中图分类号：TP391　　单位代号：10280
密　　级：公开　　学　　号：23721514

上海大学

硕士学位论文

SHANGHAI UNIVERSITY
MASTER'S DISSERTATION

题目	面向农业果蔬多属性联合评估的小样本多任务学习方法研究

作　　者　葛嘉浩
学科专业　计算机应用技术
导　　师　韩越兴
完成日期　二〇二六年四月

姓　　名：葛嘉浩　　学号：23721514

论文题目：面向农业果蔬多属性联合评估的小样本多任务学习方法研究

上海大学

本论文经答辩委员会全体委员审查，确认符合上海大学硕士学位论文质量要求。

答辩委员会签名：

主　　席：

委　　员：

导　　师：

答辩日期：2026 年 6 月 9 日

姓　　名：葛嘉浩　　学号：23721514

论文题目：面向农业果蔬多属性联合评估的小样本多任务学习方法研究

上海大学学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师指导下，独立进行研究工作所取得的成果。除了文中特别加以标注和致谢的内容外，论文中不包含其他人已发表或撰写过的研究成果。参与同一工作的其他研究者对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：葛嘉浩

日期：2026 年 6 月 9 日

上海大学学位论文使用授权说明

本人完全了解上海大学有关保留、使用学位论文的规定，即：学校有权保留论文及送交论文复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容。

(保密的论文在解密后应遵守此规定)

学位论文作者签名：葛嘉浩　　导师签名：韩斌杰

日期：2026 年 6 月 9 日　　日期：2026 年 6 月 9 日

上海大学工学硕士学位论文

面向农业果蔬多属性联合评估的小样本多任务学习方法研究

作　　者: 葛嘉浩
导　　师: 韩越兴
学科专业: 计算机应用技术

计算机工程与科学学院
上海大学
2026 年 4 月

A Dissertation Submitted to Shanghai University for the
Degree of Master in Engineering

Research on Few-Shot Multi-Task Learning Methods for Multi-Attribute Joint Evaluation of Agricultural Fruits and Vegetables

Candidate: Jiahao Ge
Supervisor: Yuexing Han
Major: Technology of Computer Application

School of Computer Engineering and Science
Shanghai University
April, 2026

摘　要

近年来，深度学习技术的发展推动了人工智能在农业视觉分析领域的应用。在果蔬采后质量检测领域中，视觉模型能够为无损检测与自动分选提供技术支撑。然而，当前研究仍面临小样本条件下监督信息不足和异构任务联合优化困难等问题。一方面，公开数据中图像与多属性标签严格对齐的样本较少，模型容易出现过拟合；另一方面，重量、曲直度和成熟度等连续属性在采集时通常只覆盖少量离散状态，端点样本之间缺少能够表征渐变过程的中间样本，常规增强难以补充这类变化信息。

以解决小样本监督不足问题和实现多属性联合评估目标为出发点，本文提出面向农业果蔬多属性联合评估的小样本多任务学习方法，用于重量估计、关键表型特征回归与等级分类，并在黄瓜与香蕉数据上进行系统验证。具体工作如下：

（1）针对小样本条件下多任务特征干扰和回归分类难以协同的问题，提出基于多任务学习的果蔬多属性联合评估框架。该方法包括预分类路由、任务分支建模、特征增强与跨分支交互、联合损失优化。为支撑该方法训练与评估，进一步构建 FruVegSet 多属性对齐数据集，基于重量、曲直度与成熟度测量结果并结合等级映射规则，实现图像、连续属性与等级标签的一一对应。在当前数据规模与实验协议下，对比单任务模型和其他多任务模型，该方法在黄瓜和香蕉数据上取得较好的综合表现，尤其在任务平衡与等级判别方面表现稳定。

（2）针对小样本条件下连续属性的端点样本之间过渡状态缺失的问题，在多任务联合评估框架基础上进一步提出基于中间样本生成的多任务扩展学习方法。该方法通过 DiffMorpher 生成中间过渡样本，结合 SAM2 分割模型、主体掩膜筛选策略与分任务伪标签构建提升补充监督质量。在该策略作用下，该方法在关键表型特征回归与等级分类任务中取得稳定提升，并一定程度改善三项任务的整体表现。

关键词：多任务学习；多属性联合评估；中间样本生成；伪标签构建；小样本学习

ABSTRACT

In recent years, the development of deep learning has promoted the application of artificial intelligence in agricultural visual analysis. In postharvest fruit and vegetable quality inspection, vision models can provide technical support for nondestructive inspection and automated sorting. However, current studies still face two key challenges, namely insufficient supervision under few-shot conditions and difficult joint optimization across heterogeneous tasks. On one hand, publicly available datasets with strictly aligned image-level multi-attribute labels are limited; on the other hand, continuous attributes such as weight, straightness, and maturity are usually sampled at only a few discrete states during data collection, and representative intermediate samples between endpoints are often missing, which is difficult to compensate for using conventional augmentation.

To address insufficient supervision under few-shot conditions and achieve joint evaluation of multiple attributes, this thesis proposes a few-shot multitask learning method for agricultural fruits and vegetables, targeting weight estimation, key phenotypic attribute regression, and grade classification, and validates it systematically on cucumber and banana datasets. The main work is as follows:

(1) To address feature interference across tasks and the difficulty of coordinating regression and classification under few-shot conditions, a fruit and vegetable multi-attribute joint evaluation framework based on multitask learning is proposed. The framework includes preclassification routing, task-specific branch modeling, feature enhancement with cross-branch interaction, and joint loss optimization. To support training and evaluation, a multi-attribute aligned dataset, FruVegSet, is further constructed. Based on measurements of weight, straightness, and maturity together with grade mapping rules, one-to-one alignment between images, continuous attributes, and grade labels is established. Under the current data scale and experimental protocol, compared with single-task models and other multitask models, the framework achieves good overall performance on both cucumber and banana datasets, especially in task balance and grade discrimination.

(2) To address the lack of transition states between endpoint samples of continuous attributes under few-shot conditions, a multitask extended learning method based on intermediate sample generation is further proposed on top of the multitask joint evaluation framework. The method generates intermediate transition samples using DiffMorpher and improves the quality of supplementary supervision by combining a SAM2 segmentation model, a principal mask selection strategy, and task-specific pseudo-label construction. With this strategy, stable gains are achieved in key phenotypic attribute regression and grade classification, and the overall performance of the three tasks is improved to a certain extent.

Keywords: multitask learning; joint evaluation of multiple attributes; intermediate sample generation; pseudo-label construction; few-shot learning

目 录

第一章 绪 论

1.1 课题来源

本课题得到国家自然科学基金（面上，编号：52273228），中国国家自然科学基金青年科学基金项目（编号：62401352，62002215），上海市科技青年人才扬帆计划（编号：23YF1412900），教育部硅酸盐质文物保护重点实验室（上海大学）项目（编号：SCRC2024ZZ07TS，SCRC2024ZZ05TS）以及上海市浦江计划（编号：20PJ1404400）资助。

1.2 课题背景概述

随着全球人口增长、城市化进程加快以及农产品流通规模持续扩大，农业采后处理环节对自动化、标准化和高效率检测的需求不断提升。传统依赖人工经验的称重、分拣与分级流程虽然应用广泛，但在处理效率、结果一致性和长期用工成本方面存在明显瓶颈。相关研究指出，人工智能技术正在农业场景中形成由感知到决策的技术链条，应用范围已覆盖监测、识别、品质评估与流程优化[1-3]。智能农业系统不仅依赖深度学习模型本身，还需要将模型部署到持续运行的数据系统中[4-6]。围绕这一趋势，物联网与人工智能的融合、传感网络协同和系统级评估方法也在同步推进[7-8]。此外，气候适应性与多源信息融合研究进一步表明，农业智能化已经从单模态识别扩展到跨模态协同[9-11]。

在果蔬流通与商品化场景中，重量、关键表型特征和质量等级是决定定价、包装、运输和市场接受度的核心属性。重量反映交易计量与产量信息；关键表型特征反映外观状态与规格差异；质量等级反映商品品质与流通标准。长期以来，这些属性主要依赖人工称量、人工测量和目视判别获取，难以满足大批量、无损、快速与稳定检测的现实需求[12]。

随着机器视觉技术的发展，基于图像的果蔬属性分析已具备可行性。面向无接触计量的研究表明，视觉重量估计在受控条件下可以获得较稳定精度[13-14]。在形状与规格分析方面，基于轮廓和几何结构的表型测量方法也在持续完善[15]。在质量检测方面，实时检测、采后质量控制和田间鲁棒评估等研究正在并行推进[12,16-19]。但

多数方法仍围绕单一目标优化，尚未充分利用多属性之间的协同关系。

除建模问题外，数据条件也是关键约束。公开数据集中，部分数据主要支持类别识别，如 Fruits-360 一类的数据资源更侧重类别标签[20-21]。也有数据集开始覆盖体积或质量标签，例如 AppleV、FruitQ 以及若干面向质量评估的数据集[22-28]。然而，这些数据通常仍以单任务为主，难以直接支撑图像、连续属性和等级标签的严格对齐联合学习。同时，重量、曲直度、成熟度等连续属性在实际采集中常表现为离散采样，端点状态之间缺少代表性中间样本；常规翻转、旋转和裁剪等增强虽然能扩充数量，但难以补充新的状态语义[29-31]。

因此，面向果蔬多属性联合分析，当前需要同时回答两个核心问题：其一，如何在统一视觉输入下实现重量、关键表型特征与等级任务的协同建模；其二，如何在小样本条件下补充连续属性的过渡状态样本并构建可用监督。围绕上述问题开展研究，是推动果蔬采后智能检测由单项识别走向一体化评估的重要基础。

1.3 课题研究的目的与意义

果蔬在采后处理、流通运输和商品销售阶段，需要完成称重、规格判定和品质分级等多个环节。传统流程通常依赖人工称量、人工目测和简单机械分选，在效率、一致性和长期成本方面存在明显限制。随着果蔬商品化和标准化程度提高，行业对无损、快速、稳定的自动检测技术提出了更高要求。基于此，本文将研究对象统一为重量估计、关键表型特征预测和质量等级分类三项任务，并围绕三者的联合建模开展研究。

从研究目的看，重量和关键表型特征属于连续属性，品质等级属于离散判定结果，三者在同一视觉对象上存在耦合关系。若分别建模，容易出现特征重复学习、任务信息利用不足和系统部署冗余等问题。因此，本文的核心目标是：在统一视觉输入下构建可同时支持连续属性预测与等级判定的协同学习方法，并在小样本条件下提升模型对属性变化规律的学习能力。

从理论意义看，该问题同时包含回归与分类两类监督形式，能够为异构任务条件下的共享表示学习、任务协同与任务干扰分析提供具体研究场景。围绕果蔬多属性联合建模开展研究，有助于检验多任务学习和特征增强方法在农业视觉场景中的适用边界，并为后续方法改进提供可验证的问题设定。

从应用意义看，若单次成像可同时输出重量、关键表型特征和质量等级，可减少称重、测量和人工分级等环节切换，提高分选流程自动化程度。该能力可直接服务于分拣、包装、仓储、运输和零售等环节，提升定价与分级一致性，降低由重复操作和误判带来的采后损耗。

综上，本文的研究意义主要在于针对小样本条件下的多属性联合分析问题，给出可复现的建模路径与实验验证，为后续工程化应用提供更明确的方法依据。

1.4 国内外研究现状

现有果蔬智能检测研究已经从传统机器视觉逐步发展到深度学习驱动的自动分析框架。国际综述指出，该领域已形成外观质量分类、重量估计和关键表型特征分析三类主线[12]。另有综述进一步强调，深度学习正在替代传统规则方法成为主流技术路线[32]。

围绕采后质量控制，研究正由单一分类逐步扩展到实时检测与流程优化[16]。在采后环节的系统化应用方面，也有研究关注智能评估链路重构[17]。与此同时，无损检测技术体系和田间鲁棒评估研究也在持续推进[18-19]。

围绕跨场景监测与模型持续学习，已有研究探索增量更新机制[33]。也有工作关注农业视觉模型在复杂环境中的泛化问题[3,34]。国内相关工作同样形成了从基础综述到工程应用的研究脉络[35-36]。在无损检测体系方面，国内研究已覆盖关键技术与应用流程[37]。此外，产地处理与在线感知研究表明，国内研究正在向系统化与在线化方向发展[38-39]。

1.4.1 果蔬单任务视觉分析研究

果蔬视觉分析早期主要依赖颜色阈值分割、纹理描述和几何参数计算，用于尺寸测量、成熟度识别和缺陷检测。随着深度学习方法的发展，卷积神经网络逐渐成为外观品质检测与等级分类的主流路线。例如，深度模型已能够对颜色、尺寸、纹理和损伤信息进行联合表征[12,32]，国内相关研究也已从规则驱动方法逐步转向数据驱动建模[35-36]。

在重量估计方面，现有研究通常通过图像分割、面积统计、几何参数提取或体积近似建立视觉特征与重量之间的映射关系。早期方法多依赖经验公式和机器学习回归器，并在无接触称量场景中验证可行性[13]。面向多对象无损估重的研究也报告

了可部署性结果[13]。后续研究引入深度回归框架，以提升单视角图像条件下的重量估计稳定性[14]。针对具体对象的重量预测研究也在持续增加，例如芒果和李子等果实[40-41]。在国内应用层面，视觉识别与智能称重设备融合的研究也已出现[42]。

在关键表型特征分析方面，研究重点包括轮廓提取、尺寸测量、曲率描述、体积估计和表型参数分析。相关综述表明，视觉方法已由简单测径逐步发展到结合分割、检测与几何建模的综合框架[15,43]。同时，也有研究利用颜色与三维几何特征进行联合测量，推动关键表型特征分析从二维向多源信息扩展[22,44]。国内在果实尺寸和品质指标自动测量方面也已有实证工作[45]。

综合分析来看，单任务研究已分别在重量估计、关键表型特征分析和质量分类上取得进展，但三类任务大多仍独立建模。由于这些任务共享同一图像中的边缘、轮廓、颜色和纹理信息，独立建模容易导致特征重复学习，且难以充分利用任务间关联，因此仍需要面向多属性协同的统一建模框架。

1.4.2 小样本与数据受限学习研究

果蔬智能检测研究普遍面临样本规模有限、标注成本较高以及类别分布不平衡等问题。与通用视觉任务相比，农业图像采集常受季节、品种、拍摄环境和标注条件限制，导致很多研究只能在小规模数据上进行训练和验证。因此，小样本学习、迁移学习、数据增强和预训练迁移等方法逐渐成为果蔬视觉研究中的一个重要方向。

现有研究表明，在样本不足条件下，小样本学习和迁移学习能够在一定程度上缓解模型对大规模标注数据的依赖。以成熟度判别为例，小样本策略在有限样本下可获得较好的适应性[29]。针对绿色果实检测等难样本场景，研究普遍结合数据增强和检测结构改进以提升性能[30,46]。围绕复杂背景和尺度变化问题，也有研究通过结构优化进一步改进小样本检测效果[47-48]。相关研究还从网络结构改进与训练策略层面给出补充证据[49-51]。

相关综述指出，果蔬质量检测仍普遍依赖小规模或单任务数据集[12,32]。面向农业视觉任务的数据受限问题，也有综述给出一致判断[52]。为补足训练数据，迁移学习、轻量网络和合成增强策略被广泛采用[53-54]。在模型效率优化方向，轻量网络搜索方法也被引入相关研究[55]。国内在线感知研究同样强调了高质量数据获取与标注成本对模型性能的制约[39]。

综合分析来看，现有小样本研究仍主要集中在检测、成熟度识别或病害识别等

单一任务。对于同时包含回归和分类目标的多属性任务，仅依靠传统增强或简单迁移学习往往不足以解决监督稀缺和任务耦合问题。特别是在重量和成熟状态等连续属性任务中，端点样本之间中间状态不足的问题仍未被充分解决。

1.4.3 多任务学习与特征增强研究

为了提高模型对复杂农业图像的表达能力，近年来研究开始更多关注特征增强和多任务学习方法。一类研究侧重于多尺度特征融合。特征金字塔网络通过自顶向下路径和横向连接融合不同层级语义信息，可增强模型对尺度变化目标的表征能力[56]。后续研究进一步补充了该范式在检测任务中的结构实现[56]。另一类研究强调注意力重加权机制，通过空间和通道响应选择提升关键区域表达[57-58]。这些方法为关键表型特征建模与多属性联合学习提供了可复用的特征基础。

另一条主线是多任务学习，即在共享输入下同时优化多个相关目标。经典框架通常采用硬参数共享以提高特征复用效率[59]。随后，研究者提出了 Cross-Stitch、不确定性加权、分支学习和专家路由等策略以缓解任务冲突[60-63]。在任务关系建模方面，MMoE 和 MTAN 等结构进一步增强了任务间信息选择能力[64-65]。在农业质量检测场景中，已有研究尝试用多任务框架联合完成新鲜度与类别相关任务，验证了多任务路线的可行性[66]。国内分选与产地处理研究也在强调多参数协同感知，但多属性联合建模仍处于发展阶段[38,67]。

综合分析来看，果蔬多任务研究虽然已从并列输出走向关系建模，但在农业场景中的方法深度仍有限。特别是在重量回归、关键表型特征回归和质量等级分类并存时，任务监督形式与语义层级差异显著，如何在共享表示与任务特异性之间取得平衡仍是核心问题。

1.4.4 果蔬数据集研究现状

数据集是支撑果蔬智能检测研究的基础。现有公开数据集大多围绕单一任务构建，其中较常见的是面向类别识别、成熟度分类或外观质量分级的数据集。以 Fruits-360 为代表的数据集在类别识别方面应用广泛，但标签主要集中于类别信息[20-21]。面向质量或成熟度任务的专用数据集也在增加，例如火龙果相关数据资源已支持成熟度和质量分级研究[28,68]。类似地，FruitQ 相关数据资源在外观质量分类任务上具有较高应用价值[23,69]。

在物理属性标注方面，已有部分数据开始关注重量、体积或内部品质指标。AppleV 数据集主要服务于苹果体积估计[22]。针对芒果、李子等对象的研究也提供了重量建模样本[13,40-41]。此外，SpectroFood 等数据集利用高光谱反射信息提供干物质等内部质量指标[26,70]。国内相关研究也已覆盖在线感知和多传感测量，但主要聚焦系统应用和单场景验证[39,45]。

综合分析来看，现有数据资源虽然在数量和任务类型上持续增长，但整体仍以单任务标注为主。受控采集、背景简单和标注维度有限等问题依然普遍存在[12,32]。面向实际工程部署的数据集也常出现采集条件受限问题[24]。在工程化数据采集场景中，样本分布偏窄和任务标签不完整的问题较常见[25,27,71]。无论在国际还是国内研究中，图像、重量、关键表型特征和质量等级严格对齐的公开数据组织形式仍较少，这直接限制了多属性协同建模研究的可复现性与可比性。

1.5 论文主要工作

针对农业果蔬图像中重量、关键表型特征和等级联合预测任务面临的小样本、任务差异明显以及中间样本不足等问题，本文围绕多任务联合建模与扩展学习两个层面开展研究。论文主要工作如下：

（1）在小样本条件下，提出基于多任务学习的果蔬多属性联合评估框架，并构建配套的 FruVegSet 多属性对齐数据集。方法层面，针对回归与分类任务协同困难问题，设计预分类路由、双分支特征提取、特征增强与跨分支交互机制，以及联合损失优化策略。数据层面，围绕黄瓜与香蕉两类对象建立图像、重量、关键表型特征和等级标签的一一对应关系，其中关键表型特征在黄瓜子集为曲直度，在香蕉子集为成熟度。基于该数据与实验协议，所提方法在综合任务表现上优于单任务模型与其他多任务基线。

（2）在上述多属性联合评估的多任务框架基础上，提出基于中间样本生成的多任务扩展学习方法。针对端点状态之间过渡样本缺失问题，在同类端点样本对之间利用 DiffMorpher 生成中间图像，并结合 SAM2 与主体掩膜筛选策略提取稳定的目标区域，进一步构建重量、关键表型特征与等级伪标签。通过将生成样本以可控比例注入训练过程，一定程度上增强了小样本条件下模型对连续变化过程的学习能力，并改善了关键表型特征回归与等级分类表现。

1.6 论文组织架构

本文围绕果蔬图像中重量、关键表型特征与质量等级的联合评估问题展开研究，研究路径分为多任务联合建模与中间样本扩展两条主线。论文整体技术路线如图 1.1所示。

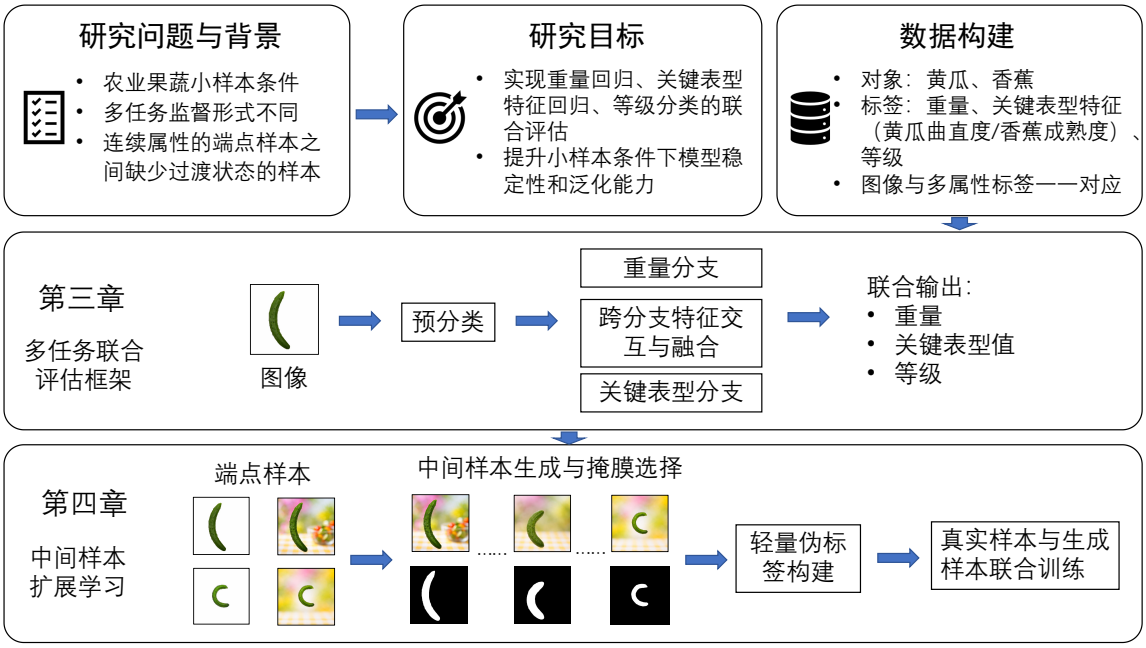


图 1.1 本文技术路线图

本文其余章节安排如下：

第二章介绍研究所需的相关理论与关键技术，包括卷积神经网络、特征增强与融合、多任务学习以及评价指标，为后续方法与实验提供理论基础。

第三章提出基于多任务学习的果蔬多属性联合评估框架。该章先介绍整体框架，由预分类路由、双分支特征提取、特征增强与跨分支交互模块组成，以及联合损失函数设计，随后给出 FruVegSet 数据集构建与标签组织方式，最后通过对比实验与消融实验，分析模型在重量、关键表型特征和等级任务上的联合学习表现。

第四章提出基于中间样本生成的多任务扩展学习方法。该章围绕端点样本配对、中间样本生成、主体掩膜筛选和分任务伪标签赋值展开，进一步给出真实样本与生成样本联合训练的优化方式，最后通过对比实验与消融实验，分析样本注入比例、主体掩模选择策略和伪标签构建流程对模型性能的影响。

第五章对全文研究工作进行总结，归纳主要结论，并结合现有不足提出后续研究方向。

第二章　相关理论和技术概述

深度学习视觉方法的研究重点已由单一结构性能提升转向系统化建模，即在统一框架内同时处理基础表示学习、特征增强与融合、联合优化以及结果判定标准四个层面。本章围绕卷积神经网络、特征增强和融合方法、多任务学习、评价指标展开，目的在于建立后续章节理论支撑链条，使方法设计、训练策略和实验结论具备统一的原理依据。

2.1　卷积神经网络

卷积神经网络（Convolutional Neural Network, CNN）解决的是高维输入如何映射为可判别中间表示的问题。该问题的难点不在于函数形式是否足够复杂，而在于模型是否能够在可训练条件下提取稳定特征。卷积结构通过局部连接和参数共享引入结构先验，显著降低参数冗余，避免在输入维度较高时出现不可控的参数膨胀。

局部连接假设意味着模型优先学习邻域模式，参数共享假设意味着同类模式在不同空间位置可由同一组参数识别。两种假设共同提高了统计效率，模型无需为每个位置学习独立参数，即可形成可迁移的局部特征检测器。该机制解释了卷积网络在视觉任务中长期稳定有效的原因。

CNN 的层级表示具备清晰的语义演化过程。浅层特征倾向于编码边缘、纹理和局部对比关系，中层特征逐步形成结构片段，高层特征则聚合为语义模式。该层级路径并非简单堆叠，而是把低层可观测信号转化为高层可分离表示，为后续模块提供输入。激活函数与下采样结构在这一过程中承担关键角色。激活函数引入非线性，使网络能够表达复杂映射关系；下采样压缩冗余并扩大有效感受域，使后续层具有更强上下文整合能力。若缺少这两类操作，卷积层即使加深也难以形成稳定的语义抽象。

普通深层卷积网络常出现优化退化，表现为层数增加后训练误差不降反升。该现象主要来自优化路径不稳定，而非表示能力不足。残差学习通过捷径连接重构信息流，其基本形式为

$$\boldsymbol{y} = F(\boldsymbol{x}) + \boldsymbol{x}, \tag{2.1}$$

其中 $\boldsymbol{x}$ 表示输入特征，$F(\cdot)$ 表示参数化变换分支，$\boldsymbol{y}$ 表示残差块输出。该公式的含义在于网络学习的对象不再是完整映射，而是相对输入的增量修正。其对应的训练价值在于保留路径与修正路径并存。若某层尚未学习到有效模式，恒等分支仍可保持信息连续传递；若某层学习到有效增量，输出将通过残差累积逐步改善表示质量。该机制直接提升了深层训练稳定性[72-73]。

