doi:10.6041/j.issn.1000-1298.2026.17.027

基于改进 Vision Transformer 与环境权重融合的金耳生长阶段识别方法

孙先鹏[1,2]　寇龙龙[1]　关博文[3]　王新轲[3]　刘　娜[4]　宁明阳[1]

(1. 西北农林科技大学园艺学院, 陕西杨凌 712100; 2. 西北农林科技大学智能装备研究院, 陕西杨凌 712100;
3. 西安交通大学人居环境与建筑工程学院, 西安 710054; 4. 新疆维吾尔自治区农业科学院, 乌鲁木齐 830091)

摘要: 针对金耳工厂化生产中人工判断生长阶段存在的主观性强、效率低、标准难统一等问题,本文提出一种融合视觉与环境信息的多模态金耳生长阶段识别方法。首先构建包含图像与温度、相对湿度、光照强度、CO_2 浓度同步采集的金耳多模态数据集,涵盖幼耳期、转色期、成熟期 3 个阶段共 50 420 份样本;其次设计改进 Vision Transformer 图像特征提取网络,引入多尺度特征嵌入、位置增强编码模块与 ResNet 局部感知增强模块,强化金耳菌包纹理细节和空间关联的建模能力;同时提出环境因子自适应权重融合策略,通过可学习权重对环境参数差异化加权,实现视觉与环境多模态特征的联合优化。实验结果表明,该方法测试集总体识别准确率达 93.2%;宏观平均 F1 值 93.3%,较原始 ViT 和 ResNet50 基线模型的 F1 值分别提升 7.2、8.7 个百分点,其中幼耳期 F1 值高达 98.8%。消融实验验证了多尺度特征嵌入、局部感知增强模块及环境融合策略对模型性能的显著提升作用,环境因子重要性分析量化了金耳不同生长阶段的环境敏感性差异,转色期对相对湿度(权重 0.40)和光照强度(0.25)的敏感性最高。该研究实现了金耳生长阶段的高精度、可解释识别,模型参数量仅为原始 ViT 的 52.1%,推理速度达 22.2 f/s,满足工厂化生产实时监测需求,为食用菌工厂化生产的智能化、精准化管理提供了高效的技术解决方案。

关键词: 金耳; 生长阶段识别; 改进 Vision Transformer; 多模态融合

中图分类号: S646; TP391.4　　**文献标识码**: A　　**文章编号**: 1000-1298(2026)17-0288-11　　**OSID**:

Method for Identifying *Tremella aurantialba* Growth Stages Based on Improved Vision Transformer and Environmental Weight Fusion

Sun Xianpeng[1,2]　Kou Longlong[1]　Guan Bowen[3]　Wang Xinke[3]　Liu Na[4]　Ning Mingyang[1]

(1. *College of Horticulture, Northwest A&F University, Yangling, Shaanxi* 712100, *China*
2. *Institute of Intelligent Agricultural Equipment, Northwest A&F University, Yangling, Shaanxi* 712100, *China*
3. *School of Human Settlements and Civil Engineering, Xi'an Jiaotong University, Xi'an* 710054, *China*
4. *Xinjiang Academy of Agricultural Sciences, Urumqi* 830091, *China*)

Abstract: Aiming at the inherent limitations of manual growth stage judgment in industrial *Tremella aurantialba* (golden ear) cultivation, including strong subjectivity, low efficiency, and difficulty in standardizing evaluation criteria, a multimodal recognition method for golden ear growth stages was proposed by integrating visual and environmental information. Firstly, a multimodal dataset of golden ear was constructed, which incorporated synchronously collected images and key environmental parameters (temperature, humidity, light intensity, and CO_2 concentration). This dataset covered 50 420 samples across three distinct growth stages, namely the young ear stage, color conversion stage, and mature stage. Secondly, an improved Vision Transformer (ViT) network was designed for image feature extraction, where multi-scale feature embedding, dynamic positional encoding, and a ResNet-based local perception enhancement module were introduced to enhance the modeling capability of texture details and spatial correlations of golden ear cultivation bags. Meanwhile, an adaptive weight fusion strategy for environmental factors was developed, which employed learnable weights to perform differential weighting on environmental parameters, thereby achieving joint optimization of visual and environmental multimodal features. Experimental results show that the overall recognition accuracy of the proposed method on the

