

朱德利, 林智健. 基于 MF-SSD 卷积神经网络的玉米穗丝目标检测方法 [J]. 华南农业大学学报, 2020, 41(6): 109-118.  
ZHU Deli, LIN Zhijian. A corn silk detection method based on MF-SSD convolutional neural network[J]. Journal of South China Agricultural University, 2020, 41(6): 109-118.

# 基于 MF-SSD 卷积神经网络的玉米穗丝目标检测方法

朱德利<sup>1,2</sup>, 林智健<sup>1</sup>

(1 重庆师范大学 计算机与信息科学学院, 重庆 401331; 2 重庆市数字农业服务工程技术研究中心, 重庆 401331)

**摘要:**【目的】玉米穗丝是玉米的授粉器官, 生长发育状况会影响玉米的产量。为了在玉米生长状态监测和产量预测工作中实时准确识别玉米穗丝, 提出一种基于多特征融合 SSD (MF-SSD) 卷积神经网络的玉米穗丝检测模型。【方法】基于特征图对玉米穗丝进行检测, 在 VGG16-SSD 的基础上, 用 MobileNet 替换特征提取器, 加入多层特征融合结构, 得到 MF-SSD 网络模型; 通过网络优化调整, 试验了 MF-SSD-cut-3、MF-SSD 和 MF-SSD-add-3 共 3 种网络结构, 优选出检测性能最好的网络结构用于玉米穗丝检测。基于玉米穗丝图像数据集, 应用 0~180°随机旋转原始图像和水平翻转、平移原始图像 2 种数据增广技术提升模型训练效果。对是否使用二次训练策略和是否使用 Focal loss 解决样本不平衡问题进行了试验, 并对比分析 Loss 的下降过程。【结果】通过加入多层特征融合结构对 SSD 模型改进后能够提高网络的检测能力, 提升识别速度。与 VGG16-SSD 相比, MF-SSD 在交并比指标方面的平均精度提高 7.2%, 对玉米穗丝小目标检测的平均召回率提高 19.6%, 检测速度最高能提升 18.7%。在存储空间和运行时间有较高要求的嵌入式环境下, MF-SSD-cut-3 模型在满足检测效果的前提下, 以较小的空间代价获得了相对较短的运行时间; 在不考虑空间和时间因素的情况下, MF-SSD 模型获得更好的检测效果。二次训练策略提高了网络的收敛速度和模型的稳定性; Focal loss 有效解决了 SSD 算法中正负样本数量不平衡问题, 使网络模型的训练更容易收敛。【结论】MF-SSD 模型对小目标的检测能力能满足农业生产中对玉米穗丝的实时检测需要, 可以用于玉米生长状态的自动监控和产量的精准预测。

**关键词:** 玉米穗丝; 目标检测; 卷积神经网络; 特征融合; MF-SSD  
**中图分类号:** TP391      **文献标志码:** A      **文章编号:** 1001-411X(2020)06-0109-10

## A corn silk detection method based on MF-SSD convolutional neural network

ZHU Deli<sup>1,2</sup>, LIN Zhijian<sup>1</sup>

(1 Department of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China;  
2 Research Center of Chongqing Digital Agricultural Service Engineering Technology, Chongqing 401331, China)

**Abstract:** 【Objective】Corn silk is the pollination organ of maize, its growth state will affect the yield of corn. In order to identify the corn silk in real-time and accurately in corn growth state monitoring and yield prediction, a corn silk detection model based on multi-feature fusion SSD (MF-SSD) convolutional neural network was proposed. 【Method】Corn silk detection was based on feature images. The MF-SSD network model was

modified from VGG16-SSD through replacing feature extractor by MobileNet and integrating multi-layer feature fusion structure. By optimizing and adjusting the network, three kinds of MF-SSD with different network structures (MF-SSD-cut-3, MF-SSD and MF-SSD-add-3) were tested, and the structure with the best detection performance was selected for corn silk detection. Based on the image data set of corn silk, two kinds of data augmentation techniques (randomly rotating original image from 0 to 180°, horizontally rolling over or translating original image) were applied to improve the training effect of the model. Whether using secondary training strategy and Focal loss to solve sample number imbalance was investigated, and the decrease process of Loss was compared and analyzed. 【Result】 The improved SSD model added with multi-layer feature fusion structure could improve the detection ability and recognition speed of network. Compared with VGG16-SSD, the average accuracy of intersection over union increased by 7.2%, the average recall of small target detection of corn silk increased by 19.6%, and the detection speed increased by 18.7%. In the embedded environment having high demand for storage space and run time, MF-SSD-cut-3 obtained shorter run time with smaller storage space in the premise of satisfying detection effect. MF-SSD obtained better detection effect in the condition of taking no account of storage space and run time. The secondary training strategy improved the network convergence speed and model stability. Focal loss effectively solved number imbalance problem of positive and negative samples, and made the training of network model more convergent. 【Conclusion】 The detection effect of the proposed MF-SSD model for small targets can meet the needs of real-time detection of corn silk in agricultural production. It can be used for automatic monitoring of corn growth state and yield prediction.

**Key words:** corn silk; object detection; convolutional neural network; feature fusion; MF-SSD

玉米是中国主产农作物之一，在我国国民经济中具有举足轻重的作用。现阶段对玉米生长状态的监控和产量的预测主要还是人工方式<sup>[1]</sup>，主要原因是玉米生长状态特征不容易自动提取，果穗性状在生长周期内变化较大。玉米是雌雄同体异花，雌花在植株的叶腋位置，雄花在植株的顶部位置<sup>[2]</sup>。玉米穗丝作为授粉器官，不同生长阶段的性状对玉米生长状况的监控<sup>[3]</sup>和果穗产量的预测有重要的参考意义。监控玉米穗丝的重要一步即是对其所处的位置进行检测。

不同生长阶段的玉米穗丝的特性和性状有很大不同，不具有通用特征，用传统机器视觉方法进行目标提取和定位的效果不佳。近年来，大量大型图像数据建设和硬件计算力的增强促使卷积神经网络在计算机视觉的各个方面获得突破；王临铭等<sup>[4]</sup>提取甘肃大麦病斑的颜色和纹理特征，以特征向量为输入向量构造大麦病害神经网络分类器模型，对甘肃大麦病害的整体识别正确率达到 86.7%；王璨等<sup>[5]</sup>提出了基于卷积神经网络提取多尺度分层特征的玉米识别方法。此外，利用神经网络可对农作物的生长状态进行科学分析；周云成等<sup>[6]</sup>提出一种基于面向通道分组卷积网络的番茄主要器官实时识别网络模型；刘永波等<sup>[7]</sup>提出一种基于深度卷积神

