

基于半监督伪标签学习的小麦头检测算法



陈辉凯^{1,*}, 熊璇², 鞠蔚华²

¹西部州长大学信息科学学院, 美国盐湖城 84107

²南昌市第九医院传染科, 江西南昌 330046

摘要: 小麦是全球性的粮食作物, 利用基于深度学习的小麦头检测算法有利于简化种植流程, 降低种植成本, 提高小麦产量。但是由于全球小麦的性状多样、生长周期不一致, 构建一个能够在各种场景都保持高鲁棒性以及高精度的检测模型至关重要。对此, 提出了基于深度学习的小麦头检测算法, 对于缺少丰富领域数据的小麦头检测场景, 提出利用半监督伪标签生成学习算法对无标注的数据进行半监督标注并用于模型训练, 提高在新特征域上的泛化性能。此外, 为了平衡小麦头检测算法的速度与精度, 设计了多模型融合机制, 对具有不同特性的检测算法进行融合。最后, 介绍了公开小麦头检测数据集, 并将其用于算法验证, 提出的算法在其验证集上获得 95.2% 的平均准确率, 在测试集上获得 76.1% 的平均准确率, 表明提出的算法具备对跨域场景的鲁棒能力以及优异的检测性能。

关键词: 目标检测; 半监督学习; 伪标签学习; 农业; 深度学习; 神经网络

DOI: [10.57237/j.cst.2023.04.006](https://doi.org/10.57237/j.cst.2023.04.006)

Semi-supervised Pseudo Label Learning-based Wheat Head Detection

Huikai Chen^{1,*}, Xuan Xiong², Weihua Ju²

¹Department of Information Technology, Western Governors University, Salt Lake City 84107, USA

²Department of Infection Diseases, Nanchang No. 9 Hospital, Nanchang 330046, China

Abstract: Wheat is a global food crop. Deep learning-based wheat head detection algorithm is helpful to simplify the planting process, reduce the planting cost and improve the wheat yield. However, due to the diversity of wheat traits and in-consistent growth cycles around the world, it is very important to build a detection model that can maintain high robustness and high accuracy in various scenarios. A deep learning-based wheat head detection algorithm was proposed. For the situation that domain data was scarce for wheat head detection, a semi-supervised pseudo label generation algorithm was proposed to semi supervised label the unlabeled data and use it for model training to improve the generalization performance in the new feature domain. In addition, in order to balance the speed and accuracy of wheat head detection algorithm, a multi model fusion mechanism was designed to fuse the detection algorithms with different characteristics. Finally, the open-source wheat head detection dataset was introduced and was used for algorithm verification. The algorithm obtained an average accuracy of 95.2% on its validation set and 76.1% on the test set, which

*通信作者: 陈辉凯, hkchen.dom@gmail.com

showed that the proposed algorithm had robustness to cross domain scenes and excellent detection performance.

Keywords: Object Detection; Semi-supervised Learning; Pseudo Label Learning; Agriculture; Deep Learning; Neural Network

1 引言

小麦是世界上种植最为广泛的粮食作物之一，经过人类多年持续的努力，小麦育种的研究不断进步，并直接保障了上亿人口免于饥饿[1]。近年来，小麦种植规模扩大，小麦种植工作量剧增，例如在小麦培育过程中必须对其生长情况进行及时记录和分析，通过分析小麦生长状况与环境的关系提高小麦的质量，增加产量；此外，估计小麦头生长密度，即计算单位面积下小麦头的个数也十分重要，有利于务农人员控制种植环境，减少因种植密度过大而造成的光照不足以及营养缺乏，或者种植密度稀疏而造成的产量低下。但是，由于务农人员数量有限，上述工作依靠人力完成不仅费时费力，通常还会带来无法估计的人为误差，因此减少务农人员机械化工作的需求愈发强烈。计算机视觉技术[2, 3]在智慧农业[4]，工业[5-12]，与经济技术[13-18]领域的引入将为这一困境提供思路。其中基于深度学习的目标检测技术适用于许多上述基于人力完成的工作，小麦头检测算法就是其中重要的应用。通过检测算法提取作物图像中独立小麦头的信息可以监控农作物的生长，改善作物的生长状况。

真实的小麦生长场景中通常给小麦头检测算法带来许多挑战，包括光照变化、复杂种植背景、多样的种植分布、农作物自身以及与其他植物的重叠、拍摄角度和距离的不确定。这些挑战使得图像检测算法的设计十分困难，但也有许多研究人员为此做出了努力。Hamuda E 等人[19]使用卡尔曼滤波和匈牙利算法检测农田中的作物，作者实验中使用的数据没有重叠作物，背景也为简单的土壤，因此该算法在真实场景中不具有足够的鲁棒性。Lu J 等人[20]提出一种用于柑橘作物检测的算法，其研究场景中存在大棚带来的颜色和光照变化，因此可以应对复杂的环境背景和不同层次的光照场景，但是对于作物尺度的处理无法满足需求，在目标表观不清晰、尺寸小的情况下表现不好。Linker R 等人[21]提出在自然场景中的苹果检测算法，该算法通过使用颜色和平滑性特征的像素级检测方法获得高效的检测速度，可是由于场景单一，仍然无法

很好地适应密集场景和存在大量重叠的场景。

一方面，随着深度学习技术的进步[22-24]，基于深度学习的智慧农业技术逐渐涌现[25, 26]。农作物分类是其他模型的基础，Zhang Y D 等人[27]提出使用 13 层的卷积神经网络（convolutional neural network, CNN）进行水果分类，并获得 94.94%的精度。农作物的检测和分割任务则是在分类的基础上要对目标边界框级别或者像素级别的位置进行标定。Chen S W 等人[28]基于全连接网络提出水果的斑点检测算法，并通过后续的计数算法对目标进行计数。Dyrmann M 等人[29]使用全连接网络设计草地以及叶子检测算法。为了能在负载的自然环中对目标区域进行更加精准的识别，Chen S 等人[30]提出使用卷积神经网络以及高效检测算法在食品售卖环节进行智能化操作。

