基于自生成标签的玉米苗期图像实例分割

赵露露[1]，邓寒冰[1,2※]，周云成[1,2]，苗 腾[1,2]，赵 凯[1]，杨 景[1]，张羽丰[1]

（1. 沈阳农业大学信息与电气工程学院，沈阳 110866；2. 辽宁省农业信息化工程技术研究中心，沈阳 110866）

摘 要：在植物图像实例分割任务中，由于植物种类与形态的多样性，采用全监督学习时人们很难获得足量、有效且低成本的训练样本。为解决这一问题，该研究提出一种基于自生成标签的玉米苗期图像实例分割网络（automatic labelling based instance segmentation network，AutoLNet），在弱监督实例分割模型的基础上加入标签自生成模块，利用颜色空间转换、轮廓跟踪和最小外接矩形在玉米苗期图像（俯视图）中生成目标边界框（弱标签），利用弱标签代替人工标签参与网络训练，在无人工标签条件下实现玉米苗期图像实例分割。试验结果表明，自生成标签与人工标签的距离交并比和余弦相似度分别达到 95.23% 和 94.10%，标签质量可以满足弱监督训练要求；AutoLNet 输出预测框和掩膜的平均精度分别达到 68.69% 和 35.07%，与人工标签质量相比，预测框与掩膜的平均精度分别提高了 10.83 和 3.42 个百分点，与弱监督模型（DiscoBox 和 Box2Mask）相比，预测框平均精度分别提高了 11.28 和 8.79 个百分点，掩膜平均精度分别提高了 12.75 和 10.72 个百分点；与全监督模型（CondInst 和 Mask R-CNN）相比，AutoLNet 的预测框平均精度和掩膜平均精度可以达到 CondInst 模型的 94.32% 和 83.14%，比 Mask R-CNN 模型的预测框和掩膜平均精度分别高 7.54 和 3.28 个百分点。AutoLNet 可以利用标签自生成模块自动获得图像中玉米植株标签，在无人工标签的前提下实现玉米苗期图像的实例分割，可为大田环境下的玉米苗期图像实例分割任务提供解决方案和技术支持。

关键词：图像处理；深度学习；实例分割；弱监督学习；苗期玉米；植物表型

doi：10.11975/j.issn.1002-6819.202301085

中图分类号：S513；TP391.41；S24 **文献标志码：**A **文章编号：**1002-6819(2023)-11-0201-11

赵露露，邓寒冰，周云成，等. 基于自生成标签的玉米苗期图像实例分割[J]. 农业工程学报，2023，39(11)：201-211. doi: 10.11975/j.issn.1002-6819.202301085 http://www.tcsae.org

ZHAO Lulu, DENG Hanbing, ZHOU Yuncheng, et al. Instance segmentation model of maize seedling images based on automatic generated labels[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(11): 201-211. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202301085 http://www.tcsae.org

0 引 言

随着信息技术与农业生产过程的不断融合，计算机视觉技术被广泛用于获取植物表型信息。智能化植物表型监测技术能够监测农作物生长情况，通过分析表型特征并采取对策，有效缓解由气候变化、耕地减少等原因导致的粮食安全问题，加速育种与现代化农业的进步。相较于人工获取植物表型信息，传统的计算机视觉技术可以针对目标区域提供一种基于图像的非接触式检测手段[1]，如基于 Otsu 与分水岭结合的两级分割算法结合梯度 Hough 圆变换[2]，基于 LAB 颜色空间的图像分割[3]、基于统计直方图 K-means 聚类算法的图像聚类分割[4] 等，这些方法可以在图像背景信息相对简单时发挥分割优势。而当图像背景复杂度提高时，可以使用基于朴素贝叶斯分类的图像分割方法，通过引入概率因素提高分割准确率，实现复杂背景下植物特定区域的信息提取[5]。另外，基于多阶段柯西灰狼算法的多阈值图像分割优化器也成功实现了分割[6]。然而，利用传统计算机视觉方法获得较好的分割结果需要图像质量、构图内容、主要目标占比和位置等素满足一定要求，因此这些方法往往不具有普适性和通用性。

近些年深度学习技术蓬勃发展，特别是在图像的实例分割技术领域取得了实质性的进步[7]。自 HARIHARAN 等[8] 首次利用深度学习模型同步实现“目标检测+分割”任务以来，基于深度学习技术的实例分割方法开始迅速发展，并通过不断优化使实例分割模型的性能得到显著提升[9-12]。研究人员利用图像实例分割技术能够实现更精准的单体植株、叶片、器官、果实等信息提取，通过使用深度卷积神经网络提高实例分割模型对图像复杂背景的适应能力[13-14]。如孙红等[15] 利用 SSDLite-MobileDet 网络模型实现了玉米冠层的快速检测；王璨等[16] 通过改进双注意力机制结合形态学处理方法，实现玉米图像中的杂草目标分割；TURGUT 等[17] 采用基于注意力机制的深度学习架构，通过提取上下文特征并进行特征传播，以分层方式处理点区域实现植物器官分割；ZENKL 等[18] 利用 DeepLab V3+模型实现室外条件下冬小麦植物分割。

然而，目前的深度学习方法多采用全监督学习模式，即模型训练时需要提供“图像+像素级标签”，模型精度依赖于大规模的、精细到像素粒度的人工标注数据集。而植物表型领域的公开数据集较少，且数据种类单一，不

收稿日期：2023-01-17 修订日期：2023-02-22

基金项目：国家自然科学基金项目（31901399）；十四五国家重点研发计划项目子课题（2021YFD1500204）；国家重点研发计划项目子课题（2022YFD2002303-01）

作者简介：赵露露，研究方向为机器学习，计算机视觉。Email：zhaolulu_zll@163.com

※通信作者：邓寒冰，博士，副教授，研究方向为机器学习，计算机视觉，植物表型检测。Email：denghanbing@syau.edu.cn

具有普适性，研究人员往往要根据需求创建个性化的数据集，导致人工标注成本一直居高不下。为了缓解这一问题，一些研究人员尝试降低模型训练过程对标签精度的要求，通过使用图像级标签[19-22]或边界框标签[23-25]这种非精准标签训练深度学习模型，在弱监督学习、无监督学习的模式下实现分割，DiscoBox[26]以及 FreeSOLO[27]等方法显著缩小了与完全监督学习的差距，为农业图像实例分割技术提供了有效的技术方法；赵亚楠等提出基于边界框标注掩膜的深度卷积神经网络，利用伪标签代替像素级标签作为训练样本实现玉米植株图像的高精度分割[28]；ZHUANG 等[29]利用深度卷积神经网络，基于框级标注和颜色相似性的弱监督学习方式对绿叶蔬菜实现了实例分割；周云成等[30]提出基于稠密卷积自编码器的无监督深度估计模型，以番茄植株的双目图像为训练数据，通过估计深度误差以及阈值的精度为依据实现番茄植株图像的深度估计；LU 等[31]以无人机航拍图像获取冠层面积、冠幅、位置等信息，提出一种无监督图像分割方法，用于在自然光照条件下快速获取果树冠层。

实际上，弱监督学习仍然依赖于包含强大本地化信息的标注，尽管采取边界框标注图像的时间少于像素级标注，但为了获取更精准的分割结果，往往需要大量的训练数据，对于大田玉米图像，由于拍摄中存在光照、叶片重叠、杂草等影响，弱监督学习需要的标签依旧存在标注成本高的问题。而无监督学习不存在标注成本问题，但是由于没有标签对目标区域的范围界定，其分割精度不足以支撑对植物特征信息的描述，尤其是玉米植株这种形态复杂的对象，其分割效果不具优势，因此弱监督与无监督学习不适用于大部分的农作物图像实例分割任务。为了有效降低人工标注成本，又能较好地描述图像细节得到高精度的分割结果，本研究设计了一种基于自生成标签的实例分割网络，以弱监督实例分割模型为基础，在主干网络前加入弱标签自生成模块，利用颜色空间转换、轮廓跟踪和最小外接矩形在玉米苗期图像（顶视图）中生成目标边界框（自生成标签），利用自生成标签代替人工标签参与弱监督模型训练，最终在无人工标签条件下实现玉米苗期图像的实例分割。

