基于深度学习与目标跟踪的苹果检测与视频计数方法

高芳芳[1]，武振超[1]，索 睿[1]，周忠贤[1]，李 瑞[2]，傅隆生[1,3,4※]，张 昭[5]

（1. 西北农林科技大学机械与电子工程学院，杨凌 712100； 2. 绥德县兰花花生态食品有限责任公司，绥德 718000；
3. 农业农村部农业物联网重点实验室，杨凌 712100； 4. 陕西省农业信息感知与智能服务重点实验室，杨陵 712100；
5. 北达科他州立大学农业与生物系统工程系，美国法戈 58102）

摘 要：基于机器视觉技术自动检测苹果树上的果实并进行计数是实现果园产量测量和智慧果园生产管理的关键。该研究基于现代种植模式下的富士苹果视频，提出基于轻量级目标检测网络 YOLOv4-tiny 和卡尔曼滤波跟踪算法的苹果检测与视频计数方法。使用 YOLOv4-tiny 检测视频中的苹果，对检测到的果实采用卡尔曼滤波算法进行预测跟踪，基于欧氏距离和重叠度匹配改进匈牙利算法对跟踪目标进行最优匹配。分别对算法的检测性能、跟踪性能和计数效果进行试验，结果表明：YOLOv4-tiny 模型的平均检测精度达到 94.47%，在果园视频中的检测准确度达到 96.15%；基于改进的计数算法分别达到 69.14%和 75.60%的多目标跟踪准确度和精度，较改进前算法分别提高了 26.86 和 20.78 个百分点；改进后算法的平均计数精度达到 81.94%。该研究方法可有效帮助果农掌握园中苹果数量，为现代化苹果园的测产研究提供技术参考，为果园的智慧管理提供科学决策依据。

关键词：视频计数；YOLOv4-tiny；卡尔曼滤波器；匈牙利算法；果实匹配
doi：10.11975/j.issn.1002-6819.2021.21.025
中图分类号：TP391.41 **文献标志码**：A **文章编号**：1002-6819(2021)-21-0217-08

高芳芳，武振超，索睿，等. 基于深度学习与目标跟踪的苹果检测与视频计数方法[J]. 农业工程学报，2021，37(21)：217-224. doi：10.11975/j.issn.1002-6819.2021.21.025 http://www.tcsae.org
Gao Fangfang, Wu Zhenchao, Suo Rui, et al. Apple detection and counting using real-time video based on deep learning and object tracking[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(21): 217-224. (in Chinese with English abstract) doi：10.11975/j.issn.1002-6819.2021.21.025 http://www.tcsae.org

0 引 言

苹果等经济作物产量信息的获取是精细技术体系中的重要环节，也是果农进行果园生产管理的关键指标。苹果是中国大规模种植的农产品之一，是农业经济发展的重要支柱[1]。随着苹果种植模式的改进，矮化密植、小树冠、宽行距、窄株距的现代苹果园栽培模式已成为主流，大幅提高了果园土地利用率[2]。果园产量的预测有助于有效利用资源并合理分配劳动力、仓储空间和采收设备等资源[3]。但是，当前商业果园中果实数量主要由人工抽样计数估测，耗时费力[4]。因此，果实的自动计数对获取果园产量信息至关重要。

近年来，面向果实测产的果实自动计数方法得到了广泛研究。Dorj 等[5]利用果实的颜色特征对柑橘图像进行检测与计数研究，开发的算法与人工计数之间的决定系数 R^2 为 0.93。Mekhalfi 等[6]基于 Viola-Jones 目标检测算法检测田间环境下的猕猴桃图像并进行计数，试验在两个果园内的计数误差分别为 6%和 15%。尽管许多相关研究获得了较好的计数性能，但是这些研究都集中在静态图像[7-8]，并未对视频中的果实进行动态的计数研究。

快速有效检测苹果树上的果实是实现果实视频计数的前提。传统的目标检测算法多依赖于目标的颜色、纹理和形状等特征，并结合形态学操作进行图像分割以检测出目标区域[9-12]。但是传统的算法难以兼顾实时性和准确性，且研究结果不具有通用性。基于深度学习的目标检测算法对数据集的表达更高效和准确，泛化能力更强，更容易应用于实际场景[13]。其中 YOLO 等的一阶检测算法与 Faster R-CNN 等的二阶检测算法相比具有较快的检测速度[14]，被广泛应用于果实的检测研究[15-18]。

目标跟踪可以建立视频中相同果实的联系，是实现果实视频自动计数的另一重要环节。近年来，诸多学者提出了多种跟踪算法[19]，主要用于军事航空航天、安全监控和智能驾驶领域[20-22]。其中卡尔曼滤波算法因为可以在连续变化的系统中进行状态估计，且占用内存较小而被广泛应用[23]。Wang 等[24]基于卡尔曼滤波器对果园视频中的芒果果实进行运动跟踪，从而实现视频中果实的自动计数，其计数结果与人工计数结果比较产生了 9.9%的重复计数和 7.3%的计数错误，导致计数高出约 2.6%。刘军等[25]通过匈牙利匹配和卡尔曼滤波算法实现车辆跟踪，改善了重叠遮挡时的车辆跟踪效果。

收稿日期：2021-07-14 修订日期：2021-10-10
基金项目：国家自然科学基金（32171897）；陕西省创新人才推进计划-青年科技新星项目（2021KJXX-94）；中国博士后科学基金资助项目（2019M663832）；中国科学技术部国家外国专家局高端外国专家引进计划（G20200027075）
作者简介：高芳芳，研究方向为果实测产方法与技术。
Email: gaofangfang@nwafu.edu.cn
※通信作者：傅隆生，博士，副教授，博士生导师，研究方向为智慧农业技术与装备。Email：fulsh@nwafu.edu.cn
中国农业工程学会会员：傅隆生（E042600025M）