经网络的在自然环境条件下识别玉米病害的方法；Koirala 等<sup>[8]</sup>以芒果为研究对象，使用深度学习方法对研究目标进行检测并预测产量；Tian 等<sup>[9]</sup>通过改进的 YOLO-V3 模型对果园中不同生长阶段苹果的检测进行了研究。

目前目标检测算法一般分为 2 种，基于候选区域的两步目标检测算法（如 Faster R-CNN 系列<sup>[10]</sup>）和基于回归的单步目标检测算法（如 YOLO、SSD<sup>[11]</sup>）。两步目标检测算法的精度高但检测速度稍慢，单步目标检测算法的检测速度快但精度略低。其中，SSD 同时借鉴了 YOLO 的回归思想和 Faster R-CNN 的锚（Anchor）机制，检测性能相对较好，效果也超越了两者。但是 SSD 的模型参数过多，运行时内存占用量过大，在显存容量较小的 GPU 设备上无法运行<sup>[12]</sup>。

针对农业生产实践中玉米生长状况监控和产量预测中对玉米穗丝实时视觉识别的需求及问题，本研究通过改进深度学习中的 SSD 目标检测算法以完成玉米穗丝的位置确定，提出了基于多特征融合 SSD（Multi-feature fusion SSD, MF-SSD）的玉米穗丝检测模型。该模型有较高的小目标识别性能，在识别准确度和性能消耗方面均可满足实际要求。

1 SSD 网络模型及其改进

1.1 SSD 网络模型

SSD 属于单步目标检测算法, 在多尺度特征图上以回归的方式直接得到目标的类别和位

置。如图 1 所示, 传统的 SSD 网络模型架构包括 2 个部分, 基础网络采用 VGG16 用于目标特征提取, 在基础网络后是额外添加的特征层, 最后从 6 个不同尺度的特征图上提取特征为目标检测使用。

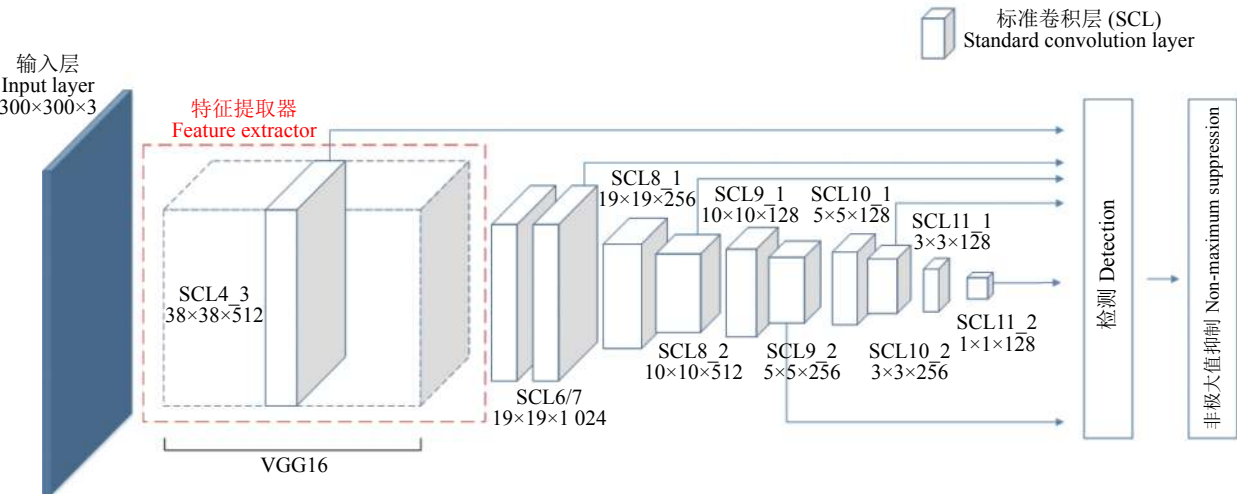


图 1 VGG16-SSD 的网络结构  
Fig. 1 VGG16-SSD network structure

传统的 VGG16-SSD 模型结构较长, 操作相对复杂, 不易训练。其中 VGG16 的参数过多, 模型较大, 在有限资源的工程应用设备上使用会造成困扰。另外, 在 VGG16-SSD 训练过程中, 小目标特征经过多层卷积后会发生特征消失现象, 甚至不会检测到目标。因此, 如何改善 SSD 网络参数规模大、小目标识别能力不佳等问题是本文研究的重点。

1.2 使用 MobileNet 作为特征提取器

VGG16 的权重参数约  $1.38 \times 10^8$  个<sup>[13]</sup>, 是为解决类似 ImageNet 这类大规模图像任务而设计的深度卷积神经网络。玉米穗丝的检测在类型和样本数量方面都远小于 ImageNet, 所以可使用比 VGG16 更轻量化的特征提取器。

Alawad 等<sup>[14]</sup> 提出的 MobileNet 卷积网络模型属于一种轻量级的卷积网络, 它基于深度可分离的卷积架构, 将常规标准模型结构分解成深度卷积与逐点卷积 2 类, 如图 2 所示。深度卷积作用于每个通道, 而逐点卷积用来组合各个通道的输出。通过卷积分解, 可以有效减少计算量, 加快运算速度, 减轻因过度拟合导致的训练问题。

深度可分离卷积较传统卷积所降低的参数比例可以定义为公式 (1):

$$\frac{D_k^2 MD_F^2 + MND_F^2}{D_k^2 MND_F^2} = \frac{1}{N} + \frac{1}{D_k^2}, \quad (1)$$

式中,  $D_k$  为卷积核的宽高,  $D_F$  为输出特征图的宽

高,  $M$  为输入的特征图数量,  $N$  为卷积核数量。其中传统卷积的计算量为  $D_k^2 MND_F^2$ , 深度可分离卷积的计算量为  $MND_F^2 + D_k^2 MD_F^2$ 。

