

文章编号: 1002-1582(2025)05-0617-09

基于遮挡增强知识蒸馏的果实识别

史源平^{1,2,3}, 马彦恒^{1*}, 耿亮^{2,3}, 褚丽娜¹, 李伟¹, 李秉璇¹

(1. 陆军工程大学石家庄校区, 河北 石家庄 050003)

(2. 石家庄学院 机电工程学院, 河北 石家庄 050035)

(3. 石家庄市农业机器人智能感知重点实验室, 河北 石家庄 050035)

摘 要: 果实被枝叶遮挡时会显著削弱采摘机器人视觉系统的识别精度。针对这一问题, 文章提出一种基于“遮挡增强知识蒸馏”的果实识别框架。该框架在构建的遮挡数据集上, 通过引入遮挡信息约束多尺度语义特征对齐, 并结合指数移动平均策略稳定训练过程。在 MinneApple 数据集上的实验结果表明, AP@[0.5:0.05:0.95]、AP@0.5、AP@0.75、小目标 AP 与大目标 AP 分别达到 0.744、0.946、0.793、0.674 和 0.976, 较主流检测模型平均提升 0.178。研究证明, 方法能够在复杂遮挡场景下显著提高采摘机器人的视觉感知精度, 为智慧农业果实自动采摘提供有效技术支撑。

关 键 词: 图像处理; 遮挡增强; 知识蒸馏; 特征对齐

中图分类号: TP391

文献标识码: A

Fruit recognition based on occlusion enhanced knowledge distillation

SHI Yuanping^{1,2,3}, MA Yanheng^{1*}, GENG Liang^{2,3}, CHU Lina¹, LI Wei¹, LI Bingxuan¹

(1. Shijiazhuang Campus of Army Engineering University of PLA, Shijiazhuang 050003, China)

(2. School of Mechanical and Electrical Engineering, Shijiazhuang College, Shijiazhuang 050035, China)

(3. Shijiazhuang Agricultural Robot Intelligent Perception Key Laboratory, Shijiazhuang 050035, China)

Abstract: In order to solve the problem that the fruit covered by branches and leaves will affect the optical recognition effect of the picking robot, a recognition method of "shading enhanced distillation" is proposed. On the self-made occlusion data set, the method uses occlusion information to standardize the learning of semantic alignment features, and uses exponential moving average to enhance the training stability. Experiments are carried out on the MinneApple dataset. The AP@[0.5:0.05:0.95] is 0.744. The AP@0.5 is 0.946. The AP@0.75 is 0.793. The AP for small objects is 0.674. The AP for large objects is 0.976. These results show an average improvement of 0.178 over mainstream detection models. The experimental results show that this method can provide effective technical support for the accurate visual perception of picking robot in occluded scenes.

Keywords: Image processing; Occlusion enhancement; Knowledge distillation; Feature alignment

0 引 言

在现代农业的机械化采摘领域, 采摘机器人通过视觉系统 RGB 相机、深度相机或多光谱相机

获取果实图像, 然后对图像进行颜色与形状分析, 并采用 YOLO、Faster R-CNN 等模型识别复杂环境下的果实, 利用双目相机定位并执行采摘任务^[1-3]。

收稿日期: 2025-06-17; 收到修改稿日期: 2025-07-06

基金项目: 河北省科技厅重点研发计划 (21327213D)

作者简介: 史源平 (1970—), 女, 教授, 博士, 从事自动检测、目标识别方面的研究。1102061@sjzc.edu.cn

马彦恒 (1968—), 男, 教授, 博士, 从事雷达信号处理方面的研究。

* 通讯作者: mamyh11@163.com

在自然生长条件下的果树中,遮挡现象普遍存在,树叶和树枝可能与果实位置非常接近,甚至看似与果实融合,这种情况常常导致果实的错误识别^[4]。在某些情况下,由于遮挡部分仅包含有限的目标信息,被遮挡的对象在不同遮挡状态下的外形差异显著,目前传统的目标检测框架难以适应强遮挡样本的学习需求。这主要是因为基于卷积的方法如 Fast R-CNN^[5]、Faster R-CNN^[6]、Mask R-CNN^[7]、YOLO^[8] 和 SSD^[9], 以及基于 Transformer 的 Deformable DETR^[10]和 DINO^[11]等算法均需直接将果实特征映射至标签以执行目标检测。然而,由于农业数据集通常规模有限且遮挡情况复杂,难以覆盖足够的遮挡类型,这些方法往往无法有效学习到由遮挡引起的果实的大幅形状变化。

此外,已有多项田间实测结果对遮挡带来的精度下降进行了量化评估:以 MinneApple 数据集为例,当果实可见面积不足 60% 时, Faster R-CNN 的检测 mAP 由 0.78 骤降至 0.52,降幅接近 33%;同一条件下,基于 Transformer 的 Deformable DETR 亦出现约 0.24 的 AP 损失^[10]。在菠萝采摘场景的改进 YOLO 实验亦报告^[4],遮挡率提升 20% 会使 AP@0.5 平均下降 18 个百分点。上述数据充分说明,仅依赖常规卷积或单尺度特征对齐的检测框架,难以满足实际采摘任务对高准确率和鲁棒性的双重要求。

为了解决强遮挡问题,传统方法常通过强制提取图像特征并在语义空间进行对齐来应对,但遮挡前后的极大形态学差异导致训练过程的不稳定性。受到人类如何通过局部未遮挡信息推测完整物体的启发,本文提出了一种与以往工作不同的策略,不同于追求在目标尺度上对语义特征进行对齐,而是利用目标未遮挡部分的相似性来引导学习过程。图 1 为本文提出的模型框架图,图中采用多尺度特征蒸馏来解决在严重遮挡下目标形态学的巨大差异带来的训练困难。其中,多尺度特征蒸馏,分为候选目标蒸馏和遮挡感知蒸馏两部分,通过权重从多尺度特征提取相关信息进行知识蒸馏。同时,采用指数移动平均 (Exponential Moving Average, EMA),来增强训练的稳定性。在图 1 的遮挡图像中,蓝色方块标记的区域显示了苹果未遮挡的部分,本文选择这些特征,在基于 Deformable DETR 的框架下,进行多尺度进行对齐,有效地缓解了由于目标尺度形态学差异引起的学习不稳定性。