本文基于深度学习目标检测算法结合卡尔曼滤波器和改进匈牙利算法的目标跟踪框架实现视频中苹果的自动计数。为保证苹果检测跟踪算法的实时性与准确性，采用 YOLOv4 的简化版本，即 YOLOv4-tiny，作为目标检测网络，以期在保证检测精度的前提下实现更快的处理速度。基于卡尔曼滤波算法对目标进行预测跟踪。基于欧式距离和交并比（Intersection over Union，IoU）改进匈牙利算法对目标进行匹配，降低目标产生过度计数的概率，提高果实计数精度。

1 材料与方法

1.1 试验数据采集

试验果园位于陕西省宝鸡市扶风县法门镇（107°9′E，34°38′N），以广泛种植的红富士品种为研究对象，于2020年的收获季节9月下旬采集了苹果图像和视频数据。现代化果园通过使用铁丝将树枝进行约束以规范树木生长，其株间距在 0.3～1.5 m 之间，行间距约为 3.5～4.0 m，如图 1a 所示。研究人员通过控制搭载 RealSense D435 相机的遥控小车在树行间行走获取果园数据。为保证获取果树的完整视野，相机距离地面最低高度为 1.43 m，距离树行最少 2.07 m。共获取果园图像 800 张（如图 1b 所示），分辨率为 720×1 280，采集时段为 8：00–18：00，涵盖了顺光、逆光与侧光等可能的光照情况。采集同一果园的有效视频 10 个，保存为 MP4 格式，分辨率为 720×1 280，视频帧率为 30 帧/s。

a. 苹果种植环境 b.采集图像
a. Apple planting environment b. Obtained image

图 1 陕西省宝鸡市扶风县红富士苹果试验园

Fig.1 Experimental orchard of Fuji apple in Fufeng County, Baoji City, Shaanxi Province

1.2 数据集制作

根据实际计数需求，人工标注了 800 张图像用于训练与测试检测网络。使用 LabelImg 工具对原始数据集中苹果目标进行边界框标注。标注时将苹果的最小外接矩形框作为真实框，仅对可人工识别或能推测出果实轮廓的苹果目标进行标记，其中处于树行后排和图像边界的果实以及掉落在地和绑在树上的果实不进行标注。

采用旋转、镜像翻转、运动模糊和亮度变换对原始图像数据进行扩增。数据扩增是深度学习中常用的方法，即在不改变图像类别的情况下，增加训练数据集，让数据集尽可能的多样化，这有助于改善学习过程并提高模型的泛化能力[26]。使用自行开发的软件将 800 张原始数据扩增至 8 000 张后，并经人工检查确认扩增结果无误。通过对原始图像进行旋转和镜像翻转处理改变苹果的分布方向，对原始图像进行 90°、180°和 270°度旋转，以及水平和垂直翻转操作扩大数据集为 4 800 张；为更好的检测运动中的果实目标，研究通过运动模糊操作扩增数据集至 6 400 张；通过亮度变换增强原始数据集的亮度范围，将数据集扩增为 8 000 张。按照 4：1 随机划分为训练集和测试集[27]。

1.3 果实多目标跟踪计数方法

图 2 为本文方法的处理流程图。算法主要包含以下步骤：1）检测视频：使用 YOLOv4-tiny 网络检测视频中的苹果，得到目标检测框和相应特征；2）预测目标：使用卡尔曼滤波算法对目标在视频中下一帧的位置和状态进行预测；3）匹配目标：使用改进的匈牙利算法对视频前后两帧之间若干目标进行最优匹配，得到目标在视频中的轨迹，匹配失败的目标轨迹将被暂时保存，并继续参与后续帧预测匹配，直到该目标连续 30 帧都匹配失败后被视为消失果实，进而删除该轨迹。匹配成功则输出结果，并进行参数更新，重新开始目标检测。

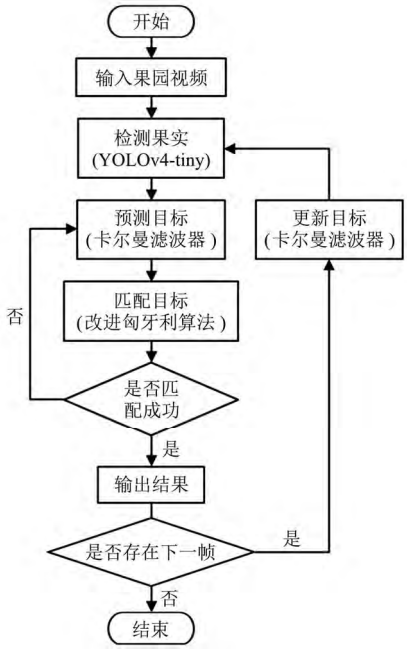


图 2 果实计数流程图

Fig.2 Flow chart of fruit counting

1.3.1 基于 YOLOv4-tiny 的果实检测

在使用基于 YOLO 深度学习网络进行目标检测方面，YOLOv4-tiny 可以快速且高精度的检测果实。YOLO 于 2016 年被首次提出，近年来不断改进，产生了 YOLOv2、YOLOv3 和 YOLOv4 等系列网络。相应的专为低端图像处理器（Graphic Processing Unit，GPU）设备而设计的轻量级网络，如 YOLOv3-tiny 和 YOLOv4-tiny，也随之被提出。其中，YOLOv4-tiny 在具有 80 个分类类别的 MS COCO 数据集中实现了 40.2%的 mAP（mean Average Precision）[28]，比 YOLOv3-tiny 高 7.1%，。因此，本研究拟选择 YOLOv4-tiny 进行果实检测。

YOLOv4-tiny 网络包括 Darknet 层和 YOLO 层两部分，Darknet 是 YOLOv4-tiny 的特征提取层，YOLO 层是

目标检测层。该网络层主要由卷积层和池化层构成，如图 3 所示，网络中每层的输出特征图尺寸表示为“分辨率宽×分辨率高×通道数/步长”。网络主要包含 1 个输入层（图像输入尺寸为 416×416），21 个带激活的卷积层（激活函数为 Leaky 和 Linear），11 个路由层，3 个最大池化层和 2 个输出层（输出层特征图尺度为 13×13，26×26）。其中 11 个路由层均分布在卷积层之后，用于在当前层引出之前卷积所得到的特征层。

