doi:10.6041/j.issn.1000-1298.2025.01.008

基于视觉触觉双重迁移学习的番茄成熟度检测方法

张　鹏[1]　杜东峰[1]　李　爽[1]　单东日[2]　陈振学[3]

(1.齐鲁工业大学(山东省科学院)信息与自动化学院,济南 250353;
2.齐鲁工业大学(山东省科学院)机械工程学院,济南 250353;
3.山东大学控制科学与工程学院,济南 250061)

摘要:针对当前自动化采摘过程中仅依赖视觉技术无法准确识别番茄成熟度的问题,提出了一种基于视觉触觉双重迁移学习的番茄成熟度检测方法。该方法首先采用视觉触觉双重迁移学习融合算法作为特征提取融合模块,解决无法有效提取番茄特征信息的问题。其次,将软参数共享-多标签分类方法作为分类模块,通过增加不同分类任务之间的关联性,避免出现过拟合的现象。本文主要针对成熟后为红、黄果等单一颜色的番茄品种,并在新开发的视觉触觉数据集进行实验研究。实验表明,软参数共享-多标签检测模型参数量为 1.882×10^7,成熟度 AUC 分值达到 0.977 3,对比不确定性加权损失、自适应硬参数共享、十字绣网络和软参数共享等检测模型,参数量分别下降 3.08×10^6、6.16×10^6、3.08×10^6和 3.08×10^6,成熟度 AUC 分值分别提高 0.017 5、0.017 9、0.026 7 和 0.008 9。这表明该方法在一定程度上提高了自动化采摘过程中对番茄成熟度的检测能力,为番茄成熟度检测问题提供了一种有效的解决方法。

关键词:番茄成熟度;机器视觉;机器触觉;双重迁移学习;软参数共享-多标签

中图分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1000-1298(2025)01-0074-10　　**OSID**:

Tomato Maturity Detection Method Based on Visual and Haptic Double Transfer Learning

ZHANG Peng[1]　DU Dongfeng[1]　LI Shuang[1]　SHAN Dongri[2]　CHEN Zhenxue[3]

(1. *College of Information and Automation Engineering*, *Qilu Technology University* (*Shandong Academy of Sciences*), *Jinan* 250353, *China*
2. *College of Mechanical Engineering*, *Qilu Technology University* (*Shandong Academy of Sciences*), *Jinan* 250353, *China*
3. *College of Control Science and Engineering*, *Shandong University*, *Jinan* 250061, *China*)

Abstract: Aiming at the problem that tomato ripeness cannot be accurately recognized by relying only on visual technology in the current automated picking process, a tomato ripeness detection method based on visual-haptic dual migration learning was proposed. The method firstly adopted the visual and haptic double transfer learning fusion algorithm as the feature extraction fusion module to solve the problem of not effectively extracting tomato feature information. Secondly, the soft parameter sharing - multilabel classification method was used as the classification module to avoid overfitting by increasing the correlation between different classification tasks. Focusing on tomato varieties that ripened to a single color, such as red and yellow fruits, experimental studies on a newly developed visual and tactile dataset were conducted. The experiments showed that the parameter count of the soft parameter sharing - multilabel detection model was 1.882×10^7, and the ripeness AUC score reached 0.977 3. Compared with the detection models of uncertainty weighted loss, adaptive hard parameter sharing, cross-stitch network, and soft parameter sharing, the parameter counts dropped by 3.08×10^6, 6.16×10^6, 3.08×10^6, and 3.08×10^6, and the ripeness AUC scores increased by 0.017 5, 0.017 9, 0.026 7 and 0.008 9, respectively. This indicated that the method improved the detection of tomato ripeness during automated picking to a certain extent and provided an effective solution to the tomato ripeness detection problem.

Key words: tomato maturity; machine vision; machine haptics; double transfer learning; soft parameter sharing - multilabel

收稿日期:2024-03-25　修回日期:2024-05-15

基金项目:山东省重大科技创新工程项目(2019JZZY010444)和齐鲁工业大学(山东省科学院)科教产融合培优基金项目(2023PY006)

作者简介:张鹏(1977—),男,副教授,博士,主要从事智能检测与控制技术和信息融合等研究,E-mail:zp@qlu.edu.cn

0　引言

目前，自动化采摘过程中的番茄成熟度检测方法主要有传统的方法和基于深度学习的方法。传统方法主要根据人为观察图像中果实与背景之间的差异，通过设计算法提取图像特征从而实现果实的识别[1]。但当果实的光照条件、背景环境等发生变化时，果实识别准确性会受到较大的影响[2]。

深度学习的方法通过构建模型，利用数据集对模型进行训练，自动学习有用的特征，从而实现对番茄成熟度识别。其中，杨信廷等[3]提出了一种基于 Swin Transformer 与 GRU 的低温贮藏番茄成熟度识别方法，番茄成熟度正确识别率为 95.783%。BEGUM 等[4]提出了一种基于迁移学习的番茄成熟度检测方法，实现了 3 种迁移学习方法，即 VGG、Inception 和 ResNet，实验结果表明，VGG19 在迭代 50 轮和批次大小 32 时表现最佳。PANGILINAN 等[5]分别比较了基于 InceptionV3、ResNet50 和 VGG19 的番茄成熟度检测方法，评估了 3 种卷积神经网络算法在识别番茄成熟度的准确性，其中 VGG19 算法获得了 95.00% 的最高准确率。RAFIQ 等[6]应用人工神经网络对番茄进行分类，准确率达到 96.00%。与传统方法相比，基于深度学习的方法避免了人工特征提取等复杂步骤，并在某种程度上有效地提升了果实识别分割的准确性，鲁棒性更强，但该方法依赖于数据集拥有大量标注准确的训练图像，且对硬件设备算力具有较高要求。

在自动化采摘过程中使用视觉技术仅从外观识别成熟度，对番茄内部硬度特征缺乏检测手段，而硬度特征也体现了番茄的不同成熟度，同时对后期的加工、运输和贮藏方式以及番茄的价格及风味会造成极大的影响[7]，所以使用一些检测番茄硬度的技术，辅助检测成熟度。例如光谱差分技术和空间分辨光谱技术[8]，以及 Sinclair iQ 硬度计[9]和触觉传感器[10]等。其中，HE 等[11]使用 GelSight 触觉传感器有效测量樱桃番茄的硬度，并准确预测其短期的硬度变化。王俊平等[12]利用视觉传感器与电子鼻融合的方法检测番茄，分别提取颜色与气味信息，准确率为 93.31%。然而，在开放环境的情况下，无法有效收集番茄的气体，因此该方法在实际应用中仍面临挑战。

为了提高自动化采摘过程中检测番茄成熟度的能力，进行更准确的判断，本文以成熟后为红、黄果等单一颜色的番茄品种为研究对象，采用机器视觉技术和机器触觉技术分别获取番茄颜色和硬度特征，结合迁移学习策略建立一种基于视觉触觉双重迁移学习的番茄成熟度检测方法。

