

基于注意力机制金字塔网络的麦穗检测方法

章权兵 胡姗姗 舒文灿 程 鸿

(安徽大学农业生态大数据分析与应用技术国家地方联合工程研究中心, 合肥 230601)

摘要: 为了准确预测小麦产量,提出了一种基于特征金字塔网络改进的小麦穗部检测方法。针对检测结果中存在的误检或漏检等问题,本文首先在原始特征提取网络的编码和解码区域分别引入通道注意力机制和空间注意力机制,以增加对麦穗空间信息和语义信息的提取,有效提升网络对遮挡麦穗的检测性能;其次对原始区域建议网络的输入进行改进,设计了一种加权区域建议网络,在通道级别上将高层具有强语义信息的低分辨率特征图融合在一起,经过一系列的全连接层和激活函数生成对应维度的概率后,对底层高分辨率特征图进行加权以增强有用的信息通道,为难以检测的较小麦穗生成更精确的检测框。关于实地采集的灌浆期麦穗图像的实验结果表明,本文方法明显改善了对遮挡麦穗和较小麦穗的检测效果,其检测精确度、召回率和平均精度分别达到 80.53%、87.12% 和 88.53%。通过对公开 ACID 数据集上不同时期麦穗检测结果的对比分析,进一步验证了本文方法的有效性。

关键词: 麦穗检测; 金字塔网络; 特征提取; 注意力机制; 加权区域建议网络

中图分类号: TP391.41 文献标识码: A 文章编号: 1000-1298(2021)11-0253-10 OSID: 

Wheat Spikes Detection Method Based on Pyramidal Network of Attention Mechanism

ZHANG Quanbing HU Shanshan SHU Wencan CHENG Hong

(National Engineering Research Center for Agro-Ecological Big Data Analysis and Application,
Anhui University, Hefei 230601, China)

Abstract: With the aim to predict the wheat yield accurately, an improved wheat spikes detection method based on feature pyramid network was proposed. In order to solve the problem of misdiagnosis or omission in the detection results, channel attention mechanism and spatial attention mechanism were introduced into the coding and decoding regions of the original feature extraction network, which increased the extraction of spatial information and semantic information on the wheat spikes and effectively improved the detection performance of the network for obscured wheat spikes. At the same time, a weighted-region proposal network was designed to improve the input of the original region proposal network, in which several low-resolution feature maps with strong semantic information characteristics were fused together on channel levels. After a series of full connection layers and activation functions, the fused feature map was converted to probability of the corresponding channels, which were used to weight the underlying high-resolution feature maps to enhance useful information channels. Thus, a more accurately detection frame was generated for smaller spikes which were difficult to detect. The experimental results of the collected wheat spikes images showed that the method could significantly improve the detection effect of the shaded and smaller wheat spikes, where the precision of recognition, recall rate and average precision were 80.53%, 87.12% and 88.53%, respectively. Through the comparative analysis of wheat spikes detection results in different periods on the public ACID data set, the validity of the proposed method was further verified.

Key words: wheat spikes detection; pyramidal networks; feature extraction; attention mechanism; weighted region proposal network

0 引言

全球粮食年产量高达 7 亿 t,而在众多的粮食作物中,小麦是最重要的农作物品种之一,且是世界 1/3 人口的主要食物来源,其种植面积和总产量均居谷类作物之首^[1]。随着城市化进程的加快和人均收入的增长,全球小麦消费也在快速增长。预计到 2050 年,小麦需求量将比现在增加 60%^[2]。在我国,小麦是主要粮食作物和战略储备粮,其产量对国家粮食安全具有重要影响。准确的产量预测不仅是小麦生产管理的重要环节之一,而且可以为农业管理决策者提供参考^[3]。穗数是决定单位面积产量的关键因素,快速、准确地对麦穗进行检测和计数,不仅在作物管理和作物产量评价中具有重要作用,而且在作物育种、植株表型分析等方面也具有十分深远的意义。

传统的小麦产量预测主要依靠人工计数和图像处理的方法。人工计数方法繁琐,容易出错且不具有代表性^[4];而图像处理方法是可实现麦穗自动计数的一种快速简便的方法,因此得到了大量研究^[5-7]。上述传统方法需要大量实验寻找有用的特征,工作量和复杂度都很高。此外,在实际的田间环境中,小麦紧密地生长在一起,麦穗间及麦穗与叶片间相互遮挡,传统方法对麦穗的检测效果并不理想。

随着现代植物表型技术的快速发展,传统方法已经无法满足需求,而深度学习方法逐渐被大多数研究者应用到农作物的产量预测中,特别是卷积神经网络(Convolution nerual network, CNN)在图像分析中的应用开始成为主流,这使得自动化麦穗计数成为可能^[8]。LU 等^[9]提出了一种名为 TasselNet 的深度卷积神经网络来建立麦穗计数回归模型。为了研究不同生长阶段的小麦对检测结果的影响,HASAN 等^[10]使用相同的 R-CNN 结构在 4 个不同生长阶段捕获的不同数据集上生成 4 个不同的模型,从而实现小麦穗粒计数。SADEGHI-TEHRAN 等^[11]使用简单线性迭代聚类(Simple linear iterative clustering, SLIC)提取冠层相关特征,构建特征模型,然后将其添加到用于小麦穗部语义分割网络中,但该方法对密集麦穗的检测效果并不理想。张领先等^[12]基于图像处理和深度学习技术,设计并实现了基于卷积神经网络的冬小麦麦穗检测计数系统。将构建的冬小麦麦穗识别模型与非极大值抑制结合,进行冬小麦麦穗计数。MISRA 等^[13]提出了一种新的深度学习网络 SpikeSegNet,该网络结合了两种新的特征网络:局部补丁提取网络(Local patch

network, LPNet)和全局掩码细化网络(Global mask refined network, GMRNet),并应用 imageJ 的分析粒子函数计算出了图像中的麦穗数。与早期的图像处理等方法相比,以上方法对麦穗的检测精度有所提高。然而,上述算法大多都只能在麦穗独立无遮挡且背景比较简单的情況下,才能取得较好的检测结果,因此如何在麦穗目标较小且相互遮挡的情况下提高网络的检测性能成为关键性问题。

本文以麦穗为研究对象,首先在金字塔特征提取网络中引入注意力机制模块,建立通道空间特征金字塔网络(Channel spatial feature pyramid network, CSFPN),有选择地增强网络的语义信息和位置信息,减少噪声信息对检测结果的影响;其次根据麦穗图像的实际输入尺寸,提出一种加权区域建议网络(Weighted-region proposal network, WRPN),生成更准确的待检测框,提高对小目标麦穗的检测效果。最后,为进一步验证本文方法的性能,将其与更快的卷积神经网络(Faster R-CNN)^[14]、特征金字塔网络(Feature pyramid network, FPN)^[15]以及近期提出的 YOLO v4 网络^[16]进行比较。