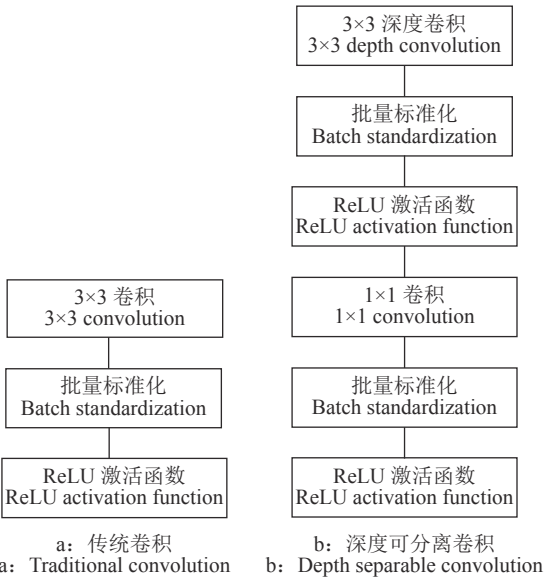


图 2 传统卷积和深度可分离卷积  
Fig. 2 Traditional convolution and depth separable convolution

本文使用 MobileNet 网络作为特征提取器替换 SSD 中的 VGG16, 能够减少整体网络的计算开销, 减少内存占用量, 改进后的网络结构如图 3 所示。MobileNet-SSD 模型网络结构与 VGG16-

SSD 的网络结构类似，都是抽取多个不同尺度的特征图进行目标检测。该模型结合 MobileNet 和 SSD

2 种类型网络的优势，保持了原有 SSD 网络结构，降低了参数模型大小。

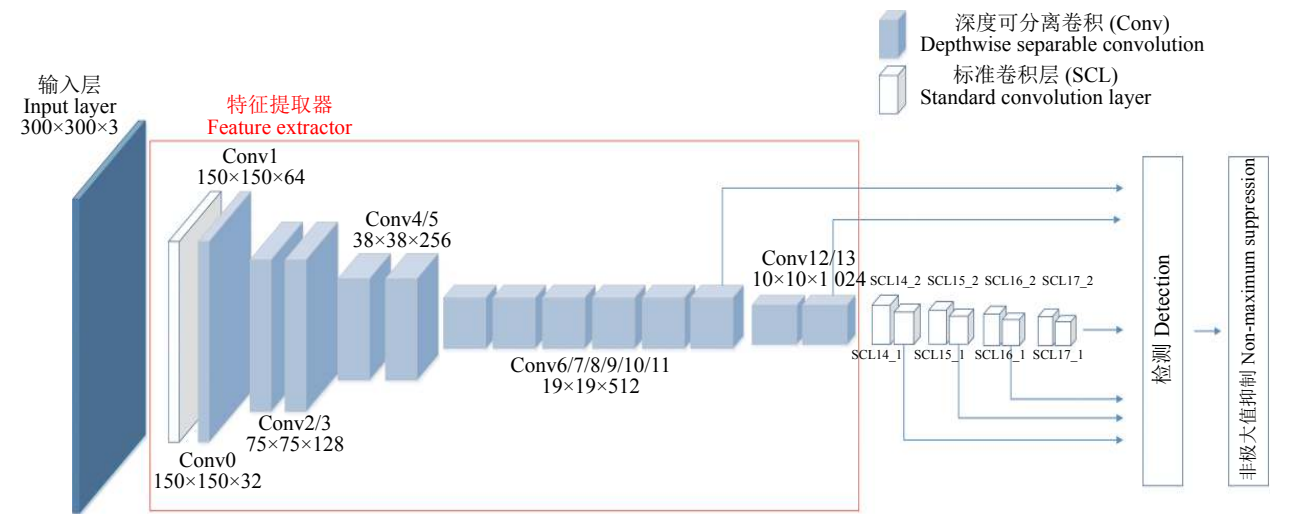


图 3 MobileNet-SSD 的网络结构

Fig. 3 MobileNet-SSD network structure

1.3 MF-SSD 的网络模型架构设计

MobileNet 替换 VGG16 后的网络 MobileNet-SSD 仍然存在小目标检测能力不佳的问题，原因是该网络不能充分利用局部的细节特征和全局的语义特征。随着网络层次的增加，卷积层提取特征的语义越来越强，SSD 却直接用不同语义的特征图进行预测。图 4a 为原始 SSD 结构，这种方式各个层之间没有联系。最底层的特征图分辨率高但是语义

值不够，导致不能被使用。SSD 只使用较上层的特征图用于目标检测，因此对小物体的检测性能不佳。

Lin 等<sup>[15]</sup>提出一种旨在提高准确率的特征提取器——特征金字塔网络 (Feature pyramid network, FPN)，如图 4b。FPN 提供了一条自顶向下的路径，从语义丰富的层构建高分辨率的层。把浅层的细节特征和高层的语义特征结合起来进行预测，使得算法的精度明显提升。