1　软参数共享-多标签检测模型

为了实现小样本情况下对番茄成熟度检测，以视觉触觉双重迁移学习融合算法作为特征提取融合模块，软参数共享-多标签分类方法作为分类模块构建检测模型，结构如图 1 所示。

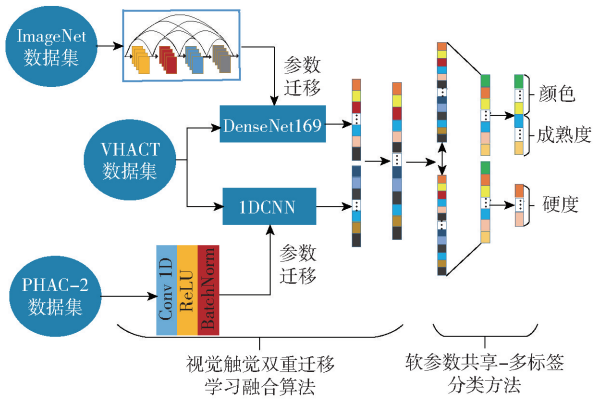

图 1　总体结构图

Fig. 1　Overall structure diagram

源域训练：分别将 ImageNet 数据集和 PHAC－2（Penn Haptic Adjective Corpus 2）数据集输入到视觉、触觉神经网络，进行迭代训练，得到最优的参数。

目标域训练：首先将源域模型的权重参数进行迁移，迁移到目标域；再添加分类模块，利用目标域训练集数据对分类模块进行训练；然后采用微调技术对模型参数进行微调，提高模型检测的准确率。最后利用测试集数据进行测试，完成番茄成熟度检测。

1.1　视觉触觉双重迁移学习融合算法

1.1.1　视觉迁移学习模型

分别将 Alexnet[13]、GoogLeNet[14]、VGGNet[15]、ResNet[16]、DenseNet[17]的迁移学习模型作为视觉特征提取模型，提取视觉特征向量 $\boldsymbol{d}_c$，并使用公开可用的番茄数据集进行对比实验，该数据集是从 http://www.kaggle.com 上在线发布的图像公开可用存储库中收集。实验发现 DenseNet169 迁移学习模型实验结果在准确率和稳定性上优于其他模型，结构如图 2 所示，$\boldsymbol{d}_c$ 由 DenseNet169 模型最后一层卷积层展平得到。

1.1.2　触觉迁移学习模型

由于本文使用的触觉数据均是一维信号，所以使用 5 层一维卷积迁移学习模型提取触觉特征向量 $\boldsymbol{d}_e$。该模型使用的激活函数为 ReLU，如图 3 所示。选择 PHAC－2 触觉数据集[18]作为触觉模型的源域数据集。PHAC－2 数据集由宾夕法尼亚大学收集，提供较为丰富触觉信息。首先在源域数据集上对触

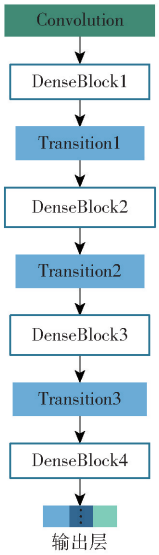

图 2 视觉迁移学习模型

Fig. 2 Visual transfer learning model

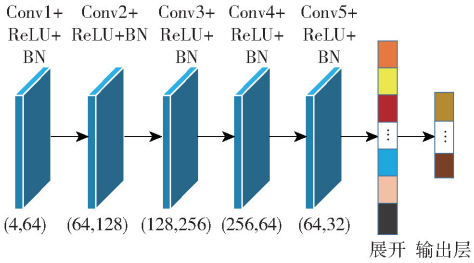

图 3 触觉迁移学习模型

Fig. 3 Tactile transfer learning model

觉模型进行迭代训练,得到最优的参数,然后将参数迁移到目标域数据集。触觉模型输出是触觉特征向量 $\boldsymbol{d}_e$,由第 5 层卷积层展平得到。

1.1.3 融合方法

视觉特征向量和触觉特征向量采用特征融合方法[19],其中最常见的做法是将两个特征向量串联。具体而言,特征向量 $\boldsymbol{d}_c$、$\boldsymbol{d}_e$ 分别来自视觉模型和触觉模型的最后一层展平操作,然后串联起来形成视触融合特征向量 $\boldsymbol{d}_w$,公式为

$$\boldsymbol{d}_w = [\boldsymbol{d}_c^{\mathrm{T}} + \boldsymbol{d}_e^{\mathrm{T}}]^{\mathrm{T}} \tag{1}$$

其中:视觉特征向量 $\boldsymbol{d}_c$ 和触觉特征向量 $\boldsymbol{d}_e$ 都是列向量。本文使用迁移学习的 DenseNet169 模型和 5 层一维卷积模型进行组合,构建的双重迁移学习融合算法如图 4 所示。

1.2 软参数共享-多标签分类方法

为了避免模型出现过拟合现象,采用了软参数共享[20]的分类方法。该方法主要对模型的下层参数之间的距离进行正则化,以迫使不同分类任务之间的参数相似,从而提高模型泛化能力。由于视觉和触觉的模态不同,在特征提取部分并没有进行参数共享,权重共享仅发生在分类器的全连接层部分,本文选择使用两个全连接层用于番茄成熟度检测。

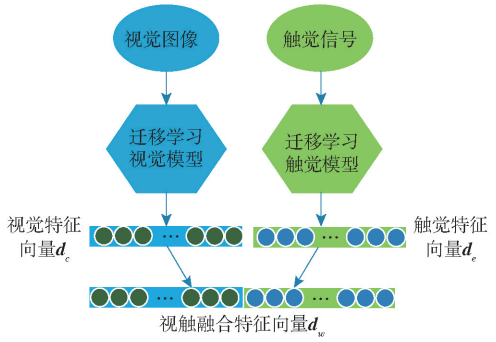

图 4 双重迁移学习融合算法原理图

Fig. 4 Schematic of double transfer learning fusion algorithm

传统的软参数共享方法是多种不同的任务具有特定的全连接层,这样增加整个模型参数量,如图 5b 所示。

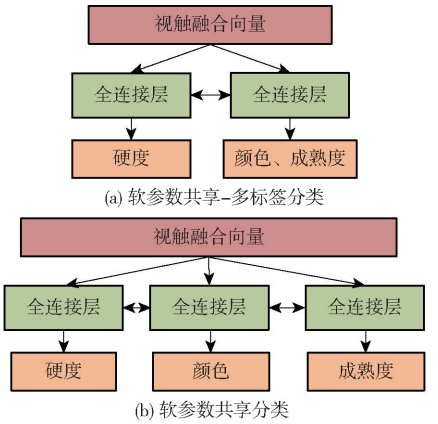

图 5 分类方法对比

Fig. 5 Comparison of classification methods

为了增加不同任务之间的关联性并减少模型的参数量,提出了一种软参数共享-多标签分类方法,如图 5a 所示,将硬度 $y^{(1)}$ 作为一个单任务学习,而将颜色 $y^{(2)}$ 与成熟度 $y^{(3)}$ 两个不同任务的标签合并成一个多标签学习,形成一个包含两个子任务的多标签任务学习。硬度与颜色均是成熟度相关的因素,但颜色与成熟度之间的共性特征较多。通过合并颜色和成熟度的标签,使得不同分类任务之间的关联性增强,从而提高了准确率。若把硬度、颜色、成熟度作为 3 个不同的分类任务,任务相对独立,其中一个任务为其他任务提供有用的知识有限,降低了不同任务同时分类正确的概率。该分类方法将原先的三任务学习变成双任务学习,仅需两个全连接层进行参数共享,从而减少了参数量,并提高了运算效率,具体实现方法公式为