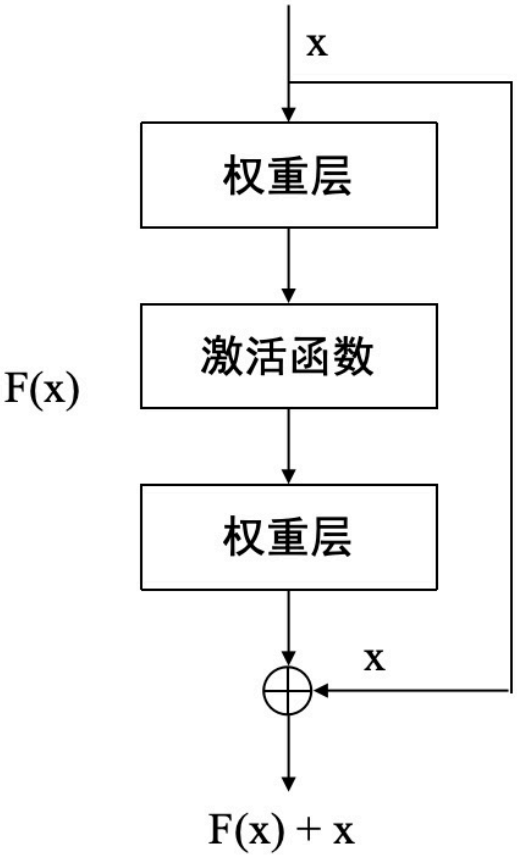


图 2.1　残差结构示意图

批归一化与初始化策略是残差机制稳定落地的配套条件。批归一化减少层间分布漂移，初始化策略抑制训练早期梯度波动，两者共同降低优化不确定性[74-75]。主干网络在这一组合下更容易形成平滑、可复用的中间表示。

卷积主干的理论边界同样明确。卷积机制擅长局部模式建模，但对远距离依赖的直接表达能力有限；参数共享提升效率，但也可能削弱位置特异信息。主干输出通常仍需后续结构做信息重分配与跨层整合，才能满足复杂目标的建模需求。另外，卷积网络性能并不只由深度决定，表示质量与训练可控性更关键。网络越深并不必然越优，若优化路径不稳定或特征分布波动过大，额外深度可能转化为训练噪声放

大器。

就结构而言，卷积主干负责提供基础表示，其作用是建立统一特征底座，而非直接完成全部预测目标。后续特征增强、交互与联合优化模块建立在该底座之上，主干质量将直接影响这些模块的有效性上限。

2.2 特征增强和融合方法

卷积特征存在天然层级差异，浅层特征细节丰富但语义弱，深层特征语义强但细节弱，单层特征难以同时满足多类判别需求。特征增强和融合方法针对这一矛盾提出，核心目标是提高中间表示的均衡性与可用性。

在机制层面，特征增强可划分为两条主线：一条是跨层尺度融合，解决信息来源不完整问题；另一条是动态权重分配，解决信息使用不充分问题。前者处理可用信息范围，后者处理信息优先级。两条主线共同构成现代视觉模型的中间表示优化框架。结构设计若忽略这两类问题，模型容易出现典型失衡：仅依赖浅层时语义判别不足，仅依赖深层时细节证据不足；仅做静态拼接时特征干扰明显，仅做单通路传播时信息覆盖不完整。增强与融合模块的引入，本质上是为了把表示能力从可提取推进到可利用。

2.2.1 特征金字塔

特征金字塔方法 (FPN) 针对层间语义断层提出。深层特征语义抽象能力高，浅层特征空间细节保持度高，若两者隔离使用，模型难以形成稳定统一表示。FPN 通过自顶向下路径与横向连接实现跨层融合，其标准形式为

$$\boldsymbol{P}_l = \mathrm{Conv}(\boldsymbol{C}_l + \mathrm{Up}(\boldsymbol{P}_{l+1})), \tag{2.2}$$

其中 $\boldsymbol{C}_l$ 表示第 l 层主干特征，$\boldsymbol{P}_{l+1}$ 表示上一级金字塔特征，$\mathrm{Up}(\cdot)$ 表示上采样，$\mathrm{Conv}(\cdot)$ 表示融合平滑，$\boldsymbol{P}_l$ 表示当前层融合输出[56]。

该公式包含两步逻辑。第一步是语义回流，高层语义通过上采样传递到低层尺度；第二步是细节补偿，低层细节与回流语义融合，得到语义强度更均衡的特征表示。公式中的加法操作反映信息对齐融合，卷积平滑用于抑制融合后的局部噪声。

特征金字塔的直接收益体现在三方面：跨层复用提升信息利用率，语义对齐降低层间表达割裂，融合输出增强后续模块的输入稳定性。对后续决策层而言，金字

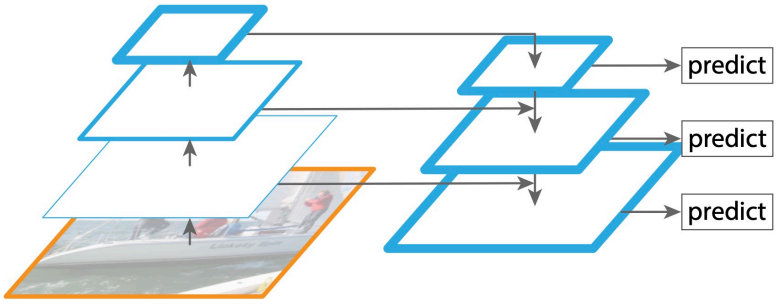


图 2.2 特征金字塔示意图[56]

塔输出通常比单层特征更具鲁棒性。

横向融合方式通常包括逐元素相加与通道拼接。逐元素相加结构简洁、数值稳定，通道拼接保留信息更完整。二者选择不属于固定优劣关系，关键在于后续模块对融合特征的消费方式与训练稳定性要求。

然而，金字塔机制并不意味着融合层越多越好。层间分布差异过大时，融合可能引入额外不一致噪声；若输入本身噪声显著，跨层传播也可能放大噪声链路。有效应用通常依赖分布对齐、归一化与稳定训练策略的配合。特征金字塔提升的是表示质量，而不是直接定义优化目标。它改善的是信息组织方式，不能替代目标函数设计与优化控制。若目标之间存在冲突，仅靠尺度融合无法自动消解冲突。

特征金字塔还具备较强可组合性。其输出能够与注意力、分支交互、多头预测结构直接衔接，形成清晰的中间接口层。这种接口属性使金字塔方法在复杂系统中具有较高工程复用价值。

2.2.2 注意力机制

注意力机制处理的是信息分配问题。卷积特征并不天然区分重要与次要成分，若所有通道和位置被同等处理，关键模式容易被冗余响应稀释。注意力通过可学习权重实现动态筛选，使模型容量优先服务于高价值信息。

通道注意力、空间注意力和跨源注意力对应三类不同问题。通道注意力处理语义维度权重分配，空间注意力处理位置维度权重分配，跨源注意力处理特征源之间的条件化关联。三者共同构成注意力机制的基本范式[76-77]。

交叉注意力用于显式建模不同特征源之间的依赖关系，标准形式为

$$\mathrm{Attn}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\top}}{\sqrt{d_k}}\right)\boldsymbol{V}, \tag{2.3}$$

其中 $\boldsymbol{Q}$ 为查询特征，$\boldsymbol{K}$ 为键特征，$\boldsymbol{V}$ 为值特征，d_k 为键向量维度。$\boldsymbol{Q}\boldsymbol{K}^\top$ 给出相关性评分，softmax 将评分归一化为权重，右乘 $\boldsymbol{V}$ 得到加权汇聚输出，如图 2.3所示。

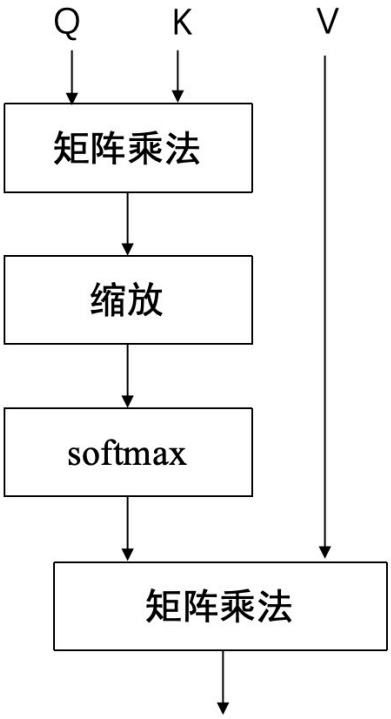


图 2.3　交叉注意力机制示意图[78]

公式中的每个部件对应明确功能：查询决定当前信息需求，键提供可匹配索引，值提供被汇聚内容，归一化权重决定信息注入比例。该过程把静态连接转化为输入条件驱动的动态交互。与直接拼接相比，交叉注意力能够表达按相关性进行选择性交互；与固定加权相比，权重随输入变化而变化，表达灵活性更高。该机制尤其适用于需要跨分支、跨层或跨模态建立对应关系的场景。

注意力机制与卷积机制能够互相补充，卷积提供稳定局部先验，注意力提供动态全局重分配。二者联合时，卷积负责基础模式编码，注意力负责信息重排与关联强化，系统整体表达能力通常更完整。同样，注意力机制提高的是表示利用效率，不直接替代目标定义与优化控制。若损失目标设置不合理，或训练协议不稳定，注意力模块难以单独保证性能提升。

2.3 多任务学习

多任务学习（Multi-Task Learning, MTL）关注多个目标在统一模型中的协同优化。核心命题不是任务数量增加，而是共享与分化如何同时成立。有效多任务系统需要在参数效率、目标一致性和训练稳定性之间取得平衡。

多任务目标函数的一般形式为

$$\mathcal{L}_{\text{mtl}} = \sum_{t=1}^{T} \lambda_t \mathcal{L}_t, \tag{2.4}$$

其中 T 为任务数量，$\mathcal{L}_t$ 为第 t 个任务损失，λ_t 为任务权重。该式揭示联合训练的本质是多目标加权优化，任务关系会通过权重和梯度共同作用于共享参数更新。

共享机制通常分为硬共享与软共享。硬共享直接复用主干参数，参数效率高，结构简单；软共享通过分支交互或约束机制控制共享强度，冲突可控性更好。两者差异不在于是否共享，而在于共享颗粒度与信息流路径。硬共享的风险在于目标差异较大时梯度冲突更明显，软共享的风险在于结构复杂度提高带来的训练不稳定。机制选择应依据目标相关性和可控性需求，而非固定偏好。

联合优化中的常见问题是损失尺度失衡。某些任务损失数值天然较大，若直接加权求和，更新过程会被单一任务主导，其他任务学习被压制。固定权重方案实现简单，但超参数敏感，且对训练阶段变化适应性较弱。

基于不确定性的权重学习把权重参数化并纳入训练，其形式为[61]

$$\mathcal{L}_{\text{uw}} = \sum_{t=1}^{T} \left(\frac{1}{2\sigma_t^2} \mathcal{L}_t + \log \sigma_t \right), \tag{2.5}$$

其中 σ_t 为第 t 个任务的不确定性参数。第一项提供自适应缩放，第二项约束权重避免无界扩张。该机制将静态调权转为动态学习，降低人工调参与阶段不匹配问题。不确定性加权不是冲突消除器。若目标方向本身矛盾，权重学习只能缓和冲突强度，无法消除冲突来源。此时仍需结合分支解耦、受控交互或阶段化训练策略，形成结构与优化协同。

多任务系统评价应强调整体平衡而非单项峰值。某一任务显著提升但其他任务明显退化时，系统收益并不成立。联合优化分析应报告多目标表现、波动范围与平衡关系，而非只比较单一指标极值。

2.4 评价指标

评价体系的作用是把模型行为转化为可比较、可解释、可复核的定量结论。完整评价包含三层信息：精度水平、稳定程度与差异可信度。回归指标描述连续输出误差，分类指标描述离散判别质量，显著性检验判断比较结论是否可靠。

回归评价关注预测值与真实值之间的偏差关系。决定系数 R^2 与均方根误差 RMSE 形成互补指标组：R^2 反映趋势解释能力，RMSE 反映绝对误差量级。

决定系数定义为

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{2.6}$$

其中 y_i 为第 i 个样本真实值，$\hat{y}_i$ 为预测值，$\bar{y}$ 为真实值均值，n 为样本数。分子是残差平方和，分母是总离差平方和。该指标衡量模型对真实波动的解释程度。

RMSE 定义为

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{2.7}$$

其中符号含义与上式一致。RMSE 与目标变量同量纲，便于直接解释平均误差规模；平方项使其对大偏差更敏感，能够更明显反映异常误差样本影响。

两类指标联合使用可避免解释偏差。单独使用 R^2 可能忽略绝对误差，单独使用 RMSE 可能忽略趋势拟合能力。数据分布会影响指标解释。目标值方差较小时，R^2 对微小偏差敏感；目标值范围较大时，RMSE 容易受少量大偏差样本牵引。

2.4.1 分类任务指标

分类评价关注离散标签预测正确性与类别均衡性。准确率提供总体正确比例，宏平均指标提供类别层面的平衡视角。

准确率定义为

$$\text{Acc} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(\hat{y}_i = y_i), \tag{2.8}$$

其中 $\hat{y}_i$ 为预测标签，y_i 为真实标签，$\mathbb{I}(\cdot)$ 为示性函数，n 为样本数。准确率直观，但在类别不平衡时可能高估模型能力。

第 k 类精确率、召回率与 F_1 定义为

$$\text{Prec}_k = \frac{TP_k}{TP_k + FP_k}, \tag{2.9}$$

$$\text{Recall}_k = \frac{TP_k}{TP_k + FN_k}, \tag{2.10}$$

$$F_{1,k} = \frac{2\,\text{Prec}_k\,\text{Recall}_k}{\text{Prec}_k + \text{Recall}_k}, \tag{2.11}$$

其中 TP_k、FP_k、FN_k 分别表示第 k 类真阳性、假阳性与假阴性数量。精确率对应预测可靠性，召回率对应检出能力，F_1 提供二者平衡刻画。

宏平均定义为

$$\text{Prec}_{\text{macro}} = \frac{1}{K}\sum_{k=1}^{K}\text{Prec}_k, \tag{2.12}$$

$$F_{1,\text{macro}} = \frac{1}{K}\sum_{k=1}^{K}F_{1,k}, \tag{2.13}$$

其中 K 为类别总数。宏平均为各类别赋予等权重，能够降低类别不平衡对总体结论的掩盖效应。多类问题下，准确率与宏平均指标应联合报告。前者反映总体正确率，后者反映类别均衡表现。两者结合可形成更完整、更公平的分类评价结论。

2.4.2 显著性检验

方法比较若仅依据均值差异，结论易受随机波动影响。显著性检验用于判定差异是否具有统计可靠性，其作用是为比较结论提供概率意义上的支持。

McNemar 检验[79]用于同一测试集样本级配对分类比较，统计量为

$$\chi^2 = \frac{(b-c)^2}{b+c}, \tag{2.14}$$

其中 b 表示模型 A 误判且模型 B 正判的样本数，c 表示模型 A 正判且模型 B 误判的样本数。该检验关注错误模式差异，而非仅比较总体准确率。

Wilcoxon 符号秩检验[80]用于重复实验形成的配对指标序列。该方法不要求差值服从正态分布，对小样本重复实验具有较高适用性，其统计思想是比较差值秩分布而非原值大小，因而对异常值更稳健。在本文中，该检验用于比较多次重复实验得到的分类准确率序列。

比较结果报告包含均值、标准差与显著性结论。均值描述中心水平，标准差描述波动范围，显著性检验描述差异可靠度，三类信息共同构成完整证据链。显著性结果仍需结合差异幅度解释。统计显著不自动等价于实际意义显著。若增益量级很小，方法价值可能有限；若增益显著且稳定，结论可信度更高。

2.5 本章小结

本章介绍了本论文使用到的相关理论和技术。首先介绍了卷积神经网络的基本原理与残差学习机制，然后介绍了特征增强和融合方法中的特征金字塔与注意力机制，随后介绍了多任务学习中的共享机制与联合优化思想，最后介绍了回归任务与分类任务的评价指标定义以及显著性分析，为后续章节的方法构建与实验分析提供理论基础。

第三章　基于多任务学习的果蔬多属性联合评估框架

针对不同农产品判别依据不一致、多任务特征需求存在差异、以及小规模数据条件下任务相互干扰的问题，本章提出一种基于多任务学习的果蔬多属性联合评估框架，聚焦果蔬多属性联合评估任务，目标是在统一框架下同时完成重量回归、关键表型特征回归与质量等级分类。该框架先通过预分类模块完成类别路由，再由重量分支与关键表型分支分别学习对应表征，并结合特征融合与交叉注意力机制实现联合预测，用于黄瓜与香蕉两类样本的协同建模与综合评估。本章先介绍模型总体设计与训练目标，再说明数据集构建与标签组织过程，最后给出对比实验与消融分析。

3.1　方法概述

3.1.1　框架概述

本节先从整体结构出发说明多任务框架的组成与信息流，再逐步给出各模块的计算过程和损失定义。所提出的框架采用分而治之的策略，提出分层结构，首先进行水果和蔬菜类型识别，然后应用针对该类型定制的多任务预测模型。如图 3.1所示，预分类模块首先分析输入图像以确定农产品类别，例如黄瓜或香蕉。基于此结果，框架动态地将图像路由到相应的特定任务子网络。

每个子网络旨在同时执行三项任务：重量估计、关键表型特征回归和品质等级分类。这些任务通过子网络内的专用分支实现，每个分支接收特定类型的视觉特征。在黄瓜和香蕉模型中，输入图像首先分别进行重量 f_{w} 和关键表型特征 f_{t} 提取，例如黄瓜的曲直度和香蕉的成熟度。然后，利用这些特征预测重量和相应的关键表型指标。最后，将特征与中间表示融合，生成品质等级分类得分，最终预测品质等级。具体过程在后文描述。这种模块化和类型感知的设计使框架能够适应不同类别的产品，同时在每种类型中保持多任务学习的优势。

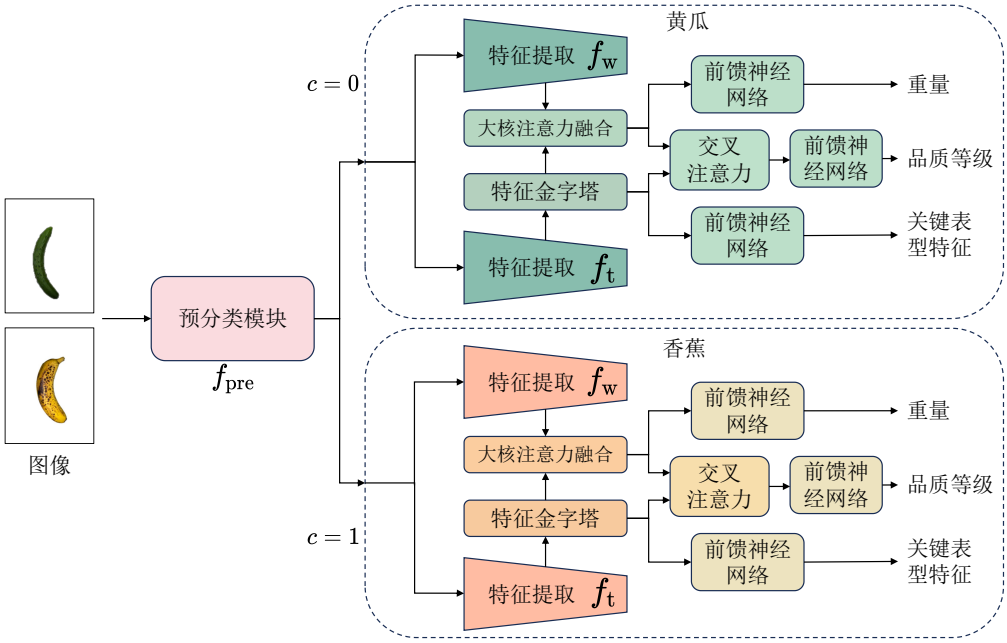


图 3.1 本章方法整体框架

3.1.2 预分类模块

预分类模块作为一种路由机制，用于确定输入图像的农产品类别，并将其导向专门的多任务预测子网络。这种分层结构使得整个框架能够适应不同类型的水果和蔬菜，同时保持对每种水果和蔬菜的专门处理。输入图像 $\boldsymbol{I} \in \mathbb{R}^{H\times W\times 3}$ 首先由预分类网络 $f_{\text{pre}}(\cdot)$ 进行处理，该网络输出预测的类别标签 c。基于预测的类别 $c \in \{0,1\}$，其中 $c=0$ 表示黄瓜，$c=1$ 表示香蕉，从一组特定任务的子网络 $\{f_0, f_1\}$ 中选择相应的专家模型 $f_c(\cdot)$。最终预测向量为：

$$\boldsymbol{Y} = f_c(\boldsymbol{I}), \quad \boldsymbol{Y} = [\hat{w}, \hat{t}, \hat{\boldsymbol{g}}]. \tag{3.1}$$

其中，$\hat{w}$ 为预测的重量标量，$\hat{t}$ 为预测的关键表型相关标量，也就是曲直度或成熟度，$\hat{\boldsymbol{g}} \in \mathbb{R}^4$ 表示品质的分类 logits。这些预测的详细公式在后文描述。

本章采用 ResNet18[72] 作为预分类模块的架构，因为它在效率和准确率之间取得了良好的平衡。首先在 ImageNet [81] 上进行预训练，然后在 FVS 的混合部分上进行微调。在当前数据划分下，该模块在预留测试集上达到 100% 分类准确率，说明其在本实验设置中能够满足路由需求。

引入可学习的预分类模块是为了从结构角度支持统一且可扩展的系统。虽然简单的像素级规则可以区分黄色香蕉和绿色黄瓜，但在一些模糊情况下，例如未成熟

的绿色香蕉与黄瓜颜色特征相似时，这些规则可能会失效。可学习的分类器利用纹理和轮廓线索来解决此类区别，而固定规则则依赖有限的视觉信号。

此外，黄瓜和香蕉的关键表型目标遵循不同的规则，因为曲直度描述的是几何形状，而成熟度描述的是表面斑点。这些目标无法在共享空间中同时学习而不相互干扰。预分类器充当路由步骤，将样本发送到正确的分支，从而防止负迁移，此设计旨在支持本章范围内各种采集条件下的可靠路由。

为了增强模块化并减少潜在的任务干扰，预分类模块与下游子网络分开训练。然后，使用预分类器分配的样本对每个特定任务的模型进行优化，从而实现针对每种农产品类别的定向监督和结构设计。这种设计还确保每个子网络中的关键表型回归模块专注于相关特征，避免不同类型之间的混淆，例如，黄瓜关注曲直度，香蕉关注成熟度。

3.1.3 多任务子网络结构

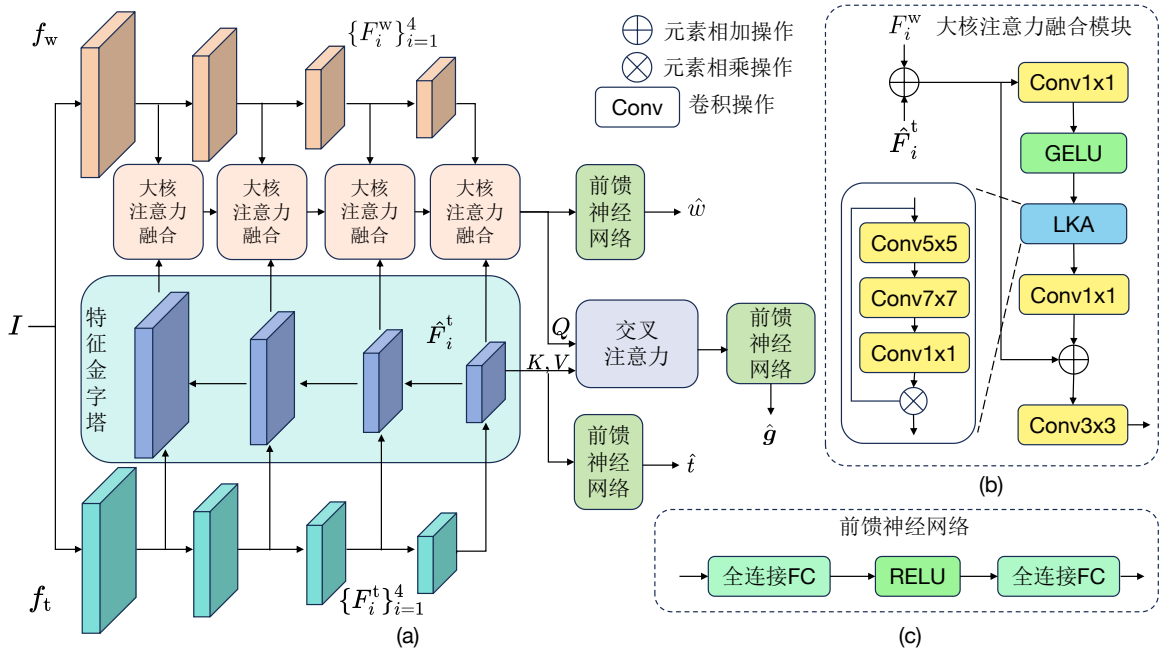


图 3.2　多任务子网络结构

多任务子网络采用并行双分支设计与特征融合策略进行多任务预测，如图 3.2所示。经过预分类模块路由后，每个输入图像 $\boldsymbol{I}$ 被送入一个特定任务的子网络，该子网络包含两个专门的 CNN 骨干网络，也就是重量骨干网络 $f_{\mathrm{w}}(\cdot)$ 和关键表型骨干网络 $f_{\mathrm{t}}(\cdot)$。这些骨干网络从四个阶段提取多尺度特征图，分别表示为 $\{\boldsymbol{F}_i^{\mathrm{w}}\}_{i=1}^4$ 和 $\{\boldsymbol{F}_i^{\mathrm{t}}\}_{i=1}^4$，

其中 $\boldsymbol{F}_i^{\mathrm{w}}, \boldsymbol{F}_i^{\mathrm{t}} \in \mathbb{R}^{H_i \times W_i \times C_i}$ 分别表示重量骨干网络和关键表型骨干网络第 i 阶段的中间特征。

