南京农业大学学报 2024,47(4):772-781
Journal of Nanjing Agricultural University
http://nauxb.njau.edu.cn
DOI:10.7685/jnau.202312020

唐伯青,赵大勇,熊锋,等. 基于视觉自注意力模型的苗期玉米与杂草检测方法[J]. 南京农业大学学报,2024,47(4):772-781.
Tang Boqing,Zhao Dayong,Xiong Feng,et al. Detection method of maize and weeds at seedling stage based on visual self-attention model[J]. Journal of Nanjing Agricultural University,2024,47(4):772-781.

基于视觉自注意力模型的苗期玉米与杂草检测方法

唐伯青[1,3],赵大勇[1,2*],熊锋[1,2],李德强[1,2]

(1.中国科学院沈阳自动化研究所,辽宁 沈阳 110016;2.中国科学院机器人与智能制造创新研究院,辽宁 沈阳 110169;
3.中国科学院大学计算机科学与技术学院,北京 100049)

摘要:[**目的**]识别作物和杂草是农业智能化中自动除草的关键步骤。本文旨在解决作物与杂草识别精度低、检测模型实时性和鲁棒性差等问题。[**方法**]以叶龄处于3~8叶期的玉米及其伴生杂草为研究对象,提出一种高效准确的玉米苗与杂草的检测方法。该方法以实时端到端目标检测视觉自注意力模型为基础框架,用小尺度卷积等效替代大尺度深度卷积的思想,以较小的精度损失降低推理耗时。引入一种包含上下文信息的自顶向下注意力机制,强化模型对小目标的检测效果。应用组合图像增强策略,提升模型精度与泛化能力。[**结果**]改进后模型的平均检测精度为90.11%,推理阶段单张图片耗时33.67 ms,模型参数量44.86 MB。改进后的模型比主流目标检测模型总体精度更高,且推理速度快。[**结论**]所提方法对于玉米苗与伴生杂草的整体检测性能优秀,能够提高杂草识别的准确性和效率。

关键词:玉米;杂草;检测;实时视觉自注意力模型;等效卷积;图像增强
中图分类号:S24;S513 **文献标志码**:A **文章编号**:1000-2030(2024)04-0772-10

Detection method of maize and weeds at seedling stage based on visual self-attention model

TANG Boqing[1,3],ZHAO Dayong[1,2*],XIONG Feng[1,2],LI Deqiang[1,2]

(1.Shenyang Institute of Automation,Chinese Academy of Sciences,Shenyang 110016,China;
2.Institutes for Robotics and Intelligent Manufacturing,Chinese Academy of Sciences,Shenyang 110169,China;
3.School of Computer Science and Technology,University of Chinese Academy of Sciences,Beijing 100049,China)

Abstract:[**Objectives**]Identifying crops and weeds are crucial aspects of advancing intelligent and automated weeding. This article aimed to improve the accuracy of crop and weed identification,to enhance the real-time performance of detection models,and to enhance robustness. [**Methods**]Focusing on maize crops and their corresponding weeds in the leaf age range of 3-8 leaves,this research endeavored to devise a detection method for maize seedlings and associated grasses. The seedling detection method leveraged an improved real-time end-to-end object detection with transformers(RT-DETR)for maize and weed detection in field conditions. The novel concept of replacing large-scale deep convolution with small-scale convolution equivalence within RT-DETR was introduced,reducing training complexity and inference time while maintaining detection accuracy. Furthermore,a self-attention mechanism with contextual information was integrated to enhance target attention and improve small target detection. Additionally,a combined image enhancement strategy was employed to enhance model accuracy and generalization. [**Results**]The improved model effectively distinguished weeds from crops in complex field scenarios,achieving an average detection accuracy of 90.11%. In the inference stage,each image took 33.67 ms for processing,with a model size of 44.86 MB. Compared with the mainstream target detection model,the improved model had higher overall accuracy and fast speed. [**Conclusions**]The proposed method had excellent overall detection performance for corn seedlings and associated weeds,which could improve the accuracy and efficiency of weed identification.

Keywords:corn;weed;detection;end-to-end object detection with transformers;convolutional equivalence;image augmentation

玉米是我国北方的重要作物。玉米生产容易受杂草侵害,玉米苗期(出苗后28 d内)遭受杂草侵害的影响最大,通常苗期杂草导致26.75%的玉米产量损失,严重的可达90%以上[1]。当前我国作物除草的主要方法是喷洒化学除草剂(约占40%)[2]。大量使用化学除草剂耗费人力和物力,还造成杂草抗药性增加和有害物残留[3]。准确识别作物和杂草是解决作物草害损失和减少农药危害的前提与核心,深度学习

收稿日期:2023-12-15
基金项目:中国科学院战略性先导专项(XDA28040400)
*通信作者:赵大勇,副研究员,主要从事人工智能算法及应用研究,E-mail:zhaodayong@sia.cn。

的识别方法准确性和泛化性能优秀,是当前农田复杂环境下热门的检测方法[4-5]。

主流深度学习方法可以分为卷积神经网络模型和视觉自注意力模型。卷积模型的 two-stage 类算法将检测分为锚框的定位与分类两个过程,如 Faster-RCNN[6]。樊湘鹏等[7]基于 Faster-RCNN 识别棉田幼苗与伴生杂草,验证集平均识别精确率 88.67%,耗时 0.385 s,但验证数据量较小,难以说明泛化能力的有效性。Cascade R-CNN[8]是级联多个不同交并比阈值检测框的模型,Song 等[9]基于 Cascade R-CNN 提出组合 RGB 与深度图像的多通道下幼芽定位方法,每张图片的平均精度和处理速度分别为 97.5%和 12.9 $f\cdot s^{-1}$(77 ms),然而该模型自身的复杂度较高,难以满足田间除草的轻量快捷要求。卷积模型的 One-stage 算法将检测转化为回归问题,如 YOLO[10]和 SSD[11]等。亢洁等[12]将多尺度融合思想引入 SSD 模型检测甜菜与杂草,平均检测精度与速度分别为 88.84%和 38.46 $f\cdot s^{-1}$;Gallo 等[13]用 YOLOv7 识别由无人机采集的大尺寸田地图像实现杂草检测,平均精度为 74.1%,但该方法将模型直接用于处理包含多个目标的大型航拍图像,识别精度较低。无锚框卷积算法无需设计目标检测框的参数和优化方法,训练难度更低。如 FCOS[14]模型是一种一阶的全卷积结构模型,Peng 等[15]进行了 FCOS 模型对稻田中杂草的检测试验,对包含 8 类杂草检测的平均精度与速度分别为 81.3%和 13.0 $f\cdot s^{-1}$,该方法对比了其他常见模型的识别精度,但没有不同杂草种类、分布和背景条件对识别效果的分析,且推理速度有限。