1 卷积神经网络

1.1 Faster R-CNN

Faster R-CNN 是一种端到端的目标检测算法,它将候选区域生成、特征提取、分类和位置精修 4 个步骤统一到一个网络框架内。该网络首先利用共享的残差网络 ResNet101^[17]提取图像的特征,并将提取到的特征图传送至区域建议网络(Region proposal network, RPN)中,生成所有可能的候选框。其次利用 Softmax 函数判断候选框内是待检测的目标还是背景,并采用边框回归方法对候选框进行初步回归和校正。然后利用感兴趣区域(Region of interest, ROI)池化层^[18]将特征图与 RPN 输出的候选框集成,并选择与特征图上的每个感兴趣区域相对应的特征。最后使用全连接层来确定目标类别,此时再次对候选框进行回归校正,以获得目标的精确位置和类别。Faster R-CNN 虽然对大目标的检测效果较好,但由于其特征提取网络不断下采样,当进行卷积池化到最后一层时,损失了大量的语义信息,对小目标的检测效果并不理想。

1.2 FPN

特征金字塔网络(FPN)是一种自下而上的结构,主要解决物体检测中的多尺度问题,通过对网络连接做简单改变,在基本不增加计算量的情况下,提升小物体检测的性能。它先进行一系列的下采样操作,得到原始图像尺寸 1/4、1/8、1/16、1/32 的特征

图,这是网络的前馈操作过程。然后逐步进行上采样,生成相应尺寸的特征图。在对每一层信息进行处理时,将前一层信息作为输入,并将上采样结果与自下而上生成的具有相同分辨率的特征图进行水平连接。两层特征图融合后,对其结果进行 3×3 卷积,消除上层采样的混叠效应。

FPN 在进行特征融合时,将前一层的粗糙特征直接转化为后一层的细化特征。在经过一系列的下采样后,有一些高层次的特征图(如原始图像尺寸的 $1/16$ 、 $1/32$),相对于底层特征而言,位置信息较弱。经过采样并与具有相同分辨率的特征图融合后,得到的特征图仍然存在位置信息弱、位置信息融合不足等潜在问题,对网络性能影响较大。

1.3 RPN

区域建议网络(RPN)是一个专门用于提取候选区域的全卷积网络,对前期提取到的特征图进行处理,寻找可能包含目标区域的候选框,并预测每个框的类别得分。输入 RPN 的特征图与原始图像有一定的映射关系,首先使用滑动窗口(Sliding window)在图像上滑动,生成大量的锚(anchor),然后对每个锚做前背景的二分类判断,再通过非极大值抑制(NMS)^[19]算法对候选框进行调整以消除冗余框,并将调整后的候选框作为感兴趣区域输出。

RPN 的输入图像尺寸为原始图像尺寸的 $1/4$ 、 $1/8$ 、 $1/16$ 、 $1/32$,从而利用多尺度空间信息生成待检测区域。然而,在麦穗检测任务中,初始图像中的最大麦穗尺寸约为 $32 \text{ 像素} \times 20 \text{ 像素}$ 。在经过

大的逐步下采样(例如 $1/16$ 、 $1/32$)后,很难通过 RPN 在这些特征图中生成待检测框。但如果只使用原始图像尺寸的 $1/4$ 作为特征图,则会浪费网络顶层的分类信息,从而对网络的性能有很大的影响。

1.4 注意力机制

视觉注意力机制是人类视觉中一种特殊的脑信号处理机制。它通过快速扫描图像来捕获目标区域,然后更加注意获取更详细的信息,并抑制其他无用信息。对人类来说,这是一种以有限的注意力快速筛选大量信息的方法。在现有的语义分割体系中,金字塔结构可以提取不同尺度的特征信息,但缺乏全局上下文的优先关注。因此,使用注意力机制^[20]在传统神经网络的基础上增加新的连接,就可以自动确定对输入的每个部分应该分配多少注意,从而为卷积神经网络提取的特征提供准确的像素级关注度。卷积神经网络中常用的两种注意力机制是通道注意力机制(Channel attention mechanism, CAM)和空间注意力机制(Spatial attention mechanism, SAM)^[21]。通道注意力机制是一个一维的特征图谱,特征图中的每个通道被赋予一个权重;空间注意力机制以特征图上的每个像素为单位,为其赋予一个权重,它是一个二维特征图谱。

2 卷积神经网络改进

本文选择 Faster R-CNN 作为网络框架,并引入注意力机制和加权区域建议网络对麦穗进行检测。卷积神经网络模型架构如图 1 所示。

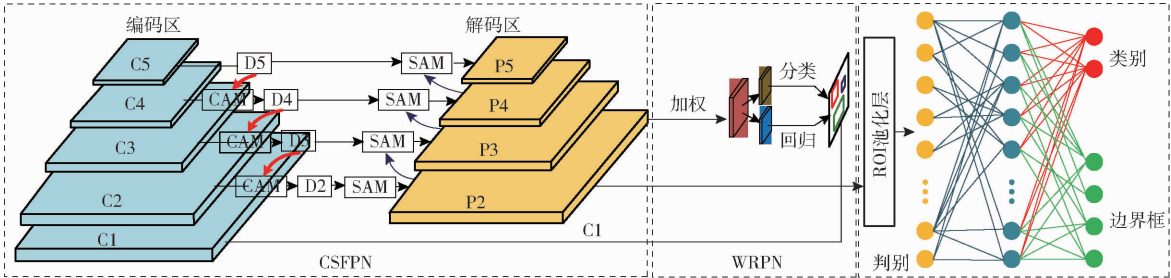


图 1 卷积神经网络模型架构
Fig. 1 Architecture of CNN model

图 1 主要由 3 部分组成,其中 CSFPN 模块在金字塔网络的编码区添加了通道注意力机制,以帮助提取信道中有意义的信息并抑制噪声信道;解码区引入了空间注意力机制,增强有用的位置信息以加强解码部分空间信息的整合,从而避免编码部分的粗糙特征直接流向解码区域。设计了一种新的加权区域建议网络 WRPN,使其生成更精确的待检测框,再执行第 1 次候选框分类回归操作,确定候选框并进行前背景的判别。经过 ROI 池化层之后,对候选