这种双主干设计源于任务截然不同的视觉特性。重量的估计主要依赖于全局几何属性，而关键表型回归则依赖于精细的局部边缘或纹理。共享同一主干会导致负迁移，即一个任务会干扰另一个任务的特征提取。正如本章消融实验中验证的那样，使用两个不共享的主干可以让每个分支专注于与其目标最相关的特征，同时仍然可以通过融合模块实现交互。

为了捕捉多尺度语义线索，关键表型相关特征通过自顶向下的特征金字塔网络结构进行细化。第 i 层上采样和合并的特征计算如下：

$$\hat{\boldsymbol{F}}_i^{\mathrm{t}} = \boldsymbol{F}_i^{\mathrm{t}} + \mathrm{Upsample}(\hat{\boldsymbol{F}}_{i+1}^{\mathrm{t}}), \quad i = 3, 2, 1; \quad \hat{\boldsymbol{F}}_4^{\mathrm{t}} = \boldsymbol{F}_4^{\mathrm{t}} \tag{3.2}$$

其中，Upsample 表示上采样操作。这里的 $\boldsymbol{F}_i^{\mathrm{t}}$ 表示关键表型骨干网络在第 i 层输出的特征图，$\hat{\boldsymbol{F}}_i^{\mathrm{t}}$ 表示经由自顶向下路径融合后的关键表型特征图。符号 i 为层级索引，取值为 $1, 2, 3, 4$，其中 $i = 4$ 对应语义层级更高的深层特征。

为了融合重量与关键表型信息，在每个层级 i 上应用由大核注意力 (LKA)[58] 调整的大核注意力融合模块 (LKAF)：

$$\boldsymbol{S}_i = \boldsymbol{F}_i^{\mathrm{w}} + \hat{\boldsymbol{F}}_i^{\mathrm{t}} + \boldsymbol{\delta}_i, \qquad \boldsymbol{\delta}_i = \begin{cases} 0, & i = 1, \\ \mathrm{Upsample}\left(\tilde{\boldsymbol{F}}_{i-1}\right), & i > 1, \end{cases} \tag{3.3}$$

$$\boldsymbol{U}_i = \mathrm{GELU}\big(\mathrm{Conv}_{1\times1}(\boldsymbol{S}_i)\big), \tag{3.4}$$

$$\bar{\boldsymbol{F}}_i = \boldsymbol{S}_i \ + \ \mathrm{Conv}_{1\times1}\big(\mathrm{LKA}(\boldsymbol{U}_i)\big), \tag{3.5}$$

$$\tilde{\boldsymbol{F}}_i = \mathrm{Conv}_{3\times3}(\bar{\boldsymbol{F}}_i). \tag{3.6}$$

其中，$\mathrm{Conv}_{1\times1}(\cdot)$ 表示卷积核大小为 1×1 的卷积操作，主要用于通道映射与特征重组；$\mathrm{Conv}_{3\times3}(\cdot)$ 表示卷积核大小为 3×3 的卷积操作，用于在局部邻域内进一步整合空间上下文。$\mathrm{GELU}(\cdot)$ 表示高斯误差线性单元激活函数，用于引入非线性表达能力。$\mathrm{LKA}(\cdot)$ 表示大核注意力操作，用于增强与目标相关的空间响应区域。

最深层的融合特征 $\tilde{\boldsymbol{F}}_4$ 经过池化处理后，被送入前馈网络以预测重量：

$$\boldsymbol{z}_{\mathrm{w}} = \mathrm{GAP}(\tilde{\boldsymbol{F}}_4), \quad \hat{w} = \mathrm{FFN}_{\mathrm{w}}(\boldsymbol{z}_{\mathrm{w}}) \tag{3.7}$$

其中，GAP$(\cdot)$ 表示全局平均池化操作。对于输入特征图 $\boldsymbol{X} \in \mathbb{R}^{H\times W\times C}$，其输出向量定义为

$$\mathrm{GAP}(\boldsymbol{X})_c = \frac{1}{HW}\sum_{u=1}^{H}\sum_{v=1}^{W}\boldsymbol{X}_{u,v,c}, \quad c = 1,\ldots,C. \tag{3.8}$$

输出 $\boldsymbol{z}_{\mathrm{w}}$ 为重量分支的全局特征向量，$\hat{w}$ 为预测重量。

同理，最上层的关键表型特征 $\hat{\boldsymbol{F}}_4^{\mathrm{t}}$ 经过池化处理后，被送入前馈网络来预测所表示的关键表型值：

$$\boldsymbol{z}_{\mathrm{t}} = \mathrm{GAP}(\hat{\boldsymbol{F}}_4^{\mathrm{t}}), \quad \hat{t} = \mathrm{FFN}_{\mathrm{t}}(\boldsymbol{z}_{\mathrm{t}}) \tag{3.9}$$

其中，$\hat{t}$ 表示预测的关键表型值，对应于曲直度或成熟度，具体取决于对应的类别。

本章三处前馈网络均采用两层全连接结构，即 FC、ReLU、FC，其形式写为

$$\mathrm{FFN}(\boldsymbol{z}) = \mathrm{FC}_2(\mathrm{ReLU}(\mathrm{FC}_1(\boldsymbol{z}))), \tag{3.10}$$

其中，$\mathrm{FC}_1(\cdot)$ 表示第一层全连接映射，$\mathrm{FC}_2(\cdot)$ 表示第二层全连接映射。该结构在重量回归头、关键表型回归头和等级分类头中保持一致。当用于重量与关键表型回归时输出为标量，当用于等级分类时输出为四维 logits。

为了对果蔬的质量品级进行分类，本章应用了交叉注意力机制，其中查询 $\boldsymbol{Q}$ 由 $\tilde{\boldsymbol{F}}_4$ 生成，键 $\boldsymbol{K}$ 和值 $\boldsymbol{V}$ 均由 $\hat{\boldsymbol{F}}_4^{\mathrm{t}}$ 导出：

$$\begin{aligned} \boldsymbol{Q} &= \mathrm{Conv}_{1\times1}(\tilde{\boldsymbol{F}}_4), \quad \boldsymbol{K} = \boldsymbol{V} = \hat{\boldsymbol{F}}_4^{\mathrm{t}} \\ \boldsymbol{F}_{\mathrm{att}} &= \mathrm{CrossAttn}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) \end{aligned} \tag{3.11}$$

其中，CrossAttn$(\cdot)$ 表示交叉注意力操作，$\boldsymbol{Q}$ 为查询特征，$\boldsymbol{K}$ 为键特征，$\boldsymbol{V}$ 为值特征。

交叉注意力的结果经过投影和全局平均池化处理后，被送入前馈网络来预测质量品级：

$$\boldsymbol{z}_{\mathrm{l}} = \mathrm{GAP}(\mathrm{Conv}_{1\times1}(\boldsymbol{F}_{\mathrm{att}})), \quad \hat{\boldsymbol{g}} = \mathrm{FFN}_{\mathrm{cls}}(\boldsymbol{z}_{\mathrm{l}}) \tag{3.12}$$

其中 $\hat{\boldsymbol{g}} \in \mathbb{R}^4$ 是果蔬品级分类的 logit 向量。

最终，每个子网络能够输出三个预测结果：重量值 $\hat{w} \in \mathbb{R}$、关键表型值 $\hat{t} \in \mathbb{R}$ 和品级分类的 logit 向量 $\hat{\boldsymbol{g}} \in \mathbb{R}^4$。这些预测对应于式 (3.1) 中描述的向量 $\boldsymbol{Y}$ 的分量。

3.1.4 损失函数

本章模型需要同时完成重量回归、关键表型特征回归和品质等级分类三个任务。其中，关键表型特征在黄瓜子集中对应曲直度，在香蕉子集中对应成熟度。为实现三项任务的联合优化，本章将总损失定义为两个回归损失与一个分类损失的组合。考虑到不同任务在目标形式、数值范围和学习难度上存在差异，若直接对各损失项进行固定权重求和，容易导致某一任务在训练过程中占据主导，从而影响整体优化效果。为此，本章采用基于任务不确定性的加权策略，对多任务目标进行自适应平衡。

设 $\hat{w}$ 和 w 分别表示预测重量与真实重量，$\hat{t}$ 和 t 分别表示预测关键表型值与真实关键表型值，$\hat{\boldsymbol{g}}$ 表示质量等级分类的预测 logits，$\boldsymbol{g}$ 表示对应的真实类别标签。本章首先定义重量回归损失、关键表型特征回归损失以及等级分类损失为

$$\mathcal{L}_{\text{weight}} = \text{Huber}(\hat{w}, w), \qquad \mathcal{L}_{\text{trait}} = \text{Huber}(\hat{t}, t), \qquad \mathcal{L}_{\text{grade}} = \text{CE}(\hat{\boldsymbol{g}}, \boldsymbol{g}), \tag{3.13}$$

其中，Huber$(\cdot)$ 表示 Huber 损失，CE$(\cdot)$ 表示交叉熵损失。相较于均方误差，Huber 损失在误差较小时保持平滑优化特性，在误差较大时对异常值不那么敏感，因此更适合本章中存在一定标注波动和样本规模有限的回归任务。

在此基础上，本章采用不确定性加权的方式构建总损失函数：

$$\mathcal{L}_{\text{total}} = \frac{1}{2\sigma_1^2}\mathcal{L}_{\text{weight}} + \frac{1}{2\sigma_2^2}\mathcal{L}_{\text{trait}} + \log\sigma_1 + \log\sigma_2 + \gamma\mathcal{L}_{\text{grade}}, \tag{3.14}$$

其中，σ_1 和 σ_2 为可学习的任务不确定性参数，分别对应重量回归任务和关键表型特征回归任务，γ 为分类损失的权重系数，本章中设为 1。

上述设计的核心思想在于通过引入可学习的不确定性参数，模型可以在训练过程中自动调整不同回归任务的贡献大小，从而减弱因任务量纲差异和学习难度差异所带来的优化失衡问题。同时，分类任务仍通过交叉熵损失对等级预测进行直接约束，使模型能够在连续关键表型估计与离散等级判别之间建立稳定的联合学习过程。该损失设计与本章的双分支多任务框架相匹配，有助于提高整体训练稳定性与多任务协同优化效果。

3.2 数据集构建

在完成方法设计说明后，本节介绍用于训练与评估的数据集来源、采集流程和标签构建方式。为了支持重量估计、关键表型特征分析和质量分类的多任务学习，本章构建了一个名为 FruVegSet (FVS) 的自定义数据集。该数据集包含黄瓜和香蕉的图像和物理属性对，分别代表两种典型的水果和蔬菜类别。下文将详细介绍黄瓜与香蕉两类样本的采集和处理流程。

3.2.1 黄瓜数据构建

为确保黄瓜成像条件的一致性，本章搭建了一个尺寸为 60 厘米 ×40 厘米的定制成像平台。一台 4K 摄像机固定在平台上方 30 厘米的高度，如图 3.3所示。图像采集在玻璃温室中进行，温室可提供稳定的自然光照并减少阴影。每个黄瓜样本都放置在白色背景的墙纸上，以增强对比度。

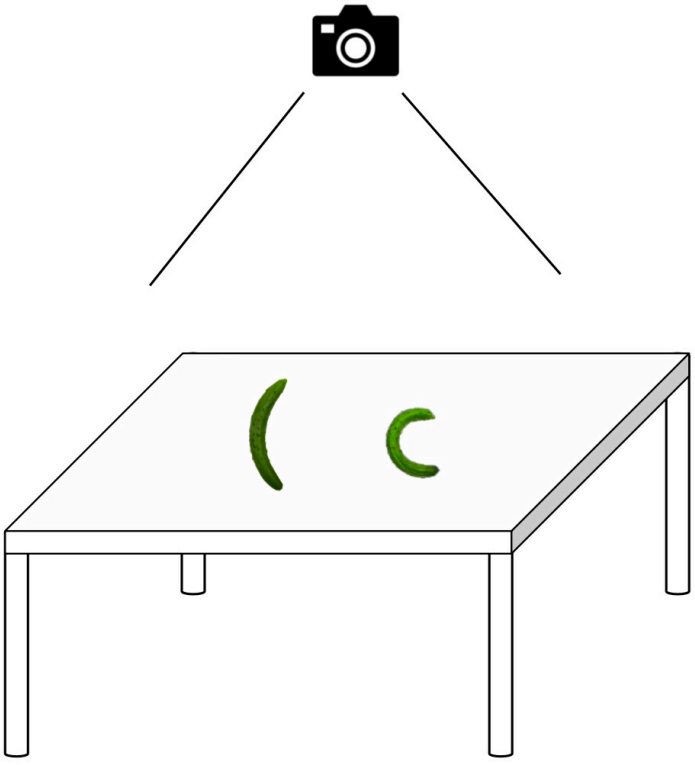


图 3.3　采集示意图

采集得到的黄瓜原始图像如图 3.4左侧所示。图像采集后，将每根黄瓜从其所属的组图像中分割出来，生成单独的分割黄瓜图像，用于进一步分析。

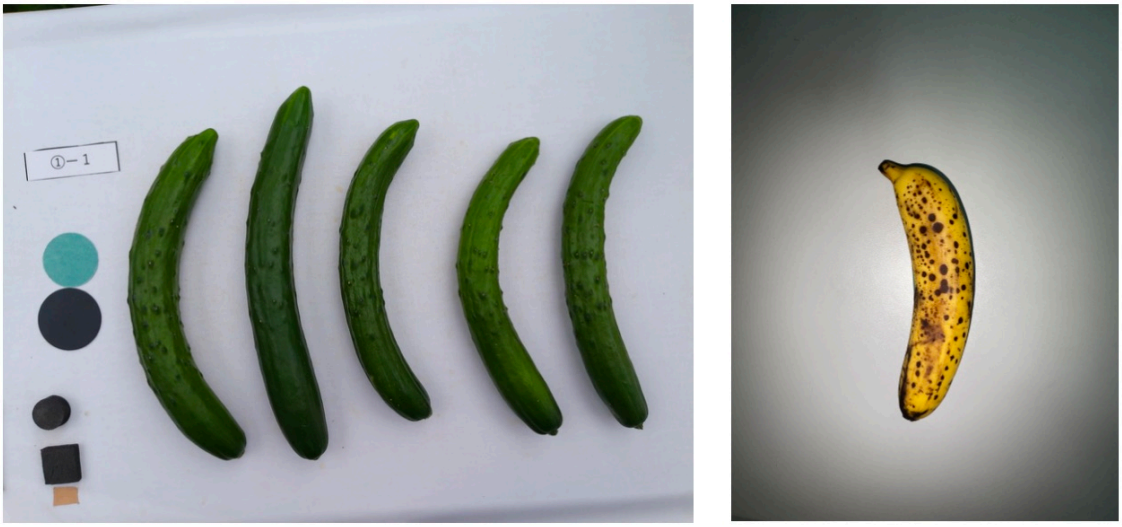


图 3.4　采集的原始图像示例

对于每一根黄瓜，本章使用电子秤测量其鲜重，并且还基于端点之间最大的线性距离 d，和从该线段到外曲线顶点的最大垂直距离 h 这两个像素测量值来计算黄瓜的曲直度 C。然后，曲直度指标定义为 $C = d/h$，如图 3.5所示。最终得到的黄瓜数据共有 220 张黄瓜图像。

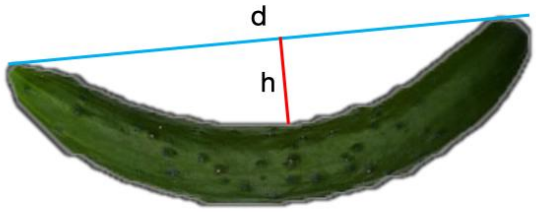


图 3.5　黄瓜曲直度计算示意图

表 3.1　黄瓜样本标签示例

样本	重量 (g)	关键表型特征 (曲直度)	等级
(a)	83.0	38.897	特级
(b)	148.0	13.903	一级
(c)	131.0	6.079	二级
(d)	30.0	2.901	三级

本章还分析了视觉外观与物理属性之间的关系，以验证数据集的设计。图 3.6和

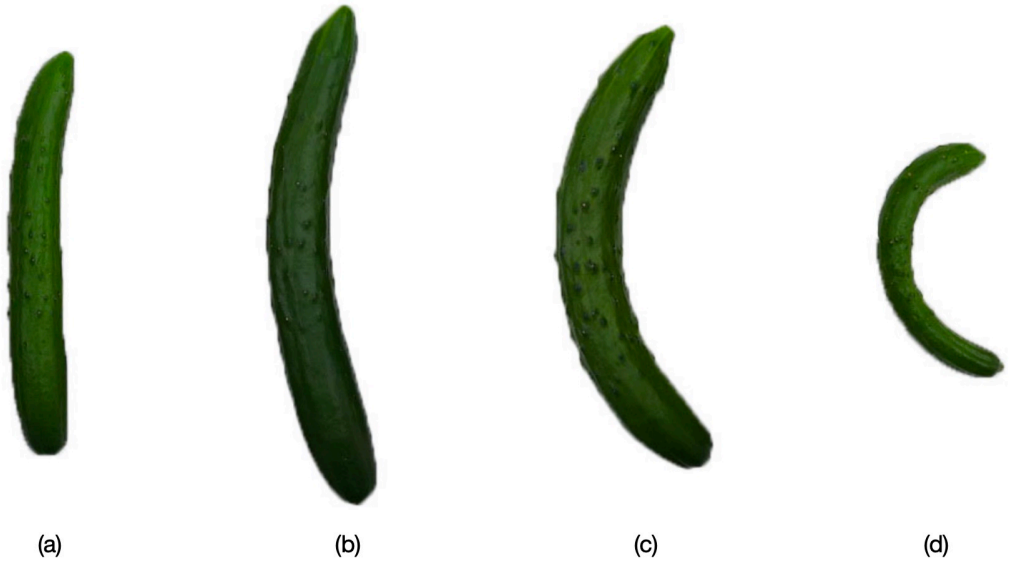


图 3.6　不同等级黄瓜样本示例

表 3.1展示了数据集中的四个代表性黄瓜样本。其中，图 3.6 中子图 (a) 至 (d) 依次对应特级、一级、二级和三级。特级样本几乎是直的，其曲直度值最高；而三级样本弯曲最为明显，曲直度值最低。上述样例主要用于说明标签与外观之间的对应关系，不能单独作为总体结论依据；关于曲直度与重量关系及其区分能力的判断，仍需结合后续统计分布结果进一步分析。

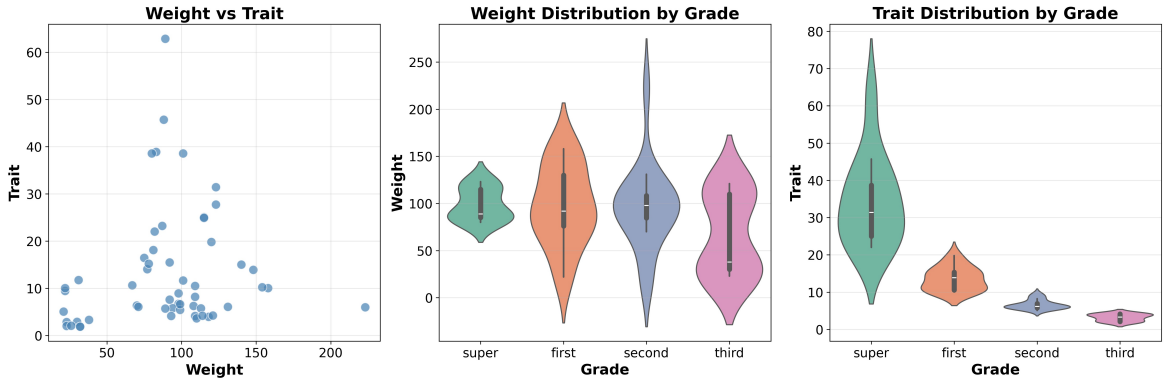


图 3.7　黄瓜重量、关键表型特征（曲直度）与等级关系

本章在图 3.7中进一步总结了这些关系。左侧散点图显示，重量相近的黄瓜的曲直度差异很大。重量在 100 克左右的样本，其曲直度值可能从非常小到超过 30，表明这两个属性之间没有较高的相关性。中间和右侧分布图显示，虽然不同等级的黄瓜重量范围重叠度很高，但曲直度从特级到三级单调递减，且各等级之间界限分明。这提示在本数据集中，曲直度相较重量具有更高的等级区分潜力。基于这一统计现

象，本章将重量与曲直度作为两个相互补充的回归目标进行联合建模。为保证等级标签定义的一致性，本章参考《黄瓜等级规格》[82]并结合本章数据分布进行扩展，将曲直度阈值 20、10 和 5 作为四级划分边界。具体地，$C > 20$ 对应特级，$10 < C \leq 20$ 对应一级，$5 < C \leq 10$ 对应二级，$C \leq 5$ 对应三级。

3.2.2 香蕉数据构建

本章在实验室环境下独立采集了香蕉的样本。虽然实验室的光照条件相对稳定，但为了减少天花板灯光的阴影，使用手机的手电筒功能，并以适当角度照射。摄像头位于样本上方约 40 厘米处。每个香蕉都放置在纯白色背景上，以确保能够清晰地分离出物体，如图 3.4右侧所示。同样地，对每张原始香蕉的采集图像进行分割，提取出香蕉区域，最终得到 159 张处理后的香蕉图像。

对于每一根香蕉，本章用电子秤记录其鲜重，并且还计算每根香蕉的成熟度指标 m。首先，将 159 张图像转换为灰度图像，然后应用阈值分割识别深色斑点。接着，统计斑点像素个数 N_{b} 和香蕉总像素个数 N_{a}，并将成熟度计算为 $m = N_{\mathrm{b}}/N_{\mathrm{a}}$。此过程如图 3.8所示。

图 3.8　香蕉成熟度计算示例

表 3.2　香蕉样本标签示例

样本	重量 (g)	关键表型特征 (成熟度)	等级
(a)	175.9	0.0112	特级
(b)	120.3	0.0819	一级
(c)	114.9	0.1497	二级
(d)	118.6	0.3208	三级

图 3.9　不同等级香蕉样本示例

图 3.9展示了四个香蕉样品，它们的表面斑点存在明显差异。其中，图 3.9 中子图 (a) 至 (d) 依次对应特级、一级、二级和三级。特级样品表面光滑呈黄色，几乎没有深色区域。随着等级降低，深色斑点的数量和密度均有所增加。三级样品的斑点最多，且斑点聚集面积最大。如表 3.2所示，成熟度从特级样品的 0.0112 增加到三级样品的 0.3208，而重量则没有呈现一致的变化趋势。这些样例仅用于展示可观察到的变化趋势，尚不足以直接推出总体规律；成熟度与等级关系的总体判断需结合后续统计分布结果给出。

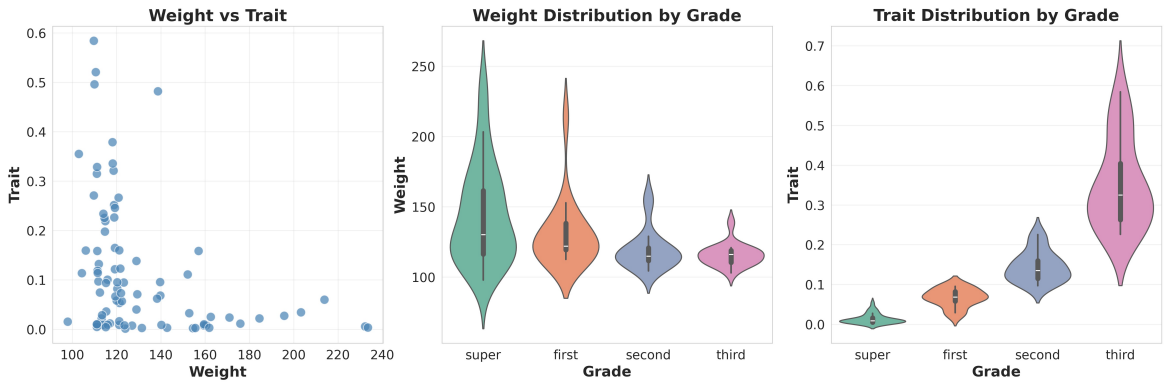


图 3.10　香蕉重量、关键表型特征（成熟度）与等级关系

图 3.10进一步展示了这些统计数据。左侧散点图显示，重量相近的香蕉成熟度差异显著。这意味着重量和成熟度分别代表了香蕉物理状态的不同方面，应该独立预测。中间和右侧分布图显示，虽然不同等级的香蕉重量重叠度较高，但成熟度却能为质量评估提供清晰的单调信号。这些统计结果支持将重量和成熟度作为两个独立但互补的回归目标。本章参考《香蕉等级规格》[83]并结合本章数据分布进行扩展，

将成熟度阈值 0.025、0.1 和 0.2 作为四级划分边界。具体地，$m \leq 0.025$ 对应特级，$0.025 < m \leq 0.1$ 对应一级，$0.1 < m \leq 0.2$ 对应二级，$m > 0.2$ 对应三级。

值得注意的是，数据采集环境代表了不同的农业生产阶段。黄瓜图像是在温室自然光照下采集的，代表现场评估；香蕉图像是在实验室人工光照下采集的，代表标准化分级。这些设置导致背景和光照质量存在差异，因为自然光照会变化，而人工光照则相对稳定。尽管如此，处理流程仍然保持一致。具体而言，分割步骤去除背景，网络提取特征以应对不同的表面光照。这些设置主要用于提升数据采集与预处理的一致性，为后续模型比较提供可比条件。

考虑到原始香蕉测试集样本数量有限，为了解决数据规模可能存在的问题，本章进一步扩展了香蕉测试集，以获得更可靠的评估结果。本章额外采集并拍摄了 70 张带有相应标签的香蕉分割图像。这些图像被添加到现有的测试子集中，最终得到 100 张香蕉测试图像。此次扩展提供了更广泛的外观变化范围，并有助于进行更稳定的评估。由于黄瓜的外形的曲直与黄瓜的生长周期和季节性供应有关，使得在短时间内收集并扩展数据变得十分困难，因此没有额外对黄瓜测试数据作出扩展。

由于数据集规模有限，本章特意采用手工设计的关键表型指标来表征曲直度和成熟度，而非端到端学习的特征。虽然深度特征能够捕捉复杂的语义信息，但在训练数据集有限的情况下，它们可能存在过拟合的问题。相比之下，这些清晰的几何指标能够提供稳定、低噪声的监督信号，该信号在数学上具有鲁棒性，并且易于在农业实践中解释。因此，本章特意使用显式的关键表型指标，以支持在数据有限的情况下进行稳定学习，同时保持可扩展性，以便在未来获得更大数据集时学习关键表型表示。