视觉自注意力模型 DETR[16]将目标检测视为包含物体种类与边界框的设定置信度任务,如 Abuhani 等[17]进行甜菜、油葵与杂草的目标检测试验,检测精度为 63.9%。基于 DETR 改进的实时端到端目标检测视觉自注意力模型 RT-DETR[18],采用高效编码器架构加快模型的推理速度,加入卷积结构的主干网提升精度,但是仍然存在小目标和密集目标检测困难的问题。

注意力机制是模仿人类视觉聚焦过程[19],能通过学习强化有效特征信息并抑制不相关特征调制模型的输出,其效果优秀且方便模块化被引入到各类深度学习任务中[20]。其中自顶向下注意力模块模仿动物的反馈式神经连接,在高层特征中得到目标的种类信息,在低层特征中得到目标局部信息,通过反馈通道产生具有更精确局部信息高层目标的种类信息,提升模型的小目标和密集目标检测效果[21]。

目前基于高性能试验平台的苗草检测方法逐渐完善,但应用于小目标、密集且分布极不平衡的杂草检测方法难以兼顾高精度与高实时性。因此,本文提出一种综合目标检测模型高性能与快速性能优化(用一种小尺度卷积替代大尺度卷积 ConvNeXt[22])的玉米与杂草检测方法,并用一种自顶向下注意力模块 TDAM[23]联合低级与高级特征,提升模型的小目标检测效果。

1　材料与方法

1.1　材料与仪器

玉米品种为‘九玉 103’。数据集为自行采集的玉米与杂草田间数据,采集自内蒙古自治区扎兰屯市大河湾农场,采集装置选用分辨率为 2 400×2 400 的 ISOCELL-5F1 相机,安装在施药机架上,调整相机安装高度,使得单张图像采集范围为边长 1 m 的正方形。数据全部为自然环境下的玉米苗期植株及其常见伴生杂草,数据采集时间为 2023 年 6 月 15—30 日,图片包括不同的采集时间、光照和天气条件、田地土壤背景。使用图像标注工具对数据集进行标注,将图像分辨率尺寸调整为 640×640,标注过程中忽略图像中露出 20%或以下的目标和周长小于 5 cm 的极小杂草。完成标注数据图像 923 张,总共 9 536 个玉米目标与 14 607 个杂草目标框,平均每张图片有 26 个目标框,框长宽比主要分布在 0.75~1.5,大部分目标框占全图片的面积小于 5%。将标注数据按照 8∶2 的比例划分训练集与验证集。

1.2　算法测试平台

本模型在 AI-Studio 云平台使用深度学习工具箱 PaddlePaddl 2.5.0 进行训练,cuDNN 版本为 8.4。硬件配置描述:CPU 为 Intel(R)Xeon(R)Gold 6148 CPU@2.40 GHz,使用核心为 2 核,内存 16 GB,硬盘大小 100 GB;GPU 型号为单卡 NVIDIA-Tesla-V100,可用显存为 16 GB。

1.3　评价标准

以平均检测精度(mAP)作为模型检测精度的评价指标。其中检测精度评价指标是交并比(IoU):即真实框与预测框的交并面积的比值,真实框与预测框的交并比大于等于该比值即为正确检测。本文中使用 mAP 表示交并比在 0.5~0.95 的平均精度。此外用单张图片检测的平均推理耗时评价模型的检测速度;使用模型大小(Params 参数量)评价模型的内存占用情况。

1.4 玉米苗与杂草识别模型设计

玉米苗杂草识别模型参照 RT-DETR 架构的颈部网和快速解码器，设计识别模型用更轻量的卷积和更少的计算达成高精度识别效果。

1.4.1 大尺度卷积的等效替换 大尺度卷积具有更大的“感受野”，可以带来模型精度的提升，但这种提升是以内存访问成本为代价。根据 Yu 等[24]的研究，仅将部分通道输入卷积层可达到与大尺度卷积同样效果。本文将 ConvNeXt 块中的大尺寸深度卷积层进行分离，用小尺度卷积结构等效替代大卷积。

图 1 是 ConvNeXt 块的基本结构，其中存在卷积核为 7×7 的大尺度深度卷积结构[25]。与更小尺寸卷积核的结构相比，大尺度结构可以改善模型精度，但大尺度结构会带来更多的内存访问与计算成本，卷积核为 7×7 的结构 FLOPS 比 3×3 的大 1.22 倍[26]，难以满足玉米与杂草检测的实时性要求。

假设 ConvNeXt 块中输入 $X \in R^{N\times C\times H\times W}$，其中：$N$ 表示输入数据组数；C 表示通道数；H 是输入数据高度；W 是宽度。忽略 H 与 W 变化仅计算通道数变化，则数据经过深度卷积操作后的输出：

$$X_1 = \mathrm{DwConv}_{7\times 7}^{C\to r_1 C}(X) \tag{1}$$

式中：X 表示输入；$\mathrm{DwConv}_{7\times 7}^{C\to r_1 C}(\cdot)$ 表示 7×7 的深度卷积；C 表示输入通道数；r_1 表示通道变化系数。

