

基于尺度交互蒸馏网络的薄壳山核桃品种识别方法

赵 宁¹, 陈智坤^{1*}, 杨朋飞¹, 王瑞多², 张计育³, 李永荣⁴

(1. 陕西省西安植物园 陕西省植物研究所, 西安 710061; 2. 中国科学院西安光学精密机械研究所光子功能材料与器件研究室, 西安 710119; 3. 江苏省中国科学院植物研究所, 南京 210014; 4. 江苏省南京绿宙薄壳山核桃科技有限公司, 南京 210007)

摘 要: 薄壳山核桃是一种重要的经济坚果, 由于品种繁多, 对其进行快速科学的鉴定是进行种质资源保护与品种选育的重要基础。为了实现薄壳山核桃品种的快速鉴定, 该研究针对品种鉴定提出了基于尺度交互蒸馏网络的薄壳山核桃品种识别方法, 通过学习薄壳山核桃的可判别性特征实现品种鉴定。研究选择波尼等 12 种薄壳山核桃, 建立了 9 048 张实拍图像的品种识别数据集; 针对薄壳山核桃图片取样中距离变化导致的目标尺度多样性问题, 设计了一种全局-局部特征协同学习方案, 用于提取尺度不变特征; 与此同时, 该研究结合尺度知识蒸馏方案, 通过训练提取的不同尺度数据进行预测保证模型训练的有效性。结果表明, 通过训练该方法对上述 12 个品种的薄壳山核桃品种识别准确率均达到了 96.98%, 显著提高了薄壳山核桃的品种鉴定准确率。该研究开发的薄壳山核桃品种自动识别模型对于未来果实鉴定及产品分选提供了技术手段。

关键词: 薄壳山核桃; 品种鉴定; 尺度交互蒸馏; 多尺度上下文注意融合; 知识蒸馏

doi: 10.11975/j.issn.1002-6819.202408186

中图分类号: S126

文献标志码: A

文章编号: 1002-6819(2025)-05-0209-08

赵宁, 陈智坤, 杨朋飞, 等. 基于尺度交互蒸馏网络的薄壳山核桃品种识别方法[J]. 农业工程学报, 2025, 41(5): 209-216. doi: 10.11975/j.issn.1002-6819.202408186 <http://www.tcsae.org>

ZHAO Ning, CHEN Zhikun, YANG Pengfei, et al. Identification method for *Carya illinoensis* varieties based on scale interactive distillation network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2025, 41(5): 209-216. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202408186 <http://www.tcsae.org>

0 引 言

薄壳山核桃 (*Carya illinoensis*), 属胡桃科山核桃属植物, 原产于美国, 是世界著名的坚果及木本油料树种^[1]。其果实又名美国山核桃、碧根果, 具有营养价值高、果型大、果皮薄、出仁率高、风味好等特点, 深受消费者的喜爱^[2-3]。中国引进薄壳山核桃已有 120 多年的历史, 目前该产业主要分布在江苏、浙江、云南、安徽、陕西等地, 总面积 80 000 hm² 以上^[4], 坚果年产量约为 1 000 t 左右^[3]。

薄壳山核桃品种繁多, 达上千种, 其中用于生产的有 40 多种。中国栽植的主要品种为国外引进, 由于各地气候条件不同, 适宜生产的品种也存在差异^[5]。准确快速地识别品种对品种选育意义重大^[2]。不同品种坚果形态及坚果品质均有差异。尽管国际植物新品种保护联盟发布的《山核桃属薄壳山核桃品种特异性、一致性、稳定性测试指南》中将坚果形状、大小、颜色、纵横比、果尖长等作为薄壳山核桃品种测试的重要指标^[6]。但是,

由于薄壳山核桃的种植区域分散、引种时名称翻译不规范等原因, 易造成品种混淆, 导致品种鉴定工作面临挑战。

早期的薄壳山核桃品种识别主要依靠人工观察坚果的外形、大小、颜色等特征进行分类。然而这类方法需要鉴定人员具有丰富的实践经验, 存在效率低、准确性差等缺点。目前已有学者利用 DNA 分子标记方法构建薄壳山核桃主要品种的分子信息^[7], 用于薄壳山核桃品种鉴定, 然而此方法成本大且时效性差, 不宜用于实际生产。因此, 开发一种快速、廉价、环保的薄壳山核桃种子品种鉴定方法具有重要的研究及产业意义^[8]。

近年来, 国内外学者开展了大量基于机器视觉的果实品种识别的研究^[9-10], 主要分为基于传统机器学习和基于深度学习的果实品种识别方法。传统基于机器学习的果实品种识别方法主要采用果实的颜色、纹理、形状等特征, 利用不同的算法和分类器来提取和识别这些特征, 如支持向量机、Hough 变换等。ANUJA 等^[11]提出了基于机器学习的多种水果自动检测与分级方法, 使用支持向量机等 4 种不同的分类器, 对苹果等 4 种不同的水果进行分级检测, 水果检测准确率达 80% 以上。这些方法虽然在一定程度上可以实现品种识别, 但缺点是特征提取器的构建需要大量的人工干预和专业知识, 特征提取的效果受到图像质量和环境因素的影响, 分类器的泛化能力和准确率不高, 检测速度不够快等。深度学习模型在水果分类^[12-13]、植物鉴定^[14-15]、作物分类^[16-17]、杂草分类^[18-19]、病虫害监测^[20-21]、质量评价^[22-24] 等多个农业

收稿日期: 2024-08-26 修订日期: 2024-12-02

基金项目: 陕西省重点研发计划项目 (2022FP-03), 陕西省林业科技重点专项 (SXLK2022-02-5); 陕西省科学院科技计划项目 (2024K-04)

