

# 基于 UANP-MT 的半监督菜心杂草分割方法

蔡雨霖<sup>1</sup>, 肖佳仪<sup>1</sup>, 余超然<sup>2\*</sup>, 宋 钊<sup>2</sup>, 李 静<sup>2</sup>, 岳学军<sup>1</sup>

(1. 华南农业大学电子工程学院/人工智能学院, 广州, 510642; 2. 广东省农业科学院蔬菜研究所/  
广东省蔬菜新技术研究重点实验室, 广州, 510640)

**摘 要:** 杂草作为一种常见的农业问题, 对农作物的生长造成比较严重的影响, 控制和管理杂草是农业生产活动中的重要一环。近年来, 随着无人机技术和人工智能技术的快速发展, 基于无人机平台的特定区域杂草管理是目前除草作业的主流研究, 而精确高效地对田间杂草进行识别和检测是实现自动化杂草管理的重要前提。但高效的识别模型往往意味着大量的农业数据。为了降低对农业标签数据的依赖性, 该研究提出了一种 UANP-MT (uncertainty aware and network perturbed mean teacher) 的半监督语义分割网络。该模型基于 PSPNet 结构与 MT (mean teacher) 的思想, 首先通过对教师网络做扩增输出, 令该部分做出若干次推理并取其均值, 以此来保证网络预测的鲁棒性, 其次在网络的一致性学习部分构建不确定性系数来约束不同网络间的输出差异, 提高预测的置信度和可靠性, 从而提高模型的识别准确度。为了验证所提出的模型的有效性, 设计消融试验, 包括对网络参数的取值设置, 特征提取网络 backbone 的选取, 以及在不同数据集量的数据集上对模型进行性能测试, 试验过程中确定了模型的一些最佳的参数设置。结果表明, 在与原监督网络的对比试验中, 在所提出的 UANP-MT 模型在标签数据低于原监督网络的前提下, 其 F1 分数, 像素精确度 PA (pixel accuracy) 以及交并比 Iou (intersection over union) 3 个评估指标或皆比原监督网络更高, 性能更优。此外, 在与常用的语义分割模型的对比中, UANP-MT 也体现出了其优越性, 在 1/4 数据集的标签数据量参与训练的情况下 F1 分数为 81.83%, 像素准确度为 95.84%, 交并比为 90.70%。评估指标分别优于次之的 Deeplabv3+ 模型 4.71, 7.94, 8.27 个百分点。该模型能够较好地在低标签数据量情况下对杂草数据集做出高质量的检测和识别, 极大地减少物力和时间成本, 对后续开发无人机平台的自动化除草作业有一定的参考作用。

**关键词:** 图像处理; 杂草识别; 半监督学习; 语义分割; 无人机平台

doi: 10.11975/j.issn.1002-6819.202302096

中图分类号: TP391.41

文献标志码: A

文章编号: 1002-6819(2023)-11-0183-09

蔡雨霖, 肖佳仪, 余超然, 等. 基于 UANP-MT 的半监督菜心杂草分割方法[J]. 农业工程学报, 2023, 39(11): 183-191.

doi: 10.11975/j.issn.1002-6819.202302096 <http://www.tcsae.org>

CAI Yulin, XIAO Jiayi, YU Chaoran, et al. UANP-MT based semi-supervised image segmentation method for identifying weeds in cabbage field[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(11): 183-191. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202302096 <http://www.tcsae.org>

## 0 引 言

农田杂草具有极强的繁殖能力, 并通过和作物争夺养分、引入作物疾病、吸引有害昆虫等方式直接或间接影响作物产量、质量, 给农业生产造成严重损失。随着农业无人机的发展, 田间低空喷施的工作得以进一步研究和广泛应用, 逐渐取代人为施药成为二次除草工作的主要方式。就目前而言无人机施药针对的是整块大田区域以全覆盖和均匀喷施的方式进行化学除草作业, 对于

田间“片状”与“点状”的不规则杂草分布规律来说, 大范围全面性的喷施难免造成药剂浪费甚至对环境产生污染<sup>[1-3]</sup>。对这种针对性不强且效率不高的传统方法存在的问题, 基于无人机平台的杂草识别与精确除草成为当前研究应用的主要方向。

在计算机视觉与图像处理领域中, 对田间杂草区域的识别涉及到不同场景区域的分割和区分问题。根据图像目标内容的不同特征, 包括颜色特征, 纹理特征, 形状特征, 空间关系特征<sup>[4-10]</sup>等, 将图像划分成若干个互不相交的区域, 使得这些区域内的像素表现出一致性或相似性, 从而能区分杂草所在区域。近年来面向无人机遥感平台的农业识别任务中, 因其极广的视野深度和图像维度, 是对农田农情信息进行整体判断的重要依据。对这些图像进行处理需要一些更加精确和细致的机器视觉方法。

而得益于深度神经网络的发展, LONG 等<sup>[11]</sup>提出了全卷积网络 FCN (fully convolutional network) 的结构, 实

收稿日期: 2023-02-16 修订日期: 2023-03-19

基金项目: 广东省重点领域研发计划项目 (2019B020214003); 广东省驻镇帮镇扶村农村科技特派员项目 (粤科函农字 [2021]1056 号); 广州市科技计划项目 (202206010088, 20212100068); 广东省农业科学院创新基金项目 (202203); 广东省农业科学院协同创新中心课题 (XTXM202201)。

作者简介: 蔡雨霖, 研究方向为无人机杂草检测与识别。

Email: yullin@stu.scau.edu.cn

\*通信作者: 余超然, 博士, 助理研究员, 研究方向为蔬菜智慧化应用研发。Email: yuchaoran@gdaas.cn

现了图像内容分割和语义识别的端到端训练,也同时奠定了语义分割算法的发展基调。对于农田杂草遥感平台识别的研究领域,构建语义分割识别模型也是目前使用较多的方法。一些经典的语义分割网络如 SegNet<sup>[12]</sup>、U-Net<sup>[13]</sup> 和 Deeplab<sup>[14]</sup> 系列等被陆续应用于水稻田,棉花地和玉米等地的杂草识别之中,皆取得了不错的分割和识别效果<sup>[15-22]</sup>。比如 GENZE 等<sup>[23]</sup> 用 FCN, U-Net 和 Deeplabv3+等语义分割网络和残差网络 ResNet<sup>[24]</sup> 系列的不同组合在高粱地上的杂草分割效果做了对比,结果表明 U-Net 与 ResNet34 的组合分割对杂草和作物的分割效果最好,其 f1-score 达到了 89%。另外,针对原始网络在特定的数据集中表现不足的问题,一些研究会对网络做适应性的改进,如 WANG 等<sup>[25-27]</sup> 对 U-Net 都做了不同程度的数据集适应性改进,以得到更准确的杂草分割模型。KIM 等<sup>[28]</sup> 基于 U-Net 的基本结构,提出了一种两阶段的杂草识别模型,并在 3 个公共数据集 (BoniRob dataset, crop/weed field image dataset (CWFID), rice seedling and weed dataset) 上做了测试,验证了两阶段模型的优越性。