$$\boldsymbol{y} = [\boldsymbol{y}^{(1)}] + [\boldsymbol{y}^{(2)} + \boldsymbol{y}^{(3)}] \tag{2}$$

1.3 定义损失函数

软参数共享-多标签分类方法产生的损失函数,包括 pyTorch 神经网络架构提供的多标签分类损失函数 MultilLabelSoftMarginLoss 与软参数共享产生的

损失函数L_s，其中多标签损失函数 Loss(Multilabel)公式为

$$L_{\mathrm{MI}}(x_l,y_l)=-\frac{1}{C}\sum_j y_l[j]\lg\left(\frac{1}{1+\mathrm{e}^{-y_l[j]}}\right)+(1-y_l[j])\lg\left(\frac{\mathrm{e}^{-x_l[j]}}{1+\mathrm{e}^{-x_l[j]}}\right) \tag{3}$$

式中　x_l——全连接层输出

C——类别数量

y_l——监督标签

$x_l[j]$——x_l 中第 j 个值

$y_l[j]$——y_l 中第 j 个值

从公式(3)可以得出，在神经网络反向传播过程中：当 $y_l[j]=1$，$x_l[j]$是一个比较大的正数才能保证损失取最小值；当 $y_l[j]=0$，$x_l[j]$是一个比较大的负数才能保证取损失最小值，即

$$x_l[j]=\begin{cases}1 & (x_l[j]>0)\\ 0 & (x_l[j]<0)\end{cases} \tag{4}$$

因此多标签分类的输出需要选择一个阈值将其调整为正负样本，本文以 0 为阈值对 $x_l[j]$进行调整。

采用 L2 范数对两个全连接层的参数进行正则化，实现软参数共享分类方法。W_1 和 W_2 分别表示全连接层下层参数，则软参数共享产生的损失函数为

$$L_s=\alpha[\,\|W_1-W_2\|\,] \tag{5}$$

式中　W_1、W_2——两个全连接层权重

α——L2 范数正则化参数，取 10^{-8}

采用各损失函数直接相加的方式获得总的损失函数，则软参数共享-多标签分类方法的损失函数 L_{sp}可表示为

$$L_{sp}=L_{\mathrm{MI}}+L_s \tag{6}$$

2　实验

2.1　VHACT 数据集

由于实验平台的触觉传感器为压力传感器，接触部分虽然采用弹性皮肤包裹着开孔泡沫塑料，具备一定的柔性特质，但仍需要施加一定的压力。因此对小果番茄并不适用，所以本文选择抗压能力较强的大红果番茄与大黄果番茄。目前的番茄成熟度数据集缺乏各级成熟度的硬度特征，而硬度特征又对番茄加工、运输和贮藏方式产生较大影响。为了解决这一问题，建立了 VHACT(Visual and Haptic Adjective Corpus of Tomatoes)的数据集。VHACT 数据集共收集 30 个番茄在抓取环境下的视觉和触觉数据，数据采集平台包括 RealSense D435i 型相机与 NumTac 型触觉传感器，分别安装在 Kinova 机械臂腕部与手指末端，如图 6 所示。尽管获取数据可能需要更多设备和人力成本，但可以获取番茄全面的信息。

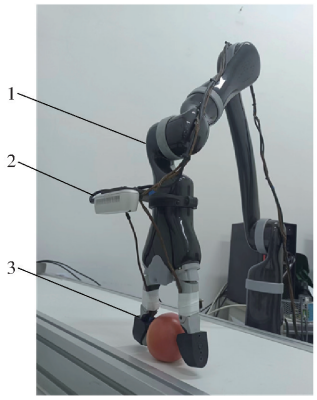

图 6　VHACT 数据集收集平台

Fig. 6　VHACT dataset collection platform

1. Kinova 机械臂　2. Intel RealSenseD435 型相机　3. NumTac 型触觉传感器

在视觉数据收集过程中，将番茄随意摆放在实验台上，以俯视的方式对番茄进行多角度拍照，模拟采摘过程中不同方向接近番茄的情况。通过这种方式，可以获得每个番茄的 10 幅不同角度图像，为后续视觉数据提供多角度的番茄信息，如图 7 所示。在触觉数据收集过程中，从番茄的上方进行抓取操作，每个番茄进行 10 次抓取，每次抓取都会旋转一定的角度，以避免在同一位置进行重复抓取。NumTac 型触觉传感器能够收集机械臂抓取物体过程中的低频流体压力 P_{DC}和高频流体振动信号 P_{AC}，其中 P_{AC}的采样频率为 2 200 Hz，P_{DC}的采样频率为 100 Hz。一次抓取可以分为夹取、保持和打开 3 个过程，模拟了采摘过程中与番茄接触情况。

在硬度方面，由于缺乏专业的硬度检测设备，无法直接对番茄硬度进行实际测量。参照文献[21]，发现将不同成熟阶段的番茄硬度划分为 3 个类别：软、中和硬。本文沿用这一分类标准。同时，从 3 个类别中各随机选择 4 个触觉数据，如图 8 所示。根据图中数据得知，不同硬度番茄对应着不同变化幅度，通过观察触觉信号的幅度变化，可以判断番茄硬度。根据国家标准GH/T 1193—2021[22]进行成熟度与颜色的标签制定，将番茄划分为 5 种颜色特征(绿色、白绿色、淡红色、红色、黄色)与 5 种成熟度(绿熟期、转色期、半熟期、成熟期、完熟期)。这种分类方式描述各级成熟度与颜色和硬度的关系，有助于机器视觉和触觉检测。根据不同成熟度体现出的硬度与颜色的变化，进行相应的加工、存储与管理，降低番茄腐败率，提高经济效益，具体描述如表 1 所示[23]。