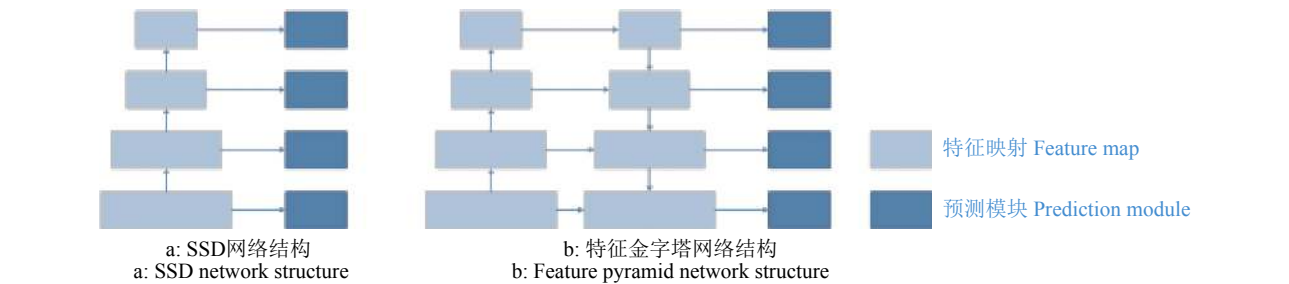


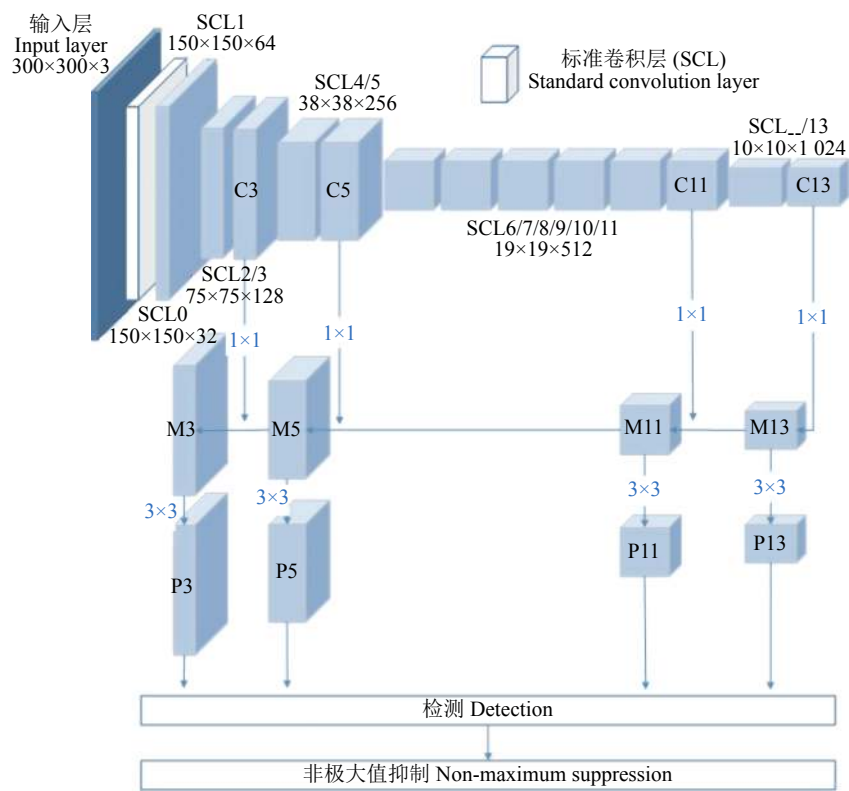
图 4 SSD 与特征金字塔网络结构对比

Fig. 4 Structure comparison of SSD and feature pyramid network

本文受 FPN 网络结构思想的启发，重新构建了 MobileNet-SSD 中的特征提取器，图 5 为重建的网络模型。该网络的前向过程实现了自底向上特征提取，在前向过程中，特征图大小在经过某些层后会改变，而经过某些层不会改变。将不改变特征图大小的卷积层归为一个阶段，将每个阶段的最后一层得到的特征图进行组合能够构成特征金字塔。

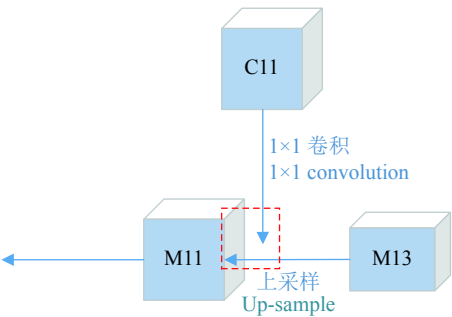
对 MobileNet-SSD 不同阶段的最后一层 C3、C5、C11、C13 构建一条自顶向下反向路径。自顶向下构建层的过程采用最近邻上采样法进行，从语义

丰富的层构建高分辨率的层，得到重建层的语义较强，但目标的位置不精确。为检测更加准确，本文通过添加横向连接将上采样的结果和自底向上生成的相同大小的特征图进行融合得到融合结果 M3、M5、M11、M13。在横向连接前对前一层特征图要使用 1×1 的卷积核，目的是改变卷积核个数，使其和后一层的特征图个数相同，如图 6。在融合后使用 3×3 的卷积核对每个融合结果进行卷积，消除上采样的混叠效应，最终生成的特征图结果是 P3、P5、P11、P13，然后使用 P3 到 P13 进行多尺度目标检测。



C3、C5、C11、C13 表示卷积层，M3、M5、M11、M13 表示融合结果，P3、P5、P11、P13 表示特征图结果  
C3, C5, C11 and C13 indicate convolution layers; M3, M5, M11 and M13 indicate fusion results; P3, P5, P11 and P13 indicate feature map results

图 5 MF-SSD 的网络结构  
Fig. 5 MF-SSD network structure



C11 表示卷积层，M11、M13 表示融合结果  
C11 indicates convolution layer; M11 and M13 indicate fusion results

图 6 横向连接过程  
Fig. 6 Horizontal connection process

该模型对网络结构的关键改进在于把多层特征融合到了特征图的提取层，故最终的模型本文称之为 MF-SSD 网络模型。相比于传统的 VGG16-SSD 网络模型，该网络模型减少了网络参数，能够提升检测速度；同时融合底层特征的高分辨率和高层特征的高语义信息，提高检测准确度，改善模型检测小目标的性能。

为获得更好的玉米穗丝检测网络，考虑到轻量级特征提取器参数较少，可以通过改变特征提取单元修改网络结构的特征提取能力<sup>[16]</sup>。本文以 MF-SSD 模型为基准，通过增加和减少特征提取层数获

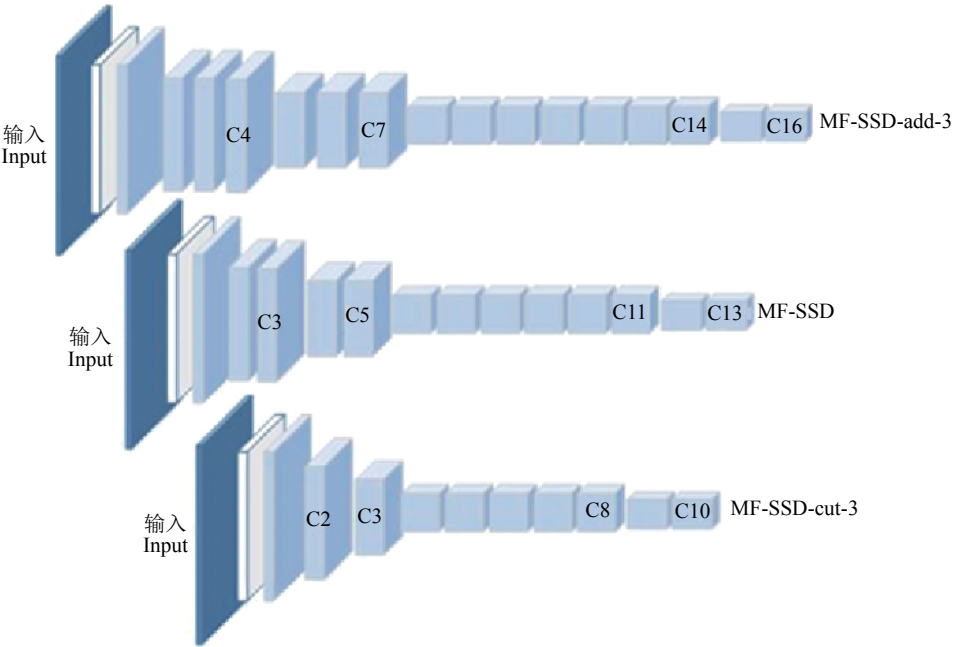
得了 MF-SSD 的另外 2 种变体结构，分别是 MF-SSD-add-3 和 MF-SSD-cut-3。其中，MF-SSD-add-3 是对 MF-SSD 模型的每个阶段增加 1 层深度可分离卷积，并选取 C4、C7、C14、C16 这 4 层进行特征融合，将最终生成特征金字塔 P4、P7、P14、P16 进行多尺度的目标检测；MF-SSD-cut-3 是对 MF-SSD 模型的每个阶段减少 1 层深度可分离卷积，并选取 C2、C3、C8、C10 这 4 层进行特征融合，将最终生成特征金字塔 P2、P3、P8、P10 进行多尺度的目标检测。图 7 是 3 种不同特征提取结构的模型说明。在本文的试验中，把玉米穗丝图像作为输入，通过对比 3 种不同网络结构的 MF-SSD 性能来筛选最优网络模型。