最终的果蔬数据集（FVS）包含 220 张黄瓜图像和 159 张香蕉图像用于基本的实验，以及额外 70 张香蕉图像用于扩展测试，共计 449 个样本。每张图像都标注了新鲜重量、关键表型特征和质量等级。这种多模态标注结构为多任务模型的开发提供了丰富的监督信息。

3.3 实验分析

3.3.1 实验设置

所有实验均在 FVS 数据集上进行，具体构建方法见第 3.2节。该混合数据集用于训练和评估预分类模块，根据预测的类别，每张图像都会被分配到相应的子网络。

为了训练主要的多任务子网络模型，所有输入图像都被动态地调整为 224 × 224 像素。黄瓜子集被分为 195 张用于训练和验证，25 张用于测试。类似地，香蕉子集包含 129 张用于训练和验证，30 张用于测试，并且与额外的 70 张合并为 100 张作为扩展测试。为了增强模型的泛化能力，在训练过程中应用了一种在线增强策略，即引入最大 15 度的随机旋转。每个输入样本包含 RGB 图像及其对应的真实标注。这些属性作为模型回归头和分类头的预测目标。

所有模型均使用 AdamW 优化器 [84] 进行训练，学习率为 0.00005，批大小为 20，训练轮数为 100。所有实验均在 NVIDIA 3090 GPU 上进行。这些超参数的选择基于初步实验，以确保模型在两个子集上均能稳定收敛。

3.3.2 对比实验

为了体现本章提出的多任务框架的结果，本章与单任务模型、农业领域模型以及多任务模型进行对比实验。

首先，本章与单任务骨干网络进行比较。本章比较了不同卷积骨干网络在三种彼此独立的单任务模型中训练时的表现，以及将其集成到多任务框架中时的表现。单任务模型分别针对重量回归、类别特定的关键表型回归和品质等级分类单独训练。单任务结果通过训练三个彼此独立的模型得到，每个模型都配置为仅针对一个任务优化的单输入单输出网络。子网络在两个分支中使用相同的骨干网络。

表 3.3　黄瓜数据上的单任务与本章框架对比

单任务模型	**重量** R^2	**关键表型特征** R^2	**等级准确率**
ResNet18	0.734	0.884	0.680
ResNet34	0.964	**0.915**	0.760
ResNet50	0.963	0.725	**0.920**
VGG11	0.950	0.728	0.800
VGG16	**0.971**	0.867	**0.920**
本章方法	0.969	0.897	**0.920**

表 3.3 给出了黄瓜子集上的结果。对于回归任务，报告 R^2；对于分类任务，给出准确率。表中加粗表示该列最优结果，下划线表示该列次优结果。从重量回归看，单任务 VGG16 取得最高 R^2，本章方法为 0.969，为该指标次优；从曲直度回归看，单任务 ResNet34 取得最高 R^2，本章方法为 0.897，同样为次优，二者差值为 0.018。等级分类方面，本章方法准确率为 0.920，与单任务 ResNet50 和 VGG16 并列最优。该结果表明，在黄瓜数据上，本章方法未在两个回归指标上取得绝对最优，但保持了接近最优的回归性能，同时在等级分类上达到最优水平。

为了控制骨干网络容量的影响，本章在黄瓜数据上使用 ResNet34 对框架和单任务基线进行聚焦比较。所有模型都使用相同的数据划分、优化器和训练配置。在相同骨干网络和训练设置下，提出的框架取得了 0.920 的分类准确率，相比单任务分类模型的 0.760 提升了 0.160。重量回归的 R^2 从 0.964 略微提升到 0.969，关键表型特征回归的 R^2 从 0.915 略微下降到 0.897。结合这三项结果可见，在当前数据规模与实验协议下，多任务框架在回归任务上保持了与单任务模型接近的结果，同时在等级分类任务上获得了更明显的提升。该提升是在统一骨干与统一训练设置下得到的，并且对应于将三个独立模型整合为一个模型的训练与推理流程。详细对比可见图 3.11 左图。

表 3.4　香蕉数据上的单任务与本章框架对比

单任务模型	**重量 R^2**	**关键表型特征 R^2**	**等级准确率**
ResNet18	0.281	0.713	0.833
ResNet34	0.769	0.822	0.900
ResNet50	0.778	0.765	0.867
VGG11	0.779	**0.966**	0.800
VGG16	0.688	0.964	0.867
本章方法	**0.831**	0.906	**0.933**

表 3.4 给出了香蕉子集上的结果，表格结构与前文一致。回归任务包括重量回归和成熟度回归，报告指标为 R^2；分类准确率以 0 到 1 之间的小数表示。表中加粗表示该列最优结果。由表 3.4 可见，在香蕉数据上，本章方法在重量回归和等级分类上均取得最优结果：重量 R^2 为 0.831，高于单任务最优值 0.779；等级准确率为 0.933，高于单任务最优值 0.900。成熟度回归方面，单任务 VGG11 的 R^2 最高，本章方法为 0.906，低于最优值，但仍高于多数单任务骨干。该结果说明，在当前香蕉子集划分

下，多任务模型结构能够同时维持较高的回归与分类表现。

为了控制骨干网络容量的影响，本章在香蕉数据上对提出框架和单任务基线统一使用 ResNet18 骨干网络进行聚焦比较。在相同骨干网络和训练设置下，所提框架将重量 R^2 从 0.281 提升到 0.831，将关键表型特征回归 R^2 从 0.713 提升到 0.906，并将分类准确率从 0.833 提升到 0.933。由此可见，在该受控比较中，多任务框架在三项指标上均高于对应的单任务基线。对应结果在图 3.11 右图中给出了可视化对比。

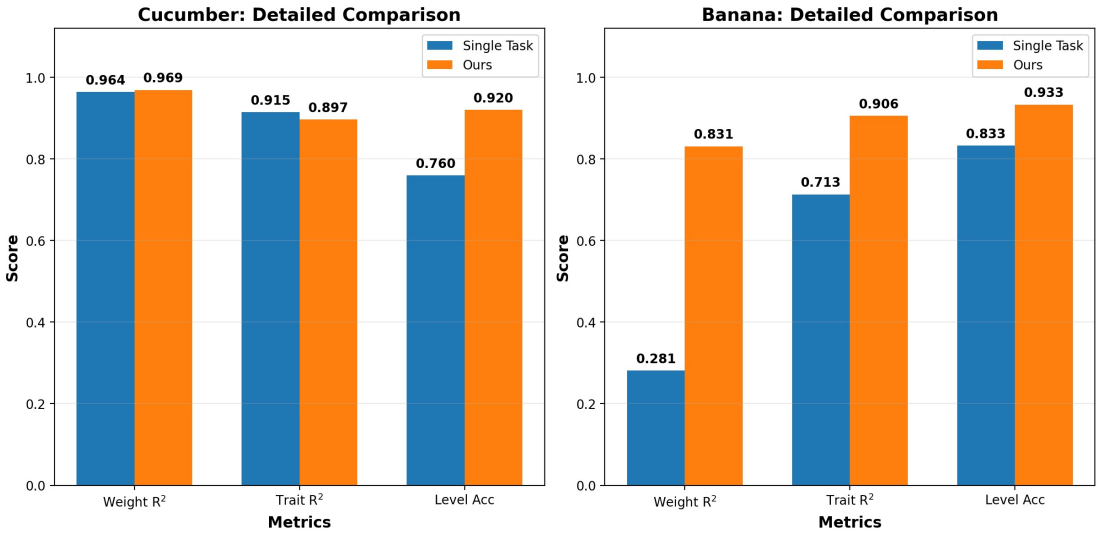


图 3.11　单任务方法与本章框架使用同一骨干的聚焦比较

本章还将提出的框架与专门为果蔬品质分类任务设计的模型进行比较。CIDIS 模型[85] 被选为具有代表性的外部基线，因为它在水果质量分级方面已经表现出较好的效果。CIDIS 采用专门的卷积结构，用于检测表面缺陷和表征成熟度的视觉模式。与本章提出的框架不同，它是一个单任务网络，仅处理输入图像并输出分类结果。为了适配本章数据集，本章将 CIDIS 最后的全连接层修改为四维输出，对应预定义的四个品质等级。

本章还纳入了 FruitVision[86]，这是一个为水果分类任务设计的轻量模型。本章考虑了两种训练设置：一种是不冻结任何层的全量微调，另一种是冻结骨干网络、仅训练分类头的部分微调。与 CIDIS 类似，FruitVision 中最后的全连接层也被替换为四维输出层，以匹配数据集中的品质分级方案。所有模型都在与提出的框架相同的数据集和数据划分上训练和评估，使用相同的超参数。

表 3.5 报告了在保留测试集上使用标准 RGB 图像输入时的准确率。提出的框架在两种数据上都取得了最高准确率。为了更严格地验证这些结果，本章对配对预

表 3.5　本章框架与农业模型等级分类性能比较

模型	**黄瓜准确率**	**p**	**香蕉准确率**	**p**
CIDIS	0.520	0.044	0.867	0.617
FruitVision（未冻结）	0.440	0.001	0.733	0.041
FruitVision（冻结）	0.640	0.070	0.667	0.043
本章方法	**0.920**		**0.933**	

测结果计算了 McNemar 检验的 p 值。对于黄瓜数据，提出的模型显著优于 CIDIS 和 FruitVision。对于香蕉数据，由于测试集较小，相比 CIDIS 的提升在统计上相当，$p = 0.617$，但相较 FruitVision 的优势仍然显著，$p < 0.05$。为了进一步分析分类行为，图 3.12 给出了黄瓜和香蕉数据集上的混淆矩阵。图 3.12 中子图 (a) 至 (d) 是在黄瓜数据上的结果，子图 (e) 至 (h) 是在香蕉数据上的结果，按从上到下的顺序依次对应 CIDIS、FruitVision（未冻结）、FruitVision（冻结）和本章方法。可视化结果表明，提出的模型具有更少的非对角误差，尤其在区分相邻等级时优于各基线模型。

这些结果说明，在当前实验设置下，提出的框架通过回归任务与分类任务的联合学习，能够获得更丰富的特征表示，并在两类数据上取得较一致的分类收益。此外，提出的框架将多种能力整合进一个统一框架中，减少了对多个独立模型的需求，并提高了集成农业评估流程中的部署效率。

结合结构设计与实验现象，本章给出三点可能原因。第一，回归任务与分类任务之间的共享特征表示促使模型同时提取整体轮廓线索和细粒度表面细节，而这两类信息对于品质分级都具有参考价值。第二，多任务训练提高了数据利用效率，使同一样本可以同时服务于多个目标，因此在小样本条件下具有一定的过拟合缓解作用。第三，相关任务的联合优化使模型在被评估的两类农产品上呈现较一致的性能趋势。

一个可能的疑问是，提出的框架受益于多任务监督，而基线模型仅依赖图像输入。为了解决这一问题并保证比较更加公平，本章在十个随机种子上进行了更稳健的评估，并对基线模型进行了增强。具体来说，对于 CIDIS 和 FruitVision，本章将图像特征与真实的标量重量和关键表型值进行拼接，从而保证所有方法都能够访问相同的信息。

表 3.6 显示，与三个基线模型相比，Wilcoxon 符号秩检验均得到 $p < 0.05$。这些结果说明，在扩展输入条件下，提出的多任务框架仍取得了更高的平均准确率。该现象提示，性能差异不仅与是否使用辅助标量有关，也与模型内部的特征交互机制

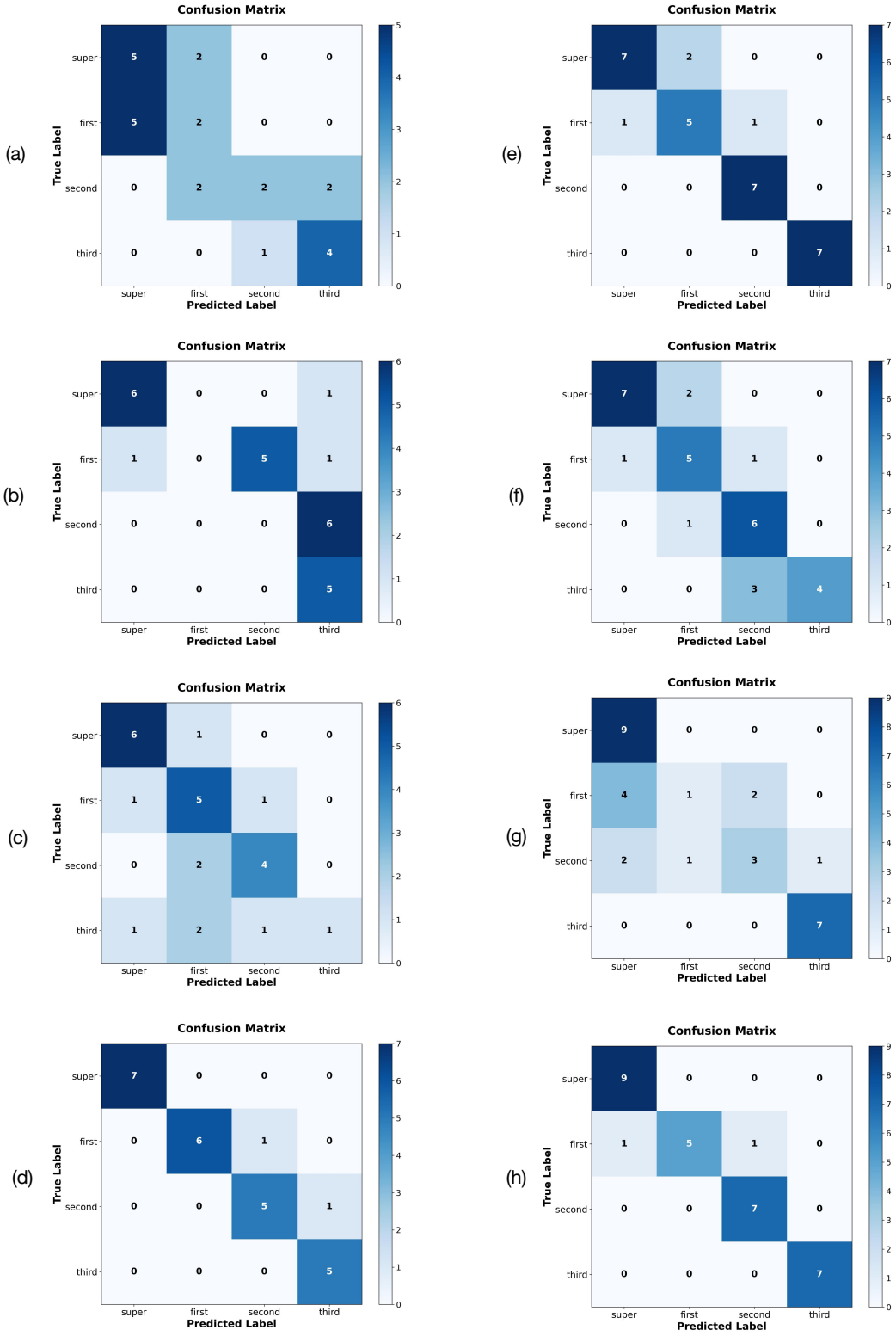


图 3.12　黄瓜与香蕉数据集混淆矩阵对比

表 3.6　扩展输入条件下各模型等级准确率比较

模型	**黄瓜准确率**	**p**	**香蕉准确率**	**p**
CIDIS	0.856 ± 0.043	0.010	0.827 ± 0.056	0.014
FruitVision（冻结）	0.693 ± 0.080	0.002	0.753 ± 0.079	0.004
FruitVision（未冻结）	0.540 ± 0.218	0.002	0.677 ± 0.180	0.002
本章方法	$\mathbf{0.944 \pm 0.040}$		$\mathbf{0.900 \pm 0.052}$	

相关。

为了评估提出的框架设计选择的有效性，本章还将框架与五种具有代表性的多任务学习基线进行比较：Hard Parameter Sharing(HPS)[87], DSelect-k[63], Learning to Branch(LTB)[62], Multi-gate Mixture-of-Experts(MMoE)[64] 以及 Multi-Task Attention Network(MTAN)[65]，并对它们的输出头进行了修改，以适配数据集。所有模型都在相同超参数下使用 10 个随机种子进行训练。

表 3.7　黄瓜数据上多任务学习基线对比

	重量		**关键表型特征**		**等级**			
模型	R^2	**RMSE**	R^2	**RMSE**	**Prec.**	**F1**	**Acc.**	**p**
DSelect-k	0.941 ± 0.045	7.170 ± 2.630	0.699 ± 0.218	8.018 ± 2.538	0.849 ± 0.173	0.842 ± 0.160	0.852 ± 0.144	0.125
HPS	0.960 ± 0.015	$\mathbf{6.202 \pm 1.138}$	0.753 ± 0.063	7.561 ± 0.966	0.806 ± 0.045	0.798 ± 0.048	0.808 ± 0.047	0.002
LTB	0.922 ± 0.050	8.444 ± 2.521	0.777 ± 0.120	7.033 ± 1.749	0.843 ± 0.035	0.800 ± 0.044	0.816 ± 0.041	0.002
MMoE	0.885 ± 0.217	8.515 ± 6.453	0.811 ± 0.046	6.619 ± 0.804	0.751 ± 0.244	0.756 ± 0.220	0.776 ± 0.193	0.002
MTAN	0.958 ± 0.019	6.308 ± 1.395	0.718 ± 0.091	8.028 ± 1.354	0.878 ± 0.070	0.867 ± 0.073	0.876 ± 0.070	0.084
本章方法	$\mathbf{0.962 \pm 0.015}$	7.394 ± 1.740	$\mathbf{0.883 \pm 0.043}$	$\mathbf{3.954 \pm 1.205}$	$\mathbf{0.951 \pm 0.034}$	$\mathbf{0.944 \pm 0.040}$	$\mathbf{0.944 \pm 0.040}$	

表 3.7 的结果表明，标准多任务学习架构通常难以平衡几何回归与质量分类这两类任务的竞争需求。HPS 在重量回归上表现最好，取得了最低的均方根误差，但在分级任务上表现较差。这说明单一共享骨干网络往往更关注主要的回归目标，无法为准确分类提供足够的特征。相反，MTAN 虽然取得了有竞争力的分类准确率，但在关键表型回归上明显较弱，这表明它的注意力机制可能更偏向语义类别可分性，而不是曲直度估计所需的细粒度几何精度。MMoE 和 LTB 等其他基线的表现介于两者之间，但在不同运行之间波动较大，这说明任务交互不够稳定。

相比之下，提出的框架在多任务结果上整体处于前列。它在关键表型回归和质量分类上相对最接近的基线具有优势。虽然其重量的均方根误差略高于 HPS，但它取得了最高的重量 R^2，说明预测结果与真实值之间仍保持较强相关性。这些结果表明，在融合之前将重量分支与关键表型分支分开，有助于减少任务干扰，并提升几

何线索与外观线索的学习效果。

Wilcoxon 符号秩检验进一步验证了提出框架的准确率优势在统计上是显著的，即相较 HPS、LTB 和 MMoE 时有 $p = 0.002$。对于 DSelect-k 和 MTAN，虽然由于基线运行方差较大，其 p 值超过了 0.05 阈值，但提出的方法仍取得了更高的平均准确率和更低的标准差，体现出较好的稳定性。

本章进一步在香蕉子集上使用 30 张测试图像评估多任务基线。表 3.8 给出了十个随机种子下的结果。表 3.8 与表 3.9 中加粗表示最优结果，下划线表示次优结果，p 值由 Wilcoxon 符号秩检验计算得到。

表 3.8 多任务学习基线在 30 张香蕉测试数据上的对比

	重量		关键表型特征		等级			
模型	R^2	**RMSE**	R^2	**RMSE**	**Prec.**	**F1**	**Acc.**	**p**
DSelect-k	-1.387 ± 5.677	26.996 ± 29.782	$\underline{0.768 \pm 0.254}$	0.064 ± 0.033	0.863 ± 0.201	0.852 ± 0.175	0.873 ± 0.126	0.824
HPS	-0.044 ± 2.248	19.090 ± 18.493	0.189 ± 0.032	0.135 ± 0.003	0.262 ± 0.007	0.333 ± 0.011	0.507 ± 0.020	0.002
LTB	0.134 ± 0.777	22.447 ± 9.065	$\mathbf{0.861 \pm 0.048}$	$\mathbf{0.055 \pm 0.010}$	0.857 ± 0.033	0.833 ± 0.027	0.843 ± 0.030	0.027
MMoE	$\underline{0.719 \pm 0.131}$	$\mathbf{13.530 \pm 2.702}$	0.524 ± 0.345	0.095 ± 0.041	0.668 ± 0.329	0.679 ± 0.284	0.750 ± 0.205	0.224
MTAN	0.567 ± 0.287	16.356 ± 5.072	0.551 ± 0.219	0.097 ± 0.026	$\mathbf{0.907 \pm 0.060}$	$\mathbf{0.899 \pm 0.064}$	0.900 ± 0.062	0.828
本章方法	$\mathbf{0.735 \pm 0.156}$	$\underline{13.560 \pm 4.490}$	0.744 ± 0.144	$\underline{0.064 \pm 0.013}$	$\underline{0.902 \pm 0.063}$	$\underline{0.885 \pm 0.064}$	$\mathbf{0.900 \pm 0.052}$	

在这个较小的测试划分上，不同基线处理三项任务的方式呈现出明显差异。MMoE 能较好拟合重量分布，这一点从较低的均方根误差可以看出，但它在关键表型与等级任务上表现不佳。其关键表型 R^2 和准确率都较低，说明它学习这两个目标的能力有限。相比之下，LTB 取得了最高的关键表型 R^2，似乎能够较好捕捉局部斑点模式，但其重量预测较弱。MTAN 得到了最佳准确率，但在回归任务上出现下降，尤其是在重量预测上。这说明它的注意力图更关注与类别相关的区域，而不是重量和成熟度所需的连续物理线索。

虽然提出的框架没有在每一个指标上都取得最佳值，但它给出了较稳定的多任务结果。它取得了最高的重量 R^2，并在准确率上与最佳基线持平，同时保持了较强的关键表型回归性能，位列第二。与通过牺牲分类性能来提高重量预测的 MMoE，以及通过降低重量预测性能来提高分类性能的 MTAN 相比，提出的带有注意力融合的双分支设计在多个任务之间实现了相对更均衡的表现。

Wilcoxon 符号秩检验表明，提出的框架相比 HPS 和 LTB 在准确率上取得了统计显著的提升。然而，与 DSelect-k、MMoE 和 MTAN 的比较得到了较高的 p 值，这说明在这个有限测试数据上，性能差异尚不能在统计上区分。因此，仅依赖这个小测

试集不足以进行严格评估。这一发现直接促使本章在下一节构建扩展的 100 图像测试集，以提供更稳健的评估。

在上述比较基础上，进一步分析了测试集扩展的影响。为了评估泛化能力，本章在扩展后的 100 张香蕉图像划分上测试了各模型。表 3.9 揭示了一个关键现象，即测试集扩展暴露了基线多任务学习方法的脆弱性。

表 3.9　多任务学习基线在 100 张香蕉测试数据上的对比

	重量		关键表型特征		等级			
模型	R^2	**RMSE**	R^2	**RMSE**	**Prec.**	**F1**	**Acc.**	**p**
DSelect-k	-4.090 ± 11.921	26.266 ± 28.069	$\underline{0.658 \pm 0.201}$	$\mathbf{0.070 \pm 0.019}$	0.712 ± 0.174	0.646 ± 0.126	0.677 ± 0.089	0.361
HPS	-1.407 ± 4.717	20.369 ± 16.851	0.213 ± 0.041	0.110 ± 0.003	0.190 ± 0.022	0.249 ± 0.027	0.385 ± 0.033	0.002
LTB	-0.964 ± 0.770	23.448 ± 4.518	-28.922 ± 40.751	0.536 ± 0.415	0.065 ± 0.014	0.101 ± 0.013	0.236 ± 0.009	0.002
MMoE	$\underline{0.271 \pm 0.267}$	$\underline{14.315 \pm 2.568}$	0.479 ± 0.261	0.086 ± 0.024	0.535 ± 0.279	0.498 ± 0.209	0.567 ± 0.156	0.098
MTAN	0.092 ± 0.506	15.704 ± 4.134	0.197 ± 0.354	0.108 ± 0.026	$\mathbf{0.785 \pm 0.040}$	$\mathbf{0.720 \pm 0.048}$	$\mathbf{0.731 \pm 0.045}$	0.004
本章方法	$\mathbf{0.451 \pm 0.099}$	$\mathbf{12.580 \pm 1.120}$	$\mathbf{0.662 \pm 0.083}$	$\underline{0.072 \pm 0.009}$	$\underline{0.758 \pm 0.031}$	$\underline{0.656 \pm 0.028}$	$\underline{0.677 \pm 0.022}$	

表 3.9 的结果显示，基线模型在回归任务上出现了明显崩溃。最突出的问题出现在重量预测上，DSelect-k、HPS 和 LTB 等方法都给出了负的 R^2，并伴随很大的标准差。这说明它们完全无法拟合扩展后更大测试集中的重量模式。即使是分类准确率高于提出的模型且对应 $p = 0.004$ 的 MTAN，在回归结果上也非常弱。其重量 R^2 仅约为 0.092，关键表型 R^2 仅约为 0.197。这表明 MTAN 过于关注类别信号，使得模型更像一个分类器，而不是一个真正的多任务模型。

相比之下，提出的框架在三项任务上表现得更加稳定。它在重量预测上取得了最高的 R^2，数值为 0.451，在关键表型预测上也取得了最高的 R^2，数值为 0.662。这些结果说明，在该测试划分下，提出框架的预测总体上更接近真实值，而部分基线模型的回归可用性较弱。提出的模型牺牲了少量分类准确率，以保持两个回归任务的可靠性，这与联合评估方法的核心目标一致。Wilcoxon 检验显示，提出的方法与 HPS 和 LTB 在分类结果上的差异十分明显，其 p 值为 0.002。尽管对于 MMoE 和 DSelect-k，准确率上的 p 值高于 0.05，但其回归结果相对较弱，因此在本章的多任务目标下适用性受限。