作者简介: 赵宁, 博士生, 助理研究员, 研究方向为薄壳山核桃引种选育。

Email: zn1002618336@xab.ac.cn

*通信作者: 陈智坤, 博士, 副研究员, 研究方向为薄壳山核桃种质资源研究与创新。Email: zhikunchen@ms.xab.ac.cn

领域得到了广泛应用,并取得了显著性能。KATARZYNA 等^[25]开发了一个卷积神经网络(convolutional neural network, CNN)模型,可以在对 6 个苹果品种进行分类,准确率达到 99.78%。MURESAN 等^[26]采用迁移学习策略,比较了 AlexNet、GoogLeNet 等预训练模型在水果识别任务上的性能。ZHU 等^[27]利用 AlexNet 网络模型对蔬菜图像进行分类,并研究了不同数量的蔬菜图像对 CNN 模型的影响;与此同时,基于 CNN 与反向传播神经网络和支持向量机(support vector machine, SVM)分类器进行了准确率比较。张三林等^[8]提出了一种基于 YOLOv5 的核桃品种识别与检测方法。该方法构建了核桃检测数据集,并在此基础上训练 YOLOv5 模型。试验结果表明,该方法在核桃品种识别与检测方面具有较高的精度和速度,为机器人分拣核桃提供了技术支撑。QIAN 等^[28]通过利用水稻种子表面三维点云数据,提出一种基于深度学习的水稻品种快速准确识别的分类方法。为应对在胡麻种子自动化考种过程中遇到的数据统计准确率低的挑战,毛永文等^[29]提出了一种基于深度神经网络的胡麻种子智能考种方法,对不同品种胡麻种子的识别准确率达 97.28%。

虽然深度学习模型在多种农业应用中取得了较大成功,但针对薄壳山核桃品种识别这一特定研究对象,由于果实类别间差异细微,且易受光照、拍摄角度和距离等外部因素干扰,鲜有专门进行薄壳山核桃品种识别的模型。为此,本文提出了一种基于尺度交互蒸馏网络的薄壳山核桃品种识别方法,学习具有判别性的关键特征作为快速品种鉴别的依据。本文提出的方法主要由全局分支和局部分支组成,分别用于学习全局和局部上下文特征;通过多尺度上下文注意力融合模块,融合全局和局部上下文特征,以期学习具有判别性的关键特征用于薄壳山核桃品种识别。

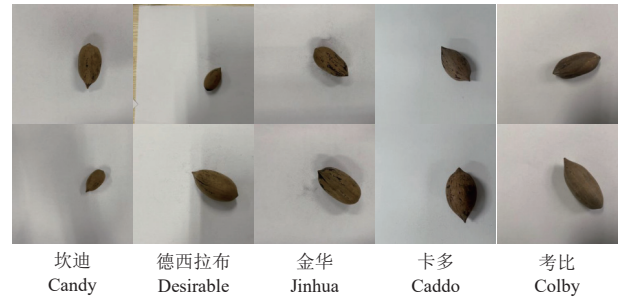
1 试验与方法

1.1 图像采集与预处理

以收获期去除果皮的坎迪 Candy、德西拉布 Desirable、卡多 Caddo、考比 Colby、马罕 Mahan、萨婆 Sauber、特贾斯 Tejas、威奇塔 Wichita、金华 Jinhua、绿亩 1 号 Lvzhou 1、亚林 100Yalin100、钟山 25 Zhongshan25 等 12 个品种薄壳山核桃为样本,每个品种随机抽取 30 个样本(由南京绿亩薄壳山核桃科技有限公司的李永荣研究员鉴定),构建了一个薄壳山核桃品种识别数据集。构建数据集过程中考虑了薄壳山核桃光照、角度和拍摄距离等因素,从多角度拍摄了 12 个品种的薄壳山核桃图像,拍摄图像尺寸为 1 000 像素×2 000 像素,获得符合试验条件的图像共 2 262 张,其中马罕和威奇塔各 186 张,其他类别各 189 张。

针对获得的 2 262 张薄壳山核桃图像,依次对每张图像进行了类别标注。同时,为了增大数据集规模以避免训练过拟合现象,对获得的图像进行了随机裁剪、随机旋转、添加噪声等处理,最终将数据集规模扩充到 9 048

张,按照 8:1:1 的比例划分训练集、验证集和测试集。图 1 为构建薄壳山核桃品种识别数据集部分代表性样本。



注:每个品种各展示 2 个样本,上下行属于同一品种。

Note: Each variety displays 2 samples, and the upstream and downstream belong to the same variety.

图 1 薄壳山核桃品种识别数据集部分代表性样本

Fig.1 Some representative samples of pecan nut recognition dataset

1.2 基于尺度交互蒸馏网络的品种识别方法框架

图 2 所示为本文所述基于尺度交互蒸馏网络的薄壳山核桃品种识别方法框架图。该方法的特别之处在于采用双分支架构学习不同尺度输入的全局和局部上下文特征。其中,多尺度上下文注意力融合模块旨在整合全局和局部语义特征,用来增强不同尺度和不同空间位置之间的信息交互。此外,引入了尺度间知识蒸馏方案旨在多尺度预测结果的一致性。

1.3 全局和局部上下文特征提取

1.3.1 全局分支和局部分支

目前主流的深度学习模型通过叠加多个神经网络层来提取全局上下文特征,大大提高分类模型的性能。随着网络层数的增加,需要学习的参数急剧增加,从头开始学习参数需要大量带有标注信息的数据。目前还没有人工标注的专门针对薄壳山核桃品种识别大规模图像数据集。为了解决这一问题,采用迁移学习策略提取高质量的全局上下文语义特征。具体来说,本文采用预训练的 ResNet18 中的卷积层作为全局分支的主要架构。用 I_g 表示输入的全局图像,则全局上下文特征 $F_{global} \in R^{H \times W \times C}$ 可以通过以下方式获得:

$$F_{global} = f_g(I_g; \theta_g), \quad (1)$$

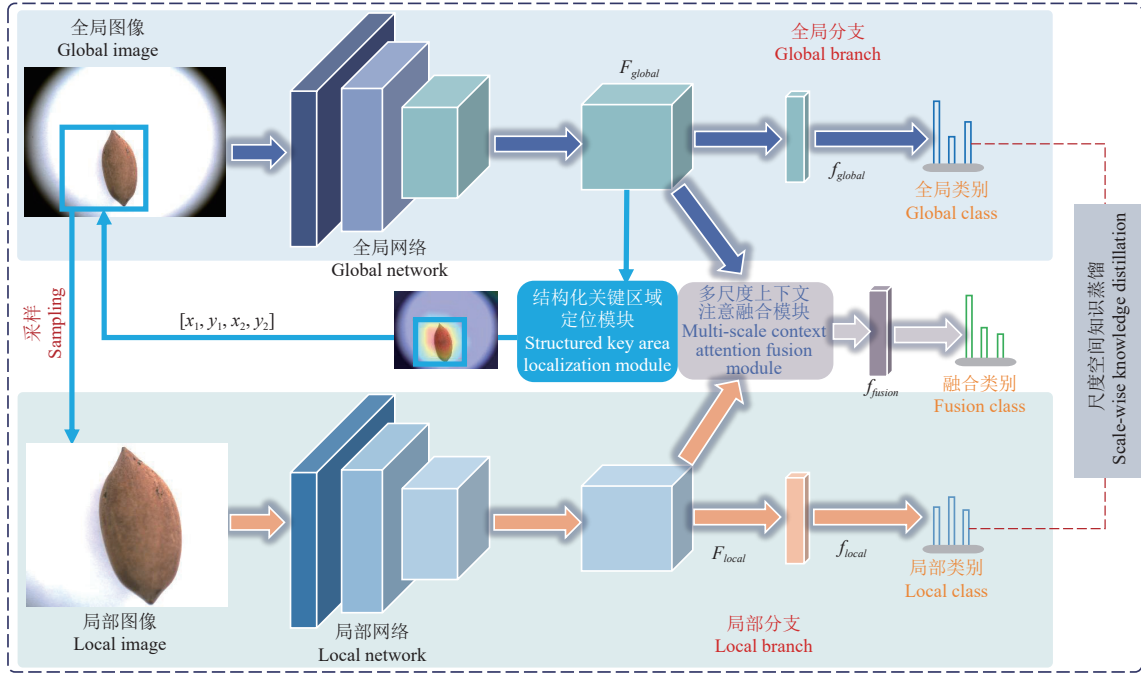
其中 f_g 表示全局分支子网络, θ_g 表示 f_g 中待学习的参数。局部分支与全局分支共享相同的网络结构,局部上下文特征 F_{local} 可以通过以下方式获得:

$$F_{local} = f_l(I_l; \theta_l), \quad (2)$$

其中 I_l 表示输入的局部图像, f_l 代表局部分支子网络, θ_l 表示 f_l 中待学习的参数。

1.3.2 结构化关键区域定位模块

为了采样得到局部感兴趣目标图像 I_l , 本文利用结构化关键区域定位模块对局部目标进行定位^[30]。该模块由能量聚合、能量图构造和贪婪边界搜索 3 个连续步骤组成,用于生成边界框并对局部感兴趣目标图像进行采样。首先,进行能量聚集运算,通过式(3)定量描述全局上下文特征 F_{global} 中各空间元素的重要程度:



注：\$F_{global}\$和\$F_{local}\$分别代表全局和局部上下文特征；\$f_{global}\$和\$f_{local}\$分别代表全局和局部上下文特征向量；\$f_{fusion}\$代表融合后的特征向量；\$[x_1, y_1, x_2, y_2]\$代表定位的边界框左上及右下点坐标。
 Note: \$F_{global}\$ and \$F_{local}\$ represent the global and local context features, respectively; \$f_{global}\$ and \$f_{local}\$ denote the global and local context feature vectors, respectively; \$f_{fusion}\$ represents the fused feature vector; \$[x_1, y_1, x_2, y_2]\$ denotes the coordinates of the top-left and bottom-right points of the localized bounding box.

图2 基于尺度交互蒸馏网络的薄壳山核桃品种识别方法框架图

Fig.2 Framework of pecan nut recognition method based on scale interactive distillation network

$$M_e = \sum_i^c F_{global}(i, H, W), \quad (3)$$

式中\$M_e \in R^{H \times W}\$表示能量图（\$H, W\$分别表示特征的高和宽），\$i\$表示通道，\$c\$为\$F_{global}\$的通道数。为了消除负元素的干扰，实现更精确的定位，将\$M_e\$的元素按最小-最大比例缩放到\$[0, 1]\$范围内，并进行上采样，从而得到缩放后的能量图\$\hat{M}_e\$。

其次，为了优化操作，对能量图进行结构化操作，将二维\$\hat{M}_e\$沿空间高度和宽度聚合为一维结构化能量向量：

$$\begin{cases} V_b = \sum_{j=0}^W \hat{M}_e(j, H) \\ V_w = \sum_{k=0}^H \hat{M}_e(W, k) \end{cases}, \quad (4)$$

式中\$V_b\$和\$V_w\$分别表示沿空间高度和宽度方向的一维结构化能量向量。\$j\$和\$k\$分别表示缩放后的能量图\$\hat{M}_e\$的空间宽度和高度。

第三步是进行贪婪边界搜索，从而确定局部目标的边界框。用\$E_{[0:W]}\$表示\$V_w\$的能量和，\$E_{[x_1:x_2]}\$表示从\$x_1\$到\$x_2\$沿空间宽度的能量，它们可以通过下述公式计算：

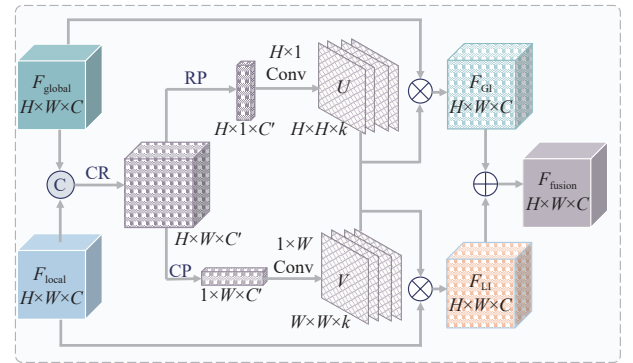
$$\begin{cases} E_{[0:W]} = \sum_{s=0}^W V_w(s) \\ E_{[x_1:x_2]} = \sum_{s=x_1}^{x_2} V_w(s) \end{cases}, \quad (5)$$

式中\$s\$表示\$V_w\$向量的空间位置索引，宽度边界\$x_1\$和\$x_2\$可以通过在约束条件\$E_{[x_1:x_2]}/E_{[0:W]} > \xi\$下确定最小的\$[x_1 : x_2]\$面积来求解，其中\$\xi\$表示预定义的能量阈值。高度边界\$y_1\$和\$y_2\$可以类似地进行求解。一旦得到整个边界\$[x_1, x_2, y_1, y_2]\$，即可从输入的全局图像\$I_g\$中采样局部感兴趣目标图像\$I_l\$。