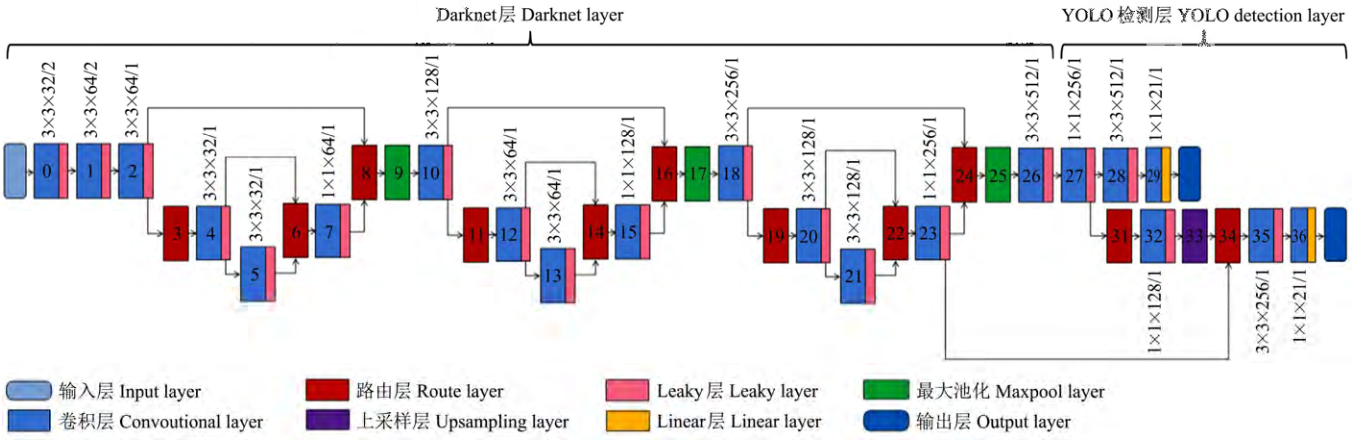


图 3　YOLOv4-tiny 网络结构

Fig.3　The structure of YOLOv4-tiny network

YOLOv4-tiny 网络与 YOLO 系列其他网络一样采用 darknet 深度学习框架实现输入图像端到端的训练。YOLO 网络首先将输入图像划分为 $N\times N$ 网格（cell）[29]。如果一个对象的中心落在某 cell 内，则该 cell 负责检测该对象。每个 cell 都预测边界框（bounding box）和这些框的置信度得分。每个 bounding box 包含 5 个预测参数：x、y、w、h 和 confidence，其中（x，y）表示目标检测框的坐标，（w，h）是目标检测框的宽度和高度。最后通过非极大值抑制选择置信度最高的 bounding box 作为最后结果。

1.3.2　基于卡尔曼滤波的果实跟踪

卡尔曼滤波器是一种广泛使用的线性系统的最优跟踪算法。由于本文获取的果园视频帧率达到 30 帧/s，视频序列之间果实目标的位置变化很小,可基本视为匀速运动[28]，因此可假设果实跟踪系统随时间变化是线性相关的。算法采用匀速和线性观测模型预测和更新目标轨迹[30]。在预测阶段，卡尔曼滤波器使用视频上一帧的估计，做出对当前帧的预测，即根据 YOLOv4-tiny 网络的检测位置预测对应的果实跟踪位置和其协方差矩阵。

$$\hat{\boldsymbol{x}}_{k|k\text{-}1}=\boldsymbol{A}\hat{\boldsymbol{x}}_{k\text{-}1|k\text{-}1}+\boldsymbol{B}\boldsymbol{u}_{k\text{-}1} \tag{1}$$

$$\boldsymbol{P}_{k|k\text{-}1}=\boldsymbol{A}\boldsymbol{P}_{k\text{-}1|k\text{-}1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{Q} \tag{2}$$

式中 $\hat{\boldsymbol{x}}_{k|k\text{-}1}$ 为第 k 帧的先验状态估计值，是滤波的中间计算结果，即根据上一帧（k–1 帧）的最优估计预测的 k 帧的结果，是预测方程的结果；$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为系统的参数矩阵；$\hat{\boldsymbol{x}}_{k\text{-}1|k\text{-}1}$ 为 k–1 帧的后验状态估计值，是滤波的结果之一；$\boldsymbol{u}_{k\text{-}1}$ 为 k–1 帧的过程噪声；$\boldsymbol{P}_{k|k-1}$ 为第 k 帧的先验估计协方差，是滤波的中间计算结果；$\boldsymbol{P}_{k-1|k-1}$ 为 k–1 帧的后验估计协方差，表示状态的不确定度，是滤波的结果之一；Q 为系统过程噪声的协方差，是卡尔曼滤波器用于估计离散时间过程的状态变量。

在更新阶段，滤波器利用对当前帧的观测值优化在预测阶段获得的预测值，以获得一个更精确的新估计值，即根据检测果实和预测果实之间的匹配关系更新果实跟踪位置及其协方差矩阵。

$$\boldsymbol{K}_k=\boldsymbol{P}_{k|k-1}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{P}_{k|k-1}\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{R})^{-1} \tag{3}$$

$$\hat{\boldsymbol{x}}_{k|k}=\hat{\boldsymbol{x}}_{k|k\text{-}1}+\boldsymbol{K}_k(\boldsymbol{Z}_k-\boldsymbol{H}\hat{\boldsymbol{x}}_{k|k\text{-}1}) \tag{4}$$

$$\boldsymbol{P}_{k|k}=(\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H})\boldsymbol{P}_{k|k\text{-}1} \tag{5}$$

