

基于融合 GhostNetV2 的 YOLO v7 水稻籽粒检测

刘庆华^{1,2} 杨欣仪² 接浩² 孙世诚² 梁振伟³

(1. 江苏科技大学自动化学院, 镇江 212100; 2. 苏州科技大学电子与信息工程学院, 苏州 215004;
3. 江苏大学农业农村部智能农机装备重点实验室, 镇江 212013)

摘要: 水稻籽粒检测在粮食储存中凸显重要作用, 直接影响粮食销售的价格。针对一般机器视觉检测算法在水稻籽粒小目标的密集场景下存在难以识别且网络模型参数大, 检测速度较慢、成本高等问题, 提出一种基于 YOLO v7 优化的水稻籽粒检测算法。首先将部分高效聚合网络模块 (Efficient layer aggregation network, ELAN) 替换成轻量级网络模块 GhostNetV2 添加到主干及颈部网络部分, 实现网络参数精简化的同时也减少了通道中的特征冗余; 其次将卷积和自注意力结合的注意力模块 (Convolution and self-attention mixed model, ACmix) 添加到 MP 模块中, 平衡全局和局部的特征信息, 充分关注特征映射的细节信息; 最后使用 WIoU (Wise intersection over union) 作为损失函数, 减少了距离、纵横比之类的惩罚项干扰, 单调聚焦机制的设计提高了模型的定位性能。在水稻籽粒图像数据集上验证改进后的模型检测水平, 实验结果表明, 改进后的 YOLO v7 模型的 mAP@0.5 达 96.55%, mAP@0.5:0.95 达 70.10%, 训练模型参数量也有所下降, 在实际场景以暗黑色为背景的水稻杂质检测中的效果优于其他模型, 满足了水稻籽粒的实时检测要求, 可将此算法应用于自动化检测粮食系统中。

关键词: 水稻籽粒检测; YOLO v7; 轻量级网络; 注意力模块

中图分类号: TP391.41 文献标识码: A 文章编号: 1000-1298(2023)12-0253-08 OSID: 

Rice Grain Detection Based on YOLO v7 Fusing of GhostNetV2

LIU Qinghua^{1,2} YANG Xinyi² JIE Hao² SUN Shicheng² LIANG Zhenwei³

(1. College of Automation, Jiangsu University of Science and Technology, Zhenjiang 212100, China
2. College of Electronics and Information Engineering, Suzhou University of Science and Technology, Suzhou 215004, China
3. Key Laboratory of Intelligent Agricultural Machinery Equipment, Ministry of Agriculture and Rural Affairs, Jiangsu University, Zhenjiang 212013, China)

Abstract: Rice grain detection plays an important role in grain storage, directly affecting the price of grain sales. In response to the problems of difficult recognition, large network model parameters, slow detection speed, and high cost of general machine vision detection algorithms in dense scenes with small rice grain targets, a rice grain detection algorithm was proposed based on YOLO v7 optimization. Firstly, some efficient layer aggregation network (ELAN) modules were replaced with lightweight network module GhostNetV2 and added them to the backbone and neck network sections, achieving precise simplification of network parameters while reducing feature redundancy in channels. Secondly, the attention module (ACmix) that combined convolution and self attention was added to the MP module, balancing global and local feature information and fully paying attention to the details of feature mapping. Finally, wise intersection over union (WIoU) was used as the loss function to reduce penalty term interference such as distance and aspect ratio. The design of monotonic focusing mechanism improved the positioning performance of the model. The improved model detection level was verified on the rice grain image dataset, and the experimental results showed that the improved YOLO v7 model was high, mAP@0.5 was up to 96.55%, mAP@0.5:0.95 reached 70.10%, and the training model parameters were also decreased. In practical scenarios, the effect of rice impurity detection with a dark black background was

收稿日期: 2023-05-26 修回日期: 2023-08-21

基金项目: 国家自然科学基金项目 (52275251)

作者简介: 刘庆华 (1977—), 男, 教授, 博士, 主要从事智能交通、人工智能和计算机视觉研究, E-mail: liuqh@just.edu.cn

通信作者: 梁振伟 (1988—), 男, 教授, 博士, 主要从事农业机器人和联合收获机智能控制系统等研究, E-mail: zhenwei_liang@ujs.edu.cn

better than other models, meeting the real-time detection requirements of rice grains. This algorithm can be considered for application in automated grain detection systems.