但在深度学习领域,大量的标注数据能训练出高效的模型。但往往大量获取标注数据并非一件易事。一种非监督的思想便由此而来,如 Mean Teacher (MT) 方法<sup>[29]</sup>,旨在令教师网络和学生网络在未标注的数据预测上输出一致。其后续 Perturbed Strict MT<sup>[30]</sup>, Uncertainty aware MT<sup>[31]</sup> 是在 MT 基础之上的相应改进,通过增加对模型输出的约束来达到提高模型性能的效果。再如 VERMA 等<sup>[32]</sup> 提出的“插入一致性”(interpolatiuon consistency)方法,其本质上也是一种 MT 的思想。网络期望两张图像以不同权重混合之后的预测输出,能和原本各自图像不同权重预测输出加起来一致。由于农业生产活动周期普遍较短,耗费人力物力去做数据标注是一件成本高但时效性低的工作。但就目前而言,农业数据集上鲜有采用半监督方式去降低标签数据的依赖性且提高图像识别表现的研究,而是利用无监督的思想去进行数据增强。如 ZOU 等<sup>[33]</sup> 提出了一种具有无监督思想的数据增强方法,其对不同的数据集进行有效合成,最后这人工合成的数据集添加进训练网络后,在原测试集上其交并比 Iou (intersection over union) 达到了 89.45%。PICON 等<sup>[34]</sup> 则也提出了一些半监督思想的数据增强方法,其一旨在得到多种单一植株的混合合成图像,其二是在复杂农田中先对其中作物做标注,再利用背景相减得到杂草的分割图像。结合两种方法进行模型训练后,其对 7 种杂草识别的平均平衡准确度 BAC (balance accuracy) 为 75.71%。

上述方式固然能减少对农业数据的标注成本问题,且有较为不错的表现,但面对复杂的田间环境时,其局限性较大。如本文研究的菜心地田间环境,其中菜心和

杂草生长空间高度重合,较为密集,采用数据增强方式来做杂草识别较为困难,而用于监督学习的标注工作又过于庞大,所以需要研究一种高效的半监督杂草分割模型。针对这个问题,本文提出一种 UANP-MT (uncertainty aware and network perturbed mean teacher) 结构模型,通过增加教师网络的扰动来保证网络预测的鲁棒性,并利用不确定性系数来提高网络输出的置信度,从而提高网络的识别准确度和泛化能力,确保在少部分标注数据下杂草的高效识别。

## 1 材料与方法

### 1.1 数据获取与数据集构建

#### 1.1.1 数据图像获取

本文试验对象为菜心地试验田,位于广东省农业科学院蔬菜研究所大丰试验基地(113°37'E, 20°16'N)。菜心的生长周期较短,一般从播种到商品收获期在 35~45 d。本文数据采集时间为 2022 年 9 月 4 日,此时该菜心地正处于叶片生长期,也是除草的关键时期。数据获取设备为大疆“御”Mavic 2 专业版相机,1 英寸 CMOS,有效像素 2 000 万,视角 77°,等效焦距 28 mm。内置电子快门,照片分辨率为 5 472 × 3 648 像素。

数据获取过程中,设置无人机为区域自动拍摄模式,飞行高度为 5 m,飞行速度为 1.5 m/s,拍摄间隔设置为 1 s。为了避免无人机阴影给画面带来的影响,无人机飞行相机俯仰角与地面呈 45°夹角,并不垂直于地面。

最终获取 130 张尺寸为 5 472 × 3 648 像素的菜心地图像,每张图像实际分辨率为 0.1 cm/像素。

#### 1.1.2 数据集构建

数据集构建描述如图 1 所示。将一张 5 472 × 3 648 尺寸的图像分割成 64 张 684 × 456 像素的子图,减少后续训练时的内存消耗以及较大程度保证图像不失真。经去重和预处理后,最终得到 2 304 张子图。

在所有子图中,存在一些全是背景像素的图像,可分为两类,一类是土壤像素,另一类是黑色地膜像素,其中基本不包含前景部分,所以归为负样本。在深度学习训练中,正负样本数的平衡是确保模型预测性能的一大前提。过多的正样本会让模型泛化能力不足,而过多的负样本则会降低模型预测的准确度。在本文数据中,负样本数量也相对较多,所以需要剔除一些冗余的负样本数据,增加正样本数据在训练中的权重。

剔除的冗余负样本子图为 148 张,剩下 2 156 张子图则构成本次试验的全部数据。将其划分为训练集和测试集。训练集占 95%,为 2 048 张子图;测试集占 5%,为 108 张子图。