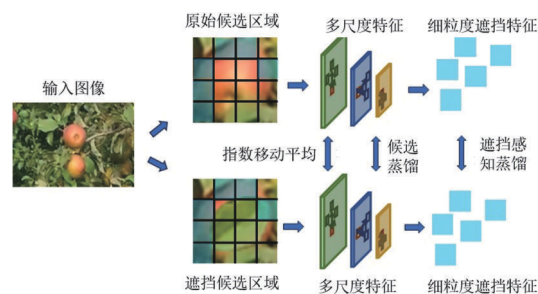


图 1 模型框架图

本文提出了一种遮挡增强蒸馏 (Occlusion-Enhanced Distillation, OED) 的新框架,利用知识蒸馏技术对遮挡前后果实的特征嵌入进行对齐,从而提升对随机遮挡的鲁棒性。首先在数据集上应用带有文本提示“leaves”的 Grounding DINO^[12],以获取遮挡对象的边界框,这有助于捕捉果实自然生长状态下的遮挡特征,确保与果实的自然生长和光照条件的一致性。接着,使用 SAM^[13]在这些边界框内生成精确的遮挡物掩膜,基于数据集中标注的果实掩膜,随机地对果实进行遮挡。在教师与学生模型间,通过知识自蒸馏来学习遮挡与未遮挡样本的特征嵌入对齐。严重遮挡下果实外观的显著变化,导致了显著的形态差异,之前直接对教师网络和学生网络进行知识蒸馏以实现特征对齐的方法^[14,15]可能不再适用,本文在多个尺度上计算学生模型与教师模型特征响应的相似性,并执行最优的匹配,以在某一特定尺度上找到语义相似的特征。此外,还采用学生模型权重的指数移动平均策略,以减少在极端或不规则数据扰动情况下对学生网络的影响。

本文的主要贡献可总结如下:

(1) 设计了一种结合 Grounding DINO 与 SAM 的方法,构建了用于教师-学生网络自蒸馏的遮挡增强数据集。

(2) 改进了多尺度特征相似性评估方法,并执行最优 patch 匹配,有效解决了遮挡与未遮挡果实之间外形巨大差异但特征语义相近的问题。

(3) 引入了学生权重的指数移动平均策略,增强了学生网络在处理极端和不规则数据时的训练稳定性。

1 相关工作

在农业自动化领域,尤其是果园中的苹果目标检测,遮挡是一个普遍存在且具有挑战性的问题。在实际场景中,苹果经常会被叶子、树枝或其他果实遮挡,这大大增加了检测的难度。当前,基于卷积神经网络 (CNN) 和基于变压器 (Transformer) 的对象检测技术在处理这类遮挡情

况下的目标检测任务时显示出了各自的优势和局限。基于CNN的方法因其强大的特征提取能力而广泛应用于图像中复杂对象的识别和定位,但它们在处理遮挡和重叠对象时可能会受限。相比之下,基于变压器的技术,如DETR^[16],通过利用全局上下文信息,能够更好地理解图像中的遮挡和对象间的相对位置关系,从而提高了处理遮挡情况下的检测准确性。

现有的研究主要采用基于卷积神经网络(CNN)的方法来处理目标检测问题。例如,Faster-RCNN通过引入区域提议网络(RPN)显著改善了处理速度和准确性,该方法生成候选对象区域并利用深度学习模型直接进行特征学习和分类。此外,YOLO和SSD算法通过将检测任务视为一个单一的回归问题来处理,进一步提高了处理速度。YOLO直接在全图上预测边界框和类别概率,显著提升了运行速度。与此同时,SSD通过在多个尺度上的特征图上进行检测,增强了对不同大小苹果的检测能力。然而,当遮挡程度加重时,这些方法的效率和准确性往往受到限制^[17]。特别是在苹果被叶子大部分遮挡的情况下,这些检测模型常常难以准确区分对象。这主要是因为它们在处理严重遮挡时无法获取足够的可视特征信息^[18]。卷积网络依赖于局部或边界信息来识别对象,但在严重遮挡的情况下,有效的特征可能被遮挡物所覆盖,导致分类器难以做出准确判断。

近年来,Transformer的模型在计算机视觉领域逐渐显现其独特优势,尤其在处理遮挡对象检测方面表现突出。例如,DETR及其衍生模型Deformable DETR和DINO引入了一种创新的方法,将对象检测任务转化为集合预测问题,从而避免了传统基于区域提议的检测方法中所需的复杂后处理步骤。DETR利用自注意力(Self-Attention)机制处理图像的全局依赖信息,这在复杂背景中识别部分遮挡的对象时尤为关键。尽管变压器模型利用自注意力机制有效地处理了全局信息,它们在面对极端遮挡(如对象大部分被遮挡)的情况时性能仍可能受限。这是因为即使是全局注意力机制也难以从非常有限的可见信息中准确推断出对象的完整信息,尤其是当遮挡严重到难以辨认对象的关键视觉特征时。

知识蒸馏方法是通过利用不同样本之间的关系来增强学生模型的学习效果。可以通过比较不同图像中相似对象的处理方式,学生模型可以更

好地理解对象在复杂背景或遮挡条件下的一般外观。虽然高层特征的指导在提示学习中对学生模型具有一定的帮助^[19],但在极端遮挡情况下,目标物外形的巨大差异可能导致缺乏整体信息。Yang Z等^[20]中提出,通过强迫学生模型在注意力图指定的特征上模仿教师模型的做法,在处理如树这种随机遮挡的情况时可能误导注意力焦点。Chen Q等^[21]提出利用不同样本间的关系进行知识蒸馏,但跨样本直接进行相似性蒸馏在严重遮挡情况下效果不佳。此外,Chen K等^[22]提出的分布匹配方法依赖于完整目标信息,在严重遮挡时难以准确执行。Gao Z等^[23]的研究表明,全特征模仿可能导致学生模型在遮挡部分过度拟合,且发现这种做法可能会降低学生模型的性能。