收稿日期: 2026-03-10　修回日期: 2026-04-15

基金项目: 2025 陕西省农业财政专项和新疆维吾尔自治区重点研发项目(2024B04028)

作者简介: 孙先鹏(1979—),男,副教授,博士,主要从事设施农业智慧环境调控研究,E-mail: sunxianpeng@nwsuaf.edu.cn

test set reaches 93.2%, with a macro-average F1-score of 93.3%. It outperforms the original ViT and ResNet50 baseline models by 7.2 and 8.7 percentage points in terms of F1-score, respectively. In particular, the F1-score for the young ear stage is as high as 98.8%. Ablation experiments further verified that the multi-scale feature embedding, local perception enhancement module, and environmental fusion strategy all contributed significantly to the improvement of model performance. Moreover, the importance analysis of environmental factors quantified the differences in environmental sensitivity among different growth stages of golden ear; notably, the color conversion stage exhibited the highest sensitivity to humidity (weight 0.40) and light intensity (weight 0.25). The research realized high-precision and interpretable recognition of golden ear growth stages. With only 52.1% of the parameters of the original ViT and an inference speed of 22.2 f/s, the proposed model met the real-time monitoring requirements of industrial production, providing an efficient technical solution for the intelligent and precise management of industrial edible fungi cultivation.

Key words: *Tremella aurantialba*; growth stage recognition; improved Vision Transformer; multimodal fusion

0　引言

金耳(*Tremella aurantialba*)是一种具有重要食用与药用价值的珍稀菌类作物,富含多糖、蛋白质、氨基酸等活性成分,具有抗氧化、免疫调节、降血糖等多种生理功能[1-3]。近年来,随着工厂化栽培技术的突破,金耳在食用菌生产中实现规模化推广,成为高附加值特色农产品[4-5]。金耳的产量和品质与其生长阶段密切相关,不同阶段在温度、湿度、光照强度和 CO_2 浓度等环境参数上存在差异化需求[6]。目前,生产实践中生长阶段的划分主要依赖经验丰富的技术人员通过目测菌包形态和生长状态来判断,存在劳动强度高、主观判断差异大、标准难以统一等问题,难以满足规模化、标准化生产的实际需求[7]。因此,开展基于视觉与环境信息的金耳生长阶段自动化识别研究,对于实现精准管理和智能调控具有重要意义。

随着计算机视觉和深度学习技术的快速发展,基于图像的作物生长监测取得了显著进展,相关研究涵盖了果蔬目标检测、病害识别、作物分割及产量预测等多个方向[8-12]。卷积神经网络(CNN)在 ImageNet 等大规模数据集上取得了突破性进展,在提取局部纹理特征方面表现优异,但其感受野受网络深度和卷积核大小限制,对远距离依赖关系的建模能力有限[13]。近年来提出的 Vision Transformer(ViT)基于自注意力机制,在图像分类和目标检测等任务中表现出优异性能[14-16]。然而,标准 ViT 多针对自然图像设计,直接应用于金耳菌包图像时,容易受到复杂背景干扰、光照条件变化、菌丝纹理细碎且分布不均等因素影响,导致特征提取不充分,识别精度难以满足实际应用需求[17-18]。