另一方面，通用检测任务也在快速发展。Girshick R 等人[31]提出的 R-CNN 系列算法通过使用候选区域网络（region proposal network, RPN）来获得感兴趣区域（region of interest, RoI），结合分类器对每个感兴趣区域进行分类，同时利用回归器获得精确的边界。Redmon J 等人[32]提出的 YOLO 系列算法将分类和定位统一划分为一个回归任务，因此检测速度十分快。就此，也有不少学者提出相应的智能农业检测算法。其中 Bargouti S [33]提出基于 Faster R-CNN 框架的多类别水果检测算法，并取得了较好的检测性能。

然而在构建检测农作物的算法模型的过程中，还有一个重要的问题没有被纳入考虑范围，即农作物自身的差异和变化。之前的研究工作受限于农作物目标的形状类别、有限的生长地理位置和观测条件，测试数据集单一，造成模型识别性能较差，鲁棒性不佳，难以应用于真实场景。通常在深度模型训练的过程中，研究人员将训练数据分为训练数据集和验证集，并期望通过这些数据训练的模型能够泛化到未知的测试数据上。但实际上并没有理论能够保证模型可以应用于未知域，因此模型的性能也就无从保障。再者，获取场景类别多样、目标个体性状差异大、符合跨领域数

据采集规范的农作物目标检测数据集也存在诸多挑战。因此,想要设计泛化性好、精确度高的小麦头农作物检测模型存在多方面的挑战。

为此,本文提出基于半监督伪标签学习的小麦头检测算法。先使用基于标签的全监督训练方式对模型进行训练,然后使用未标注的数据对模型进行训练,并重复整个过程。对于无标签的数据,本文使用伪标签技术为其实时构建标签,每次选择预测概率最大的样本进行模型训练。在伪标签训练的基础上,通过融合多个高效的检测算法设计一个完善的检测流程。本文在公开的小麦头数据集上进行模型测试,并与其他算法进行了对比,结果表明,本文提出的算法具有明显的优势。

本文的贡献如下。

- (1) 本文提出基于半监督伪标签学习的小麦头检测算法,通过将未知领域的特征知识融入模型训练过程,检测模型能够更好地适应新的领域,进而处理小麦头性状、生长时期、品种不一致的问题,提高检测精度和泛化性。
- (2) 本文提出多模型融合的检测流程,包含单模型训练、测试时增强、加权边框融合等,结合不同模型自身的优势,将多模型的检测结果进行多层次的融合,在保证算法速度的前提下极大地提高算法精度和鲁棒性。
- (3) 本文在公开的多样化小麦数据平台上进行训练和测试,并与其他检测算法进行对比,区别于先前的农作物数据集,该小麦头数据集拥有复杂多样的真实场景,进而表明了本文提出的算法在实际应用中的优势。

2 相关工作

2.1 目标检测

通用目标检测任务拥有很长的研究历史,目前的目标检测器大致分为两大类:两阶段检测器和单阶段检测器,即是否存在感兴趣区域提取模块。其中两阶段检测器通常更加灵活并且更加精确,而单阶段检测器通常更加简单和高效。近年来,研究人员聚焦于设计高效的检测器架构,例如单阶段[34]和无锚框[35]的检测器。在目标检测任务中,其中一个主要挑战是如何高效表达和处理多尺度特征。早期的检测器通常

直接在骨干网络后面连接多层特征金字塔结构[36]进行多尺度提取。Lin T Y 等人[37]提出自上到下的多层特征通道,用于结合多尺度的特征信息。Liu S 等人[38]在多层特征金字塔的基础上添加额外的自底向上的通路聚合网络。Zhang Z 等人[39]提出一个动态卷积核运算子。Liang Z 等人[40]使用全局与局部特征融合提升检测鲁棒性。最近, Ghiasi G 等人[41]将神经网络搜索算法与目标检测任务结合,用于自动搜索特征网络架构的拓扑结构。尽管通过神经网络搜索方法得到的架构获得了更好的性能,但该搜索算法的计算过程需要上千小时,并且搜索出来的网络是不规则的、难以解释的。在本文中,笔者使用两个单阶段检测器(YOLOv5 以及 EfficientDet [42])进行小麦头检测算法设计。

2.2 半监督学习

半监督学习使用标注以及未标注的数据进行模型训练,半监督学习的训练方法包含自训练、协同训练以及标签传递等[43]。许多工作通过使用一致性损失进行半监督图像分类模型的训练,或通过减小噪声图像和原图像之间的预测差值来在有限的标注数据下获得具有优良性能的模型。在目标跟踪任务下,半监督学习的训练方式也有多种形式。Cinbis R G 等人[44]通过混合边框级别的标注和图像级别的标注进行共同训练,Hoffman J 等人[45]通过按类别划分的方式进行半监督训练,其中一部分类别使用边框级别的标注,其余类别使用图像级别的标注。通过这些混合训练,模型能够学习到新领域的知识,但是图像级别的标注过于粗糙,存在给模型带来过多噪声的问题。相比之下,笔者通过伪标签的方式为未知的图像进行半监督标注,筛选出其中置信度较高的图像并将其作为新的训练样本,以此往复来获得越来越多可信的新数据,进而提高模型在新小麦头数据的泛化性能,使模型更加精确地检测到各种场景下的小麦头。