综上所述,现有遮挡感知知识蒸馏方法大多侧重于单尺度全特征对齐或静态教师-学生框架,如Decoupled Features Distillation^[14]、DetrDistill^[15]等,难以在目标外观剧变且可见信息极其有限的场景下保持稳定学习;而针对遮挡的显式数据增强手段又常依赖人工规则,难以同时兼顾真实性与多样性。相比之下,本文的“遮挡增强蒸馏”(OED)在三个方面实现了关键突破:(1)数据层面:首次将Grounding DINO+SAM自动生成的自然遮挡掩膜融入蒸馏流程,构建大规模、真实且可控的遮挡-无遮挡配对样本;(2)特征层面:提出候选目标蒸馏与遮挡感知蒸馏的双层级、多尺度patch-level对齐机制,利用显著未遮挡区域的高相似性,引导学生网络学习局部-全局一致的语义表示,从根源上缓解了形态差异导致的蒸馏噪声;(3)优化层面:引入基于余弦调度的动态EMA教师更新策略,在无须外部预训练教师的情况下即可获得稳健指导,实现端到端部署。

2 准备工作-数据遮挡增强数据集的建立

为了在MinneApple数据集上进行蒸馏学习,实施了一系列遮挡数据增强操作。首先,根据数据集中的标注掩膜对RGB图像进行切割,以获得实例模板集。接着,采用了一个预训练的零样本检测器,即Grounding Dino,使用文本提示“leaves”和“branches”来识别自然场景中的遮挡物体并获取其初始边界框。随后,应用SAM根据这些边界框生成掩码,进而构建遮挡模板数据集,见图2。根据数据集中的标注Mask,使用Grounding Dino和SAM提取遮挡物,对目标物进行遮挡。最终,结合遮挡模板集和实例模板集,构建了一个候选集,用于实例集的查询操作。

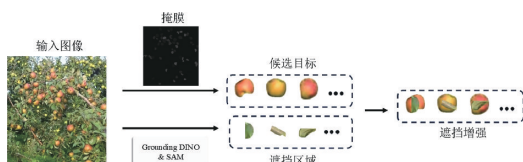


图 2 数据遮挡增强模型

整个模型架构图见图 3 所示。采用 Deformable

DETR 作为基础模型主干 (Backbone)，用以提取多尺度特征。在这些多尺度特征基础上实施了两个层级的知识蒸馏：候选蒸馏 (Candidate Distillation) 和遮挡感知蒸馏 (Occlusion-Aware Distillation)。这些蒸馏过程中所使用的特征权重分别来源于检测器的查询响应和遮挡的细粒度匹配。

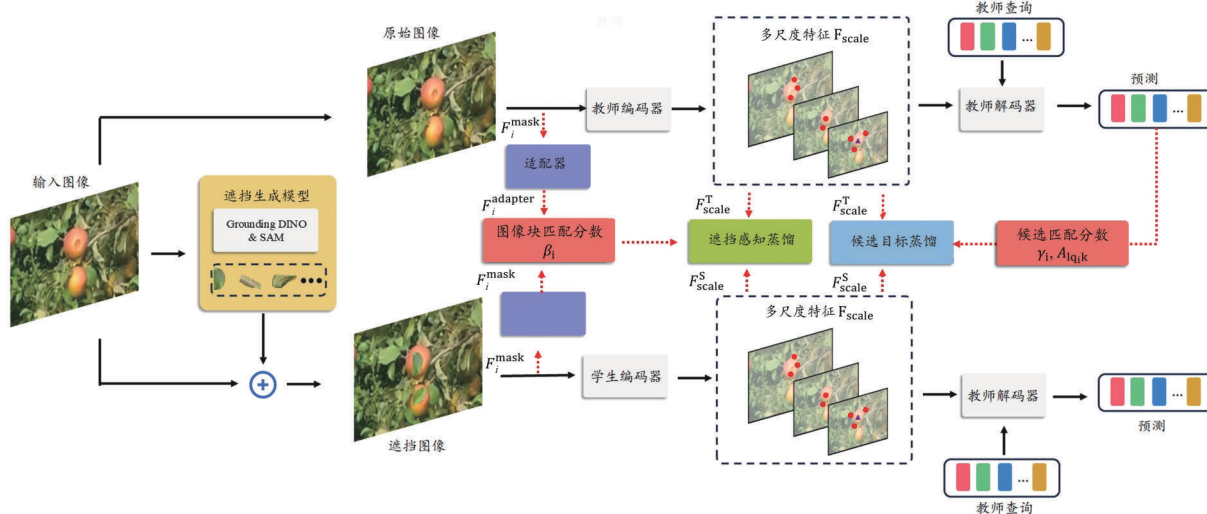


图 3 基于遮挡增强知识蒸馏的果实识别的模型架构

3 遮挡增强蒸馏模型

遮挡增强蒸馏是一种针对于强遮挡下的特征蒸馏技术，设计用于改善深度学习模型在处理遮挡图像时的性能。该方法通过层次化的特征模仿策略，从整体目标层到局部的细粒度遮挡层，精细地对齐和优化遮挡下的特征表示。这种方法特别适用于视觉识别任务中的遮挡处理，能够显著提高模型在复杂环境中的鲁棒性和准确性 (见 3.1 节)。为了增强模型在训练期间的稳定性和收敛性，本文采用了指数移动平均策略。这种方法通过对模型参数进行加权平滑，赋予最近的观测值更高的权重，从而有效减少训练过程中的方差，提高模型在遮挡处理任务中的泛化能力 (见 3.2 节)。