式中 $\boldsymbol{K}_k$ 为滤波增益矩阵，是滤波的中间计算结果；$\boldsymbol{H}$ 是状态变量到观测变量的转换矩阵，表示将状态和观测连接起来的关系；$\boldsymbol{R}$ 为测量噪声协方差；$\hat{\boldsymbol{x}}_{k|k}$ 为第 k 帧的后验状态估计值，是滤波的结果之一；$\boldsymbol{Z}_k$ 为测量值，是滤波的输入；$\boldsymbol{P}_{k|k}$ 为第 k 帧的后验估计协方差，是滤波的结果之一；$\boldsymbol{I}$ 为单位矩阵。

1.3.3　基于改进匹配算法的果实匹配

匈牙利算法是一种在多项式时间内求解任务分配问题的组合优化算法，在本研究中被用来建立检测果实和预测果实之间的关系，即匹配果实。分别通过 2.1 与 2.2 小节算法得到果园视频中果实的检测结果与预测结果，并采用欧氏距离衡量卡尔曼预测结果和 YOLOv4-tiny 检测结果之间的相似性，如公式（6）所示。

$$D_{i,j}=\sqrt{(x_i-x_j)^2+(y_i-y_j)^2} \tag{6}$$

式中(x_i, y_i)和(x_j, y_j)分别表示目标跟踪框坐标和检测框坐标。

在果实匹配阶段，使用匈牙利算法寻求预测结果与检测结果的匹配最优解，如公式（7）和（8）所示。匹配成功则进入卡尔曼滤波的更新阶段。

$$\min Z=\sum_{i=1}^{m}\sum_{j=1}^{n}D_{i,j}x_{i,j} \tag{7}$$

$$\text{s.t.}\begin{cases}\sum_{i=1}^{m}x_{i,j}=1,\ i=1,2,\ldots,m\\ \sum_{j=1}^{n}x_{i,j}=1,\ j=1,2,\ldots,n\\ x_{i,j}=0\text{或}1,\ i(j)=1,2,\ldots,m(n)\end{cases} \tag{8}$$

式中 m 和 n 分别是跟踪到的果实数目以及检测到的果实数目。

匹配失败主要分为跟踪目标匹配失败以及检测目标匹配失败。跟踪目标匹配失败是由 YOLOv4-tiny 网络的漏检或视频中果实的消失导致的，两种原因下视频帧中都不存在该果实的检测框，所以匹配失败。检测框中的目标匹配失败产生的原因也有两个可能，一是该果实是视频中新出现的，二是该果实长时间被遮挡。新出现的果实尚未存在轨迹，而长时间被遮挡但尚未判定为消失状态的果实尽管一直在被跟踪，但是由于跟踪框中并不存在该果实，使得卡尔曼滤波器预测结果的不确定性逐步增加，导致该果实的轨迹与检测结果不能成功匹配。

针对上述匹配失败问题，对匹配失败的跟踪目标与匹配失败的检测目标进行 IoU 匹配，如公式（9）所示，同时试验确定最大阈值以去除相关性较低的检测框与跟踪框之间的匹配，若匹配成功则进入卡尔曼滤波的更新阶段。

$$\mathrm{IoU}=\frac{|D\cap T|}{|D\cup T|} \tag{9}$$

式中 D 和 T 分别表示检测框和跟踪框区域。

再次匹配失败的跟踪目标的轨迹将被暂时保存，并继续进行预测，直到该轨迹连续 30 帧都匹配失败才被视为消失果实，进而删除该轨迹。再次匹配失败的检测框中的目标将被视为新果实，进行跟踪。最后按照果实在视频帧中的出现顺序赋予果实数字 ID（从 1 开始）实现果实的计数。

1.4 试验设计

本文试验分为以下 3 部分：

1）果实检测：首先应用 5 120 张果园图像基于 Darknet 框架分别训练出 YOLOv3-tiny 和 YOLOv4-tiny 网络模型，平台为配备有 Intel Core i5-6400（2.70 GHz）的 CPU，16 GB 的 RAM 和 NVIDIA GTX 1080 8 GB 的 GPU 的台式计算机。软件工具包括 CUDA 9.0，cuDNN 7.1.3，Microsoft Visual Studio 2015，CMake-3.16，Python 3.6 和 OpenCV 3.1.0。根据网格搜索算法设置网络中的最优超参数为初始学习率为 0.001，权重衰减系数为 0.000 5，训练策略采用动量项为 0.9 的动量梯度下降算法，迭代次数设置为 50 000。随后使用训练得到的网络模型分别测试果园图像，比较检测精度。

2）视频跟踪：使用卡尔曼滤波算法结合未改进和改进的匈牙利匹配算法对果园视频分别进行跟踪，对比效果，测试果实跟踪精度。测试平台为配备有 Intel Core i7-8565U（1.80 GHz）CPU，8 GB RAM 和 NVIDIA GeForce MX250 2 GB GPU 的笔记本电脑。软件工具包括 CUDA 10.0，cuDNN 7.6.5，Microsoft Visual Studio 2017，Python 3.8 和 OpenCV 4.4.0。

3）果实计数：首先由 3 名研究人员分别对获取的 10 个果园视频进行计数，通过对 3 名研究人员所得计数结果进行均值处理，得到视频中果实的实际数目。计数时，研究人员将视频逐帧播放，首先记录第一帧中果实数目，然后在后续帧中记录新出现的果实数目，直至视频结束，得到视频中总的果实数目，实现视频中果实的计数。随后基于算法分别获取视频中的果实数目，并将结果与人工计数结果进行比较，测试果实计数精度。

1.5 试验评价指标

1.5.1 果实检测试验

本文使用准确率（Precision，P）、召回率（Recall，R）、准确率和召回率的调和平均数（$F1$ score，$F1$）、平均检测精度值（Average Detection Precision，ADP）和检测一张图像所用时间（Detection time）来评价训练得到的网络模型性能，计算公式如下所示。此外，使用视频平均检测准确度（Video Detection Accuracy，VDA）表示模型正确检测到的目标占所有目标的比值，数值越高表示模型性能越好。