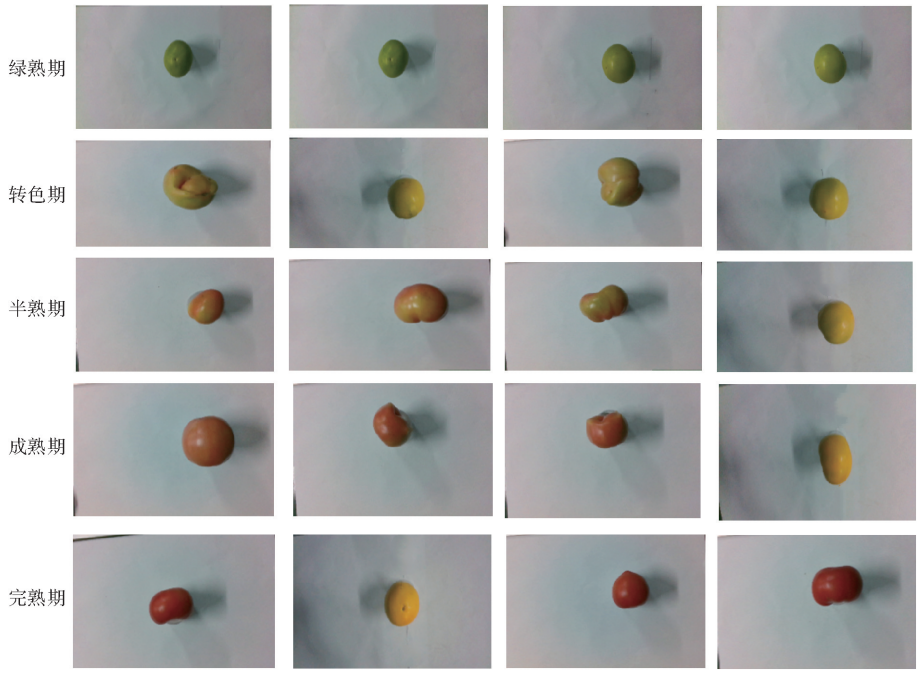

图 7 VHACT 数据集中部分视觉数据

Fig. 7 Selected visual data from VHACT dataset

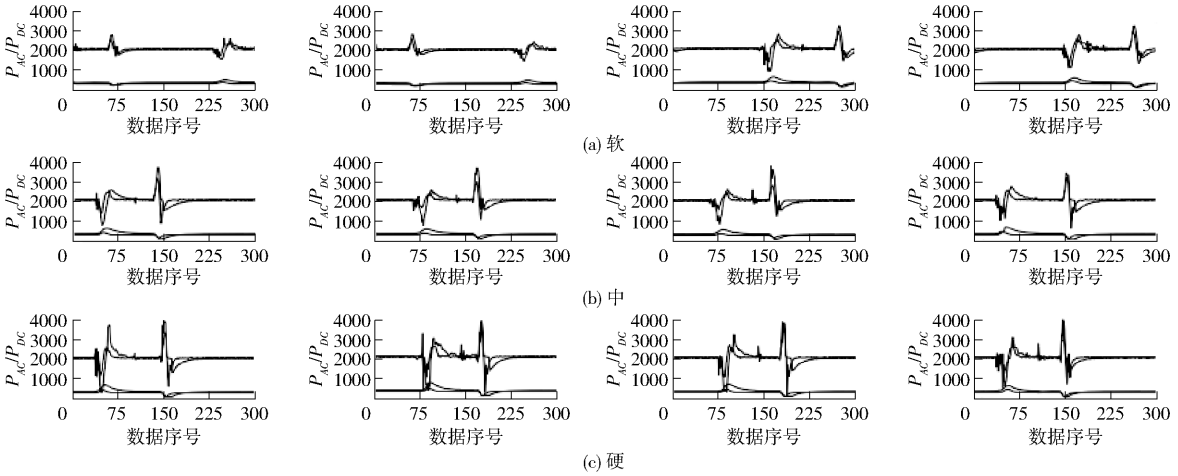

图 8 VHACT 数据集中部分触觉数据

Fig. 8 Selected haptic data from VHACT dataset

表 1 番茄成熟度描述

Tab. 1 Description of tomato maturity

成熟度	描述	采摘情况
绿熟期	果实定形,果实有光泽,由绿色变为白绿色,果实硬	可人工催熟,采摘贮藏或远距离运输
转色期	绿熟到成熟的过渡期,果脐周围开始出现黄色或淡红色晕斑,果实着红或着黄面积占比为 10% ~30%,果实硬	可采摘贮藏,中短距离运输
半熟期	果实着红或着黄面积 30% ~60%,果实硬度中等	可采摘,隔 2 d 上市鲜销
成熟期	果实着红或着黄面积 60% ~90%,果实硬度中等	可采摘,隔 1 d 上市鲜销
完熟期	果实表面 90% 以上为红色或黄色,果实软	可采摘,当日上市鲜销

2.2 数据预处理

本文沿用 GAO 等[24]数据处理方法,将 PHAC-2 数据集的 P_{AC} 下采样至 100 Hz,并使用主成分分析法将 19 个电极阻抗信号(E_1,E_2,…,E_{19})降维至 4 维阻抗信号,即经过处理后触觉信号由 P_{DC}、P_{AC}、T_{DC}、T_{AC}以及 4 维阻抗信号组成。将挤压、保持、慢速滑动、快速滑动 4 个动作过程都下采样到 300 个数据点长度,然后将 4 个动作过程串联得到 1 200 个数据点触觉数据,即每个物体触觉大小为 $C_P \times T_P$,其中 $C_P=4$ 表示触觉通道数,$T_P=1\ 200$ 表示数

据长度，最后对每一维数据进行归一化处理得到归一化值

$$s' = \frac{s - \bar{s}}{\sigma} \tag{7}$$

式中　$\bar{s}$——数据均值　　σ——数据偏差

s——原始数据

与 PHAC－2 数据集相比，VHACT 数据集的触觉数据只包含 P_{DC} 和 P_{AC} 两维数据，将 P_{AC} 下采样到 100 Hz，将每次抓取过程获得的触觉数据采样为 300 个数据点的长度，然后将左右手的触觉数据并联得到 $C_V \times T_V$ 的触觉表示，其中 $C_V = 4$、$T_V = 300$，部分的触觉数据如图 8 所示。在训练过程中将图像调整为 300 像素并从中心截取 224 像素的图像作为模型输入，为了避免在处理过程中丢失信息，未对图像进行任何处理，确保原始图像完整性，使得模型能够更好地捕捉图像中的特征和细节。

经过处理后的触觉数据，可以观察到变化幅度。当接触软的番茄时，触感传感器捕捉到的触觉数据呈现出振幅较小的特点，这反映了番茄表面的柔软性和一定的缓冲能力。同时，这些信号的上升和下降速度相对较慢。当碰到硬度中等的番茄时，触觉数据与软的番茄相比，显示出较大的振幅，信号的变化速度较快。而当触感传感器接触到硬度高的番茄时，触觉数据呈现出的变化就是数据振幅达到最大，信号变化速度较为迅速。

