文章编号:1003－6199(2025)04－0094－08
DOI:10.16339/j.cnki.jsjsyzdh.202504016

改进 YOLOv7 和 DeepSort 的视频苹果数量检测

龚圳玮,彭伟†,田雅暄
(湖南大学物理与微电子科学学院,湖南 长沙 410082)

摘 要:单颗苹果树上苹果数量是准确估计苹果园产量的重要参数。由于苹果树上的苹果密度大并且相互叠加,很难对苹果进行自动准确的计数。大多数基于深度学习的方法是通过静态图像进行苹果检测和计数,如果拍摄的区域重复,则重复区域的苹果会被重复计数。为此,提出了一种基于视频的多目标跟踪的计数方法,首先,对 YOLOv7 模型进行改进。将注意力机制和网络的 backbone 相结合,额外增加一个小目标检测头并且在原边框回归损失中引入归一化的 Wassestein 距离(Normalized Wasserstein Distance,NWD)以提高算法对微小物体的检测能力。结果表明,改进后的 YOLOv7 模型 mAP 比原模型提高了 1.61%,达到 84.42%。其次,过滤掉 DeepSort 在跟踪目标时出现的重复目标 ID。最后,结合改进后的 YOLOv7 检测算法和 DeepSort 跟踪算法,计算出视频中不同的 ID 个数即是苹果个数,提升了整套算法检测的准确率,准确率达到 88.3%。

关键词:目标检测;多目标跟踪;损失函数;注意力机制;YOLOv7;DeepSort
中图分类号:TP39 **文献标识码**:A

Improved YOLOv7 and DeepSort for Counting Apples via Video

GONG Zhenwei, PENG Wei†, TIAN Yaxuan
(School of Physics & Electronics, Hunan University, Changsha, Hunan 410082, China)

Abstract: The number of apples on a single apple tree is an important parameter for accurately estimating the yield of an apple orchard. Due to the high density of apples on an apple tree and their overlapping, it is difficult to count apples automatically and accurately. Most deep learning-based methods detect and count apples through static images. If the captured area is repeated, the apples in the repeated area will be counted repeatedly. In response to the above problems, a counting method based on video multi-target tracking is proposed. First, the YOLOv7 model is improved. The attention mechanism is combined with the backbone of the network, an additional small target detection head is added, and the normalized Wasserstein distance (NWD) is introduced into the original bounding box regression loss to improve the algorithm's ability to detect tiny objects. The results show that the mAP of the improved YOLOv7 model is 1.61% higher than that of the original model, reaching 84.42%. Secondly, the repeated target IDs that appear when DeepSort tracks targets are filtered out. Finally, by combining the improved YOLOv7 detection algorithm and DeepSort tracking algorithm, the number of different IDs in the video is calculated to be the number of apples, which improves the detection accuracy of the entire algorithm to 88.3%.

Key words: object detection; multi-target tracking; loss function; attention mechanism; YOLOv7; DeepSort

收稿日期:2024－12－18
作者简介:龚圳玮(2000—),男,湖南常德人,硕士研究生,研究方向:深度学习,视觉方向。
† **通信联系人**,E-mail: pengwei@hnu.edu.cn

近年来现代农业结合计算机视觉技术的各种智能设备在农业生产中的应用日益普及，并且成为农业领域的研究热点。具体到苹果生产方面，基于深度学习目标检测技术可以通过相机或无人机航拍的图片或视频进行苹果识别[1]、产量预测、病虫害识别，这些技术极大地推动了苹果生产管理的智能化发展。

用于农作物产量估算的传统技术在一定程度上仍然存在诸多问题，如滑动窗口导致图像边界不清晰[2]、目标特征设计复杂、可移植性差、人工设计烦琐等问题[3]。因此越来越多的基于深度学习的目标检测算法被应用于农业生产中。这些算法大致可以分为单阶段和两阶段算法。相较于两阶段算法，单阶段算法冗余计算更少，检测速度更快，对硬件性能要求更低。代表性的单阶段算法有 YOLO 系列[4-6]、SSD 等。单阶段算法直接在输入图像上预测目标的类别或位置，检测速度得到了较大的提升，因此在农业领域得到了广泛应用。Zhao 等[7]基于 YOLOv5 算法做了改进，能够准确检测出 UAV 图像中的小麦数，准确率达到了 94.1%，比原始 YOLOv5 模型高出了 10.8%。Wang 等[5]提出 YOLOv7 比之前的 YOLO 系列具有更高的准确率。Zhu 等[8]将 transformer 预测头和卷积注意力模块（CBAM[9]）添加到 YOLOv5 中，有效提升了模型对小目标的检测精度。