Key words: rice grain detection; YOLO v7; lightweight network; attention module

0 引言

农业图像检测 (Agricultural image detection, AID) 已成为现代农业检测方法中非常重要的步骤^[1]。水稻是最主要的粮食作物之一, 随着农业机械化的应用与普及, 水稻的含杂质率直接影响到粮食销售价格, 较高的谷物破碎率会使谷物容易发霉并且难以储存。因此, 水稻籽粒中的病害杂质检测是保证粮食合格的重要举措。

针对水稻籽粒病害检测问题, 国内外开展了广泛的研究, 通过采用机器视觉检测的各类方法实现了水稻籽粒病害检测。马德贵等^[2]在 21 世纪初利用椭圆形拟合水稻病斑, 并且分析其主轴长和人工检测病斑的主轴长的关系, 识别准确率为 89%。TAHIR 等^[3]等利用一种基于长度、颜色和形状等特征的图像处理方法对水稻进行质量分级, 通过分析几种利用图像处理进行稻米品质分级的分类器, 再通过应用一些计算算法和图像滤波器, 展示了机器视觉对于水稻质量分类的有效性。

但是以上基于形态处理的机器视觉检测方法存在一定的弊端, 当水稻更换病害品种或者背景环境复杂时, 识别能力和效率会大幅度降低, 无法按照要求完成水稻籽粒采集任务。随着深度学习技术越来越广泛地应用于现代农业领域, 目前已经取得很多成果。如姚青等^[4]利用改进后的 RetinaNet 对水稻冠层植株上的两种害虫类别进行检测, 其平均精度达 93.76%。OLSEN 等^[5]使用 Inception - v3 和 ResNet - 50 CNN 模型在数据集上展示了强大的基线性能, 实验证明 ResNet - 50 模型也在目标机器人水稻杂草控制应用的实时要求下进行了很好的推理, 其准确率达到 95.1%。陈进等^[6]利用 U - Net 语义分割算法, 对水稻籽粒及背景等 4 种像素类别进行图像分割, 识别正确率在 88.56% 以上。卷积神经网络 (CNN) 作为一种具有代表性的深度学习算法, 已广泛应用于工业图像的分类^[7]、定位^[8]、检测^[9]和分割^[10]。因此, 本研究采用当前模型精度与性能优势较为均衡的 YOLO v7 网络模型对水稻籽粒进行检测。

1 YOLO v7 概述

YOLO 算法是单阶段 (one-stage) 目标检测算法中的领先算法, 具有快速、高精度的特点。2015 年

YOLO v1^[11]被提出, 单级检测算法首次出现在人们的视野中。有效地解决了两级检测网络推理速度慢的缺点, 对检测精度保持了良好的效果。YOLO v3^[12]是 YOLO v1 的改进版本。其最大的特点是引入了剩余模块 Darknet - 53 和特征金字塔网络 (FPN) 架构, 对 3 种不同尺度的对象进行预测, 实现了多尺度融合。此后, YOLO v4^[13]和 YOLO v5 在 YOLO v3 的基础上增加了一些创新点。2022 年 YOLO v7 提出了 Extended - ELAN 架构, 在不破坏原有梯度路径的前提下提高网络的自学习能力。此外, MP 模块是将图像特征层进行 MaxPool 操作, 其感受野得到扩张, 再经 1 × 1 卷积得到特征信息拼接, 提高了网络泛化能力。SPPCSPC 模块是在多个卷积之后进行 3 个并行的 MaxPool 操作, 减少图像失真或图像特征重复的问题。

但 YOLO v7 算法框架在实际检测水稻籽粒的过程中还存在某些局限可加以改进: ①虽然网络中的 ELAN 模块具有很强的特征聚合能力, 但过量的重参数会压制训练速度, 因此需要考虑引入轻量化网络改进算法。②MP 模块在池化操作的过程中, 由于水稻籽粒受到光线、色彩差异、位置遮挡等因素影响造成特征信息丢失的现象, 因此考虑加入注意力机制改善该模块。③由于水稻籽粒的尺寸不一, 同一幅图像中物体像素分布复杂多变, 导致位置和形状的偏差在检测中影响较大, 因此考虑替换默认损失函数。

本文提出一种基于 YOLO v7 优化的水稻籽粒检测方法, 利用机器视觉算法的先进性, 缩小网络模型尺寸从而提升检测速率。

2 YOLO v7 模型改进

2.1 主干及颈部网络优化

2.1.1 GhostNetV2 网络

目前引进轻量化网络是模型减少计算成本和参数的有效策略, 最具有代表性的轻量级网络有 MobileNet^[14]、ShuffleNet^[15]和 GhostNet^[16]。MobileNet 将标准卷积分解为深度卷积和点卷积, MobileNetV2^[17]和 MobileNetV3^[18]通过引入逆残差块的方法改善网络结构。ShuffleNet 采用通道重排的思想实现信息交换, ShuffleNetV2^[19]在此基础上减少网络分支达到泛化模型推理速率的效果。而在 ImageNet 数据集上的效果显示, GhostNetV2 在分类