框进行第 2 次分类回归操作。由于网络中引入的注意力机制是一个轻量级的模块,因此可以把它们集成到卷积神经网络的框架中一起进行端到端的训练。实验结果表明,引入的注意力机制模块可以提升网络对麦穗的检测性能,且增加的运算量可以忽略不计。

2.1 通道注意力机制

通道注意力机制主要关注图像中的语义信息。在网络训练过程中,利用众多通道的卷积核对原始

图像进行映射时,会生成与通道数相同数量的特征图。然而,并不是每一个通道都对检测结果具有积极的影响。在整个网络的每个阶段中,都会有许多噪声通道与有意义的通道相混合。因此,有效地过滤噪声通道是提高网络性能的关键。本文在 FPN 的解码区引入了通道注意力机制模块,利用该模块对特征图中的不同通道赋予不同的权重,增加通道间的差异,从而提高编码器部分的特征提取性能。如图 2a 所示,首先通过全局平均池化层 (Global average pooling, GAP) 来集成每个通道的全局信息,将尺寸为 $W \times H \times C$ (W 、 H 、 C 分别为特征图的宽、高

通道数)的上层特征图 D5 拉伸为 $1 \times 1 \times C$ 的一维向量。然后利用收缩全连接层和 ReLU 激活函数建立通道间的相互依赖关系,并将所得的一维向量降维至 $1 \times 1 \times C/4$,再通过扩展全连接层将其映射回原始维度 $1 \times 1 \times C$,并发送到 Sigmoid 激活函数,获得对应通道的概率。最后用所得的概率对底层特征图 C4 进行加权,再将加权处理后的结果与高层特征图 D5 进行通道级的融合 (concat),并将融合结果作为下一层的输入 D4。这样,网络就可以通过全局信息的映射,有选择地增强含有有用语义信息特征的通道,抑制含有无用信息的通道。

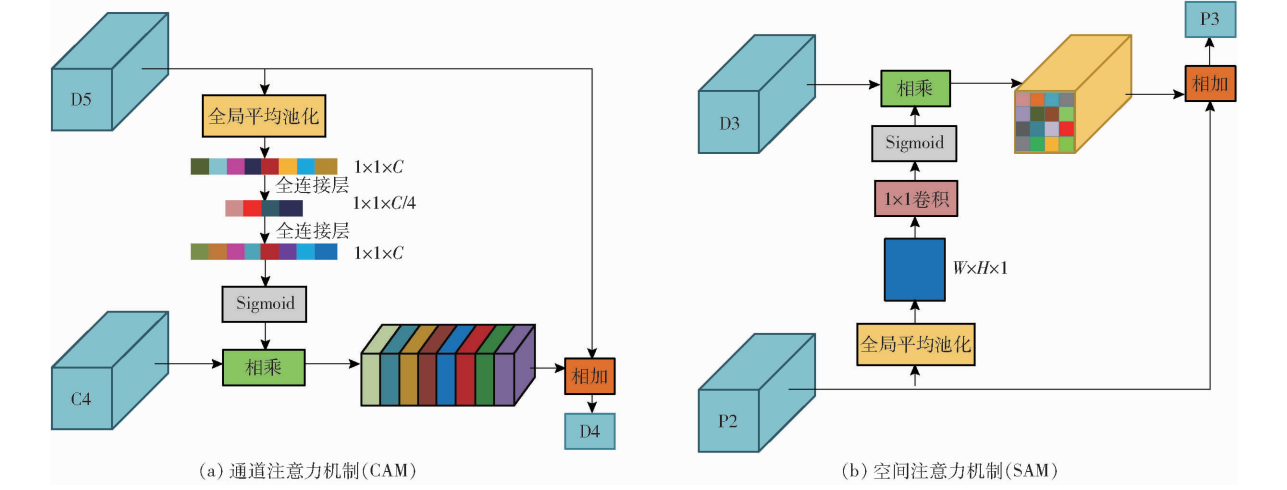


图 2 改进的特征融合部分

Fig. 2 Improved feature integration component

2.2 空间注意力机制

与通道注意力机制不同,空间注意力机制侧重于空间维度中的位置信息。在当前目标检测结构中,金字塔结构被广泛应用于通过在编码部分下采样和在解码部分恢复空间分辨率来提取不同尺度的特征信息。然而,常用的上采样插值方法在分辨率恢复中引入了大量的位置信息噪声^[22]。同时,如果编码器部分的粗糙特征映射直接与解码器部分的细化特征相结合,则会导致性能下降^[23]。为此,本文引入空间注意力机制对上采样阶段的噪声信息进行滤波,减少低维特征对检测结果的影响。首先将高分辨率特征图 D2 发送到全局平均池化层,与 CAM 不同的是,其将通道聚集起来生成全局空间信息图,特征图的维数从 $W \times H \times C$ 变为 $W \times H \times 1$ 。其次添加一个核尺寸为 1×1 的卷积层来调整每个像素的值,并将调整后的特征图输入到 Sigmoid 激活函数中,生成空间概率图。然后将编码器部分的特征图 D2 乘以其生成的概率图来加权自身,生成特征图 P2。最后进行与 D2 类似的操作,先由 P2 生成空间概率图,并利用其对 D3 进行加权,再将加权结果与 P2 融合为下一层 P3,如图 2b 所示。这里,空间注

意力模块选择性地增强了有用的位置信息,抑制了无意义的位置信息。

2.3 加权区域建议网络

RPN 的设计是通过取不同尺寸的特征图作为输入,并形成一系列的区域建议网络。在麦穗检测任务中,小麦的目标非常小,且相互重叠,从而在一个小尺寸的输入图像中很难产生几十个精确的麦穗检测框^[24]。使用大尺度的特征图作为区域建议网络是提升小目标检测结果的一种有效的可行方案。同时,为了充分利用网络中的语义信息,本文在原始 RPN 的输入部分设计了加权模块,提出了一种修正后的加权区域建议网络 (WRPN),使其更适合于小目标的检测,如图 3 所示。与空间分辨率较大的特征图相比,高层较小的特征图具有更强的语义信息。为了更好地生成待检测区域,将高层较小的低空间分辨率特征图 (P3、P4、P5) 融合在一起,融合方式为通道级上的相加。融合后的特征图经过全局平均池化层,生成特征向量。再通过收缩全连接层将特征向量降维至 256,并由 Sigmoid 激活函数生成对应的概率。这样具有较强语义信息的高层较小的特征图 (P3、P4、P5) 就可以对低层较大的特征图 (P2) 进行