上述研究都是基于静态图像进行实验的，然而在自然环境中，存在苹果相互遮挡或重叠的问题并且拍摄的静态图像视角单一，视野有限。此外，在拍摄过程中会出现相同区域的苹果出现在不同的图像中，从而导致重复计数，直接影响了产量估计的准确性，因此本研究首先在常规目标检测算法 YOLOv7 的基础上，根据苹果图像目标小且密集的特点，有针对性地改进以提高其对苹果的检测精度。然后结合 DeepSort 算法直接基于视频对苹果进行计数。Zheng 等[10]结合了 YOLO 检测算法和 DeepSort 跟踪算法对柑橘视频里的柑橘进行跟踪计数，F1 得分为 89.07%。Ge 等[11]在不同时间段检测并跟踪番茄，包括刚开花的番茄、未成熟的绿色番茄和成熟的红色番茄，精度分别为 93.1%、96.4%和 97.9%。

注意力机制（Attention Mechanism）是深度学习中一个重要的概念，是帮助模型进行注意力分配的一种技术。本文提出使用无参的三维注意力机制 SimAM[12]（Simple Attention Mechanism），将其集成于网络 backbone 输出的不同尺度的特征图之后，用于增加网络对小目标特征的表征能力。

模型损失函数的性能往往会直接影响目标检测算法的精度。YOLOv7 的损失函数包括三个部分：分类损失（Classification Loss），使用二元交叉熵（Binary Cross Entropy，BCE）计算预测框对应每个类别的概率和真实类别之间的损失；目标置信度损失（Objectness Loss），同样使用二元交叉熵（BCE）计算预测框是否包含物体的概率和真实情况之间的损失；边框回归损失（Bbox Regression Loss）：使用 CIoU 损失（Complete Intersection over Union Loss）计算预测框和真实框之间的坐标偏差。其中边框回归损失，对小物体边框的细微移动十分敏感。对于小目标而言，微小的位置偏差会导致 IoU 的急剧下降，使标签分配准确率下降。最近一些研究表明，Wasserstein 距离可以很好地反映小目标间的位置关系[13]。因此本文在计算 CIoU 损失后再融入 NWD（归一化的 Wassertein 距离），具体来说就是将边框建模为二维正态分布，然后使用 NWD 来衡量微小物体边框之间的差异，从而提升模型对微小物体的检测能力。

综上所述，针对苹果计数场景目标小且密集容易重复计数的特点，一方面本工作在现有研究的基础上，首先将三维注意力机制（SimAM）整合进兼备速度和精度的 YOLOv7 网络，以增强网络对特征的提取能力。其次，在原网络检测大、中、小三个尺寸目标检测头的基础上，额外再增加了一个专门用于检测微小物体的检测头。最后在 YOLOv7 计算 CIoU 损失的基础上融入 NWD，有效提高了模型检测小目标的能力。另一方面本工作对 DeepSort 算法的超参数做了调整，包括重新训练 ReID（Re-Identification，特征重识别）模型，让其对苹果的跟踪效果更加理想。同时对算法也做了调整，增加了计数模块，改善了 ID 跳变问题。

1　YOLOv7

当前无论单阶段或是两阶段的目标检测算法，通常都由 Input、Backbone、Neck、Prediction 这四部分构成。YOLOv7 是在 YOLOv5 的基础上改进得到的，所以在网络架构和训练、推理流程上与 YOLOv5 相似，YOLOv7 的速度与精度超越了现有大多数的目标检测算法，在检测速度和精度之间达到了很好的平衡。

YOLOv7 的 Input 模块对输入图像进行 Mosaic 数据增强和尺寸的自适应缩放（默认为 680×

680)后输入到 Backbone 模块。Backbone 模块由一系列的 CBS 模块(Conv、BatchNormalization、Silu)、卷积池化模块高效层聚合网络和扩展的高效层聚合网络等模块组成。Neck 模块采用自上而下的特征金字塔结构和自下而上的路径聚合结构即路径聚合特征金字塔网络来进行多尺度特征融合,Neck 模块最后分别输出 80×80、40×40、20×20 三种不同尺度的特征图,用以分别预测小、中、大三种目标。Head 模块主要由重参数卷积(RepConv)组成,它可以将多个计算模块合并成一个,提高模型的效率。Head 模块将 Neck 模块经过融合的特征进行解耦,得到最后所需要的类别与位置信息。

2 改进 YOLOv7

针对检测目标小且密集的特点对 YOLOv7 模型的 Backbone 模块、Head 模块和边框损失函数做了相应的改进。在 BackBone 输出的四种尺度特征图(因为额外增加了一个检测头,所以相应的 BackBone 的输出也要增加)后加上三维注意力机制 SimAM 提高网络的特征提取能力,在 Head 模块增加一个额外的小目标检测头,用以提升算法对小目标的回归精度;使用基于 CIoU 和 Wasserstein 距离的混合边框损失来代替原有的基于 CIoU 的边框损失。改进后的 YOLOv7 网络结构如图 1 所示,其中网络改进部分已重点标识。