1.4 网络训练目标

MF-SSD 的训练同时对目标位置和种类进行回归，其目标损失函数是分类损失与定位损失之和，表达式如公式 (2) 所示。

$$L(x, c, l, g) = \frac{1}{n} [L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)], \quad (2)$$

式中， $n$  是与真实框相匹配的预测框个数； $L_{\text{conf}}(x, c)$  是分类损失，采用 Softmax 分类器； $L_{\text{loc}}(x, l, g)$  是定位损失，采用 Smooth L1 loss<sup>[17]</sup>； $x$  为预测框与不同类别的真实框的匹配结果； $c$  为预测物体框的



C2、C3、C4、C5、C7、C8、C10、C11、C13、C14、C16 表示卷积层  
C2, C3, C4, C5, C7, C8, C10, C11, C13, C14 and C16 indicate convolution layers

图 7 MF-SSD-add-3、MF-SSD 和 MF-SSD-cut-3 的特征提取结构

Fig. 7 Feature extraction structures of MF-SSD-add-3, MF-SSD and MF-SSD-cut-3

置信度;  $l$  为预测框的位置信息;  $g$  为真实框的位置信息;  $\alpha$  是 2 种损失的均衡系数, 一般设置为 1。

在训练过程中, 减小损失函数值可以确保在提升预测框分类置信度的同时提高预测框的位置准确度, 从而训练出性能较好的预测模型。

1.5 解决正负样本数量失衡问题

MF-SSD 在各检测层生成不同尺寸的预测框, 选取交并比 (Intersection over union, IoU) 大于 0.50 的预测框作为正样本, IoU 小于 0.50 的预测框作为负样本。大目标物体上 IoU 大于 0.50 的预测框多, 正负样本均衡; 小目标物体上 IoU 大于 0.50 的预测框少, 正负样本数量失衡, 训练效果不佳。IoU 是候选框与正确标记框 (Ground truth) 的交叠率, 即它们的交集与并集的比值, 如图 8 所示。

在 VGG16-SSD 中使用困难负例挖掘策略从困难样本中选取负样本, 实现了正负样本数量 1:3 的比例, 但却忽略了 Loss 较小的简单样本。本文认为训练一个有效的网络应该平衡简单样本和困难样本的权重, 不能完全舍弃简单样本对 Loss 的贡献。因此本文使用 Lin 等<sup>[17]</sup>提出的 Focal loss 同时解决正负样本不平衡以及简单样本与困难样本不平衡的问题。Focal loss (FL) 的表达式如公式 (3) 所示。

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log_e p_t, \tag{3}$$

式中,  $p_t$  是不同类别的分类概率;  $\alpha_t$  是  $[0, 1]$  区间的小数, 用于调节正负例比例, 正例为  $\alpha_t$ , 负例为  $1-\alpha_t$ ;  $1-p_t$  用于对简单样本进行权重抑制;  $\gamma$  用于平



黄色框: 正确的穗丝目标标记框; 红色框: 目标检测的预测框, 1: 简单正例, 2: 困难正例, 3: 困难负例, 4: 简单负例

Yellow box: Correct marker box of silk; Red box: Predictive box of target detection, 1: Simple positive sample, 2: Difficult positive sample, 3: Difficult negative sample, 4: Simple negative sample

图 8 4 种不同类型的样本

Fig. 8 Samples of four different types

滑权重抑制。对于简单样本,  $p_t$  较大, 权重  $1-p_t$  较小; 而对于困难样本,  $p_t$  较小, 权重  $1-p_t$  较大。  $\gamma$  和  $\alpha_t$  都是固定值, 本文采用  $\alpha_t=0.25, \gamma=2$ 。

Focal loss 函数可以通过减少简单样本的权重让简单样本对 Loss 做贡献, 但又不主导 Loss, 使得 MF-SSD 网络模型在训练时更专注于困难样本的分类。用数据集进行多次结果优化训练模型可不断提高模型的检测性能, 从而训练出性能较好的 MF-SSD 检测模型。

## 2 玉米穗丝目标检测网络的训练

### 2.1 玉米穗丝图像数据采集

现场采集并结合网络采集共搜集了 2 980 幅玉米穗丝图像作为模型训练和验证的数据集。其中, 现场采集设备使用的是 Nikon D7200, 使用 AF-S 尼克尔镜头, 焦距范围是 18~105 mm, 该设备使用 35.9 mm × 23.9 mm 的 CMOS 图像传感器, 有较好的图像采集质量。采集地点选择在重庆璧山种植玉米的农地。为避免所收集数据存在随机性、图片尺寸大小不一、重复率较高等问题, 数据输入网络前对图像进行了预处理, 去除了冗余图像样本以及目标特征不清等杂糅图像 300 幅, 最终获得有效现场图像共 2 200 幅, 包括吐丝期 650 幅, 灌浆期 700 幅, 成熟期 850 幅。网络采集图像 780 幅, 包括吐丝期 200 幅, 灌浆期 280 幅, 成熟期 300 幅。在上述玉米穗丝数据集中, 选择 2 400 幅作为训练集, 580 幅作为测试集。通过人工对每幅玉米穗丝图像进行标注, 圈选出穗丝的真实区域, 并标注区域的类别。为降低训练样本多样性不足导致网络过拟合的概率, 图像收集过程中采取了一些措施。考虑到温室环境光条件不同会造成成像结果差异, 分别搜集晴天和阴天的玉米穗丝图片; 考虑到玉米穗丝形态和遮挡情况不同, 对不同形态、不同成熟度的玉米穗丝从不同高度 (1.0 m、1.5 m、2.0 m), 距离 (0.50 m、0.75 m、1.00 m) 和拍摄方向等多个角度进行搜集, 以增加样本的多样性。