3 小麦头检测算法

本节介绍小麦头检测算法,首先介绍整体算法流程以及其形式化表示,其次介绍用于模型融合的两个高效检测算法,最后介绍半监督伪标签算法的原理。

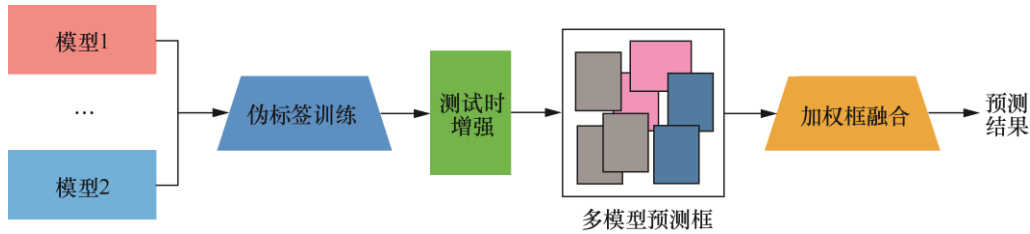


图1 基于半监督伪标签学习的小麦头检测算法流程

3.1 算法流程

基于半监督伪标签学习的小麦头检测算法流程如图1所示。首先，通过初始模型训练和测试后筛选出性能相当的多个模型，即检测器。接着，利用这些在部分有标签的数据上训练的模型进行伪标签训练。具体的流程是首先通过对无标签图像进行检测，即进行伪标注，其次选择其中检测结果置信度高的图像按照一定比例加入原始训练集，并对模型进行微调（fine-tuning），这个过程多次循环直至平均精度提升小于一定的范围后停止（具体参考算法1）。在获得经过伪标签微调的模型后，使用模型对测试数据进行推断，首先对测试数据进行测试时增强，即对原图像进行翻转、光照变化等，让模型对同一样本的不同增强图像分别

进行预测；在获得多模型预测框后，使用加权框融合算法，提取其中置信度最高的候选框作为最终目标位置，从而得到小麦头检测的预测结果。

算法1 伪标签训练算法

- ①使用标注数据训练有监督模型 M ;
- ②使用有监督模型 M 对无标签数据进行预测，得出预测概率 P ;
- ③通过预测概率 P 筛选高置信度样本;
- ④使用有标注数据以及伪标签数据训练新模型 M' ;
- ⑤用 M' 替换 M ，重复以上步骤直至模型效果不出现提升。

本文将模型伪标签训练的停止条件设置为平均精度提升在 0.2% 以内，经过上述过程，本文获得了对目标域具有泛化性能的检测模型。

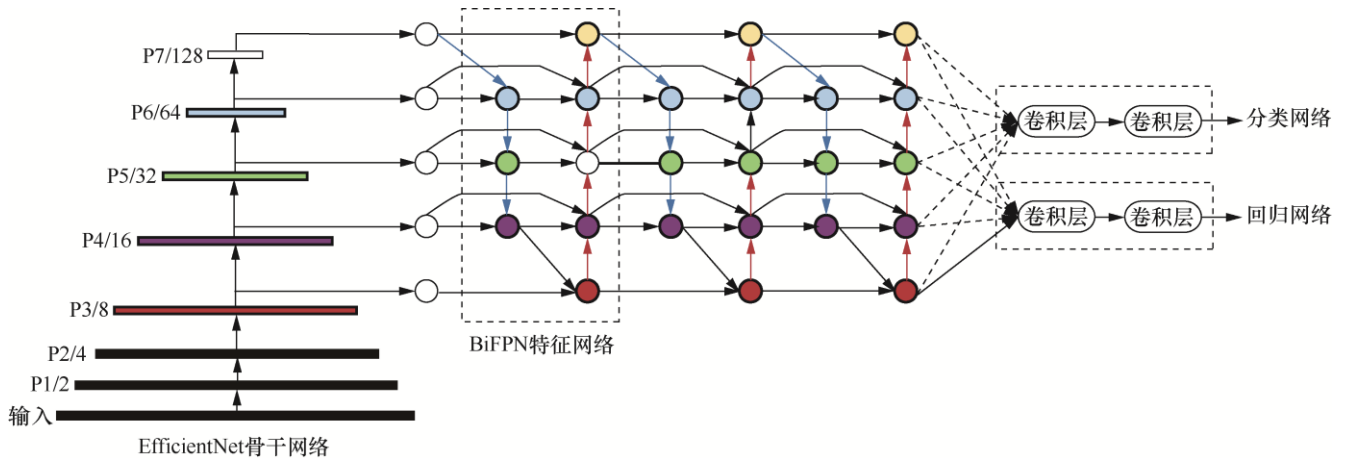


图2 EfficientDet 的整体结构

3.2 检测算法

(1) EfficientDet

EfficientDet [42] 作为 EfficientNet [46] 在检测任务上的扩展，其主要特性体现在针对影响检测任务性能的两个重要因素进行改进。

改进一：高效的多尺度特征融合。在 FPN 的基础上，又发展出了 PANet [38] 和 NAS-FPN [41] 以进行跨

尺度特征融合。但是在融合特征时，上述方法只是将特征进行简单的相加，然而不同特征的分辨率是不同的，它们对融合后的特征的贡献程度也是不同的。基于这个思路，EfficientDet 提出了加权双向特征金字塔网络（weighted bi-directional feature pyramid network, BiFPN），它引入了可学习的权重以感知不同特征的重要性，同时多次进行自上而下和自下而上的多尺度特征融合，如式(1)所示：

$$\begin{aligned} P_7^{out} &= Conv(P_7^{in}) \\ P_6^{out} &= Conv(P_6^{in} + Re\ size(P_7^{out})) \\ &\dots \\ P_3^{out} &= Conv(P_3^{in} + Re\ size(P_4^{out})) \end{aligned} \quad (1)$$