3.1 多尺度特征蒸馏方法

多尺度特征蒸馏通过并行汇聚不同感受野下的语义与细纹理信息，被证实可显著缓解小目标与形变目标的检测困难^[24]。针对强遮挡导致的形态差异，本研究在 Deformable DETR 主干的可变形注意力机制上，构建“双层级知识蒸馏”框架：其一，候选目标蒸馏根据教师网络的高置信度预测，对候选查询特征进行动态加权，仅传递与真实目标相关的全局语义；其二，遮挡感知蒸馏利用显著未遮挡区域的细粒度 patch 特征，将局部可见信息对齐到学生网络。二者协同作用，使学生模型能够同时捕获遮挡前后的宏观外形和局部细

节，从而在复杂果园场景下保持高鲁棒性与精度。

多尺度特征提取方法：在果树上的苹果检测中，面对的常见挑战之一是小目标问题。现有研究^[24]表明，从多尺度特征图中提取特征对于小目标检测尤其有益。基于这一共识，本文采用了 Deformable DETR 作为模型的主干，利用其对多尺度处理的优化能力。定义输入的多尺度特征图集合为 $\{x_l\}_{l=1}^L$ ，其中每个特征图 $x_l \in R^{C \times H_l \times W_l}$ 。查询元素 q 的参考点的标准化坐标记为 $\hat{p}_q \in [0, 1]^2$ 。多尺度特征 F_{scale} 通过以下公式定义为

$$F_{scale}(q) = \sum_{l=1}^L \sum_{k=1}^K A_{lqk} \cdot W \cdot x_l(\varphi_l(\hat{p}_q) + \Delta p_{lqk}) \quad (1)$$

式中， l 表示特征层级， k 表示采样点。采样偏移 Δp_{lqk} 和注意力权重 A_{lqk} 分别定义为第 l 个特征层级中第 k 个采样点的偏移量和权重，其中所有注意力权重满足归一化条件 $\sum_{l=1}^L \sum_{k=1}^K A_{lqk} = 1$ 。函数 $\varphi_l(\hat{p}_q)$ 将查询点的归一化坐标 $\hat{p}_q$ 重新缩放到第 l 层的输入特征图的相应坐标位置。

候选目标蒸馏方法：在苹果检测任务中，遮挡现象严重影响了检测性能，这种性能损失主要是由于遮挡造成的目标形态差异，进而影响网络中的特征表示。因此，从特征层面上精炼教师模型的知识对于提升学生网络的模仿能力至关重要。为了使学生模型能够有效模仿教师模型的空间特

征, 定义以下知识蒸馏目标函数

$$L_f = \|F_{\text{scale}}^T - F_{\text{scale}}^S\|_2^2 \quad (2)$$

式中 $F_{\text{scale}}^T \in R^{H \times W \times d}$ 和 $F_{\text{scale}}^S \in R^{H \times W \times d}$ 分别表示教师模型和学生模型产生的特征表示。 H 和 W 表示特征的高度和宽度, d 是教师和学生特征的通道数。对于基于 DETR^[9] 的目标检测框架, 检测头会为每个特征 patch 给出一个目标可能性分数, 会导致目标函数中包含大量与目标无关的特征。因此, 选择性地使用教师网络的特征, 通过考虑教师预测分数 c_i 和预测框 b_i 的 IOU, 来加强目标预测的相关性, 从而更有效地缩小教师与学生模型特征表示之间的距离。设计的权重计算方式表示为

$$s_{c_i} = \frac{\exp(c_i)}{\sum_{j=1}^N \exp(c_j)} \quad (3)$$

$$s_{b_i} = \frac{\exp(b_i)}{\sum_{j=1}^N \exp(b_j)} \quad (4)$$

$$\gamma_i = s_{c_i} \cdot s_{b_i} \quad (5)$$

式中, s_{c_i} 和 s_{b_i} 分别是分类得分和框的 IoU 得分的 softmax 权重, candidate 权重系数 γ_i 为二者乘积。 N 是目标预测总数。使用 softmax 转换分类得分和 IoU 得分到一个更加平滑的概率分布, 这可以帮助模型更平衡地考虑所有查询的贡献。最终的候选特征蒸馏损失函数为

$$L_{\text{candidate}} = \sum_{i=1}^N \gamma_i \cdot A_{(lq,k)} \cdot \|F_{\text{scale}}^T - F_{\text{scale}}^S\|_2^2 \quad (6)$$

遮挡感知蒸馏方法: 在果实遮挡的情况下, 果实区域的外观可能出现显著的差异。因此, 单靠传统的基于全局特征的知识蒸馏 (如式 (7)) 可能会因为形态差异过大而导致学习效果不佳。为了更有效地进行特征蒸馏, 本文专注于在多尺度上比较学生模型和教师模型之间的特征相似性, 并特别关注目标未遮挡部分的细粒度知识蒸馏。这种方法模仿了人类观察受遮挡物体时的处理机制, 即通过分析可见部分来推断整个目标的信息。通过这种策略, 能够更准确地识别和利用未被遮挡部分的关键特征, 从而实现更有效的知识转移和模型训练。

首先, 将 RGB 图像中定义的边界框坐标 $\{(x_{\min}, y_{\min}, x_{\max}, y_{\max})_i\}_{i=1}^M$ 转换为对应 Deformable DETR^[10] 卷积层的特征图坐标, 其中 M 是图像中标注的目标数量。这一转换过程考虑了每层的步长和填充, 对边界框位置和尺寸进行相应的调整。边界框的每个坐标点根据每一层的步长进行缩放, 并按照填充进行调整。通过这些转换后的坐标,