1.4 多尺度上下文注意力融合

为了促进不同尺度和不同空间位置之间的信息交互，

提出一种多尺度上下文注意融合模块，用于全局和局部特征之间的上下文信息协同聚合。工作流程如图3所示。



注：\$U\$和\$V\$分别代表行和列变换矩阵；\$F_{GI}\$和\$F_{LI}\$分别代表全局和局部感知交互特征；\$F_{fusion}\$代表融合后的特征；\$C'\$代表列池化特征通道数。©拼接、⊗矩阵相乘、⊕逐元素相加、CR：通道缩减、RP：行池化、LP：列池化。
 Note: \$U\$ and \$V\$ denote row and column transformation matrices, respectively; \$F_{GI}\$ and \$F_{LI}\$ are global and local perception interaction features, respectively; \$F_{fusion}\$ represents the fused feature; \$C'\$ represents the number of column-pooled feature channels. © Concatenation, ⊗ matrix multiplication, ⊕ element by element addition, CR: channel reduction, RP: row pooling, LP: column pooling.

图3 多尺度上下文注意力融合流程图

Fig.3 The flow chart of multi-scale contextual attention fusion

首先，将全局上下文特征\$F_{global}\$和局部上下文特征\$F_{local}\$通过通道缩减进行拼接和压缩：

$$F^{cm} = h([F_{global}, F_{local}]), \quad (6)$$

式中\$F^{cm} \in R^{H \times W \times C}\$代表压缩得到的特征，\$h(\cdot)\$表示包括批归一化和 ReLU 激活函数的\$1 \times 1\$卷积层用于通道缩减，\$[:, :]\$表示两个特征的拼接操作。

其次, 分别沿着 F^{cm} 的空间宽度和高度进行行池化和列池化, 得到行池化特征 $F^{rp} \in R^{H \times 1 \times C'}$ 和列池化特征 $F^{cp} \in R^{1 \times W \times C'}$ (F^{cm} 中上标 cm 代表压缩, F^{cm} 表示通过公式(6)压缩得到的特征矩阵, F_m^{rp} 和 F_n^{cp} 中的 m 和 n 分别代表对这个矩阵的行索引和列索引)。该操作通过下式实现:

$$F_m^{rp} = \max \{F_{m,n}^{cm} | 1 \leq n \leq W\} \quad (7)$$

$$F_n^{cp} = \max \{F_{m,n}^{cm} | 1 \leq m \leq H\} \quad (8)$$

第三步是基于池化特征 F^{rp} 和 F^{cp} , 对变换矩阵 $U \in R^{H \times H \times C'}$ 和 $V \in R^{W \times W \times C'}$ 进行估计:

$$U = \text{rearrange}(F^{rp} * W^{rp}), \quad (9)$$

$$V = \text{rearrange}(F^{cp} * W^{cp}), \quad (10)$$

式中 $*$ 表示卷积运算, W^{rp} 表示大小为 $H \times 1$ 的卷积核, W^{cp} 表示大小为 $1 \times W$ 的卷积核, rearrange 代表调整矩阵形状的重塑形操作。

第四步采用双线性注意变换分别从全局上下文特征 F_{global} 和局部上下文特征 F_{local} 中获取全局感知交互特征 $F_{GI} \in R^{H \times W \times C}$ 和局部感知交互特征 $F_{LI} \in R^{H \times W \times C}$:

$$F_{GI} \in U \cdot F_{global} \cdot V \quad (11)$$

$$F_{LI} \in U \cdot F_{local} \cdot V \quad (12)$$

式中 $\cdot$ 为矩阵乘积运算。为方便起见, 将式(11)和式(12)中的矩阵转置省略。

最后, 将全局感知交互特征 F_{GI} 和局部感知交互特征 F_{LI} 相结合, 学习融合特征 F_{fuse}

$$F_{fuse} = F_{GI} \oplus F_{LI} \quad (13)$$

因此, 通过多尺度上下文注意融合模块可以有效地增强全局上下文特征和局部上下文特征之间的交互。此外, 该模块是即插即用的, 可扩展到各种多尺度特征的融合任务中。

1.5 尺度知识蒸馏

为了解决全局特征和局部特征之间预测不一致的问题, 提出一种尺度间知识蒸馏方案以产生更一致的预测。具体来说, 令 I 和 y 分别表示输入图像和相应的类标签。预测分布 $P(y|I; Z, T)$ 可表示为

$$P(y|I; Z, T) = \frac{\exp(f_y(I; Z)/T)}{\sum_{k=1}^K \exp(f_k(I; Z)/T)}, \quad (14)$$

式中 f_k 表示网络对第 k 类的预测分数, Z 代表待学习的参数, $T > 0$ 表示蒸馏温度参数。为了匹配同一类的局部图像和全局图像之间的预测分布, 引入了一种尺度正则化损失, 可表示为 L_{kd} :

$$L_{kd}(I_g, I_l; Z, T) = KL(P(y|I_l; Z_{lb}, T) \| P(y|I_g; Z_{gb}, T)), \quad (15)$$

式中 $KL(\cdot \| \cdot)$ 表示 Kullback-Leibler (KL) 散度。 Z_{lb} 和 Z_{gb} 分别为局部分支和全局分支中的参数。在上述损失的约束下, 可以有效缓解不同尺度之间预测的不一致性问题。

1.6 试验环境及模型评价指标

1.6.1 试验环境及细节

试验开发环境基于 Pytorch 深度学习框架进行, 使用 Adam 优化器对所提出的尺度交互蒸馏网络进行优化, 将学习率初始化为 $1e-4$, 并采用余弦退火学习率衰减法设置学习率。试验批量大小设置为 64, 训练周期设置为 100。结构化关键区域定位模块中预定义的能量阈值 ξ 设为 0.8。采用 31G 内存的 Windows 10 工作站, 配置 NVIDIA GeForce RTX 3090 24 GB 显卡进行训练和结果验证。试验通过 Python 语言和 Pytorch 深度学习工具实现。为了验证所提方法在实际应用场景下的效果, 本研究对薄壳山核桃额外拍摄了一些实际场景图片用于模型泛化性能的测试。部分数据如图 4 所示 (共 8 个数据样本)。



注: 随机展示的样本实际应用场景图。

Note: Sample actual application scenario diagram displayed at random.