对训练集数据和测试集数据做语义范围标注,标注工具为 Labelme 3.16.7 版本,平均一张 684 × 456 像素的子图标注时间为 3~5 min。



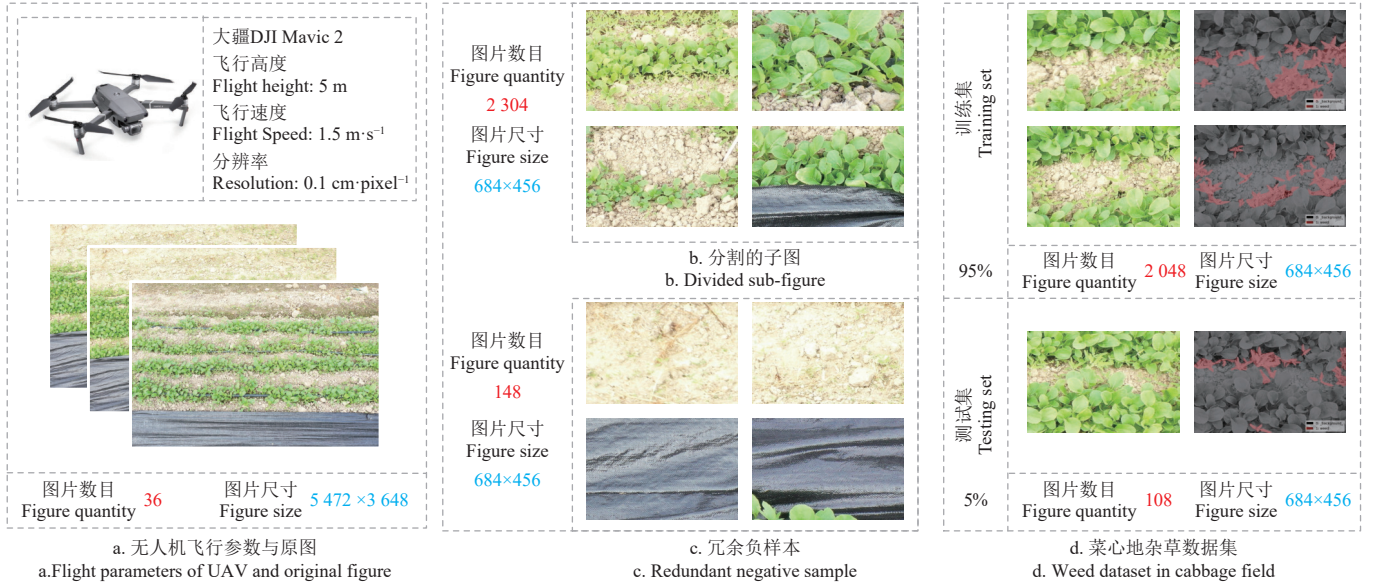


图 1 菜心地杂草数据集构建  
Fig.1 Construction of weed dataset in the cabbage field

## 1.2 语义分割算法

### 1.2.1 PSPNet

PSPNet 是 ZHAO 等<sup>[35]</sup> 提出的一种基于全局信息的语义分割模型。其网络结构如图 2 所示。输入图像在经过特征提取网络“CNN”后得到输入特征图。紧接着对该特征图做 4 种尺寸的全局池化“Pool”，得到 4 种不同尺度的池化层。图中用蓝色的立方体表示全局信息最细致的层级，其他颜色次之。为了保持全局特征的权重，采用 1×1 的卷积“Conv”对这 4 个层级的特征层做特征降维。然后通过双线性插值直接对低维特征图进行上采样“Upsampling”，使其与原始特征图尺度一致。最后，将不同层级的特征图拼接为最终的金字塔池化全局特征。

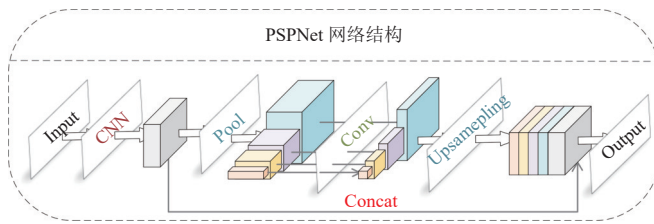


图 2 PSPNet 网络结构  
Fig.2 Network structure of PSPNet

### 1.2.2 Backbone

本文将探讨 3 种不同量级的“CNN”特征提取网络 Mobilenetv2, ResNet50 和 ResNet101 在杂草识别模型中的表现。

Mobilenetv2 是一个轻量型的卷积神经网络，其中一个模块叫 Inverted residual block, 主要包括扩展层，投影层和深度可分离卷积层。扩展层和投影层的目的是做特征升维和特征降维。而深度可分离卷积则完成卷积计算，它将 3×3 的传统卷积分成两个阶段，首先用与输入通道数相等的 3×3 卷积核对输入做卷积运算，然后再用与输出通道数相等的 1×1 卷积对上一步得到的特征图

进行通道数的调整。如此便大大降低了计算量和参数量。

ResNet 系列是近年来深度学习邻域较为主流的特征提取网络。其网络中设计了两种残差模块，其一是以两个 3×3 卷积网络串接在一起作为一个残差模块，其二是一种以 1×1, 3×3, 1×1 这 3 个卷积网络串联成为一个残差模块。ResNet 根据残差层的数量来划分不同的网络。ResNet18, 20, 34, 50, 101, 152 等。其中 ResNet50 和 ResNet101 的使用较为普遍，两者的区别主要是在网络的第四个卷积阶段残差模块数目不同。ResNet50 的浮点运算次数 Flops 为 3.8×10<sup>9</sup>, ResNet101 的浮点运算次数 Flops 为 7.6×10<sup>9</sup>。

### 1.3 Mean Teacher (MT) 网络

Mean Teacher (MT) 是一个比较经典的半监督模型。该模型由“教师网络”和“学生网络”两个网络组成，其核心思想是鼓励两个网络的输出相似，也即是半监督学习其中一个重要的思想，“预测一致性”。

MT 的网络结构如图 3 所示。其中  $x^u$  和  $x^l$  分别表示无标签数据和标签数据， $y$  则表示作为监督学习部分的 ground truth。

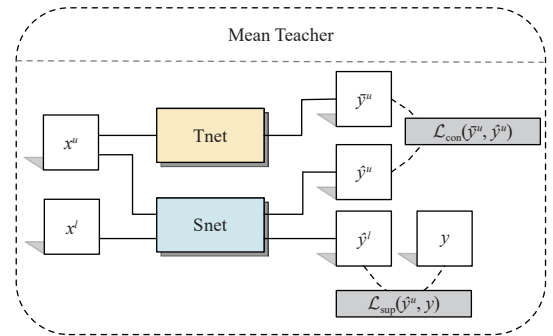


图 3 Mean Teacher (MT) 网络结构

Fig.3 Network structure of Mean Teacher (MT)

$x^l$  经过“学生网络”Snet 后得到输出  $\hat{y}^l$ ，这部分与  $y$  做监督学习，损失函数为  $\mathcal{L}_{\text{sup}}(\hat{y}^l, y)$ 。

而  $x''$  则分别输入“教师网络” Tnet 和“学生网络” Snet 得到输出  $\hat{y}''$  和  $\hat{y}'$ ，这一部分做一致性学习。损失函数为  $\mathcal{L}_{\text{con}}(\hat{y}'', \hat{y}')$ 。

所以最终 Snet 的损失函数构成为

$$\mathcal{L}_{\text{Snet}} = \mathcal{L}_{\text{sup}}(\hat{y}', y) + \mathcal{L}_{\text{con}}(\hat{y}'', \hat{y}') \quad (1)$$

Tnet 和 Snet 两个网络结构相同，在训练不断迭代的过程中，Snet 的参数通过损失函数梯度下降不断更新，Tnet 的参数则通过“学生网络”更新赋值得到。

#### 1.4 基于 UANP-MT 的杂草识别模型构建

对于 MT 的“预测一致性”思想，其首要关键点在于两个网络预测结果的可靠性。若其中一个网络的预测结果并不如期望值，那么训练过程中将会导致整个模型拟合结果变差，甚至拟合失败。由于“学生网络” Snet 有 Ground truth 实际标签做监督，所以主要对“教师网络” Tnet 部分做预测结果可靠性的提升。

本文提出了一种 UANP-MT 杂草识别模型。UANP-MT 模型相对于 MT，其主要的提升有两点。

其一是对 Tnet 部分做了扩增，使用了两个 Tnet。我们在每一个训练轮次的训练内只更新其中一个 Tnet 网络的参数，来增加两个网络差异性，然后将两个独立更新权值的教师网络的预测结果进行重复推理并平均，可以减少单一模型的随机性和偏差，从而提高一致性学习中伪标签的准确性和可靠性，能对学生网络做出更加可靠的指导，提高泛化性能，如此来进一步增强预测结果的鲁棒性。

其二是对“预测一致性”部分增加了对不确定信息的约束。在一致性预测中，两个网络对同一输入做出的预测难免出现偏差，在不断的迭代过程中，这样的偏差将对模型的“预测一致性”部分造成影响，这即是不确定性的因素。