可以看到，与 30 图像测试划分相比，多个指标都有明显下降。为了理解这种变化，本章考察了图 3.13 所示的数据分布。较大的测试集在重量和成熟度两个连续变量上都具有更窄的取值范围，其中非常重或非常成熟的样本更少。这种更窄的范围增加了 R^2 指标的敏感性。现在，即使绝对误差较小，也会产生更大的相对误差，从

而降低 R^2 数值，即使模型整体行为仍然稳定。更大数据集中的等级分布也更加均衡，这消除了较小测试集中的类别不平衡，而这种不平衡此前会抬高准确率分数。这些发现说明，评估结果的变化是由测试集本身的性质造成的。提出的框架在这个更均衡且更具挑战性的划分上仍然保持稳定表现，这表明它在当前评估条件下具有稳定性，但未来仍然需要在更大规模的数据集上进一步验证。

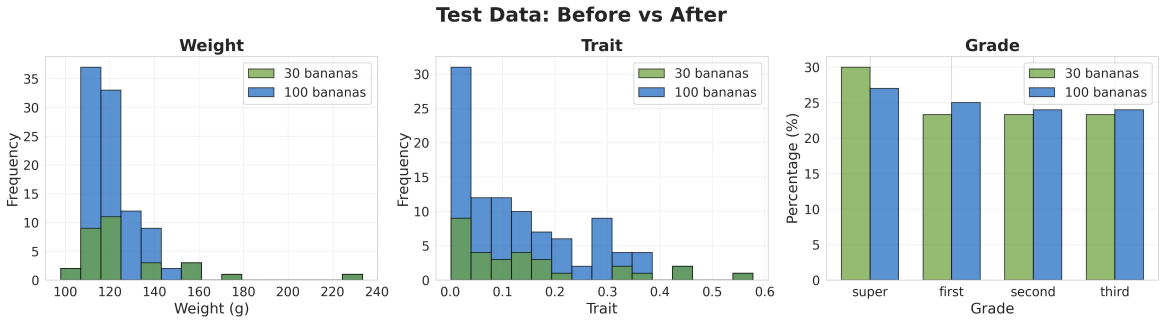


图 3.13　香蕉测试集扩展前后数据分布对比图

3.3.3　消融实验

为了探究提出的框架内部模块的有效性及其影响，本章设计了一系列消融实验，包括骨干网络、损失函数、特征增强和融合的消融实验。首先，分析骨干网络选择对多任务性能的影响。在该组实验中，仅替换主干网络类型，其余训练与推理设置保持一致，以便直接比较网络容量与特征偏好对三任务联合优化的影响。

表 3.10　黄瓜数据上的骨干网络消融实验

骨干网络	**重量** R^2	**曲直度** R^2	**等级准确率**
ResNet18	0.962	0.833	0.880
ResNet34	0.969	**0.897**	**0.920**
ResNet50	**0.971**	0.787	0.720

表 3.11　香蕉数据上的骨干网络消融实验

骨干网络	**重量** R^2	**成熟度** R^2	**等级准确率**
ResNet18	0.831	**0.906**	**0.933**
ResNet34	**0.882**	0.863	0.833
ResNet50	0.767	0.809	0.867

由表 3.10 可见，黄瓜数据上三种主干呈现出明显的任务权衡关系。ResNet50 在重量回归上取得最高 R^2，结果为 0.971，但其曲直度 R^2 仅为 0.787，等级准确率仅为 0.720。相较之下，ResNet34 的重量 R^2 为 0.969，仅低 0.002，却在曲直度回归上提升 0.110，在等级分类上提升 0.200。与 ResNet18 相比，ResNet34 在重量、曲直度和等级准确率上分别提升 0.007、0.064 和 0.040。上述结果说明，在黄瓜数据上，ResNet34 在三任务之间提供了更稳健的综合平衡。

由表 3.11 可见，香蕉数据上的最优骨干与黄瓜不同。ResNet18 在成熟度回归与等级分类上分别达到 0.906 和 0.933，均为该数据集最优。虽然 ResNet34 在重量回归上更高，结果为 0.882，而 ResNet18 为 0.831，但其成熟度回归下降 0.043，等级准确率下降 0.100，整体任务协同性明显减弱。ResNet50 在三个任务上也未超过 ResNet18。该差异表明，不同农产品类别对特征层级与模型容量的需求并不一致。黄瓜更依赖中等容量主干对形态与等级信息的联合建模，香蕉在当前样本规模下更适合轻量主干对纹理与成熟度线索的稳定提取。

基于以上结果，后续消融实验采用分数据集的骨干设置：黄瓜使用 ResNet34，香蕉使用 ResNet18。在此基础上，进一步分析不同损失函数设置对多任务性能的影响。为验证本章所采用损失函数设计的合理性，本节比较三种多任务损失设置对模型性能的影响，并在上述骨干设置下分别于黄瓜和香蕉数据上进行评估。对于回归任务，报告 R^2 与均方根误差 RMSE；对于分类任务，报告精确率 Precision、F_1 值和准确率 Accuracy。所有指标均表示为多次重复运行下的均值 ± 标准差。本节消融表格中加粗数值表示对应设置下的最优结果。

本章考虑三种设置。均方误差记为 MSE。交叉熵记为 CE。

第一种设置是两个回归头使用均方误差 MSE，分类头使用交叉熵 CE。总损失为

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE\_weight}} + \mathcal{L}_{\text{MSE\_trait}} + \mathcal{L}_{\text{CE\_grade}}. \tag{3.15}$$

其中，$\mathcal{L}_{\text{MSE\_weight}}$ 和 $\mathcal{L}_{\text{MSE\_trait}}$ 分别表示重量回归与关键表型特征回归的 MSE 损失，$\mathcal{L}_{\text{CE\_grade}}$ 表示质量分类的 CE 损失。

第二种设置是两个回归头使用 Huber 损失，分类头使用交叉熵 CE。总损失为

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{Huber\_weight}} + \mathcal{L}_{\text{Huber\_trait}} + \mathcal{L}_{\text{CE\_grade}}. \tag{3.16}$$

其中，$\mathcal{L}_{\text{Huber\_weight}}$ 和 $\mathcal{L}_{\text{Huber\_trait}}$ 分别表示重量回归与关键表型特征回归的 Huber 损失。

第三种设置，也是本章的设置，回归头使用任务不确定性对损失进行加权[61]。进一步，将总损失修改为

$$\mathcal{L}_{\text{total}} = \frac{1}{2\sigma_1^2}\mathcal{L}_{\text{weight}} + \frac{1}{2\sigma_2^2}\mathcal{L}_{\text{trait}} + \log\sigma_1 + \log\sigma_2 + \gamma\,\mathcal{L}_{\text{grade}}. \tag{3.17}$$

其中，$\mathcal{L}_{\text{weight}}$ 和 $\mathcal{L}_{\text{trait}}$ 分别表示重量与关键表型特征的 Huber 损失项，$\mathcal{L}_{\text{grade}}$ 表示分类任务的 CE 损失，σ_1 和 σ_2 是可学习的对数方差参数，γ 是一个标量，若无特别说明则设为 1。

表 3.12　损失函数设置消融实验

		重量		关键表型特征		等级		
数据	损失	R^2	RMSE	R^2	RMSE	Prec.	F1	Acc.
黄瓜	MSE + CE	0.955±0.012	6.648±0.893	0.873±0.042	5.362±1.007	0.789±0.100	0.715±0.095	0.740±0.087
	Huber + CE	0.968±0.007	5.636±0.614	0.879±0.051	5.218±1.106	0.919±0.048	0.900±0.058	0.904±0.054
	Uncertainty-Weighted	**0.979±0.003**	**4.543±0.321**	**0.909±0.034**	**4.546±0.891**	**0.931±0.039**	**0.918±0.046**	**0.920±0.044**
香蕉	MSE + CE	0.823±0.059	10.782±1.803	0.114±0.224	0.140±0.017	0.376±0.267	0.365±0.192	0.473±0.148
	Huber + CE	**0.851±0.040**	**9.961±1.379**	0.758±0.084	0.073±0.013	0.818±0.115	0.760±0.083	0.800±0.056
	Uncertainty-Weighted	0.849±0.035	10.035±1.177	**0.881±0.021**	**0.052±0.005**	**0.871±0.078**	**0.849±0.077**	**0.853±0.078**

表 3.12 给出了实验结果。在黄瓜数据上，不确定性加权设置在所有任务上都取得了最佳结果，包括回归与分类。在香蕉数据上，Huber 加交叉熵在重量回归上取得最佳结果，而不确定性加权设置在关键表型特征回归与分类指标上取得最佳结果。均方误差加交叉熵的损失设置在香蕉数据上的关键表型特征任务与分类任务上明显落后。

这些现象表明，在存在标签噪声时，鲁棒回归更有利于重量估计，因此 Huber 损失更合适。不确定性加权方案能够在训练过程中平衡各任务贡献，并利用关键表型线索与等级之间的相关性，因此有利于关键表型估计与质量预测。

在完成损失函数设置分析后，本节进一步评估分类头中不同融合方法的影响，重点考察交叉注意力融合策略的作用。共比较四种策略，即逐元素加法、逐元素乘法、通道拼接和交叉注意力。实验在黄瓜和香蕉数据上进行，骨干网络分别采用上一小节确定的 ResNet34 与 ResNet18。结果汇总于表 3.13。

在黄瓜数据上，交叉注意力在等级分类上取得了最高准确率，并在曲直度回归上取得了最高的 R^2。它在重量估计上也达到了有竞争力的性能。通道拼接在重量回归上取得了最高的 R^2，但同时带来了分类准确率的明显下降。这些结果说明，交叉

注意力更善于利用来自关键表型分支的空间信息来支持质量分级任务，同时保持稳定的回归性能。

在香蕉数据上，交叉注意力在重量回归、成熟度回归和等级分类上都优于其他方法。特别是，它在成熟度回归上取得了最高的 R^2，并在分类任务上取得了最佳的精确值、F1 和准确率。这说明，交叉注意力提供的选择性关键表型线索整合方式增强了共享特征的判别能力，因此同时有利于回归和分类。

总体上，交叉注意力相比更简单的融合方法，在多个任务之间取得了更好的平衡。它能够自适应地从关键表型分支传递信息，从而增强特征表示，并支持回归与分类任务的联合优化。

表 3.13　交叉注意力融合方式消融实验

		重量		关键表型特征		等级		
数据	方法	R^2	RMSE	R^2	RMSE	Prec.	F1	Acc.
黄瓜	add	0.971±0.005	5.357±0.473	0.864±0.045	5.564±0.971	0.897±0.035	0.867±0.043	0.872±0.039
	concat	**0.980±0.002**	**4.438±0.194**	0.902±0.040	4.696±1.028	0.841±0.042	0.799±0.052	0.808±0.047
	multiply	0.974±0.005	4.996±0.559	0.837±0.058	6.086±1.123	0.872±0.012	0.821±0.042	0.840±0.025
	cross-attn	0.979±0.003	4.543±0.321	**0.909±0.034**	**4.546±0.891**	**0.931±0.039**	**0.918±0.046**	**0.920±0.044**
香蕉	add	0.818±0.093	10.783±2.677	0.588±0.142	0.095±0.016	0.158±0.081	0.219±0.096	0.373±0.090
	concat	0.825±0.072	10.651±2.190	0.690±0.076	0.083±0.010	0.090±0.000	0.139±0.000	0.300±0.000
	multiply	0.840±0.056	10.246±1.839	0.751±0.143	0.072±0.019	0.804±0.098	0.782±0.063	0.813±0.034
	cross-attn	**0.849±0.035**	**10.035±1.177**	**0.881±0.021**	**0.052±0.005**	**0.871±0.078**	**0.849±0.077**	**0.853±0.078**

随后，本节分析大核注意力融合 LKAF 的放置位置与层数影响。具体做法是改变 LKAF 在特征融合路径中的放置位置及其层数。本章考虑了四个可能的位置，从浅到深记为 L1 层到 L4 层。不同设置在黄瓜和香蕉数据上进行测试。结果见表 3.14。表中勾号表示对应层启用 LKAF。

在黄瓜数据上，加入 LKAF 后，所有任务的性能都逐步提升。当 LKAF 被应用在全部四层时，结果最好。该配置在重量回归上取得了最高的 R^2，在两个回归任务上都取得了最低的均方根误差，并在分类任务上取得了最佳的精确值、F1 和准确率。仅在较深层使用 LKAF 相比只带来了有限收益，这说明，在多个尺度上引入长程依赖对于稳定的多任务性能很重要。

在香蕉数据上，完整的 LKAF 配置同样取得了最佳分类性能以及成熟度回归中最佳的 R^2。对于重量回归，仅在最深的两层使用 LKAF 可以得到最高的 R^2，但会导致分类准确率下降。这说明，尽管更深层的 LKAF 有利于回归，但多层级 LKAF 更

有利于在所有任务之间取得平衡。

这些结果验证了 LKAF 是本章框架中的一个关键组成部分。多层级放置方式使网络能够在整个特征层级中持续建模长程空间依赖，从而提高回归精度，并增强质量分类所需的共享表示。

表 3.14　LKAF 放置位置与层数消融实验

	LKAF 放置位置				重量		关键表型特征		等级		
数据	L1	L2	L3	L4	R^2	RMSE	R^2	RMSE	Prec.	F1	Acc.
黄瓜					0.974±0.006	5.017±0.598	0.906±0.055	4.456±1.491	0.920±0.029	0.902±0.021	0.904±0.020
				✓	0.970±0.009	5.440±0.752	0.898±0.040	4.789±1.009	0.912±0.033	0.893±0.042	0.896±0.041
			✓	✓	0.970±0.006	5.377±0.583	0.895±0.014	4.956±0.327	0.920±0.045	0.912±0.054	0.912±0.053
		✓	✓	✓	0.974±0.006	5.062±0.607	0.899±0.033	4.802±0.802	0.906±0.054	0.892±0.052	0.896±0.048
	✓	✓	✓	✓	**0.979±0.003**	**4.543±0.321**	**0.909±0.034**	**4.546±0.891**	**0.931±0.039**	**0.918±0.046**	**0.920±0.044**
香蕉					0.840±0.049	10.291±1.604	0.757±0.090	0.073±0.014	0.822±0.125	0.772±0.085	0.807±0.053
				✓	0.848±0.035	10.082±1.214	0.761±0.072	0.072±0.011	0.861±0.020	0.760±0.054	0.800±0.037
			✓	✓	**0.864±0.026**	**9.564±0.950**	0.695±0.167	0.080±0.022	0.714±0.162	0.688±0.159	0.747±0.119
		✓	✓	✓	0.837±0.037	10.431±1.158	0.707±0.031	0.081±0.004	0.805±0.073	0.720±0.094	0.747±0.075
	✓	✓	✓	✓	0.849±0.035	10.035±1.177	**0.881±0.021**	**0.052±0.005**	**0.871±0.078**	**0.849±0.077**	**0.853±0.078**

进一步地，本节分析 FPN 在不同分支中的放置位置影响。具体做法是调整特征金字塔网络 FPN 在网络中的作用分支。考虑了四种配置，即同时应用于重量分支和关键表型分支、两个分支都不使用、仅应用于重量分支以及仅应用于关键表型分支。实验在黄瓜和香蕉数据上进行。结果见表 3.15。表中 W 与 T 分别表示重量分支与关键表型分支是否启用 FPN。

在黄瓜数据上，仅在关键表型分支中使用 FPN 取得了最佳的整体平衡。该配置在曲直度回归上取得了最高的 R^2，在两个回归任务上都取得了最低的均方根误差，并取得了最佳分类准确率。当 FPN 同时作用于两个分支时，重量回归性能略有提升，但代价是分类准确率下降。完全去除 FPN 也会导致关键表型回归和分类性能下降。这些结果说明，FPN 提供的多尺度特征聚合在关键表型分支中最有效，因为该分支需要更细致的空间表示。

在香蕉数据上，趋势保持一致。仅在关键表型分支中使用 FPN 时，成熟度回归的 R^2 最高，分类结果也最好。相比之下，仅在重量分支中使用 FPN 时，重量回归性能最好，但分类和成熟度预测较弱。在两个分支上同时使用 FPN 并没有超过仅关键表型分支的设置。这些发现说明，将 FPN 集中用于关键表型分支有助于捕捉与曲直度或成熟度相关的细微视觉线索，而这些线索对分类很关键；而在重量分支中过多使用 FPN 可能会削弱这种效果。

总体来看，这一消融实验验证了最有效的 FPN 放置方式是在关键表型分支内部。这种设计为关键表型相关任务利用了多尺度空间特征，同时也支持了更好的分类性能。

表 3.15　FPN 放置位置消融实验

	FPN 放置位置		重量		关键表型特征		等级		
数据	W	T	R^2	RMSE	R^2	RMSE	Prec.	F1	Acc.
黄瓜	✓	✓	0.972±0.007	5.201±0.696	0.897±0.031	4.881±0.688	0.922±0.036	0.902±0.042	0.904±0.041
			0.980±0.009	**4.380±0.941**	0.846±0.053	5.931±1.021	0.894±0.073	0.873±0.071	0.880±0.067
	✓		0.972±0.010	5.197±0.884	0.862±0.066	5.554±1.290	0.920±0.036	**0.930±0.034**	0.919±0.036
		✓	0.979±0.003	4.543±0.321	**0.909±0.034**	**4.546±0.891**	**0.931±0.039**	0.918±0.046	**0.920±0.044**
香蕉	✓	✓	0.842±0.055	10.178±1.752	0.797±0.084	0.066±0.013	0.856±0.065	0.810±0.054	0.827±0.053
			0.842±0.027	10.323±0.862	0.756±0.113	0.072±0.017	0.852±0.037	0.789±0.053	0.813±0.045
	✓		**0.851±0.040**	**9.961±1.379**	0.758±0.084	0.073±0.013	0.818±0.115	0.760±0.083	0.800±0.056
		✓	0.849±0.035	10.035±1.177	**0.881±0.021**	**0.052±0.005**	**0.871±0.078**	**0.849±0.077**	**0.853±0.078**

最后，本节分析共享与非共享骨干网络的差异。为了研究骨干网络设计对三个任务的影响，本章比较了一个在任务之间共享骨干网络的变体与提出的架构。提出的架构为每个任务使用彼此独立的任务特定骨干网络。表 3.16 给出了黄瓜和香蕉两个子集上的结果。

表 3.16　共享与非共享骨干网络消融实验

		重量		**关键表型特征**		**等级**		
数据	**骨干设置**	R^2	RMSE	R^2	RMSE	Prec.	F1	Acc.
黄瓜	共享	0.960 ± 0.011	6.271 ± 0.837	0.826 ± 0.025	6.389 ± 0.464	0.740 ± 0.092	0.724 ± 0.089	0.736 ± 0.083
	非共享	$\mathbf{0.979 \pm 0.003}$	$\mathbf{4.543 \pm 0.321}$	$\mathbf{0.909 \pm 0.034}$	$\mathbf{4.546 \pm 0.891}$	$\mathbf{0.931 \pm 0.039}$	$\mathbf{0.918 \pm 0.046}$	$\mathbf{0.920 \pm 0.044}$
香蕉	共享	0.846 ± 0.035	10.164 ± 1.159	0.682 ± 0.213	0.082 ± 0.025	0.848 ± 0.071	0.826 ± 0.068	0.833 ± 0.067
	非共享	$\mathbf{0.849 \pm 0.035}$	$\mathbf{10.035 \pm 1.177}$	$\mathbf{0.881 \pm 0.021}$	$\mathbf{0.052 \pm 0.005}$	$\mathbf{0.871 \pm 0.078}$	$\mathbf{0.849 \pm 0.077}$	$\mathbf{0.853 \pm 0.078}$

在黄瓜数据上，使用非共享骨干网络在所有任务上都带来了明显提升。共享模型虽然在重量预测上仍然可接受，但在分级准确率上出现了较大下降，数值约为 73.6%。这说明，为重量预测学习到的特征主导了共享特征空间，从而削弱了正确分级所需的有效线索。当两个骨干网络彼此分离后，网络可以同时学习用于重量预测的全局轮廓信息以及用于曲直度预测的局部弯折线索，并且不会让一个任务干扰另一个任务。这种设计将分类准确率提升到 92.0%，并降低了关键表型回归的均方根误差。这些结果说明，独立特征提取对于学习黄瓜不同物理属性很重要。

在香蕉数据上，分离骨干网络的效果在成熟度相关的关键表型预测上最为明显。两种设计在重量指标上的变化不大，但非共享结构将关键表型 R^2 从 0.682 提升到了

0.881。这说明，在共享模型中，有利于重量预测的强全局信号会掩盖香蕉表面细小纹理模式，而这些局部纹理正是成熟度估计所需要的。将两个分支分开后，关键表型路径便可以专注于这些局部纹理线索，从而恢复原先丢失的性能。改进后的关键表型特征也进一步提高了分级准确率，这说明当不同任务依赖的视觉线索差异较大时，非共享设计能够提供更好的任务平衡。

3.4 本章小结

本章提出了基于多任务学习的果蔬多属性联合评估框架，围绕黄瓜与香蕉两类样本，构建了包含重量、关键表型特征值与等级标签的 FruVegSet 数据集。该框架通过重量分支与关键表型分支分别建模不同任务所需线索，再结合 FPN、LKAF 与交叉注意力完成重量回归、关键表型回归和等级分类的联合优化，并通过不确定性加权损失提升训练过程的稳定性与任务平衡性。在当前数据规模与实验协议下，对比实验表明本章方法在两类数据上均具有较好的综合表现，尤其在关键表型相关任务和等级分类任务上体现出更稳定的优势。消融实验进一步说明，非共享双分支骨干、关键表型分支中的多尺度特征增强以及跨分支融合模块是性能提升的主要来源。上述结果为下一章基于中间样本扩展与伪标签学习的方法提供了可复用的基础模型与训练框架。

第四章　基于中间样本生成的多任务扩展学习方法

针对上一章农业果蔬图像数据及其对应标注数据规模有限的问题，本章提出一种基于中间样本生成的多任务扩展学习方法。方法流程包括中间样本生成、主体区域提取、伪标签赋值和联合训练四个阶段。本章所称的“轻量”，主要指伪标签构建阶段直接利用已训练分支输出、显式几何测量与确定性映射规则完成标签构建，并仅引入少量可调超参数。该设计旨在在不显著增加人工标注成本与训练开销的条件下，为下游多任务学习提供稳定的补充监督。

4.1　方法概述

4.1.1　整体框架

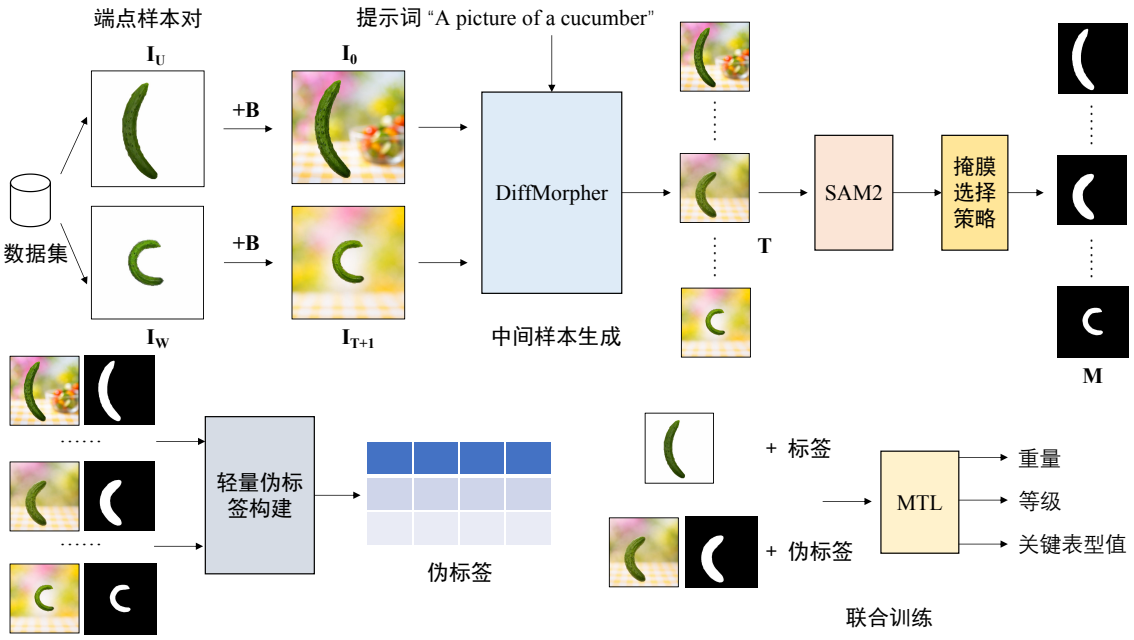


图 4.1　本章方法整体框架

本章方法的整体框架如图 4.1 所示，整体流程由中间样本生成、掩膜主体区域提取、伪标签构建以及联合训练四个部分组成。首先，仅在同一类别的训练子集中构造端点样本对，并在端点之间生成中间图像，以补充原始数据中难以直接采集的过渡状态。在送入生成模型之前，先从数据集中选取同类样本的图像对，作为端点

样本，并对端点样本执行统一背景补全，将主体图像叠加到背景模板 $\boldsymbol{B}$ 上，形成生成模型的输入图像，以减弱原始裁剪差异和背景缺失对后续生成过程的干扰。随后，对生成图像执行自动分割并通过主体掩膜选择策略来确定生成图像中果蔬的主体区域，以降低生成图像中的背景干扰对后续测量的影响。

在获得主体区域后，本章分别构建重量、关键表型特征和等级伪标签：重量标签由预测值与端点区间门控共同确定；关键表型特征标签由显式测量直接计算；等级标签由连续属性通过明确映射规则得到。最后，将伪标注中间样本与原始样本联合用于多任务模型训练。相关端点配对规则、生成参数、伪标签边界处理与映射公式在后续小节中给出。