作物生长过程受环境因子阶段性、累积性和耦合作用的长期影响,仅依赖图像信息难以全面反映其生理阶段特征[19-20]。部分学者已开始尝试将温湿度、光照强度等数据与图像特征进行多模态融合,用于设施环境评估和产量预测等任务,但针对食用菌特别是金耳的相关研究仍较少,且对环境因子重要性差异的建模不够充分,缺乏对因子间交互作用的有效表征[21-23]。因此,亟需开展金耳生长阶段智能识别方法研究,突破单一模态信息局限,实现视觉与环境信息的有效融合,为金耳工厂化生产提供技术支撑。

针对金耳生长阶段精准识别的现实需求与技术瓶颈,本文提出一种基于改进 Vision Transformer 与环境权重融合的金耳单目标生长阶段识别方法:构建图像与环境参数同步采集的金耳生长阶段多模态数据集,将金耳生长过程划分为幼耳期、转色期和成熟期 3 个阶段,为后续模型训练与评估提供可靠的数据支撑;设计改进 ViT 图像特征提取网络,在标准 Vision Transformer 基础上引入多尺度特征嵌入、位置增强编码模块与局部感知增强模块,兼顾金耳菌包的全局形态与局部纹理细节,提升对菌包外观细微变化的表征能力;提出环境因子权重融合策略,通过可学习权重对温度、湿度、光照强度和 CO_2 浓度等关键环境因子进行自适应加权,并与图像深度特征在融合层进行联合建模,实现金耳单目标生长阶段的多模态精准识别;最后通过多组对比实验与消融实验,对所提方法进行验证。

1　材料与方法

1.1　实验材料与栽培环境

实验在陕西省咸阳市杨陵区富隆产业园进行,栽培方舱内尺寸为 6 m(长度)×3 m(宽度)×3 m(高度),配备加湿、通风、制冷/制热和补光等栽培装备,方舱实物图如图 1 所示。

实验所用金耳菌棒由西北农林科技大学食用菌中心统一提供,规格为 10 cm(直径)×42 cm(长度)。菌棒经灭菌接种后完成培养熟化,转入方舱



(a) 方舱外景图 (b) 方舱内部图

图 1 方舱实物图

Fig. 1 Physical drawings of cultivation chamber

进行出耳管理。金耳子实体生长主要分为 3 个阶段,分别为幼耳期 S1、转色期 S2、成熟期 S3,各阶段所需的环境参数如表 1 所示。

1.2 数据采集系统与标注方法

1.2.1 图像数据采集

为兼顾成本和实际应用推广性,本研究采用智能手机作为图像采集终端。使用红米 K70 型手机(主摄像头像素 5 000 万,5 000 万像素模式下分辨率 8 192 像素 ×6 144 像素)作为图像采集设备,采集图像时距菌棒 0.3 m;结合手机操作特性与方舱内光照规律,每日固定 2 个时段采集图像(08:30—09:30 和 15:30—16:30),每天采集图像 200 幅左右,采集周期覆盖 2 个生长周期(共 56 d),总共采集了 1.12×10^4 幅金耳图像数据。每次拍摄时,操作者沿通道缓慢移动,在每排菌架前对代表性菌棒进行拍摄,确保图像中至少包含 1 或 2 个金耳实体,并尽量避免严重遮挡。

表 1 金耳各阶段所需的环境参数

Tab. 1 Environmental parameters required for different growth stages of *Tremella aurantialba*

阶段编号	阶段名称	形态特征概述	温度/℃	相对湿度/%	CO_2 浓度/($\mu mol\cdot mol^{-1}$)	光照强度/lx
S1	幼耳期	直径 2 ~ 3 cm,白色或淡黄色	22 ~ 24	75 ~ 85	1 200 ~ 1 500	无
S2	转色期	直径 5 ~ 10 cm,黄色白色相间	20 ~ 22	85 ~ 95	800 ~ 1 200	300 ~ 500(散射光)
S3	成熟期	直径 10 ~ 12 cm,完全金黄色,基本没有白色	18 ~ 20	80 ~ 90	800 ~ 1 000	300 ~ 500(散射光)