2.3　实验设置

采用 AUC（Area under curve）分值作为成熟度、硬度、颜色识别的评价指标。为了验证视触融合检测技术的有效性，进行 3 组消融实验，包括视觉检测技术、触觉检测技术以及视触融合检测技术，对比同等实验条件下的成熟度识别效果。

为了验证提出的视觉触觉双重迁移学习融合算法的有效性，进行了 4 组消融实验。在实验中，视觉模型、触觉模型和分类模块等参数保持一致，通过比较不同占比下的性能表现，评估迁移学习对番茄成熟度识别任务的作用。

为验证软参数共享-多标签检测模型的有效性，以图 5 作为特征提取融合模快，构建了自适应硬参数共享[25]、不确定性加权损失[26]、软参数共享[19]和十字绣网络[27]的检测模型，通过分析各个检测模型的识别效果，验证软参数共享-多标签检测模型的有效性。

实验中视觉和触觉是成对输入，即一幅图像配合一个触觉数据输入神经网络，每个番茄的视觉和触觉数据都一一对应。为加快神经网络的训练速度，对全连接层输出进行归一化操作，公式为

$$y = \frac{x - E[x]}{\sqrt{\mathrm{Var}[x] + G}}\gamma + \beta \tag{8}$$

式中，$\gamma = 1$，$\beta = 0$，$G = 1 \times 10^{-5}$，γ、β 是输入大小为 C 的参数向量。

为了验证在小样本情况下，模型能有效提取番茄特征信息，本文将 VHACT 数据集按照 8∶2、6∶4、4∶6 和 2∶8 的比例随机分成 4 种不同比例训练集和测试集，以验证模型的性能。

使用 PyTorch 构建神经网络模型，训练和测试过程在 Intel（R）HD Graphics 530 显卡上进行。在训练过程中，使用基于 Adam 的优化方法，批处理量大小设定为 12，学习率设置为 0.000 02，其余参数为默认值，周期设置为 250。

研究表明，微调参数可以提高模型的准确性，模型的性能由获得最高分类精度的参数值决定。本文的周期设置为固定值 250，批处理量大小（Batch size）设置 12、16、20，并评估了其对模型精度的影响。同样，学习率分别设置为 0.000 015、0.000 02 和 0.000 025。

2.4　实验结果分析

2.4.1　不同视觉模型实验结果对比

图 9 显示了在公共番茄数据集上不同视觉模型对成熟度分类结果。从图中可以看出，随着训练次数不断增加，模型性能都在不断的提升。但在不同训练周期内，各模型之间的差异也较为明显。在所有的视觉模型中，DenseNet169 表现出了较高的识别性能，比其他模型具有更好的精度和稳定性。无论是在训练初期还是后期，DenseNet169 精度都明显高于其他模型。这种优势不仅仅体现在最终精度上，更体现在整个训练过程中的稳定性。即使在训练初期，DenseNet169 就已经展现出了较高的精度，并且

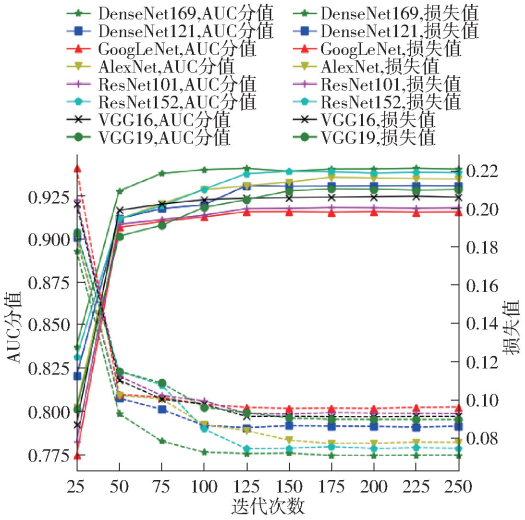

图 9　不同视觉模型实验结果对比

Fig. 9　Comparison of experimental results of different visual models

随着训练进行,其精度还在稳步提升。

2.4.2 视触融合检测技术消融实验

为了验证视触融合技术的检测效果,进行与视觉检测技术、触觉检测技术的消融实验,实验结果如表 2 所示。

表 2 检测技术消融实验结果对比

Tab. 2 Comparison of ablation experimental results of detection techniques

训练集与测试集比例	检测技术	参数量	训练时间/min	成熟度(AUC 分值)
8:2	视觉	1.446×10^7	278	0.969 9
	触觉	1.580×10^6	134	0.921 7
	视触融合	1.564×10^7	350	0.978 4
6:4	视觉	1.446×10^7	242	0.950 7
	触觉	1.580×10^6	106	0.918 7
	视触融合	1.564×10^7	326	0.972 7
4:6	视觉	1.446×10^7	213	0.930 6
	触觉	1.580×10^6	85	0.898 0
	视触融合	1.564×10^7	291	0.952 0
2:8	视觉	1.446×10^7	185	0.900 5
	触觉	1.580×10^6	69	0.878 7
	视触融合	1.564×10^7	264	0.942 5

由表 2 可知,视触融合检测技术具有最大的参数量和最长的训练时间。这是因为该技术融合了视觉和触觉检测技术,在提取特征信息时涵盖了更多的数据,因此需要更多的时间处理数据。视觉检测技术比触觉效果要好。主要由于视觉通过外观颜色差异直接识别成熟度,而机器触觉检测硬度特征受番茄物质的影响,可变因素较多。视触融合检测技术的效果最好,融合了硬度与颜色信息,获取的特征信息更加全面。因此,视触融合检测技术的识别能力整体最优,在番茄成熟度检测中可行且有效。

2.4.3 双重迁移学习融合算法消融实验

为了验证视觉触觉双重迁移学习融合算法在识别成熟度中的有效性,进行了 4 组消融实验,分别为算法Ⅰ(视觉、触觉均无迁移学习融合)、算法Ⅱ(迁移学习视觉与无迁移学习触觉融合)、算法Ⅲ(迁移学习触觉与无迁移学习视觉融合)、算法Ⅳ(视觉触觉双重迁移学习融合)。为了减少实验结果的随机性,每种比例下进行了 4 次实验,分别获取 4 次实验结果及平均值,实验结果如表 3 所示。

从表 3 中数据可以看出,迁移学习的识别效果均优于无迁移学习。当训练数据比例较大时,模型无论是否使用迁移学习,AUC 分值均能达到 0.90 以上。然而,随着测试数据增加,迁移学习与无迁移学习的检测效果差距变大,这表明即使训练集中只有少量数据,预训练模型的知识仍然对模型的性能有一定影响。

表 3 双重迁移学习融合算法消融实验结果对比

Tab. 3 Comparison of ablation experimental results of dual migration learning fusion algorithms