加权,以增强低层特征图中的有用通道,进而更好地检测出目标。考虑到提高空间分辨率可以提高检测精度,将 P2 上采样为和 C1 同样尺寸的特征图,再经过 1×1 卷积层处理至相同通道后与 C1 相融合以获得更丰富的位置信息。最后将特征向量加权到融合后的特征图,并将加权结果输入到 RPN 中。

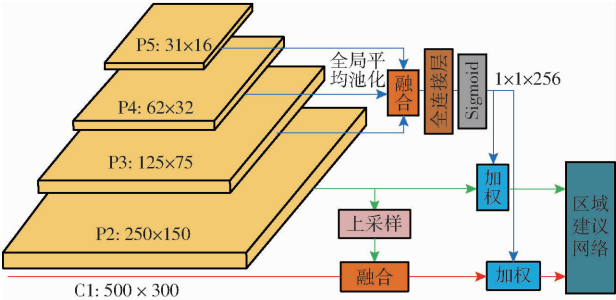


图 3 修正后的加权区域建议网络

Fig. 3 Revised weighted-regional proposal network

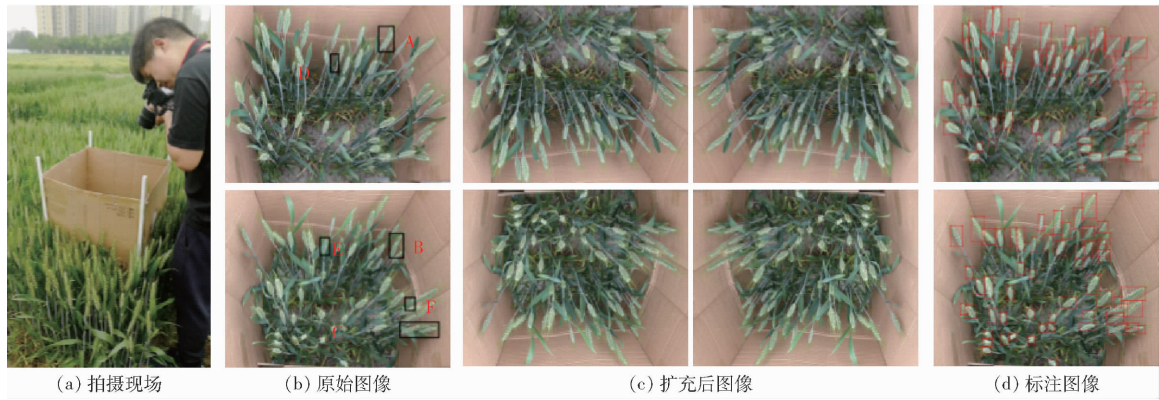


图 4 部分麦穗图像及其标注示例

Fig. 4 Some images of wheat spikes and examples of their labeling

本文从获取的 140 幅小麦灌浆期图像中随机选取 120 幅图像作为训练集,余下 20 幅作为测试集。由于实验室设备计算能力有限,将原始图像压缩为 1 000 像素 $\times$ 600 像素。为了提高最终训练模型的稳定性和鲁棒性,对训练集进行了数据扩充。数据集扩充是在保持图像数据标签不变的前提下对原始图像进行变换,从而增加样本的多样性。本文主要采用水平翻转和镜像翻转的方式将训练集扩充至 480 幅。使用图像标注工具 Labellmg 绘制图像中目标的矩形边框,麦穗的所有像素都在包围框的范围内。部分小麦数据集如图 4 所示,图 4d 中红色矩形框是对图像中每个目标麦穗的标注。由图 4 可见,所拍图像中既有麦穗粘连(如图 4b 中标注的 A、B、C 区域),又有叶片遮挡麦穗的情形(如图 4b 中标注的 D、E、F 区域)。

第 2 组公开的 ACID 麦穗图像数据集包含嫩芽期、灌浆期及成熟初期 3 个时期的小麦图像,共 1 398 幅。该数据集中包含训练集图像 1 374 幅,其中

3 模型训练

3.1 数据集

为了验证本文模型的性能,使用两组数据集进行了训练和测试,一组是实地拍摄的麦穗图像,一组是诺丁汉大学计算机视觉实验室公开发布的带有标注的 ACID 麦穗图像数据集。

实地采集的麦穗图像拍摄于安徽省合肥市庐阳区安徽省农业科学院实验基地。于 2019 年 5 月 25 日在晴朗无风的天气条件下,使用佳能(Canon)EOS 80D 型手持数码相机,采用自然曝光模式以多种倾斜角度对麦穗进行拍摄(图像分辨率为 3 984 像素 $\times$ 2 656 像素)。为便于图像获取,采集时使用硬纸板围成一个长、宽、高分别为 0.5、0.5、1.2 m 的长方体框,将目标区域与外界小麦隔离开(图 4a);同时为了排除光照因素的影响,在多个不同时间段进行了图像采集。

嫩芽期 108 幅、灌浆期 1 034 幅、成熟期 232 幅;测试集图像 24 幅,其中嫩芽期 6 幅、灌浆期 12 幅、成熟期 6 幅。不同时期的小麦图像示例如图 5 所示。原始数据集中的图像分辨率为 1 024 像素 $\times$ 1 024 像素,受计算能力的限制,将其降采样到 512 像素 $\times$ 512 像素。实验中将不同物候时期的训练集图像放在一起进行模型训练,然后对不同候物时期的测试图像分别进行测试。

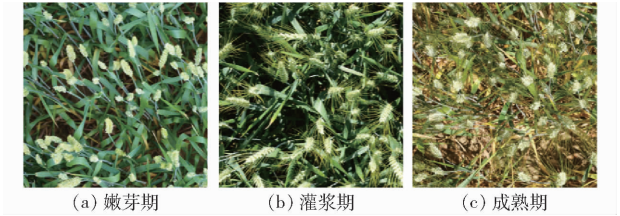


图 5 ACID 不同时期小麦图像

Fig. 5 ACID images of wheat in different periods

3.2 参数设置

实验是在 Linux 操作系统上使用 Tensorflow 1.10

框架和 Python 3.5.6 实现的,硬件系统为 NVIDIA GEFORCE RTX2070 GPU、Intel i7-8700 处理器和 16 GB RAM。为了得到更好的训练效果,以 Faster R-CNN 为基础网络,采用 VOC2007 预训练模型中的权重参数。网络中关于通道注意力机制、空间注意力机制及加权区域建议网络等的权值都由网络自身学习而确定,通过反向传播不断调整权值,从而学习到合适的权值以使网络达到最优性能。其他训练参数如表 1 所示。