而增加对这部分的约束则可以在计算一致性损失时只对两个网络预测偏差较小的结果进行学习，而对预测偏差较大的结果进行抑制。因此，Snet 得到了优化和更为可靠的监督，并反过来鼓励 Tnet 部分做出质量更高的预测。

UANP-MT 的模型结构如图 4 所示。“教师网络” Tnet 部分分为 PSPNet (Tnet1) 和 PSPTNet (Tnet2)，两个网络皆以 PSPNet 为基础网络，二者结构一致， $x''$  输入后分别做出  $\hat{y}_1''$  和  $\hat{y}_2''$  的预测。为了约束其预测结果的可靠性，令这两个网络在每一个训练轮次都做出  $M$  次预测，进行均值整合后形成最终 Tnet 部分的输出  $\hat{y}''$ 。

$$\hat{y}'' = \frac{1}{M} \times \sum_{i=1}^M 0.5[(\hat{y}_1'')^i + (\hat{y}_2'')^i] \quad (2)$$

因此一致性学习部分的损失函数构成为  $\mathcal{L}_{\text{con}}(\hat{y}'', \hat{y}')$ 。 $\hat{y}''$  表示  $x''$  输入 Snet 后的预测结果。监督学习部分的损失函数为  $\mathcal{L}_{\text{sup}}(\hat{y}', y)$ 。 $\hat{y}'$  表示  $x'$  输入 PSPNet (Snet) 后的预测结果。PSPNet (Snet) 同样以 PSPNet 做为基础网络，在结构上与 Tnet 部分的两个网络结构一致。

总体的损失函数构成为

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{sup}}(\hat{y}', y) + \mathcal{L}_{\text{con}}(\hat{y}'', \hat{y}') \text{mask} \quad (3)$$

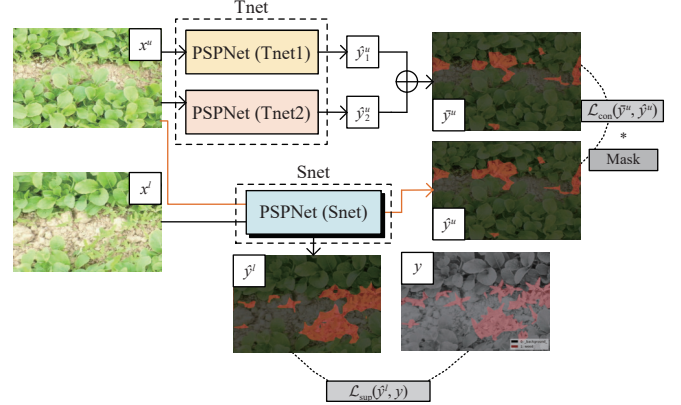


图 4 UANP-MT 杂草识别模型

Fig.4 UANP-MT weed identification model.

其中 var 表示两个网络在重复推理过程中的差异量均值，mask 用以计算 Tnet 部分目标预测结果的不确定性，作为一个约束性的系数，过滤掉不可靠的预测。其计算如下式：

$$\text{mask} = \frac{1}{1 + e^{\text{var}}} \quad (4)$$

$$\text{var} = 0.5 \times \frac{\sum_{i=1}^M [(\hat{y}_1'')^i - \bar{y}'']^2 + [(\hat{y}_2'')^i - \bar{y}'']^2}{M} \quad (5)$$

其主要思想是计算出  $M$  次预测的均方差作为不确定性系数，从而约束 Tnet 预测结果对损失函数的影响，鼓励 Tnet 做出鲁棒性强，质量高的预测。

从整体模型来说，Snet 利用  $\mathcal{L}_{\text{total}}$  做梯度下降不断更新其参数。而 Tnet 部分的参数则通过 Snet 做 EMA (exponential moving average) 赋值得到。具体计算如下：

$$W_t \leftarrow \alpha W_t + (1 - \alpha) W_s \quad (6)$$

其中  $W_t$  和  $W_s$  分别表示 Tnet 部分的权重参数和 Snet 部分的权重参数。在每个训练轮次，PSPNet (Sent) 只对其中一个 Tnet 进行 EMA 赋值，PSPNet (Tnet1) 或者 PSPNet (Tnet2)。

## 2 试验与结果分析

### 2.1 试验环境及训练参数设置

本文的所有试验均在 Intel(R) Core(TM) i9-10920X CPU @ 3.50 GHz 运行内存为 64 GB 的计算机平台完成，平台操作系统为 Window10，机身显卡型号为 NVIDIA GeForce RTX 3 090，中央处理器为 Inter(R) Core (TM) i9-10920X @ 3.50 GHz。软件环境在 window10 Pycharm 专业版客户端中搭建，采用 python 3.8.10 版本在深度学习 Tensorflow2 框架上实现。

训练时参数设置如下：输入尺寸为  $473 \times 473$ ；初始学习率为 0.01；梯度下降优化器选用 SGD，其内部 momentum 参数设置为 0.9；损失函数选用经典的交叉熵损失计算式。Batchsize 设置为 4，训练轮次为 80。

### 2.2 模型评估指标

本文试验效果的验证采用语义分割中常用的指标进行评估。在下文中，需要首先对所用参数值做出标明：其中  $i$  表示类别（文中指代杂草）， $t_i$  表示实际像素为  $i$



类的数量,  $p_i$  表示预测像素为  $i$  类的数量。  $n_{ii}$  表示预测为  $i$  实际也是  $i$  类的数量。

像素准确率  $P_A$ , 其含义为在类别  $i$  的预测中, 真实属于  $i$  类像素的准确率。计算式如下:

$$P_A = \frac{n_{ii}}{t_i} \quad (7)$$

交并比  $Iou$ , 其含义为模型对某一类别预测结果和真实值的交集与并集的比值。计算式如下:

$$Iou = \frac{n_{ii}}{t_i + p_i - n_{ii}} \quad (8)$$

F1 分数 F1 score, 其含义是查准率和查全率的综合平衡指标。计算式如下, 其中  $\beta$  设为 1。

$$F1 \text{ score} = \frac{n_{ii}}{(t_i + p_i)\beta^2} \times (1 + \beta^2) \quad (9)$$

## 2.3 结果与分析

### 2.3.1 Tnet 部分不同 $M$ 值测试

Tnet 部分的两个网络在每个训练轮次都做出  $M$  次预测, 其均值结果不仅能用于判断两个网络预测的不确定性, 且能最终让 Tnet 部分的输出更加稳定。所以, 对于  $M$  的取值需要在试验中进行测试验证。

在这一环节采用训练集的 1/2 (1 024) 数据以及它们对应的标签作为标签数据, 剩下部分 1/2 (1 024) 仅采用原始数据而无标签作为无标签数据, 进行模型训练并测试模型在测试集上的表现。特征提取网络采用 Mobilenetv2, 其轻量化的结构能极大减少训练时长, 节约时间成本。而对于  $M$ , 取了 5 个值进行测试。