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{10}$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{11}$$

$$F1=\frac{2\mathrm{PR}}{\mathrm{P}+\mathrm{R}} \tag{12}$$

$$\mathrm{ADP}=\int_0^1 P(R)\mathrm{d}R \tag{13}$$

式中 TP（True Positive）为正确检测出来的目标果实；FP（False Positive）为误检出来的目标果实；FN（False Negative）为漏检的目标果实。

1.5.2 果实跟踪试验

本文使用果实 ID 切换率（ID Switch Rate，IDSR）、多目标跟踪准确度（Multiple Object Tracking Accuracy，MOTA）和多目标跟踪精度（Multiple Object Tracking Precision，MOTP）来评价跟踪算法性能。果实 ID 切换率指视频中 ID 发生变换的果实占所有计数果实的比值，值越小越好。多目标跟踪准确度表示跟踪算法在保持跟踪轨迹时的性能，值越大越好，如公式（14）所示。多目标跟踪精度指所有跟踪目标匹配成功率，精度越高表示算法性能越好，如公式（15）所示。

$$\mathrm{MOTA}=M/S \tag{14}$$

$$\mathrm{MOTP}=M/T \tag{15}$$

式中 S 表示视频中跟踪到的果实总数目，即算法计数数目；M 表示视频中跟踪匹配正确的果实个数；T 表示视频中跟踪匹配的果实总数目，即检测数目。

1.5.3 果实计数试验

本文使用 ACP 值（Average Counting Precision）来评价算法的计数精度，计算公式如式（16）所示。

$$\mathrm{ACP}=\frac{\sum_1^n\left(1-\dfrac{|S-G|}{G}\right)}{n_{\mathrm{v}}} \tag{16}$$

式中 G 表示人工计数数目；n_{v} 是视频个数。

2 试验结果与分析

2.1 果实检测试验结果

分别使用训练得到的 YOLOv3-tiny 和 YOLOv4-tiny 网络在相同的测试数据集下对比试验。测试集中共包含 1 600 张图像，78 776 个标注样本。YOLOv4-tiny 和

YOLOv3-tiny 网络分别检测到 73 266 和 71 501 个 TP 样本，11 346 和 12 291 个 FP 样本，5 510 和 7 275 个 FN 样本。基于式（10）～（13）计算得到相应指标如表 1 所示，YOLOv4-tiny 比 YOLOv3-tiny 网络检测精度高 1.76 个百分点，且检测用时少 0.009 s/帧，证明研究所选用 YOLOv4-tiny 网络的优越性。此外，该网络 94.47%的 ADP 可为卡尔曼滤波器提供可靠的果实检测位置信息，为果实计数的实现奠定基础。

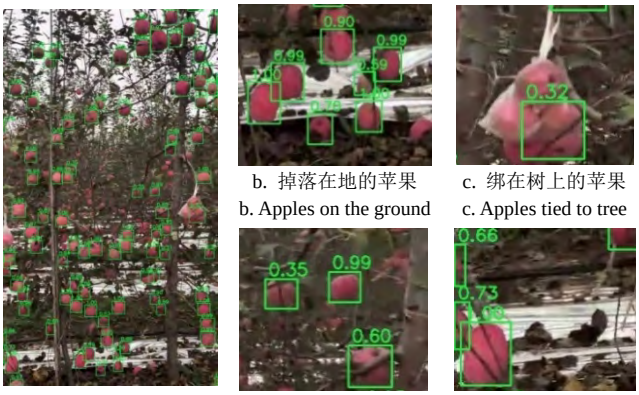
表 1　不同网络的检测结果对比

Table 1　Comparison of detection results of different networks

方法 Method	准确率 Precision	召回率 Recall	*F*1 分数 *F*1score	检测精度 Detection precision/%	检测时间 Detection time/(s·帧$^{-1}$)
YOLOv3-tiny	0.85	0.91	0.88	92.71	0.027
YOLOv4-tiny	0.87	0.93	0.90	94.47	0.018

YOLOv4-tiny 网络的检测效果如图 4 所示，部分果实被漏检或错误检测。错误检测主要是检测出了一些不需要被计数的果实样本，主要包括掉落在地的果实（图 4b）、绑在树上的果实（图 4c）、后排果树上的果实（图 4d）和图像边界处遮挡较为严重的果实（图 4e）。错误检测将会导致算法对一些非目标果实进行计数，从而增大果实计数数目。为降低错误检测带来的计数误差，算法在检测视频中的果实时，将会提高置信度阈值，以避免对检测精度较低的果实执行跟踪匹配过程。漏检果实多为被果实或树叶遮挡严重的果实，这是因为难以人工识别或推测出果实轮廓的苹果目标默认不做标记。漏检会导致果实的计数数目小于实际数目。但在本文中，研究对象为视频中的果实，它会因为连续出现在多个视频帧中而增加被检测到的概率。因此，视频中极少出现漏检果实。

使用训练得到的 YOLOv4-tiny 模型对同一果园的 10 段果园视频分别进行检测，过程中提高置信度阈值为 0.8，其 VDA 达到 96.15%，证明训练出的模型具备高精度检测果园视频的能力。视频中很少出现果实漏检的情况，因为多数果实都因连续出现在视频帧中而被成功检测。但是依然存在少量非目标果实被错误检测。未来研究将通过比较果实的检测框特征进一步去除非目标果实的错误检测，例如通过设置检测框长宽比阈值去除边界处非目标检测框、设置检测框面积阈值去除果树后排的非目标检测框等措施来改善果实计数效果。