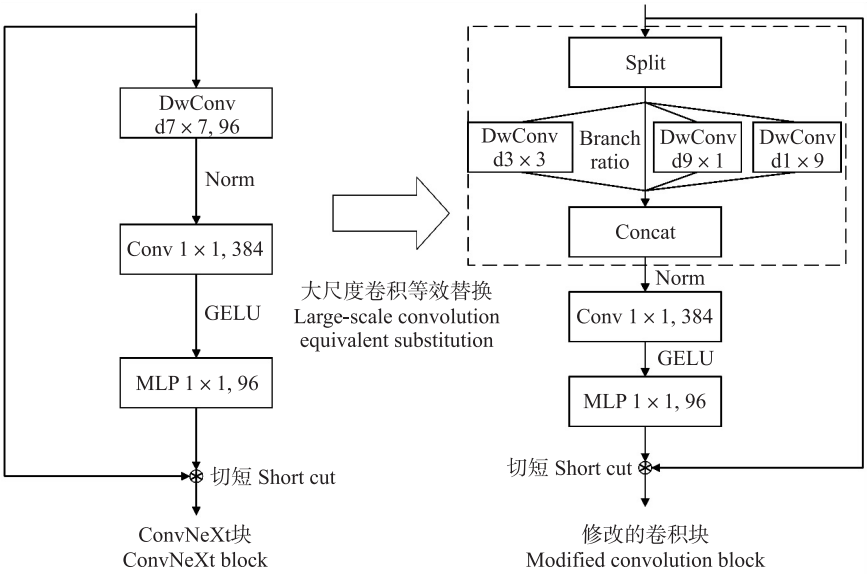

图 1 ConvNeXt 块与等效卷积结构

Fig. 1 Block structure of ConvNeXt and equivalent convolution

通道数为 C 的数据经过卷积后通道数变化为 $r_1 C$。随后经过批量归一化处理，分别经过 1×1 卷积层、GELU 激活函数和 1×1 卷积层（MLP）和残差连接，得到修改的卷积块输出：

$$Y = \mathrm{Conv}\{F[\mathrm{Conv}(X_2)]\} + X \tag{2}$$

式中：$\mathrm{Conv}(\cdot)$ 表示卷积运算；$F[\cdot]$ 表示 GELU 激活函数；X_2 表示公式 1 经过批量归一化处理的结果。

改进的等效卷积结构如图 1。若将输入深度卷积的数按照通道进行划分，设定分割因子 n，得到一组分割的通道，一部分划分后的通道直接通过该卷积层。剩余的通道均等分为 3 块，分别经过 $k_b\times k_b$、$1\times k$、$k\times 1$ 的深度卷积。通道划分：

$$(X_a, X_b, X_c, X_d) = \mathrm{Split}(X) \tag{3}$$

式中：$\mathrm{Split}(\cdot)$ 表示通道划分；X_a 表示按划分因子划分的通道数，会直接通过本层卷积；X_b、X_c、X_d 表示余下通道作均分结果，分别经过 $k_b\times k_b$、$1\times k$、$k\times 1$ 的深度卷积。

各条通道的输出：

$$X'_a = X_a \tag{4}$$

$$X'_b = \mathrm{DwConv}_{k_b\times k_b}^{C(n-1)/3n\to r_b C(n-1)/3n}(X_b) \tag{5}$$

$$X'_c, X'_d = \mathrm{DwConv}_{1\times k}^{C(n-1)/3n\to rC(n-1)/3n}(X_c, X_d) \tag{6}$$

式中：k_b 表示方型卷积层卷积核大小；k 表示长型卷积层卷积核大小；r_b 表示 k_b 卷积的通道变化系数；r 表示 k 卷积通道变化系数。

再将以上 4 个结果拼接，得到输出：

$$X_1' = \mathrm{Concat}(X_a, X_b, X_c, X_d) \tag{7}$$

式中：Concat(·)表示拼接操作；X_1'表示拼接后的输出。

拼接后得到的输出结果 X_1'的通道数为：

$$C' = \left[\frac{1}{n} + \frac{1}{3n}(n-1)(r_b+2r)\right]C \tag{8}$$

式中：C'表示输出的通道数。

输出可以等效为 $r'C$，公式(1)中 X_1 通道为 r_1C。且因为 1×1 大小的卷积核是全连接层不构成深度卷积，则当 k_b 与 k 均不等于 1 时，本文提出的改进结构可等效为单个卷积核的深度卷积，等效缩放因子为 r'。若忽略深度卷积层中的偏置与输入、输出的通道数变化，假设普通卷积的参数量为 $k'2C^2$、计算量为 $2k'2C^2HW$，k'为卷积运算的卷积核大小，深度卷积的参数量与计算量分别为 $k'2C$ 和 $2k'2CHW$。本文改进后等效的深度卷积结构分别具有 $k_b\times k_b$、$1\times k$、$k\times 1$ 的卷积核，原输入通道被分为四路依次输入，若通道全部为均等划分，则通道数(C)为原有的 1/4，修改后的等效深度卷积层结构参数和计算量分别为$(2k+k_b2)C/8$和$(2k+k_b2)CHW/4$。改进后的结构具有更少的参数量与计算量，同时满足实时性要求，且不降低模型复杂度。

1.4.2　TDAM 注意力模块　自然田间条件下杂草分布表现出零散或集群分布，即杂草零星随机分布或者若干杂草密集生长成一个杂草簇，成簇杂草内各种杂草与作物相互重叠、遮挡，难以区分，使苗草检测增加很大难度。注意力机制可以快速扫描图像，筛选出感兴趣的区域，对特定区域进行更多运算，同时抑制其他区域，是提高检测效率的有效方法[27]。对于成簇分布的杂草，运用注意力机制可模仿眼球聚焦与联系图像上下文特征，提高困难情况检测效果。图 2 的 TDAM 注意力模块是一种自顶向下注意力机制。该模块的输入是处于不同维度的 2 个特征块的输出，注意力模块的输入分别做池化、拼接、卷积操作得到“注意力探照灯(visual searchlight)”。“注意力探照灯”依次通过激活函数和卷积层得到通道注意力。通道注意力与低级特征逐点卷积后再与“注意力探照灯”卷积，随后降维得到空间注意力。模块的最终输出是空间注意力的逐点卷积，输出添加在低级特征块前，循环次数(T)等于高级特征块与低级特征块间差的块数，本文设定 $T=2$。