4.1.2 中间样本生成

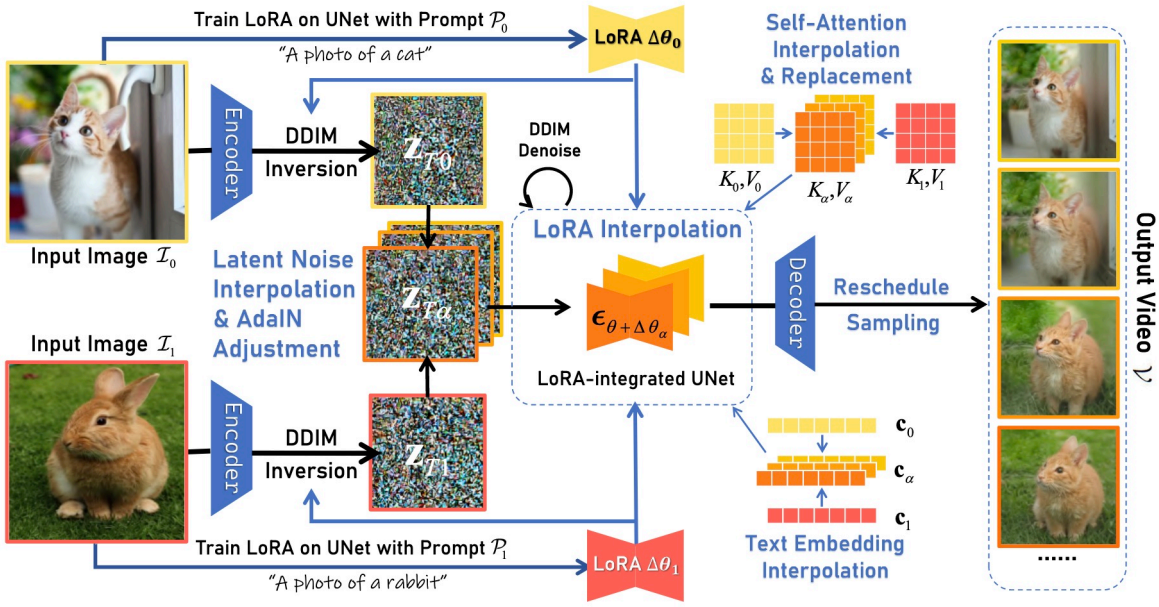


图 4.2 DiffMorpher 框架图[88]

中间样本生成的目的在于为原本规模较小的训练数据提供扩展样本，并在一定程度上补充端点样本之间缺失的过渡状态。对于农业果蔬对象而言，重量、关键表型特征以及品质等级等属性通常并不是完全孤立变化的，而是随着外观状态的改变呈现出一定的连续变化趋势。然而，受采集条件和标注成本限制，原始数据只覆盖少量离散状态，这会限制模型的学习能力。为此，本章在两个真实端点样本之间构造过渡图像，以实现中间样本补充，从而扩展训练样本。

本章采用 DiffMorpher [88] 作为中间样本生成模型。DiffMorpher 是一种基于预训

练扩散模型的图像变化方法，其目标是在两张输入图像之间生成自然且平滑的中间过渡序列。与传统依赖几何配准和像素混合的图像形变方法不同，DiffMorpher 并不是直接在像素空间中对两张图像做简单融合，而是借助扩散模型的生成先验，在更高层的表示空间中建立由起始图像到目标图像的连续过渡路径，因此生成得到的中间图像通常具有较好的结构合理性与视觉自然性。根据其整体流程，DiffMorpher 以两张输入图像及其文本提示作为输入，首先对两端图像进行 DDIM 反向的操作以获得对应潜空间表示，同时分别拟合各自的 LoRA 模块。随后，在 LoRA 参数空间和潜变量空间中进行联合插值，使中间结果在高层语义和低层外观两个层面都表现出逐步过渡的特征。此外，原方法还通过自注意力插值、AdaIN 调整和采样顺序控制，进一步提升相邻中间帧之间的纹理连续性和视觉一致性。

在本章中，DiffMorpher 被用于农业果蔬同类别样本之间的中间样本构造。首先在同一类别内部选取两个真实样本作为一组端点输入，其中黄瓜样本仅与黄瓜样本配对，香蕉样本仅与香蕉样本配对，以避免跨类别生成带来的语义偏移。端点选择基于训练子集中的组图像组织方式进行，每组仅在类别内部配对。考虑到原始样本多为主体裁剪结果，设该端点样本对后续共生成 T 个中间样本，则对应序列索引为 $0, 1, \ldots, T, T+1$，其中 0 与 $T+1$ 分别表示两端点位置。在输入 DiffMorpher 之前，本章先对每个端点图像补加背景模板 $\boldsymbol{B}$，并将两端输入定义为

$$\boldsymbol{I}_0 = \boldsymbol{I}_U + \boldsymbol{B}, \tag{4.1}$$

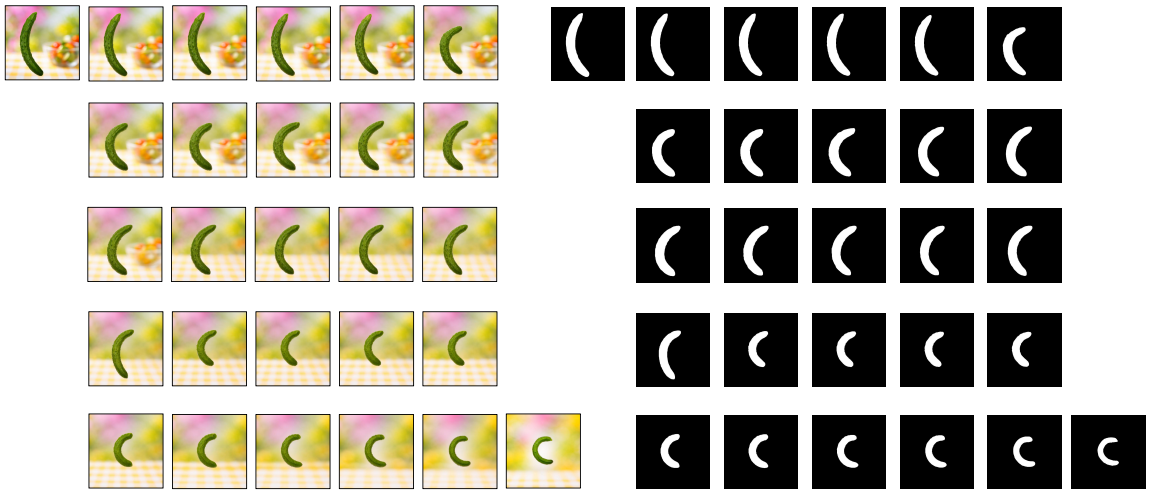
$$\boldsymbol{I}_{T+1} = \boldsymbol{I}_W + \boldsymbol{B}. \tag{4.2}$$

其中，$\boldsymbol{I}_U$ 和 $\boldsymbol{I}_W$ 分别表示同一类别下选取的一对真实端点样本图像，$\boldsymbol{B}$ 表示背景模板，$\boldsymbol{I}_0$ 和 $\boldsymbol{I}_{T+1}$ 表示加入背景后的两端输入图像，“$+\boldsymbol{B}$” 的操作表示在保留主体外观不变的前提下添加背景。该处理用于统一端点图像的背景分布，降低生成序列中的伪影，并尽量减少生成失真图像对后续步骤的影响。

在端点样本对确定之后，本章将其与提示词输入 DiffMorpher，对于一组香蕉样本，使用提示词 “A picture of a banana”；对于一组黄瓜样本，使用提示词 “A picture of a cucumber”。将图像和对应提示词输入 DiffMorpher 后，训练其中的 LoRA 模块，并通过插值和采样，生成中间图像。对于同一对端点图像，生成得到的中间图像在

视觉上表现为由起始端点向目标端点逐步过渡的过程。这些中间图像并不是对原图进行翻转、旋转或裁剪等简单增强后得到的结果，而是对应于两个真实状态之间构造出的新样本，因此既能够在一定程度上扩充原始小样本数据，也能够为端点之间原本未被采集到的中间状态提供补充。

图 4.3 和图 4.4 给出了黄瓜和香蕉的中间样本生成示例。其中，子图 (a) 展示了由两端真实样本之间生成的一组中间图像，可以观察到目标外观沿序列呈现出较为连续的变化过程。对于黄瓜样本，中间序列主要反映了形态和弯曲状态的逐步变化；对于香蕉样本，中间序列则体现了整体外观和表面状态的渐变特征。



(a) (b)

图 4.3 黄瓜的中间样本生成示例及其分割掩膜

4.1.3 主体掩膜选择策略

在完成中间样本生成之后，所得图像虽然在视觉上构成了由端点样本向目标样本逐步过渡的序列，但这些图像并不能直接用于后续的关键表型特征测量与伪标签构建。其原因在于，农业果蔬目标通常只占据图像中的局部区域，而生成过程中可能同时引入背景干扰等现象。如果直接在整幅图像上进行属性计算，容易受到无关区域影响，从而降低后续几何测量和标签构建的稳定性。因此，在中间样本生成之后，本章首先对生成图像进行目标分割，以提取尽可能准确的主体区域，为后续分析提供统一基础。

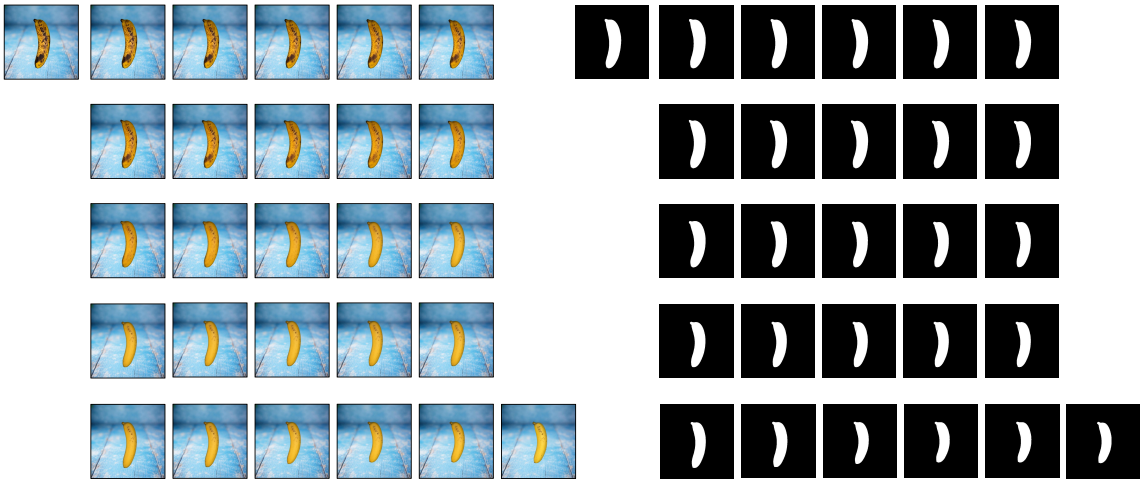


图 4.4　香蕉的中间样本生成示例及其分割掩膜

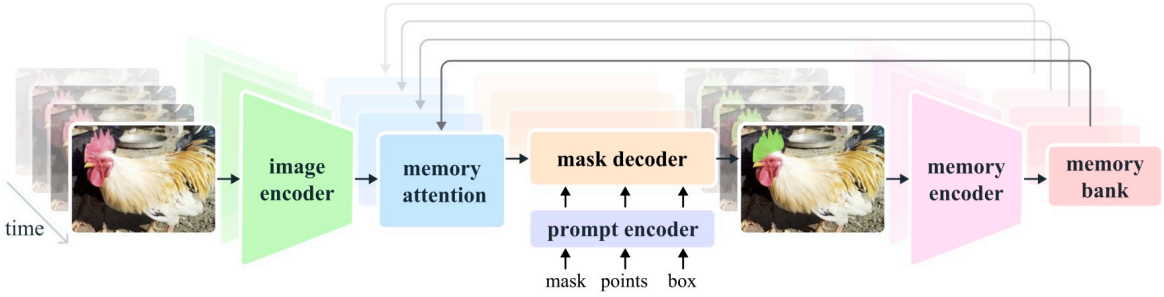


图 4.5　SAM2 框架图[89]

本章采用 Segment Anything Model 2（SAM2）[89] 作为自动分割模型。SAM2 是一种面向图像与视频统一建模的可提示分割基础模型，其核心思想是在当前帧图像特征的基础上，结合提示信息与历史记忆，预测目标区域对应的分割掩膜。根据原论文给出的模型结构，SAM2 主要由图像编码器、记忆注意力模块、提示编码器、掩膜解码器以及记忆编码器等部分组成。其中，图像编码器负责提取输入图像的视觉特征，提示编码器用于接收点、框或掩膜等提示信息，掩膜解码器则根据融合后的特征输出目标区域的分割结果；在视频任务中，模型还通过记忆模块利用前序帧信息增强时序一致性，而在单张图像场景下，该过程可以视为记忆为空的特殊情况。

对于本章任务而言，尽管中间样本是单张生成图像而非原始视频序列，但 SAM2 仍然适合作为自动分割工具。原因在于 SAM2 具备较强的开放域目标分割能力，能够在缺乏专门训练的情况下对果蔬主体区域给出较合理的候选掩膜；基于这一考虑，

本章使用 SAM2 对每张中间样本图像进行自动分割，获得对应的候选掩膜集合。然而，SAM2 对单张图像进行自动分割时，通常会输出多个候选掩膜。这些候选结果中既可能包含真实目标主体，也可能包含背景区域等其他区域。如果直接将其中任意一个候选掩膜作为后续分析对象，容易引入额外噪声。

因此，本章设计了一种基于多因素加权的主体掩膜选择策略，从所有候选掩膜中选取最可能对应主体的区域。该策略综合考虑候选区域的面积、中心位置、长宽比、分割质量以及是否贴近图像边界等因素。

设输入图像尺寸的高度为 H，宽度为 W，则图像面积 A_{img} 为

$$A_{\text{img}} = H \cdot W. \tag{4.3}$$

对于自动生成的候选掩膜集合 $\mathcal{M} = \{\boldsymbol{M}_i\}_{i=1}^{N}$，将每个候选掩膜 $\boldsymbol{M}_i$ 对应的面积记为 A_i，并将其外接框记为 $\boldsymbol{b}_i$ 表示为

$$\boldsymbol{b}_i = (x_i, y_i, w_i, h_i), \tag{4.4}$$

其中 (x_i, y_i) 为外接框左上角坐标，w_i 和 h_i 分别表示外接框的宽和高。

为去除明显过小的噪声区域，仅保留面积不低于图像总面积一定比例的候选掩膜，即

$$A_i \geq \tau A_{\text{img}}, \tag{4.5}$$

其中 τ 为面积阈值。本章中设定 $\tau = 0.01$。满足式 (4.5) 的掩膜候选集合记为 $\mathcal{M}'$。对于任一保留候选掩膜 $\boldsymbol{M}_i \in \mathcal{M}'$，其外接框中心坐标 (c_i^x, c_i^y) 定义为

$$c_i^x = x_i + \frac{w_i}{2}, \qquad c_i^y = y_i + \frac{h_i}{2}. \tag{4.6}$$

图像中心坐标 $(c_{\text{img}}^x, c_{\text{img}}^y)$ 定义为

$$c_{\text{img}}^x = \frac{W}{2}, \qquad c_{\text{img}}^y = \frac{H}{2}. \tag{4.7}$$

同时，为衡量候选区域与图像中心的接近程度，定义中心距离 d_i 为

$$d_i = \sqrt{\left(c_i^x - c_{\text{img}}^x\right)^2 + \left(c_i^y - c_{\text{img}}^y\right)^2}, \tag{4.8}$$

并将其归一化为中心位置得分 s_i^{center}：

$$s_i^{\text{center}} = -\frac{d_i}{\max(H, W)}. \tag{4.9}$$

由式 (4.9) 可知，候选区域越接近图像中心，其得分越高。

其次，考虑到生成目标通常呈细长形态，进一步定义长宽比偏好项。计算候选外接框的对称长宽比 r_i 为

$$r_i = \max\left(\frac{w_i}{h_i + \varepsilon}, \frac{h_i}{w_i + \varepsilon}\right), \tag{4.10}$$

其中 ε 是防止除零的极小常数。随后将其截断并归一化，得到长宽比得分 s_i^{aspect}：

$$s_i^{\text{aspect}} = \frac{\min(r_i, r_{\max})}{r_{\max}}, \tag{4.11}$$

其中 $r_{\max}$ 为长宽比上限，本章中设定 $r_{\max} = 8.0$。该定义使得细长的候选区域获得更高得分，但同时限制过大长宽比对总分的影响。

此外，自动分割模型还会在分割掩膜的同时输出每个候选区域的质量估计，例如预测交并比（predicted IoU）和稳定性得分（stability score）。记这两个分数分别为 u_i 和 v_i，则候选区域的质量得分 s_i^{quality} 定义为

$$s_i^{\text{quality}} = \alpha u_i + (1 - \alpha) v_i, \tag{4.12}$$

其中 $\alpha \in [0, 1]$ 为权重系数。本章中设定 $\alpha = 0.5$。

为了避免选择过度贴近图像边界的区域，引入边界惩罚项。设边界容差为 k，当候选外接框满足下列任一条件时，

$$x_i \le k, \quad y_i \le k, \quad x_i + w_i \ge W - k, \quad y_i + h_i \ge H - k, \tag{4.13}$$

则认为该候选区域与图像边界接触。其边界惩罚 s_i^{border} 定义为

$$s_i^{\text{border}} = \begin{cases} -\beta, & \text{若候选区域贴近图像边界,} \\ 0, & \text{否则,} \end{cases} \tag{4.14}$$

其中 $\beta > 0$ 为惩罚系数。本章中设定 $k = 5$，$\beta = 0.3$。

综合以上各项，候选掩膜 $\boldsymbol{M}_i$ 的总评分 S_i 如式 (4.15) 所示：

$$S_i = \lambda_1 s_i^{\text{quality}} + \lambda_2 s_i^{\text{center}} + \lambda_3 s_i^{\text{aspect}} + s_i^{\text{border}}, \tag{4.15}$$

其中 $\lambda_1, \lambda_2, \lambda_3$ 分别表示质量项、中心项和长宽比项的权重。本章中设定 $\lambda_1 = 1.2$，$\lambda_2 = 0.8$，$\lambda_3 = 0.8$。

最终，选择综合评分最高的候选掩膜作为主体目标区域，即选择候选掩膜 $\boldsymbol{M}^*$ 为

$$\boldsymbol{M}^* = \arg\max_{\boldsymbol{M}_i \in \mathcal{M}'} S_i. \tag{4.16}$$

该策略利用目标通常位于图像中心、具有细长外形且具备较高分割质量这一先验，对候选掩膜进行排序并选择最终主体区域。其效果将在第 4.2.3节的掩膜消融实验中给出。

4.1.4 轻量伪标签构建

在完成中间样本生成与主体区域确定之后，这些样本虽然在视觉上体现了由起始端点到目标端点的连续过渡，但并不具备人工标注信息，因此需要进一步为其构建可用于训练的伪标签。考虑到重量、关键表型特征和等级三类任务的监督来源并不相同，本章采用分任务的伪标签构建策略。其中，重量伪标签采用区间门控与预测补全相结合的方式生成；关键表型特征伪标签沿用第三章中定义的几何测量方法直接计算；等级伪标签则按照第三章构建数据集时的标注规则，根据中间样本的属性结果进行赋值。通过这种设计，可以使生成样本同时获得连续属性监督和离散分级监督，从而参与后续联合训练。

对于重量任务，本章不直接将端点重量在线性区间内平均分配给所有中间样本，也不完全依赖预测模型对全部样本直接赋值，而是采用一种结合先验约束与模型预测的两阶段伪标签构建方法。其基本考虑在于一方面，中间样本由两个真实端点样本生成，其合理重量通常应落在端点重量附近的有限范围内；另一方面，图像生成过程中可能出现局部形变，导致预测模型对某些中间样本给出偏离真实变化趋势的异常值。因此，若不加约束地直接使用预测结果，容易引入较大噪声。

设某一组端点样本分别为 $\boldsymbol{I}_U$ 和 $\boldsymbol{I}_W$，其真实重量标签分别为 w_U 和 w_W。设该组端点之间共生成 T 个中间样本，记为 $\{\boldsymbol{I}_i\}_{i=1}^{T}$，其中 i 为中间样本在序列中的索引。

首先定义该组样本的端点重量上下界 $w_{\min}$ 和 $w_{\max}$：

$$w_{\min} = \min(w_U, w_W), \qquad w_{\max} = \max(w_U, w_W). \tag{4.17}$$

考虑到生成样本可能存在轻微超出端点范围的合理波动，本章在端点区间基础上向两侧外扩容差值 δ_w：

$$\delta_w = 0.05\,(w_{\max} - w_{\min}), \tag{4.18}$$

从而构造重量门控区间 $\mathcal{G}_w$：

$$\mathcal{G}_w = [\, w_{\min} - \delta_w,\ w_{\max} + \delta_w \,]. \tag{4.19}$$

为保证后续补全过程可执行，将端点视为同一序列中的两个锚点，记

$$\hat{w}_0 = w_U, \quad \hat{w}_{T+1} = w_W. \tag{4.20}$$

其中，端点标签来自真实标注，直接作为固定监督，不参与门控筛选。

对于该组中的第 i 个中间样本 $\boldsymbol{I}_i$，先利用前一阶段在真实标注数据上训练得到的重量预测分支 $f_w(\cdot)$ 计算其候选预测值 $\tilde{w}_i$：

$$\tilde{w}_i = f_w(\boldsymbol{I}_i). \tag{4.21}$$

随后，根据门控区间 $\mathcal{G}_w$ 对候选预测值进行筛选。若预测结果满足式 (4.22)

$$\tilde{w}_i \in \mathcal{G}_w, \tag{4.22}$$

则认为该预测与端点先验一致，可直接作为该样本的重量伪标签 $\hat{w}_i$，即

$$\hat{w}_i = \tilde{w}_i. \tag{4.23}$$

若预测结果落在门控区间之外，则暂不立即赋值，而是将该位置记为待补全状态。这样做的原因在于，超出门控范围的预测值往往意味着模型输出与该组样本应有的连续变化趋势不一致，不能直接用于训练。

对于所有待补全位置，本章进一步利用其序列中最近的两个已接受伪标签进行线性插值补全。设第 i 个样本尚未获得有效重量标签，在索引集合 $\{0, 1, \ldots, T, T+1\}$

上寻找其前后方向最近的两个已赋值位置 p 和 q，其中 $p < i < q$，并已知对应伪标签分别为 $\hat{w}_p$ 和 $\hat{w}_q$。这里 p, q 可能取端点锚点 0 或 $T+1$。由于两端锚点 $\hat{w}_0$ 与 $\hat{w}_{T+1}$ 始终有效，上述 p, q 对任意待补全位置均存在，因此补全过程在本章设定下是闭合的。则第 i 个样本的补全重量 $\hat{w}_i$ 定义为

$$\hat{w}_i = \hat{w}_p + \frac{i-p}{q-p}\left(\hat{w}_q - \hat{w}_p\right). \tag{4.24}$$

式 (4.24) 的含义是：当某些位置的模型预测不可靠时，利用相邻可靠样本之间的线性趋势进行补全，通常比直接接受异常预测更稳定。

与重量不同，关键表型特征具有较强可解释性，因此本章不使用学习式预测构建关键表型特征伪标签，而是直接采用显式测量。对黄瓜，使用第三章定义的曲直度公式 $C = d/h$；对香蕉，同理使用第三章定义的暗斑比例 $m = N_b/N_a$ 作为成熟度。由此得到关键表型特征伪标签 $\hat{t}_i$：

$$\hat{t}_i = \begin{cases} C_i, & \text{黄瓜}, \\ m_i, & \text{香蕉}. \end{cases} \tag{4.25}$$

该定义与数据构建阶段保持一致，并且避免在小样本条件下额外引入回归噪声。

对于等级任务，本章同样不直接依赖分类模型对中间样本进行硬预测，而是沿用第三章构建数据集时所采用的等级标注规则，根据样本的属性结果生成对应的等级伪标签。若直接以分类模型输出作为伪标签，容易把模型自身偏差进一步放大；而若回到数据集构建时的原始规则，则可以保证中间样本的等级定义与真实样本保持一致。具体地，黄瓜样本按照曲直度阈值规则映射：$C > 20$ 为特级，$10 < C \leq 20$ 为一级，$5 < C \leq 10$ 为二级，$C \leq 5$ 为三级；香蕉样本按照成熟度阈值规则映射：$m \leq 0.025$ 为特级，$0.025 < m \leq 0.1$ 为一级，$0.1 < m \leq 0.2$ 为二级，$m > 0.2$ 为三级。

4.1.5 联合训练与损失函数

在完成中间样本生成、主体区域确定以及伪标签构建之后，本章将生成样本作为对原始训练集的补充数据来源，并与真实样本共同用于多任务模型训练。本章通过控制生成样本的注入比例来研究其对模型训练的影响。在构建完成全部伪标注中间样本之后，本章按照一定比例从生成数据中随机抽取样本，并与原始人工标注样

本共同组成扩展训练集。通过这种方式，可以在保留原始监督稳定性的基础上，逐步分析生成样本数量对模型性能的影响。实验部分将进一步比较不同生成数据注入比例下的结果，以观察中间样本在多任务学习中的实际贡献。

在训练过程中，原始样本和生成样本分别形成两部分监督信号。其中，原始样本对应的标签来自人工测量与人工标注，具有较高的可靠性；生成样本对应的标签则来自前文构建的伪标签，虽然能够提供额外的监督信息，但其准确性仍可能受到中间样本生成误差和伪标签误差的影响。因此，在损失设计上，本章对原始数据和生成数据分别建模，再通过加权方式进行联合优化。

设原始样本上的重量预测、关键表型特征预测和等级预测对应的损失分别记为 $\mathcal{L}_{\mathrm{w}}^{\mathrm{real}}$，$\mathcal{L}_{\mathrm{t}}^{\mathrm{real}}$，$\mathcal{L}_{\mathrm{g}}^{\mathrm{real}}$。其中，重量和关键表型特征任务采用 Huber 损失，等级任务采用交叉熵损失。则原始数据部分的多任务损失定义为

$$\mathcal{L}_{\mathrm{real}} = \frac{1}{2(\sigma_{\mathrm{w}}^{\mathrm{real}})^2}\mathcal{L}_{\mathrm{w}}^{\mathrm{real}} + \frac{1}{2(\sigma_{\mathrm{t}}^{\mathrm{real}})^2}\mathcal{L}_{\mathrm{t}}^{\mathrm{real}} + \log \sigma_{\mathrm{w}}^{\mathrm{real}} + \log \sigma_{\mathrm{t}}^{\mathrm{real}} + \mathcal{L}_{\mathrm{g}}^{\mathrm{real}}, \tag{4.26}$$