训练集与测试集比例	算法	成熟度(AUC 分值)				
		实验 1	实验 2	实验 3	实验 4	平均值
8:2	Ⅰ	0.930 2	0.929 1	0.916 3	0.924 1	0.924 9
	Ⅱ	0.964 4	0.967 2	0.962 3	0.963 2	0.964 3
	Ⅲ	0.942 2	0.942 8	0.952 3	0.942 2	0.944 9
	Ⅳ	0.983 9	0.985 6	0.983 0	0.981 8	0.983 6
6:4	Ⅰ	0.890 7	0.907 9	0.903 9	0.906 8	0.902 3
	Ⅱ	0.951 1	0.950 1	0.952 4	0.951 0	0.951 2
	Ⅲ	0.911 4	0.917 6	0.921 2	0.921 4	0.917 9
	Ⅳ	0.976 1	0.975 7	0.977 6	0.973 7	0.975 8
4:6	Ⅰ	0.865 3	0.869 6	0.889 0	0.872 9	0.874 2
	Ⅱ	0.938 8	0.930 6	0.941 0	0.937 6	0.937 0
	Ⅲ	0.910 0	0.914 4	0.912 1	0.907 8	0.911 1
	Ⅳ	0.960 9	0.964 2	0.956 5	0.956 2	0.959 5
2:8	Ⅰ	0.740 2	0.757 9	0.734 1	0.703 9	0.734 0
	Ⅱ	0.921 0	0.914 5	0.915 0	0.907 6	0.914 5
	Ⅲ	0.894 5	0.901 4	0.892 4	0.876 6	0.891 2
	Ⅳ	0.940 9	0.947 5	0.940 9	0.942 0	0.942 8

进一步对比数据发现,双重迁移学习的识别效果最好,AUC 分值平均达到 0.965 4,其次迁移学习视觉网络比触觉网络效果要好,这是因为不同成熟度的番茄颜色差异明显,视觉网络提取的颜色特征可以直接反映番茄各级成熟度存在的差异。随着番茄越来越熟,硬度虽然会逐渐降低,但是触觉网络提取到的硬度特征易受其他干扰因素的影响,所以实验效果较差。综上所述,在保留较多训练数据的情况下,使用迁移学习可以提高成熟度检测的准确率;而在训练集数据较少时,迁移学习仍显著提升模型性能。

2.4.4 检测模型参数微调实验

为了进一步研究批处理量和学习率对模型性能的影响,本文在保证单一变量不变的情况下,更改其他变量,比较单一变量对模型性能的影响,实验结果如图 10 所示。

由图 10 可知,不同的参数组合对模型检测效果影响较大。随着批处理量的增加,模型的整体准确率也随着增加,在批处理量为 16 时,准确率在不同比例下均达到较高值,主要是较大的批处理量意味着每个批次包含更多的样本,可以更好地代表整体数据分布。这有助于减少随机性的影响,稳定模型的训练过程,并降低模型的方差。通过在每个较大批次上计算梯度的平均值,可以减小梯度的方差,从而更稳定地更新模型参数。当批处理量增加到 20 时准确率出

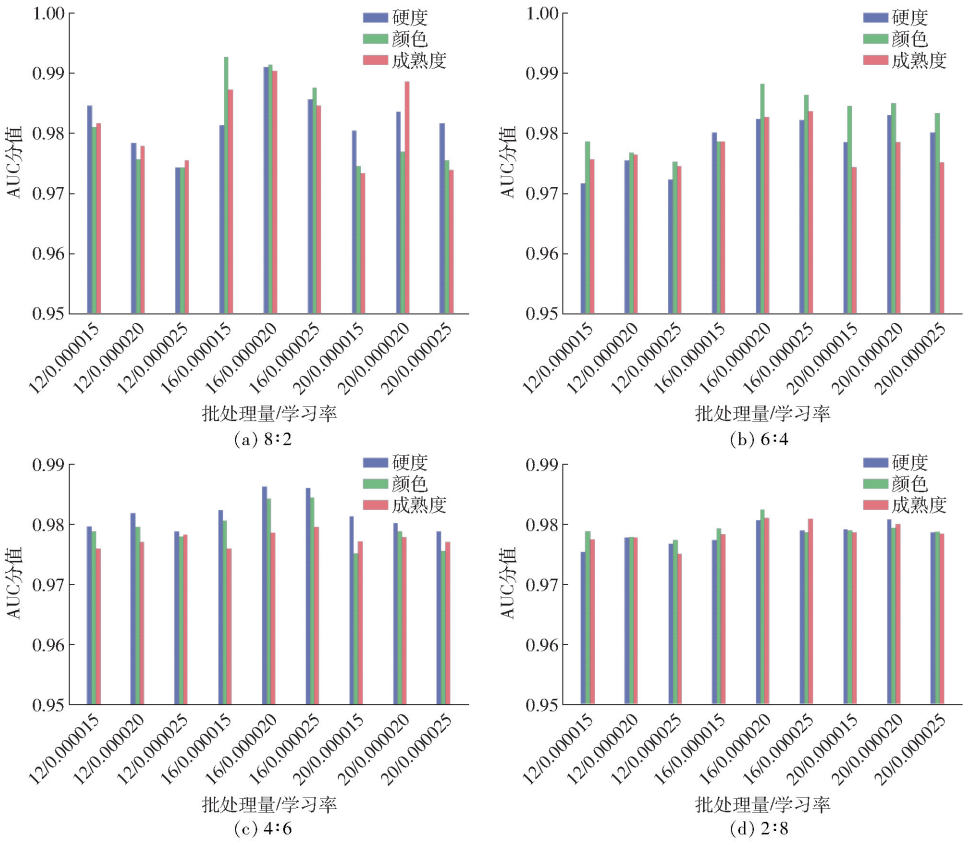

图 10　参数微调实验结果对比

Fig. 10　Comparison of experimental results of parameter fine-tuning

现下降,主要是过大的批处理量可能会使模型陷入局部最优解而无法达到更好的泛化能力。

从图 10 可以看出,在批处理量为 16 时,学习率为 0.000 02 比 0.000 015 的效果较好,意味着模型在每次参数更新时采取稍大的步幅。这会加快模型的收敛速度,并在较短的时间内达到较好的性能。在训练过程中,模型可能会陷入局部最优点的不利情况,较大的学习率可以帮助模型跳出这些局部最优点,向着更好的全局最优点前行。同时较大的学习率具有一定的正则化效果,可以减少模型对训练集的过拟合。过拟合往往是由于模型过于复杂或数据集噪声导致的,较大的学习率有助于使模型的参数更加平滑,并抑制过拟合现象。学习率如果设置过大,会导致模型在训练过程中不稳定地收敛。模型参数更新的步幅太大,会错过最优解,导致训练过程无法达到理想的收敛状态。