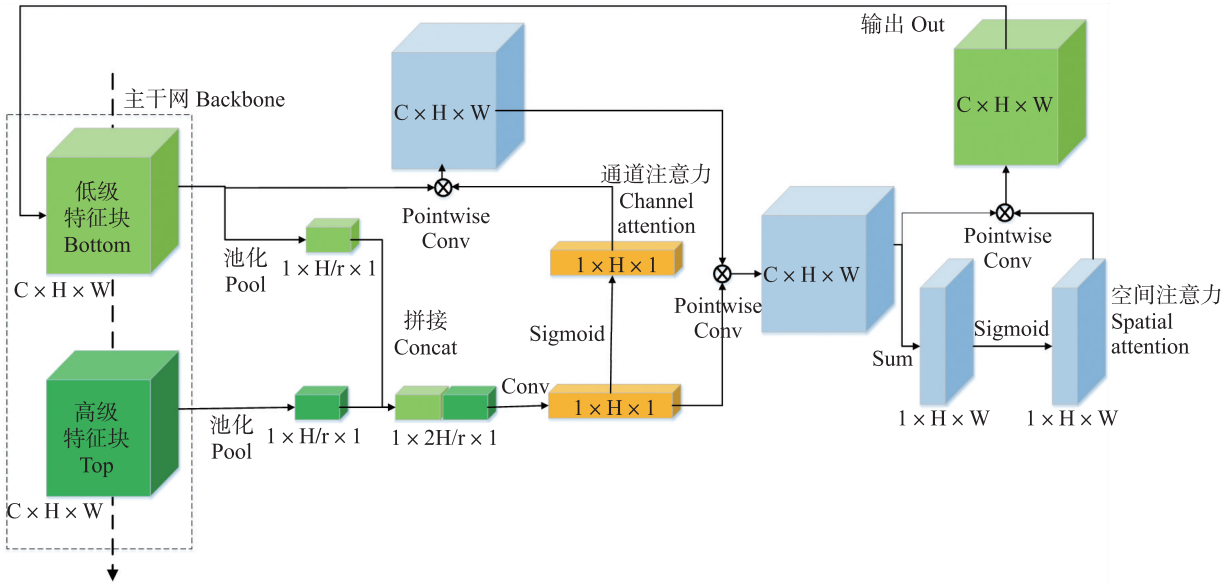

图 2　TDAM 注意力模块示意图

Fig. 2　Diagram of TDAM attention mechanism

1.4.3　模型整体结构　图 3 是本文模型整体结构。将改进后的特征提取块串联组成主干网，组成 4 个不同尺寸的特征图输出阶层(Stage)，Stage 1、Stage 2、Stage 4 由 3 个改进的特征提取块串联组成，Stage 3 则由 9 个串联组成。各个特征图间有下采样层，通过下采样层后特征图数根据采样步长(stride)发生变化(本文减半)。视觉自注意力模型的 Transformer 架构参照 RT-DETR。

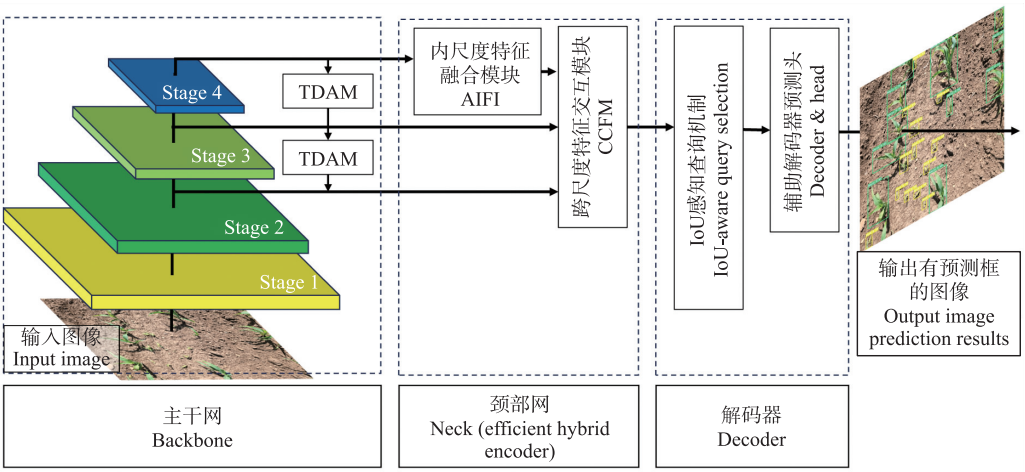

图 3 本文模型整体结构

Fig. 3 The overall model structure of this paper

1.4.4 损失函数 本模型的损失函数为边界框回归损失与分类损失之和。其中边界框的回归损失定义为平滑绝对误差与泛化交并比损失之和[28]。而分类损失采用焦点损失[29],其中正、负样本控制系数取值范围为[0,1],该系数通过交叉验证选取,本文中取正样本控制系数为 0.25。聚焦参数取值范围要求大于等于零,通过增加该系数减少易分类样本的权重,训练时对易分类样本惩罚力度更大,使得模型更专注于困难样本,当聚焦参数取零时,焦点损失退化为交叉熵损失,本文取 0.2。

1.4.5 组合图像增强方法 农田图像中的杂草目标多,自然田间图像采集过程中会受到运行机构摇晃颠簸、图像采集装置模糊失焦、田地背景差异、不同天气光照下亮度色度差、图像信息不完整等不利因素影响。在模型训练中添加图像增强预处理过程,对已有的数据进行扩充,能够增加模型检测精度与泛化能力。根据 Cubuk 等[30]的研究,自动图像增强策略(auto augment)将多种基本图像处理方法组合成数据增强策略集,将寻找合适的组合方式视为一个搜索问题,用强化学习方法(单层 LSTM[31])寻找合适的组合策略,包括 2 类参数:执行该操作的概率与操作强度。本文在训练读取数据时应用自动图像增强策略进行图像增广,保证每次训练中图像增强得到的结果均不同,提高了模型的泛化能力。本文选用的组合图像增强策略(包括 RT-DETR 的 3 种图像增强方法)为:随机扭曲、随机扩展、随机遮挡、目标框水平平移、均衡、剪切翻转、色调分离、剪切混合、曝光、目标框旋转,这 10 种图像增强方法组成组合图像增强策略。

2 结果与分析

2.1 模型训练设置

训练模型开始前加载 ConvNeXt 的主干网预训练模型(COCO-DataSet 数据集)。采用单卡训练,批处理大小设定为 4;优化器选择 AdamW 方法,训练前期使用渐进预热方法学习 2 000 步,配置初始学习率为 1×10^{-4};学习率变化策略:Piecewise-Decay,其衰减率为 0.98,衰减周期为 100 步;网络模型训练时最大迭代次数为 150。训练过程中每 1 代记录 1 次训练集损失,每 3 代进行 1 次评估并保存 1 次验证集损失。图 4 所示的是损失和平均精度随训练代数变化的曲线。随着训练次数增加,训练集损失逐渐接近验证集损失(图 4-a),表明设定的迭代次数内模型训练已达到稳定。平均精度变化曲线(图 4-b)则反映了