测试结果如表 1 所示。从表中可以看到, 模型的性能大体上是随着  $M$  值的升高而升高, 尤其是在评估指标 F1 分数上最能体现出来, 当  $M=10$  时, F1 分数为 84.50%。但从交并比和像素准确度上看, 当  $M=8$  时模型性能则更佳, 其交并比为 92.53%, 像素准确度为 97.29%。且从总体上看, 由于网络结构一致, 所以模型帧率 Fps (frames per second) 相差无几。当设置  $M=8$  时进行训练得到的模型其 Fps 为 86.49 帧/s。

表 1 对不同  $M$  值的测试

Table 1 Model test of coefficient  $M$

$M$ 值 $M$ value	F1 分数 F1 score /%	像素准确度 Pixel accuracy /%	交并比 Intersection over union /%	帧率 Frames per second / (帧·s <sup>-1</sup> )
2	82.80	96.42	92.46	85.67
4	83.10	96.55	92.48	86.17
6	83.30	96.68	92.52	85.59
8	84.20	97.29	92.53	86.49
10	84.50	96.23	92.50	85.44

由于用来做均值拟合的数据越多, 出现不确定性数值的概率会越大, 在后续的不确定性计算有些数据不能被学习到, 导致模型性能下降。所以并非对 Tnet 部分做重复预测越多越好, 而且  $M$  值越大, 越消耗计算资源。所以综合考虑, 在后续的试验中, 都将  $M$  值设置为 8。

### 2.3.2 不同 backbone 下的模型性能测试

不同量级的特征提取网络对数据集也有很重要的影响。选取 3 种网络, 分别是 Mobilenetv2, ResNet50, ResNet101, 测试不同的 backbone 下模型的性能。

采用训练集的 1/2 (1 024) 数据以及它们对应的标签作为标签数据, 剩下部分 1/2 (1 024) 仅采用原始数据而

无标签作为无标签数据, 进行模型训练并测试模型在测试集上的表现。测试结果如表 2 所示。

表 2 不同 backbone 下的模型性能

Table 2 Performances of the model with different backbones

骨干网络 Backbone	F1 分数 F1 score /%	像素准确度 Pixel accuracy /%	交并比 Intersection over union /%	帧率 Frames per second / (帧·s <sup>-1</sup> )
Mobilenet	84.17	97.29	92.53	86.49
ResNet50	85.40	97.22	93.17	58.76
ResNet101	79.00	97.92	91.68	27.79

从表 2 中可以看出, ResNet50 的表现较为优异和稳定, 在 3 个网络中的 F1 分数和交并比都是最佳的, 分别为 85.40% 与 93.17%。且从图 5 中看出, ResNet50 的损失函数收敛更快更稳定。所以, Backbone 设置为 ResNet50 对本菜心杂草数据集来看, 更能提取和学习到关键的特征。而从推理速度上看, Mobilenetv2 因其网络的轻量化, 所以有着最高的推理速度, 其帧率达到 86.49 帧/s。ResNet50 的帧率也达到了 58.76 帧/s, 也有着较为不错的推理速度, 已然能够满足实时性检测的要求。至于量级更大的 ResNet101 表现不如人意的原因, 可能是在训练的过程中, Snet 将参数权重通过 EMA 的方式赋值给 Tnet, 这是一个参数累加的过程, 在累加初期当前值的权重小, 要随着累加变化才能体现, 所以会有一个延迟的效应。而深层网络在参数权重有限的情况下, 其学习和预测能力不能较好地发挥。所以 ResNet101 在菜心杂草数据集上表现并不如 ResNet50。在后续的试验中, 都将 backbone 设置为 ResNet50。

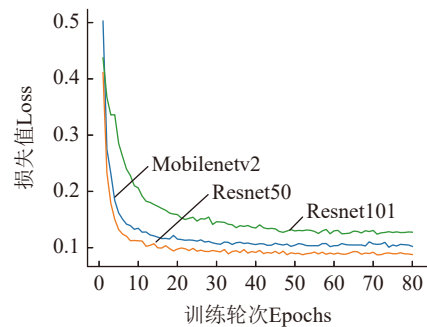


图 5 不同 backbone 下的训练集损失函数

Fig.5 Loss function of training set with different backbones

### 2.3.3 UANP-MT 模型消融试验

将 UANP-MT 拆解成独立的 UA-MT, NP-MT 和 MT 部分, 并与 UANP-MT 和原始监督网络 Sup 做消融性对比试验。对半监督网络, 将分别利用到训练集 1/2 (1 024)、1/4 (512)、1/8 (256)、1/16 (128) 数据以及它们对应的标签作为标签数据, 1/2 (1 024)、3/4 (1 536)、7/8 (1 792)、15/16 (1 920) 仅采用原始数据而无标签作为无标签数据; 对原始监督网络, 分别利用到训练集的 1/2 (1 024)、1/4 (512)、1/8 (256)、1/16 (128) 数据以及其对应的标签作为标签数据, 此部分没有无标签数据。

为了减少由随机因素引起的误差和噪声, 并提高试验结果的可靠性和稳定性, 对每一个网络进行两次模型训练, 并最终对它们在测试集上的表现进行平均取值。旨在验证不同数据量的情况下 UANP-MT 模型性能的优

异性。

试验结果如表 3 所示。可以看到, 其一, 随着标签数据量的减少, 所有模型的性能都逐渐下降。其二, 半监督模型 MT, UA-MT, NP-MT 与 UANP-MT 在 4 个不同

数据量的数据集上的表现都优于原始的监督网络。其三, 增加了推理约束的模型 UA-MT 和 NP-MT 的性能在 4 个不同数据量的数据集上皆优于 MT。最后, 在所有模型中, UANP-MT 模型在菜心杂草数据集上的性能表现最佳。

表 3 模型消融试验

Table 3 Model ablation test

模型 Model	F1 分数 F1 score /%				交并比 Intersection over union /%			
	1/2 (1 024)	1/4 (512)	1/8 (256)	1/16 (128)	1/2 (1 024)	1/4 (512)	1/8 (256)	1/16 (128)
Sup	80.80	75.54	67.12	58.78	86.46	80.81	73.26	59.83
MT	81.70	78.20	70.28	63.50	90.70	84.30	79.80	65.34
UA-MT	84.90	81.60	73.20	66.24	91.42	88.51	83.75	66.73
NP-MT	84.70	81.30	73.45	66.29	92.00	88.49	84.20	66.70
UANP-MT	85.40	81.83	75.50	67.50	93.17	90.70	87.43	68.91

如在 1/2 (1 024) 数量的标签数据情况下, 原监督网络 Sup 在测试集上的表现为 F1 分数为 80.80%, 交并比为 86.46%。而 UANP-MT 模型在只有 1/4 (512) 数量的标签数据情况下, 其在测试集上的表现已为 F1 分数为 81.83%, 交并比为 90.70%, 皆优于 Sup。说明 UANP-MT 方法对标签数据的利用性较高, 在低于正常标签数据量的情况下, UANP-MT 模型的表现甚至还优于原始网络, 验证了其在杂草检测上的有效性。