2.1 SimAM 注意力机制

注意力机制受人类视觉系统的启发,通过对输入图像特征的权重进行动态调整实现倾向于关注特定区域[14]。SimAM 是一种简单但非常有效的注意力模块,用于卷积神经网络。与现有的基于通道和空间的注意力模块不同,SimAM 模块在不向原始网络添加参数的情况下,为特征图推断出三维注意力权重。SimAM 模块在各种视觉任务上的定量评估表明,该模块灵活且有效,可以提高许多卷积神经网络的表征能力。SimAM 根据视觉神经科学的研究,最具信息量的神经元通常是那些与周围神经元显示明显不同的激活模式的神经元。此外,一个活跃的神经元还可以抑制周围神经元的活动,这种现象被称为空间抑制。为了计算每个神经元的重要性,定义了以下能量函数[14]:

$$e_t(w_t,b_t,\boldsymbol{y},x_i)=\frac{1}{M-1}\sum_{i=1}^{M-1}[-1-(w_t x_i+b_t)]^2+[1-(w_t t+b_t)]^2+\lambda w_t^2 \tag{1}$$

其中 e_t 是目标神经元的能量函数,M 是单个通道中神经元的总数,w_t 是目标神经元的权重,b_t 是目标神经元的偏置,x_i 是除目标神经元外其他神经元的特征。$[-1-(w_t x_i+b_t)]^2$ 表示目标神经元 t 的输出与其标签之间的差异,$[1-(w_t t+b_t)]^2$ 表示目标神经元与其他神经元之间的差异,λw_t^2 是正则项,防止过拟合。

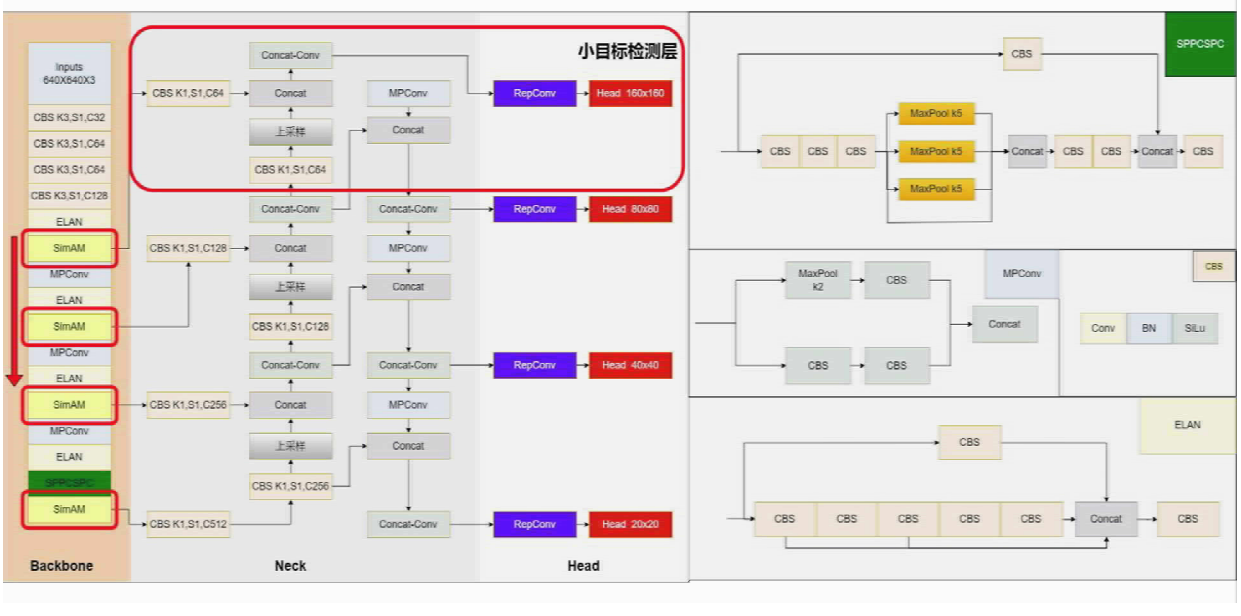


图 1 改进后 YOLOv7 的网络结构

理论上，对于每个通道，有 M 个能量函数需要求解。假设一个通道中的所有神经元都服从相同的分布，因此可以用全局均值和方差来代替局部的均值和方差，从而避免了为每个位置迭代计算均值和方差的计算成本。因此，最小能量可以通过以下公式计算：

$$e_t^* = \frac{4(\hat{\sigma}^2+\lambda)}{(t-\hat{\mu})^2+2\hat{\sigma}^2+2\lambda} \tag{2}$$