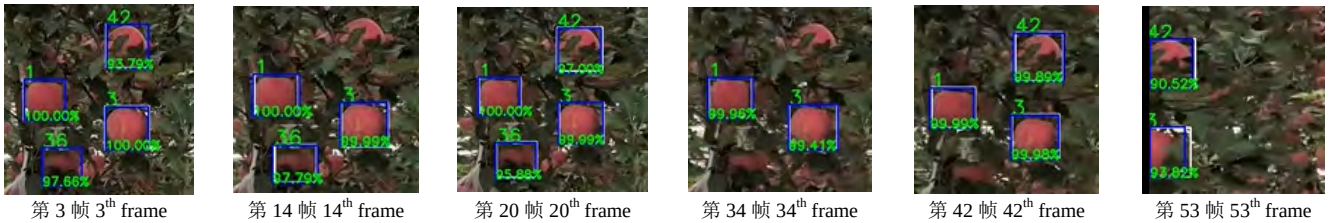
a. 果园图像检测示例 a. Detection example of orchard image　b. 掉落在地的苹果 b. Apples on the ground　c. 绑在树上的苹果 c. Apples tied to tree　d. 后排果树上的苹果 d. Apples on the back row of trees　e. 图像边界处的苹果 e. Apples at the boundary of image

图 4 基于 YOLOv4-tiny 的苹果检测图

Fig.4　Image detection of apples based on YOLOv4-tiny

2.2　果实跟踪试验结果

基于开发的算法对果园视频进行跟踪试验并获得了良好的跟踪效果。图 5 是基于本文算法对果园视频序列跟踪结果的示例，图 5a 中 1 号以及 3 号果实在视频连续 50 帧中都一直被检测并跟踪，36 号果实一直处于遮挡状态下直到该果实的置信度低于 0.8 不再执行跟踪匹配过程，可以看出，本文的算法可以有效跟踪果实。图 5a 中 42 号果实因为被遮挡导致该果实间断性的被检测，但是该果实的匹配过程并未出现差错。图 5b 中 99 号果实在运动过程中接连被两片树叶遮挡导致在视频的 61 帧到 100 帧之间出现间断性检测情况，第 3 片树叶的出现更是导致该果实在连续 20 多帧中都未能成功检测，但是该果实的匹配过程也并未出现差错。因此可以看出，尽管在视频运动过程中存在果实被间断性检测甚至长时间检测不到，但是算法依然可以跟踪到该目标，并保持原有 ID 不变。



第 3 帧 3th frame　第 14 帧 14th frame　第 20 帧 20th frame　第 34 帧 34th frame　第 42 帧 42th frame　第 53 帧 53th frame

a. 间断性检测的果实的跟踪示例

a. Tracking example of intermittently detected fruits

第 61 帧 61th frame　第 93 帧 93th frame　第 96 帧 96th frame　第 100 帧 100th frame　第 107 帧 107th frame　第 111 帧 111th frame　第 132 帧 132th frame

b. 长时间检测不到的果实的跟踪示例

b. Tracking example of fruits that have not been detected for a long time

图 5　不同检测情况下果实跟踪结果示例

Fig.5　Examples of fruit tracking results under different detection conditions

分别使用改进前与改进后计数算法对 10 个果园视频（220～524 帧）进行跟踪计数，所得跟踪指标如表 2 所示。改进后算法的 MOTA 为 69.14%，较改进前算法提高了 26.86 个百分点，MOTP 提高了 20.78 个百分点，跟踪可信度得到大幅提高。果园环境的复杂以及视频拍摄的精度导致果实 ID 极易发生跳变，甚至存在部分果实的 ID 会多次发生跳变，这也是产生跟踪误差的主要原因。匹配算法的改进使得视频IDSR由25.91%降低到了11.92%，在一定程度上降低了果实跟踪误差。

表 2 算法改进前后跟踪结果的比较

Table 2 Comparison of tracking results before and after algorithm improvement %

提出的算法 Proposed algorithm	视频检测准确度 Video detection accuracy	ID 变换率 ID switch rate	多目标跟踪准确度 Multiple object tracking accuracy	多目标跟踪精度 Multiple object tracking precision
改进前 Before improvement	96.15	25.91	42.28	54.82
改进后 After improvement	96.15	11.92	69.14	75.60

2.3 果实计数试验结果

由 3 名研究人员对采集到的原始视频进行计数获得果实数量真实值。研究人员秉持客观原则，对视频中所有可见果实进行计数。对 3 名研究人员所得计数结果进行均值处理得到相应视频中果实的实际数目。

视频中人工计数数值与算法计数数值的拟合结果如图 6 所示。基于算法获取的果实数目与人工计数获得的果实数目之间的 RMSE 为 33.711。基于公式（16）计算 10 个果园视频的计数精度均在 80%左右，平均精度为 81.94%。从拟合结果看，决定系数 R^2 值为 0.986，表明本文算法对现代果园视频中的果实计数值与真实值具有显著的线性相关性。由图 6 的数据可知，视频计数精度不会随着视频中果实数量的增多而降低，表明研究提出的方法具有较好的稳定性。