#### 2.3.4 语义分割算法对比

为了进一步验证 UANP-MT 的有效性, 选用 SegNet, U-Net, Deeplabv3+ 与 UANP-MT 进行算法性能对比。这 3 网络模型的原理和方法在不同的领域得到了广泛的验证和探究, 且近几年亦被陆续地应用于杂草检测领域并取得了不错地分割和识别效果, 用以对比模型有着较好的参考价值。

将每个模型的 backbone 特征提取网络都设置为 ResNet50, 固定输入尺寸为 473×473; 初始学习率为 0.01。Batchsize 皆设置为 4, 训练轮次为 80。在相同的试验条件下分别测试各个模型的表现。

在这一环节中, 对半监督网络 UANP-MT, 利用 1/4 (512) 数据以及它们对应的标签作为标签数据, 剩下

部分 3/4 (1 536) 仅采用原始数据而无标签作为无标签数据; 对其余语义分割网络, 利用 1/4 (512) 数据以及它们对应的标签作为标签数据, 此部分没有无标签数据。测试结果如表 4 所示。

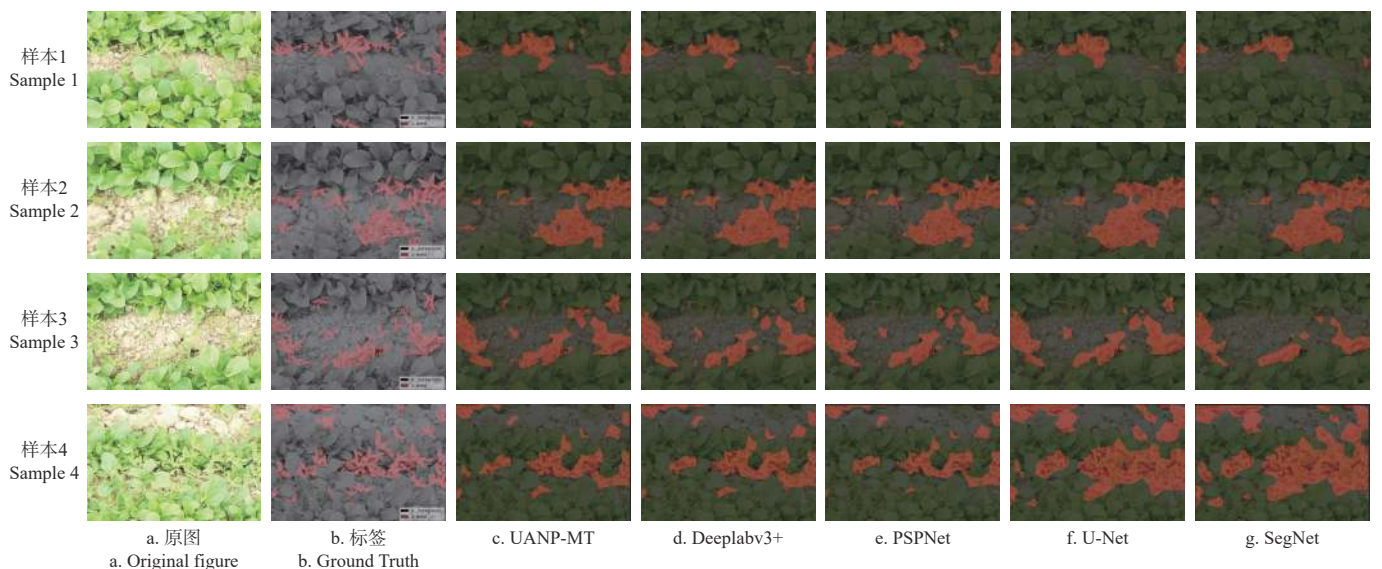
表 4 不同语义分割模型对比

Table 4 Comparison of different semantic segmentation models

网络 Network	F1 分数 F1 score /%	像素准确度 Pixel accuracy /%	交并比 Intersection over union /%	帧率 Frames per second / (帧·s <sup>-1</sup> )
SegNet	74.33	83.97	81.04	70.52
U-Net	75.22	85.01	81.45	69.71
Deeplabv3+	77.12	87.90	82.43	38.04
PSPNet	75.54	84.95	80.81	55.70
UANP-MT	81.83	95.84	90.70	56.48

在所有模型中, UANP-MT 表现最佳, F1 分数为 81.83%, 像素准确度为 95.84%, 交并比为 90.70%。这几个评估指标分别优于次之的 Deeplabv3+ 模型 4.71, 7.94, 8.27 个百分点, 且比原始模型提高 6.29, 10.89, 9.89 个百分点。从推理速度上看, 由于 SegNet 和 U-Net 网络本身的轻量性, 所以推理速度相对较高。而 UANP-MT 在识别准确度较高的前提下, 也有着不错的推理速度, 其帧率为 56.48 帧/s。

以上 5 种模型在数据集上的实际预测情况如图 6 所示。



注: 图中红色区域表示杂草像素, 灰色区域表示背景像素。

Note: In the figure, the red area represents weed pixels, and the gray area represents background pixels.

图 6 不同模型的预测结果

Fig.6 Predictive results of different models



从图 6 中可以看到, 其一, 所有模型都基本能将数据集中地杂草区域检测出来。其二, 在数据量为 1/4 (512) 的情况下, 所有模型的实际预测细致度都不高, 只有一个粗略的轮廓, 不能很好地勾勒细节。其三, 半监督模型 UANP-MT 比其它 4 个监督模型来说像素涵盖面积更广, 精度更高且细节部分更丰富一些。这也进一步验证了 UANP-MT 模型在菜心地杂草数据集上的有效性。

### 3 结 论

为了减少杂草识别模型对农业数据集的依赖性, 本文提出了一种 UANP-MT 的半监督学习方法。该方法基于 MT 的思想, 扩增了“教师网络”Tnet, 对其输出进行均值整合, 且增加了不确定性系数来判断“一致性学习”过程中的预测可靠性。通过增加对网络推理的约束来提高模型的性能, 实现了低标签数据量下模型对杂草的高效识别。

1) 本文提出了一种 UANP-MT 的杂草识别模型。对不同的特征提取网络的应用进行了测试, 结果证明在菜心杂草数据集上, ResNet50 更适合该模型。在 Tnet 部分, 测试了系数  $M$  不同取值情况下模型的性能表现, 当  $M$  等于 8 时结果较好。其交并比为 92.53%, 像素准确度为 97.29%。

2) 对于 UANP-MT 模型做了消融试验。结果表明半监督学习的表现皆优于原监督网络 Sup, 而 UANP-MT 表现最佳, 其 F1 分数为 85.40%, 交并比为 93.17%。