精度、参数数量和检测速度方面都领先于其他网络^[20],因此本文将 GhostNetV2 引入 YOLO v7 的主干及颈部网络部分,从而替换 ELAN 模块并测试其效果。

GhostNetV1 对少量传统卷积生成的部分特征图作线性变换,既能增加特征图的冗余效果,也能继承传统卷积的优点。对于给定输入特征 $X \in \mathbf{R}^{H \times W \times C}$, H 、 W 、 C 分别为特征图的高度、宽度、通道数,首先使用 1×1 卷积生成固有特征

$$Y' = X * F_{1 \times 1} \tag{1}$$

式中 $*$ ——卷积运算符号 Y' ——固有特征

$F_{1 \times 1}$ ——逐点卷积像素

$Y' \in \mathbf{R}^{H \times W \times C'_{out}}$ 表示固有特征,一般小于原始输出特征 $Y \in \mathbf{R}^{H \times W \times C_{out}}$,即 $C'_{out} < C_{out}$,随后使用简单的深度卷积操作生成更多特征,再拼接这两部分的信道维度特征

$$Y = \text{Concat}([Y', Y' * F_{dp}]) \tag{2}$$

式中 F_{dp} ——深度卷积滤波器

Concat——拼接操作函数

但是只有一半的空间信息实现信息交互,而另外的特征信息经过 1×1 逐点卷积之后并无信息交互。因此 GhostNetV2 提出一种新型的解耦全连接注意力机制(DFC),对于一个给定特征 $Z \in \mathbf{R}^{H \times W \times C}$,特征图矩阵为 $A = [\alpha_{11} \quad \alpha_{12} \quad \cdots \quad \alpha_{HW}]$,完全连接层(FC)被分解为水平和垂直并各自沿其方向聚合卷积神经网络中特征图的像素。具体操作为

$$\alpha_{hw} = \sum_{h',w'} F_{hw,h',w'} \odot z_{h'w'} \tag{3}$$

$$\alpha'_{hw} = \sum_{h'=1}^H F_{h,h'w}^H \odot z_{h'w} \tag{4}$$

($h = 1, 2, \cdots, H; w = 1, 2, \cdots, W$)

$$\alpha_{hw} = \sum_{w'=1}^W F_{w,hw'}^W \odot \alpha'_{hw} \tag{5}$$

($h = 1, 2, \cdots, H; w = 1, 2, \cdots, W$)

式中 $\odot$ ——相乘运算符号

$F_{hw,h',w'}$ ——水平和垂直方向可学习权值

h ——垂直方向元素序号

w ——水平方向元素序号

$F_{h,h'w}^H$ ——垂直方向可学习权值

$F_{w,hw'}^W$ ——水平方向可学习权值

$z_{h'w}$ ——垂直方向特征

α'_{hw} ——经过垂直变换之后的输出特征

α_{hw} ——经过垂直和水平变换之后的输出特征

$z_{h'w'}$ ——水平和垂直方向特征

经水平和垂直变换的解耦操作后,注意力模块的计算复杂度降低。GhostNetV2 依旧采用 Inverted

bottleneck 的设计,利用 2 个 Ghost 模块使特征维度先升维再降维,DFC 注意力机制与第 1 个 Ghost 模块同步进行以增强捕捉长距离空间信息的可靠性,图 1 为本文设计的 C3GhostV2 结构示意图。

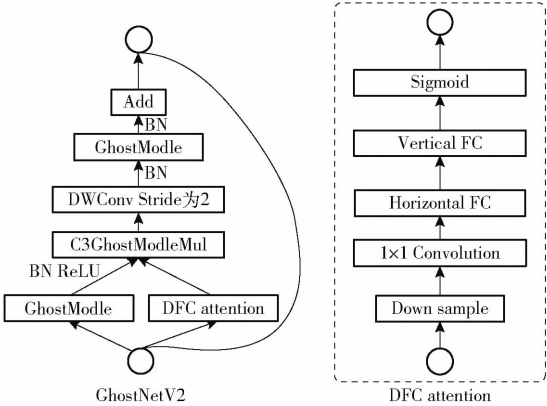


图 1 C3GhostV2 结构示意图

Fig. 1 Structure diagram of C3GhostV2

为了更加融合主干及颈部网络提取到的不同尺度的图像特征,减少传递层之间的信息损失问题,本文对 GhostNetV2 进行优化,设计出新的模块 GhostModule、C3 和 C3GhostModuleMul,如图 2 所示。C3 模块通过集联廉价卷积和拼接通道数相同的特征减少信息损失,C3GhostModuleMul 将池化后的特征信息先升维后降维之后再和其余廉价卷积融合,充分泛化网络性能。