表 1 模型参数设置
Tab.1 Setting of model parameters

参数	数值
卷积核尺寸	3×3
池化层卷积核尺寸	2×2
滑动窗口尺寸	16×16
学习率	0.001
批次	1
权重衰减率	0.000 1
梯度更新权值	0.9
最大迭代次数	30 000

3.3 性能指标

本文使用的卷积神经网络的输出是一个包围框列表及其分类。理想情况下,列表应该包含图像中的所有目标,但实际很难做到^[25]。本文目标是使测试图像中检测框的数量和位置与原始图像中手动标记框的数量和位置更加一致^[26]。对于每幅图像,检测模型返回一组置信度介于 0 和 1 之间的回归框,并利用常用指标交并比 (Interest of union, IoU) 对候选框进行筛选。

通常情况下,将 IoU 为 0.7 作为标注为前景的阈值(如果 IoU 大于 0.7,则认为检测到的对象为前景),将 IoU 为 0.3 作为标记为背景的阈值(如果 IoU 小于 0.3,则将检测到的对象划分为背景);IoU 在 0.3~0.7 之间的候选框选择丢弃^[27]。因此,在图像中检测到的前景或背景可以产生以下 4 种可能的结果:被正确分类的前景;被分类器分类为前景,但实际是背景;被正确分类为背景;被分类器分类为背景,但实际是前景。为了量化实验结果,使用精确度 (Precision)、召回率 (Recall)、平均精度 (Average precision, AP) 和 $P-R$ (Precision-recall) 曲线等指标,对建立的小麦穗数检测模型进行性能评价^[28-29]。精确度又称查准率,表示在所有预测出的正样本中,真正的正样本所占的比例;召回率又称查全率,表示在所有正样本中,正确预测出的正样本所占的比例; $P-R$ 曲线是由精确度和召回率所构成的曲线;平均精度是由 $P-R$ 曲线和坐标轴包围区域

的面积,该指标综合考量了精确度和召回率的影响,反映了模型的性能。

3.4 消融实验设置

本文主要在 Faster R-CNN 模型的基础上进行了两方面的改进:一是将其特征提取网络换成特征金字塔网络 (FPN),并在 FPN 的基础上引入了通道注意力机制 (CAM) 和空间注意力机制 (SAM),二是对原始网络中的 RPN 设计了加权模块,提出加权区域建议网络 (WRPN)。为了定量分析不同模型对麦穗检测结果的影响,分别利用 3.1 节介绍的两种数据集对以下 6 种不同组合模型进行实验,并与 Faster R-CNN 和 YOLO v4 进行比较。

FPN-RPN:将 Faster R-CNN 中的特征提取模块换成特征金字塔网络,将底层特征图和高层特征图中的信息累加,但多次的上采样和下采样操作使得深层网络对目标的定位信息存在误差。

CFPN-RPN:在 FPN-RPN 模型的基础上引入通道注意力机制,对不同通道赋予不同权重以增强有意义通道的权重,减少无意义或噪声通道的影响,提高编码器对麦穗特征提取的性能。该实验结果与 FPN-RPN 对比可验证通道注意力机制的有效性。

SFPN-RPN:在 FPN-RPN 模型的基础上引入空间注意力机制,提升解码器对特征图空间信息的利用能力,从而增强解码器生成特征图的空间位置信息的准确性。与 FPN-RPN 相比,该实验结果可验证空间注意力机制的有效性。

CSFPN-RPN:将通道注意力机制和空间注意力机制同时融入 FPN 的编码器和解码器中,使网络模型具有更强的语义信息提取能力和位置信息保留能力。将该实验结果与 CFPN-RPN 及 SFPN-RPN 对比可以体现出通道注意力机制和空间注意力机制联合后模型性能的优越性。

FPN-WRPN:对 FPN-RPN 模型中的 RPN 进行修改,将空间分辨率较小的特征图与空间分辨率较大的特征图加权融合,以对麦穗目标较小的特征生成更精确的待检测框。与 FPN-RPN 的实验结果相比,可体现 WRPN 的性能。

CSFPN-WRPN:将通道注意力机制和空间注意力机制同时融入 FPN 的编码器和解码器中,并在 FPN 中引入加权机制,以同时提升对有遮挡的麦穗和小目标麦穗的检测效果。对比其他实验结果,可体现综合改进的性能。

4 实验结果与分析

首先利用实地采集的麦穗图像数据集进行训练和测试。对于每幅测试图像,卷积神经网络返回检

测到的麦穗位置及其相应的类别得分。根据 3.3 节中定义的检测精确度、召回率和平均精度等指标来衡量模型的性能。

为鉴别不同特征提取网络对麦穗识别精度的影响,首先分别比较了 Faster R - CNN、FPN - RPN 以及本文提出的 CSFPN - RPN 这 3 种方法的检测性能,实验结果如表 2 所示。Faster R - CNN 在不断下采样的过程中发生了信息丢失,因此对麦穗的检测效果并不理想。FPN - RPN 结合上下文信息,获得信息更加丰富的特征图,相对于 Faster R - CNN 来说显著提升了检测效果,但其精确度只达到了 66.38%。本文建立的 CSFPN - RPN 利用通道和空间注意力机制有选择地增强图像的语义信息和位置信息,滤除无用的噪声信息,测试结果明显优于 FPN - RPN 和 Faster R - CNN,其检测精确度为 67.70%,召回率为 90.19%,平均精度为 85.98%。考虑到各种光照条件、阴影效果、自遮挡等与野外成像相关的影响,测试结果较优。为了量化分析 CSFPN - RPN 中引入的通道注意力机制和空间注意力机制对网络性能的影响,分别对 CFPN - RPN(FPN - RPN 中引

入通道注意力机制 CAM) 和 SFPN - RPN(FPN - RPN 中引入空间注意力机制 SAM) 进行了实验,检测结果如表 2 所示。由该实验结果可知,单独引入通道注意力机制或空间注意力机制,模型的性能有一定的提升;而将两者结合起来,分别添加到 FPN - RPN 的编码区和解码区,所建立的 CSFPN - RPN 的检测性能更佳。