其中 $\hat{\mu}=\frac{1}{M}\sum_{i=1}^{M}x_i$ 和 $\hat{\sigma}^2=\frac{1}{M}\sum_{i=1}^{M}(x_i-\hat{\mu})^2$ 分别是在通道中所有神经元上计算的均值和方差即全局均值方差。最终 SimAM 计算公式如下：

$$\tilde{X}=\mathrm{sigmoid}(\frac{1}{E})\odot X \tag{3}$$

其中 $\tilde{X}$ 是输出特征，X 是输入特征，$\frac{1}{E}$ 是能量函数的倒数，用来表征每个神经元的重要性，$\odot$ 表示点乘运算。

2.2　增加微小目标检测头

感受野是指神经网络中每个神经元连接和感知到的输入图像区域。它反映了网络提取图像信息的范围。在卷积神经网络中，越深层的神经元，其感受野越大，这意味着它编码的特征越抽象并且全局信息越多，这有助于复杂特征的提取，但对于特征信息本来就少的小目标而言，网络的感受野越小效果越好。大的感受野容易包含大量不相关的背景信息，这些无关信息会干扰小目标的特征提取，降低检测精度。

Tang 等[15]在 YOLOv5 网络架构之上进行创新性拓展，通过引入注意力机制并增设专门针对小目标的检测头，有效提升其对小目标的检测精准度与召回率等关键指标。本文根据感受野原理以及 Tang 等的启发，在 YOLOv7 网络 head 层额外增加一个感受野更小的微小目标检测头，增加网络对微小目标的检测精度。新增检测头检测特征图大小为 160×160，如图 2 所示。它将来自浅层 backbone 部分 SimAM 提取的特征信息 1 和来自网络深层 Neck 部分特征信息 2 进行融合，得到既有小感受野信息又有高层次语义信息的融合特征，提高了网络对微小目标的敏感性。

2.3　修改边框回归损失函数计算方法

YOLOv7 使用完全交并比（Complete Intersection over Union，CIoU）作为边框回归损失函数。但是单一 CIoU 对微小目标像素位置变化过于敏感，这会导致模型微小目标检测性能急剧下降。相比之下，归一化的 Wassertein 距离（NWD），对位置偏移的变化更加平滑，更适用于微小目标的检测。

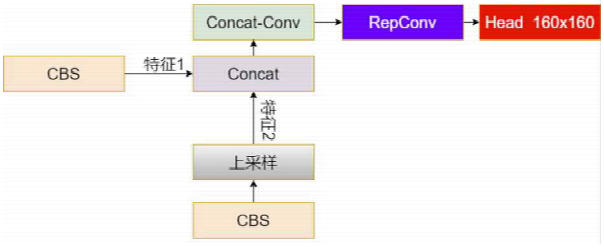


图 2　微小目标检测头

2.3.1　将边界框建模为高斯分布

在深度学习网络的训练数据中，标记的边界框或多或少会包含背景像素，其中边界框中心点的前景信息是最多，所以权重最高。像素的重要性从中心往边界递减。因此可以把边界框建模为一个二维的高斯分布[13]，具体如下：

确定一个边界框需要知道其中心点坐标（c_x，c_y），宽度（w）和高度（h）。任意一个水平边界框 $R=(c_x,c_y,w,h)$，其内切椭圆方程可表示为：

$$\frac{x-c_x}{(w/2)^2}+\frac{y-c_y}{(h/2)^2}=1 \tag{4}$$

二维高斯分布概率密度函数如下：

$$f(x\mid\mu,\Sigma)=\frac{\exp\left[-\frac{1}{2}(x-\mu)\right]\Sigma^{-1}(x-\mu)}{2\pi\mid\Sigma\mid^{\frac{1}{2}}} \tag{5}$$

其中 x 是二维高斯分布的坐标，Σ 是协方差矩阵，μ 是均值向量。当 $(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)=1$ 时，任意边界框 R 的内切椭圆就是二维高斯分布的密度轮廓，因此边界框 $R=(c_x,c_y,w,h)$ 可以建模为二维高斯分布 $N(\mu,\Sigma)$，$\mu=\begin{bmatrix}c_x\\c_y\end{bmatrix}$，$\Sigma=\begin{bmatrix}\frac{w^2}{4}&0\\0&\frac{h^2}{4}\end{bmatrix}$。至此，两个边界框之间的相似度就可以转换为两个边界框对应高斯分布之间的分布距离。