图 4 部分实际应用场景数据示例

Fig.4 Some practical application scenario data examples

1.6.2 评价指标

试验使用 2 个典型的评价指标用于定量评价试验结果。1) 总体准确率 (overall accuracy, OA): OA 通过正确分类的样本数除以测试样本总数来计算。OA 值反映了分类模型的整体性能。2) 混淆矩阵 (confusion matrix, CM): CM 是一个二维信息表, 每一行表示预测类, 每一列表示真实类。基于 CM 评价指标可以分析类间的分类误差。

2 结果与分析

2.1 消融试验

本研究中, 除提出的尺度交互蒸馏网络之外, 还设计了不同的变体以验证: 1) 多尺度上下文注意力融合模块的有效性; 2) 尺度间知识蒸馏方案的重要性。消融研究的详细过程如下。

首先, 实现了没有尺度间知识蒸馏方案的单分支基线 (变体 A)。其次, 将全局和局部上下文特征与所提出的多尺度上下文注意力融合模块相结合, 而不使用尺度间知识蒸馏方案对预测分布 (变体 B)。第三, 在预测分布之间使用尺度间知识蒸馏方案, 而全局和局部上下文特征图以串联方式进行聚合而非所提出的多尺度上下文注意力融合模块 (变体 C)。最后, 实现了所提出的尺度交互蒸馏网络的完整版本。各种变体方法的对比结果详见表 1, 通过观察可得出如下结论。

表 1 消融试验结果对比
Table 1 Comparison results for ablation experiments

对比方法 Comparative methods	多尺度上下文 注意力融合模块 Multi-scale context attention fusion module	尺度间知识 蒸馏方案 Interscale knowledge distillation scheme	总体准确率 Overall accuracy OA%
变体 A Version A			89.74
变体 B Version B	√		94.19
变体 C Version C		√	93.82
所提方法 The method of this study	√		96.98

1) 多尺度上下文注意力融合模块对提高薄壳山核桃的特征表示能力有显著作用。通过表 1 中的对比结果可得，在去掉多尺度上下文注意力融合模块后，识别性能急剧下降，OA 从 96.98% 下降到 93.82%。这是因为多尺度上下文注意力融合模块可以通过增强不同尺度和不同空间位置之间的相互作用，并学习具有尺度不变性和背景无关信息的判别性表征。此外，单分支基线（变体 A）与所提出的尺度交互蒸馏网络之间的比较结果进一步证明了所述尺度交互蒸馏网络在聚合薄壳山核桃的全局和局部信息方面的有效性。

2) 尺度间知识蒸馏方案可以有效地整合全局和局部上下文特征之间的预测分布。如表 1 所示，变体 B 的结果和所提出的尺度交互蒸馏网络相差较大。当采用尺度间知识蒸馏方案时，OA 从 94.19% 提高到 96.98%。此结果是因为当去掉尺度间知识蒸馏时，会导致薄壳山核桃品种识别模型的泛化能力有限，使得全局和局部上下文特征之间的预测分布产生不一致现象。相比之下，本文所述方法可以有效地实现这一点，这表明前述提出的尺度间知识蒸馏方案能够产生更一致的预测。

2.2 识别方法对比

为了证明所提方法的优越性，进一步和一些典型的图像分类模型进行比较，这些对比模型包括 AlexNet^[31]、VGG16^[32]、InceptionV3^[33]、ResNet101^[34]、MobileNetV2^[35]、CrossViT^[36]、TL-MobileNetV2^[37]、ViTAEv2^[38]。

表 2 和表 3 分别展示了不同方法在测试数据集和实际应用场景数据上的总体准确率结果对比。所述方法在 2 种类型数据上均取得了最优的识别准确率，这说明所述方法的性能优于其他对比方法，且有较好的实用性。该方法能够有效地提取具有尺度不变性和与背景无关信息的判别性特征，从而提高薄壳山核桃品种识别的准确性和鲁棒性。相比之下，其他对比方法的性能都有一定的差距，主要原因如下：1) AlexNet、VGG16、InceptionV3 和 ResNet 等传统的卷积神经网络在处理尺度多样性和背景干扰问题上存在一定的局限性，无法充分利用全局和局部特征之间的上下文信息。2) MobileNetV2、CrossViT 和 TL-MobileNetV2 等轻量化的网络虽然具有较低的计算复杂度和参数量，但是牺牲了一定的特征表达能力，导

致在尺度差异较大和背景干扰复杂的情况下无法达到理想的效果。3) ViTAEv2 等基于自注意力机制的网络虽然能够捕捉全局特征之间的长距离依赖关系，但是忽略了局部特征之间的细节信息，且没有考虑不同尺度之间的特征交互，导致对薄壳山核桃品种识别的精度有限，仅取得 90.11 % 的识别准确率。综上所述，所述方法通过引入多尺度上下文注意力融合模块和尺度知识蒸馏方案，有效地解决了薄壳山核桃品种识别中存在的尺度多样性和背景干扰问题，实现了对薄壳山核桃果形的精准识别。

表 2 不同方法在测试集上的结果比较
Table 2 Comparison for different methods on the testing set

对比方法 Comparative methods	年份 Year	总体准确率 OA%
AlexNet	2012	87.24
VGG16	2014	89.41
InceptionV3	2016	91.14
ResNet101	2016	92.08
MobileNetV2	2018	90.75
CrossViT	2021	93.17
TL-MobileNetV2	2023	91.57
ViTAEv2	2023	94.62
所提方法 The method of this study	2024	96.98

表 3 不同方法在实际应用场景的结果比较
Table 3 Comparison for different methods on the Actual application scenario data

对比方法 Comparative methods	年份 Year	总体准确率 OA%
AlexNet	2012	81.09
VGG16	2014	82.92
InceptionV3	2016	84.64
ResNet101	2016	87.41
MobileNetV2	2018	85.37
CrossViT	2021	88.64
TL-MobileNetV2	2023	86.52
ViTAEv2	2023	90.11
所提方法 The method of this study	2024	94.26