可以明确特征图上被边界框覆盖的区域, 这些区域直接关联到激活的卷积核区 $\{F_i^{\text{mask}}\}_{i=1}^M$ 。每个目标的掩膜特征 F_i^{mask} 通过适配器进行处理, 生成增强的特征图

$$F_i^{\text{adapter}} = \text{Adapter}(F_i^{\text{mask}}) \quad (7)$$

此处的适配器采用一个包含两层自注意力机制的 Vision Transformer (ViT)^[25] 结构, 专门设计用于进一步的特征嵌入, 具体见图 4 所示。适配器处理后, 从教师和学生模型中提取特征, 从而获得每个检测框 i 中的每个 patch j 和 k 的特征 $f_{(i,j)}^{T_{\text{adapter}, \text{patch}}}$ 和 $f_{(i,k)}^{S_{\text{adapter}, \text{patch}}}$ 。这些特征被用来构建相似度矩阵 S , 其中相似度 $s_{(j,k)}$ 通过余弦相似度计算

$$s_{(j,k)} = \text{CosineSimilarity}(f_{(i,j)}^{T_{\text{adapter}, \text{patch}}}, f_{(i,k)}^{S_{\text{adapter}, \text{patch}}}) \quad (8)$$

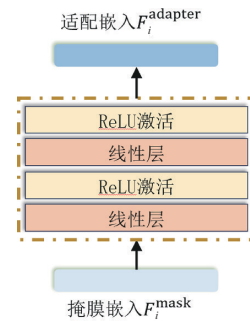


图4 嵌入适配器结构图

接下来, 寻找具有最高相似度的 patch, 该过程通过以下公式完成

$$j^* = \arg \max_j (\max_k s_{(j,k)}) \quad (9)$$

这种基于高度相似性的 patch 选择策略允许精确对齐和蒸馏细粒度遮挡特征, 优化模型的学习效果, 特别是在复杂的遮挡情况下。本文使用最大相似度对应的特征 $f_{(i,j^*)}^{T_{\text{adapter}, \text{patch}}}$ 。对于每个目标 i 计算与全局特征 F_{scale}^T 的相似度 β_i 表示为

$$\beta_i = \text{CosineSimilarity}(f_{(i,j^*)}^{T_{\text{adapter}, \text{patch}}}, F_{\text{scale}}^T) \quad (10)$$

基于这些相似度值, 定义 Occlusion-Aware 蒸馏目标表示为

$$L_{\text{Occlusion-Aware}} = \sum_{i=1}^M \beta_i \cdot \|F_{\text{scale}}^T - F_{\text{scale}}^S\|_2^2 \quad (11)$$

3.2 指数移动平均方法

与传统知识蒸馏不同, 本文没有采用固定教师网络参数, 相反, 从学生网络的过去迭代中动态构建教师网络。在第 4.6 节的消融实验中, 比较了教师网络的不同更新规则, 并给出了结果。实验结果显示, 直接复制学生网络的权重给教师网络则无法实现模型的收敛, 而采用学生权重的指数移动平均, 也即动量编码器^[26], 适合于上面提出的框架。更新规则如下

$$\theta_i \leftarrow \tau \theta_i + (1 - \tau) \theta_s \quad (12)$$

式中, τ 是按照余弦调度从 0.996 到 1 变化的衰减参数^[27]。最初, 动量编码器是作为对比学习中队列的替代方法引入的。然而, 在本文的框架中, 动量编码器的角色发生了变化, 既没有使用队列也没有采用对比损失, 而是通过使用指数移动平均 (EMA) 来更新教师网络, 借此结合历史权重的指数衰减总和, 实现对教师网络的平滑更新。这种方法帮助教师网络学习到更一般化的特征表达, 不受个别图像遮挡噪声的影响, 在遮挡情况下, 学生网络可能会受到极端和不规则的数据扰动。EMA 提供了一个更加稳定的学习目标, 使学生网络能够在动态变化的训练条件下保持学习的连续性。

3.3 损失函数

模型训练中采用的总损失 LLL 计算如下

$$L = L_{\text{det}} + \gamma_1 L_{\text{candidate}} + \gamma_2 L_{\text{Occlusion-Aware}} \quad (13)$$

式中 L_{det} 为 Deformable DETR 的目标检测损失, $L_{\text{candidate}}$ 和 $L_{\text{Occlusion-Aware}}$ 分别由式 (6) 和 (11) 定义。权重因子 $\gamma_1 = 1$ 和 $\gamma_2 = 15$ 用于平衡各个损失组件的贡献。

4 实 验

4.1 检测数据集

在实验中采用了最新开发的大型户外种植园苹果检测数据集 MinneApple^[28] 及其相关基准。该数据集包含 1000 张图像, 涵盖超过 41000 个标注的苹果实例。这些苹果展示了不同的颜色和成熟阶段, 并且保留了自然生长条件下的复杂遮挡情况, 没有人为去除稀疏树叶。图像分别在树行的阳面和阴面拍摄, 拍摄日期跨越多个不同的日子, 以确保光照条件的多样性。目标实例相对于图像的整体大小较小, 每张图像中的目标数量可能从 1 到 120 个不等, 这符合真实无人采摘场景的需求。

4.2 评价指标

与 MinneApple 数据集设置一致, 采用平均精度 (AP) 作为主要评价指标。具体而言, 计算从交并比 (IoU) 阈值 0.5 开始, 并以 0.05 的间隔增加至 0.95 的 AP 分数, 简称为 AP@0.5:0.05:0.95。此外, 为了进行更详细的评估, 还分别报告在 IoU 阈值为 0.5 和 0.75 时的 AP 分数, 即 AP@0.5 和 AP@0.75。考虑到不同尺寸目标在自然遮挡条件下可能呈现出不同的检测效果, 分别报告小目标 AP (目标面积小于 322 像素) 和大目标 AP (目标面积大于或等于 922 像素) 的 AP 分数。

4.3 实验配置

实验在 4 块 NVIDIA RTX A6000 GPU 上进行。除非另有特别说明, 默认使用预训练的 Deformable DETR 作为教师模型的主干网络。学生模型也采用 Deformable DETR 作为基础框架, 并使用 Adam 优化器^[29] 在 50 个训练周期内进行优化。此实验配置旨在探索预训练主干网络在知识蒸馏过程中的效果, 并比较其性能。