2.3.2　归一化 Wasserstein 距离

选择用归一化的 Wasserstein 距离当做两个高斯分布之间的距离。在实际的检测任务中，预测框 P 和真实框 T 对应的高斯分布为 $P\sim N(\mu_P,\Sigma_P)$，$T\sim N(\mu_T,\Sigma_T)$. 定义这两个高斯分布之间的 Wasserstein 距离如下：

$$W_2^2(\boldsymbol{N}_P,\boldsymbol{N}_T)\left\|\left(\left[cx_P,cy_P,\frac{w_P}{2},\frac{h_P}{2}\right]^{\mathrm{T}}\right)\right\|_2^2 \quad (6)$$

其中 cx_P，cy_P，表示预测框的中心点坐标，w_p，h_p 表示预测框的宽度和高度；cx_T，cy_T，表示真实框的中心点坐标，w_T，h_T 表示真实框的宽度和高度。$\|*\|_2$ 表示 2 范数。

因为 Wasserstein 距离是距离度量不能直接用来度量边界框之间的相似性，会造成损失函数难以收敛。所以对 Wasserstein 距离做归一化处理：

$$\mathrm{NWD}(\boldsymbol{N}_P,\boldsymbol{N}_T)=\exp\left(-\frac{\sqrt{W_2^2(\boldsymbol{N}_P,\boldsymbol{N}_T)}}{C}\right) \quad (7)$$

其中 C 是与训练数据集密切相关的常数，在本次苹果检测任务中，取经验值 12.8。得到 NWD 后，乘权重因子 λ 再和 CIoU 相加即得到新的边框回归损失函数：

$$\mathrm{lbox}=\lambda\times\mathrm{NWD}+(1-\lambda)\mathrm{CIoU} \quad (8)$$

如果任务中小目标占比较大，可以相应增大 NWD 的权重 λ，本次苹果检测任务 λ 取 0.5。

3 微调 DeepSort 算法

DeepSort[16] 是 Sort 算法的扩展，其主要步骤如图 3 所示。

第一步：检测器输出目标框（Detections）信息，生成新的轨迹（Tracks），通过卡尔曼滤波算法进行下一帧轨迹的预测。

第二步：检测框和预测框通过 IOU 匹配，匹配成功后会更新目标的轨迹，继续下一帧的匹配。如果匹配不成功，会出现两种情况，一种是未成功匹配的检测框（Unmatched Detections），会创建新的轨迹（Tracks）；另外一种是未成功匹配的预测框（Unmatched Tracks），会直接删除。

第三步：当一个目标连续 3 帧（默认值，可以视情况更改）都 IOU 匹配成功后，就会进行级联匹配，然后再重复第一、二步。

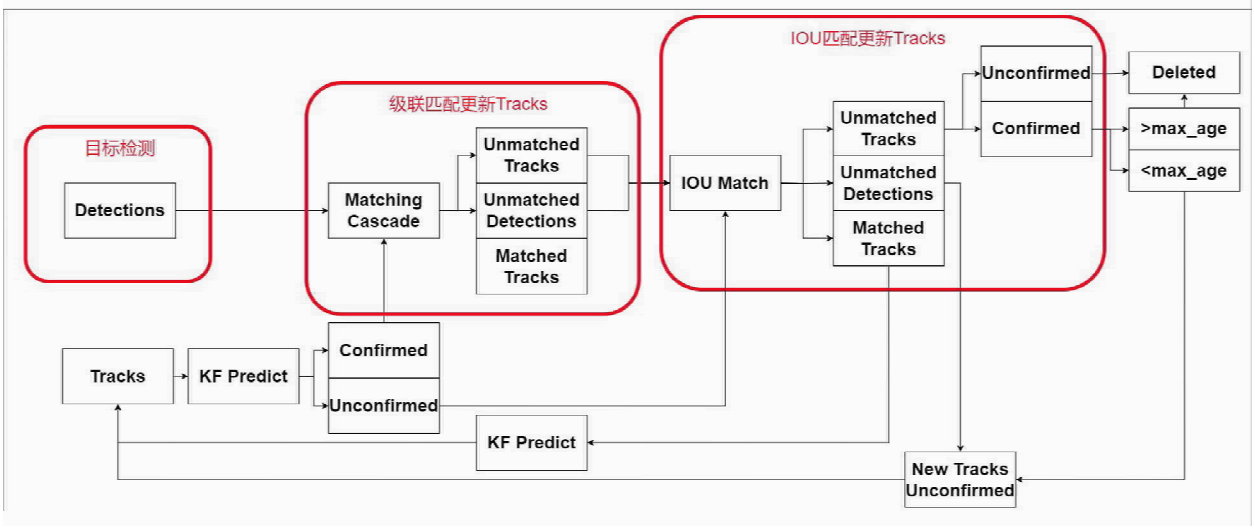


图 3 DeepSort 跟踪步骤

为了增加 DeepSort 跟踪苹果的稳定性，需要重新训练 ReID 网络模型，并且修改相应超参数。ReID 默认是用来做行人追踪的，所以其参数设置并不适合苹果追踪，首先，把网络特征输入尺寸从默认的 128×64 改为 64×64；平均池化的形状从默认的 8×4 改为 4×4。然后，修改跟踪参数：

