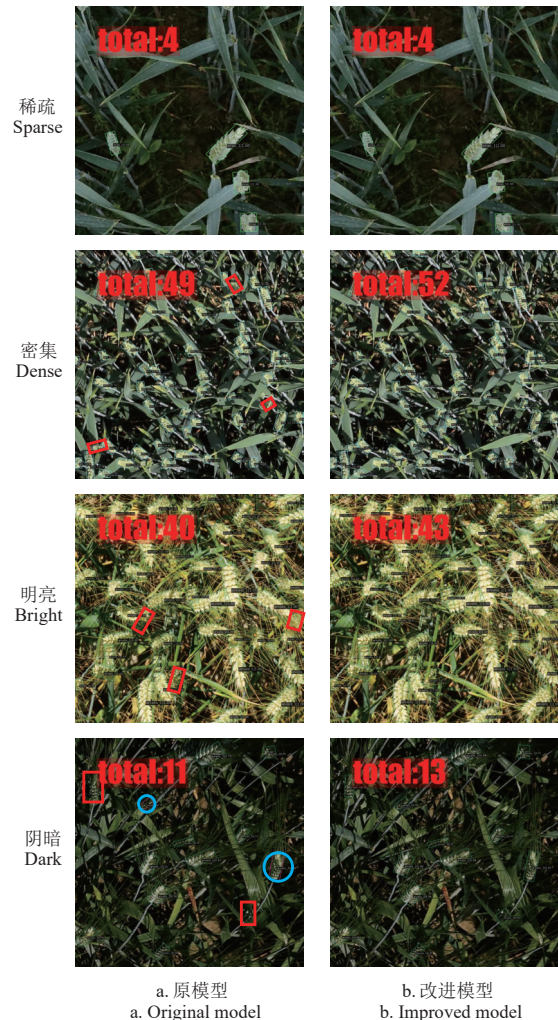
为了验证改进后模型在麦穗检测与计数任务中的表现, 在测试集中 7 类麦穗样本 (arvalis_1、arvalis_2、arvalis_3、ethz_1、inrae_1、rres_1、usask_1) 中分别取 20 张图像, 共计 140 张图像, 这些样本在光照强度以及密集程度上存在差异以便更全面地评估改进后模型性能。原模型与改进后模型对比试验结果如图 10 所示, 麦穗分布较为稀疏时, 改进前后模型对麦穗的检测效果并无明显差距, 两者都能较好的识别麦穗类别及位置; 麦穗分布较为密集或环境明亮时, 原模型在识别遮挡麦穗时出现漏检, 而改进后的 Oriented R-CNN 模型检测效果有了明显提高; 环境阴暗时, 原模型在识别过程中出现漏检和误检, 而改进后的 Oriented R-CNN 模型对麦穗的分类和定位能力均有所提高。

在计数任务中, 将预测框得分阈值设为 0.7, 成功地过滤了那些置信度较低的预测框。对比原 Oriented R-CNN 模型和改进后的模型麦穗预测结果, 表 4 清晰展示了改进后的模型在试验中取得的显著优势。改进后的 Oriented R-CNN 模型所有类别麦穗计数总和的正确率超过 90%。相较于原模型, 改进后的模型计数准确率提高 1.18 个百分点, 为实际应用提供了更可靠的结果。特别值得注意的是, 针对 arvalis_2 类别, 改进后的模型在计数正确率方面取得了显著的提升, 比原模型高出 5 个多百分点, 这突显了改进后模型在阴暗环境下检测性能的提高, 为在复杂条件下的农田应用提供了更为精确的计数结果。

2.4 不同检测模型的对比试验

将改进后的 Oriented R-CNN 模型与当前主流的 5 种旋转目标检测模型 Gliding vertex^[27]、R3 det^[28]、Rotated Faster R-CNN^[29]、S2anet^[30] 和 Rotated Retinanet^[31] 进行对比试验, 按照模型的训练策略分别训练 5 种模型, 对比识别效果如图 11 所示, 5 种对比模型在检测图片边界目标时均出现漏检, 因为边界附近麦穗目标通常是不完整, 模型获取到的麦穗特征信息过少; 此外, R3 det、