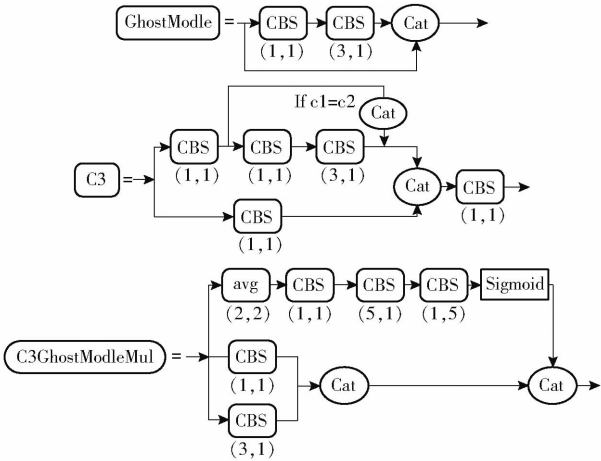


图 2 GhostModule、C3 和 C3GhostModuleMul 模块结构

Fig. 2 GhostModule, C3, and C3GhostModuleMul module structure

2.1.2 ACmix 注意力机制

在添加 GhostNetV2 网络之后,本文为进一步改善头部和颈部模型,在 MP 模块中加入注意力机制。注意力机制的作用是让网络聚焦重要特征信息而又忽略非重要特征信息,同时还能与其他模块实现信息交换与传递。近些年注意力机制一直受到广大学者的关注,HU 等^[21]利用一种新型的网络获取到特征图每个通道的重要信息并赋予权值,然后再让

网络着重关注这些重点特征通道,称为 Squeeze and Excitation block (SE)。但是通道数的反复变化在引起信息特征丢失的同时也增加了参数量和计算量,因此 WOO 等^[22]在此基础上提出一种结合空间和通道的卷积注意力模块(Convolutional block attention module, CBAM)。对于给定特征图, CBMA 会沿着通道和空间 2 个维度推断重要特征图,再和输入特征图相乘实现特征优化,但由于

CBMA 仅考虑了局部信息的结构,且引入大尺度的卷积核提取空间信息使模型结构复杂,HOU 等^[23]提出了一种轻量级模块 Coordinate Attention(CA),不仅考虑了通道和空间的相关性,还关注了位置信息,达到参数少、计算量小的特点。PAN 等^[24]提出了一种将卷积和自注意力结合的模块 ACmix,效果一样领先于 SE 和 CBMA,图 3 为 ACmix 模块整体处理过程。

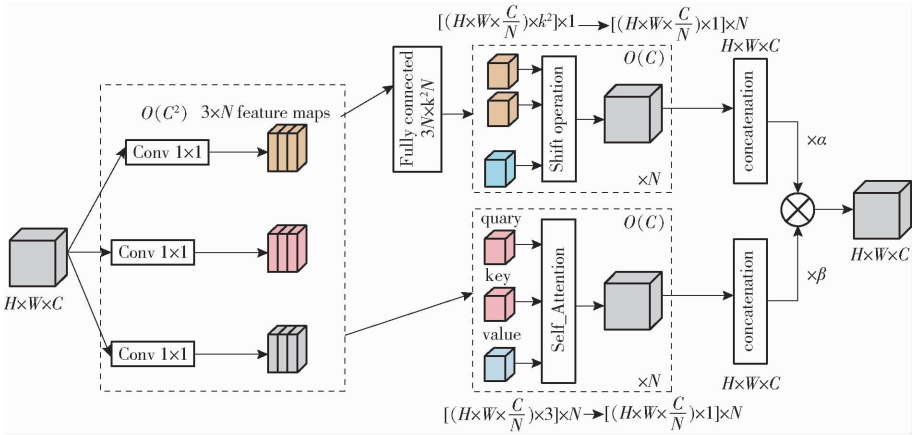


图 3 ACmix 注意力机制结构图

Fig.3 Structure diagram of ACmix Attention mechanism

首先将大小为 $H \times W \times C$ 的特征通过 3 个 1×1 卷积投影后获得 $3 \times N$ 个特征图,然后分别进行卷积和自注意力操作。对于卷积操作可以分为 2 个阶段:

阶段 1
$$g_{ij}^{-(p,q)} = K_{p,q} f_{ij} \tag{6}$$

阶段 2
$$g_{ij} = \sum_{p,q} g_{ij}^{(p,q)} \tag{7}$$

$g_{ij}^{(p,q)} = \text{Shift}(g_{ij}^{-(p,q)}, p - \lfloor k/2 \rfloor, q - \lfloor k/2 \rfloor)$
式中 f_{ij} ——输入像素的特征向量
 $\lfloor \rfloor$ ——移位操作
 k ——内核大小
 $g_{ij}^{(p,q)}$ ——投影前输出的特征图
 $g_{ij}^{-(p,q)}$ ——投影后输出的特征图
Shift——移动转换 p, q ——线性投影
 $K_{p,q}$ ——在某个位置上的线性投影
 g_{ij} ——聚合操作后得到的特征总和