其中, $P_k^{out} (k=3\dots7)$ 和 $P_k^{in} (k=3\dots7)$ 分别为 BiFPN 中不同层级的特征输出与输入特征, $Conv$ 为卷积层, $Re\ size$ 为上采样操作。

改进二: 检测模型的尺寸。以往的目标检测方法依赖较大的骨干网络 (backbone) 或较大的输入图像以提高检测精度, 而 EfficientDet 的研究人员发现, 当同时考虑检测精度和检测效率时, 需要对特征网络和分类/回归网络进行缩放。因此本文提出一种复合缩放方法 (compound scaling method), 它可以为所有的骨干网络、特征网络和分类/回归网络的分辨率/深度/宽度进行缩放 (一般都是放大)。

EfficientDet 的整体结构如图 2 所示, 其中骨干网络旁边的序号表示特征层级以及步幅大小。EfficientDet 将在 ImageNet 上经过预训练的 EfficientNet 作为骨干网络, 将 BiFPN 作为特征网络, 后续特征网络部分选取 EfficientNet 中 3 到 7 层输出的特征图 $\{P_3, P_4, P_5, P_6, P_7\}$, 随后, 在 BiFPN 中重复应用自顶向下 (top-down) 和自底向上 (bottom-up) 双向特征融合。融合后的特征分别被送入分类网络和回归网络中, 以生成目标类别并预测目标边界框 (bbox), 其中分类网络与回归网络在所有特征层级上的权重是共享的。考虑到模型的训练效率和模型推断效率, 本文在小麦头检测的应用中使用 EfficientDet-D5 作为基础模型。但由于缺乏不同地区、不同小麦品种以及不同生长周期的小麦头检测训练数据, 因此直接使用全监督的方式对初始模型进行训练获得的模型并不能获得令人满意的结果。

(2) YOLOv5

YOLOv5 网络主要由 3 个主要组件组成。①骨干网络: 在不同图像细粒度上聚合并形成图像特征的卷积神经网络。②颈部特征提取网络: 一系列混合和组合图像特征的网络层, 负责将图像特征传递到预测层。③检测头: 对图像特征进行预测, 生成边界框并预测类别。YOLOv5 官方代码给出的目标检测网络共有 4 个版本, 分别是 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x。其中 YOLOv5s 的网络结构如图 3 所示, YOLOv5s 是 YOLOv5 系列中深度最小、特征图的宽度最小的网络。其他 3 种都是在此基础上不断加深、不断加宽得到的网络。

YOLOv5s 网络的训练和设计相比之前的版本具有几点优势: ①输入端在常用数据增强方法的基础上增加 Mosaic 数据增强方法, 用来辅助增加训练样本的领域知识; ②自适应锚框计算以及自适应图像缩放; ③在骨干网络中增加 Focus 结构以及 CSP 结构, 增加特征表达能力; ④在颈部网络中使用 FPN+PAN 结构, 使用自底向上的特征融合方式, 增加网络中的定位信息; ⑤在预测端使用 GIoU_Loss, 帮助网络学习有效信息。

YOLOv5 中 Focus 结构的切片操作如图 4(a)所示, 其中 $4\times4\times3$ 的图像切片后变成 $2\times2\times12$ 的特征图。

以 YOLOv5s 的结构为例, 切片操作中特征尺寸变化如图 4(b)所示, 假设原始 $608\times608\times3$ 的图像输入 Focus 结构, 采用切片操作, 类似于近邻下采样方法, 在一张图像中每隔一个像素拿到一个值, 因此得到 4 张互补的图像, 4 张图像外观类似但是不存在信息丢失, 接着将 4 张图像拼接起

来组成 $304\times304\times12$ 的特征图, 再经过一次 32 个卷积核的卷积操作以及激活层, 最终获得 $304\times304\times32$ 的特征图。

3.3 半监督伪标签技术原理

伪标签技术是用于增强深度网络模型泛化性的半监督训练方法。本文中, 笔者以 M 层隐藏层的多层感知机为例对半监督伪标签技术的原理进行说明。

以一个简单的目标分类任务为例, 其模型的形式化表示如式(2)所示:

$$\begin{aligned} h_i^k &= s^k \left(\sum_{j=1}^d w_{ij}^k h_j^{k-1} + b_i^k \right) \\ k &= 1, \dots, M+1 \end{aligned} \quad (2)$$

其中, h_i^k 是输出单元, 即中间层特征或者目标类别的概率, s^k 是非线性激活函数 (例如 Sigmoid), h_j^k 是输入图像或者上一特征层的输入值, w_{ij}^k 为网络权重, b_i^k 为网络偏置。

对于目标分类任务, 伪标签实际上就是将预测出来的目标类别作为真实的标签。通常, 使用者选择具有最大预测概率标签对应的样本进行训练集的扩增。

对于存在领域知识不均衡的任务, 通常预先基于全监督的方法训练基础模型, 随后基于伪标签技术进行微调, 并配合使用随机丢弃方法 (dropout), 相比于全监督算法, 基于半监督伪标签训练的算法会重新