1 试验材料与数据采集

本试验选择的玉米品种为“先玉 335”，该品种的植物性状表现为在幼苗期长势较强，成株叶片数在 20 片左右，具有高抗茎腐病，中抗黑粉病、弯孢菌叶斑病，大斑病、小斑病、矮花叶病等，其优越的抗病性可以让玉米在其营养生长期保持个体健康和株形完整。玉米播种时间在 4 月份，播种方式为机播，播种行距为 50 cm，株距为 30 cm。

在苗期阶段（单体株高在 20～30cm，叶片数 3−4 叶），试验数据由无人机（大疆“精灵 4-RTK”）高空俯视平行地面拍摄获取。为保证玉米植株的基本形态稳定以及数据采集时光照条件的相似性，航拍时选择在晴朗无风的天气，采集时间在 9:00−11:00，无人机航飞高度距地面 8m，采集数据过程中，飞行航线自动覆盖玉米植株生长的整片试验田。

试验中获取的原始图像像素大小为 5472×3078，人工筛选出 500 张满足试验要求玉米苗期图像（俯视图且去除拥有大面积杂草的图像），每张图像包含 7~9 列玉米幼苗。为了适应模型的网络深度，降低模型的过拟合几率，提高网络的泛化能力，试验对基础数据集中的原始图像进行数据增强处理。对全部原始图像进行镜像翻转、添加高斯噪声、随机改变亮度等操作实现数据增强，其中增强操作可以叠加使用，默认至少有一种增强生效，每张图像增强两次，将数据集扩增到 1 500 张，在基础数据集和数据增强的数据集中分别按照 8∶1∶1 的比例随机划分训练集、验证集与测试集，以保障数据分布的合理差异性，其中 1 200 张作为训练数据，150 张作为验证数据集，150 张作为模型的测试数据集。

2 基于自生成标签的实例分割模型

2.1 总体模型框架和训练平台

本研究旨在构建具备标签自动生成的弱监督图像实例分割模型，以实现大田环境下低成本、高精度的玉米苗期图像实例分割，总体模型框架主要包括：1）图像采集与预处理：通过无人机采集大田玉米苗期顶视图，并根据试验需要人工筛选可以进行训练和测试的图像，通过数据增强方法提高样本多样性；2）标签自生成模块：在 HSV 颜色空间中进行阈值分割、膨胀前景植株区域并删除小噪声点得到仅含前景玉米植株的二值图像，利用二值图像信息进行轮廓检测并生成最小外接矩形，在原始图像中自动产生主要目标对象的边界框信息，并利用阈值筛选最终边界框，自动生成图像的弱标签；3）构建并优化弱监督深度卷积神经网络模型：利用弱标签对弱监督深度学习模型进行训练，最终获得可用于玉米苗期图像实例分割的网络模型。具体如图 1 所示。

为了保证模型训练过程的公平性，同时提高训练效率，试验中使用 2 种计算平台，平台的具体参数如表 1 所示，其中平台 1 由于计算卡性能更优，承担所有模型的预训练任务；平台 2 用于所有预训练模型的迁移学习和调优。

2.2 图像标签自生成方法

在标记熟练的情况下使用 LabelMe 软件标注单幅玉米苗期图像（顶视图），像素级标注时间约为 1 127 s，边界框标注时间约为 100 s，虽然采用边界框标注能节省标记的时间，但当数据增多时，即便边界框标注也需要耗费大量的时间成本，因此，本研究设计了图像标签自生成方法，整个过程不需要对图像进行人工标注，便能够根据图像中玉米植株的位置自动生成边界框（弱标签）。

弱标签自生成模块主要包括 2 个部分：1）颜色阈值分割：将图像由 RGB 转换为 HSV 颜色空间，通过设定玉米植株的颜色阈值范围将图像背景区域去除，消除地面影子、土地等对前景信息的影响；2）基于轮廓跟踪的最小外接矩形法：将阈值分割后的二值图像进行边缘检测，得到前景植株的轮廓点集，最后利用得到的轮廓点集坐标生成前景目标的最小外接矩形从而获得边界框标签。

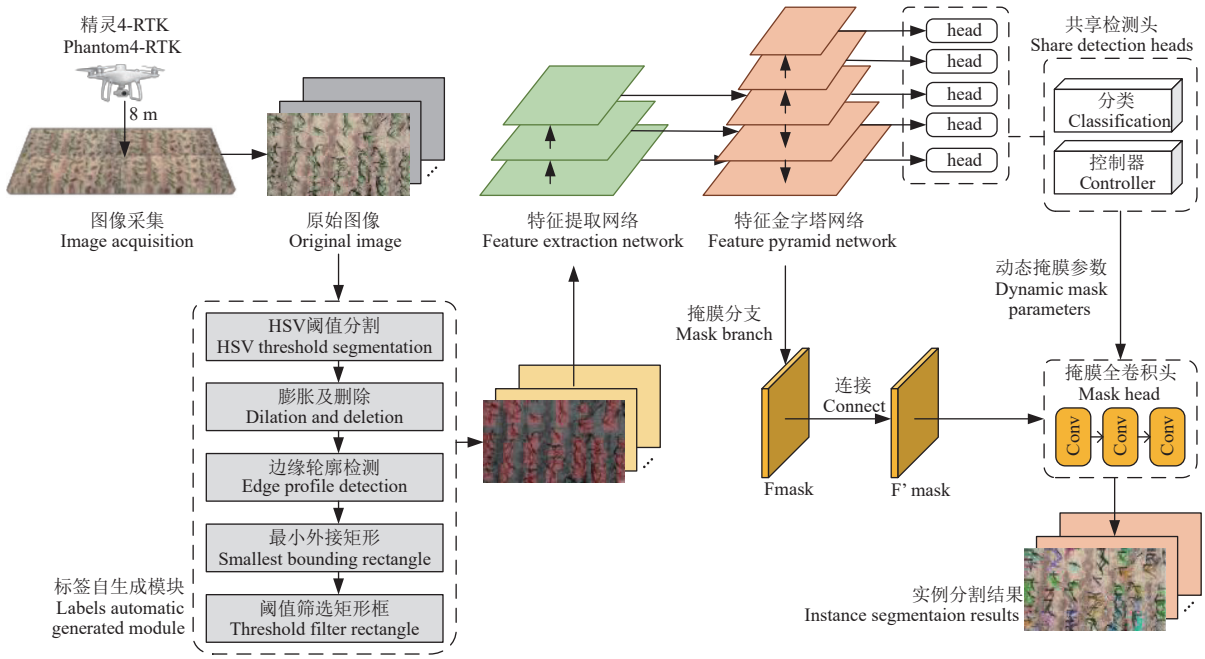


图 1　总体模型框架图

Fig.1　Overall model framework

表 1　试验平台参数

Table 1　Parameters of experimental platform

平台 Platforms	中央处理器主频 Central processing unit frequency/GHz	中央处理器核心数 Central processing unit cores	内存 Memory/GB	图形处理器 Graphic processing unit	图形处理器显存 Graphic processing unit memory/GB	图形处理器核心数 Graphic processing unit cores
平台 1 Platform 1	4.8	18	128	NVIDA 3 090	24	10 752
平台 2 Platform2	2.1	16	64	NVIDA 2080Ti	11	4 352

2.2.1　颜色阈值分割

在大田环境下，玉米植株受光照影响会在植株周围地面形成影子，这些影子在顶视图中会呈现出与植株形态相似的像素区域。试验中发现，直接在 RGB 图像上进行阈值分割容易将植株影子划入前景信息，导致后续生成边界框时将影子也框入边界框，降低边界框标签质量，并且阴影与植株边缘信息极度相似，影响分割精度。为解决阴影对前景植株信息的影响，本试验将图像由 RGB 颜色空间转为 HSV 颜色空间再进行阈值分割。HSV 能够描述图像的色调（H）、饱和度（S）以及明度（V），在 HSV 颜色空间上可以准确地对指定的颜色进行分割。对于存在多种颜色的数据集，可参考 HSV 基本颜色分量增设不同颜色的阈值进行分割。图 2 为本试验中玉米苗期图像分别在 H、S 与 V 分量上的像素值分布，根据图 2 设置 H 的像素值范围为 [15, 35] 和 [45, 70]，S 的像素值范围为 [15, 255]，V 的像素值范围为 [40, 255]。