Rotated Faster R-CNN 以及 Rotated Retinaent 对重叠目标的检测效果也不理想, 而改进后的 Oriented R-CNN 模型对于图像边界框附近的麦穗目标以及目标分布密集区域的遮挡麦穗均有着更强的识别能力, 说明其对麦穗目标的特征提取能力更强, 在实际的田野环境中具有更优异的检测精度。



注: 红色方框标记区域为漏检麦穗; 蓝色圆圈标记为该麦穗的种类被误判; 图中红色数字表示图片中麦穗总数; 下同。
Note: The red box marks the area where the wheat ear was missed; the blue circle marks the category of the ear that was misclassified; the number in red indicates the total number of ears in the picture; the same as below.

图 10 改进前后模型检测效果对比
Fig.10 Comparison of model detection effect before and after improvement

表 4 改进前后模型麦穗计数结果试验
Table 4 Tests of modeled wheat count results before and after improvement

麦穗种类 Wheat variety	人工计数 Manual counting	原模型 Original model		改进后模型 Improved model	
		计数 Counting	准确率 Accuracy/%	计数 Counting	准确率 Accuracy/%
arvalis_1	824	768	93.20	775	94.04
arvalis_2	389	303	77.89	325	83.55
arvalis_3	506	446	88.14	462	91.30
ethz_1	1 469	1 297	88.29	1 314	89.45
inrae_1	371	348	93.80	345	92.99
rres_1	901	827	91.79	836	92.79
usask_1	614	549	89.41	541	88.11
总计 Total	5 074	4 538	89.44	4 598	90.62

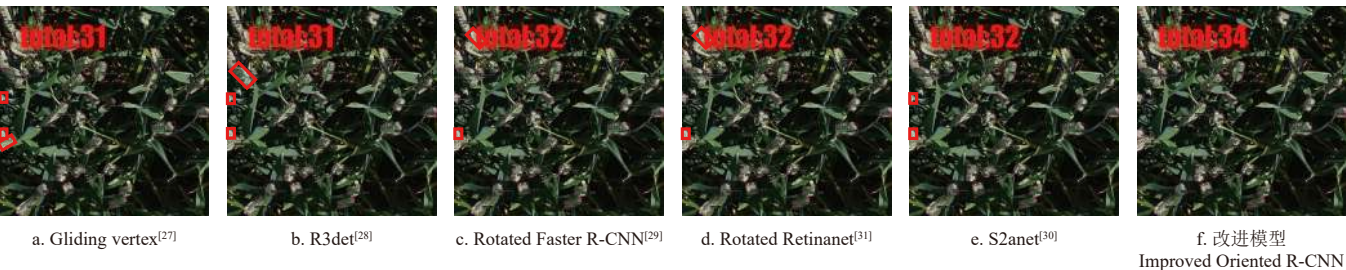


图 11 不同模型检测效果对比
Fig.11 Comparison of detection of different models

表 5 为不同旋转目标检测模型性能指标对比，由表 5 可知，Gliding vertex、R3 det、Rotated Faster R-CNN、S2anet 和 Rotated Retinanet 的 mAP 分别为 82.80%、85.30%、83.85%、85.54% 和 83.67%，而改进的 Oriented R-CNN 模型 mAP 为 87.79%，较其他 5 种对比模型分别提高了 4.99、2.49、3.94、2.25 和 4.12 个百分点。

表 5 不同模型试验结果对比							
Table 5 Comparison of test results of different models							
模型 Model	P/%	R/%	mAP/%	P _a /M	FLOPs/G	每秒帧数 Frames per second, FPS/(帧·s ⁻¹)	
Gliding vertex	77.10	88.27	82.80	41.13	198.41	27.2	
R3 det	21.66	90.46	85.30	41.72	331.72	16.6	
Rotated Faster R-CNN	76.30	89.67	83.85	41.13	198.40	27.2	
S2anet	34.54	90.91	85.54	38.56	196.81	22.7	
Rotated Retinanet	27.64	89.96	83.67	41.13	198.54	20.1	
Improved Oriented R-CNN	72.50	92.54	87.79	47.23	314.55	23.2	