第 1 阶段是让输入特征按照核权重 K 在各个坐标位置 (p, q) 做投影操作。第 2 阶段让投影后的映射根据 $K_{p,q}$ 分别进行水平和垂直方向的移位运算,最后将所有映射特征信息聚合在一起。

同理,自注意力操作也可分为 2 个阶段:

阶段 1
$$\begin{cases} q_{ij}^{(l)} = W_q^{(l)} f_{ij} \\ k_{ij}^{(l)} = W_k^{(l)} f_{ij} \\ v_{ij}^{(l)} = W_v^{(l)} f_{ij} \end{cases} \tag{8}$$

阶段 2

$$g_{ij} = \parallel_{l=1}^N \left(\sum_{a,b \in N_k(i,j)} A(q_{ij}^{(l)}, k_{ab}^{(l)}) v_{ab}^{(l)} \right) \tag{9}$$

式中 $W_q^{(l)}$ ——query 在像素 (i, j) 上的输入特征映射矩阵
 $W_k^{(l)}$ ——key 在像素 (i, j) 上的输入特征映射矩阵
 $W_v^{(l)}$ ——value 在像素 (i, j) 上的输入特征映射矩阵
 $k_{ij}^{(l)}$ ——key 投影后的特征映射
 $v_{ij}^{(l)}$ ——value 投影后的特征映射
 $q_{ij}^{(l)}$ ——query 投影后的特征映射
 $\parallel$ —— N 个注意力头输出的级联
 $N_k(i, j)$ ——以像素 (i, j) 为中心,空间宽度为 k 的区域

$A(q_{ij}^{(l)}, k_{ab}^{(l)})$ ——区域 $N_k(i, j)$ 内对应的权重
将第 1 阶段通过 3 个 1×1 卷积投影后的特征映射记为 query、key 和 value,最后将两条操作的路径合并相加输出为

$$F_{\text{out}} = \alpha F_{\text{conv}} + \beta F_{\text{att}} \tag{10}$$

式中 F_{out} ——路径最终输出
 F_{conv} ——卷积注意力分支输出
 F_{att} ——自注意力分支输出

其中 α 和 β 分别是卷积和自注意力学习的参数,用于衡量输出比重。为了使卷积和自注意力操作中输

出的全局和局部特征信息达到平衡,本文将 α 和 β 的值都设置为 1,提高 2 种模式对中间映射信息的聚合能力,使网络更加适用于检测小目标物体。

2.2 损失函数优化

YOLO v7 的坐标损失函数是用 CIoU 进行计算,计算公式为

$$L_{CIoU} = L_{IoU} - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v \quad (11)$$

其中

$$\alpha = \frac{v}{1 - L_{IoU} + v}$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2$$

式中 L_{CIoU} ——边界框回归损失函数

b ——预测框的中心点

b^{gt} ——真实框的中心点

ρ ——计算两个中心点之间的欧氏距离

w, h ——边界框的宽度、高度

α ——适用于做权衡的参数

v ——衡量长宽比一致性的参数

c ——能够同时包含预测框和真实框的最小闭包区域的对角线距离

w^{gt}, h^{gt} ——真实框的宽度和高度

L_{IoU} ——普通质量锚框的 IoU 损失函数

式 (11) 虽然性能高,但有一个不可避免的弊端:衡量纵横比值的参数 v 在真实框和预测框的比值一样大时取值为 0,会造成损失函数的惩罚项不稳定。针对此问题,ElIoU^[25]的做法是将惩罚项的影响因子拆开,分别计算真实框和预测框的长宽比。SIoU^[26]的做法是将 3 个方面(角度、距离、形状)的成本加入计算,边界框采用中心对齐方式,具有更快的收敛速度。但是训练数据中除了距离、纵横比值之外的其他度量值依然会干预惩罚项的泛化能力,为了削弱真实框与预测框重合时的干扰度量值,本文采用 R_{WIoU} 替换原来的 CIoU 作为新的损失函数,首先根据距离度量构造距离注意力,得到具有两层注意力机制的 L_{WIoUv1} 为

$$L_{WIoUv1} = R_{WIoU} L_{IoU} \quad (12)$$

$$\text{其中 } R_{WIoU} = \exp \left(\frac{(x - x^{gt})^2 + (y - y^{gt})^2}{(W_g^2 + H_g^2)^*} \right) \quad (13)$$