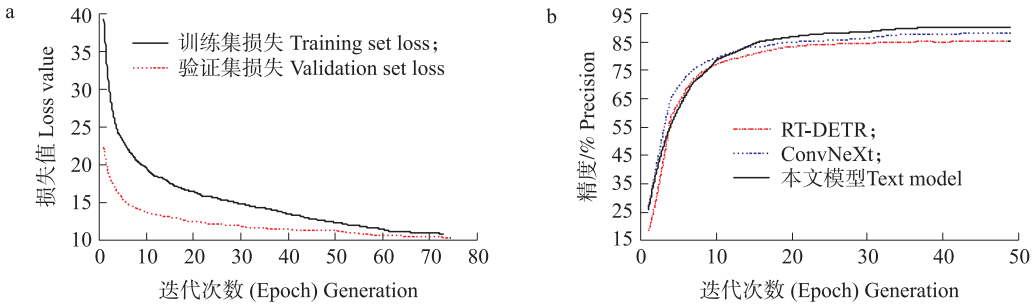

图 4 损失变化曲线(a)与平均精度变化曲线(b)

Fig. 4 Loss curve(a)and accuracy curve(b)

RT-DETR 模型、主干网替换为 ConvNeXt 的模型和本文提出的大尺度深度卷积的等效替代结构在相同条件下的模型精度变化,本文提出的等效结构加快了模型收敛速度且能保证模型精度无明显下降。

2.2　网络结构改进效果试验

2.2.1　等效卷积参数的影响　大尺寸的深度卷积层等效通道分离后模型的训练收敛速度与推理耗时均会提高,但精度会损失,因此研究深度卷积等效通道分离参数设置的影响,需寻找一组速度与精度达到最优平衡点的参数。修改后的等效深度卷积层结构中参数和计算量应当分别是$(2k+k_b2)C/8$和$(2k+k_b2)CHW/4$,参数 k 与 k_b 取值越小,网络的轻量化效果越明显。为了保证卷积层不退化且卷积计算有效,k 与 k_b 均为大于 1 的奇数,由于对比的 ConvNeXt 其卷积尺度为 7×7,因此卷积结构的轻量化设定小于 7×7。最后选取满足 $2k+k_b2\leqslant 35$ 的参数组合进行试验,寻找合适的参数组合。满足条件的等效深度卷积参数组合如表 1 所示。

表 1　等效卷积参数组合取值

Table 1　Possible parameter values for equivalent convolution

k_b 可能取值 k_b possible value	确定 k_b 后 k 可能取值 The possible value of k after k_b of determine
3	3,5,7,9,11,13
5	3,5

k 与 k_b 的参数组合试验结果如图 5 所示,选取推理耗时少、推理精度高的参数组,即图中右下方向的点。可以看出参数组合(3,9)明显优于其他参数组合,故本文取 $k_b=3$,$k=9$ 的块结构建特征提取网络。由图 5 可知,随着 2 组参数 k 与 k_b 取值的增大,模型的推理耗时呈减少趋势,与试验前推测一致,但参数组合取(3,3)时模型推理时间与精度均大幅增加;而模型推理精度并未表现出规律的变化趋势,也不随参数组接近等效前形状而表现出相近的推理精度。

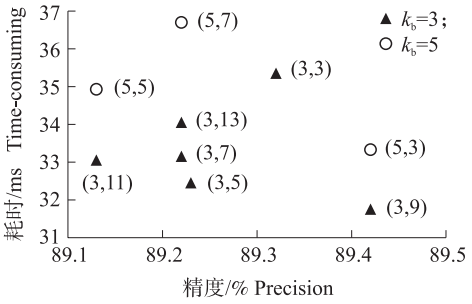

图 5　等效卷积参数取值的影响

Fig. 5　The effect of the value of equivalent convolution parameters

2.2.2　主干特征提取网络的影响　得到等效卷积最优参数后与主流主干网对比,验证改进的效果(表 2)。对主干网中的 7×7 深度可分离卷积进行上文所述的通道分离,其精度略微降低 0.58%,推理耗时减少 11.69 ms,整体性能强于其他结构,故运用本文所述的通道分离方法可以有效改善模型的推理耗时同时保证精度影响小,更加符合除草操作中检测杂草实时性的要求。试验涉及的其他主干网内卷积均换为深度可分卷积,且保证模型规模或深度级别相近。

表 2　不同网络主干网对模型检测效果的影响

Table 2　The impact of different network backbones on model detection results

主干网 Backbone network	平均检测精度/% Mean average precision(mAP)	参数量/MB Params	处理时间/ms Processing time
ResNet50	86.38	54.22	34.08
MobileNetv3[32]	83.39	50.13	40.81
Mobileone[33]	77.39	42.78	21.91
ConvNeXt	90.00	56.15	44.42
本文模型 Text model	89.42	42.96	32.73

2.2.3　不同注意力机制对模型检测效果的影响　在主干中引入注意力模块进一步改善模型精度。本文模型中主干网 3 个等级的 Stage 分别作为低、中、高级信息输入高效混合编码器(颈部网),3 个尺度的信息中仅最高层级输出同时通过内部尺度的特征交互融合模块和跨尺度融合处理模块。而 2 个低级特征,会输入到跨尺度融合模块中不同位置,此处具备添加 TDAM 条件。

为检验改进模型有效,将其与常用的 3 种注意力模块对比,将不同注意力机制添加到相同位置进行试验,其结果如表 3 所示。激励通道注意力模块(sequeeze and excite,SE)[34]可以提升模型平均精度,但增加参数量和推理耗时;而卷积注意力模块(convolutional block attention module,CBAM)[35]并未提升模型精度;坐标注意力模块(coordinate attention,CA)[36]对模型识别效果的提升总体不如 SE;自顶向下注意力模块(top down attention module,TDAM)[37]对模型精度提升最有效,比无机制增加 0.7%,而实时性的主要指标推理耗时增加较少。TDAM 的效果主要是提升大植株重叠的一类小目标,相对其他目标提升 1.3%,小