对比试验说明改进后模型在光照强度存在差异的田野环境中对分布不均匀且遮挡的麦穗目标有更好的识别能力，能够更好地满足实际麦穗检测与计数任务对模型精度的需求。

2.5 田野麦穗实例检测

将 2020 年 4 月份在南京白马镇所拍摄的麦穗图像输入改进模型，观察强光照、阴暗环境、麦穗分布密集和重叠的麦穗图像检测结果，如图 12 所示。检测结果为：强光照麦穗图像中 35 个麦穗检测到 32 个，误检 1 个；阴暗麦穗图像中 30 个麦穗检测到 28 个；密集麦穗图像中 36 个麦穗检测到 35 个；重叠麦穗图像中 37 个麦穗检测到 35 个；138 个麦穗总计正确检测数量为 130，计数准确率为 94.20%。可知，改进后的 Oriented R-CNN 模型能够满足实际麦穗检测与计数任务需求，在阴暗、强光照、密集目标和遮挡等场景中表现良好。

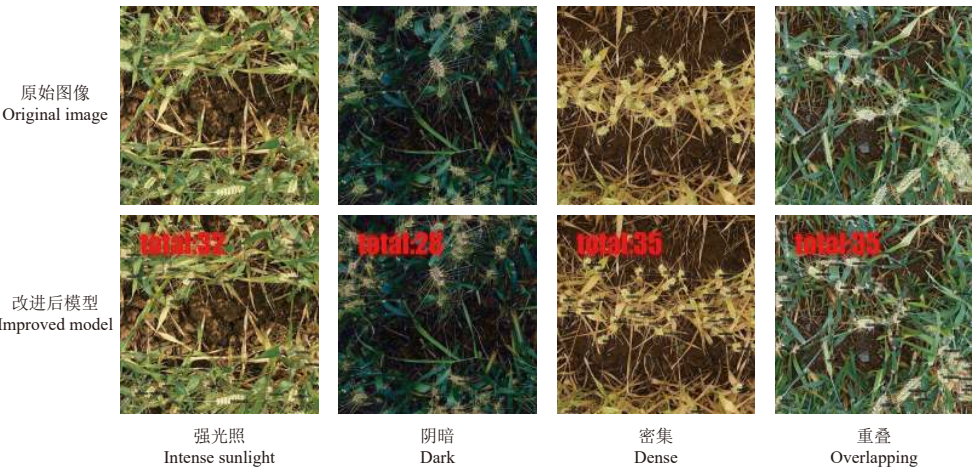


图 12 田野麦穗实例检测
Fig.12 Field wheat instance detection

3 结 论

该研究改进 Oriented R-CNN 模型，满足田野复杂环境下的麦穗目标检测与计数，主要结论如下：

- 1) 在主干网络 ResNet 中引入 SPPCSPC 金字塔模块生成主干网络输出层 F5，通过最大池化操作，能够使模型处理不同尺度的感受野对象，对于麦穗目标有更强的检测能力，模型的 mAP 提高了 0.44 个百分点。
- 2) 在颈网络 FPN 中加入路径聚合网络 PANet 并在特征融合分支中添加 E2CBAM 注意力模块，路径聚合网

络通过添加自底向上的特征融合通道进一步丰富了特征图包含的特征信息，E2CBAM 模块使模型注重对特征图通道分配关系，同时减少了对非重点的背景区域的关注，颈网络整体改进使得模型 mAP 提高 0.50 个百分点。

- 3) 使用 Soft-NMS 算法优化模型预测框筛选过程，避免漏检。本文改进的 Oriented R-CNN 最终平均精度均值 mAP 为 87.79%，与主流的旋转目标检测模型 Gliding vertex、R3 det、Rotated Faster R-CNN、S2anet 和 Rotated Retinanet 相比，mAP 分别提高了 4.99、2.49、3.94、2.25