2.3 混淆矩阵分析和可视化试验

为了进一步直观地评估每一类的识别结果，本文所提出在尺度交互蒸馏网络基础上构建的薄壳山核桃品种识别数据集上的混淆矩阵，如图 5 所示。该模型对大多数品种的识别准确率较高。其中，除“威奇塔”外，其他品种的准确率都达到了 94% 以上，“坎迪”“德西拉布”和“金华”等品种的准确率甚至达到了 100%。“威奇塔”识别准确率较低的原因可能是由于该品种和其他品种区别性不强，主要表现为“威奇塔”果形和纹理与其他类别存在相似度较高，导致模型难以区分。总体而言，本文所提出的尺度交互蒸馏网络在薄壳山核桃品种识别方面具有良好的性能。该模型能够有效地提取薄壳山核桃果形的尺度不变特征和与背景无关的信息，从而提高识别准确率。

为了更好地解释所提出的尺度交互蒸馏网络对薄壳山核桃品种识别的分类机制，使用 Grad-CAM++可视化了薄壳山核桃图像中感兴趣目标区域的特定品种位置。通过 10 组样本的可视化结果（图 6，共 12 组样本数据），针对图片中背景干扰和不同尺度差异的感兴趣目标，所述方法均可准确定位其中的关键区域。这表明所

述方法可以有效抑制背景信息和感兴趣目标的尺度变化带来的干扰, 显示了所述方法的鲁棒性。

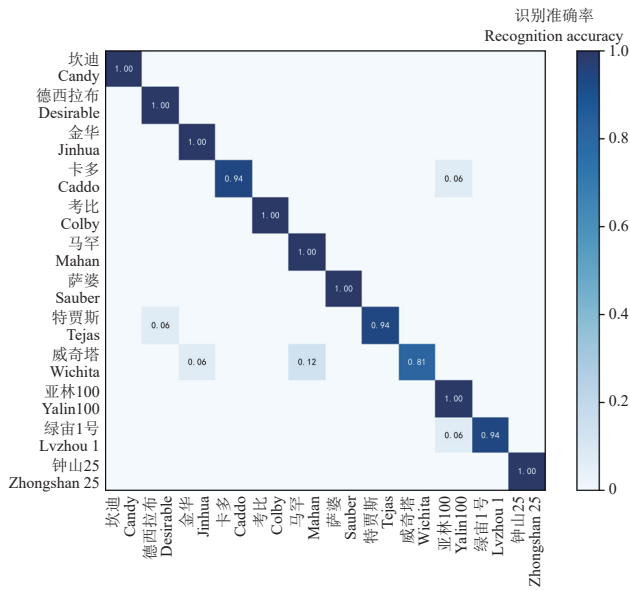
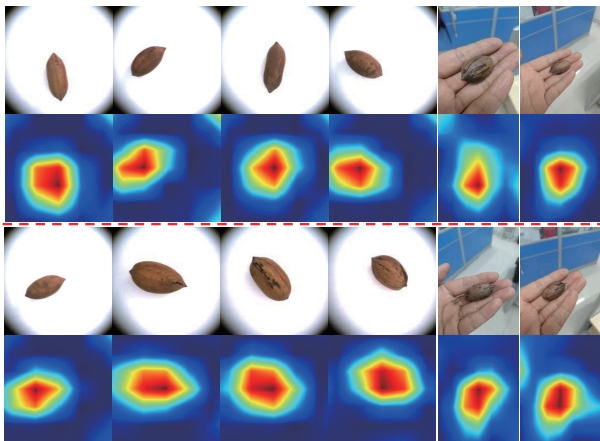


图5 混淆矩阵结果

Fig.5 Results for confusion matrix



注: 第1行和第3行为随机挑选的样本图片, 第2行和第4行为对应的提取特征图像。

Note: The first and third rows are randomly selected sample images, while the second and fourth rows are the corresponding feature-extracted images.

图6 感兴趣目标区域定位结果可视化

Fig.6 Visualization of localization results for region of interest

3 结论

本研究针对薄壳山核桃品种识别的困难, 以12个品种薄壳山核桃坚果为研究对象, 采集并标注了包含9 048张图像的薄壳山核桃的品种识别数据集。提出了一种基于尺度交互蒸馏网络的薄壳山核桃品种识别方法。为了协同融合全局和局部特征之间的上下文信息, 提出了即插即用的多尺度上下文注意力融合模块以促进不同尺度和不同空间位置之间的交互。同时, 设计了一种尺度间知识蒸馏方案来提高薄壳山核桃品种识别模型的泛化能力, 通过提取全局和局部上下文特征之间的预测分布, 从而产生更一致的识别结果。本文提出的尺度交互蒸馏网络对薄壳山核桃的品种识别总体准确率高达96.98%,

可有效解决实际生产中品种识别困难的问题, 可以帮助科研人员、种植户及加工企业快速、准确地识别和分类薄壳山核桃果实品种, 从而提高品种选育的效率和准确性, 为优化测试指南提供科学依据。另外, 也可应用于市场监管和商业交易中, 帮助消费者快速鉴别薄壳山核桃的品种, 保障消费者的权益。本研究为进一步研发高效、经济、无损的品种鉴定技术提供理论依据。

[参考文献]