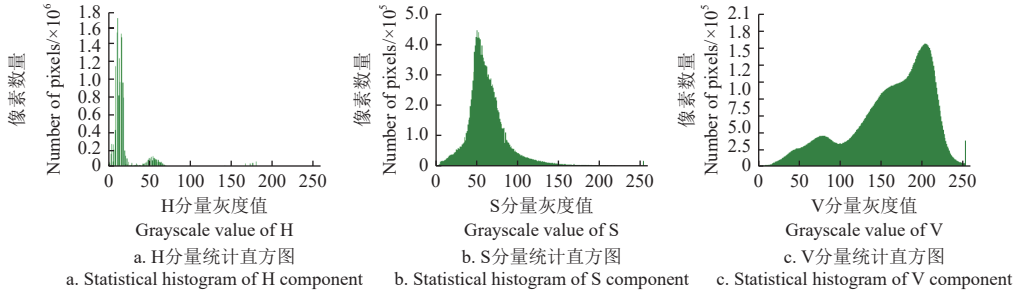
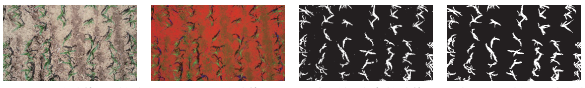


图 2　玉米苗期图像的 H、S、V 分量统计直方图

Fig.2　Statistical histogram of H, S, V components of maize seedling images

基于 HSV 阈值的分割结果如图 3 所示，通过观察发现图像中存在一些离散的小面积噪声点，因此本试验在颜色阈值分割后增加一次膨胀处理，删除小联通区域的冗余噪声点，分割结果如图 3d，由图 3c 与 3d 中前景植株内部孔洞对比可以看出，通过膨胀处理可以有效解决阈值分割造成的小部分像素缺失，保持了玉米植株的实例完整，删除小连通区域既去除了冗余噪声点，又有效抑制了膨胀后未成功连接到玉米植株实例的小面积区域，避免了一个实例被隔开的情况。



a. RGB图像/原图 a. RGB image /original image　b. HSV图像 b. HSV image　c. 阈值分割图像 c. Threshold segmentation image　d. 去噪后的二值图像 d. Binary image with de-noising

图 3　基于 HSV 颜色空间的阈值分割

Fig.3　Threshold segmentation based on HSV color space

2.2.2 基于轮廓跟踪的最小外接矩形法

在获得图像前景信息（玉米植株）后，需要根据玉米植株的位置获得对应的边界框信息，本研究采取的方法为在二值图像中对前景目标进行边缘检测，识别前景目标边缘后再绘制其最小外接矩形。为了获取前景目标轮廓，本研究采用文献 [32] 的方法解析二值图像的拓扑结构，获取二值图像前景的边界的包围关系。具体算法如下：

1）确定点边界类型。定义输入的二值图像 $F=\{f_{ij}\}$，初始化边界序号 N_{BD}=1，前一个边界的编号 LN_{BD}=1，并且每一行扫描开始，LN_{BD} 重置为 1。使用光栅扫描输入的二值图像，找到点 (i, j) 满足边界跟踪初始点的条件则终止扫描。条件为：若 $f_{ij}=1$ 并且 $f_{i,j-1}=0$，则 (i, j) 是外边界开始点，$N_{BD}=N_{BD}+1$，$(i_2, j_2)=(i, j-1)$，该点是一个外边界；若 $f_{ij}\geqslant 1$ 并且 $f_{i,\ j+1}=0$，则 (i, j) 是孔边界开始点，$N_{BD}=N_{BD}+1$，$(i_2, j_2)=(i, j+1)$。如果 $f_{ij}\geqslant 1$，则 $LN_{BD}=f_{ij}$，该点是一个孔边界。

如果点 (i, j) 同时满足以上 2 个条件，则该点作为外边界的起始点。

2）基于边界类型决定当前边界的父边界。判断规则如表 2 所示。

表 2 新边界的父边界判断规则

Table 2 Judgment rules for parent boundary of new boundary

边界类型 Boundary	前边界类型 Front boundary type	
	外边界 Outer border	孔边界 Hole boundaries
外边界 Outer boundary	边界的父边界	边界
孔边界 Hole boundary	边界	边界的父边界

3）从边界起始点 (i, j) 开始，跟踪已检测到的边界。以 (i_2, j_2) 为起始点，按顺时针方向查找以 (i, j) 为中心的 8 邻域的第一个非 0 像素点（前景目标），记录为 (i_1, j_1)；再以 (i_1, j_1) 的下一个点为起始点，按逆时针方向查找以 (i_3, j_3) 为中心的 8 邻域的第一个非 0 像素点为 (i_4, j_4)；更新边界序号 N_{BD}，迭代更新起始点，直至扫描到图像的右下角顶点时结束，边界序号相同的像素点属于同一个边界。

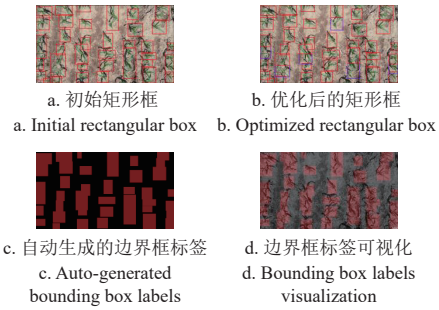
利用该算法得到的玉米苗期轮廓图像如图 4 所示，通过提取图像前景目标的边界信息，得到玉米植株的外轮廓点集 $S=\{S_1, S_2, \ldots, S_n\}$后，遍历一个植株点集 S_k 内所有像素点，将 S_k 内 i 值最小与 j 值最大的点记为 (i_{min}, j_{max})，作为外接矩形的左上顶点，将 S_k 内 i 值最大与 j 值最小的点记为 (i_{max}, j_{min})，作为外接矩形的右下顶点。

图 4 基于二值图像生成的内部轮廓图像

Fig.4 Interior contour images based on binary images

通过这 2 个顶点可以得出外接矩形的顶点与长宽，绘制出轮廓的垂直边界最小矩形，这个矩形与图像上下边界平行，保证了与手动标注矩形框在方向上的一致性，同时解决了由手动标注的随机性导致的冗余背景、前景植株框定不完全等问题。

此外，本研究对自动生成的边界框做优化处理：1）自动生成的边界框（初始矩形框，如图 5a）中会有一些不包含植株的矩形区域（例如小面积杂草），这些信息是在颜色阈值分割后残留的信息，因此在自动生成矩形时要设定矩形长与宽的阈值，使小于阈值的矩形框不被绘制；2）相邻玉米植株叶片部分可能出现互相遮挡的情况，因此在生成轮廓点集时会有多株玉米被分在一个轮廓里，导致生成的边界框包含多个玉米植株，影响分割效果。为此，本研究统计了全部自动生成矩形边框的长、宽值，利用长、宽的均值作为边界框大小的阈值，将长、宽值过大的矩形框进行均分，以保证一个边界框只有一株玉米植株，处理结果如图 5b 所示。将自动生成的边界框信息作为弱监督学习模型的标签信息（伪标签），由于标签生成方过程中没有人工标注，因此不产生人工成本，伪标签及其可视化图如图 5c 和 5d 所示。



a. 初始矩形框 a. Initial rectangular box

b. 优化后的矩形框 b. Optimized rectangular box

c. 自动生成的边界框标签 c. Auto-generated bounding box labels

d. 边界框标签可视化 d. Bounding box labels visualization

注：图 5b 中红色部分为初始生成的矩形框，紫色部分为经过阈值处理后进一步分割的矩形框。

Note: The red part in Fig.5b is the initially generated rectangular box, and the purple part is the rectangular box that is further divided after threshold processing.