3) 与较为常用的语义分割模型 SegNet, U-Net, Deeplabv3 的对比, 进一步验证了 UANP-MT 模型的高效性。在只有 1/4 (512) 标签数据的情况下, 其 F1 分数为 81.83%, 像素准确度为 95.85%, 交并比为 90.70%。分别比原始模型提高 6.29, 10.89, 9.89 个百分点。

本文的研究成果对于田间杂草识别上具有一定借鉴意义, 是一个有效降低对农业数据集依赖性, 减少物力和时间成本的方法。我们对使用该方法在不同时期的菜心杂草数据集也做了一些研究, 发现在源数据集训练后的模型在不同时期的目标数据集中表现平平, 具体表现在源模型能基本识别目标数据集中的杂草所在区域, 但对杂草像素的预测和还原都比较有限, 精度不高。今后拟采用领域自适应的方法来提高模型的泛化能力。并解决杂草识别工作中现存的一些问题, 实时部署使用, 实现基于无人机平台的自动化特定区域杂草管理。

### 【参 考 文 献】

- [1] 王璨, 武新慧, 张燕青, 等. 基于移位窗口 Transformer 网络的玉米田间场景下杂草识别[J]. 农业工程学报, 2022, 38(15): 133-142.  
WANG Can, WU Xinhui, ZHANG Yanqing, et al. Recognizing weeds in maize fields using shifted window transformer network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(15): 133-142. (in Chinese with English abstract)
- [2] 赵辉, 曹宇航, 岳有军, 等. 基于改进 DenseNet 的田间杂草识别[J]. 农业工程学报, 2021, 37(18): 1136-142.  
ZHAO Hui, CAO Yuhang, YUE Youjun, et al. Field weed recognition based on improved DenseNet[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(18): 1136-142. (in Chinese with English abstract)
- [3] 任锦芬. 基于图像处理的麦田杂草识别方法研究[D]. 西安: 西京学院, 2020.  
REN Jinfen. Image Processing based Weed Identification Method in Wheat Field[D]. Xi'an: Xijing University, 2020. (in Chinese with English abstract)
- [4] SUN L, LUO B, LIU T, et al. Algorithm of adaptive fast clustering for fish swarm color image segmentation[J]. IEEE Access, 2019, 7: 178753-178762.
- [5] CHEN P, LIANG F. Cotton nitrogen nutrition diagnosis based on spectrum and texture feature of images from low altitude unmanned aerial vehicle[J]. Scientia Agricultura Sinica, 2019, 52(13): 2220-2229.
- [6] 吴兰兰, 熊利荣, 彭辉. 基于 RGB 植被指数的大田油菜图像分割定量评价[J]. 华中农业大学学报, 2019, 38(2): 109-113.  
WU Lanlan, XIONG Lirong, PENG Hui. Quantitative evaluation of in-field rapeseed image segmentation based on RGB vegetation indices[J]. Journal of Huazhong Agricultural University, 2019, 38(2): 109-113. (in Chinese with English abstract)
- [7] 苗荣慧, 杨华, 武锦龙, 等. 基于图像分块及重构的菠菜重叠叶片与杂草识别[J]. 农业工程学报, 2020, 36(4): 178-184.  
MIAO Ronghui, YANG Hua, WU Jinlong, et al. Weed identification of overlapping spinach leaves based on image sub-block and reconstruction[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(4): 178-184. (in Chinese with English abstract)
- [8] 赵静, 李志铭, 鲁力群, 等. 基于无人机多光谱遥感图像的玉米田间杂草识别[J]. 中国农业科学, 2020, 53(8): 1545-1555.  
ZHAO Jing, LI Zhiming, LU Liquan, et al. Weed identification in maize field based on multi-spectral remote sensing of unmanned aerial vehicle[J]. China Agricultural Science, 2020, 53(8): 1545-1555. (in Chinese with English abstract)
- [9] 杭艳红, 苏欢, 于滋洋, 等. 结合无人机光谱与纹理特征和覆盖度的水稻叶面积指数估算[J]. 农业工程学报, 2021, 37(9): 64-71.  
HANG Yanhong, SU Huan, YU Ziyang, et al. Estimation of rice leaf area index combining UAV spectrum, texture features and vegetation coverage[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(9): 64-71. (in Chinese with English abstract)
- [10] 任龙龙, 冯涛, 翟传龙, 等. 基于 MATLAB 图像处理的苹果大小、颜色、圆形成度及缺陷度特征融合分级研究[J]. 数字技术与应用, 2021, 39(7): 90-95.  
REN Longlong, FENG Tao, ZHAI Chuanlong, et al. Research on feature fusion grading of apple size, color, roundness and defect based on MATLAB image processing[J]. Digital Technology and Application, 2021, 39(7): 90-95. (in Chinese with English abstract)
- [11] LONG J, SHELHAMER E, DARRELL T. Fully convolutional