最后,综合选择批处理量为 16、学习率为 0.000 02 作为参数值,既可以平衡各分类任务准确性的同时,又达到最佳的整体性能。

2.4.5　不同检测模型实验结果对比

在确定模型最佳参数情况下,进一步验证软参数共享-多标签检检测模型(模型Ⅴ)的效果,复现了不确定性加权损失[23](模型Ⅰ)、自适应硬参数共享[24](模型Ⅱ)、十字绣网络[26](模型Ⅲ)和软参数共享[20](模型Ⅳ)等检测模型,并在 VHACT 数据集上进行对比实验,实验结果如表 4 所示。

通过对比数据发现,软参数共享-多标签检测模型参数量最少,仅有 1.882×10^7。这是因为在一个全连接层中合并两个分类任务,实现统一模型架构,使得模型设计更加简单,减少了一些重复工作和冗余结构。实验中不确定性加权损失、软参数共享和十字绣网络经过统计发现三者参数量仅相差 1 000,对整体网络参数影响较小。自适应硬参数共享检测模型参数量最大,达到 2.498×10^7,主要是自适应硬参数共享分类方法添加了一个自适应节点层,不同的任务各自具有独立的全连接层,这就造成模型参数量相对较大。

从表 4 可以看出,软参数共享-多标签检测模型在不同比例下软硬度、颜色、成熟度分类方面均略好于其他方法,AUC 分值平均在 0.94 以上,将不同分类任务的特征融合在全连接层中可以促进不同任务之间的信息交流和发现共性,这有助于模型更好地理解和捕捉跨任务关联性。十字绣网络分类方法整体效果最差,主要由于不同任务之间的相关性较低或者不存在,则共享特征会降低模型的分类性能。随着测试集数据的增加,各种检测模型的准确率均出现不同程度的下降,主要是随着测试数据的增多,增加了分类的难度,但是

表 4 不同检测模型性能结果对比

Tab. 4 Comparison of performance results of different detection models

训练集与测试集比例	检测模型	参数量	AUC 分值		
			硬度	颜色	成熟度
8∶2	Ⅰ	2.190×10^7	0.981 2	0.977 4	0.971 2
	Ⅱ	2.498×10^7	0.983 9	0.988 2	0.977 5
	Ⅲ	2.190×10^7	0.986 2	0.953 7	0.965 4
	Ⅳ	2.190×10^7	0.985 3	0.985 5	0.984 1
	Ⅴ	1.882×10^7	0.991 0	0.991 3	0.990 4
6∶4	Ⅰ	2.190×10^7	0.972 9	0.969 9	0.964 3
	Ⅱ	2.498×10^7	0.976 6	0.966 2	0.964 2
	Ⅲ	2.190×10^7	0.974 9	0.955 4	0.955 4
	Ⅳ	2.190×10^7	0.974 5	0.977 2	0.974 3
	Ⅴ	1.882×10^7	0.982 4	0.988 2	0.982 7
4∶6	Ⅰ	2.190×10^7	0.975 3	0.952 4	0.953 0
	Ⅱ	2.498×10^7	0.972 3	0.963 6	0.952 4
	Ⅲ	2.190×10^7	0.971 5	0.952 3	0.950 5
	Ⅳ	2.190×10^7	0.972 1	0.962 1	0.961 8
	Ⅴ	1.882×10^7	0.983 5	0.981 1	0.974 3
2∶8	Ⅰ	2.190×10^7	0.955 5	0.950 1	0.950 7
	Ⅱ	2.498×10^7	0.952 1	0.953 5	0.943 5
	Ⅲ	2.190×10^7	0.943 1	0.945 5	0.931 0
	Ⅳ	2.190×10^7	0.953 2	0.956 7	0.953 6
	Ⅴ	1.882×10^7	0.961 3	0.964 8	0.961 9

软参数共享-多标签检测模型下降最少，仍然保持较高的准确率，说明具有较强的泛化能力。

图 11 为 5 种检测模型在硬度、颜色和成熟度方面的实验结果，其中图 11 中的数据为不同检测模型在不同比例下的平均值。从图 11a 可知，不同检测模型 AUC 分值均在 0.96 以上。主要由于硬度的等级划分较少，降低了识别难度。从图 11b 可知，绿色、红色与黄色的准确率较高，主要是它们颜色为深色，具有较高的识别度，容易提取这些特征信息，而白绿色与淡红色为浅色，掺杂其他色素的影响，导致识别准确率较低。由图 11c 可知，在成熟度方面，绿熟期、成熟期、完熟期识别效果较好，主要是这些成熟度颜色特征较为统一，并不掺杂着其他色素的影响，硬度特征也较为明显。而转色期与半熟期的颜色为多种颜色混合，识别较为困难，硬度特征与绿熟期、成熟期差异也不明显。但是，软参数共享-多标签检测模型具体表现较为均衡，主要增加不同分类任务之间的相关性，不同分类任务相互依赖，其中一个任务可以为其他任务提供有益的知识，从而提高整体分类精度。

3 结束语

为了提高自动化采摘过程中番茄成熟度识别的准确率，提出了一种基于视觉触觉双重迁移学习的番茄成熟度检测方法。为了验证该方法的性能，首先，进行视触融合技术消融实验，验证了视触融合技术在番茄成熟度识别上的优越性；其次，进行了触觉双重迁移学习融合算法的消融实验，实验表明，双重迁移学习融合算法的识别精度均高于其他 3 组算法；然后，采用参数微调技术进一步提高软参数共享-多标签模型的检测能力；最后，通过与不确定性加权损失、自适应硬参数共享、十字绣网络和软参数共享的检测模型对比，软参数共享-多标签检测模型取得最优的效果，参数量分别下降 3.08×10^6、6.16×10^6、3.08×10^6 和 3.08×10^6，成熟度 AUC 分值分别提高 0.017 5、0.017 9、0.026 7、0.008 9。实验充分证明该方法以最小的参数量实现了最精准的识别效果。制作了一个番茄的视觉触觉数据集，相较于其他番茄数据集，增加了番茄不同成熟度的硬度特征，更加全面地反映了番茄的成熟特征。

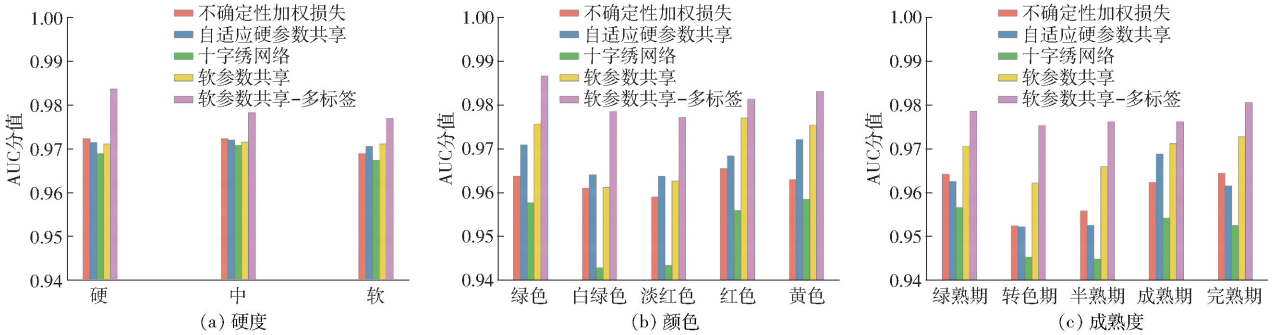

图 11 不同检测模型在硬度、颜色和成熟度方面实验结果

Fig. 11 Experimental results on hardness, color, and maturity of different detection models

参 考 文 献

[1] 黄玉萍, LU Renfu, 戚超, 等. 基于空间可分辨光谱的番茄成熟度判别方法研究[J]. 光谱学与光谱分析, 2018, 38(7): 2183 - 2188.