图 5 最小外接矩形和自动生成的边界框标签

Fig.5 Smallest enclosed rectangle and automatic generated bounding box labels

2.3 弱监督图像实例分割网络及关键评价指标

以基于边界框标注的弱监督实例分割模型（BoxInst[33]）为基础，增加标签自生成模块，对玉米植株 RGB 图像上的目标区域（植株）自动生成边界框，替代构建训练样本中的人工标注过程。以全卷积的方式利用动态卷积滤波器在全图上动态获取每一个实例的掩膜，由于对每个实例都是在全图尺度上的预测，因此可以更好地分割不规则的形状，适用于大田环境下玉米苗期图像的实例分割。而大田场景下的玉米苗期图像背景复杂，需要对株型、叶片细节等部分的特征进行精确提取，虽然越深的网络分类准确度越高，但考虑到本试验的数据样本类别单一，数据集相较于公开数据集（COCO）小很多，为了减少网络过深造成的过拟合风险，因此选择 ResNet50 作为主干特征提取网络，并采用双向特征金字塔网络（bidirectional feature pyramid network，BiFPN）

作为特征提取网络，BiFPN 能够在不增加原有模型计算量的情况下在不同特征层进行加权以平衡不同尺度的特征信息，达到更高效的多尺度融合，其中高分辨率的特征保留空间位置信息，低分辨率的特征保留类别相关的抽象信息，能够对目标进行准确分类并减少小目标的漏检情况。

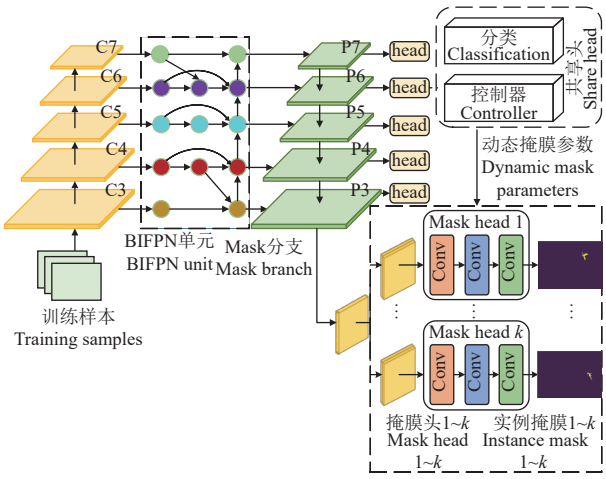
模型结构如图 6 所示，主要分为 2 部分：第一部分利用全卷积网络提取特征做逐像素回归，根据共享检测头得出目标实例的类别和动态生成滤波器参数，其中分类分支（classification）预测每个像素的类别，控制器分支（controller）用于产生掩膜分支（mask branch）的网络参数，该参数可在全局上对每个实例生成一个掩膜。

第二部分为掩膜分支，根据第一部分检测头动态产生的掩膜参数，结合经过主干网络特征提取后卷积生成的掩膜特征图作为输入，且由于每个实例都独有对应掩膜分支，包含实例的形状和大小等信息，所以当掩膜分支作用于全局掩膜特征图上时，就可以区分当前实例和背景信息，从而预测出每一个实例的掩膜。图 6 中 Mask head 有 3 个 1×1 卷积，每个卷积有 8 个通道，采用 ReLU 函数作为激活函数，不使用归一化层，最后一层有一个输出通道，并使用 Sigmoid 预测每个类别的概率。具体步骤如下：

1）原始图像经过标签自生成模块得到用于网络训练的带有边界框掩膜的训练样本；

2）利用卷积神经网络提取特征并在分类分支上实现分类和中心度检测，过滤效果不好的检测框；

3）利用动态卷积滤波器对多个实例动态生成多个不同的掩膜参数，结合经主干网络特征提取再卷积生成的掩膜特征，区分当前实例和背景信息，从而预测出每一个实例的掩膜。



注：C3～C7 为主干网络特征层，P3～P7 为特征提取网络的特征层。
Note: C3-C7 are the feature layers of the backbone network, P3-P7 are the feature layers of the feature extraction network

图 6　动态掩膜过程
Fig.6　Dynamically mask process

对于深度学习模型，网络越深模型收敛所需要的训练样本数越多。为了避免样本数量少导致模型普适性差的问题，采用公开数据集对模型进行预训练，再将学习到的特征迁移到新的学习任务中。本文首先利用 COCO 公开数据集对主干网络进行预训练，将预训练权重迁移到本文网络模型中，将模型卷积层参数初始化，以提高卷积层的特征提取能力和泛化能力。采用随机梯度下降进行网络训练，由于训练中使用的 GPU 显存有限，而原始图像的分辨率较高，因此设置较小的 batch size，具体参数及其初始值为：图像批量数为 2，学习率 为 0.01，训练步数 2 000，初始动量 0.9。

在试验中，为了保证模型评估的公平性，AutoLNet 以及所有其他弱监督模型的训练过程都使用自生成的边界框标签，有监督模型使用像素级人工掩膜标签。

本文的分割任务是实现大田场景下的玉米苗期图像实例分割，为了解决图像像素不均衡问题，本研究采用的损失函数计算式为

$$\begin{cases} L = L_{fcos} + \lambda L_{mask} \\ L_{fcos} = L_{cls} + L_{loc} + L_{ctr} \\ L_{mask} = L_{proj} + L_{pairwise} \end{cases} \tag{1}$$

式中L_{fcos}表示检测头产生的损失，L_{mask}表示实例分割的掩膜损失，通过权重λ平衡两个损失，L_{cls}表示分类损失，L_{loc}表示回归损失，L_{ctr}表示中心度损失，L_{proj}表示投影损失，$L_{pairwise}$表示成对损失。

由于训练模型使用自动生成的边界框掩膜作为训练样本，需要验证自动生成边界框的精度。本文选取距离交并比 D_{IoU} 作为评价指标，如式 (2) 所示，距离交并比不仅考虑 2 个边界框的交并比，同时考虑边界框的距离、重叠率及尺度，可以更好地衡量自动生成边界框的精度。

$$D_{\mathrm{IoU}} = \frac{1}{\mathrm{num}} \sum_{i \in \mathrm{num}} \left[\frac{1}{k} \sum_{j=1}^{k} \frac{S_i^j(G \cap A)}{S_i^j(G \cup A)} - \frac{\rho^2(b_i^A, Gb_i^G)}{c_i^2} \right] \times 100\% \tag{2}$$

式中 num 表示样本集合，k 表示一个样本中的实例个数，G 代表真值标注图像，A 代表自动标注图像，$S_i^j(G \cap A)$ 表示第 i 个样本中第 j 个实例真值标注与自动生成边界框的交集面积，$S_i^j(G \cup A)$表示第 i 个样本中第 j 个实例真值标注与自动生成边界框的并集面积，b 代表边界框的中心点，c 为包含 2 种边界框的最小闭合区域的对角线距离，ρ代表两点间的欧氏距离。

余弦相似度用向量空间中两个向量夹角的余弦值衡量两个个体间的差异大小，将图像表示成一个向量，通过计算向量间的余弦距离表征两张图像的相似度，可以检测二维空间中两张图像的相似度，用于不同分割方法生成的二值图像与真值图像的对比，其中前景像素值为 1 (白色)，背景像素值为 0(黑色)。按照式 (3) 计算余弦值。

$$\cos\theta = \frac{1}{\mathrm{num}} \sum_{i \in \mathrm{num}} \frac{\sum_{j=1}^{k} \left(P_i^j \cdot T_i^j \right)}{\sqrt{\sum_{j=1}^{k} \left(P_i^j \right)^2} \cdot \sqrt{\sum_{j=1}^{k} \left(T_i^j \right)^2}} \times 100\% \tag{3}$$

式中 P 代表预测生成的二值图像，T 代表真值图像，P_i^j 表示第 i 个预测的分割图像中第 j 个向量，T_i^j表示第 i 个真值标签图像中的第 j 个向量，k 表示每个图像的像素点

个数。

为了验证网络模型的图像分割精度，本文采用的评价指标为平均精度（average precision，AP），如式 (4) 所示。

$$\mathrm{AP}=\int_0^1 p(r)\mathrm{d}r \tag{4}$$

式中 AP 的值为 PR 曲线下的面积，p 为精度，r 为召回率，计算式如式（5）和式（6）所示。

$$p=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}\times 100\% \tag{5}$$

$$r=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}\times 100\% \tag{6}$$

式中 TP 表示真正例，表示预测掩膜置信度大于置信度阈值且真实掩膜覆盖度最高的像素集合；FP 为预测掩膜像素集合减去 TP 像素集合；FN 为真值掩膜像素集合减去 TP 像素集合。精度 p 代表被预测为正例的结果占真正例的比值，召回率 r 代表真样本中被预测为正例的比值，AP 值越大，模型的性能越好。