- networks for semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 39(4): 640-651.
- [12] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
- [13] RONNEBERGER O, FISCHER P, BROX T, et al. U-net: Convolutional networks for biomedical image segmentation[C]// *Medical Image Computing and Computer-assisted Intervention (MICCAI)*, 2015: 234-241.
- [14] CHEN L, PAPANDREOU G, KOKKIONS L, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40(4): 834-848.
- [15] 邓泓, 杨滢婷, 刘兆朋, 等. 基于深度学习的无人机水田图像语义分割方法[J]. *中国农机化学报*, 2021, 42(10): 165-172.  
DENG Hong, YANG Yingting, LIU Zhaopeng, et al. Semantic segmentation of paddy image by UAV based on deep learning[J]. *Journal of Chinese Agricultural Mechanization*, 2021, 42(10): 165-172. (in Chinese with English abstract)
- [16] 彭明霞, 夏俊芳, 彭辉. 融合 FPN 的 Faster R-CNN 复杂背景下棉田杂草高效识别方法[J]. *农业工程学报*, 2019, 35(20): 202-209.  
PENG Mingxia, XIA Junfang, PENG Hui. Efficient recognition of cotton and weed in field based on Faster R-CNN by integrating FPN[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2019, 35(20): 202-209. (in Chinese with English abstract)
- [17] 朱逸航, 张延宁, 张小敏, 等. 基于语义分割的棉花垄间实时道路识别[J]. *浙江农业科学*, 2021, 62(9): 1721-1725.  
ZHU Yihang, ZHANG Yanning, ZHANG Xiaomin. Semantic segmentation based real-time road recognition between cotton ridges[J]. *Zhejiang Agricultural Science*, 2021, 62(9): 1721-1725. (in Chinese with English abstract)
- [18] 李玉田. 基于无人机视觉的油菜苗期田间杂草识别研究[D]. 长沙: 湖南农业大学, 2019.  
LI Yutian. UAV Vision Based Identification of Weeds in Rape Seedling Field[D]. Changsha: Hunan Agricultural University, 2019. (in Chinese with English abstract)
- [19] 孙俊, 谭文军, 武小红, 等. 多通道深度可分离卷积模型实时识别复杂背景下甜菜与杂草[J]. *农业工程学报*, 2019, 35(12): 184-190.  
SUN Jun, TAN Wenjun, WU Xiaohong, et al. Real-time recognition of sugar beet and weeds in complex backgrounds using multi-channel depth-wise separable convolution model[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2019, 35(12): 184-190. (in Chinese with English abstract)
- [20] HUANG H, DENG J. A fully convolutional network for weed mapping of unmanned aerial vehicle imagery[J]. *PLOS ONE*, 2018, 13(4): e0196302.
- [21] 王生生, 王顺, 张航, 等. 基于轻量级网络及无人机遥感图像的大豆田杂草识别[J]. *农业工程学报*, 2019, 35(6): 81-89.
- WANG Shengsheng, WANG Sun, ZHANG Hang, et al. Soybean field weed recognition based on light sum-product networks and UAV remote sensing images[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2019, 35(6): 81-89. (in Chinese with English abstract)
- [22] MA X, DENG X, QI L, et al. Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields[J]. *PLOS ONE*, 2019, 14(4): e0215676.
- [23] GENZE N, AJEKWE R. Deep learning-based early weed segmentation using motion blurred UAV images of sorghum fields[J]. *Computers and Electronics in Agriculture*, 2022, 202: 107388.
- [24] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// *Computer Vision and Pattern Recognition (CVPR)*. Caesars Palace: IEEE, 2016: 770-778.
- [25] WANG Q, CHENG M, XIAO X, et al. An image segmentation method based on deep learning for damage assessment of the invasive weed *Solanum rostratum* Dunal[J]. *Computers and Electronics in Agriculture*, 2021, 188: 106320.
- [26] ZOU K, LIAO Q, ZHANG F, et al. A segmentation network for smart weed management in wheat fields[J]. *Computers and Electronics in Agriculture*, 2022, 202: 107303.
- [27] NASIRI A, OMID M, TAHERI-GARAVAND A, et al. Deep learning-based precision agriculture through weed recognition in sugar beet fields[J]. *Sustainable Computing: Informatics and Systems*, 2022, 35: 100759.
- [28] KIM Y, PARK K. MTS-CNN: Multi-task semantic segmentation-convolutional neural network for detecting crops and weeds[J]. *Computers and Electronics in Agriculture*, 2022, 199: 107146.
- [29] TARVAINEN A, VALPOLA H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[C]// *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017: 1195-1204.
- [30] LIU Y, TIAN Y, CHEN Y, et al. Perturbed and strict mean teachers for semi-supervised semantic segmentation[C]// *Computer Vision and Pattern Recognition (CVPR)*. New Orleans: IEEE, 2022: 4248-4257.
- [31] HEGDE D, SINDAGI V, KILIC V, et al. Uncertainty-aware mean teacher for source-free unsupervised domain adaptive 3D object detection[C]// *Computer Vision and Pattern Recognition (CVPR)*. Nashville: IEEE, 2021.
- [32] VERMA V, KAWAGUCHI K, LAMB A, et al. Interpolation consistency training for semi-supervised learning[J]. *Neural Networks*, 2022, 145: 96-106.
- [33] ZOU K, CHEN X, WANG Y, et al. A modified U-Net with a specific data argumentation method for semantic segmentation of weed images in the field[J]. *Computers and Electronics in Agriculture*, 2021, 187: 106242.
- [34] PICON A, SAN-EMETERIO M, et al. Deep learning-based segmentation of multiple species of weeds and corn crop using synthetic and real image datasets[J]. *Computers and Electronics in Agriculture*, 2022, 194: 106719.
- [35] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]// *Computer Vision and Pattern Recognition (CVPR)*. Las Vegas: IEEE, 2017: 6230-6239.



## UANP-MT based semi-supervised image segmentation method for identifying weeds in cabbage field

CAI Yulin<sup>1</sup>, XIAO Jiayi<sup>1</sup>, YU Chaoran<sup>2✉</sup>, SONG Zhao<sup>2</sup>, LI Jing<sup>2</sup>, YUE Xuejun<sup>1</sup>

(1. College of Electronic Engineering (College of Artificial Intelligence), South China Agricultural University, Guangzhou, 510642, China;

2. Vegetable Research Institute, Guangdong Academy of Agricultural Sciences / Guangdong Key Laboratory for New Technology Research of Vegetables, Guangzhou, 510640, China)

**Abstract:** Weeds pose a great threat to modern agricultural production. The unrestrained growth of weeds has also a significant impact on crop yields and quality. Therefore, it is a high demand to effectively manage and control weeds in recent years, in order to optimize agricultural production activities. The specific area of weed management can be expected to serve as the promising research direction for weed control operations, with the rapid development of unmanned aerial vehicle (UAV) platforms and artificial intelligence (AI). Accurate and efficient identification of weeds can be one of the most key steps to realizing automated weed management in the field. However, the current efficient models of weed identification are often required a significant amount of labeled agricultural data. The major bottleneck can be also limited in the model. In this study, a semi-supervised semantic segmentation network was proposed, called UANP-MT (uncertainty aware and network perturbed mean teacher), in order to reduce the dependence on labeled data. The PSPNet network model was also selected to leverage the teacher-student network concept from MT (Mean Teacher). Several key improvements were then incorporated during this time. Firstly, the improved model reduced the errors that caused by random network factors. The teacher network output was also augmented to take the mean, thereby ensuring the robustness of the network prediction. Secondly, the improved model constructed the uncertainty constraints for the different uncertainties between networks and assigns weightings, in order to balance the differences in the output between different networks. This approach enhanced the confidence and reliability of the prediction, ultimately leading to the higher recognition accuracy of an improved model. A series of ablation experiments were conducted to evaluate the effectiveness of the improved model. These experiments included the setting values of network parameters, the selection of feature extraction network backbones, and the model performance testing on datasets with varying volumes of data. Some optimal parameter settings were identified for the improved model, with the ResNet50 as the preferred choice for the model backbone. The comparison experiments showed that the UANP-MT model outperformed the original supervised network, in terms of the three evaluation indicators of F1 score, pixel accuracy (PA), and intersection over union (IoU), when the labeled data was less than that of the original supervised network, indicating the superiority of the semi-supervised network. Furthermore, the effectiveness of the improved model was further validated to compare with the classic semantic segmentation networks, including SegNet, U-Net, and Deeplabv3+ on the CaiXin weed dataset. UANP-MT model achieved better performance with an F1 score of 81.83%, a PA of 95.84%, and an IoU of 90.70% when only 1/4 of the labeled data was used for training. These evaluation metrics were superior to those of the Deeplabv3+ model by percentage point of 4.71, 7.94, and 8.27, respectively. The UANP-MT model was achieved in the high-quality weed detection and recognition with the low labeled data, thus significantly reducing the dependence on labeled data with the manpower and time cost saving. The findings can provide valuable implications for the subsequent weed recognition research and the development of automated weeding operations on UAV platforms.

**Keywords:** image processing; weed identification; semi-supervised learning; semantic segmentation; unmanned aerial vehicle (UAV) platform