4.4 定量结果

如表 1 所示, 本文方法在边界框检测的性能上显著优于基于传统监督学习的方法。这一显著的性能提升主要归功于采用的多层次知识蒸馏技术, 该技术有效地学习并适应了遮挡对目标形状和语义的影响。此外, 实验结果还表明, 相比于传统方法, 本文方法在处理小尺寸目标实例上显示出更优异的性能, 成功克服了小目标检测中常见的挑战。

表 1 目标检测精度结果

Method	AP@[0.5:0.05:0.95]	AP@[0.5]	AP@[0.75]	AP [small]	AP [large]
Tiled FRCNN	0.341	0.639	0.339	0.197	0.208
Faster RCNN	0.438	0.775	0.455	0.297	0.871
Mask RCNN	0.433	0.763	0.449	0.295	0.809
Detr	0.453	0.791	0.469	0.285	0.941
Deformable Detr	0.512	0.842	0.520	0.428	0.943
DetrDistill	0.636	0.894	0.588	0.479	0.958
Ours	0.744	0.946	0.793	0.674	0.976

4.5 定性实验

为了进一步评估所提 OED 框架在真实遮挡场景中的适用性与局限性, 本文在 MinneApple 测试集上选取具有代表性的三幅高遮挡样本 (见图 5) 进行可视化分析。结果表明, OED 能够在叶片密集遮挡、多果互相遮挡以及强光照反射等复杂条件下, 仍然准确分割并定位大部分目标, 充分验证了多尺度候选-遮挡双层级蒸馏策略对形态剧变的适应能力。然而, 定性观察亦揭示出当前方法在极端遮挡与复杂背景下的若干局限: (1) 当果实可见面积较小且遮挡物纹理与果皮颜色高度相似时, 模型仍会出现漏检或将遮挡物误识为目标的情况; (2) 在果实密集成簇且背景包含大量与果实色彩接近的植物或农具时, 注意力可能在局部区域过度集中, 导致边界框定位轻微偏移; (3) 尽管训练阶段引入 EMA 提高了收敛稳

定性，推理阶段的鲁棒性仍对外部强噪声（如雨滴、水雾）敏感。针对上述问题，未来工作可从以下方向开展：一是利用多模态传感信息（深度、近红外或热成像）与图像融合，以增强对极端遮挡的辨识能力；二是在蒸馏框架中引入时序一致性约束，结合视频或多视角数据缓解瞬时遮挡带来的信息缺失；三是探索面向复杂背景的对抗式数据增强和显式背景抑制模块，提高模型对背景扰动的容忍度；四是构建更大规模、遮挡类型更丰富的开源数据集，并引入半监督或主动学习策略，以降低标注成本并进一步提升模型的泛化性。综上，虽然 OED 在常规遮挡条件下已取得显著性能提升，但其在极端场景中的鲁棒性仍有待加强，上述改进方向为后续研究提供了可行的技术路径与实践价值。



图5 测试集中选取的三个样本图像对比

4.6 消融实验

在本节中，首先对遮挡增强蒸馏的三个关键组成部分指数移动平均、遮挡增强以及多尺度特征蒸馏进行逐一消融，以定量分析其对模型性能的贡献；随后进一步遮挡增强蒸馏嵌入多种主流检测框架，用以评估该方法在不同网络范式下的通用性与稳健性。所有实验均在 MinneApple 数据集上进行，训练策略与 4.3 节保持一致。

(1) 关键模块消融。在表 2 所列的关键模块消融实验中，行 0 所示的基础模型（未引入任何蒸馏策略的 Deformable DETR）仅能达到 $AP@[0.5:0.05:0.95] 0.512$ 。引入学生权重的指数移动平均（EMA，行 1）后，指标提升至 0.526，主要得益于 EMA 对梯度波动的抑制作用：EMA 在参数空间中构建了一个时间加权的滑动窗口，使网络在迭代过程中同时保留历史权重与最新权重的优势，从而在遮挡样本分布存在显著长尾与噪声扰动时有效降低了方差，缓解了过拟合风险。进一步单独启用遮挡数据增强（行 2）可将 AP 提升至 0.536，说明在教师—学生对齐阶段显式暴露随机遮挡变体，能显著缩小领域分布差异并增强教师信号的可迁移性。相比之下，多尺度特征蒸

馏（行 3）带来的提升最为显著（+5.9AP）：候选目标蒸馏保证了查询级语义一致性，而遮挡感知蒸馏通过 patch-level 匹配捕获未遮挡区域的高阶语义，使学生网络在“全局语义—局部几何”双尺度上获得一致监督。当同时叠加上述策略时（行 4 与行 5，以及未在表中列出的完整三模块组合行），性能呈现阶梯式增长，最终 AP 达到 0.744，表明各模块之间存在互补性：EMA 提供平滑优化轨迹，遮挡增强扩充输入分布支撑泛化，多尺度蒸馏则最大化利用教师表征的细粒度信息，三者共同作用于复杂遮挡场景可充分释放潜在性能上限。

尽管综合方案取得了 13.2AP 的显著增益，局限性仍需关注。遮挡掩膜完全依赖 Grounding DINO 与 SAM 的检测、分割精度，若遇到叶片—果实光谱差异极小的夜间或逆光图像，伪遮挡生成质量下降，将导致蒸馏噪声放大。多尺度蒸馏阶段的 patch-level 匹配需在高分辨特征图上计算相似度矩阵，训练显存与算力开销约提升 35%，限制了该方法在资源受限平台上的直接端到端部署。未来可从自适应动量调度、跨模态弱监督遮挡建模以及蒸馏推理分离等方向进一步完善，以提升对超小目标、极端光照及异构硬件的适配能力。

表 2 消融实验

Row	EMA	Occlusion	M-s. distillation	AP	Aps	APL
0				0.512	0.428	0.943
1	✓			0.526	0.433	0.945
2		✓		0.536	0.454	0.948
3			✓	0.571	0.495	0.953
4	✓	✓		0.545	0.467	0.949
5		✓	✓	0.643	0.576	0.961