3 结果与分析

3.1 自生成标签质量评估

为了评估标签自生成模块得到的边界框标签质量，选取大津阈值分割、全局阈值分割和自适应阈值分割方法，分别用分割后的二值图像以及生成的边界框与真值进行对比。边界框质量与分割产生的二值图掩膜质量密切相关，根据 2.2 节，使用式 (3) 获得掩膜与真值间的余弦相似度，使用式 (2) 获得掩膜对应的边界框与真值标注的边界框间的距离交并比。图 7 给出了相关分割方法得到的二值图像及对应的边界框信息与真值的对比情况。

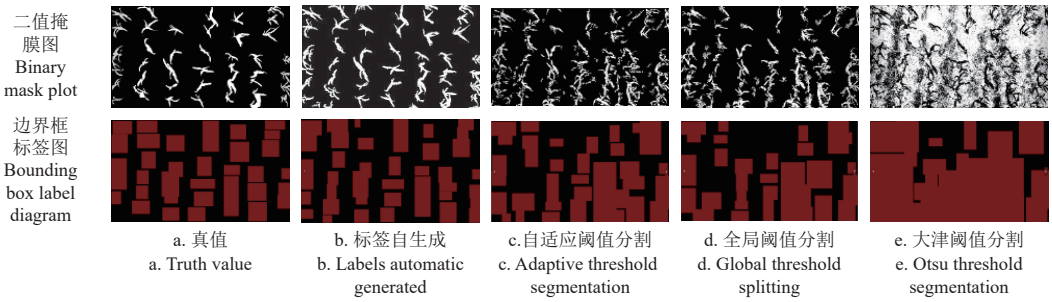


图 7 不同分割方法的二值掩膜与边界框标签对比

Fig.7 Comparison of binary masks and bounding box labels with different segmentation methods

如图 7 所示，标签自生成模块能够生成与真值图像基本一致的二值掩膜图，并且由此得到的边界框标签也与真值有相同的空间分布。表 3 给出了不同方法产生的掩膜及标签与真值之间的余弦相似度与距离交并比。标签自生成模块对应的掩膜图像与真值图像的余弦相似度达到 94.10%，自生成的边界框与真值边界框的距离交并比达到 95.23%，2 个指标都远高于其他方法。

表 3 自动标注与人工标注的标签质量对比

Table 3 Label quality comparison of automatic labeled with manual labeled （%）

方法 Segmentation method	余弦相似度 Cosine similarity	距离交并比 Distance intersection over union
大津阈值 Otsu threshold	29.44	32.07
全局阈值 Global threshold	43.97	69.42
自适应阈值 Adaptive threshold	48.48	75.83
本文 This paper	94.10	95.23

标签质量对比结果表明，标签自生成模块可以生成用于弱监督训练的有效标签，在无人工参与情况下能够实现大田图像的高质量自动标注，而高质量的标签可以保证模型训练过程的稳定性。如图 8 所示，在相同硬件平台和网络训练参数条件下，分别使用自生成标签与人工标签进行训练，两种样本的 6 种训练损失的变化趋势基本一致，即具有相似的收敛趋势，可见标签自生成方法产生的边界框标签样本完全可以代替真值标注进行模型训练，能够获得与真值标注样本相似的稳定性。

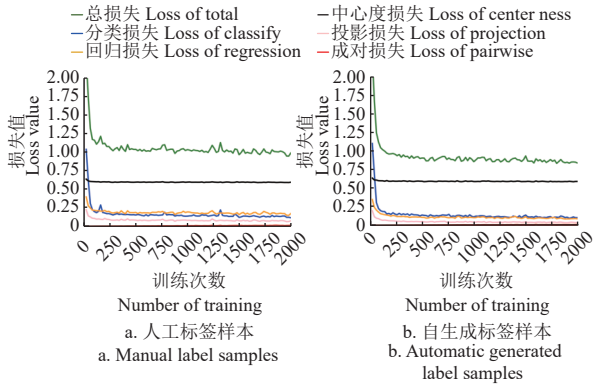


图 8 人工标签与自动标签样本的训练损失

Fig.8 Training loss of manual label and automatic generated label samples

3.2 模型测试结果分析

选择 AP50～AP75、$\mathrm{AP_L}$ 与 AP 验证网络模型的实例分割精度（详见表 4）。在构建 AutoLNet 模型时，分别选用 4 种主干网络进行测试。如表 4 所示，使用自生成标签为样本进行训练时，AutoLNet 模型使用 ResNet50+BiFPN 主干网络得到的预测框和掩膜精度最好，模型的 AP 值分别为 68.69% 和 35.07%，其中，模型在交并比阈值大于等于 0.75 时（AP75），预测框精度为 73.67%，掩膜精度为 12.03%，当交并比阈值大于等于 0.5 时（AP50），预测框精度达到 96.39%，掩膜精度达到 91.75%，表明随着预测值与真值间交并比阈值的降低，AutoLNet 预测

的平均精度随之升高；主干网络为 ResNet50+BiFPN 时，由于 BiFPN 可以引入学习权重学习不同输入特征，并应用自顶向下和自底向上的多尺度融合方式进行特征提取，因此相较于 ResNet50+FPN，其预测框和掩膜的平均精度提高了 4.47 个百分点（其中 ResNet50+BiFPN 的 AP 值为 68.69%，ResNet50+FPN 的 AP 值为 64.22%）和 2.4 个百分点（其中 ResNet50+BiFPN 的 AP 值为 35.07%，ResNet50+FPN 的 AP 值为 32.67%）；主干网络为 ResNet101+BiFPN 时，预测框和掩膜精度的 AP 值分别下降了 2.63（其中 ResNet50+BiFPN 的 AP 值为 68.69%，ResNet101+BiFPN 的 AP 值 66.06%）和 6.51 个百分点（其中 ResNet50+BiFPN 的 AP 值为 35.07%，ResNet101+BiFPN 的 AP 值 28.56%），AP 值并未随主干网络的加深而提高，这是由于本试验样本数据仅有 1 类，且样本总体数量不及公开数据集，在仅使用边界框标注进行训练时，网络深度的增加反而会导致某些浅层的学习能力下降，限制深层网络的学习，从而降低特征提取能力。另外，AutoLNet 模型在交并比超过 70% 后的掩膜精度降幅明显，而预测框精度降幅稳定，其原因是 AutoLNet 模型以边界框为监督，通过网络学习的像素级细节不如有监督模型，当交并比阈值较高时，预测的掩膜精度通常不如预测框精度，易出现较大幅度下降。

由表 4 可知，自生成标签与人工标签皆在主干网络为 ResNet50+BiFPN 时有最好的精度表现，且自生成标签在预测框与掩膜的平均精度（AP）分别高出人工标签 10.83（其中 ResNet50+BiFPN 的自生成标签 AP 值为 68.69%，人工标签 AP 值为 57.86%）与 3.42 个百分点（其中 ResNet50+BiFPN 的自生成标签 AP 值为 35.07%，人工标签 AP 值为 31.65%），这是因为相较于人工标签，自生成标签能够对目标实例的边界进行更精准的定位，极大地减少了人工标注的随机性与不确定性，使更精准的前景信息被模型网络学习，从而获得更高的精度。

表 4　AutoLNet 模型在不同主干网络下的平均精度

Table 4　Average precision of AutoLNet under different backbone network　（%）

主干网络 Backbone network	标签 Label	评价类型 Evaluation type	AP50	AP55	AP60	AP65	AP70	AP75	AP_L	AP
ResNet50+FPN	自生成标签	预测框	88.21	77.93	70.61	69.41	66.37	65.83	64.76	64.22
		掩膜	89.21	76.44	61.92	49.63	26.16	8.17	33.05	32.67
	人工标签	预测框	87.05	84.71	77.64	77.52	71.16	63.23	57.13	57.03
		掩膜	73.56	71.92	56.93	32.01	11.64	4.36	28.63	28.07
ResNet50+BiFPN	自生成标签	预测框	96.39	86.79	81.96	80.93	79.67	73.67	68.94	68.69
		掩膜	91.75	81.92	78.17	57.31	36.62	12.03	35.44	35.07
	人工标签	预测框	95.58	95.56	84.42	83.35	75.00	67.25	57.86	57.86
		掩膜	92.19	89.72	66.90	44.33	14.71	5.70	31.69	31.65
ResNet101+FPN	自生成标签	预测框	87.05	82.72	78.62	74.90	74.86	66.90	59.16	59.09
		掩膜	87.67	81.09	58.44	45.91	20.71	3.13	29.73	29.57
	人工标签	预测框	93.84	93.81	89.65	84.62	78.75	63.59	57.83	57.83
		掩膜	91.59	81.20	61.22	34.63	16.84	7.92	29.53	29.49
ResNet101+BiFPN	自生成标签	预测框	94.27	84.90	80.78	77.42	77.36	71.94	66.14	66.06
		掩膜	83.66	76.51	55.12	42.03	11.22	2.07	28.79	28.56
	人工标签	预测框	89.60	89.55	84.01	80.42	78.36	73.02	57.04	57.04
		掩膜	81.10	75.92	54.00	27.77	15.34	6.97	26.13	26.13