计算网络权重更新的值和方向，朝着更适合目标域的方向进行模型优化。

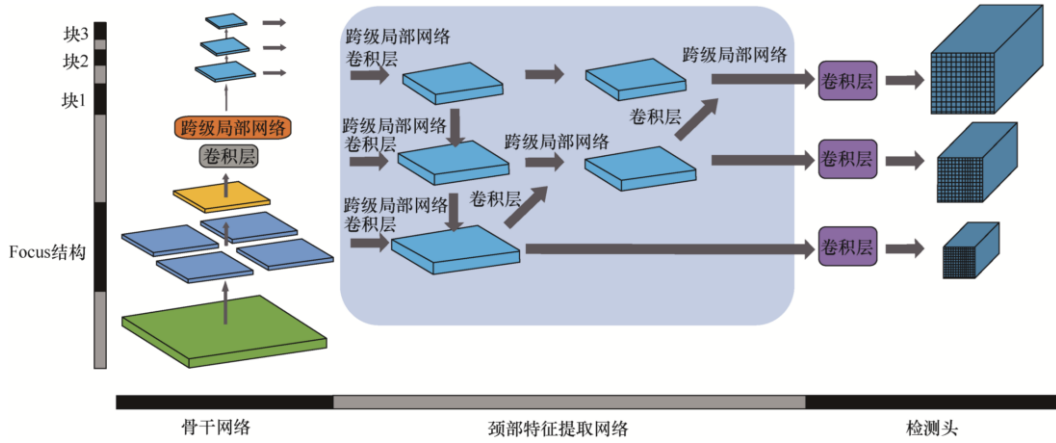


图3 YOLOv5s的网络架构

在一些情况下，由于已标注图像和伪标注图像的数量差距非常巨大，因此如何平衡二者来获得足够的性能就变得至关重要，其中，总的损失函数如式(3)所示：

$$L = \frac{1}{n} \sum_{m=1}^n \sum_{i=1}^C L(y_i^m, f_i^m) + \alpha(t) \frac{1}{n'} \sum_{m=1}^{n'} \sum_{i=1}^C L(y_i^m, f_i^m) \quad (3)$$

其中， n 是批量数， n' 代表无标签数据的批量数（后续带有上标的符号均代表无标签数据的对应表示）， C 代表每张图像中目标的数量， f_i^m 是输出单元， y_i^m 是标签， $\alpha(t)$ 是用于平衡有标签数据和无标签数据的参数，详细解释参考下文。

一个合适的 $\alpha(t)$ 通常十分重要，如果 $\alpha(t)$ 过大，训练过程会严重偏向无标签数据，从而极大地影响有标签数据；然而，如果 $\alpha(t)$ 过小，算法就无法从无标签数据中学到有价值的特征。为了防止网络在训练优化的过程中在局部最优附近震荡，研究人员使用回火（annealing）的方式，随着训练轮次逐渐改变无标签数据对模型的影响程度。 $\alpha(t)$ 如式(4)所示：

$$\alpha(t) = \begin{cases} 0 & t < T_1 \\ \frac{t-T_1}{T_2-T_1} \alpha_f & T_1 \leq t \leq T_2 \\ \alpha_f & T_2 \leq t \end{cases} \quad (4)$$

其中， T_1 、 T_2 以及 α_f 为常数，根据实际情况调节，在笔者的实验中，由于无标签数据也同样缺乏，将 T_2 设置为 0， α_f 设置为 1。

半监督学习的目标是使用无标签数据提高泛化性能。簇假设理论[47]表明，使用伪标签技术在低密度空间的决策边界可以提升模型泛化性能。

4 实验分析

4.1 数据集介绍

小麦头检测任务来自于欧洲计算机视觉国际会议（European Conference on Computer Vision, ECCV2020）的研讨会竞赛——有关植物表型的计算机视觉问题（computer vision problems in plant phenotyping, CVPPP）。

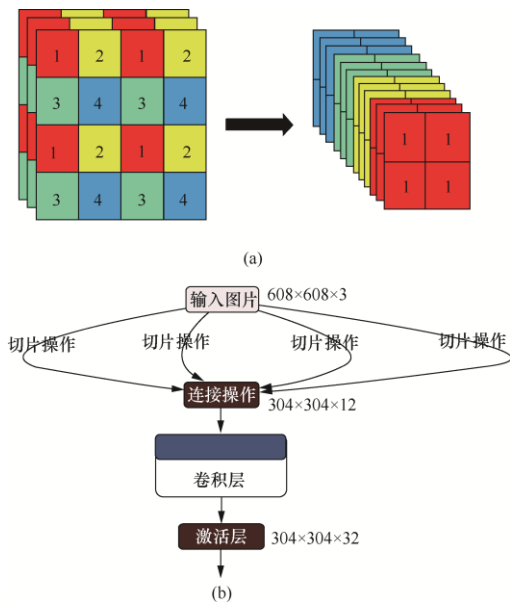


图4 YOLOv5s中Focus的切片操作以及切片操作中特征尺寸的变化

小麦作为粮食在全球范围内广泛流行，并得到了广泛的研究。为了获得有关麦田的大量准确数据，植物科学家使用“小麦头”（包含小麦顶部的穗）的图像进行检测。这些图像用于估计不同品种的小麦头的密度和大小，从而帮助农民对自己的农场进行管理规划，同时可以通过使用这些数据评估小麦的健康状况和成熟程度。

但是，在野外图像中进行准确的小麦头检测十分具有挑战性。小麦的重叠、外界因素（如风吹）会造成照片模糊不清，还有其他许多不可预知的问题都会使识别单个小麦头变得困难。此外，小麦的外观会因成熟度、颜色、基因型和头部方向而存在巨大差异。由于小麦在世界范围内种植，因此还必须考虑小麦的品种、种植密度、种植模式和田间状况。小麦头检测模型需要获得对不同生长环境的泛化性能。当前的检测方法涉及一阶段检测器 Yolo-V3 和两阶段检测器 Faster-RCNN，但是即使在使用大型数据集进行训练时，仍然存在对领域数据分布的偏见（bias），性能无法令人满意。