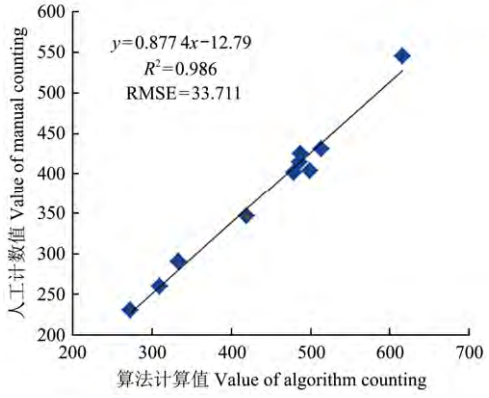


图 6 视频中人工计数值与算法计数值的拟合结果

Fig.6 Fitting results of manual counting and algorithm counting in video

3 结 论

本文分别基于 YOLOv4-tiny 网络、卡尔曼滤波器和改进的匈牙利匹配算法对自然环境下的苹果进行检测、跟踪和匹配，提出了一种基于现代化苹果园视频的果实计数方法。基于所获取的果园图像训练 YOLOv4-tiny 网络模型，其平均检测精度值（Average Detection Precision，ADP）和视频平均检测准确度（Video Detection Accuracy，VDA）值分别达到 94.47%和 96.15%，检测一张图像仅用时 0.018 s。通过该模型对视频中的果实进行逐帧检测，得到果实的像素坐标。基于卡尔曼滤波器对果实位置进行预测，并先后基于欧式距离和交并比（Intersection over Union，IoU）衡量预测果实和检测果实之间的相似性，通过匈牙利算法匹配连续视频帧中的相同果实，多目标跟踪准确度（Multiple Object Tracking Accuracy，MOTA）和多目标跟踪精度（Multiple Object Tracking Precision，MOTP）分别为 69.14%和 75.60%。最后，基于视频中果实出现顺序赋予果实数字 ID 实现果实的计数，且 ACP 值（Average Counting Precision）达到 81.94%，在获得的 10 个果园视频中基于算法得到的果实数目与人工计数得到的果实数目的决定系数为 0.986，RMSE 为 33.711，表明了算法的有效性，研究结果可以为实际应用过程中自动估计现代化苹果园以及其他现代化果园的果实数目提供参考。

［参 考 文 献］

[1] Gao F, Fu L, Zhang X, et al. Multi-class fruit-on-plant detection for apple in SNAP system using Faster R- CNN[J]. Computers and Electronics in Agriculture, 2020, 176: 105634.

[2] Fu L, Majeed Y, Zhang X, et al. Faster R-CNN-based apple detection in dense-foliage fruiting-wall trees using RGB and depth features for robotic harvesting[J]. Biosystems Engineering, 2020, 197: 245-256.

[3] Rahnemoonfar M, Sheppard C. Deep count: Fruit counting based on deep simulated learning[J]. Sensors, 2017, 17(4): 0905.

[4] Liu X, Chen S W, Liu C, et al. Monocular camera based fruit counting and mapping with semantic data association[J]. IEEE Robotics and Automation Letters, 2019, 4(3): 2296-2303.

[5] Dorj U O, Lee M, Yun S. An yield estimation in citrus orchards via fruit detection and counting using image processing[J]. Computers and Electronics in Agriculture, 2017, 140: 103-112.

[6] Mekhalfi M L, Nicolò C, Ianniello I, et al. Vision system for automatic on-tree kiwifruit counting and yield estimation[J]. Sensors, 2015, 20(15): 4214.

[7] Bargoti S, Underwood J P. Image segmentation for fruit detection and yield estimation in apple orchards[J]. Journal of Field Robotics, 2017, 34(6): 1039-1060.

[8] Koirala A, Walsh K B, Wang Z, et al. Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of ‘MangoYOLO’[J]. Precision Agriculture, 2019, 20(6): 1107-1135.

[9] Lv J, Ni H, Wang Q, et al. A segmentation method of red apple image[J]. Scientia Horticulturae, 2019, 256: 108615.

[10] 傅隆生，孙世鹏，Manuel V，等. 基于果萼图像的猕猴桃果实夜间识别方法[J]. 农业工程学报，2017，33(2)：199-204.
Fu Longsheng, Sun Shipeng, Manuel V, et al. Kiwifruit recognition method at night based on fruit calyx image[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(2): 199-204. (in Chinese with English abstract)

[11] 马翠花，张学平，李育涛，等. 基于显著性检测与改进 Hough 变换方法识别未成熟番茄[J]. 农业工程学报，2016，32(14)：219-226.
Ma Cuihua, Zhang Xueping, Li Yutao, et al. Identification of immature tomatoes base on salient region detection and improved Hough transform method[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2016, 32(14): 219-226. (in Chinese with English abstract)

[12] Liu X, Zhao D, Jia W, et al. A detection method for apple fruits based on color and shape features[J]. IEEE Access, 2019, 7: 67923-67933.

[13] 薛月菊，黄宁，涂淑琴，等. 未成熟芒果德改进 YOLOv2 识别方法[J]. 农业工程学报，2018，34(7)：173-179.
Xue Yueju, Huang Ning, Xu Shuqin, et al. Immature mango detection based on improved YOLOv2[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(7): 173-179. (in Chinese with English abstract)

[14] Mazzia V, Khaliq A, Salvetti F, et al. Real-time apple detection system using embedded systems with hardware accelerators: An edge AI application[J]. IEEE Access, 2020, 8: 9102-9114.

[15] 刘芳，刘玉坤，林森，等. 基于改进型 YOLO 的复杂环境下番茄果实快速识别方法[J]. 农业机械学报，2020，51(6)：229－237.
Liu Fang, Liu Yukun, Lin Sen, et al. Fast recognition method for tomatoes under complex environments based on improved YOLO[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(6): 229-237. (in Chinese with English abstract)

[16] 刘天真，滕桂法，苑迎春，等. 基于改进 YOLOv3 的自然场景下冬枣果实识别研究[J]. 农业机械学报，2021，52(5)：17-25.
Liu Tianzhen, Teng Guifa, Yuan Yingchun, et al. Winter jujube fruit recognition based on improved YOLOv3 under natural scene[J]. 2021, 52(5):17-25. (in Chinese with English abstract)

[17] 吕石磊，卢思华，李震，等. 基于改进 YOLOv3-LITE 轻量级神经网络的柑橘识别方法[J]. 农业工程学报，2019，35(17)：205-214.
Lv Shilei, Lu Sihua, Li Zhen, et al. Orange recognition method using improved YOLOv3-LITE lightweight neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(17): 205-214. (in Chinese with English abstract)

[18] Kuznetsova A, Maleva T, Soloviev V. Using YOLOv3 algorithm with pre- And post-processing for apple detection in fruit-harvesting robot[J]. Agronomy, 2020, 10(7): 1016.

[19] Lukežič A, Vojíř T, Čehovin Z L, et al. Discriminative correlation filter tracker with channel and spatial reliability[J]. International Journal of Computer Vision, 2018, 126(7): 671-688.

[20] Lee H, Cho A, Lee S, et al. Vision-based measurement of heart rate from ballistocardiographic head movements using unsupervised clustering[J]. Sensors, 2019, 19: 3263.