- [1] CHEN, B., SHI, B., GONG, J., et al. Quality detection and variety classification of pecan seeds using hyperspectral imaging technology combined with machine learning[J]. Journal of Food Composition and Analysis, 2024, 131: 106248.
- [2] 姜冰燕, 蒋昕晨, 张涛, 等. 基于文献计量分析的我国薄壳山核桃研究现状[J]. 果树资源学报, 2024, 5(2): 60-63. JIANG Bingyan, JIANG Xinchun, ZHANG Tao, et al. Research status of pecan based on bibliometrics[J]. Journal of Fruit Tree Resources, 2024, 5(2): 60-63. (in Chinese with English abstract)
- [3] 张计育, 李永荣, 陈智坤, 等. 薄壳山核桃产业发展中存在的问题及其对策[J]. 中国果业信息, 2021, 38(11): 21-23. ZHANG Jiyu, LI Yongrong, CHEN Zhikun, et al. Problems and countermeasures in the development of *Carya illinoensis* industry[J]. China Fruit News, 2021, 38(11): 21-23. (in Chinese with English abstract)
- [4] ZHU H J. Challenges for the expanding pecan industry in China[J]. Pecan South, 2018, 51(9): 16-23.
- [5] 姚小华. 中国薄壳山核桃[M]. 北京: 科学出版社, 2014.
- [6] International Union For The Protection of New Varieties of Plants. Pecan nutguidelines for the conduct of tests for distinctness, uniformity and stability: TG/308/1-2015 [S]. Geneva: UPOV, 2015: 5-13.
- [7] BENTLEY N, GRAUKE L J, KLEIN P. Genotyping by sequencing (GBS) and SNP marker analysis of diverse accessions of pecan (*Carya illinoensis*)[J]. Tree Genetics & Genomes, 2019(15): 1-17.
- [8] 张三林, 张立萍, 郑威强, 等. 基于YOLOv5的核桃品种识别与定位[J]. 中国农机化学报, 2022, 43(7): 167-172. ZHANG Sanlin, ZHANG Liping, ZHENG Weiqiang, et al. Identification and localization of walnut varieties based on YOLOv5[J]. Journal of Chinese Agricultural Mechanization, 2022, 43(7): 167-172. (in Chinese with English abstract)
- [9] 冯妍舟, 刘建霞, 王海翼, 等. 基于多级残差信息蒸馏的真实图像去噪方法[J]. 计算机工程, 2024, 50(3): 216-223. FENG Yanzhou, LIU Jianxia, WANG Haiyi, et al. Real image denoising method based on multi-level residual information distillation[J]. Computer Engineering, 2024, 50(3): 216-223. (in Chinese with English abstract)
- [10] 李茂, 肖洋铁, 宗望远, 等. 基于改进YOLOv8模型的轻量化板栗果实识别方法[J]. 农业工程学报, 2024, 40(1): 201-209. LI Mao, XIAO Yangyi, ZONG Wangyuan, et al. Detecting chestnuts using improved lightweight YOLOv8[J]. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 2024, 40(1): 201-209. (in Chinese with English abstract)
- [11] ANUJA B, ATUL B. Automatic detection and grading of multiple fruits by machine learning[J]. Food Analytical Methods, 2020, 13(3): 751-761.
- [12] KOIRALA A, WALSH K B, WANG Z, et al. Deep learning

- for mango (*Mangifera indica*) panicle stage classification[J]. *Agronomy*, 2020, 10(1): 143.
- [13] LU S, LU Z, AOK S, et al. Fruit classification based on six layer convolutional neural network[C]//2018 IEEE 23rd International Conference on Digital Signal Processing (DSP). Shanghai China:IEEE, 2018: 1-5.
- [14] GHAZI M M, YANIKOGLU B, APTOULA E. Plant identification using deep neural networks via optimization of transfer learning parameters[J]. *Neurocomputing*, 2017, 235: 228-235.
- [15] 朱轶萍, 吴华瑞, 郭旺, 等. 基于改进 UperNet 的结球甘蓝叶球识别方法[J]. 智慧农业, 2024, 6(3): 128-137.
- ZHU Yiping, WU Huarui, GUO Wang, et al. Identification method of cabbage leaf ball based on improved UperNet[J]. *Intelligent Agriculture*, 2024, 6(3): 128-137. (in Chinese with English abstract)
- [16] POURDARBANI R, SABZI S, GARCÍA-AMICIS V M, et al. Automatic classification of chickpea varieties using computer vision techniques[J]. *Agronomy*, 2019, 9(11): 672.
- [17] 张瑞青, 李张威, 郝建军, 等. 基于迁移学习的卷积神经网络花生荚果等级图像识别[J]. 农业工程学报, 2020, 36(23): 171-180.
- ZHANG Ruiqing, LI Zhangwei, HAO Jianjun, et al. Image recognition of peanut pod grades based on transfer learning with convolutional neural network[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2020, 36(23): 171-180. (in Chinese with English abstract)
- [18] KNOLL F J, CZYMMEK V, HARDERS L O, et al. Real-time classification of weeds in organic carrot production using deep learning algorithms[J]. *Computers and Electronics in Agriculture*, 2019, 167: 105097.
- [19] HU W, WANE S O, ZHU J, et al. Review of deep learning-based weed identification in crop fields[J]. *International Journal of Agricultural and Biological Engineering*, 2023, 16(4): 1-10.
- [20] 周罕觅, 陈佳庚, 代智光, 等. 基于知识蒸馏和轻量级卷积神经网络的植物病害识别方法[J]. 南京农业大学学报, 2024, 47(6): 1189-1201.
- ZHOU Hanmi, CHEN Jiageng, DAI Zhiguang, et al. Plant disease recognition method based on knowledge distillation and lightweight convolutional neural network[J]. *Journal of Nanjing Agricultural University*, 2024, 47(6): 1189-1201. (in Chinese with English abstract)
- [21] 兰玉彬, 孙斌书, 张乐春, 等. 基于改进 YOLOv5s 的自然场景下生姜叶片病虫害识别[J]. 农业工程学报, 2024, 40(1): 210-216.
- LAN Yubin, SUN Binshu, ZHANG Lechun, et al. Identifying diseases and pests in ginger leaf under natural scenes using improved YOLOv5s[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2024, 40(1): 210-216. (in Chinese with English abstract)
- [22] PRZYBYLAK A, KOZIOŃSKI R, OSUCH E, et al. Quality evaluation of potato tubers using neural image analysis method[J]. *Agriculture*, 2020, 10(4): 112.
- [23] 黄海松, 陈星燃, 韩正功, 等. 基于多尺度注意力机制和知识蒸馏的茶叶嫩芽分级方法[J]. 农业机械学报, 2022, 53(9): 399-407,458.
- HUANG Haisong, CHEN Xingran, HAN Zhenggong, et al. Tea buds grading method based on multiscale attention mechanism and knowledge distillation[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 53(9): 399-407,458. (in Chinese with English abstract)
- [24] 陈锋军, 陈闯, 朱学岩, 等. 基于改进 YOLOv7 的油茶果实成熟度检测[J]. 农业工程学报, 2024, 40(5): 177-186.
- CHEN Fengjun, CHEN Chuang, ZHU Xueyan, et al. Detection of *Camellia oleifera* fruit maturity based on improved YOLOv7[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2024, 40(5): 177-186. (in Chinese with English abstract)
- [25] KATARZYNA R, PAWEŁ M. A vision-based method utilizing deep convolutional neural networks for fruit variety classification in uncertainty conditions of retail sales[J]. *Applied Sciences*, 2019, 9(19): 3971.
- [26] MURESAN H, OLTEAN M. Fruit recognition from images using deep learning[J]. *Acta Universitatis Sapientiae, Informatica*, 2018, 10(1): 26-42.
- [27] ZHU L, LI Z, LI C, et al. High performance vegetable classification from images based on alexnet deep learning model[J]. *International Journal of Agricultural and Biological Engineering*, 2018, 11(4): 217-223.
- [28] QIAN Y, XU Q, YANG Y, et al. Classification of rice seed variety using point cloud data combined with deep learning[J]. *International Journal of Agricultural and Biological Engineering*, 2021, 14(5): 206-212.
- [29] 毛永文, 韩俊英, 刘成忠. 基于机器视觉的胡麻种子自动化考种方法[J]. 智慧农业, 2024, 6(1): 135.
- MAO Yongwen, HAN Junying, LIU Chengzhong. Automatic testing method of flax seed based on machine vision[J]. *Intelligent Agriculture*, 2024, 6(1): 135. (in Chinese with English abstract)
- [30] WANG Q, HUANG W, XIONG Z, et al. Looking closer at the scene: Multiscale representation learning for remote sensing image scene classification[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 33(4): 1414-1428.
- [31] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. *Advances in Neural Information Processing Systems*, 2012, 25(12): 1097-1105.
- [32] E. MAGGIORI, Y TARABALKA, G. CHARPIAT, et al. Convolutional neural networks for large-scale remote-sensing image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(2): 645-657.
- [33] SZEGEDY C, VANHOUCHE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition, 2016: 2818-2826.
- [34] SHAFIQ M, GU Z. Deep residual learning for image recognition: A survey[J]. *Applied Sciences*, 2022, 12(18):8972.
- [35] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [36] CHEN C F R, FAN Q, PANDA R. Crossvit: Cross-attention multi-scale vision transformer for image classification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE, 2021: 357-366.
- [37] GULZAR Y. Fruit image classification model based on MobileNetV2 with deep transfer learning technique[J]. *Sustainability*, 2023, 15(3): 1906.
- [38] ZHANG Q, XU Y, ZHANG J, et al. Vitaev2: Vision transformer advanced by exploring inductive bias for image recognition and beyond[J]. *International Journal of Computer Vision*, 2023, 131(5): 1141-1162.