目标物体的平均正确率数值(average precision value)为 64.0%,表明 TDAM 的自顶向下多级特征联合能够更好细化特征。

表 3 不同注意力机制对模型检测效果的影响

Table 3 The impact of different attention mechanism on model detection performance

注意力机制 Attention mechanism	平均检测精度/% mAP	参数量/MB Params	处理时间/ms Processing time
无机制 No mechanism	89.42		32.73
挤压激励通道注意力模块(SE)[34]	89.62	0.13	32.75
卷积注意力模块(CBAM)[35]	89.87	0.24	33.17
坐标注意力模块(CA)[36]	89.55	0.18	33.02
自顶向下注意力模块(TDAM)	90.12	1.89	33.60

2.3 改进模型的消融试验

为了评估模型改进的效果,设计消融对比试验,RT-DETR 模型设定为 A,各改进方法的模型如表 4 所示,消融试验的结果如表 5 所示。消融试验表明改进方法对模型的性能均有提升。相较于原模型的 3 种通用图像预处理方法,模型推理评估阶段组合图像增强方法的精度提高 0.64%,该方法不参与推理评估且不影响推理速度。主干网替换能利用数据的更多特征,更好识别多情况的玉米与杂草,精度进一步提升,但模型处理时间增加。TDAM 注意力机制提高模型对目标聚集区域及小目标的敏感程度,解决玉米与杂草识别任务中目标小及植株伴生重叠、遮挡的目标不完整、不清晰的问题。等效卷积降低了模型内存占用,处理速度提高 25.65%,精度损失小,有效改进了模型整体性能。消融试验结果表明改进的模型在采集的玉米苗与杂草数据集上平均精度更高,实时性效果更好,能更好完成检测任务。

表 4 各改进方法的消融试验

Table 4 Methods involved in ablation tests

改进方法 Improvement method	模型 Model				本文模型 Text model
	A	B	C	D	
自动数据增强 Auto augment		√	√	√	√
替换主干网 Replace the backbone			√	√	√
注意力模块 Attention module				√	√
等效卷积 Equivalent convolution					√

注:A 为 RT-DERT 模型,B~D 为改进方法的模型。下同。√:采用了改进方法。

Note: A indicates RT-DERT model, and B-D indicate methods involved models. The same as follows. √: Improved method adopted.

表 5 不同方法对模型的影响

Table 5 The influence of different methods on the model

模型 Model	平均检测精度/% mAP	参数量/MB Params	处理时间/ms Processing time
A	85.74	54.22	33.28
B	86.38	54.22	33.28
C	89.42	56.15	44.42
D	90.12	58.04	45.29
本文模型 Text model	90.11	44.86	33.67

2.4 不同目标检测模型性能比较试验

在相同测试数据集下将本文模型与 7 个主流目标检测模型进行对比试验,结果如表 6 所示。改进后的模型平均检测精度(mAP)最高达 90.11%,处理时间 33.67 ms,对比其他模型整体性能提升。本文提出的改进方法具有最高的识别准确率,减少玉米杂草管理过程中的误检与漏检。同时该方法具有较快的单幅图像处理速度,在计算资源受限条件下保证玉米苗与杂草识别的实时性能。

表 6 不同目标检测模型性能比较

Table 6 Performance comparison of different target detection models

模型 Model	平均检测精度/% mAP	参数量/MB Params	处理时间/ms Processing time
YOLOv5	79.68	7.07	244.00
YOLOv7	81.79	37.20	30.10
Cascade-RCNN	82.76	69.23	100.75
FCOS	84.68	32.16	60.91
Faster-RCNN	82.37	99.36	123.43
Rt-DETR	85.74	54.22	34.08
本文模型 Text model	90.11	44.86	33.67

2.5　多种情况下识别效果试验

实际玉米田间环境复杂，杂草生长状况多样且分布不均匀，常有玉米与杂草伴生重叠、枝叶遮挡情况。因此对少量目标、伴生目标、多种类杂草混合、存在异物的 4 类种情况下的图像进行模型检测效果验证试验。将本文模型与视觉自注意力模型的 RT-DETR 及表 6 中识别精度较高的卷积模型 FCOS 对比，其对作物和杂草目标的识别效果如表 7 所示。

表 7　模型对两类目标检测结果

Table 7　Two type results of object detection

模型 Model	玉米识别精度/% Maize identification accuracy	杂草识别精度/% Weeds identification accuracy
FCOS	94.4	72.1
Rt-DETR	99.3	72.4
本文模型 Text model	99.6	79.9

图 6 是 3 种模型在不同情况下的识别效果展示。少量目标且不交叠情况下杂草与玉米苗的识别最简单，各类识别模型均能得到不错效果。杂草与玉米伴生重叠或存在大量目标的情况下，目标多、形态与大小差异巨大，被遮挡的目标不完整，容易漏检。从图 6 可见，本文模型的杂草和玉米识别效果最好。此外本文模型对于存在垃圾等异物影响的情况不容易错误识别，抗干扰与泛化能力较强。