1.2.2 环境数据采集及预处理

采用多参数传感器阵列采集金耳生长的环境数据,具体传感器布置如图 2 所示,传感器参数如表 2 所示。方舱内共划分 30 个栽培区,每个栽培区指一个独立的菌架层,共 30 个传感器布置点(P1 ~ P30),每个栽培区均布设一组传感器,同步采集温度、相对湿度、光照强度、CO_2 浓度 4 项关键因子,采集周期为 1 次/min,整个采集周期共获取约 2.42×10^6 组环境数据。同时,在图像采集时,会记录其所在栽培区信息,确保每幅图像均能与其所在栽培区对应的环境传感器数据进行精准匹配,从而捕捉金耳生长的微环境特征。

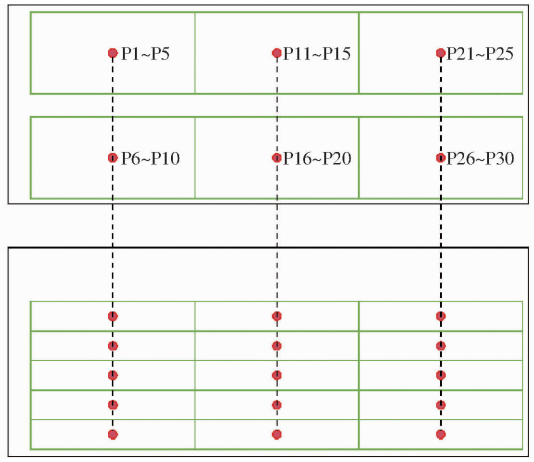

图 2 金耳方舱传感器布置图

Fig. 2 Sensor layout of *Tremella aurantialba* shelter

表 2 传感器参数

Tab. 2 Sensor parameters

仪器名称	测量项目	量程	精度
温湿度传感器	温度	-20 ~ 60℃	±0.3℃
	相对湿度	0 ~ 100%	±2%
二氧化碳传感器	CO_2 浓度	400 ~ 10 000 μmol/mol	±50 μmol/mol
光照传感器	光照强度	0.1 ~ 188 000 lx	±10%

为实现环境数据与图像样本的时间同步,针对在时刻 t 拍摄的每幅图像,本文采用时间窗均值匹配策略:在 t 附近构建时间窗 $[t-\Delta,\ t+\Delta]$,Δ 为时间窗半宽,将时间窗内的环境观测取均值,作为该图像对应的环境特征向量。设 $\boldsymbol{s}(\tau)\in\boldsymbol{R}^m$,$\boldsymbol{R}^m$ 表示时刻 τ 的 m 维环境观测向量,则 t 时刻拍摄的图像对应的环境向量 $\boldsymbol{v}(t)$ 表达式为

$$\boldsymbol{v}(t)=\frac{1}{|\Omega(t)|}\sum_{\tau\in\Omega(t)}\boldsymbol{s}(\tau)$$
$$(\Omega(t)=\{\tau\mid\tau\in[t-\Delta,t+\Delta]\}) \quad (1)$$

式中,$|\Omega(t)|$ 为时间窗内的有效采样点数,由于采样周期为 1 min,当无缺失时 $|\Omega(t)|\approx2\Delta+1$。本文综合考虑采样周期、图像时间戳可能存在的对齐误差以及验证集性能,最终取 $\Delta=5$ min。

当时间窗内环境数据存在缺失时,采用线性插值(Linear interpolation)对缺失点进行补全;若缺失发生在序列首尾导致无法插值,则采用最近邻值填充进行补齐。完成缺失值处理后,再进行时间窗均值计算以获得 $\boldsymbol{v}(t)$,从而降低个别时刻传感器丢包对环境特征表示的影响。为消除不同环境变量量纲与取值范围差异,各环境变量仅在训练集上进行标