式中 H_g ——最小包围框的高

W_g ——最小包围框的宽

$*$ ——分离操作

x, y ——边界框中心点坐标

x^{gt}, y^{gt} ——真实框中心点坐标

其中 $R_{WIoU} \in [1, e)$, 明显提升了普通预测框 L_{IoU} 的性能质量。 $L_{IoU} \in [0, 1]$, 降低了 R_{WIoU} 干扰项的性

能质量,同时也在真实框和预测框重合度较高的情况下显著降低其对中心点距离的关注。为了防止 R_{WIoU} 产生阻碍收敛的梯度,将 W_g, H_g 从计算中分离出来,因此不必引入新的度量指标。

Focal Loss 设计了一种针对交叉熵的单调聚焦机制,有效降低了普通度量值对损失值的干扰。这使得模型能够聚焦于重要度量值,获得分类性能的提升。 $WIoUv2$ 构造了单调聚焦系数 L_{IoU}^* 的表达式为

$$L_{WIoUv2} = L_{IoU}^* L_{WIoUv1} \quad (\gamma > 0) \quad (14)$$

式中 γ ——学习参数

在模型训练过程中,梯度增益 L_{IoU}^* 随着 L_{IoU} 的减小而减小,导致训练后期收敛速度较慢。因此,引入 L_{IoU} 的均值作为归一化因子

$$L_{WIoUv2} = \left(\frac{L_{IoU}^*}{\bar{L}_{IoU}} \right)^\gamma L_{WIoUv1} \quad (15)$$

式中 $\bar{L}_{IoU}$ ——动量为 m 的滑动平均值

动态更新为归一化因子使梯度增益 $\left(\frac{L_{IoU}^*}{\bar{L}_{IoU}} \right)^\gamma$ 整体保持在较高水平,解决了训练后期收敛速度慢的问题。

此外,为了能让边界框回归聚焦在普通质量的预测框,为其分配一个小的梯度增益,对离群度较大的预测框分配较小的梯度增益,将有效防止低质量因素产生较大的有害梯度,因此构造了一个非单调聚焦系数并将其应用于 $WIoUv1$,即

$$L_{WIoUv3} = r L_{WIoUv1} \quad (16)$$

其中

$$r = \frac{\theta}{\delta \alpha^{\theta - \delta}} \quad (17)$$

$$\theta = \frac{L_{IoU}^*}{L_{IoU}} \quad (\theta \in [0, +\infty)) \quad (18)$$

式中 θ ——非单调聚焦系数

δ ——梯度增益

r ——锚框半径

当 $\theta = \alpha$ 时, $r = 1$, 当预测框的离散群程度满足 $\theta = C$ (C 为定值) 时,预测框将获得最高的梯度增益。由于 $\bar{L}_{IoU}$ 是动态的,预测框的质量划分标准也是动态的,这使得 $WIoUv3$ 在任何时刻都能做出最符合当前情况的梯度增益分配策略。

为了防止低质量锚框在训练初期落后,初始化 $\bar{L}_{IoU} = L_{IoU} = 1$, 让预测框具有更高的增益,此外为了在训练的早期阶段保持这样的策略,需要设置一个小的动量 m 来延迟 $\bar{L}_{IoU}$ 接近真实值的时间, m 设置为 $1 - \sqrt[n]{0.05}$, t 为训练次数, n 为迭代周期。这种设置使得经过 t 轮训练后有 $\bar{L}_{IoU} \approx \overline{L_{IoU} - real}$, 其中 $real$ 表示真实的交并比损失函数值,同时会聚焦于普通

质量的预测框,提高模型的定位性能。

2.3 改进后的网络模型

经过以上分析,将 YOLO v7 模型加以改进:首先在主干和颈部引入 GhostNetV2 轻量级网络,替换部分 ELAN 模块;其次在 MP 模块中添加 ACmix 注

意力机制,让检测目标的全局和局部特征信息达到平衡,充分提取浅层特征细节,增强小目标检测准确率的同时也减小了误判率;最后优化损失函数,将 CIoU 替换为 WIoU。改进后的 YOLO v7 总体结构如图 4 所示。