### 2.2 试验环境

试验软件环境操作系统为 Window 10, 深度学习软件框架为 TensorFlow。试验硬件环境 CPU 型号为 IntelCorei7-8750H, 内存为 16 G, GPU 型号为 NVIDIA GTX 1 070 (Max-Q Design)。为提高模型的训练效率, 本文使用 COCO 数据集对模型进行预训练, 再使用玉米穗丝数据集训练本文模型, 能在一定程度上提高用玉米穗丝数据集训练模型的效率<sup>[18]</sup>。

### 2.3 数据增广方法

作为一种数据驱动的深度学习模型, 训练样本越多, 覆盖的对象种类越全面, 越能保证 MF-SSD 模型的检测效果。本研究受时间和条件的限制, 采集的玉米穗丝数据集样本偏少, 故在研究中使用了数据增广技术。数据增广是在保持图像数据标签不变的前提下, 对原图像进行变换以获取新的样本。在已有样本的基础上制作更多的图像样本, 可以增加样本的多样性, 降低网络过拟合的风险。

研究采用的第 1 种数据增广方法是通过以一定角度 (0~180°) 随机旋转原始图像来形成新图像; 虽然在玉米穗丝图像采集过程中考虑了样本形态、角度等的多样性, 但玉米穗丝的生长以及成像角度是随机的; 通过随机角度旋转原始图像, 可部分消除样本多样性不足的问题; 研究在训练集中随机选取了 800 幅图像进行了这种增广操作。采用的第 2 种数据增广方法是水平翻转和平移图像; 将搜集到的图像进行等比例缩放, 然后从缩放后的图像中随机裁剪 300 像素×300 像素的图像作为结果, 同时对裁剪结果做随机水平翻转和平移; 研究在训练集中随机选取了 600 幅图像进行了第 2 种增广操作。

经过数据增广处理, 使训练样本的数量增加到了 3 800 幅, 可以增强训练的有效性, 同时提高网络推理的泛化能力。

### 2.4 网络训练策略

本文的网络模型使用二次训练策略。首先用 COCO 数据集训练本文的网络结构, 得到预训练网络模型, 保存模型参数。COCO 数据集图像目标均有明确信息标注, 并具有目标位置以及类别等信息, 且单张图像包含丰富的信息, 便于目标检测与识别。预训练模型学习率初始值为 0.001, 网络结构各层参数均初始化为 0。然后使用玉米穗丝数据集再次进行训练, 先对网络模型参数初始化, 对各个特征提取层参数进行赋值, 参数来源于 COCO 数据集预训练模型所产生的参数保存值; 训练后输出模型分类检测结果。该策略可提高模型训练性能<sup>[19]</sup>。

### 2.5 评价指标

物体检测中, 精度的评估通常用平均精度 (Average precision, AP) 表示, 覆盖面的评估通常用平均召回率 (Average recall, AR) 表示<sup>[20-21]</sup>。本文采用 COCO 数据集使用的目标检测评价指标<sup>[19]</sup>对算法效果进行评测, 以期能够客观、全面、多样化地评价本文的检测算法性能。

在评价指标中, AP 从 2 个方面来衡量。其一是从 IoU 方面, 对不同的 IoU 阈值分别计算 AP, 再综合得到平均值, 比如 AP (IoU=0.50~0.95) 表示对 IoU 范围在 0.50~0.95、0.05 为步长的检测结果分别计算精度并取平均值, AP (IoU=0.50) 表示 IoU=0.50 时的平均精度; 其二是以目标在图像中的尺寸为依据来看目标检测精度, 本文依据尺寸把目标对象分成小目标、中型目标和大目标, 小目标在图像中的尺寸小于 32 像素×32 像素, 中型目标在图像中的尺寸在 32 像素×32 像素~96 像素×96 像素之间, 大目标在图像中的尺寸大于 96 像素×96 像素。AR

是针对 3 类不同的目标尺寸分别计算每个图像中的穗丝目标在 IoU 方面的召回率的平均值。

使用 AP 和 AR 指标对算法的检测精度有效性进行评价,是目标检测领域对算法进行评估的常用方法<sup>[10, 22]</sup>,能较好地体现模型的检测性能。本试验使用该指标分析各个网络模型对小目标的检测效果和对所有目标的平均检测精度,优选出适合于玉米穗丝检测的网络模型。

### 3 结果与分析

#### 3.1 检测性能分析

利用玉米穗丝的数据集进行训练和测试,用对 VGG16-SSD、MobileNet-SSD、MF-SSD、MF-SSD-add-3、MF-SSD-cut-3 的训练结果分别进行检测,选取的置信度阈值为 0.50,即所有检测出物体的类别

置信度均在 0.50 以上。检测结果如表 1 所示,检测时间表示在 300 像素×300 像素图像上的检测时间。

由表 1 可知,MobileNet-SSD 在检测精确度上略低于 VGG16-SSD 网络,但仍表现出良好稳定性与精确性,在基本保证准确率的情况下减少了计算时间和参数数量。使用 MF-SSD 网络模型时,在 IoU 取 0.50 时,其 AP 可达到 96.6%,与 VGG16-SSD 模型基本一致,但在高 IoU(0.50~0.95)下的 AP 相比 VGG16-SSD 提高了 7.2%。在小目标检测精度上,相比 VGG16-SSD 模型,MF-SSD 有 20.4% 的提高,对玉米穗丝小目标检测的 AR 提高了 19.6%,说明把浅层的细节特征和高层的语义特征融合起来能够提升算法对小目标的检测能力。从统计的数据分析来看,本文所提出的 MF-SSD 模型及其变体的检测效果均比 VGG16-SSD 和 MobileNet-SSD 模型更好。