MAX_DIST：余弦距离阈值，本实验设为 0.2，实验效果较好。

MIN_CONFIDENCE：检测器置信度阈值，本实验设为 0.3，实验效果较好。

NMS_MAX_OVERLAP：非极大值抑制阈值，本实验设为 0.5，实验效果较好。

MAX_IOU_DISTANCE：IOU 匹配阈值，本实验设为 0.7，实验效果较好。

N_INIT：表示新建一个 Tracks（Confirmed）需要的最少匹配成功次数。增大该参数可以减少误检对跟踪的干扰，但也可能漏掉一些真正的新目标，本实验设为 20，实验效果较好。

最后增加计数模块，并显示在视频上。在用 DeepSort 进行多目标追踪时，容易出现 id 跳变的

问题。主要原因是，没有成功匹配的目标也会被分配 id，这会导致 ID 号与目标真实数量相差较大，如果单纯用 ID 号来作为检测到的苹果数量，误差会较大。为此本文在原 DeepSort 算法额外增加了计数模块，过滤掉没有成功匹配的 ID，计算成功匹配目标的 ID 数来作为最终的计数结果，能有效提高计数的准确率。

4　实验

4.1　实验环境

本实验的运行环境是 AMD Ryzen Threadripper PRO 3945WX 12 核 CPU，Nvidia RTX 5000 GPU，Ubuntu 22.04 Linux 操作系统，Pycharm 编译器，Cuda 12.0，Python 编程语言：Python 3.10，torch 1.11＋cu115。

训练时关键超参数设置：本实验用到的预训练模型是 yolov7.pt，迭代次数为 300，每次输入到 GPU 的 Batch Size 为 16，输入图像大小统一设置为 640×640，采用 Adam 优化算法，采用余弦退火学习率衰减方法。

4.2　数据集

实验用数据集 Apple 是在网上公开数据集 VOCData 中，挑选了 1150 张苹果目标尽可能小的图片，来尽量模拟小目标场景，验证集数量为 400，并且验证集不在训练集中。另外一个数据集是网上公开无人机检测数据集 Drone，该数据集特点是检测目标像素小，大多为微小目标。为了进一步佐证本算法改进对小目标检测效果是有提升的，因此选择了 Drone 数据集。其中训练集包含 5200 张图片，验证集包含 2600 张图片。

4.3　消融实验

为了验证每个改进部分对模型性能的影响，以原始的 YOLOv7 作为基准模型，并依次加入改进部分，分别在 Apple 和 Drone 数据集上进行消融实验，结果详见表 1。其中，A 方法表示在网络的 backbone 层即特征提取阶段融入了 SimAM 注意力机制，B 方法表示在网络的 head 层加入额外的小目标检测头，C 方法表示在网络原始的边框损失基础上加上 NWD，作为新的边框损失函数。从表 1可以看出本文提出的方法均能有效提高模型的检测能力，特别是 C 方法在不增加任何网络参数的前提下，提高模型的检测能力。因此充分证明了融入 NWD 后的损失函数能提高模型对微小物体检测的能力。经 A、B、C 三种方法改进后的网络，在 Apple 数据集上，mAP@0.5 提升了 1.61%，如图 4 所示。在 Drone 数据集上，mAP@0.5 提升了 3.95%，提升效果明显。显然，本文的改进方法在两个数据上平均精度都有提升。但是同样的改进方法对于不同的数据集，平均精度提升效果是不一样的。在 Apple 数据集上，全改进（加上 A，B，C 三种改进方法）的效果是最好的，但是在 Drone 数据集上，部分改进（只加上 B，C 两处改进）mAP@0.5 达到了 91.1%，远好于加上 A，B，C 三处改进的效果（mAP@0.5 只有 85.32%）。结果如图 5 所示，不同数据特征应采用针对性不同的改进方法。