全球小麦检测（global wheat detection, GWD）数据集[48]来自 7 个国家的 9 个学校和研究所：日本东京大学、法国农业科学研究所、法国营养与环境研究所、法国技术研究机构 Arvalis、瑞士苏黎世联邦理工学院、加拿大萨斯喀彻温大学、澳大利亚昆士兰大学、中国南京农业大学和英国洛桑实验站。在该数据集中，算法将从室外的小麦植株图像（包括来自全球的小麦）中检测出小麦植株的头部。基于全球的小麦数据，本文设计用于小麦头检测的通用解决方案，用来估计小麦头的数量和大小。为了更好地评估未知的基因型、环境和种植条件多个因素对性能的影响，本文的训练数据集涵盖了多个区域，包括来自法国、英国、瑞士、加拿大的 3000 多张图像。测试数据包括来自澳大利亚、日本和中国的约 1 000 张图像。本文按照地区对 GWD 数据集进行可视化，可以看到小麦的形状、麦穗的颜色、小麦的生长时期等都有着显著的差异，如图 5 所示。



图 5 GWD 小麦头检测数据中存在的挑战

4.2 性能指标及对比

本文算法基于 Pytorch 的深度框架实现，计算机

硬件设备的配置为 HP Pro 3330, Intel 2.4GHz CPU，内存 32G，英伟达 GTX-2080Ti 显卡。算法将 EfficientDet-D5 和 YOLOv5x 模型作为基本结构，并进行调优。

在训练阶段，采取分阶段的训练方式以获得最好的收敛结果。对于原始模型，EfficientDet-D5 训练 80 轮次，YOLOv5x 训练 50 轮次，单轮伪标签训练为 10 轮次。训练过程中根据需要使用 SGD 以及 Adam 作为训练优化器，学习率为 10^{-3} 。

本文的实验结果包含 4 个部分，分别为本地 8:2 比例分割的训练集、验证集、无法预知的公开测试榜（public leaderboard）和私有测试榜（private leaderboard）。本文中出现的测试指标均为交并比（intersection over union, IoU）阈值为 50 的平均精度（mean average precision 50, mAP50）。

YOLOv5x 的性能优化见表 1，EfficientDet-D5 的性能优化见表 2。从表 1~表 2 可以看出，使用测试时增强方法（TTA）、伪标签增强（PL）、加权框融合（WBF）、数据增强（Aug）都对基准算法（Baseline）性能提升有明显的效果。

表 1 YOLOv5x 的性能优化

算法	本地验证集	公开测试榜
Baseline	94.2	70.8
Baseline +WBF	-	73.1
Baseline+TTA+WBF	-	75.9
Baseline+TTA+WBF+PL	-	76.0

表 2 EfficientDet-D5 的性能优化

算法	本地验证集	公开测试榜
Baseline	92.2	69.7
Baseline+Aug+WBF	-	71.5
Baseline+Aug+WBF+PL	-	73.4

表 3 ECCV-CVPPP 算法竞赛性能排名以及对应准确率

算法	本地验证集	公开测试榜	私有测试榜
Top1	-	77.4	68.9
Top2	-	77.4	68.7
Top3	-	77.2	68.3
Top10	-	76.8	67.3
Top20	-	76.4	66.8
本文算法	95.2	76.1	66.7

表 4 本文算法与其他检测算法的性能

算法	本地验证集	公开测试榜
YOLOv3 [49]	92.2	65.0
Fast R-CNN [31]	94.0	66.8
YOLOv4 [50]	93.4	65.6
本文算法	95.2	76.1

表 1 和表 2 展示了各个模块在 YOLOv5x 以及 EfficientDet-D5 检测算法上的性能提升, 并进行本地验证集以及公开测试榜测评。在 YOLOv5x 算法中, 已经存在大量增强方法, 因此不考虑数据增强带来的性能提升。此外, 可以看到加权框融合改进非极大值抑制算法对性能贡献 2.3%, 测试时增强贡献 2.8%, 伪标签技术贡献 0.1%, 单模型在公开测试榜上获得 76.0% 的 mAP50。在 EfficientDet-D5 算法中, 使用数据增强以及加权框融合算法, 其贡献 1.8%, 伪标签技术贡献 1.9%, 单模型在公开测试榜上获得 73.4% 的 mAP50。

表 3 中展示了在公开以及私有排名榜上, 本文的算法与部分算法的性能对比。本文最终的多模型算法在私有测试榜获得 66.7% 的 mAP50, 在公开测试榜获得 76.1% 的 mAP50。在全球 2000 余支队伍中, 本文算法排名 22, 位于前 1%。对比第一名 (Top1) 在私有测试榜中只相差 2.2%, 在公开测试榜只相差 1.3%。

本文算法与其他检测算法的性能见表 4。YOLOv3 是使用 Darknet-53 的单阶段检测算法, Fast R-CNN 通过特征剪裁 (feature crop) 操作对目标进行剪裁、缩放以及分类, YOLOv4 是在 YOLOv3 的基础上使用了大量先进的结构改进得到的。可以看到, 本文算法在真实测试环境下具有明显的优势。

4.3 检测效果展示

小麦头检测效果如图 6 所示, 可以观察到, 本文算法可以应对小麦头检测问题中存在的性状不统一 (如图 6(a)~图 6(f)、农作物重叠多 (如图 6(c)~图 6(f))、分布不规律 (如图 6(a)~图 6(f)、光照不均匀 (如图 6(a)、图 6(b)、图 6(e)、图 6(f)等诸多具有挑战性的场景。