Identification method for *Carya illinoensis* varieties based on scale interactive distillation network

ZHAO Ning¹, CHEN Zhikun^{1*}, YANG Pengfei¹, WANG Ruiduo², ZHANG Jiyu³, LI Yongrong⁴

(1. Xi'an Botanical Garden of Shaanxi Province, Institute of Botany of Shaanxi Province, Xi'an 710061, China; 2. Research Laboratory of Photonic Functional Materials and Devices, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China; 3. Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing 210014, China; 4. Nanjing Green Universe Pecan Science & Technology Limited Company in Jiangsu Province, Nanjing 210007, China)

Abstract: Pecan (*Carya illinoensis*) is one of the highly valuable nut crops in the global agricultural economy, due to its nutritional and economic advantages. There has ever a rising demand for high-quality cultivars of pecan in recent years. Precise and efficient identification is then required for germplasm resource conservation, breeding programs, and commercial production. However, some challenges still remain in identifying the pecan cultivars, owing to their morphological similarities, and environmental and image acquisition, such as lighting, angle, and distance. Traditional cultivar identification can also rely mainly on manual observation and expert knowledge; These manual approaches cannot fully meet large-scale production, due to the time-consuming susceptibility to subjective errors. It is very necessary to develop automatic, scalable, and accurate recognition of the pecan cultivars. In this study, a scale-interactive knowledge distillation network (siKD-Net) was introduced to enhance the accuracy of classification using deep learning. The discriminative features were then extracted from the pecan images, in order to facilitate the highly precise differentiation of cultivar. 9 048 images were captured from the twelve cultivars of widely cultivated pecan, such as Pawnee. An image dataset was established to support the training and evaluation of the improved model. Among them, the different imaging distances also caused variations in the object scales during data acquisition. The inconsistent representation of features seriously deteriorated the classification of the diverse image samples. Fortunately, the SIKD-Net framework performed better to solve the scale variation and feature inconsistency during image classification. The global-local feature was also incorporated with the scale-aware knowledge distillation. A series of experiments were then conducted to evaluate the performance of the SIKD-Net model. The results indicate that the classification accuracy of 96.98% was achieved in the twelve pecan cultivars, significantly surpassing conventional machine learning and deep learning. A comparison was also made on the baseline models, including the standard convolutional neural networks (CNNs) and transformer-based architectures. Moreover, the scale-interactive knowledge distillation substantially enhanced both the robustness and accuracy of the cultivar recognition. The high accuracy of classification was achieved in the pecan cultivar for agricultural and industrial applications. Automatic recognition can be extended to the rest of tree nuts and fruit crops for precise classification. Furthermore, the improved model can be deployed within smart systems for seedling selection, orchard management, and post-harvesting. Therefore, intelligent sorting can be integrated to enhance efficiency, in order to reduce the reliance on manual inspection in commercial settings. In conclusion, deep-learning identification can be expected to significantly enhance the efficiency and accuracy of pecan cultivar recognition. The hyperspectral imaging and multimodal feature fusion can also be integrated to further enhance the performance and the deployment of the SIKD-Net model for the large-scale classification in real-world orchard environment. The findings can also provide a scientifically robust and practically viable solution to pecan germplasm identification in precision agriculture, intelligent crop classification, and food processing.

Keywords: *Carya illinoensis*; variety identification; scale interactive distillation; multi scale context attention fusion; knowledge distillation