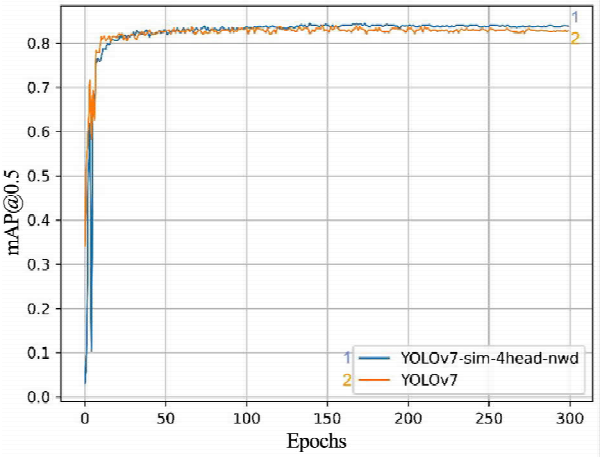


图 4　Apple 数据集全改进和原模型 map 值对比图

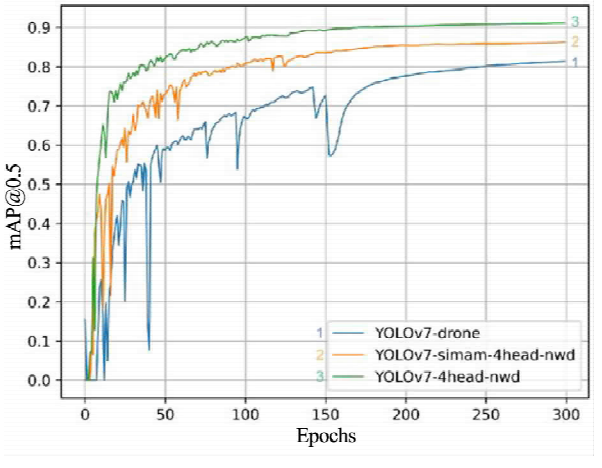


图 5　Drone 数据集全改进和部分改进 map 值对比图

为了进一步验证本文改进 YOLO 算法的效果，实验分别选取了不同场景下的苹果图像进行检测对比并且将检测到的苹果数量显示在图像左上角，现截取检测图像局部更加直观的展示改进效果，如图 6 所示。

本文的改进不仅增加了召回率，检测到了很多

原模型漏检的目标而且准确率也有所改进。原 YOLO 检测到 11 个苹果，本文改进的模型检测到了 20 个苹果，很多更小的苹果都被检测到。

4.4 苹果跟踪计数分析

验证 ReID 模型的再识别性能：如图 7 所示，图像从左到右分别是测试视频 1 的第 39，第 46，第 60 帧图片。id 为 9 的苹果，在视频第 39 帧被第一次检测到。到第 46 帧时，由于被树叶遮挡，算法丢失目标，等到第 60 帧时，再次被跟踪到。说明重新训练过的 ReID 模型能重识别苹果。

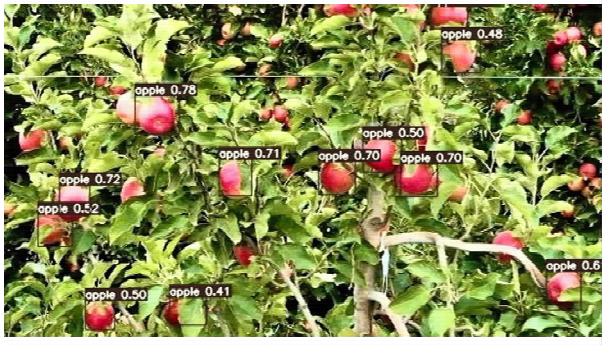
表 1 两种数据集上的消融实验

YOLOV7	A	B	C	参数量	mAP@0.5/%		mAP @0.5:0.95/%	
					Apple	Drone	Apple	Drone
√				37.20	82.81	81.37	47.90	52.32
√	√			37.46	83.74	83.03	53.26	54.30
√		√		41.68	83.03	82.65	48.53	52.58
√			√	37.20	83.12	83.01	49.95	54.05
√	√	√	√	41.99	84.42	85.32	55.04	56.12

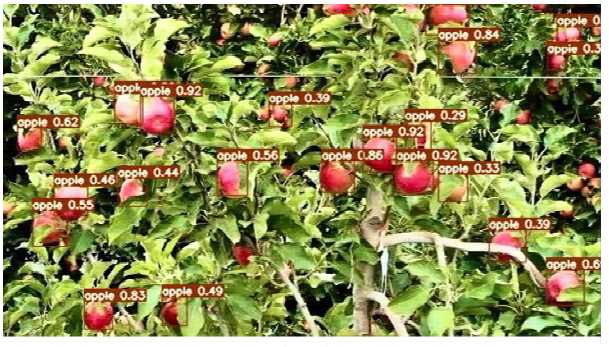
结合改进过的 YOLOv7 模型和改进过的 DeepSort 算法对拍摄的苹果视频进行计数实验，计算准确率并跟原模型进行对比，实验对比结果如表 2 所示。