和 4.12 个百分点, 能够较好满足实际麦穗检测与计数任务对检测模型精度的需求。

【参 考 文 献】

- [1] 鲍烈, 王曼韬, 刘江川, 等. 基于卷积神经网络的小麦产量预估方法[J]. *浙江农业学报*, 2020, 32(12): 2244-2252. BAO Lie, WANG Mantao, LIU Jiangchuan, et al. Estimation method of wheat yield based on convolution neural network[J]. *Acta Agriculturae Zhejiangensis*, 2020, 32(12): 2244-2252. (in Chinese with English abstract)
- [2] 杜世伟, 李毅念, 姚敏, 等. 基于小麦穗部小穗图像分割的籽粒计数方法[J]. *南京农业大学学报*, 2018, 41(4): 742-751. DU Shiwei, LI Yinian, YAO Min, et al. Counting method of grain number based on wheatear spikelet image segmentation[J]. *Journal of Nanjing Agricultural University*, 2018, 41(4): 742-751. (in Chinese with English abstract)
- [3] 谢元澄, 何超, 于增源, 等. 复杂大田场景中麦穗检测级联网络优化方法[J]. *农业机械学报*, 2020, 51(12): 212-219. XIE Yuancheng, HE Chao, YU Zengyuan, et al. Optimization method of wheat ear detection cascade network in complex field scenarios[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(12): 212-219. (in Chinese with English abstract)
- [4] 刘东, 曹光乔, 李亦白, 等. 基于颜色特征的小麦抽穗扬花期麦穗识别计数[J]. *中国农机化学报*, 2021, 42(11): 97-102. LIU Dong, CAO Guangqiao, LI Yibai, et al. Recognition and counting of wheat ears at flowering stage of heading poplar based on color features[J]. *Journal of Chinese Agricultural Mechanization*, 2021, 42(11): 97-102. (in Chinese with English abstract)
- [5] FERNANDEZ-GALLEGO J A, KEFAUVER S C, GUTIERREZ N A, et al. Wheat ear counting in-field conditions: High throughput and low-cost approach using RGB images[J]. *Plant Methods*, 2018, 14: 1-12.
- [6] 刘哲, 袁冬根, 王恩. 基于改进 Bayes 抠图算法的麦穗小穗自动计数方法[J]. *中国农业科技导报*, 2020, 22(8): 75-82. LIU Zhe, YUAN Donggen, WANG En. Automatic counting method of wheat grain based on improved bayes matting algorithm[J]. *Journal of Agricultural Science and Technology*, 2020, 22(8): 75-82. (in Chinese with English abstract)
- [7] 杜颖, 蔡义承, 谭昌伟, 等. 基于超像素分割的田间小麦穗数统计方法[J]. *中国农业科学*, 2019, 52(1): 21-33. DU Ying, CAI Yicheng, TAN Changwei, et al. Field wheat ears counting based on superpixel segmentation method[J]. *Scientia Agricultura Sinica*, 2019, 52(1): 21-33. (in Chinese with English abstract)
- [8] 刘航, 刘涛, 李世娟, 等. 基于深度残差网络的麦穗回归计数方法[J]. *中国农业大学学报*, 2021, 26(6): 170-179. LIU Hang, LIU Tao, LI Shijuan, et al. Research on wheat ear regression counting based on deep residual network[J]. *Journal of China Agricultural University*, 2021, 26(6): 170-179. (in Chinese with English abstract)
- [9] 章权兵, 胡姗姗, 舒文灿, 等. 基于注意力机制金字塔网络的麦穗检测方法[J]. *农业机械学报*, 2021, 52(11): 253-262. ZHANG Quanbing, HU Shanshan, SHU Wenchuan, et al. Wheat spikes detection method based on pyramidal network of attention mechanism[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(11): 253-262. (in Chinese with English abstract)