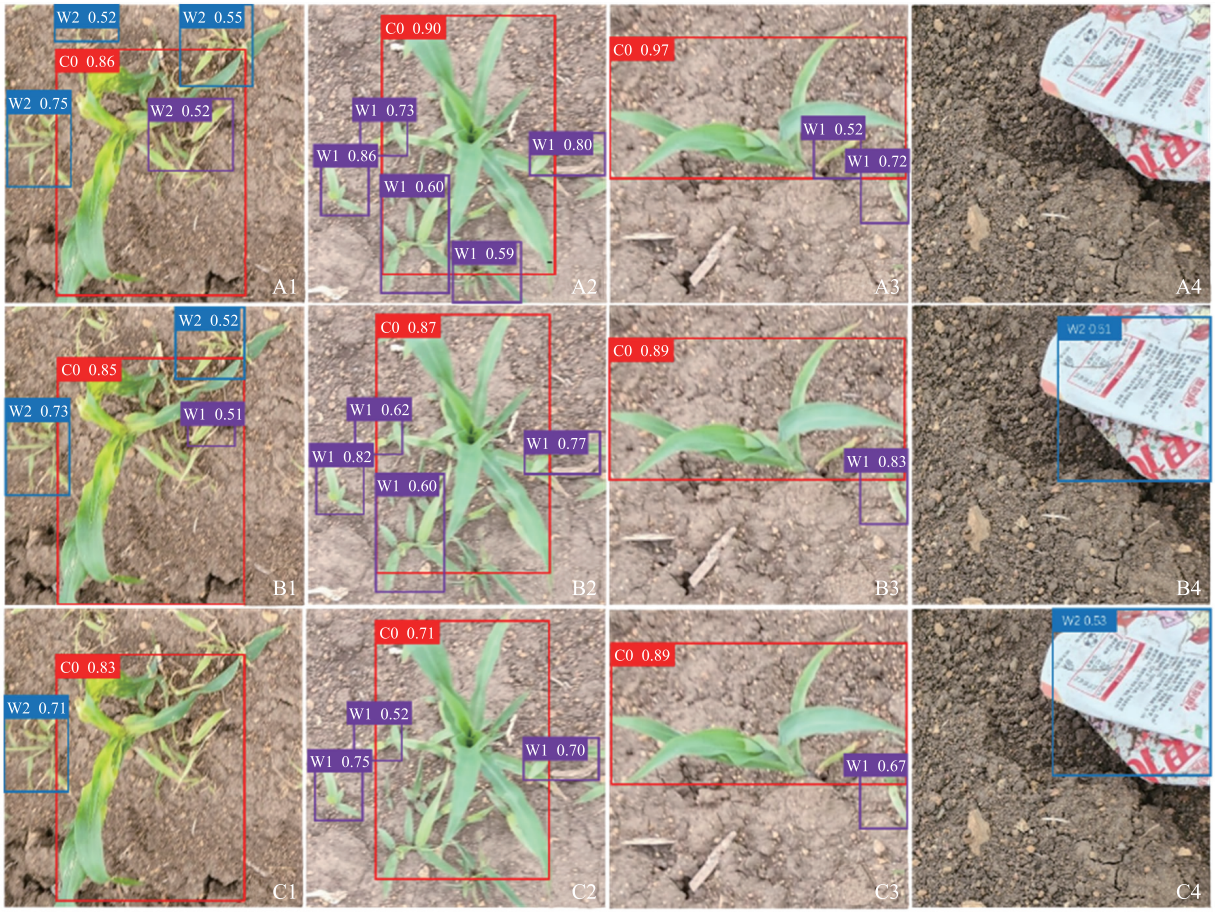

图 6　模型检测效果示例图

Fig. 6　Diagram of model checking effect

A、B、C 分别表示本文模型、RT-DETR、FCOS 检测效果；1、2、3、4 分别表示多类杂草、杂草和作物伴生、少量目标、异物条件下识别效果。目标框标注中 W1、W2 表示 2 类杂草，C 表示玉米，数字表示置信度。

A, B and C represent the detection effect of the text model, RT-DETR and FCOS, respectively. 1, 2, 3, 4 indicate the recognition effect under conditions of multiple weeds, weed and crop associated, a small number of targets, and foreign bodies, respectively. In the target box, W1 and W2 represent two types of weeds, C represents corn, and the number represents confidence.

3 结论

本文提出一种基于视觉自注意力模型的玉米苗与杂草检测方法,涉及等效大尺度深度卷积、注意力机制、组合图像增强方法,进行了等效卷积模型的参数选取试验,将本文模型的改进方法与主流方法进行对比,本文方法在多种条件下识别结果均最好。

试验中发现杂草与玉米苗识别任务的主要难点是识别杂草,各识别模型对杂草的识别精度均差于玉米苗,主要的识别错误是小目标、重叠或不完整目标的漏检。该问题难以通过调整模型消除,这表明数据集标注时对小目标和不完整目标的标注存在问题,可能忽略了部分小目标和不完整目标。尽管标注图像时要求忽略图像中极小和过于不完整目标,但有少数标注了这2类情况的目标,影响了数据集质量,而目标框数量多且混杂,难以区分有误目标。此外有时模型会将一些异物错误识别,但对田间较常见的石块、秸秆等能正常处理,其原因是标注的数据集中不存在该类目标,而包括了一些常见异物,因此导致了错误检测。由于该类物体数目少,总体影响小,且各类模型训练前均已经加载了来自大规模图像数据集(COCO数据集)的预训练模型,因此模型具有一定识别其他物体能力,因此可以假定异物对模型识别效果无影响。

参考文献 References:

[1] 王宇,黄春艳,郭玉莲,等. 春玉米田杂草防治关键期[J]. 黑龙江农业科学,2017(6):33-36.
Wang Y,Huang C Y,Guo Y L,et al. Critical period of weed control in spring maize fields[J]. Heilongjiang Agricultural Sciences,2017(6):33-36(in Chinese with English abstract).

[2] 李香菊. 我国耐除草剂转基因作物研发与产业化应用前景[J]. 植物保护,2023,49(5):316-324.
Li X J. Development of herbicide tolerant crops and their commercialization in China[J]. Plant Protection,2023,49(5):316-324(in Chinese with English abstract).

[3] 郭永丽,祁圆林,于佳星,等. 2种色素合成抑制剂防除小麦田抗药性禾本科杂草的潜力研究[J]. 南京农业大学学报,2022,45(3):529-538. DOI:10.7685/jnau.202104024.
Guo Y L,Qi Y L,Yu J X,et al. Study on the potential of two pigment synthesis inhibitors to manage herbicide resistant gramineous weeds in wheat fields[J]. Journal of Nanjing Agricultural University,2022,45(3):529-538(in Chinese with English abstract).

[4] 王大庆,禄琳,于兴龙,等. 基于深度迁移学习的EfficientNet玉米叶部病害识别[J]. 东北农业大学学报,2023,54(5):66-76.
Wang D Q,Lu L,Yu X L,et al. Maize leaf diseases identification using EfficientNet based on deeptransfer learning[J]. Journal of Northeast Agricultural University,2023,54(5):66-76(in Chinese with English abstract).

[5] 徐会杰,黄仪龙,刘曼. 基于改进YOLOv3模型的玉米叶片病虫害检测与识别研究[J]. 南京农业大学学报,2022,45(6):1276-1285. DOI:10.7685/jnau.202110039.
Xu H J,Huang Y L,Liu M. Research on pest detection and identification of corn leaf based on improved YOLOv3 model[J]. Journal of Nanjing Agricultural University,2022,45(6):1276-1285(in Chinese with English abstract).