准化处理，具体处理公式为

$$\hat{x} = \frac{x - \mu}{\sigma} \tag{2}$$

式中　$\hat{x}$——标准化后的环境观测值

x——某个环境变量的原始观测值

μ——训练集对应变量的均值

σ——训练集对应变量的标准差

验证集与测试集使用训练集得到的 μ、σ 执行相同变换，以避免数据泄露。

1.2.3　生长阶段标注与数据集构建

在完成原始图像采集与环境数据记录后，首先对金耳子实体进行目标级标注。针对每幅原始图像，采用 LabelImg 等标注工具，由具备生产经验的技术人员逐一框选出图像中所有金耳子实体，得到其最小外接矩形边界框（Bounding box）。同时，依据菌盖大小、形态饱满程度、边缘卷曲情况及颜色等外观特征，对每个金耳目标进行生长阶段标注，将其统一划分为幼耳期、转色期和成熟期 3 个阶段。对于光照极端、严重遮挡或阶段判断存在较大分歧的样本，经人工复核后予以剔除，确保标注数据的准确性和一致性。

在完成目标级边界框标注的基础上，为与后续基于 Transformer 的单目标分类模型相匹配，采用基于标注框的区域裁剪（ROI cropping）策略，将每个标注的金耳子实体从原始图像中单独裁剪出来，生成以单朵金耳为中心的局部图像块。裁剪时在边界框周围适当扩展一定像素范围，以保留少量菌棒及背景信息，增强模型对空间上下文的感知能力。将裁剪后的单朵金耳图像与时间同步匹配预处理后的环境特征向量进行关联，构建形成金耳多模态样本集。最终经样本筛选、数据关联与校正，得到有效单朵金耳多模态样本 50 420 份，其中幼耳期 14 421 份、转色期 19 768 份、成熟期 16 231 份。

本研究以原始采集大图为基本单元进行分层抽样，按比例 7∶2∶1将多模态样本集划分为训练集、测试集与验证集，确保各数据集内金耳生长阶段的样本分布与总体一致，各数据集样本数量统计如表 3 所示。

表 3　金耳各阶段数量

Tab. 3　Quantity of *Tremella aurantialba* at each growth stage

数据集	幼耳期数量	幼耳期占比/%	转色期数量	转色期占比/%	成熟期数量	成熟期占比/%	数量总计	总计占比/%
训练集	10 095	28. 60	13 838	39. 21	11 362	32. 19	35 295	70
测试集	2 884	28. 59	3 954	39. 21	3 246	32. 19	10 084	20
验证集	1 442	28. 60	1 976	39. 19	1 623	32. 19	5 041	10
总计	14 421	28. 60	19 768	39. 21	16 231	32. 19	50 420	100

1.3　改进 Vision Transformer 图像特征提取网络

传统 Vision Transformer（ViT）模型将输入图像切分为固定尺寸的 Patch 并转换为嵌入向量，整个特征提取过程仅依赖单一尺度的 Patch 表征，属于单尺度特征表征方式，缺乏多尺度特征捕捉能力，无法有效区分金耳不同生长阶段的形态特征差异。为解决这一问题，结合金耳不同生长阶段的形态尺度特性，通过引入多尺度特征嵌入[23]、增强位置编码、局部感知增强三大核心模块，实现金耳不同生长阶段图像特征的精准提取与有效表征，具体结构如图 3 所示。

1.3.1　多尺度特征嵌入

金耳生长为连续过程，幼耳期、转色期、成熟期的形态特征差异显著，且不同阶段的有效识别特征存在尺度差异。本研究提出的多尺度特征嵌入模块替代了标准 Vision Transformer 模型中的初始 Patch Embedding 层，直接对原始输入图像进行特征提取。多尺度特征嵌入通过多感受野特征提取与注意力自适应融合，强化不同生长阶段的特异性特征表达，核心逻辑如图 3 中多尺度特征嵌入模块所示。