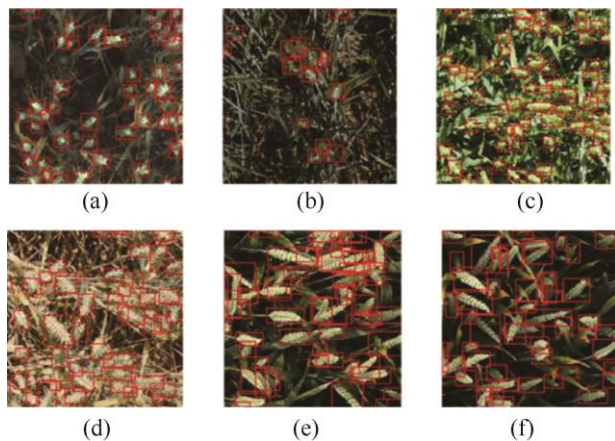


图 6 小麦头检测效果展示

5 结论

小麦是主要粮食全球供应, 因此必须考虑不同的生长条件对小麦培育的影响, 并期望构建能够在不同环境之间泛化的小麦头检测模型。在得到一个良好的基础模型后, 研究人员使用较低的成本即可准确估算出不同品种和场景下的小麦头的密度和大小, 帮助农民更好地评估农作物的生长状况, 并有助于解决全球性的粮食问题。本文中提出使用半监督的算法获得未知域的伪标签, 并多次用于训练, 逐渐提高对新场景的泛化性能。此外, 本文使用目前性能优良的检测算法 YOLOv5 和 EfficientDet 进行多模型融合, 结合当前优良的增强和融合算法, 设计了一个完备的检测数据流。最后, 本文介绍了一个分布广泛, 极具挑战性的小麦头检测数据集, 基于此与其他算法进行性能对比, 并说明本文算法的优良性。

参考文献

- [1] REYNOLDS M P, BORLAUG N E. Applying innovations and new technologies for international collaborative wheat improvement [J]. The Journal of Agricultural Science, 2006, 144(2): 95-110.
- [2] Youwei, L. (2023). Real-time eye blink detection using general cameras: a facial landmarks approach. International Science Journal of Engineering & Agriculture, 2(5), 1-8.
- [3] Lu, Y. (2023, July). REAL-TIME EYE BLINK DETECTION WITH GENERAL CAMERAS. In The 28th International scientific and practical conference "Science and development of methods for solving modern problems (pp. 198-201).
- [4] 陈妮亚, 阮佳阳, 黄金苗, 等. 结合深度学习与生物特征识别在冷链拣选中的算法研究[J]. 智能科学与技术学报, 2019, 1(1): 88-95.
- [5] CHEN N Y, RUAN J Y, HUANG J M, et al. Algorithm design for food-picking combining deep learning and biometrics recognition [J]. Chinese Journal of Intelligent Science and Technology, 2019, 1(1): 88-95.
- [6] Liu, Yiming, and Yi Bao. "Review on automated condition assessment of pipelines with machine learning." Advanced Engineering Informatics 53(2022): 101687.
- [7] Liu, Yiming, and Yi Bao. "Review of robot-based damage assessment for offshore with turbines." Renewable and Sustainable Energy Reviews 158 (2022): 112187.
- [8] Liu, Yiming, and Yi Bao. "Intelligent monitoring of spatially-distributed cracks using distributed fiber optic sensors assisted by deep learning." Measurement 220(2023): 113418.