注：AP50、AP55、AP60、AP65、AP70 与 AP75 表示分割结果与真值间的交并比（IoU）分别大于等于 0.5、0.55、0.6、0.65、0.7 和 0.75，AP_L 表示分割掩膜的像素数量大于 9 216，AP 是所有类别的平均值，即分割结果与真值间的交并比从 0.5 开始间隔 0.05 取值到 0.95 的总体均值。下同。

Note: AP50, AP55, AP60, AP65, AP70 and AP75 indicate that the intersection and union ratio (IoU) between the segmentation result and the true value are greater than or equal to 0.5, 0.55, 0.6, 0.65, 0.7 and 0.75, respectively, AP_L indicates that the number of pixels of the segmentation mask is greater than 9 216, AP is the average of all categories, that is, the intersection ratio between the segmentation result and the truth value is the overall mean from 0.05 to 0.95. The same below.

为比较 AutoLNet 模型与其他基于边界框标签的弱监督实例分割模型的平均精度，本试验选取 2 个较为成熟的弱监督模型 DiscoBox 和 Box2Mask 进行对比。为保证公平性，DiscoBox 和 Box2Mask 使用与 AutoLNet 相同的人工标签作为训练样本。由表 5 可知，AutoLNet 模型对应的预测框平均精度和掩膜平均精度都高于 DiscoBox 和 Box2Mask 模型。在全局平均精度（AP）方面，AutoLNet 预测框精度比 DiscoBox 模型高 11.28 个百分点（其中 AutoLNet 的 AP 值为 68.69%，DiscoBox 的 AP 值为 57.41%），比 Box2Mask 模型高 8.79 个百分点（其中 AutoLNet 的 AP 值为 68.69%，Box2Mask 的 AP 值为 59.90%）；而掩膜精度方面，AutoLNet 比 DiscoBox 模型高 12.75 个百分点（其中 AutoLNet 的 AP 值为 35.07%，DiscoBox 的 AP 值为 22.32%），比 Box2Mask 模型高 10.72 个百分点（其中 AutoLNet 的 AP 值为 35.07%，Box2Mask 的 AP 值为 24.35%）。可以证明，AutoLNet 的网络结构在全图的掩膜特征中动态地为每一个实例生成一个掩膜的分割方式，在实现基于边界框标注的弱监督实例分割方面是有效的，能够以更高的精度完成大田环境下玉米苗期图像的实例分割。

为了比较 AutoLNet 模型在图像实例分割精度上与全监督模型的差别，本文选择全监督模型 CondInst 和 Mask R-CNN 与 AutoLNet 进行对比。如表 6 所示，与 CondInst 模型对比，AutoLNet 模型的预测框与掩膜精度略低于主干网络为 ResNet101+BiFPN 的 CondInst 模型，分别达到 CondInst 模型的 94.32% 和 83.14%，全局平均精度（AP）分别相差 4.14（其中 CondInst 的 AP 值为 72.83%，AutoLNet 的 AP 值为 68.69%）和 7.11 个百分点（其中 CondInst 的 AP 值为 42.18%，AutoLNet 的 AP 值为 35.07%），接近于主干网络为 ResNet50+FPN 的 CondInst 模型，掩膜的 AP 值低 3.84 个百分点（其中 CondInst 的 AP 值为 38.91%，AutoLNet 的 AP 值为 35.07%），而预

测框的 AP 值则高出 4.76 个百分点（其中 AutoLNet 的 AP 值为 68.69%，CondInst 的 AP 值为 63.93%），尤其在交并比阈值大于等于 0.5 时（AP50），AutoLNet 模型的预测框精度与掩膜精度高于 CondInst 模型 2.08（其中 AutoLNet 的 AP50 值为 96.39%，CondInst 的 AP50 值为 94.31%）和 3.81 个百分点（其中 AutoLNet 的 AP50 值为 91.75%，CondInst 的 AP50 值为 87.94%），说明交并比阈值要求较低时，AutoLNet 模型优于 CondInst 模型；与 Mask R-CNN 模型对比，在预测框精度与掩膜精度上 AutoLNet 模型皆优于 Mask R-CNN 模型，相较于基于 ResNet101+FPN 主干网络的 Mask R-CNN 模型，AutoLNet 在全局平均精度（AP）上高出 7.54（其中 AutoLNet 的 AP 值为 68.69%，Mask R-CNN 的 AP 值为 61.15%）和 3.28 个百分点（其中 AutoLNet 的 AP 值为 35.07%，Mask R-CNN 的 AP 值为 31.79%）。结果表明，在玉米苗期图像实例分割任务中，AutoLNet 模型的分割精度接近全监督模型 CondInst，且优于 Mask R-CNN 模型，原因是 AutoLNet 模型虽然以边界框为监督，但在自生成边界框时利用图像的纹理、色彩等基本特征对目标实例生成了最小包围边界框，能够去除多余背景信息对模型学习效果的影响，因此 AutoLNet 模型能够达到接近于全监督的实例分割效果。

表 5 AutoLNet 与弱监督模型（DiscoBox 和 Box2Mask）的平均精度对比

Table 5 Comparison of average precision between AutoLNet and weak supervised models (DiscoBox and Box2Mask) （%）

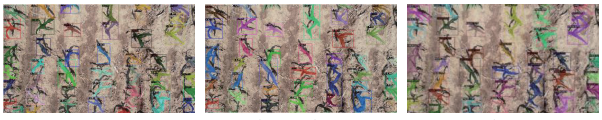
模型 Model	主干网络 Backbone network	评价类型 Evaluation type	AP50	AP55	AP60	AP65	AP70	AP75	AP_L	AP
DiscoBox	ResNet50+FPN	预测框	83.21	79.56	76.52	74.32	69.89	67.78	58.04	57.41
		掩膜	73.26	72.54	56.52	31.90	7.24	0.73	21.11	22.32
Box2Mask	ResNet50+FPN	预测框	87.34	84.13	80.32	77.71	75.22	66.33	51.24	59.90
		掩膜	78.17	70.22	56.67	27.23	5.94	0.34	22.21	23.81
	ResNet101+FPN	预测框	86.31	82.84	79.62	76.20	71.31	64.02	60.24	57.08
		掩膜	78.83	70.09	55.41	31.23	7.44	0.84	24.22	24.35
AutoLNet	ResNet50+BiFPN	预测框	96.39	86.79	81.96	80.93	79.67	73.67	68.94	68.69
		掩膜	91.75	81.92	78.17	57.31	36.62	12.03	35.44	35.07

表 6 AutoLNet 与全监督模型 (CondInst 和 Mask R-CNN) 的平均精度对比

Table 6 Comparison of average precision between AutoLNet and fully supervised models(CondInst and Mask R-CNN) （%）

模型 Model	主干网络 Backbone network	评价类型 Evaluation type	AP50	AP55	AP60	AP65	AP70	AP75	AP_L	AP
CondInst	ResNet50+FPN	预测框	94.31	92.82	91.01	88.76	83.41	70.11	64.85	63.93
		掩膜	87.94	79.58	70.93	56.04	33.02	22.20	42.27	38.91
	ResNet101+FPN	预测框	95.73	94.42	92.35	90.89	87.03	75.50	66.55	65.80
		掩膜	92.15	82.61	73.10	58.71	36.64	20.08	45.69	41.05
	ResNet101+BiFPN	预测框	97.50	93.53	92.55	90.77	88.02	85.94	73.56	72.83
		掩膜	95.14	83.11	73.23	60.02	38.67	20.66	45.82	42.18
Mask R-CNN	ResNet50+FPN	预测框	92.71	91.65	89.44	85.32	79.50	60.78	59.00	59.01
		掩膜	84.14	77.90	62.32	40.53	17.03	2.26	28.73	28.40
	ResNet101+FPN	预测框	94.03	92.32	90.81	87.20	82.92	64.26	61.25	61.15
		掩膜	86.65	81.71	69.30	49.44	23.40	2.56	32.01	31.79
AutoLNet	ResNet50+BiFPN	预测框	96.39	86.79	81.96	80.93	79.67	73.67	68.94	68.69
		掩膜	91.75	81.92	78.17	57.31	36.62	12.03	35.44	35.07