[6] Ren S Q,He K M,Girshick R,et al. Faster R-CNN:towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(6):1137-1149.

[7] 樊湘鹏,周建平,许燕,等. 基于优化Faster R-CNN的棉花苗期杂草识别与定位[J]. 农业机械学报,2021,52(5):26-34.
Fan X P,Zhou J P,Xu Y,et al. Identification and localization of weeds based on optimized Faster R-CNN in cotton seedling stage[J]. Transactions of the Chinese Society for Agricultural Machinery,2021,52(5):26-34(in Chinese with English abstract).

[8] Cai Z W,Vasconcelos N. Cascade R-CNN:high quality object detection and instance segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2021,43(5):1483-1498.

[9] Song P,Chen K Y,Zhu L F,et al. An improved cascade R-CNN and RGB-D camera-based method for dynamic cotton top bud recognition and localization in the field[J]. Computers and Electronics in Agriculture,2022,202:107442.

[10] Redmon J,Divvala S,Girshick R,et al. You only look once:unified,real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas:IEEE,2016:779-788.

[11] Liu W,Anguelov D,Erhan D,et al. SSD:single shot MultiBox detector[C]//European Conference on Computer Vision. Cham:Springer,2016:21-37.

[12] 亢洁,刘港,郭国法. 基于多尺度融合模块和特征增强的杂草检测方法[J]. 农业机械学报,2022,53(4):254-260.
Kang J,Liu G,Guo G F. Weed detection based on multi-scale fusion module and feature enhancement[J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(4):254-260(in Chinese with English abstract).

[13] Gallo I,Rehman A U,Dehkordi R H,et al. Deep object detection of crop weeds:performance of YOLOv7 on a real case dataset from UAV images[J]. Remote Sensing,2023,15(2):539.

[14] Tian Z,Shen C H,Chen H,et al. FCOS:fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision(ICCV). Seoul:IEEE,2019:9626-9635.

[15] Peng H X,Li Z H,Zhou Z Y,et al. Weed detection in paddy field using an improved RetinaNet network[J]. Computers and Electronics in Agriculture,2022,199:107179.

[16] He L,Zhou Q Y,Li X T,et al. End-to-end video object detection with spatial-temporal transformers[C]//Proceedings of the 29th ACM International Conference on Multimedia. Virtual Event:ACM,2021:1507-1516.

[17] Abuhani D A,Hussain M H,Khan J,et al. Crop and weed detection in sunflower and sugarbeet fields using single shot detectors[C]//2023 IEEE International Conference on Omni-layer Intelligent Systems(COINS). Berlin:IEEE,2023:1-5.

[18] Lv W Y,Zhao Y A,Xu S L,et al. DETRs beat YOLOs on real-time object detection[EB/OL]. (2023-04-17)[2023-10-13]. http://arxiv.org/abs/2304.08069.pdf.

[19] Crick F. Function of the thalamic reticular complex:the searchlight hypothesis[J]. Proc Natl Acad Sci USA,1984,81(14):4586-4590.

[20] 张宸嘉,朱磊,俞璐. 卷积神经网络中的注意力机制综述[J]. 计算机工程与应用,2021,57(20):64-72.
Zhang C J,Zhu L,Yu L. Review of attention mechanism in convolutional neural networks[J]. Computer Engineering and Applications,2021,57(20):64-72(in Chinese with English abstract).

[21] Deco G,Zihl J. Top-down selective visual attention:a neurodynamical approach[J]. Visual Cognition,2001,8(1):118-139.

[22] Liu Z,Mao H Z,Wu C Y,et al. A convnet for the 2020s[C]//2022 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans:IEEE,2022:11976-11986.

[23] Shantanu J,Basura F,Cheston T. TDAM:top-down attention module for contextually guided feature selection in CNNs[C]//European Conference on Computer Vision. Cham:Springer,2022:259-276.

[24] Yu W H,Zhou P,Yan S C,et al. InceptionNeXt:when inception meets ConvNeXt[EB/OL]. (2023-05-29)[2024-3-13]. http://arxiv.org/abs/2303.16900.

[25] Chollet F. Xception:deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu:IEEE,2017:1800-1807.

[26] Tan M,Le V Q. MixConv:mixed depthwise convolutional kernels[J]. CoRR,2019:abs/1907.09595.

[27] 王郝日钦. 基于深度学习的水稻知识智能问答系统理论与方法研究[D]. 沈阳:沈阳农业大学,2022.
Wang-Hao R Q. Research on theory and method of rice knowledge intelligent question answering system based on deep learning[D]. Shenyang:Shenyang Agricultural University,2022(in Chinese with English abstract).

[28] Zheng Z,Wang P,Liu W,et al. Distance-IoU Loss:faster and better learning for bounding box regression[C]//2020 Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto:AAAI Press,2020:12993-13000.

[29] Weber M,Fürst M,Zöllner J M. Automated focal loss for image based object detection[C]//2020 IEEE Intelligent Vehicles Symposium(IV). Las Vegas:IEEE,2020:1423-1429.

[30] Cubuk E D,Zoph B,Mané D,et al. AutoAugment:learning augmentation strategies from data[C]//2019 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach:IEEE,2019:113-123.

[31] Zoph B,Vasudevan V,Shlens J,et al. Learning transferable architectures for scalable image recognition[C]//2018 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City:IEEE,2018:8697-8710.

[32] Howard A,Sandler M,Chu G,et al. Searching for MobileNetV3. [J]. CoRR,2019:abs/1905.02244.

[33] Pavan K,Anasosalu V,James G. MobileOne:an improved one millisecond mobile backbone[C]//2023 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver:IEEE,2023:7907-7917.

[34] Hu J,Shen L,Sun G. Squeeze-and-excitation networks[C]//2018 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City:IEEE,2018:7132-7141.

[35] Sanghyun W,Jongchan P,Joon-Young L,et al. CBAM:convolutional block attention module[C]//European Conference on Computer Vision. Cham:Springer,2018:3-19.

[36] Hou Q,Zhou D,Feng J. Coordinate attention for efficient mobile network design[C]//2021 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville:IEEE,2021:13713-13722.

责任编辑:刘怡辰