图 6 为利用不同特征提取方法得到的部分结果。由于 Faster R - CNN 所利用的特征不足,所以只将大部分独立且无遮挡的麦穗检测出来,如图 6a 所示。FPN 将高层的强语义信息与低层的强位置信息相融合,利用全局信息优化网络性能,检测到了更多的目标,如图 6b 中红色椭圆框所示(FPN - RPN 能够正确检测出,Faster R - CNN 没有检测出)。但其在融合过程中混入了大量无意义的噪声信息,从而导致漏检(如未检测出图 6c 中粉色菱形框所示的麦穗)和误判(如图 6b 最下面一幅图中黑色椭圆框所示)现象,尤其是对有遮挡麦穗的检测效果依然不理想。而本文建立的 CSFPN 有选择地增强了高层特征的强语义信息和低层特征的强位置信息,有效消除了噪声信息的干扰,显著提高了网络对有遮挡麦穗的鉴别能力,由图 6c 可以看出,利用 CSFPN 作为特征提取网络对遮挡麦穗的检测效果有显著提升。

以上在 Faster R - CNN 框架下,基于改进的 CSFPN 特征提取方法比原始的 FPN 对遮挡麦穗的检测效果更加理想,但仍很难成功检测出图像中的较小麦穗。这是由于原网络中的 RPN 对于输入图像中的较小目标很难生成精确的待检测框。为此,本文提出 WRPN,对原始 RPN 输入进行加权处理,将信息更加全面的特征图送入该模块。为了验证该

表 2 不同模型的检测结果比较

Tab.2 Comparison of detection results of different models

模型	精确度	召回率	平均精度
Faster R - CNN	49.55	75.29	71.34
FPN - RPN	66.38	88.03	83.29
CFPN - RPN	66.63	88.79	85.02
SFPN - RPN	67.59	88.63	84.20
CSFPN - RPN	67.70	90.19	85.98
FPN - WRPN	79.86	87.39	84.82
CSFPN - WRPN	80.53	87.12	88.53
YOLO v4	66.50	77.82	82.34

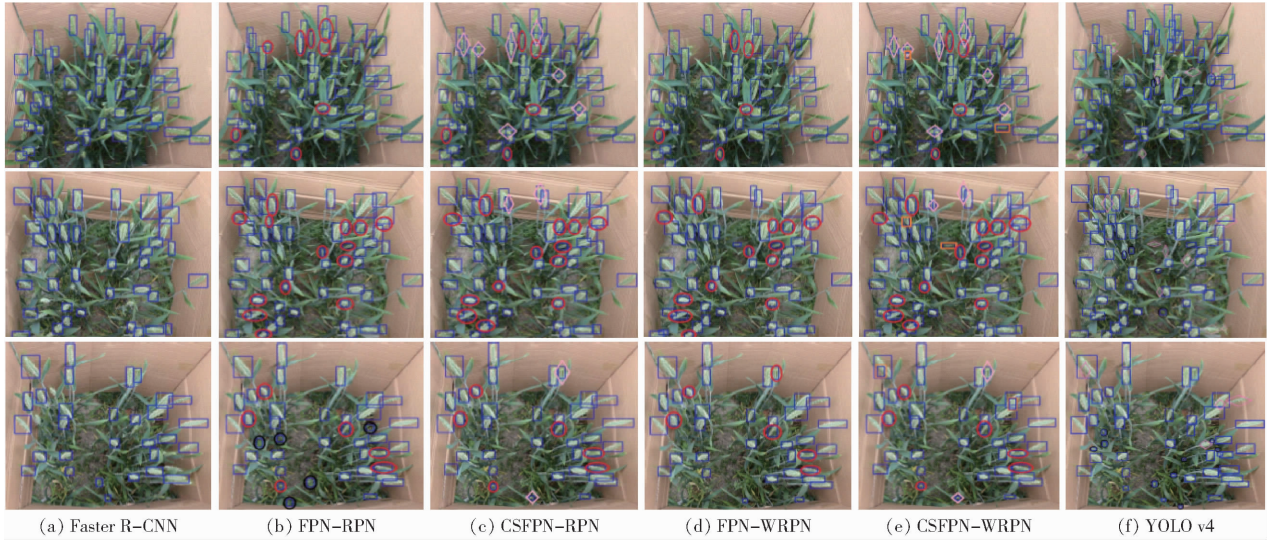


图 6 不同模型的测试结果

Fig.6 Test results with different models

方法的有效性,在相同的数据集上进行实验,分别在 FPN 和 CSFPN 的基础上添加 WRPN 网络,并与近期提出的 YOLO v4 进行了比较。不同网络的精确度、召回率和平均精度如表 2 所示,部分测试结果如图 6 所示。由实验结果可知,在 FPN 的基础上添加 WRPN 网络,尽管召回率和平均精度没有明显变化,但精确度明显提高;而在 CSFPN 的基础上添加 WRPN 网络,在提高精确度的同时,也明显提高了平

均精度,并且提升了对较小目标麦穗的检测效果(图 6e 所示,橘色框区域为其他方法未检测出的较小麦穗)。为了更直观地显示出不同模型在麦穗检测上的精确度和召回率,给出了以上不同网络模型的 $P-R$ 曲线,如图 7 所示。由图 7 可见,基于 CSFPN 提取特征与 WRPN 生成待检测框相结合的模型的 $P-R$ 曲线与坐标轴所围区域的面积大于其他模型,这也体现了本文方法的优越性。