表 1 不同网络模型试验结果对比  
Table 1 Comparison of experimental results of different network models

| 网络模型<br>Network<br>model | 平均精度/%            |      |      | 平均精度/%            |                  |                 | 平均召回率/%         |                  |                 | 模型大小/MB<br>Model<br>size | 检测时间/s<br>Detection<br>time |
|--------------------------|-------------------|------|------|-------------------|------------------|-----------------|-----------------|------------------|-----------------|--------------------------|-----------------------------|
|                          | Average precision |      |      | Average precision |                  |                 | Average recall  |                  |                 |                          |                             |
|                          | IoU=              | IoU= | IoU= | 小目标               | 中型目标             | 大目标             | 小目标             | 中型目标             | 大目标             |                          |                             |
|                          | 0.50~0.95         | 0.50 | 0.75 | Small<br>target   | Medium<br>target | Large<br>target | Small<br>target | Medium<br>target | Large<br>target |                          |                             |
| VGG16-SSD                | 64.1              | 97.3 | 74.5 | 39.0              | 63.1             | 79.4            | 48.5            | 70.7             | 84.3            | 160                      | 0.320                       |
| MobileNet-SSD            | 62.9              | 96.5 | 72.8 | 37.9              | 61.9             | 78.7            | 46.7            | 68.7             | 82.8            | 84                       | 0.261                       |
| MF-SSD-cut-3             | 70.9              | 95.8 | 85.2 | 55.3              | 70.4             | 80.2            | 65.5            | 76.1             | 83.3            | 80                       | 0.260                       |
| MF-SSD                   | 71.3              | 96.6 | 86.8 | 59.4              | 70.4             | 79.4            | 68.1            | 75.7             | 83.0            | 82                       | 0.287                       |
| MF-SSD-add-3             | 71.9              | 96.7 | 87.6 | 57.0              | 71.8             | 79.8            | 67.3            | 76.8             | 83.1            | 86                       | 0.275                       |

分析 MF-SSD 模型变体的检测试验数据可知,与 MF-SSD 相比,MF-SSD-add-3 的 AP 略有提升,但对小目标的 AP 下降了 2.4%;MF-SSD-cut-3 在不同 IoU 取值下的检测性能要弱于 MF-SSD,但差距在 1% 以内,而对大目标的 AR 还要高出 0.3%。从模型尺寸的角度来说,MF-SSD-cut-3 的最终模型是最小的。同时,在检测时间方面,在不同变体的改进模型中 MF-SSD-cut-3 耗费时间最少,与 VGG16-SSD 模型相比速度提升了 18.7%。

综合上述结果,玉米穗丝目标检测的最终网络结构的选择方法是:在存储空间和运行时间有较高要求的嵌入式环境下,选择 MF-SSD-cut-3 模型可以在满足检测效果的前提下,以较小的空间代价获得相对较快的运行时间;在不考虑空间和时间的情况下,则可以选择 MF-SSD 模型以获得更好的检测效果。

#### 3.2 不同训练策略对训练结果的影响

为分析不同训练策略对模型训练的影响,本文对是否使用二次训练策略和是否使用 Focal loss 解

决样本不平衡问题进行了试验,并比对分析 Loss 的下降过程。

图 9 的 4 种策略中,MobileNet-SSD 未使用二次训练策略和使用二次策略的模型结构完全相同,两者区别在于是否使用 COCO 数据集预训练网络模型,对这 2 种网络的训练误差下降过程进行了比较。结果表明,使用二次训练策略情况的训练误差降落速度快于未使用二次训练策略的训练方法,同时其损失百分比更小,说明使用二次训练策略更利于模型收敛。MF-SSD 使用困难负例挖掘策略和使用 Focal loss 策略的区别在于解决不平衡样本的策略。结果表明,在学习率相同的情况下,MF-SSD 网络使用 Focal loss 能让网络更容易收敛,收敛速度加快,并且误差收敛到较小的值;使用困难负例挖掘时,Loss 值的降落速度较慢,震荡幅度较大,不易收敛。因此使用二次训练策略和 Focal loss 更利于 MF-SSD 模型的训练,训练出性能更好的检测模型。图 10 是使用 MF-SSD 检测的结果例图。

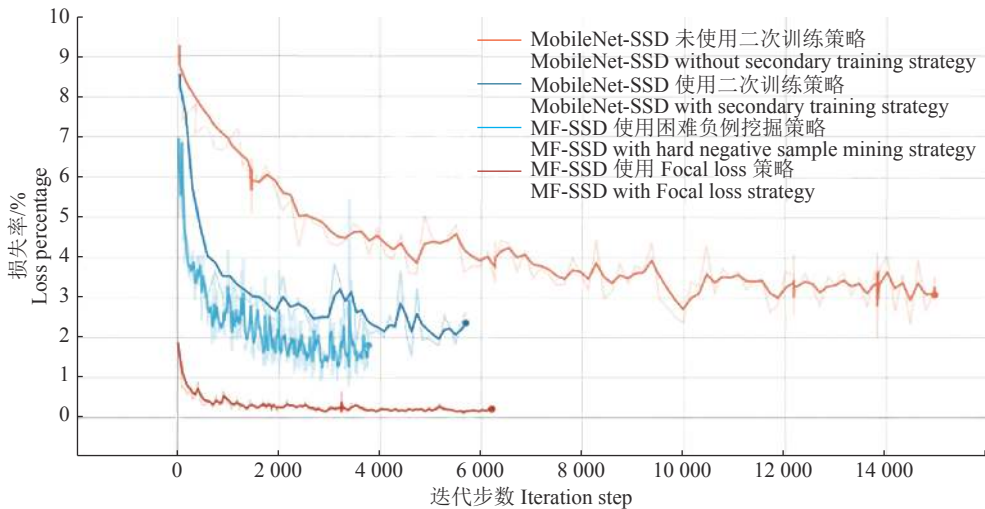
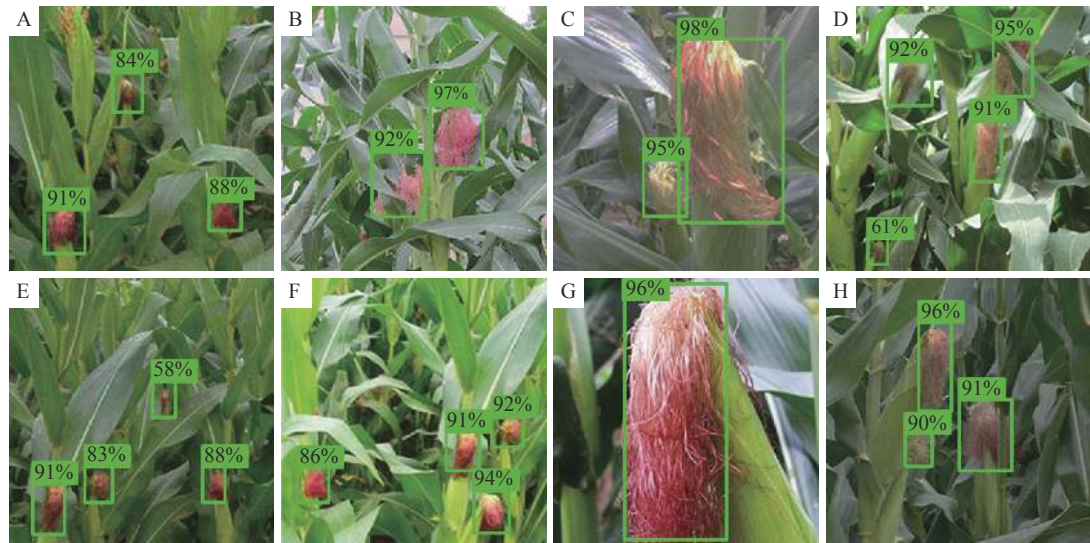