[21] Ngo T N, Wu K, Yang E, et al. A real-time imaging system for multiple honey bee tracking and activity monitoring[J]. Computers and Electronics in Agriculture, 2019, 163(1): 104841.

[22] Wawrzyniak N, Hyla T, Popik A. Vessel detection and tracking method based on video surveillance[J]. Sensors, 2019, 19(23): 5230.

[23] Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric[C]. Taipei: International Conference on Image Processing, 2018: 3645-3649.

[24] Wang Z, Walsh K, Koirala A. Mango fruit load estimation using a video based MangoYOLO — Kalman filter — hungarian algorithm method[J]. Sensors, 2019, 19(12): 2742.

[25] 刘军，后士浩，张凯，等. 基于增强 Tiny YOLOV3 算法的车辆实时检测与跟踪[J]. 农业工程学报，2019，35(8)：118-125.
Liu Jun, Hou Shihao, Zhang kai, et al. Real-time vehicle detection and tracking based on enhanced Tiny YOLOV3 algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(8): 118-125. (in Chinese with English abstract)

[26] Song Z, Zhou Z, Wang W, et al. Canopy segmentation and wire reconstruction for kiwifruit robotic harvesting[J]. Computers and Electronics in Agriculture, 2021, 181: 105933.

[27] 傅隆生，冯亚利，Elkamil Tola，等. 基于卷积神经网络的田间多簇猕猴桃图像识别方法[J]. 农业工程学报，2018，34(2)：205-211.
Fu Longsheng, Feng Yali, Elkamil Tola, et al. Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(2): 205-211. (in Chinese with English abstract)

[28] Han B, Lee J, Lim K, et al. Design of a scalable and fast yolo for edge-computing devices[J]. Sensors, 2020, 20(23): 6779.

[29] Fu L, Feng Y, Wu J, et al. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model[J]. Precision Agriculture, 2021, 22(3): 754-776.

[30] 乔虹，冯全，张芮，等. 基于时序图像跟踪的葡萄叶片病害动态监测[J]. 农业工程学报，2018，34(17)：167-175.
Qiao Hong, Feng Quan, Zhang Rui, et al. Dynamic monitoring of grape leaf disease based on sequential images tracking[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(17): 167-175. (in Chinese with English abstract)

Apple detection and counting using real-time video based on deep learning and object tracking

Gao Fangfang[1], Wu Zhenchao[1], Suo Rui[1], Zhou Zhongxian[1], Li Rui[2], Fu Longsheng[1,3,4※], Zhang Zhao[5]

(1. *College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling* 712100, *China*;
2. *Suide County Lanhuahua Ecological Food Co., Ltd., Suide* 718000, *China*;
3. *Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling* 712100, *China*;
4. *Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling* 712100, *China*;
5. *Department of Agricultural and Biosystems Engineering, North Dakota State University, Fargo* 58102, *U.S.A.*)

Abstract: Yield estimation for apples is a key for predicting stock volumes, allocating needed labor, and planning harvesting operations. Manual visual yield estimation for a small number of trees to predict the number of fruits in an orchard has traditionally been employed, resulting in inaccurate and misleading information. Therefore, an automated solution for orchard yield measurement is urgently needed. Detection and counting of fruits infield based on machine vision coupled with advanced machine learning algorithms is a key to realizing orchard yield measurement automatically, which can provide baseline information for better production management. Therefore, this study aims to develop an automated video processing method to realize the automated detection and counting of apple fruits in an orchard environment with a modern vertical fruiting-wall architecture. This study proposed a fruit counting method based on a lightweight YOLOv4-tiny network and Kalman filter algorithm toward this end. 'Fuji' variety was selected, which is widely planted in modern planting patterns. 800 images and 10 videos of apple trees were acquired using a remote-controlled car equipped with a Realsense D435 camera. Firstly, fruits in the orchard video were detected using the trained YOLOv4-tiny model. Secondly, all detected apples would be predicted based on the Kalman filter algorithm. Subsequently, all predicted and detected apples in the subsequent frame would be optimally matched based on the Hungarian algorithm of the Euclidean distance and Intersection over Union (IoU). Successfully matched fruits would be added to the tracked track, based on which the corresponding Kalman filter was updated. The trajectory that failed to match would be temporarily saved until the match failed for 30 consecutive frames, while the detection target that failed to match was regarded as a new fruit. Finally, the fruit digital ID would be assigned based on the appearance sequence of the fruit in the video frame to realize the fruit count. In order to prove the superior performance of the deep learning network trained in this study, YOLOv3-tiny was chosen to use the same test dataset for comparison with YOLOv4-tiny. The test results showed that the Average Detection Precision (ADP) based on YOLOv4-tiny reached 94.47%, which was 1.76 percentage points higher than that of YOLOv3-tiny. Besides, YOLOv4-tiny only took 0.018 s on average to detect fruits in one image with the resolution of 720×1 080 pixels, which was 0.009 s faster than YOLOv3-tiny. It could be seen that YOLOv4-tiny could achieve high-precision detection of fruits at a faster speed, which provided a good foundation for fruit counting. The Multiple Object Tracking Accuracy (MOTA) and the Multiple Object Tracking Precision (MOTP) based on Kalman and improvement Hungarian algorithms were 69.14% and 75.60%, which were 26.86 percentage points and 20.78 percentage points higher than the indicators based on Kalman and the unimproved Hungarian algorithm, respectively, indicating the reliability of the tracking algorithm. Furthermore, an average precision of 81.94% and a determination coefficient of 0.986 with counting performed by manual observation were reached in 10 orchard videos. The method developed in this study can effectively feedback the detection and counting results of apples in the orchard for growers, provide technical reference for the production measurement research of modern apple orchards, and provide scientific decision-making basis for intelligent management of orchards.

Keywords: video counting; YOLOv4-tiny; Kalman filter; Hungarian algorithm; fruit matching