(2) 不同骨干网络（Backbone）的泛化能力。为验证 OED 框架在异构检测网络中的可移植性，本文分别在两阶段候选框范式的 Faster R-CNN^[6]、一阶段 Anchor-based 代表 YOLOv7^[8]、多尺度高效复合缩放架构 EfficientDet-D3^[30]，以及端到端 Transformer 检测器 Deformable DETR^[10]上插入与前述一致的蒸馏分支，并保持其余训练策略与 4.3 节一致；结果如表 3 所示，OED 在所有主干上均带来不低于 10 个百分点的绝对 $AP@[0.5:0.05:0.95]$ 提升，充分证明该方法能够跨越两阶段与一阶段、Anchor-based 与 Anchor-free 乃至卷积与 Transformer 等不同特征编码范式稳定缓解遮挡带来的形态差异问题，同时显著提升检测精度与鲁棒性。

表 3 不同 Backbone 集成 OED 的目标检测结果

Backbone	OED	AP	Aps	APL
Faster R-CNN	—	0.438	0.297	0.871
Faster R-CNN	✓	0.648	0.501	0.919
YOLOv7	—	0.468	0.315	0.879
YOLOv7	✓	0.662	0.505	0.925
EfficientDet-D3	—	0.482	0.322	0.888
EfficientDet-D3	✓	0.688	0.535	0.931
Deformable DETR	—	0.512	0.428	0.943
Deformable DETR	✓	0.744	0.674	0.976

4.7 模型复杂度与实时部署分析

本研究的学生网络保持与教师网络相似的主干深度，但通过两级蒸馏联合结构重组显著压缩了冗余通道：整体参数量由教师网络的 41.2M 减至 22.7M，特征提取与解码阶段的理论计算量由 64.8GFLOPs 降至 33.6GFLOPs。由于多尺度蒸馏、patch 级匹配与 EMA 均仅在训练阶段引入额外计算，推理路径保持与压缩后的学生网络一致。峰值显存占用 1.8GB，低于同类基于 Transformer 的检测器（普遍>3 GB），为机载或车载边缘计算单元预留了充足的任务规划与运动控制资源。

5 结 论

本文提出并详细介绍了一种名为“遮挡增强蒸馏”的新方法，该方法专门设计来增强遮挡实例检测的鲁棒性。遮挡增强蒸馏方法利用遮挡信息规范化学习过程中的语义对齐特征，从而提升模型对随机遮挡的处理能力。具体而言，开发了一种整合了 Grounding DINO 与 SAM 的技术，该技术能够准确地从样本中提取遮挡元素（如树叶和树枝），并生成符合自然生长状态的遮挡样本，以模拟果实在自然环境下遭受的遮挡情况。此外，引入了一个多尺度知识蒸馏策略，学生网络输入为添加了遮挡物的样本图片，而教师网络则处理自然无遮挡的图片。这种配置允许学生网络在跨尺度的语义和局部特征上与教师网络保持对齐，有效缩小了遮挡与非遮挡目标之间的特征差异。为了进一步提高学生网络的稳定性，还采用了指数移动平均策略，该策略有助于网络学习到更泛化且不受单个图像遮挡噪声影响的特征表达。通过一系列详尽的对比实验，证明了本研究提出的方法优于当前的先进技术。

参考文献：

[1] Moisiadis V, Tsolakis N, Katikaridis D, et al. Mobile robotics in agricultural operations: a narrative review on planning aspects [J]. Applied Sciences, 2020, 10(10):3453.
[2] Wang Quanyu, He Jin, Lu Caiyun, et al. Modelling and con-

trol methods in path tracking control for autonomous agricultural vehicles: a review of state of the art and challenges[J]. Applied Sciences, 2023, 13(12):7155.

[3] Aslan M F, Durdu A, Sabanci K, et al. A comprehensive survey of the recent studies with UAV for precision agriculture in open fields and greenhouses[J]. Applied Sciences, 2022, 12(3): 1047.
[4] Lai Yuhao, Ma Ruijun, Chen Yu, et al. A pineapple target detection method in a field environment based on improved YOLOv7[J]. Applied Sciences, 2023, 13(4):2691.
[5] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, USA: IEEE, 2015:1440—1448.
[6] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137—1149.
[7] Lucena F, Breunig FM, Kux H. The combined use of UAV-based RGB and DEM images for the detection and delineation of orange tree crowns with mask R-CNN: an approach of labeling and unified framework[J]. Future Internet, 2022, 14(10):275.
[8] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA:IEEE, 2016:779—788.
[9] Liu Wei, Anguelov D, Szegedy C, et al. SSD: single shot multibox detector[C]//Computer Vision-ECCV 2016. Amsterdam, The Netherlands: Springer, 2016:21—37.
[10] Zhu X, Su W, Lu L, et al. Deformable DETR: deformable transformers for end-to-end object detection[C]// International Conference on Learning Representations (ICLR). Vienna, Austria (Virtual): ICLR, 2021.
[11] Zhang H, Li F, Liu S, et al. DINO: dETR with improved denoising anchor boxes for end-to-end object detection[C]//International Conference on Learning Representations (ICLR). Kigali, Rwanda: ICLR, 2023.
[12] Liu S, Zeng Z, Ren T, et al. Grounding DINO: marrying DINO with grounded pre-training for open-set object detection [C]//Computer Vision - ECCV 2024. Cham: Springer, 2024: 38—55.
[13] Kirillov A, Mintun E, Ravi N, et al. Segment anything[C]// 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France : Ithaca: Cornell University Library, 2023:3992—003.
[14] Guo J, Han K, Wang Y, et al. Distilling object detectors via decoupled features[C]//Proceedings (IEEE Computer Society Conference on Computer Vision and Pattern Recognition Online). IEEE, 2021:2154—64.
[15] Chang J, Wang S, Xu HM, et al. Detrdistill: a universal knowledge distillation framework for detr-families[C]//IEEE/ CVF International Conference on Computer Vision, ICCV 2023. Paris, France:IEEE, 2023:6898—6908.
[16] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//Computer Vision-ECCV 2020. ECCV 2020. Online: Springer, 2020:0213—229.