首先，金耳输入图像为 $X \in \mathbf{R}^{H \times W \times C}$，其中 H、W、C 分别表示图像的高度、宽度和通道数。针对金耳生长阶段的尺度特征差异，本模块设计了 3 个并行的卷积分支，分别采用 3×3、5×5、7×7 的卷积核对原始输入图像 X 进行特征提取，每个卷积层均设置步长（stride）为 4，填充（padding）为 2，并输出 256 通道的特征图（即嵌入维度 $D = 256$）。这 3 个分支分别捕捉金耳的细微纹理、局部形态与全局轮廓特征，得到多尺度特征图 $X_1 \in \mathbf{R}^{(H/4) \times (W/4) \times D}$、$X_2 \in \mathbf{R}^{(H/4) \times (W/4) \times D}$、$X_3 \in \mathbf{R}^{(H/4) \times (W/4) \times D}$，对应适配金耳不同生长阶段的特征识别需求。

其次，将各尺度特征图经展平（Flatten）后，通过独立全连接层映射为固定维度的嵌入向量，实现多尺度特征的维度统一，本研究中嵌入维度 D 为 256，因此嵌入函数 f_i 定义为：$f_i: \mathbf{R}^{H' \times W' \times C} \rightarrow \mathbf{R}^{256}$（$i = 1, 2, 3$），其中 H'、W' 为对应尺度卷积后特征图的高和宽，具体计算式为

$$\boldsymbol{e}_i = \boldsymbol{W}_i \boldsymbol{F}(X_i) + \boldsymbol{b}_i$$
$$(\boldsymbol{W}_i \in \mathbf{R}^{D \times (H' \times W' \times C)}, \boldsymbol{b}_i \in \mathbf{R}^{256}) \tag{3}$$