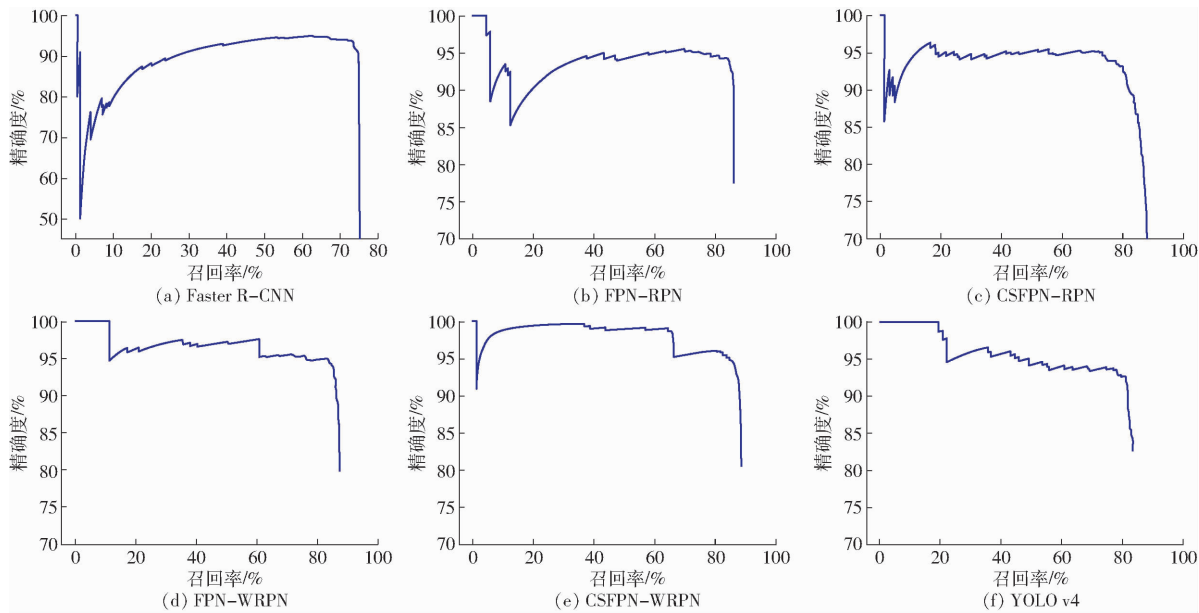


图 7 不同卷积神经网络模型下的 $P-R$ 曲线

Fig. 7 $P-R$ curves with different convolutional neural network models

利用同组模型对第 2 组公开的 ACID 麦穗图像数据集进行模型训练和测试。训练时将不同物候时期的训练集图像放在一起进行训练,测试时对不同物候时期的测试图像分类考虑。不同模型对不同物候时期麦穗图像的检测精确度如图 8 所示。由图 8 可以看出,处于灌浆期的麦穗检测精确度是 3 个时期中最高的,这是因为训练集中灌浆期的小麦图像最多,并且灌浆期的麦穗较大,编码器对其进行特征提取时更容易获取麦穗的位置信息和语义信息。而对处于嫩芽期的麦穗,训练集中图像数量有限且麦穗较小,同时麦穗颜色与嫩绿叶片颜色相近,因此对

其进行特征提取时,获取较小麦穗目标的精确位置信息及语义信息有一定的难度。与嫩芽期麦穗相似,成熟期的麦穗与同期的黄色叶片颜色一致,这给精确提取麦穗的信息增加了难度,因此精确度也略低。从该检测结果中可以看出,本文方法对复杂大田环境下的麦穗检测同样有效,这进一步验证了其优越性。

为了进一步验证麦穗遮挡对模型性能的影响,对不同遮挡程度的图像分别进行了测试。首先将实地拍摄数据集中选出的 20 幅测试图像进行划分,其中 8 幅为麦穗轻微遮挡(少量麦穗粘连和叶片遮挡),余下 12 幅为麦穗严重遮挡(大量麦穗粘连和叶片遮挡),如图 9 所示。然后分别利用 FPN-RPN

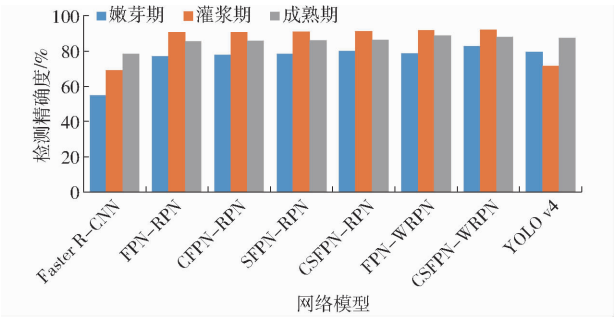


图 8 不同网络模型下 ACID 数据集检测结果比较

Fig. 8 Comparison of ACID dataset detection results of different models

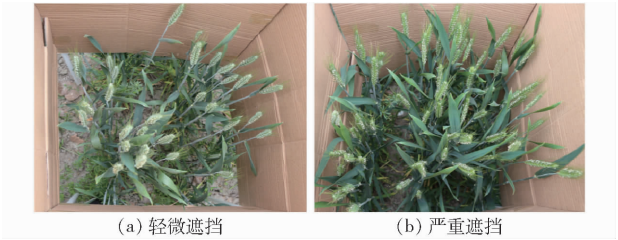


图 9 不同遮挡程度的麦穗图像

Fig. 9 Wheat images with different occlusions

和 CSFPN - WRPN 等模型进行测试,测试结果如表 3 所示。

表 3 不同遮挡程度下的检测结果

Tab.3 Detection results of wheat with different

模型	occlusions						%
	轻微遮挡			严重遮挡			
	精确度	召回率	平均精度	精确度	召回率	平均精度	
FPN - RPN	69.42	89.76	85.46	64.86	84.55	82.70	
CFPN - RPN	69.98	88.98	85.73	65.47	86.31	82.86	
SFPN - RPN	70.27	88.75	85.62	66.42	85.26	82.79	
CSFPN - RPN	73.90	89.19	86.33	67.01	83.37	82.95	
FPN - WRPN	80.39	90.42	86.82	74.91	84.85	83.54	
CSFPN - WRPN	82.53	91.62	90.90	75.44	85.14	84.87	

由表 3 可知,对于同一算法,轻微遮挡状态下的麦穗检测性能均明显优于严重遮挡情形。在 FPN - RPN 基础上单独引入通道注意力机制或空间注意力机制,对于两种不同遮挡情形下麦穗检测的精确度和平均精度都有一定的提升,但效果不明显。将两者融合后组成的 CSFPN - RPN 对于轻微遮挡和严重遮挡情形下的麦穗检测精确度分别提高了 4.48、2.15 个百分点;而改进的 FPN - WRPN 对于这两种情形的精确度分别提高了 10.97、10.05 个百分点;但 CSFPN - RPN 和 FPN - WRPN 的召回率和

平均精度变化都不明显。CSFPN - WRPN 的目标检测性能明显强于 FPN - RPN。对于轻微遮挡情形,CSFPN - WRPN 的检测精确度和平均精度分别提升了 13.11、5.44 个百分点,优势比较明显;而严重遮挡状态下精确度提升较明显,达到 10.58 个百分点,但召回率和平均精度仅提升了 0.59、2.17 个百分点,还需要进一步优化。

5 结论

(1)针对有遮挡麦穗难以检测的问题,提出了一种基于注意力机制特征金字塔网络的麦穗检测方法。在 Faster R - CNN 基础上,将原始特征提取网络替换为特征金字塔网络,并引入通道注意力机制和空间注意力机制。改进后的 CSFPN - RPN 算法有效提高了对遮挡麦穗的检测效果,使得麦穗识别精确度有了明显提升。

(2)在改进的特征提取网络基础上,设计了一种加权区域建议网络,为难以检测的较小麦穗生成更精确的检测框。CSFPN - WRPN 对于实地采集的麦穗图像的检测精确度、召回率和平均精度分别达到了 80.53%、87.12%、88.53%。同时,公开 ACID 麦穗图像数据集上不同物候时期的麦穗检测结果进一步验证了本文方法的有效性。

参 考 文 献

[1] SHI Y, WANG N, TAYLOR R K, et al. Automatic corn plant location and spacing measurement using laser line-scan technique [J]. Precision Agriculture, 2013, 14(5): 1 - 17.