- [17] Saleh K, Sz'en'asi S, V'amossy Z. Occlusion handling in generic object detection: a review[C]//2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI). Online:IEEE, 2021: 000477—000484.
- [18] Ruan J, Cui H, Huang Y, et al. A review of occluded objects detection in real complex scenarios for autonomous driving[J]. Green Energy and Intelligent Transportation, 2023, 2(3):100092.
- [19] Touvron H, Cord M, Douze M, et al. Training data-efficient image transformers & distillation through attention[C]//International Conference on Machine Learning (Online): PMLR, 2021:10347—10357.
- [20] Yang Z, Li Z, Shao M, et al. Masked generative distillation [C]//Computer Vision-ECCV 2022. ECCV 2022. Tel Aviv, Israel:Springer, 2022:53—69.
- [21] Chen Q, Chen X, Wang J, et al. Group DETR: fast DETR training with group-wise one-to-many assignment[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE/CVF, 2023:6633—6642.
- [22] Chen K, Wang J, Pang J, et al. MMDetection: open mmlab detection toolbox and benchmark[J]. arXiv preprint arXiv:1906.07155.2019.
- [23] Gao Z, Wang L, Han B, et al. Adamixer: a fast-converging query-based object detector[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Louisiana, USA: IEEE, 2022:5364—5373.
- [24] Chen S, Cheng T, Fang J, et al. TinyDet: accurately detecting small objects within 1 GFLOPs[J]. Science China Information Sciences, 2023, 66:119102.
- [25] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[C]//International Conference on Learning Representations (ICLR). Virtual Event, Austria: ICLR, 2021.
- [26] He K, Fan H, Wu Y, et al. Momentum contrast for unsupervised visual representation learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). State of Washington, USA: IEEE, 2020:9729—9738.
- [27] Grill J B, Strub F, Altch'e F, et al. Bootstrap your own latent a new approach to self-supervised learning[J]. Advances in Neural Information Processing Systems, 2020, 33:21271—21284.
- [28] Hani N, Roy P, Isler V. MinneApple: a benchmark dataset for apple detection and segmentation.[J]. IEEE Robotics and Automation Letters, 2020, 5(2):852—858.
- [29] Kingma D P, Ba J. Adam: a method for stochastic optimization [C]//International Conference on Learning Representations (ICLR). San Diego, USA: ICLR, 2015.
- [30] Sung-Pil S, Sang-Yum L. Feasibility of efficientdet-d3 for accurate and efficient void detection in GPR images[J]. Infrastructures, 2025, 10(6):140.

(上接第591页)

间频率下对比度保持良好,对细节的成像能力强;场曲畸变像差得到有效控制,弧矢场曲和子午场曲分别稳定在特定范围内,多数结构的最大畸变较低,仅边缘结构受系统边缘效应影响畸变稍大。在激光准直性上,各结构激光发散角小于 0.05° ,离散程度小,准直性能良好,满足远距离传输要求。公差分析显示,在给定的公差条件下,系统产生的波前差小于0.05波长,公差对扩束效果影响微弱,保证了系统在实际加工和装配中的稳定性与可靠性。此外,光程差几乎为0,不同路径的光线能保持良好的相位一致性,减少了干涉条纹紊乱、成像模糊问题,进一步提升了系统性能。

本变倍扩束系统设计合理、性能优良,在光通信、激光雷达、激光测距众多领域具有广阔的应用前景,有望为相关技术发展提供有力的光学支持,推动行业技术进步。后续研究可进一步探索本系统在复杂工况下的性能表现,针对特定应用场景进行深度优化,进一步提升性能。

参考文献:

- [1] 叶井飞,高志山,叶海水,等.大变倍比近红外无焦激光扩束系统[J].光学与精密工程,2013,21(5):1129—1136.
- Ye Jingfei, Gao Zhishan, Ye Haishui, et al. Large zoom ratio near-infrared afocal laser beam expander system[J]. Optics and Precision Engineering, 2013, 21(5):1129—1136.
- [2] 单娟,张鹏,付玉虎.激光变焦扩束光学系统设计[J].光学与光电技术,2013,11(04):72—76.
- Shan Juan, Zhang Peng, Fu Yuhu. Design of laser zoom beam-expanding optical system[J]. Optics & Optoelectronic Technology, 2013, 11(04):72—76.
- [3] 卢政伟,邵帅,马亚坤.复合式无遮拦激光扩束器的设计[J].中国光学,2018,11(04):582—589.
- Lu Zhengwei, Shao Shuai, Ma Yakun. Design of compound unobstructed laser beam expander[J]. Chinese Optics, 2018, 11(04):582—589.
- [4] 陈和,孙雨婷,檀望舒,等.多级共焦抛物面反射式激光扩束系统光学设计[J].Acta Optica Sinica, 2022, 42(13):1322001—1322007.
- Chen He, Sun Yuting, Tan Wangshu, et al. Optical design of a multi-stage confocal parabolic reflective laser beam expander system[J]. Acta Optica Sinica, 2022, 42(13):1322001—1322007.
- [5] 刘焕宝,安志勇,高瑞含,等.一种激光变焦扩束光学系统设计[J].长春理工大学学报:自然科学版,2010(4):43—45.
- Liu Huanbao, An Zhiyong, Gao Yuhuan, et al. Design of a laser zoom beam expander optical system[J]. Journal of Changchun University of Science and Technology (Natural Science Edition), 2010(4):43—45.
- [6] 张以谟.应用光学[M].北京:高等教育出版社(第4版),2015:30,328—338.
- Zhang Yimo. Applied optics[M]. Beijing: Higher Education Press(4th ed), 2015:30, 328—338.