其中，$\sigma_{\mathrm{w}}^{\mathrm{real}}$ 和 $\sigma_{\mathrm{t}}^{\mathrm{real}}$ 为原始数据部分重量任务和关键表型特征任务对应的可学习不确定性参数。

类似地，设生成样本上的重量预测、关键表型特征预测和等级预测对应的损失分别记为 $\mathcal{L}_{\mathrm{w}}^{\mathrm{gen}}$，$\mathcal{L}_{\mathrm{t}}^{\mathrm{gen}}$，$\mathcal{L}_{\mathrm{g}}^{\mathrm{gen}}$。则生成数据部分的多任务损失定义为

$$\mathcal{L}_{\mathrm{gen}} = \frac{1}{2(\sigma_{\mathrm{w}}^{\mathrm{gen}})^2}\mathcal{L}_{\mathrm{w}}^{\mathrm{gen}} + \frac{1}{2(\sigma_{\mathrm{t}}^{\mathrm{gen}})^2}\mathcal{L}_{\mathrm{t}}^{\mathrm{gen}} + \log \sigma_{\mathrm{w}}^{\mathrm{gen}} + \log \sigma_{\mathrm{t}}^{\mathrm{gen}} + \mathcal{L}_{\mathrm{g}}^{\mathrm{gen}}, \tag{4.27}$$

其中，$\sigma_{\mathrm{w}}^{\mathrm{gen}}$ 和 $\sigma_{\mathrm{t}}^{\mathrm{gen}}$ 为生成数据部分对应的可学习不确定性参数。式 (4.26) 与式 (4.27) 中的不确定性参数彼此独立，并不共享。这是因为原始数据与生成数据在标签来源和噪声水平上存在差异，若共享同一组参数，往往会削弱损失加权对两类数据差异的适应能力。

在此基础上，本章将总损失定义为原始数据损失与生成数据损失的加权和，即

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{real}} + \gamma \mathcal{L}_{\mathrm{gen}}, \tag{4.28}$$

其中，γ 为生成数据损失的权重系数，用于控制伪标注中间样本在整体优化目标中的贡献大小。通过设置该系数，本章能够在真实监督与伪监督之间进行平衡，从而减弱生成数据可能带来的噪声干扰。在本章设定中，伪标注样本并不与真实样本采用同等监督强度，而是通过 $\gamma < 1$ 的全局加权进行降权；同时，原始数据与生成数据分

别使用独立的不确定性参数进行任务内自适应加权。因此，伪监督在训练中承担的是补充角色，而非与真实监督等置信度的替代角色。分类损失在式 (4.26) 与式 (4.27) 中保持固定权重 1，不再纳入不确定性加权，该设置与第三章保持一致，相比将两类样本直接合并为单一损失，本章方案仅增加少量标量参数，但可以显式区分真实监督与伪监督的噪声水平。

4.2 实验分析

4.2.1 实验设置

本章实验在第三章所构建的农业果蔬数据集基础上进行，并在此基础上引入由中间样本生成方法构造的伪标注样本。原始数据仍包含黄瓜与香蕉两类子集，其中每张图像均带有重量、关键表型特征和等级标签。黄瓜样本的关键表型特征定义为曲直度，香蕉样本的关键表型特征定义为成熟度。为保证与第三章结果具有可比性，本章在原始数据上的划分与第三章保持一致，并以原始人工标注样本作为基础监督数据。

在生成数据构建方面，本章按照第 4.1.3节和第 4.1.4节的流程处理：对每对端点生成 27 张中间图像，经 SAM2 分割与主体掩膜的筛选后，再构建重量、关键表型特征和等级伪标签。从生成样本中按 25% 比例随机抽取，与原始样本组成扩展训练集。

在模型训练阶段，本章仍采用第三章的多任务预测框架作为下游模型。输入图像统一调整为 224×224 像素，优化器采用 AdamW，初始学习率为 5×10^{-5}，批大小为 20，训练轮数为 100，在线增强采用不超过 15° 的随机旋转。生成数据损失权重系数固定为 $\gamma = 0.2$。所有实验在 NVIDIA RTX 3090 GPU 及 RTX 4090 GPU 上完成。

4.2.2 对比实验

第三章已经系统比较了提出的多任务框架与多种典型多任务学习基线方法，并验证了所提多任务模型结构在农业果蔬多属性联合预测任务中的有效性。因此，第四章的对比实验不再重复完整的多任务基线比较，而是重点考察在下游模型结构保持不变的条件下，引入中间样本生成与轻量伪标签构建后是否能够进一步提升模型性能。

本节将第三章提出的多任务模型作为本章的直接基线。比较对象包括仅使用原

始人工标注样本训练的第三章多任务模型，以及在相同模型结构基础上进一步引入生成样本进行联合训练的第四章方法。为便于与已有结果建立联系，本节同时选取第三章中两个具有代表性的外部多任务学习基线作为参考。

表 4.1 不同多任务学习方法以及使用生成数据在黄瓜测试数据上的对比

	重量		关键表型特征		等级		
模型	R^2	**RMSE**	R^2	**RMSE**	**Prec.**	**F1**	**Acc.**
MMoE	0.885 ± 0.217	8.515 ± 6.453	0.811 ± 0.046	6.619 ± 0.804	0.751 ± 0.244	0.756 ± 0.220	0.776 ± 0.193
MTAN	$\underline{0.958 \pm 0.019}$	$\mathbf{6.308 \pm 1.395}$	0.718 ± 0.091	8.028 ± 1.354	0.878 ± 0.070	0.867 ± 0.073	0.876 ± 0.070
第三章方法	$\mathbf{0.962 \pm 0.015}$	$\underline{7.394 \pm 1.740}$	$\underline{0.883 \pm 0.043}$	$\underline{3.954 \pm 1.205}$	$\underline{0.951 \pm 0.034}$	$\underline{0.944 \pm 0.040}$	$\underline{0.944 \pm 0.040}$
第三章方法 + 生成数据	0.934 ± 0.047	7.983 ± 2.574	$\mathbf{0.891 \pm 0.067}$	$\mathbf{3.371 \pm 1.295}$	$\mathbf{0.952 \pm 0.034}$	$\mathbf{0.945 \pm 0.038}$	$\mathbf{0.945 \pm 0.043}$

表 4.1 展示了不同多任务学习方法以及使用生成数据在黄瓜测试数据上的对比，结果基于十个随机种子，最佳结果用加粗表示，次优结果用下划线表示。从表中结果来看，第三章提出的模型在原始数据条件下已经优于两种代表性的多任务学习基线，尤其在关键表型特征回归和等级分类任务上表现出明显优势。其中，提出的方法在重量预测的决定系数上取得最高值 0.962，在关键表型特征回归上达到 0.883 的决定系数和 3.954 的均方根误差，在等级预测上也取得了较高的精确率、F1 值和准确率。这表明在当前黄瓜数据划分下，第三章提出的双分支多任务框架在几何回归与分类任务之间表现出较好的平衡性。进一步引入生成数据后，黄瓜数据上的结果呈现出较为清晰的任务差异。具体而言，关键表型特征回归和等级分类性能继续提升，其中关键表型特征决定系数从 0.883 提高到 0.891，均方根误差从 3.954 下降到 3.371，等级任务的精确率、F1 值和准确率也小幅提升到 0.952、0.945 和 0.945。该结果提示，在当前设置下，生成样本对黄瓜关键表型特征和等级任务可能提供了更直接的补充信息。另一方面，重量任务并未继续提升，加入生成数据后重量决定系数从 0.962 下降到 0.934，均方根误差也由 7.394 上升到 7.983。这一现象表明，生成样本对不同任务的作用并不一致，在本实验中其对关键表型特征和等级任务的增益更明显，而对重量回归的帮助相对有限。

对于黄瓜而言，中间样本生成主要改变的是外观形态和曲直状态，这类变化与关键表型特征和等级标签之间具有更直接的对应关系；而重量属于物理属性，其变化不仅受轮廓外观影响，还与真实体积、密度和采集样本分布有关，难以仅由生成图像中的视觉过渡充分表征。同时，本章重量伪标签由预测门控与序列插值补全得到，并非人工实测标签，因此在小样本条件下可能引入一定伪标签误差。当生成样

本参与训练后，这类误差可能对重量回归产生干扰，使其表现弱于原始监督条件下的结果。相比之下，关键表型特征和等级标签分别由显式测量和规则映射得到，监督信号与生成图像外观变化更一致，因此更容易从中间样本中获得增益。

表 4.2　不同多任务学习方法以及使用生成数据在 30 张香蕉测试数据上的对比

	重量		关键表型特征		等级		
模型	R^2	**RMSE**	R^2	**RMSE**	**Prec.**	**F1**	**Acc.**
MMoE	0.719 ± 0.131	$\mathbf{13.530 \pm 2.702}$	0.524 ± 0.345	0.095 ± 0.041	0.668 ± 0.329	0.679 ± 0.284	0.750 ± 0.205
MTAN	0.567 ± 0.287	16.356 ± 5.072	0.551 ± 0.219	0.097 ± 0.026	0.907 ± 0.060	0.899 ± 0.064	0.900 ± 0.062
第三章方法	$\mathbf{0.735 \pm 0.156}$	13.560 ± 4.490	0.744 ± 0.144	0.064 ± 0.013	0.902 ± 0.063	0.885 ± 0.064	$\mathbf{0.900 \pm 0.052}$
第三章方法 + 生成数据	0.727 ± 0.149	13.548 ± 4.511	$\mathbf{0.747 \pm 0.142}$	$\mathbf{0.063 \pm 0.020}$	$\mathbf{0.910 \pm 0.056}$	$\mathbf{0.899 \pm 0.062}$	0.900 ± 0.060

表 4.2 展示了不同多任务学习方法以及使用生成数据在 30 张香蕉测试数据上的对比，结果基于十个随机种子，最佳结果用加粗表示，次优结果用下划线表示。香蕉 30 张测试图像上的结果显示，在较小测试规模下，不同方法之间的任务偏向性依然较为明显。MMoE 在重量回归上表现最好，取得最低均方根误差 13.530；MTAN 在等级分类任务上具有优势，精确率和 F1 值分别达到 0.907 和 0.899。相比之下，第三章提出的模型在原始数据条件下取得了最高的重量决定系数 0.735，同时在关键表型特征回归上达到 0.744 的决定系数，并在准确率上与最优方法持平，体现出较好的多任务平衡能力。在此基础上，引入生成数据后，香蕉 30 张测试图像上的关键表型特征回归和等级分类性能继续改善。关键表型特征决定系数由 0.744 提高到 0.747，均方根误差由 0.064 下降到 0.063，等级任务中的精确率提高到 0.910，F1 值提高到 0.899，准确率维持在 0.900。上述结果提示，对于香蕉这类成熟度和等级变化依赖中间状态信息的对象，生成样本在小测试集上可能有助于成熟度回归与等级判别。不过，与黄瓜类似，重量任务并未从生成数据中获得进一步增益，重量决定系数略微下降到 0.727，虽然均方根误差相比原始模型略有下降，但整体改进幅度有限。这也说明在较小测试集条件下，生成样本对香蕉重量任务的帮助仍不够稳定。

表 4.3　不同多任务学习方法以及使用生成数据在 100 张香蕉测试数据上的对比

	重量		关键表型特征		等级		
模型	R^2	**RMSE**	R^2	**RMSE**	**Prec.**	**F1**	**Acc.**
MMoE	0.271 ± 0.267	14.315 ± 2.568	0.479 ± 0.261	0.086 ± 0.024	0.535 ± 0.279	0.498 ± 0.209	0.567 ± 0.156
MTAN	0.092 ± 0.506	15.704 ± 4.134	0.197 ± 0.354	0.108 ± 0.026	0.785 ± 0.040	0.720 ± 0.048	0.731 ± 0.045
第三章方法	0.451 ± 0.099	12.580 ± 1.120	0.662 ± 0.083	0.072 ± 0.009	0.758 ± 0.031	0.656 ± 0.028	0.677 ± 0.022
第三章方法 + 生成数据	$\mathbf{0.533 \pm 0.125}$	$\mathbf{10.578 \pm 1.273}$	$\mathbf{0.694 \pm 0.092}$	$\mathbf{0.070 \pm 0.017}$	$\mathbf{0.793 \pm 0.038}$	$\mathbf{0.737 \pm 0.046}$	$\mathbf{0.741 \pm 0.039}$

表 4.3 展示了不同多任务学习方法以及使用生成数据在 100 张香蕉测试数据上的对比，结果基于十个随机种子，最佳结果用加粗表示，次优结果用下划线表示。相比之下，香蕉 100 张测试图像上的结果更加清楚地体现了本章所引入生成数据的作用。在原始数据条件下，第三章提出的模型已经优于 MMoE 和 MTAN，特别是在重量和关键表型特征两个连续属性任务上分别取得了 0.451 和 0.662 的决定系数，同时均方根误差也低于两种基线。在此基础上，进一步引入生成数据后，模型在当前 100 张香蕉测试集上的各项指标均有提升。其中，重量决定系数从 0.451 提升到 0.533，均方根误差从 12.580 下降到 10.578，关键表型特征决定系数从 0.662 提升到 0.694，均方根误差从 0.072 下降到 0.070，等级分类的精确率、F1 值和准确率也分别提升到 0.793、0.737 和 0.741。该结果提示，本章提出的中间样本生成与轻量伪标签构建方法在该测试设置下可以作为补充监督来源，并在重量、关键表型特征和等级任务上带来一致方向的改进。对于连续属性任务，生成样本在端点之间补充了过渡状态样本；对于等级分类任务，生成数据也补充了类别边界附近样本。从中可以看出，香蕉 100 张测试图像上的结果支持生成数据对多任务预测具有正向作用。

4.2.3 消融实验

对比实验结果显示，本章提出的中间样本生成与轻量伪标签构建方法在部分任务上观察到性能提升。为了进一步分析各组成部分在整体方法中的具体作用，本节从主体掩膜选择策略和生成数据注入比例两个方面开展消融实验。前者用于考察中间样本分割后主体区域选择的合理性，后者用于分析生成数据参与训练的强度对模型性能的影响。

图 4.6中子图 (a) 和子图 (b) 分别展示了使用和不使用主体掩膜选择策略的生成掩膜的比较。可以看到，在使用主体掩膜选择策略后，子图 (a) 中的掩膜能够较完整地覆盖香蕉主体区域，背景区域被有效排除，目标轮廓也较为连续，能够为后续成熟度计算和伪标签构建提供较稳定的主体区域。相比之下，未使用该策略时，子图 (b) 中的掩膜明显偏离真实主体，模型选择了大面积背景区域作为分割结果，仅保留了少量香蕉局部区域。这类错误掩膜会直接影响目标面积、暗斑比例以及关键表型特征等后续计算结果，从而向伪标签构建阶段引入较大噪声。该结果说明，仅依赖自动分割模型输出的候选结果并不能保证最终掩膜一定对应果蔬主体。尤其在生成图像中，背景纹理、边界伪影或局部生成异常可能会被误识别为候选区域，导致主

(a) (b)

图 4.6 掩膜选择策略使用和不使用的比较

体提取失败。

考虑到由端点图像 $\boldsymbol{I}_0$ 和 $\boldsymbol{I}_{T+1}$ 通过 DiffMorpher 生成的一组中间样本缺乏真实分割标注，本章设计了一种无监督评分策略来评价和量化候选掩膜的质量，并将其作为不同评分配置比较的统一依据。设该组中间图像序列为 $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_T\}$，其中 T 表示该组生成样本的数量。对于第 i 个中间样本，记其候选掩膜为 $\boldsymbol{M}_i$。为避免与式 (4.15) 中的 $\lambda_1, \lambda_2, \lambda_3$ 混淆，本节将无监督质量评分权重记为 η_1, η_2, η_3，其质量分数 Q_i 定义为

$$Q_i = \eta_1 Q_i^{\text{sam}} + \eta_2 Q_i^{\text{conn}} + \eta_3 Q_i^{\text{smooth}}, \tag{4.29}$$

其中，Q_i^{sam} 表示 SAM2 的内部置信度得分，Q_i^{conn} 表示掩膜的主体连通性得分，Q_i^{smooth} 表示该掩膜在序列中的平滑性得分，且满足 $\eta_1 + \eta_2 + \eta_3 = 1$。本章中设定 $\eta_1 = 0.3$、$\eta_2 = 0.2$、$\eta_3 = 0.5$。

首先，SAM2 内部置信度项 Q_i^{sam} 定义为

$$Q_i^{\text{sam}} = \frac{s_i^{\text{iou}} + s_i^{\text{stab}}}{2}, \tag{4.30}$$

其中，s_i^{iou} 为 SAM2 输出的预测交并比，s_i^{stab} 为稳定性分数。该项用于反映模型对当前掩膜的基础置信程度。

其次，为了保证生成的掩膜主要对应单一主体区域，连通性项 Q_i^{conn} 定义为

$$Q_i^{\text{conn}} = \frac{A_i^{\max}}{A_i}, \tag{4.31}$$

其中，$A_i^{\max}$ 表示第 i 个掩膜的最大连通区域面积，A_i 表示该掩膜的总面积。若掩膜主要由一个完整连通区域构成，则该项取值接近于 1；若掩膜中存在较多碎块噪声，则该项会降低。

再次，考虑到中间样本由端点图像连续插值得到，相邻图像之间的目标外观变化通常较为平滑，因此序列平滑性项 Q_i^{smooth} 定义为

$$Q_i^{\text{smooth}} = \frac{1}{2}\left[\text{IoU}(\boldsymbol{M}_i, \boldsymbol{M}_{i-1}) + \text{IoU}(\boldsymbol{M}_i, \boldsymbol{M}_{i+1})\right], \tag{4.32}$$

其中，$\text{IoU}(\cdot,\cdot)$ 表示两个掩膜之间的交并比，$\boldsymbol{M}_{i-1}$ 和 $\boldsymbol{M}_{i+1}$ 分别表示第 $i-1$ 帧和第 $i+1$ 帧的掩膜。对于序列边界处的样本，仅与其唯一相邻帧计算一致性。

上述评分首先在每一组端点图像对应的生成序列内部进行。也就是说，单帧分数 Q_i 是针对某一组生成样本中的第 i 帧计算的。为了进一步评价整组生成序列的伪分割质量，组级评分 Q_{group} 定义为

$$Q_{\text{group}} = \frac{1}{T}\sum_{i=1}^{T} Q_i. \tag{4.33}$$

最后，若实验中包含 N 组端点图像，则可在所有组上进一步统计整体平均质量分数：

$$Q_{\text{all}} = \frac{1}{N}\sum_{n=1}^{N} Q_{\text{group}}^{(n)}, \tag{4.34}$$

其中，$Q_{\text{group}}^{(n)}$ 表示第 n 组生成序列的平均掩膜质量分数。这种评价方式用于在无真实标注条件下对 DiffMorpher 中间样本的伪分割结果进行定量分析。

为了验证第4.1.3节中主体掩膜选择策略各评分项的作用，本章进一步设计了主体掩膜评分配置的消融实验。该实验仅关注候选掩膜选择本身的质量，不涉及下游回归或分类任务指标。具体而言，在保持中间样本生成过程、候选掩膜来源以及图像序列不变的条件下，分别改变主体掩膜选择时的评分配置，并采用无标注条件下的掩膜质量指标对结果进行统一评价。根据前文定义，对于每一帧中间样本，主体掩膜质量由单帧评分 Q_i 表征，进一步在每组序列内部计算平均质量分数，并在全部

序列上统计总体结果。实验中同时报告三项分量得分，即 Q^{sam}、Q^{conn} 和 Q^{smooth}，分别对应 SAM2 内部置信度、主体连通性和序列平滑性；同时给出综合指标 Q，用于反映不同评分配置下主体掩膜选择的整体效果。Q 越高，说明该评分配置越有利于从候选集合中选出稳定且合理的主体掩膜。

考虑到当前主体掩膜选择策略主要由质量项、中心项、长宽比项和边界项构成，本节重点比较两类配置：一类是仅保留单一评分项的配置，用于分析各单项先验的独立作用；另一类是同时启用四个评分项的完整配置，用于评估多因素联合约束下的整体效果。表 4.4 和表 4.5 分别给出了黄瓜与香蕉中间样本上的总体汇总结果。

表 4.4　黄瓜中间样本上不同掩膜筛选配置的结果

质量项	中心项	长宽比项	边界项	Q^{sam}	Q^{conn}	Q^{smooth}	Q
✓				**0.9881**	**1.0000**	0.9590	0.9760
	✓			0.9773	0.9999	0.9345	0.9604
		✓		0.9658	**1.0000**	0.9557	0.9676
			✓	0.9750	**1.0000**	0.9057	0.9453
✓	✓	✓	✓	0.9814	**1.0000**	**0.9646**	**0.9768**

表 4.5　香蕉中间样本上不同掩膜筛选配置的结果

质量项	中心项	长宽比项	边界项	Q^{sam}	Q^{conn}	Q^{smooth}	Q
✓				**0.9861**	**1.0000**	0.9750	0.9833
	✓			0.9859	**1.0000**	0.9741	0.9828
		✓		0.9525	0.9998	0.9816	0.9765
			✓	0.9850	0.9999	**0.9846**	0.9878
✓	✓	✓	✓	**0.9861**	**1.0000**	**0.9846**	**0.9881**

从黄瓜数据的结果来看，完整配置在综合指标 Q 上取得最高值 0.9768，同时在序列平滑性得分 Q^{smooth} 上也达到最高值 0.9646。这提示对于黄瓜中间样本而言，同时结合质量项、中心项、长宽比项和边界项，更有利于选出序列连续性和整体合理性较好的主体掩膜。相比之下，仅保留单一评分项时，虽然不同配置在某些分量指标上仍能取得较高得分，但综合表现均略低于完整配置。其中，仅保留质量项时，Q^{sam} 达到最高值 0.9881，且综合指标 0.9760 也最接近完整配置，说明质量项是主体掩膜选择中较有效的单项依据。仅保留长宽比项时，综合指标为 0.9676，说明长宽比先验对黄瓜主体区域筛选具有一定作用。相比之下，仅保留中心项和边界项时，综合

指标分别下降到 0.9604 和 0.9453，表明这两类先验更适合作为辅助约束，而不适合单独作为主体选择依据。

从香蕉数据的结果来看，完整配置同样在综合指标 Q 上取得最高值 0.9881，说明多因素联合约束在香蕉中间样本上也表现出较好的整体效果。与黄瓜类似，仅保留质量项时综合指标达到 0.9833，最接近完整配置，说明质量项仍然是香蕉主体掩膜选择中的关键单项因素。仅保留边界项时，综合指标达到 0.9878，并且在序列平滑性得分 Q^{smooth} 上与完整配置同为最高值 0.9846，这提示对于香蕉样本，边界约束在抑制不合理候选区域、保持序列掩膜一致性方面作用更明显。相比之下，仅保留中心项时综合指标为 0.9828，略低于仅保留质量项；仅保留长宽比项时综合指标下降到 0.9765，同时 Q^{sam} 也明显降低到 0.9525，说明单独依赖长宽比先验时，候选掩膜的整体质量不如其他配置稳定。

综合两类数据的结果可以看出，完整配置均取得最高综合指标，提示主体掩膜选择不宜仅依赖单一先验，而更依赖多种约束共同作用。在所有单项配置中，质量项在黄瓜和香蕉数据上都表现出较强的独立判别能力；同时，边界项在香蕉数据上的作用更明显，而长宽比项在黄瓜数据上相对更有效。这种差异说明，不同对象类别在主体区域筛选时对先验信息的依赖程度并不完全一致。总体而言，表 4.4 和表 4.5 的结果支持本章采用完整掩膜筛选配置。

为了进一步分析生成数据在联合训练中的作用，对生成数据注入比例进行消融实验。在前文方法部分中，本章并未将全部生成样本固定加入训练，而是从已构建的伪标注中间样本中随机抽取一定比例，与原始人工标注样本共同组成扩展训练集。实验分别设置 25%、50%、75% 和 100% 四种生成数据注入比例，并在相同训练配置下进行比较。对于每一种比例，均从全部生成样本中随机抽取对应数量，与原始训练集共同用于模型训练。通过这种方式，可以观察当生成数据逐步增加时，模型在重量预测、关键表型特征回归和等级分类任务上的性能变化，从而分析生成样本数量与模型效果之间的关系。

表 4.6 展示了不同生成数据注入比例下黄瓜数据上的消融实验。从黄瓜数据的结果来看，不同生成数据注入比例对模型性能的影响较为明显，整体呈现出比例增大后下降的趋势。其中，在本章测试区间内，25% 注入比例在重量、关键表型特征和等级三个任务上均取得最佳结果。随着注入比例由 25% 提高到 50%、75% 和 100%，各项指标整体上持续下降，说明黄瓜任务对生成样本的使用强度较为敏感。这一现

表 4.6 不同生成数据注入比例下在黄瓜数据上的消融实验

	重量		关键表型特征		等级		
生成数据比例	R^2	RMSE	R^2	RMSE	Prec.	F1	Acc.
25%	**0.934 ± 0.047**	**7.983 ± 2.574**	**0.891 ± 0.067**	**3.371 ± 1.295**	**0.952 ± 0.034**	**0.945 ± 0.038**	**0.945 ± 0.043**
50%	0.896 ± 0.090	8.747 ± 2.923	0.863 ± 0.056	3.455 ± 1.452	0.918 ± 0.058	0.910 ± 0.075	0.912 ± 0.087
75%	0.883 ± 0.093	8.931 ± 2.977	0.849 ± 0.058	4.273 ± 1.517	0.866 ± 0.097	0.847 ± 0.096	0.854 ± 0.081
100%	0.850 ± 0.142	9.024 ± 3.061	0.838 ± 0.062	4.689 ± 1.448	0.855 ± 0.059	0.828 ± 0.056	0.839 ± 0.057