[2] Wheat in the world CGIAR[EB/OL]. <https://wheat.org/wheat-in-the-world/>.

[3] JIN X, LIU S, BARET F, et al. Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery [J]. Remote Sensing of Environment, 2017, 198: 105 - 114.

[4] XIONG H, CAO Z, LU H, et al. TasselNetv2: in-field counting of wheat spikes with context-augmented local regression networks[J]. Plant Methods, 2019, 15(1): 1 - 14.

[5] 刘涛,孙成明,王力坚,等. 基于图像处理技术的大田麦穗计数[J/OL]. 农业机械学报, 2014, 45(2): 282 - 290. LIU Tao, SUN Chengming, WANG Lijian, et al. In-field wheatear counting based on image processing technology[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(2): 282 - 290. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20140247&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2014.02.047. (in Chinese)

[6] 李毅念,杜世伟,姚敏,等. 基于小麦群体图像的田间麦穗计数及产量预测方法[J]. 农业工程学报, 2018, 34(21): 185 - 194. LI Yinian, DU Shiwei, YAO Min, et al. Method for wheatear counting and yield predicting based on image of wheatear population in field[J]. Transactions of the CSAE, 2018, 34(21): 185 - 194. (in Chinese)

[7] 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): 22 - 33.

[8] MOHANTY S P, HUGHES D P, SALATHE M. Using deep learning for image-based plant disease detection[J]. Frontiers in Plant Science, 2016,7: 1419 - 1428.

[9] LU H, CAO Z, XIAO Y, et al. TasselNet: counting maize tassels in the wild via local counts regression network[J]. Plant Methods, 2017, 13(1): 1 - 17.

[10] HASAN M M, CHOPIN J P, LAGA H, et al. Detection and analysis of wheat spikes using convolutional neural networks[J]. Plant Methods, 2018, 14(1): 1 - 13.

[11] SADEGHI-TEHRAN P, VIRLET N, AMPE E M, et al. DeepCount: in-field automatic quantification of wheat spikes using simple linear iterative clustering and deep convolutional neural networks[J]. Frontiers in Plant Science, 2019,10: 1176 - 1191.

[12] 张领先,陈运强,李云霞,等. 基于卷积神经网络的冬小麦麦穗检测计数系统[J/OL]. 农业机械学报,2019, 50(3): 144 - 150.

ZHANG Lingxian, CHEN Yunqiang, LI Yunxia, et al. Detection and counting system for winter wheat ears based on convolutional neural network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(3): 144 – 150. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190315&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2019.03.015. (in Chinese)

[13] MISRA T, ARORA A, MARWAHA S, et al. SpikeSegNet—a deep learning approach utilizing encoder-decoder network with hourglass for spike segmentation and counting in wheat plant from visual imaging[J]. Plant Methods, 2020, 16(1): 1 – 20.

[14] REN S, HE K, 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.

[15] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C] // IEEE Computer Society, 2016: 106 – 115.

[16] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLO v4: optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770 – 778.

[18] ZHONG Z, JIN L, HUANG S. DeepText: a new approach for text proposal generation and text detection in natural images[C] // ICASSP 2017 – 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017: 1208 – 1212.

[19] NEUBECK A, GOOL L J V. Efficient non-maximum suppression[C] // IEEE Computer Society, 2006: 850 – 855.

[20] CHAUDHARI S, POLATKAN G, RAMANATH R, et al. An attentive survey of attention models[J]. ACM Transactions on Intelligent Systems and Technology, 2021, 12(5): 1 – 32.

[21] LI H, XIONG P, AN J, et al. Pyramid attention network for semantic segmentation[J]. arXiv:1805.10180.

[22] GHIASI G, LIN T Y, PANG R, et al. NAS – FPN: learning scalable feature pyramid architecture for object detection[C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 7029 – 7038.

[23] GUO C, FAN B, ZHANG Q, et al. AugFPN: improving multi-scale feature learning for object detection[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 12595 – 12604.

[24] LIU Y, CEN C, CHE Y, et al. Detection of maize tassels from UAV RGB imagery with Faster – RCNN[J]. Remote Sensing, 2020, 12(2): 338 – 351.

[25] HE K, GKIOXARI G, DOLLAR P, et al. Mask R – CNN[C] // IEEE International Conference on Computer Vision (ICCV) – Venice, 2017: 2980 – 2988.

[26] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [J]. arXiv preprint arXiv: 1704.04861.

[27] HUANG G, LIU Z, MAATEN L V D, et al. Densely connected convolutional networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700 – 4708.

[28] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132 – 7141.

[29] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431 – 3440.

(上接第 225 页)

[22] CHEN L S, SUN Y Y, WANG K, et al. Rapid diagnosis of nitrogen nutrition status in rice based on static scanning and extraction of leaf and sheath characteristics[J]. International Journal of Agricultural & Biological Engineering, 2017, 10(3): 158 – 164.

[23] MIZUSHIMA A, LU R. An image segmentation method for apple sorting and grading using support vector machine and Otsu’s method[J]. Computers and Electronics in Agriculture, 2013, 94: 29 – 37.

[24] ZHANG Wengang, WU Chongzhi, ZHONG Haiyi, et al. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization[J]. Geoscience Frontiers, 2021, 12(1): 469 – 477.

[25] XIA Yufei, LIU Chuanzhe, LI Yuying, et al. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring[J]. Expert Systems With Applications, 2017, 78: 225 – 241.

[26] 吴成东, 卢紫薇, 于晓升. 基于加权随机森林的图像超分辨率算法研究[J]. 控制与决策, 2019, 34(10): 2243 – 2248.

WU Chengdong, LU Ziwei, YU Xiaosheng. Research on image super-resolution algorithm based on weighted random forest[J]. Control and Decision, 2019, 34(10): 2243 – 2248. (in Chinese)

[27] 马晓君, 董碧滢, 王常欣. 一种基于 PSO 优化加权随机森林算法的上市公司信用评级模型设计[J]. 数量经济技术经济研究, 2019, 36(12): 165 – 182.

MA Xiaojun, DONG Biying, WANG Changxin. Design of a credit rating model of listed companies based on PSO optimized weighted random forest algorithm[J]. The Journal of Quantitative & Technical Economics, Technology and Economy, 2019, 36(12): 165 – 182. (in Chinese)