图 9 为 AutoLNet 与全监督实例分割模型 CondInst 和 Mask R-CNN 的分割效果对比。从图中可以看出，对于大田环境下无人机拍摄的玉米苗期图像，AutoLNet 与全监督模型的分割结果非常接近。AutoLNet 的标签自生成模块能够替代样本标签的人工标注过程，降低了人工时间成本，并能准确分割小目标的苗期玉米植株，去除图像中由于光照产生的植株影子，避免影响对玉米植株的分割，达到无需标注的高精度实例分割。



a. AutoLNet 实例分割结果 a. Instance segmentation results of AutoLNet
b. CondInst 实例分割结果 b. Instance segmentation results of CondInst
c. Mask R-CNN 实例分割结果 c. Instance segmentation results of mask R-CNN

图 9 不同模型的分割效果对比

Fig.9 Comparison of segmentation effects of different models

4 结 论

本研究设计了一种基于自生成标签的弱监督实例分割模型 AutoLNet，利用大田场景下玉米苗期图像的色彩信息自生成边界框标签作为弱监督训练样本，最终实现 AutoLNet 模型的训练。本文研究主要得到以下结论：

1）所设计标签自生成模块利用颜色空间转换、轮廓跟踪和最小外接矩形在玉米苗期图像中自动生成目标前景的边界框（自生成标签），通过自生成标签代替弱监督模型中的人工标签，其标签精度达到 95.23%。

2）设计了基于自生成边界框标签的弱监督学习模型，在弱监督模型的基础上优化了主干网络，利用 ResNet-50+BiFPN 提高特征提取能力，以动态卷积滤波器在全图上动态地为每一个实例生成一个掩膜，在基于边界框标签的弱监督模型中对形态结构复杂的玉米苗期图像实现

高精度实例分割，在主干网络为 ResNet50+BiFPN 的 AutoLNet 模型中，自生成标签相较于人工标签在预测框与掩膜的平均精度上分别高出 10.83 和 3.42 个百分点；与 DiscoBox 和 Box2Mask 弱监督模型相比，AutoLNet 的预测框精度分别高 11.28 和 8.79 个百分点，掩膜精度分别高 12.75 和 10.72 个百分点。

3）在玉米苗期图像实例分割任务中，AutoLNet 与全监督模型的分割效果相似。AutoLNet 可以在训练中通过降低投影损失与成对损失实现弱监督学习精度的提高，在模型表现最好的主干网络下，AutoLNet 的预测框与掩膜精度可以达到 CondInst 模型的 94.32%（其中 AutoLNet 的 AP 值为 68.69%，CondInst 的 AP 值为 72.83%）与 83.14%（其中 AutoLNet 的 AP 值为 35.07%，CondInst 的 AP 值为 42.18%）；而对比 Mask R-CNN 模型，AutoLNet 的预测框精度高 7.54 个百分点，掩膜精度高 3.28 个百分点。

综上所述，针对无人机玉米苗期图像（顶视图）的实例分割任务，AutoLNet 模型可以在无人工标签的条件下实现高精度实例分割。与同样基于边界框标签的弱监督模型（DiscoBox 和 Box2Mask）相比，AutoLNet 模型得到的掩膜精度和预测框精度都高于 DiscoBox 和 Box2Mask。而在交并比阈值大于等于 0.5 的前提下，AutoLNet 的分割效果要优于全监督模型 Mask R-CNN，且与 CondInst 接近。因此，利用 AutoLNet 可以实现在大田环境下玉米苗期图像的实例分割，模型利用标签自生成模块代替手动标注过程，节省了大量的人工标注成本，可为大田环境下的玉米苗期图像实例分割任务提供解决方案和技术支持。

[参　考　文　献]

[1] 严佳豪，彭辰晨，陈超凡，等. 基于机器视觉的植物表型研究现状[J]. 南方农机，2021，52(08)：195-196.
YAN Jiahao, PENG Chenchen, CHEN Chaofan, et al. Research status of plant phenotyping based on machine vision[J]. China Southern Agricultural Machinery, 2021, 52(08): 195-196. (in Chinese with English abstract)

[2] 张彦斐，刘茗洋，宫金良，等. 基于两级分割与区域标记梯度 Hough 圆变换的苹果识别[J]. 农业工程学报，2022，38(19)：110-121.
ZHANG Yanfei, LIU Mingyang, GONG Jinliang, et al. Apple recognition based on two-level segmentation and region-marked gradient Hough circle transform[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(19): 110-121. (in Chinese with English abstract)

[3] 王礼，方陆明，陈珣，等. 基于 Lab 颜色空间的花朵图像分割算法[J]. 浙江万里学院学报，2018，31(3)：67-73.
WANG Li, FANG Luming, CHEN Xun, et al. Flower image segmentation algorithm based on Lab color space[J]. Journal of Zhejiang Wanli University, 2018, 31(3): 67-73. (in Chinese with English abstract)

[4] 陈科尹，吴崇友，关卓怀，等. 基于统计直方图 k-means 聚类的水稻冠层图像分割[J]. 江苏农业学报，2021，37(6)：1425-1435.
CHEN Keyin, WU Chongyou, GUAN Zhuohuai, et al. Rice canopy image segmentation based on statistical histogram k-means clustering[J]. Jiangsu Journal of Agricultural Sciences, 2021, 37(6): 1425-1435. (in Chinese with English abstract)

[5] 束美艳，魏家玺，周也莹，等. 基于朴素贝叶斯分类的柑橘叶片溃疡病诊断[J]. 浙江大学学报（农业与生命科学版），2021，47(4)：429-438.
SHU Meiyan, WEI Jiaxi, ZHOU Yeying, et al. Diagnosis of citrus leaf canker disease based on naive Bayesian classification[J]. Journal of Zhejiang University(Agriculture and Life Sciences), 2021, 47(4): 429-438. (in Chinese with English abstract)

[6] YU H L, SONG J M, CHEN C C, et al. Image segmentation of Leaf Spot Diseases on Maize using multi-stage Cauchy-enabled grey wolf algorithm[J]. Engineering Applications of Artificial Intelligence, 2022, 109: 104653.

[7] 苏丽，孙雨鑫，苑守正. 基于深度学习的实例分割研究综述[J]. 智能系统学报，2022，17(1)：16-31.
SU Li, SUN Yuxin, YUAN Shouzhen. A survey of instance segmentation research based on deep learning[J]. CAAI Transactions on Intelligent Systems, 2022, 17(1): 16-31. (in Chinese with English abstract)

[8] HARIHARAN B, ARBELAEZ P, GIRSHICK R, et al. Simultaneous detection and segmentation[C]// European Conference on Computer Vision, Cham: Springer, 2014.

[9] CHEN H, SUN K Y, TIAN Z, C, et al. BlendMask: Top-down meets bottom-up for instance segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition, Seattle: IEEE, 2020.

[10] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]//IEEE International Conference on Computer Vision. Venice: IEEE, 2017.

[11] TIAN Z, SHEN C H, CHEN H. Conditional convolutions for instance segmentation[C]// European Conference on Computer Vision, Cham: Springer, 2020.

[12] WANG X L, KONG T, SHEN C H, et al. SOLO: Segmenting objects by locations[C]// European Conference on Computer Vision, Cham: Springer, 2020.