图 9 不同训练策略误差比较

Fig. 9 Error comparison of different training strategies



绿色方框表示网络给出的目标区域, 方框上方的百分数表示目标检测器检测到目标区域为玉米穗丝的概率

Green rectangle indicates the object area provided by network, the percentage above the rectangle indicates the probability of the object area is corn silk after target detector detection

图 10 MF-SSD 检测结果示例

Fig. 10 MF-SSD test result example

4 结论

本文通过改进目标检测算法 SSD 的不足之处, 在玉米穗丝检测的应用方面进行了试验研究和分析, 提出一种基于 MF-SSD 卷积神经网络的玉米穗丝检测网络模型。在原始 VGG16-SSD 的基础上, 用 MobileNet 替换特征提取器, 并加入多层特征融合结构改进 MobileNet-SSD, 得到 MF-SSD 网络模型; 同时, 根据不同应用场景的需求设计了 3 种结构的 MF-SSD 网络。在玉米穗丝训练样本上, 通过多种数据增广技术, 对每种网络分别进行训练, 优选一种网络模型作为玉米穗丝检测网络模型。研究结论如下:

- 1) 特征层融合的方式能够显著提高检测网络的检测能力, 提高对小目标的检测精度和召回率。与 VGG16-SSD 相比, MF-SSD 在高 IoU (0.50~0.95) 下的 AP 提高了 7.2%, 对小目标的 AP 提高了 20.4%, 对玉米穗丝小目标的 AR 提高了 19.6%, 并且降低了模型尺寸, 提高了识别速度。MF-SSD 玉米穗丝目标检测网络结构具有较高的检测性能, 能够有效用于玉米穗丝的检测。
- 2) 在存储空间和运行时间有较高要求的嵌入式环境下, 选择 MF-SSD-cut-3 模型可以在满足检测效果的前提下, 以较小的空间代价获得相对较快的运行时间; 在不考虑空间和时间因素的情况下, 可以选择 MF-SSD 模型获得更好的检测效果。

3) 二次训练策略可提高网络的收敛速度, 提高模型的稳定性和健壮性; Focal loss 能够有效解决 SSD 算法中正负样本数量不平衡问题, 使网络模型的训练更容易收敛。

参考文献:

[1] 杨笛, 熊伟, 许吟隆, 等. 气候变化背景下中国玉米单产增速减缓的原因分析[J]. 农业工程学报, 2017, 33(增刊 1): 231-238.

[2] 刘哲, 曲艺伟, 赵祖亮, 等. 玉米优良品种推广重心转移及扩散的时空规律[J]. 农业工程学报, 2018, 34(1): 178-185.

[3] 王传宇, 郭新宇, 杜建军. 基于时间序列红外图像的玉米叶面积指数连续监测[J]. 农业工程学报, 2018, 34(6): 175-181.

[4] 王临铭, 高晓阳, 李红岭, 等. 基于神经网络的大麦病害识别研究[J]. 甘肃农业大学学报, 2015, 50(2): 173-176.

[5] 王璨, 武新慧, 李志伟. 基于卷积神经网络提取多尺度分层特征识别玉米杂草[J]. 农业工程学报, 2018, 34(5): 144-151.

[6] 周云成, 许童羽, 郑伟, 等. 基于面向通道分组卷积网络的番茄主要器官实时识方法[J]. 农业工程学报, 2018, 33(15): 153-162.

[7] 刘永波, 雷波, 曹艳, 等. 基于深度卷积神经网络的玉米病害识别[J]. 中国农学通报, 2018, 34(36): 159-164.

[8] KOIRALA A, WALSH K B, WANG Z, et al. Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of ‘MangoYOLO’ [J]. *Precis Agric*, 2019, 20(6): 1107-1135.

[9] TIAN Y, YANG G, WANG Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J]. *Comput Electron Agr*, 2019, 157: 417-426.

[10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE T Pattern Anal*, 2017, 39(6): 1137-1149.

[11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//LEIBE B, MATAS J, SEBE N, et al. Lecture Notes in Computer Science. Cham:

Springer International Publishing AG, 2016, 9905: 21-37.

[12] SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. *Nature*, 2016, 529(7587): 484-489.

[13] YU Z, LI T, LUO G, et al. Convolutional networks with cross-layer neurons for image recognition[J]. *Inform Sciences*, 2018, 433: 241-254.

[14] ALAWAD M, LIN M. Stochastic-based deep convolutional networks with reconfigurable logic fabric[J]. *IEEE Transactions on Multi-Scale Computing Systems*, 2016, 2(4). doi: 10.1109/TMSCS.2016.2601326.

[15] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//IEEE. IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE, 2017: 936-944.

[16] 马浚诚, 杜克明, 郑飞翔, 等. 基于卷积神经网络的温室黄瓜病害识别系统[J]. 农业工程学报, 2018, 34(12): 186-192.

[17] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. *IEEE T Pattern Anal*, 2020, 42(2): 318-327.

[18] 彭红星, 黄博, 邵园园, 等. 自然环境下多类水果采摘目标识别的通用改进 SSD 模型[J]. 农业工程学报, 2018, 34(16): 155-162.

[19] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context[C]//FLEET D, PAJDLA T, SCHIELE B, et al. Lecture Notes in Computer Science. Cham: Springer International Publishing AG, 2014, 8693: 740-755.

[20] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//IEEE. IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 6517-6525.

[21] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[C]//IEEE. IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018.

[22] 李善军, 胡定一, 高淑敏, 等. 基于改进 SSD 的柑橘实时分类检测[J]. 农业工程学报, 2019, 35(24): 307-313.

【责任编辑 李庆玲】