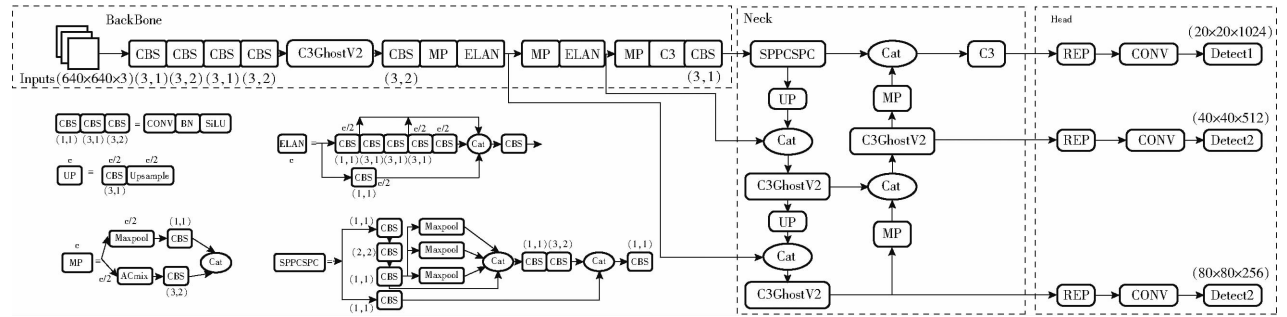


图 4 改进后的 YOLO v7 网络结构图
Fig. 4 Improved YOLO v7 network structure diagram

3 实验结果与分析

3.1 实验平台和参数设置

实验环境是 Windows 11 操作系统,在训练模型之前,初始化权重,同时根据训练结果不断地调整、优化参数,防止网络产生过拟合现象。GPU 为 RTX 3080Ti (10 GB),CPU 为 Intel (R) Platinum 8255C (40 GB),处理器为 CUDA 11.3,Python 版本为 3.8,网络学习率为 0.01,优化器为 Adam,图像训练批量大小为 16。

3.2 实验数据集

水稻籽粒样本采集于江苏省镇江市上党镇农业试验田,由专业高清摄像头拍摄,分辨率为 600 像素 × 800 像素,拍摄图像数量达到均衡。采集包含茎秆、残枝、坏粒和开口图像共计 260 幅,由于训练需要大量的图像数据,因此对数据集剪裁后进行数据增强操作(镜像翻转、缩放大小、对比度增强、高斯模糊、颜色变换等),最后使用 1 560 幅图像训练,训练集、测试集、验证集按照 8:1:1 的比例划分。

3.3 评价指标

为了更好地衡量密集场景下水稻籽粒检测的精度和实时性,本文选取了多方面指标对该目标检测模型进行评价分析。主要以精确率(Precision, P)、召回率(Recall, R)、平均精度均值(Mean average precision, mAP)、浮点运算量(FLOPs)作为评价指标。

精确率表示模型检测到正样本中实际为正样本的概率。召回率指在实际为正的样本中被预测为正样本的概率。平均准确度均值指对所有籽粒杂质类别的平均精度取平均值,一般情况下都使用 mAP 对网络模型的性能做综合性评价。mAP@0.5 定义为

4 种籽粒杂质类别的平均精度求均值。mAP@0.5:0.95 表示在不同 IOU 阈值(从 0.5 ~ 0.95,步长 0.05)上的平均 mAP,该值衡量边界回归能力和检测框与目标框的拟合精度。FLOPs 指模型复杂度,因此 FLOPs 越小越好。

3.4 消融实验

为了进一步验证改进后的 YOLO v7 在水稻籽粒算法中的优化效果,将改进后的 YOLO v7 算法逐步和初始算法做实验对比。消融实验包含 4 个步骤的改进对比:①模型对比,将 YOLO v7 与 YOLO v7-E6E 做对比,YOLO v7 的检测速率最快但精度最低,YOLO v7-E6E 的检测速率最慢但精度最高。②GhostNetV2 引入到主干和颈部网络中。③MP 模块中的 1 × 1 卷积替换为 ACmix 注意力机制模块,并且和其他注意力机制(SE, CBAM, CA)做对比。④把原损失函数 CIoU 替换为 WIoU,并且和 EIoU、SIoU 做对比。具体实验检测效果如表 1 所示。

由表 1 可知,改进的 YOLO v7 模型(编号为 10)在每个指标都有明显提升效果。相较于传统算法(编号为 1),准确率依然维持在 94% 以上,召回率提升 8.06 个百分点,mAP@0.5 提升 3.64 个百分点,mAP@0.5:0.95 提升 8.36 个百分点,网络参数也随之下降,使得检测速率也有了相应的提升。

针对消融实验可看出,编号 1、2 为模型对比实验,YOLO v7-E6E 虽然检测精度较高,但是模型复杂度高,检测速率远不如 YOLO v7。编号 3 为添加 GhostNetV2 轻量级网络之后的实验效果,mAP@0.5:0.95 提高 3.36 个百分点,而且大大降低了模型复杂度,提高了检测速度。编号 4 ~ 7 为不同注意力机

由表2可知,本文算法的 mAP@0.5 和 mAP@0.5:0.95 都高于其他算法,虽然同样作为 one-stage 的典型算法,SSD 效果较好,但本文算法更加适合于小目标密集型场景检测。