[13] 邓寒冰，许童羽，周云成，等. 基于深度掩码的玉米植株图像分割模型[J]. 农业工程学报，2021，37(18)：109-120.
DENG Hanbing, XU Tongyu, ZHOU Yuncheng, et al. Segmentation model for maize plant images based on depth mask[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(18): 109-120. (in Chinese with English abstract)

[14] 邓颖，吴华瑞，朱华吉. 基于实例分割的柑橘花朵识别及花量统计[J]. 农业工程学报，2020，36(7)：200-207.
DENG Ying, WU Huarui, ZHU Huaji. Recognition and counting of citrus flowers based on instance segmentation[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(7): 200-207. (in Chinese with English abstract)

[15] 孙红，乔金博，李松，等. 基于深度学习的玉米拔节期冠层识别[J]. 农业工程学报，2021，37(21)：53-61.
SUN Hong, QIAO Jinbo, LI Song, et al. Recognition of the maize canopy at the jointing stage based on deep learning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(21): 53-61. (in Chinese with English abstract)

[16] 王璨，武新慧，张燕青，等. 基于双注意力语义分割网络的田间苗期玉米识别与分割[J]. 农业工程学报，2021，37(9)：211-221.
WANG Can, WU Xinhui, ZHANG Yanqing, et al. Recognition and segmentation of maize seedlings in field based on dual attention semantic segmentation network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(9): 211-221. (in Chinese with English abstract)

[17] TURGUT K, DUTAGACI H, ROUSSEAU D. RoseSegNet: An attention-based deep learning architecture for organ segmentation of plants[J]. Biosystems Engineering, 2022, 221.

[18] ZENKL R, TIMOFTE R, KIRCHGESSNER N, et al. Outdoor plant segmentation with deep learning for high-throughput field phenotyping on a diverse wheat dataset [J]. Frontiers in Plant Science, 2021, 12: 774068.

[19] 谢新林，尹东旭，续欣莹，等. 基于图像级标签的弱监督图像语义分割综述[J]. 太原理工大学学报，2021，52(6)：894-906.
XIE Xinlin, YIN Dongxu, XU Xinying, et al. A survey of weakly-supervised image semantic segmentation based on image-level labels[J]. Journal of Taiyuan University of Technology, 2021, 52(6): 894-906. (in Chinese with English abstract)

[20] JIWOON A, SUHA K. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City: IEEE, 2018.

[21] JIWOON A, SUNGHYUN C, SUHA K. Weakly supervised learning of instance segmentation with inter-pixel relations[C]// IEEE Conference on Computer Vision and Pattern Recognition , Long Beach: IEEE, 2019.

[22] JUNGBEOM L, EUNJI K, SUNGMIN L, et al. Ficklenet: Weakly and semi-supervised semantic image segmentation using stochastic inference[C]// IEEE Conference on Computer Vision and Pattern Recognition , Long Beach: IEEE, 2019.

[23] DAI J F, HE K M, SUN J. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation[C]// IEEE International Conference on Computer Vision, Santiago: IEEE, 2015.

[24] ANNA K, RODRIGO B, JAN H, et al. Simple does it: Weakly supervised instance and semantic segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition , Honolulu: IEEE, 2017.

[25] SONG C F, HUANG Y, OUYANG W L, et al. Box-driven class-wise region masking and filling rate guided loss for weakly supervised semantic segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition, Long Beach: IEEE, 2019.

[26] LAN S Y, YU Z D, CHOY C, et al. Discobox: Weakly supervised instance segmentation and semantic correspondence from box supervision[C]// IEEE International Conference on Computer Vision , Montreal: IEEE, 2021.

[27] WANG X L, YU Z D, De M, et al. FreeSOLO: Learning to segment objects without annotations[C]// IEEE Conference on Computer Vision and Pattern Recognition, New Orleans: IEEE, 2022.

[28] 赵亚楠，邓寒冰，刘婷，等. 基于弱监督学习的玉米苗期植株图像实例分割方法[J]. 农业工程学报，2022，38(19)：143-152.
ZHAO Yanan, DENG Hanbing, LIU Ting, et al. Instance segmentation method of seedling maize plant images based on weak supervised learning[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(19): 143-152. (in Chinese with English abstract)

[29] ZHUANG Q, SHI J , SHI F H. Phenotype tracking of leafy greens based on weakly supervised instance segmentation and data association[J]. Agronomy, 2022, 12(7): 1567.

[30] 周云成，邓寒冰，许童羽，等. 基于稠密自编码器的无监督番茄植株图像深度估计模型[J]. 农业工程学报，2020，36(11)：182-192.
ZHOU Yuncheng, DENG Hanbing, XU Tongyu, et al. Unsupervised deep estimation modeling for tomato plant image based on dense convolutional auto-encoder[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(11): 182-192. (in Chinese with English abstract)

[31] LU Z G, QI L J, ZHANG H, et al. Image segmentation of UAV fruit tree canopy in a natural illumination environment[J]. Agriculture, 2022, 12(7): 1039.

[32] SUZUKI S. Topological structural analysis of digitized binary images by border following[J]. Computer Vision, Graphics, and Image Processing, 1985, 30(1): 32-46.

[33] TIAN Z, SHEN C H, WANG X L. BoxInst: High-performance instance segmentation with box annotations[C]// IEEE Conference on Computer Vision and Pattern Recognition, Nashville: IEEE, 2021.

Instance segmentation model of maize seedling images based on automatic generated labels

ZHAO Lulu[1], DENG Hanbing[1,2※], ZHOU Yuncheng[1,2], MIAO Teng[1,2], ZHAO Kai[1], YANG Jing[1], ZHANG Yufeng[1]

(1. *College of Information and Electrical Engineering*, *Shenyang Agricultural University*, *Shenyang* 110866, *China*;
2. *Liaoning Engineering Research Center for Information Technology in Agricultural*, *Shenyang* 110866, *China*)

Abstract: Image segmentation has been widely used for the rapid and accurate detection of plants in the various robots of modern agriculture in recent years. However, fully supervised learning cannot obtain the sufficient, effective and low-cost mask labels (manual labeling) as training samples in the segmentation task of plant image instances, due to the diversity of plant species and forms. In this study, an automatic labelling-based instance segmentation network (AutoLNet) was proposed to improve the segmentation accuracy. The weak tags were also used to train the weak supervised deep learning model. Finally, the network model was used for the image segmentation of maize seedling stage. The top view of maize seedling stage was collected by unmanned aerial vehicle (UAV). Data enhancement was then used to improve the sample diversity. A weak label self-generation module was added in front of the backbone network using the weak supervised instance segmentation model. As such, the module was composed of color space conversion, contour tracking, and the minimum peripheral rectangle. The color threshold range of corn plants was firstly set to remove the background area of the image, in order to eliminate the influence of ground shadow and land on the foreground information. The foreground corn plant area was also expanded to remove the small noise points for the binary image with only foreground corn plants. Secondly, the edge detection was carried out on the binary image after threshold segmentation. The contour point set was then set for the foreground corn plants. Finally, the minimum peripheral rectangle of the foreground object was generated automatically in the original image using the coordinates of the contour point set. The final boundary frame was obtained to filter the threshold value. The weak label was generated automatically. The weak tags were used instead of manual tags to participate in network training. The image instance segmentation of maize seedling stage was realized without the manual tags, which was greatly reduced the labor cost that required for data annotation. The test results showed that the distance intersection ratio and cosine similarity between the self-generated and manual tags reached 95.23% and 94.10%, respectively. The quality of the tags was fully met the high requirements of weak supervision training. The average accuracy of AutoLNet's output prediction frame and mask reached 68.69% and 35.07%, respectively. By contrast, the average accuracy of Autolnet's output prediction frame and mask increased by 10.83 and 3.42 percentage points, respectively, compared with the manual label models (DiscoBox and Box2Mask). The average accuracy of the forecast frame increased by 11.28 and 8.79 percentage points, respectively, whereas, that of the mask increased by 12.75 and 10.72 percentage points, respectively. The accuracy of weakly supervised learning was improved to reduce the projection and paired loss during training in the AutoLNet, compared with the fully supervised model (CondInst and Mask R-CNN). The average accuracy of prediction frame and mask in AutoLNet reached 94.32% and 83.14% of the CondInst model, 7.54 and 3.28 percentage points higher than those of prediction frame and mask R-CNN mode. Once the intersection ratio threshold was greater than or equal to 0.5, the segmentation effect of AutoLNet was better than that of the fully supervised model Mask R-CNN, similar to the CondInst. Consequently, the improved AutoLNet can be expect to automatically obtain the corn plant labels in the image using the label self-generation module. Manual labeling process was improved using the label self-generation module. Case segmentation of corn seedling images was realized for the cost saving without manual labeling. The finding can provide the solution and technical support to the high precision and low-cost segmentation task of maize seedling image instance in field environment.

Keywords: image processing; deep learning; instance segmentation; weakly supervised learning; maize seedling; plant phenotype