- [9] Zeng, Lu, et al. "Graph convolutional network with sample and feature weights for Alzheimer's disease diagnosis." *Information Processing & Management* 59.4(2022): 102952.
- [10] Xiao, Tingsong, et al. "Dual-Graph Learning Convolutional Networks for Interpretable Alzheimer's Disease Diagnosis." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer Nature Switzerland, 2022.
- [11] Wang, Xiaochen, et al. "MRMRP: multi-source review-based model for rating prediction." *Database Systems for Advanced Applications: 25th International Conference, DASFAA 2020, Jeju, South Korea, September 24–27, 2020, Proceedings, Part II* 25. Springer International Publishing, 2020.
- [12] Wang, Xiaochen, Tingsong Xiao, and Jie Shao. "EMRM: Enhanced Multi-source Review-Based Model for Rating Prediction." *Knowledge Science, Engineering and Management: 14th International Conference, KSEM 2021, Tokyo, Japan, August 14–16, 2021, Proceedings, Part III* 14. Springer International Publishing, 2021.
- [13] Hu, Jing, et al. "M-GCN: Multi-scale Graph Convolutional Network for 3D Point Cloud Classification." *2023 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2023.
- [14] Qiu, Yujuan. *ESTIMATION OF TAIL RISK MEASURES IN FINANCE: APPROACHES TO EXTREME VALUE MIXTURE MODELING*. Diss. Johns Hopkins University, 2019.
- [15] Dai, W., Mou, C., Wu, J., & Ye, X. (2023, May). Diabetic Retinopathy Detection with Enhanced Vision Transformers: The Twins-PCPVT Solution. In *2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI)* (pp. 403-407). IEEE.
- [16] Mou, C., Ye, X., Wu, J., & Dai, W. (2023, April). Automated ICD Coding Based on Neural Machine Translation. In *2023 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)* (pp. 495-500). IEEE.
- [17] Ye, X., Wu, J., Mou, C., & Dai, W. (2023, May). MedLens: Improve Mortality Prediction Via Medical Signs Selecting and Regression. In *2023 IEEE 3rd International Conference on Computer Communication and Artificial Intelligence (CCAI)* (pp. 169-175). IEEE.
- [18] Mou, C., Dai, W., Ye, X., & Wu, J. (2023, July). Research On Method Of User Preference Analysis Based on Entity Similarity and Semantic Assessment. In *2023 8th International Conference on Signal and Image Processing (ICSIP)* (pp. 1029-1033). IEEE.
- [19] HAMUDA E, MC GINLEY B, GLAVIN M, et al. Improved image processing-base crop detection using Kalman filtering and the Hungarian algorithm [J]. *Computers and Electronics in Agriculture*, 2018, 148: 37-44.
- [20] LU J, SANG N. Detecting citrus fruits and occlusion recovery under natural illumination conditions [J]. *Computers and Electronics in Agriculture*, 2015, 110: 121-130. [LinkOut]
- [21] LINKER R, COHEN O, NAOR A. Determination of the number of green apples in RGB images recorded in orchards [J]. *Computers and Electronics in Agriculture*, 2012, 81: 45-57.
- [22] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [23] Zhang, Z., & Hu, R. (2023). Byzantine-Robust Federated Learning with Variance Reduction and Differential Privacy. *arXiv preprint arXiv: 2309.03437*.
- [24] LEE S H, CHAN C S, MAYO S J, et al. How deep learning extracts and learns leaf features for plant classification [J]. *Pattern Recognition*, 2017, 71: 1-13.
- [25] Wu, J., Ye, X., Mou, C., & Dai, W. (2023, May). Fineehr: Refine clinical note representations to improve mortality prediction. In *2023 11th International Symposium on Digital Forensics and Security (ISDFS)* (pp. 1-6). IEEE.
- [26] RAHNEMOONFAR M, SHEPPARD C. Deep count: fruit counting based on deep simulated learning [J]. *Sensors*, 2017, 17(4): 905.
- [27] ZHANG Y D, DONG Z C, CHEN X Q, et al. Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation [J]. *Multimedia Tools and Applications*, 2019, 78(3): 3613-3632.
- [28] CHEN S W, SHIVAKUMAR S S, DCUNHA S, et al. Counting apples and oranges with deep learning: a data-driven approach [J]. *IEEE Robotics and Automation Letters*, 2017, 2(2): 781-788.
- [29] DYRMANN M, JØRGENSEN R N, MIDTIBY H S. RoboWeedSupport - Detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network [J]. *Advances in Animal Biosciences*, 2017, 8(2): 842-847.
- [30] Cheng, S., Chu, B., Zhong, B., Zhang, Z., Liu, X., Tang, Z., & Li, X. (2021). DRNet: Towards fast, accurate and practical dish recognition. *Science China Technological Sciences*, 64(12), 2651-2661.
- [31] GIRSHICK R. Fast R-CNN [C] // *Proceedings of the 2015 IEEE International Conference on Computer Vision*. Piscataway: IEEE Press, 2015: 1440-1448.

- [32] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] // Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 779-788.
- [33] BARGOTI S, UNDERWOOD J. Deep fruit detection in orchards [C] // Proceedings of the 2017 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2017: 3626-3633.
- [34] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [M] // Computer Vision – ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [35] LAW H, DENG J. CornerNet: detecting objects as paired keypoints [J]. International Journal of Computer Vision, 2020, 128(3): 642-656.
- [36] CAI Z W, FAN Q F, FERIS R S, et al. A unified multi-scale deep convolutional neural network for fast object detection [M] // Computer Vision - ECCV 2016. Cham: Springer International Publishing.
- [37] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] // Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 936-944.
- [38] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C] // Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8759-8768.
- [39] Zhang, Z., Zhong, B., Zhang, S., Tang, Z., Liu, X., & Zhang, Z. (2021). Distractor-aware fast tracking via dynamic convolutions and mot philosophy. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1024-1033).
- [40] Liang, Z., Cao, H., Yang, C., Zhang, Z., & Chen, G. (2022, September). Global-local feature aggregation for event-based object detection on eventkitti. In 2022 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI) (pp. 1-7). IEEE.
- [41] GHIASI G, LIN T Y, LE Q V. NAS-FPN: learning scalable feature pyramid architecture for object detection [C] // Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 7029-7038.
- [42] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection [C] // Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 10778-10787.
- [43] CHAPELLE O, SCHOLKOPF B, ZIEN E. Semi-supervised learning [J]. IEEE Transactions on Neural Networks, 2009, 20(3): 542.
- [44] CINBIS R G, VERBEEK J, SCHMID C. Weakly supervised object localization with multi-fold multiple instance learning [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(1): 189-203.
- [45] HOFFMAN J, GUADARRAMA S, TZENG E, et al. LSDA: large scale detection through adaptation [M] // Advances in Neural Information Processing Systems. [S.l.:s.n.], 2014: 3536-3544.
- [46] TAN M X, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks [C] // Proceedings of the International Conference on Machine Learning. PMLR, 2019: 6105-6114.
- [47] CHAPELLE O, ZIEN A. Semi-supervised classification by low density separation [C] // Proceedings of the International workshop on artificial intelligence and statistics. [S.l.:s.n.], 2005: 57-64.
- [48] DAVID E, MADEC S, SADEGHI-TEHRAN P, et al. Global wheat head detection (GWHD) dataset: a large and diverse dataset of high-resolution RGB-labelled images to develop and benchmark wheat head detection methods [J]. Plant Phenomics, 2020, 2020: 1-12.
- [49] REDMON J, FARHADI A. Yolov3: an incremental improvement [J]. arXiv preprint, 2018, arXiv: 1804.02767.
- [50] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection [J]. arXiv preprint, 2020, arXiv: 2004.10934.