4 结论

(1)为检测水稻籽粒杂质,提出了一种基于 YOLO v7 优化的水稻籽粒检测方法,有效解决了水稻籽粒工作高度依赖筛查者目视导致耗时耗力的

问题。
(2)通过在主干和颈部网络中引入 GhostNetV2 轻量级网络,解决了重参数较高的缺陷,同时保证了检测精度的提升。
(3)在 MP 模块中添加 ACmix 注意力机制,兼顾全局和局部二者的图像特征信息;最后优化损失函数提高模型定位能力。
(4)3 次改进的准确率及召回率均达到 90% 以上,均达到检测指标要求。

参 考 文 献

[1] 薛金林,李雨晴,曹梓建.基于深度学习的模糊农田图像中障碍物检测技术[J].农业机械学报,2022,53(3):234-242.
XUE Jinlin, LI Yuqing, CAO Zijian. Obstacle detection technology in fuzzy farmland images based on deep learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(3): 234-242. (in Chinese)

[2] 马德贵,邵陆寿,葛婧,等.水稻稻瘟病及水稻纹枯病病害程度图像检测[J].中国农学通报,2008,24(9):485-489.
MA Degui, SHAO Lushou, GE Jing, et al. Image detection of rice blast and rice sheath blight disease severity[J]. Chinese Journal of Agriculture, 2008, 24(9): 485-489. (in Chinese)

[3] TAHIR W, HUSSIN N, HTIKE Z Z, et al. Rice grading using image processing[J]. ARPN Journal of Engineering and Applied Sciences, 2015, 10(21): 10131-10137.

[4] 姚青,谷嘉乐,吕军,等.改进 RetinaNet 的水稻冠层害虫为害状自动检测模型[J].农业工程学报,2020,36(15):182-188.
YAO Qing, GU Jiale, LÜ Jun, et al. Automatic detection model for damage of rice canopy pests improved by RetinaNet[J]. Transactions of the CSAE, 2020, 36(15): 182-188. (in Chinese)

[5] OLSEN A, KONOVALOV D A, PHILIPPA B, et al. DeepWeeds: a multiclass weed species image dataset for deep learning[J]. Scientific Reports, 2019, 9: 2058.

[6] 陈进,韩梦娜,练毅,等.基于 U-Net 模型的含杂水稻籽粒图像分割[J].农业工程学报,2020,36(10):174-180.
CHEN Jin, HAN Mengna, LIAN Yi, et al. Image segmentation of mixed rice grains based on U-Net model[J]. Transactions of the CSAE, 2020, 36(10): 174-180. (in Chinese)

[7] 杨断利,王永胜,陈辉,等.基于改进 YOLO v6-tiny 的蛋鸡啄羽异常现象识别及个体分类方法[J].农业机械学报,2023,54(5):268-277.
YANG Duanli, WANG Yongsheng, CHEN Hui, et al. An improved YOLO v6-tiny based method for identifying abnormal feather pecking phenomena and individual classification in laying hens[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(5): 268-277. (in Chinese)

[8] 付友,左迅,杨凡,等.基于深度学习卷积神经网络模型的行人检测设计[J].信息技术,2021(5):34-38.
FU You, ZUO Xun, YANG Fan, et al. Design of pedestrian detection based on deep learning convolutional neural network model[J]. Information Technology, 2021(5): 34-38. (in Chinese)

[9] 张涛源,谢新林,谢刚,等.融合 Transformer 的带钢缺陷实时检测算法[J].计算机工程与应用,2023,59(16):232-239.
ZHANG Taoyuan, XIE Xinlin, XIE Gang, et al. A real-time defect detection algorithm for strip steel based on Transformer fusion[J]. Computer Engineering and Applications, 2023, 59(16): 232-239. (in Chinese)

[10] 王银宇,孟凡云,王金鹤,等.改进 ASPP 及多层次特征语义融合分割方法[J].计算机工程与应用,2023,59(13):220-228.
WANG Yinyu, MENG Fanyun, WANG Jinhe, et al. Improved ASPP and multi-level feature semantic fusion segmentation method[J]. Computer Engineering and Applications, 2023, 59(13): 220-228. (in Chinese)

[11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.

[12] REDMON J, FARHADI A. YOLO v3: an incremental improvement[J]. arXiv:1804.02767.

[13] BOCHKOVSKIY A, WANG L. YOLO v4: optimal speed and accuracy of object detection[J]. arXiv:2004.10934.

[14] ANDREW G H, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.

[15] ZHANG Xiangyu, ZHOU Xinyu, LIN Mengxiao, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6848-6856.

[16] HAN Kai, WANG Yunhe, TIAN Qi, et al. GhostNet: more features from cheap operations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580-1589.

[17] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.