表 4.7 不同生成数据注入比例下在香蕉数据上的消融实验

	重量		关键表型特征		等级		
生成数据比例	R^2	RMSE	R^2	RMSE	Prec.	F1	Acc.
25%	**0.727 ± 0.149**	**13.548 ± 4.511**	**0.747 ± 0.142**	**0.063 ± 0.020**	**0.910 ± 0.056**	**0.899 ± 0.062**	**0.900 ± 0.060**
50%	0.725 ± 0.166	13.608 ± 4.789	0.723 ± 0.238	0.074 ± 0.025	0.900 ± 0.083	0.873 ± 0.068	0.885 ± 0.040
75%	0.709 ± 0.158	13.931 ± 4.894	0.731 ± 0.161	0.066 ± 0.031	0.885 ± 0.074	0.865 ± 0.094	0.876 ± 0.105
100%	0.694 ± 0.209	14.539 ± 4.944	0.718 ± 0.195	0.076 ± 0.028	0.852 ± 0.103	0.832 ± 0.172	0.844 ± 0.062

象提示，适量生成数据有助于补充端点之间的过渡状态，但当生成数据比例进一步增大时，伪标签误差和生成误差的累积会对训练造成干扰，进而削弱模型性能。尤其是关键表型特征回归和等级分类任务下降更明显，说明在当前数据设置下，保持原始人工标注监督主导并仅引入少量生成样本更稳妥。

表 4.7 展示了不同生成数据注入比例下 30 张香蕉数据上的消融实验。从香蕉数据的结果来看，也可以观察到类似趋势。在本章测试区间内，25% 注入比例在各项指标上同样表现最好。随着生成数据比例逐步增加，重量、关键表型特征和等级任务的指标整体上均出现下降。这现象表明，对于香蕉这类成熟度和等级变化依赖中间状态的对象，少量生成样本可补充表面状态过渡信息，但当生成样本占比过高时，训练过程对伪监督依赖增强，噪声样本的负面影响也随之加大，因此整体性能不再提升。

综合黄瓜和香蕉两组结果可以看出，生成数据的作用并不是简单地通过增加样本数量线性提升性能，而是在信息增益与监督噪声之间形成平衡。较低注入比例下，生成样本可补充原始数据中缺失的过渡状态；当注入比例过高时，伪标签误差累积会逐渐抵消这部分收益，使性能下降。两类数据在本章测试区间内均显示 25% 注入比例更稳妥。该结果也说明，中间样本生成与轻量伪标签构建虽能缓解小样本训练不足，但其使用强度仍需控制。

4.3 本章小结

针对小样本条件下端点之间过渡样本不足的问题，本章提出基于中间样本生成的多任务扩展学习方法，先在同类端点样本之间利用 DiffMorpher 生成中间图像，再结合 SAM2 与主体掩膜选择策略提取主体区域，并通过重量门控补全、显式关键表型特征测量及规则映射构建重量、关键表型特征和等级伪标签，最终与原始样本联合训练。在当前数据规模与实验协议下，该方法在黄瓜数据上的关键表型特征回归与等级分类表现更稳，在香蕉 100 张测试集上对重量、关键表型特征与等级三项指标均带来一定提升，且消融结果显示完整的掩膜选择配置与 25% 生成数据注入比例在本章测试区间内表现最佳，说明该方法可作为上一章原始监督的有效补充，但收益仍依赖生成样本质量与注入比例控制。

第五章　总结与展望

5.1　结论

本文围绕小样本条件下果蔬图像多属性联合预测问题，面向重量估计、关键表型特征回归与等级分类三类任务，完成了数据构建、多任务联合建模和扩展学习研究。结合全文方法与实验结果，结论如下：

（1）提出了基于多任务学习的果蔬多属性联合评估框架，并构建了 FruVegSet 多属性对齐数据集。该数据集在黄瓜与香蕉两类对象上建立了图像、连续属性和等级标签的一一对应关系，保证了回归与分类任务在同一数据基准上的可比性。方法层面通过任务路由、分支特征学习、跨分支交互和联合损失优化，缓解了连续属性回归与等级分类之间的任务干扰。在当前数据规模与实验协议下，方法在两类数据上均呈现出较稳定的综合表现，在保持回归任务可比结果的同时，对等级判别任务表现出更明显的增益。对比与消融实验表明，非共享双分支结构、关键表型分支中的多尺度增强以及跨分支融合模块，是维持三任务平衡表现的主要结构因素。

（2）在上述多任务框架基础上，提出了基于中间样本生成的多任务扩展学习方法，并验证了其在小样本条件下的补充监督作用。该方法通过同类端点样本配对生成中间过渡图像，结合主体掩膜筛选、重量门控补全和分任务伪标签映射，构建可用于联合训练的补充样本。在当前实验设置下，方法在关键表型特征回归与等级分类任务上表现出更稳定增益，并在更大测试规模下对三项任务呈现一致方向的改进趋势。生成数据注入比例消融显示，在本章测试区间内较低注入比例更稳妥，注入比例继续增大时会因伪标签噪声累积导致性能回落。该结果说明扩展监督能够有效补充原始监督，但最终收益仍与生成样本质量和注入强度紧密相关。

5.2　工作展望

尽管本文在小样本多属性联合预测上取得了阶段性结果，但在数据规模、监督质量和跨场景验证方面仍有改进空间。后续工作可围绕以下两个方向展开：

（1）扩展数据规模与外部验证范围。本文实验主要基于黄瓜与香蕉两类数据，样本规模与场景复杂度仍有限。后续可在更多类别、更多采集条件、更大样本规模以

及更新对比模型下进行验证，并逐步建立更统一的多属性标注规范与跨数据集评估协议，以进一步检验方法的泛化能力与应用可迁移性。

（2）提升扩展监督链路与联合训练过程的稳定性。可继续完善端点配对规则、生成样本质量控制和伪标签置信度筛选机制，减少高比例注入时的噪声累积；同时可针对真实样本与生成样本的权重分配、任务损失动态平衡和训练阶段调度开展更细粒度设计，以提高不同任务之间的协同效率和模型收敛稳定性，并推动方法由受控实验设置向更复杂应用环境过渡。

插图索引

表格索引

参考文献

[1] JARARWEH Y, FATIMA S, JARRAH M, et al. Smart and sustainable agriculture: Fundamentals, enabling technologies, and future directions[J/OL]. Computers and Electrical Engineering, 2023, 110: 108799. DOI: 10.1016/j.compeleceng.2023.108799.

[2] SACHITHRA V, PERERA H, MADHUSHANI R. How artificial intelligence is used to achieve agricultural sustainability: A systematic review[J/OL]. Smart Agricultural Technology, 2023, 4: 100123. DOI: 10.1016/j.atech.2023.100123.

[3] OLIVEIRA R C, ET AL. Artificial intelligence in agriculture: Benefits, challenges, and trends[J/OL]. Applied Sciences, 2023, 13(13): 7405. DOI: 10.3390/app131374055.

[4] AKKEM Y, BISWAS S K, VARANASI A. Smart farming using artificial intelligence: A review[J/OL]. Engineering Applications of Artificial Intelligence, 2023, 120: 105899. DOI: 10.1016/j.engappai.2023.105899.

[5] ASSIMAKOPOULOS F, VASSILAKIS C, MARGARIS D, et al. Ai and related technologies in the fields of smart agriculture: A review[J/OL]. Information, 2025, 16(2): 100. DOI: 10.3390/info16020100.

[6] ONYEAKA H, TAMASIGA P, NWAUZOMA U M, et al. Using artificial intelligence to tackle food waste and enhance the circular economy: Maximising resource efficiency and minimising environmental impact: A review[J/OL]. Sustainability, 2023, 15(13). DOI: 10.3390/su151310482.

[7] MILLER T, MIKICIUK G, DURLIK I, et al. The iot and ai in agriculture: The time is now—a systematic review of smart sensing technologies[J/OL]. Sensors, 2025, 25(12): 3583. DOI: 10.3390/s25123583.

[8] REITANO M, SEGOVIA M S, NAYGA R M. A systematic review on the impact of artificial intelligence in the agri-food supply chain[J/OL]. Food Policy, 2025, 137: 102983. DOI: 10.1016/j.foodpol.2025.102983.

[9] WAQAS M, NASEEM A, HUMPHRIES U W, et al. Applications of machine learning and deep learning in agriculture: A comprehensive review[J/OL]. Green Technologies and Sustainability, 2025, 3(3): 100199. DOI: 10.1016/j.grets.2025.100199.

[10] KATHARRIA A, PANT M, VELÁSQUEZ J D, et al. Information fusion in smart agriculture: Machine learning applications and future research directions[J/OL]. Information Fusion, 2026, 129: 104040. DOI: 10.1016/j.inffus.2025.104040.

[11] JIANG C, MIAO K, HU Z, et al. Image recognition technology in smart agriculture: A review of current applications, challenges and future prospects[J/OL]. Processes, 2025, 13(5): 1402. DOI: 10.3390/pr13051402.

[12] CHUQUIMARCA L E, VINTIMILLA B X, VELASTIN S A. A review of external quality inspection for fruit grading using cnn models[J/OL]. Artificial Intelligence in Agriculture, 2024: 1-20. DOI: 10.1016/j.aiia.2024.10.002.

[13] KOÇ S, KAYRA H. Non-destructive weight prediction model of spherical fruits and vegetables using u-net image segmentation and machine learning methods[J/OL]. Journal of Agricultural Sciences, 2024, 30(4): 735-747. DOI: 10.15832/ankutbd.1434767.

[14] IZQUIERDO P, ARAGÓN-ESPINOSA P, RELAÑO DE LA GUÍA E, et al. Fruit weight estimation using image-based deep learning for use in dietary assessment [J/OL]. Next Research, 2025, 2(4): 101069. DOI: 10.1016/j.nexres.2025.101069.

[15] MIRANDA J C, GENÉ-MOLA J, GREGORIO E, et al. Fruit sizing using ai: A review of methods and challenges[J/OL]. Postharvest Biology and Technology, 2023, 205: 112587. DOI: 10.1016/j.postharvbio.2023.112587.

[16] FENG J. Promising real-time fruit and vegetable quality detection technologies: A review[J/OL]. International Journal of Agricultural and Biological Engineering, 2024. DOI: 10.25165/j.ijabe.20241702.7678.

[17] CHANG S, ZHANG M, KONG D, et al. Artificial intelligence-driven postharvest quality control of fruit and vegetable products: Intelligent detection, optimization strategies and application prospects[J/OL]. Food Bioscience, 2025, 74: 108004. DOI: 10.1016/j.fbio.2025.108004.

[18] LIU J, SUN J, WANG Y, et al. Non-destructive detection of fruit quality: Technologies, applications and prospects[J/OL]. Foods, 2025, 14(12): 2137. DOI: 10.3390/foods14122137.

[19] WANG D, LI L, JIANG X, et al. Toward robust in-field fruit quality evaluation: A critical review of emerging nondestructive technologies and devices[J/OL]. Food Research International, 2025, 225: 118058. DOI: 10.1016/j.foodres.2025.118058.

[20] MURESAN H, OLTEAN M. Fruit recognition from images using deep learning[J]. Acta Universitatis Sapientiae, Informatica, 2018, 10(1): 26-42.

[21] OLTEAN M. Fruits-360 dataset[DB/OL]. 2018[2026-05-11]. https://doi.org/10.17632/rp73yg93n8.1.

[22] BARDA S, ADITYA, KINHA R, et al. Applev: A dataset for apple fruit volume estimation[C]//Proceedings of the Fifteenth Indian Conference on Computer Vision Graphics and Image Processing. 2024: 1-8.

[23] ABAYOMI-ALLI A, ET AL. Fruitq: A new dataset of multiple fruit images for freshness evaluation[J/OL]. Data in Brief, 2023, 49: 109309. DOI: 10.1016/j.dib.2023.109309.

[24] SHARAFUDEEN M, CHANDRA V. Fruitnet and fruitbox dataset[DB/OL]. Kaggle, 2023[2026-05-11]. https://www.kaggle.com/datasets/mirlab/annotated-fruitnet-and-fruitbox.

[25] MESHRAM V, PATIL K. Fruitnet: Indian fruits image dataset with quality for machine learning applications[J]. Data in Brief, 2022, 40: 107686.

[26] NAVARRO P J, MILLER L, DÍAZ-GALIÁN M V, et al. A novel ground truth multispectral image dataset with weight, anthocyanins, and brix index measures of grape berries tested for its utility in machine learning pipelines[J]. GigaScience, 2022, 11: giac052.

[27] JAFARY P, BAZANGEYA A, PHAM M, et al. Raspberry phenoset: A phenology-based dataset for automated growth detection and yield estimation[EB/OL]. 2024 [2026-05-11]. https://arxiv.org/abs/2411.00967.

[28] ZAHURUL HAQUEA M, SARKER Y, FARHAN SADIQUE MAHI M, et al. Dragonfruitqualitynet: A lightweight convolutional neural network for real-time dragon fruit quality inspection on mobile devices[EB/OL]. 2025[2026-05-11]. https://arxiv.org/abs/2508.07306.

[29] AHMED M, ET AL. Few-shot learning for avocado fruit maturity classification under limited data conditions[J]. Computers and Electronics in Agriculture, 2024, 221: 109001.

[30] FU X, ET AL. Improved small-sample green fruit detection based on data augmentation and deep learning[J]. Computers and Electronics in Agriculture, 2024, 219: 108857.

[31] YUAN Y, CHEN Y. Vegetable and fruit freshness detection based on deep features and principal component analysis[J/OL]. Current Research in Food Science, 2024, 8: 100656. DOI: 10.1016/j.crfs.2023.100656.

[32] AKTER T, BHATTACHARYA T, KIM J H, et al. A comprehensive review of external quality measurements of fruits and vegetables using nondestructive sensing technologies[J/OL]. Journal of Agriculture and Food Research, 2024, 15: 101002. DOI: 10.1016/j.jafr.2024.101002.

[33] GUO K, CHEN H, ZHENG Y, et al. Efficient adaptive incremental learning for fruit and vegetable classification[J/OL]. Agronomy, 2024. DOI: 10.3390/agronomy14061275.

[34] ZHANG L, ZHANG M, MUJUMDAR A S, et al. Advanced model predictive control strategies for nondestructive monitoring quality of fruit and vegetables during supply chain processes[J/OL]. Computers and Electronics in Agriculture, 2024, 225: 109262. DOI: 10.1016/j.compag.2024.109262.

[35] 任永新, 单忠德, 张静, 等. 计算机视觉技术在水果品质检测中的研究进展[J/OL]. 中国农业科技导报, 2012, 14(1): 98-103. DOI: 10.3969/j.issn.1008-0864.2012.01.14.

[36] 何文斌, 魏爱云, 明五一, 等. 基于机器视觉的水果品质检测综述[J/OL]. 计算机工程与应用, 2020, 56(11): 10-16. DOI: 10.3778/j.issn.1002-8331.2001-0248.

[37] 刘妍, 周新奇, 俞晓峰, 等. 无损检测技术在果蔬品质检测中的应用研究进展[J/OL]. 浙江大学学报 (农业与生命科学版), 2020, 46(1): 27-37. DOI: 10.3785/j.issn.1008-9209.2019.09.241.

[38] 张哲, 马斌, 罗娜, 等. 果蔬农产品产地智能化处理技术研究进展与展望[J/OL]. 农业机械学报, 2025, 56(6): 1-16,46. DOI: 10.6041/j.issn.1000-1298.2025.06.001.

[39] 贾志鑫, 杨霖, 史策, 等. 农产品品质在线感知技术应用研究进展[J/OL]. 农业机械学报, 2025, 56(6): 17-32. DOI: 10.6041/j.issn.1000-1298.2025.06.002.

[40] PRABHU K, ET AL. Deep learning-based mango weight estimation using image-derived geometric features[J]. Computers and Electronics in Agriculture, 2023, 206: 107676.

[41] SABOURI A, BAKHSHIPOUR A, POORSALEHI M, et al. Machine learning techniques for non-destructive estimation of plum fruit weight[J/OL]. Scientific Reports, 2025, 15(1): 751. DOI: 10.1038/s41598-024-85051-2.

[42] 倪勤越, 吕军, 罗均毅, 等. 基于边缘计算的超市水果智能电子秤[J/OL]. 建模与仿真, 2024, 13(3): 3745-3753. DOI: 10.12677/mos.2024.133342.

[43] JERRIPOTHULA K R, SHUKLA S K, JAIN S, et al. Fruit maturity recognition from agricultural, market and automation perspectives[C]//IECON 2021–47th Annual Conference of the IEEE Industrial Electronics Society. IEEE, 2021: 1-6.

[44] WU G, ET AL. Using color and 3d geometry features to segment fruit point clouds for precision agriculture[J/OL]. Computers and Electronics in Agriculture, 2020, 174: 105469. DOI: 10.1016/j.compag.2020.105469.

[45] 赵泽华, 王庆艳, 陈俊杰, 等. 基于机器视觉与高光谱成像的西瓜考种系统研究[J/OL]. 农业机械学报, 2025, 56(12): 450-459. DOI: 10.6041/j.issn.1000-1298.2025.12.041.

[46] JIA W, ZHANG Z, SHAO W, et al. Foveamask: A fast and accurate deep learning model for green fruit instance segmentation[J/OL]. Computers and Electronics in Agriculture, 2021, 191: 106488. DOI: 10.1016/j.compag.2021.106488.

[47] DAI M, DORJOY M M H, MIAO H, et al. A new pest detection method based on improved yolov5m[J]. Insects, 2023, 14(1): 54.

[48] HAN Y, HUANG Z, SUN Y, et al. Agricultural object detection in complex environments via co-attention and self-knowledge distillation[J]. Information Sciences, 2025: 122711.

[49] GOMAA A. Advanced domain adaptation technique for object detection leveraging semi-automated dataset construction and enhanced yolov8[C]//2024 6th novel intelligent and leading emerging sciences conference (NILES). IEEE, 2024: 211-214.

[50] GOMAA A, ABDALRAZIK A. Novel deep learning domain adaptation approach for object detection using semi-self building dataset and modified yolov4[J]. World Electric Vehicle Journal, 2024, 15(6): 255.

[51] GOMAA A, SAAD O M. Residual channel-attention (rca) network for remote sensing image scene classification[J]. Multimedia Tools and Applications, 2025: 1-25.

[52] PAUDEL D, BOOGAARD H, DE WIT A, et al. Machine learning for regional crop yield forecasting in europe[J]. Field Crops Research, 2022, 276: 108377.

[53] HASSAN S, KHAN Z, SINGH A, et al. Ai-powered crop monitoring for precision agriculture[C]//International Symposium on Distributed Computing and Artificial Intelligence. Springer, 2024: 261-270.

[54] HASSAN O F, IBRAHIM A F, GOMAA A, et al. Real-time driver drowsiness detection using transformer architectures: a novel deep learning approach[J]. Scientific Reports, 2025, 15(1): 17493.

[55] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[C]// Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1314-1324.

[56] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2117-2125.

[57] PARK J, WOO S, LEE J Y, et al. Bam: Bottleneck attention module[C]//Proceedings of the British Machine Vision Conference. 2018.

[58] GUO M H, LU C Z, LIU Z N, et al. Visual attention network[J]. Computational Visual Media, 2023, 9(4): 733-752.

[59] CARUANA R. Multitask learning[J]. Machine Learning, 1997, 28(1): 41-75.

[60] MISRA I, SHRIVASTAVA 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.

[61] KENDALL A, GAL Y, CIPOLLA 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.

[62] GUO P, LEE C Y, ULBRICHT D. Learning to branch for multi-task learning[C]// International conference on machine learning. PMLR, 2020: 3854-3863.

[63] HAZIMEH H, ZHAO Z, CHOWDHERY A, et al. Dselect-k: Differentiable selection in the mixture of experts with applications to multi-task learning[J]. Advances in Neural Information Processing Systems, 2021, 34: 29335-29347.

[64] MA J, ZHAO Z, YI X, et al. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts[C]//Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. 2018: 1930-1939.

[65] LIU S, JOHNS E, DAVISON A J. End-to-end multi-task learning with attention[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 1871-1880.

[66] ZHANG Y, YANG X, CHENG Y, et al. Fruit freshness detection based on multi-task convolutional neural network[J/OL]. Current Research in Food Science, 2024, 8: 100733. DOI: 10.1016/j.crfs.2024.100733.

[67] 熊玮, 周礼辰, 余豪, 等. 薄皮果蔬分选技术与设备研究进展与展望[J/OL]. 农业机械学报, 2025, 56(8): 665-683. DOI: 10.6041/j.issn.1000-1298.2025.08.063.

[68] KHATUN T, NIROB M A S, BISHSHASH P, et al. A comprehensive dragon fruit image dataset for detecting the maturity and quality grading of dragon fruit[J/OL]. Data in Brief, 2024, 52: 109936. DOI: 10.1016/j.dib.2023.109936.

[69] ABAYOMI-ALLI O O, DAMAŠEVIČIUS R, MISRA S, et al. Fruitq: A new dataset of multiple fruit images for freshness evaluation[J/OL]. Multimedia Tools and Applications, 2024, 83: 11433-11460. DOI: 10.1007/s11042-023-16058-6.

[70] MALOUNAS I, VIERBERGEN W, KUTLUK S, et al. Spectrofood dataset: A comprehensive fruit and vegetable hyperspectral meta-dataset for dry matter estimation [J/OL]. Data in Brief, 2024, 52: 110040. DOI: 10.1016/j.dib.2024.110040.

[71] SHARAFUDEEN M, ET AL. An intelligent framework for estimating grade and quantity of tropical fruits in a multi-modal supply chain traceability system[J/OL]. Engineering Applications of Artificial Intelligence, 2023, 126: 107193. DOI: 10.1016/j.engappai.2023.107193.

[72] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[73] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks[C]// European Conference on Computer Vision. 2016: 630-645.

[74] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//Proceedings of the 32nd International Conference on Machine Learning. 2015: 448-456.

[75] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 1026-1034.

[76] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.

[77] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision. 2018: 3-19.

[78] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Advances in Neural Information Processing Systems: Vol. 30. 2017.

[79] MCNEMAR Q. Note on the sampling error of the difference between correlated proportions or percentages[J/OL]. Psychometrika, 1947, 12(2): 153-157. DOI: 10.1007/BF02295996.

[80] WILCOXON F. Individual comparisons by ranking methods[J/OL]. Biometrics Bulletin, 1945, 1(6): 80-83. DOI: 10.2307/3001968.

[81] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.

[82] 黄瓜等级规格: NY/T 1587-2008[S]. 2008.

[83] 香蕉等级规格: NY/T 3193-2018[S]. 2018.

[84] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization[EB/OL]. CoRR, 2017, abs/1711.05101[2026-05-11]. http://arxiv.org/abs/1711.05101.

[85] CHUQUIMARCA L, VINTIMILLA B, VELASTIN S. Banana ripeness level classification using a simple cnn model trained with real and synthetic datasets[EB/OL]. 2025[2026-05-11]. https://arxiv.org/abs/2504.08568.

[86] HAYAT A, MORGADO-DIAS F, CHOUDHURY T, et al. Fruitvision: A deep learning based automatic fruit grading system[J]. Open Agriculture, 2024, 9(1): 20220276.

[87] CARUANA R. Multitask learning: A knowledge-based source of inductive bias[C]// Machine learning: Proceedings of the tenth international conference. 1993: 41-48.

[88] ZHANG K, ZHOU Y, XU X, et al. Diffmorpher: Unleashing the capability of diffusion models for image morphing[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 7912-7921.

[89] RAVI N, GABEUR V, HU Y T, et al. Sam 2: Segment anything in images and videos [EB/OL]. 2024[2026-05-11]. https://arxiv.org/abs/2408.00714.

攻读硕士学位期间取得的研究成果

一、论文

1. Han Y, Ge J, Sun Y, Zhao T, Chen Q*. A multi-task learning framework for integrated assessment in agricultural applications[J]. Information Sciences, 2026: 123283. （已发表，导师一作，本人二作，中科院二区，JCR 一区）

二、软著

1. 软件名称：基于图像处理的黄瓜重量预测和分级平台 V1.0，开发人：韩越兴，葛嘉浩，王冰，登记号：2024SR1573970，登记日期：2024.10.21，申请人：上海大学。（导师一作，本人二作）

2. 软件名称：基于深度学习方法的统一果蔬重量估计和品质分级平台 V1.0，开发人：韩越兴，葛嘉浩，陈侨川，孙妍，登记号：2025SR2154617，登记日期：2025.11.05，申请人：上海大学。（导师一作，本人二作）

致　谢

时光匆匆，硕士阶段的学习与研究工作即将画上句号。回顾这段经历，从课程学习、课题开展到论文撰写，我在不断探索与反复修改中逐步完成了本论文。这个过程离不开许多老师、同学、家人和朋友的帮助与支持。在论文完成之际，我谨向所有关心、帮助过我的人表示诚挚的感谢。

首先，我要衷心感谢我的导师韩越兴老师。导师在课题和研究等方面都给予了我耐心而细致的指导。在我研究遇到困难时，导师总能给予我启发和帮助，使我能够逐步理清问题、改进方法并完成研究工作。导师严谨认真的治学态度使我受益良多，也将对我今后的学习和工作产生长远影响。

其次，我要感谢课题组的陈侨川老师和孙妍老师。在平时的学习与科研交流中，给予了我很多帮助。无论是在研究方面，还是在论文修改过程中，我都从与你们的讨论中获得了很多有价值的建议。

此外，我要感谢我的父母。感谢你们一直以来给予我的理解、包容与支持。在我面对压力和困难时，家人始终是我最坚实的后盾。我还要感谢家中最活泼的小伙伴，玄凤鹦鹉“小六”，尽管你喜欢在家中乱飞，以及不定时不定点地到处留下你消化过的食物的痕迹，但你的可爱与陪伴让我感到生活更加有趣、平淡的生活中也充满了许多活力。

最后，我要感谢师兄师姐们在我硕士阶段给予我的帮助和经验，也感谢学弟学妹们在学习、科研和日常生活中的交流与陪伴。感谢同级朋友们在科研和生活上的互帮互助和欢声笑语，让我觉得在平凡的生活中也有同行者。此外，也感谢一直约饭的饭搭子和朋友们，让我在硕士阶段的生活中吃好喝好，感受到了富足的营养转化为了体重秤上上涨的数字的同时，也享受到美食带来的快乐。

再次向所有帮助过我的人致以最诚挚的感谢。夏天就要来临，梅雨季将离去。愿所有在硕士阶段帮助和陪伴过我的人和小伙伴永远安康，愿自己永远懂得飞翔！

葛嘉浩

上海大学

2026 年 4 月 15 日