表 2 改进算法和原算法计数准确率对比

算法模型	精确率 P/%	召回率 R/%	计数准确率/%
YOLOv7+DeepSort	87.4	86.0	77.5
本文算法	88.6	87.2	88.3



(a) YOLOv7



(b) 本文方法

图 6 Apple 数据集上的检测结果对比图

图 7 ReID 重识别改善效果图

5 结　论

采用改进的 YOLOv7＋DeepSort 算法，通过对拍摄的苹果视频直接计数，与基于静态图像的计数方法相比，本文提出的方法能避免对同一区域目标重复计数并且提高了计数准确率和效率。对于检测算法 YOLOv7，首先，在感受野小的网络层处额外增加一个检测层。然后在 BackBone 输出的四种尺度的特征图后加入三维注意力机制 SimAM，以增强特征的提取能力。最后，将原来的 CIoU 的边框损失和归一化后的 Wasserstein 距离加权相加作为新的边框损失函数，来改善模型对微小目标和密集目标位置偏差的敏感性。对于跟踪算法 DeepSort，首先，重新针对苹果数据集训练 ReID 模型，然后增加计数模块以改善 ID 跳变的问题，最后设置合适的跟踪超参数，达到最佳的跟踪效果。

经实验对比，本文检测算法较原 YOLOv7 在自制数据集 Apple 和公开数据集 Drone 上 map 值分别提升了 1.61 个百分点和 3.95 个百分点，验证了所提出的改进方法的有效性。整体计数准确率提升了 10.8 个百分点。YOLO 系列算法更新迭代较快，推出了性能更好的 YOLOv8、YOLOv10 等算法，后续尝试将同样的改进运用到更新的算法上，进一步提升小目标检测能力。

参考文献

[1] 张立杰，周舒骅，李娜. 基于改进 SSD 卷积神经网络的苹果定位与分级方法[J]. 农业机械学报，2023，54(6)：223－232.

[2] ZHU H G. An efficient lane line detection method based on computer vision[J]. Journal of Physics：Conference Series，2021，1802：032006－032014.

[3] ABBAS Q，LI Y. Cricket video events recognition using HOG，LBP and multi-class SVM[J]. Journal of Physics：Conference Series，2021，1732：012036－012043.

[4] 胡皓，郭放，刘钊. 改进 YOLOX-S 模型的施工场景目标检测[J]. 计算机科学与探索，2023，17(5)：1089－1101.

[5] WANG C Y，BOCHKOVSKIY A，LIAO H Y M. YOLOv7：trainable bag of freebies sets new state of the art for realtime object detectors[J]. arXiv：2207.02696，2022.

[6] 赵振兵，王帆帆，刘良帅，等. 基于注意力特征融合 YOLOv5 模型的无人机输电线路航拍图像金具检测方法[J]. 电测与仪表，2023，60(3)：145－152.

[7] ZHAO J，ZHANG X，YAN J A. Wheat spike detection method in UAV images based on improved YOLOv5[J]. Remote Sens，2021，13：3095.

[8] ZHU X，LYU S，WANG X，et al. TPH-YOLOv5：improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision，2021.

[9] WOO S，PARK J，LEE J Y，et al. CBAM：convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision，2018.

[10] ZHENG Z H，XIONG J T，WANG Z，et al. An efficient online citrus counting system for large-scale unstructured orchards based on the unmanned aerial vehicle[J]. Field Robot，2022，10：22147.

[11] GE Y，LIN S，ZHANG Y，et al. Tracking and counting of tomato at different growth period using an improving YOLO-deepsort network for inspection robot[J]. Machines，2022，10：489.

[12] 邹振涛，李泽平. 改进 YOLOv7 的航拍图像目标检测[J]. 计算机工程与应用，2024，60(8)：173－181.

[13] WANG J，XU C，YANG W，et al. A normalized GaussianWasserstein distance for tiny object detection[J]. arXiv：2110.13389，2021.

[14] YANG L，ZHANG R Y，LI L，et al. SimAM：a simple，parameter-free attention module for convolutional neural networks[C]//Proceedings of the International Conference on Machine Learning，2021：11863－11874.

[15] TANG S，ZHANG S，FANG Y. HIC-YOLOv5：improved YOLOv5 for small object detection[C]//2024 IEEE International Conference on Robotics and Automation (ICRA). IEEE，2024：6614－6619

[16] KEUPER M，TANG S，YU Z J，et al. A multi-cut formulation for jointsegmentation and tracking of multiple objects[J]. arXiv preprint arXiv：1607.06317，2016.