HUANG Yuping, LU Renfu, QI Chao, et al. Tomato maturity classification based on spatially resolved spectra[J]. Spectroscopy and Spectral Analysis, 2018, 38(7): 2183 - 2188. (in Chinese)
[2] 马锃宏,谭励,曾伟,等. 基于视触觉感知的番茄尺寸和姿态解析方法[J]. 农业机械学报,2024,55(1):223 - 232.
MA Zenghong, TAN Li, ZENG Wei, et al. Analytical method and evaluation of tomato size and posture based on visual and tactile perception[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(1):223 - 232. (in Chinese)
[3] 杨信廷,刘彤,韩佳伟,等. 基于 Swin Transformer 与 GRU 的低温贮藏番茄成熟度识别与时序预测研究[J]. 农业机械学报, 2024,55(3):213 - 220.
YANG Xinting, LIU Tong, HAN Jiawei, et al. Low temperature storage tomato maturity recognition and time series prediction based on Swin Transformer - GRU[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(3):213 - 220. (in Chinese)
[4] BEGUM N, HAZARIKA M K. Maturity detection of tomatoes using transfer learning[J]. Measurement: Food, 2022, 7: 100038.
[5] PANGILINAN J R, LEGASPI J, LINSANGAN N. InceptionV3, ResNet50, and VGG19 performance comparison on tomato ripeness classification[C]// 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). IEEE, 2022: 619 - 624.
[6] RAFIQ A, MAKROO H A, HAZARIKA M K. Artificial neural network-based image analysis for evaluation of quality attributes of agricultural produce[J]. Journal of Food Processing and Preservation, 2016, 40(5): 1010 - 1019.
[7] 邵园园, 王永贤, 玄冠涛, 等. 基于高光谱成像的肥城桃品质可视化分析与成熟度检测[J]. 农业机械学报, 2020, 51(8): 344 - 350.
SHAO Yuanyuan, WANG Yongxian, XUAN Guantao, et al. Visual detection of SSC and firmness and maturity prediction for Feicheng peach by using hyperspectral imaging[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(8): 344 - 350. (in Chinese)
[8] HUANG Y, XIONG J, JIANG X, et al. Assessment of firmness and soluble solids content of peaches by spatially resolved spectroscopy with a spectral difference technique[J]. Computers and Electronics in Agriculture, 2022,200:107212.
[9] VALEO C, CRRISOSTO C H, SLAUGHTER D. Relationship between nondestructive firmness measurements and commercially important ripening fruit stages for peaches, nectarines and plums[J]. Postharvest Biology and Technology, 2007, 44(3): 248 - 253.
[10] SCIMECA L, MAILINO P, CARDIN-CATALAN D, et al. Non-destructive robotic assessment of mango ripeness via multi-point soft haptics[C]//2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 1821 - 1826.
[11] HE L, TAO L, MA Z, et al. Cherry tomato firmness detection and prediction using a vision-based tactile sensor[J]. Journal of Food Measurement and Characterization, 2024, 18(2): 1053 - 1064.
[12] 王俊平,徐刚. 机器视觉和电子鼻融合的番茄成熟度检测方法[J]. 食品与机械, 2022(2): 148 - 152.
WANG Junping, XU Gang. Research on tomato maturity detection method based on machine vision and electronic nose fusion [J]. Food & Machinery,2022(2):148 - 152. (in Chinese)
[13] ALOM M Z, TAHA T M, YAKOPCIC C, et al. The history began from alexnet: a comprehensive survey on deep learning approaches[J]. arXiv Preprint, arXiv:1803.01164, 2018.
[14] TANG P, WANG H, KWONG S. G - MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition [J]. Neurocomputing, 2017, 225: 188 - 197.
[15] WANG L, GUO S, HUANG W, et al. Places205-VGGNet models for scene recognition[J]. arXiv Preprint, arXiv:1508.01667, 2015.
[16] TARG S, ALMEIDAl D, LYMANY K. Resnet in resnet: generalizing residual architectures[J]. arXiv Preprint, arXiv:1603.08029, 2016.
[17] IANDOLA F, MOSKEWICZ M, KARAYEV S, et al. DenseNet: implementing efficient convnet descriptor pyramids[J]. arXiv Preprint, arXiv:1404.1869, 2014.
[18] CHU V, MCMAHON I, RIANO L. Robotic learning of biological objectives through physical interaction [J]. Robotics and Autonomous Systems, 2015, 63: 279 - 292.
[19] BEDNAREK M, KICKI P, WALAS K. On robustness of multi-modal fusion-robotics perspective[J]. Electronics, 2020, 9(7): 1152.
[20] 冯兴杰,赵新阳,冯小荣. 基于软参数共享的事件联合抽取方法[J]. 计算机应用研究, 2023, 40(1): 91 - 96.
FENG Xingjie, ZHANG Xinyang, FENG Xiaorong. Joint event extraction method based on soft parameter sharing[J]. Application Research of Computers, 2023,40(1): 91 - 96. (in Chinese)
[21] SEHGAL P, GOEL N. Auto-annotation of tomato images based on ripeness and firmness classification for multimodal retrieval [C]//2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 2016: 1084 - 1091.
[22] GH/T 1193—2021 番茄[S]. 2021.
[23] 苗荣慧,李志伟,武锦龙. 基于改进 YOLO v7 的轻量化樱桃番茄成熟度检测方法[J]. 农业机械学报,2023,54(10):225 - 233.
MIAO Ronghui, LI Zhiwei, WU Jinlong. Lightweight maturity detection of cherry tomato based on improved YOLO v7[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(10): 225 - 233. (in Chinese)
[24] GAO Y, HENDRICKS L A, KUCHENBECKER K J, et al. Deep learning for tactile understanding from visual and haptic data [C]//2016 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2016: 536 - 543.
[25] WANG H, JIN X, DU Y, et al. Adaptive hard parameter sharing method based on multi-task deep learning[J]. Mathematics, 2023, 11(22): 4639.
[26] KENDALL A, GAL Y, CIOLLA R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7482 - 7491.
[27] MISRA I, SHIVASTAVAH A, GUPTA A, et al. Cross-stitch networks for multi-task learning[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016: 3994 - 4003.