- [10] 鲍文霞, 张鑫, 胡根生, 等. 基于深度卷积神经网络的田间麦穗密度估计及计数[J]. *农业工程学报*, 2020, 36(21): 186-194. BAO Wenxia, ZHANG Xin, HU Gensheng, et al. Estimation and counting of wheat ears density in field based on deep convolutional neural network[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2020, 36(21): 186-194. (in Chinese with English abstract)
- [11] 杨蜀秦, 王帅, 王鹏飞, 等. 改进 YOLOX 检测单位面积麦穗[J]. *农业工程学报*, 2022, 38(15): 143-149. YANG Shuqin, WANG Shuai, WANG Pengfei, et al. Detecting wheat ears per unit area using an improved YOLOX[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(15): 143-149. (in Chinese with English abstract)
- [12] KHAKI S, SAFAEI N, PHAM H, et al. WheatNet: A lightweight convolutional neural network for high-throughput image-based wheat head detection and counting[J]. *Neurocomputing*, 2022, 489: 78-89.
- [13] 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 Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(16): 235-242. (in Chinese with English abstract)
- [14] 王旭, 吴艳霞, 张雪, 等. 计算机视觉下的旋转目标检测研究综述[J]. *计算机科学*, 2023, 50(8): 79-92. WANG Xu, WU Yanxia, ZHANG Xue, et al. Survey of rotating object detection research in computer vision[J]. *Computer Science*, 2023, 50(8): 79-92. (in Chinese with English abstract)
- [15] 李朝辉, 安金堂, 贾红雨, 等. 结合旋转框和注意力机制的轻量遥感图像检测模型[J]. *中国图象图形学报*, 2023, 28(9): 2706-2718. LI Zhaohui, AN Jintang, JIA Hongyu et al. Lightweight object detection model in remote sensing image by combining rotation box and attention mechanism[J]. *Journal of Image and Graphics*, 2023, 28(9): 2706-2718. (in Chinese with English abstract)
- [16] 庄文华, 唐晓刚, 张斌权, 等. 基于改进 YOLOv5 的遥感图像旋转框目标检测[J]. *电子设计工程*, 2023, 31(14): 137-141, 146. ZHONG Wenhua, TANG Xiaogang, ZHANG Binqian et al. Remote sensing image rotatable bounding box object detection based on improved YOLOv5[J]. *Electronic Design Engineering*, 2023, 31(14): 137-141, 146. (in Chinese with English abstract)
- [17] SUN J, YANG K, CHEN C, et al. Wheat head counting in the wild by an augmented feature pyramid networks-based convolutional neural network[J]. *Computers and Electronics in Agriculture*, 2022, 193: 106705.
- [18] XIE X, CHENG G, WANG J, et al. Oriented R-CNN for object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, 2021: 3520-3529.
- [19] 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[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Vancouver, BC, Canada: IEEE, 2023: 7464-7475.
- [20] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Salt Lake City, UT, USA: IEEE, 2018: 8759-8768.
- [21] 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 (ICCV). Venice, Italy: IEEE, 2017: 5561-5569.
- [22] DAVID E, SEROUART M, SMITH D, et al. Global wheat head detection 2021: An improved dataset for benchmarking wheat head detection methods[J]. *Plant Phenomics*, 2021, 2021:

- 277-285.
- [23] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [24] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) workshops. Seattle, WA, USA: IEEE, 2020: 390-391.
- [25] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). Springer, Cham, 2018: 3-19.
- [26] 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 (CVPR). Seattle, WA, USA: IEEE, 2020: 11534-11542.
- [27] XU Y, FU M, WANG Q, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. IEEE transactions on pattern analysis and machine intelligence, 2020, 43(4): 1452-1459.
- [28] YANG X, YAN J, FENG Z, et al. R3det: Refined single-stage detector with feature refinement for rotating object[C]//Proceedings of the AAAI Conference on Artificial Intelligence. AAAI Press, Palo Alto, California USA, 2021, 35(4): 3163-3171.
- [29] REN S Q, HE K M, GIRSHICK R, et al. Faster r-cnn: Towards real-time Object Detection with Region Proposal Networks[J]. Advances in Neural Information Processing Systems, 2017, 30(6): 1137-1149.
- [30] HAN J, DING J, LI J, et al. Align deep features for oriented object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-11.
- [31] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 2980-2988.

Improved Oriented R-CNN-based model for oriented wheat ears detection and counting

YU Junwei^{1,2,3}, CHEN Weiwei^{1,2,4}, GUO Yuansen^{1,2,4}, MU Yashuang^{1,2,3}, FAN Chao^{1,2,3}

(1. Key Laboratory of Grain Information Processing and Control, Ministry of Education, Henan University of Technology, Zhengzhou 450001, China; 2. Henan Key Laboratory of Grain Photoelectric Detection and Control, Henan University of Technology, Zhengzhou 450001, China; 3. College of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou 450001, China; 4. College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China)

Abstract: An accurate detection can greatly contribute to the wheat ears in field environments. Traditional object detection models with horizontal bounding boxes cannot accurately detect the densely distributed wheat ears, particularly on the significant occlusion between ears and stalks. The high miss detection of wheat ears often occurs in the variation of illumination conditions, dense distribution, and small scales, due to the overlap of prediction bounding boxes. It is a high demand to orient the wheat ears with less noise and of large background for the high performance. In this study, an improved Oriented Region-based Convolution Neural Networks (R-CNN) model was proposed to detect and count rotated wheat ears. Firstly, the spatial pyramid pooling cross-stage partial networks (SPPCSPC) was added to the backbone network to generate the last layer of the output feature map. The sensing field was then enlarged to enhance the perceptual ability of the network; Secondly, the feature aggregation network and the efficient two convolutional block attention module (E2CBAM) hybrid attention mechanism module were introduced into the neck network to enrich the feature information in the feature map; Finally, the prediction bounding boxes were optimized using the flexible non-maximal inhibition algorithm soft-non maximum suppression (Soft-NMS), in order to optimize the predicted bounding boxes screening. The E2CBAM module was improved using the convolutional block attention module (CBAM) in the E2CA module, instead of the CAM channel attention module. The E2CA module was composed of two parallel ECA branch structures: the maximum and average pooling. Two adaptive convolution kernels were then obtained to sum. Finally, the channel assignment was weighted for the important channel information. The key feature was captured to improve the detection performance of the model. To verify the E2CBAM hybrid attention module, the path aggregation network (PANet) was introduced into the neck network to enrich the semantic and target location in the feature map. The detection accuracy of the model was then improved by 0.19 percentage points. Furthermore, the detection accuracy was improved by 0.16 and 0.31 percentage points, whereas, the number of parameters increased by 0.24 and 0.20 M. respectively, in the CBAM and E2CBAM hybrid attention mechanism module. The floating-point computation remained unchanged. Compared with the CBAM, the E2CBAM hybrid attention mechanism module improved the detection accuracy of the model by 0.15 percentage points, while reducing the number of parameters by 0.04 M with the unchanged computation. The experimental results show that the improved Oriented R-CNN model accurately represented the head direction of wheat ears, indicating better detection performance. The mean mAP of average accuracy was 2.02 percentage points higher than the original model, compared with the mainstream-oriented bounding boxes detection models. Moreover, the mAP values were improved by 4.99, 2.49, 3.94, 2.25, and 4.12 percentage points, respectively, compared with the mainstream rotating target detection models, Gliding vertex, R3det, Rotated Faster R-CNN, S2anet, and Rotated Retinanet. The Oriented R-CNN was utilized to accurately represent the head direction of wheat ears. The background area was also reduced in the prediction bounding boxes. The model detection was more visually appealing. The finding can provide an effective way for the practical observation of the growth status of wheat ears and counting the number of ears.

Keywords: image recognition; crops; attention mechanism; wheat ear; Oriented R-CNN