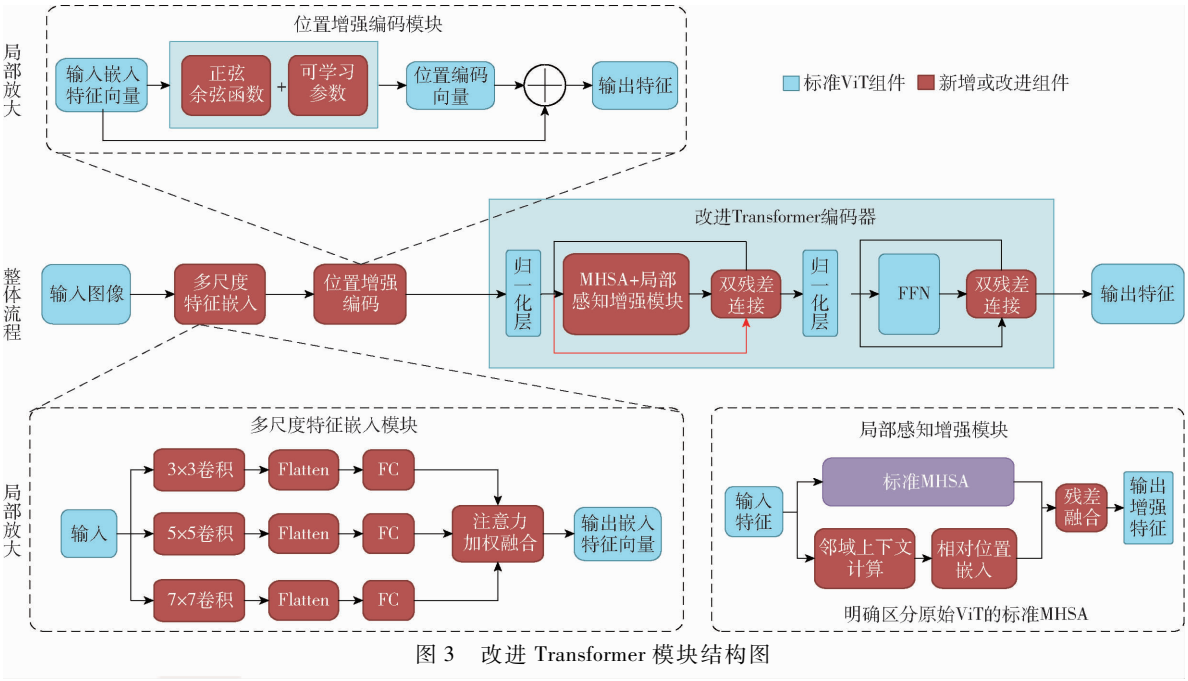

图 3 改进 Transformer 模块结构图

Fig. 3 Structure diagram of improved Transformer module

式中 $\boldsymbol{e}_i$——第 i 个尺度经过全连接层映射后的嵌入特征向量

X_i——第 i 个尺度特征图

$\boldsymbol{F}(X_i)$——展平后的一维向量

$\boldsymbol{W}_i$——可学习的权重矩阵，针对不同尺度的金耳特征分布进行自适应优化

$\boldsymbol{b}_i$——偏置向量

接着，在获得所有尺度的特征嵌入向量后，利用注意力机制对各尺度特征进行自适应加权融合，突出对金耳生长阶段识别起关键作用的特征信息，注意力权重 $\boldsymbol{\alpha}_i$ 计算公式为

$$\boldsymbol{\alpha}_i = \frac{\exp(\boldsymbol{e}_i \cdot \boldsymbol{q})}{\sum_{j=1}^{3} \exp(\boldsymbol{e}_j \cdot \boldsymbol{q})} \quad (\boldsymbol{e}_i \in \mathbf{R}^{256}, \boldsymbol{q} \in \mathbf{R}^{256}) \tag{4}$$

式中 $\boldsymbol{q}$——可学习查询向量

通过模型训练自动学习金耳不同尺度特征间的内在关联与权重分配规律。

最后，利用计算得到的注意力权重，对各尺度特征嵌入向量进行加权融合，得到包含多尺度有效信息的统一特征表示式为

$$\boldsymbol{E} = \sum_{i=1}^{3} \boldsymbol{\alpha}_i \boldsymbol{e}_i \quad (\boldsymbol{E} \in \mathbf{R}^{256}) \tag{5}$$

式中 $\boldsymbol{E}$——融合后的特征向量

$\boldsymbol{E}$ 整合了金耳不同尺度的形态与纹理特征，能够有效表征不同生长阶段的金耳图像特征。

1.3.2 位置增强编码模块

金耳图像存在纹理复杂、形态多变的特点，传统固定位置编码易忽略图像块空间关联，导致模型在学习全局形态时丢失局部空间信息。为解决这一问题，本研究提出了位置增强编码模块，通过注意力驱动的动态位置编码，强化金耳形态的空间特征捕捉，其结构如图 3 中位置增强编码模块所示。

首先为每个图像块引入与嵌入向量同维度的位置编码向量 $\boldsymbol{p}_i$，再结合正弦余弦函数的周期性与可学习参数，适配金耳不同生长阶段的空间尺度变化，计算式为

$$\boldsymbol{p}_i = \boldsymbol{W}_p \mathrm{PE}(i,:) + \boldsymbol{b}_p \quad (\boldsymbol{W}_p \in \mathbf{R}^{256\times256}, \boldsymbol{b}_p \in \mathbf{R}^{256}) \tag{6}$$

式中 $\boldsymbol{W}_p$、$\boldsymbol{b}_p$——可学习参数

$\mathrm{PE}(i,:)$——基于块索引 i 的位置编码函数

采用正弦余弦组合，捕捉金耳图像块的相对位置关系，计算式为

$$\mathrm{PE}(i,j) = \begin{cases} \sin\left(\dfrac{i}{10\,000^{2k/256}}\right) & (j = 2k) \\ \cos\left(\dfrac{i}{10\,000^{2k/256}}\right) & (j = 2k+1) \end{cases} \quad (0 \leqslant k < 128) \tag{7}$$

式中 j——位置编码向量的维度索引

i——图像块(Patch)的位置索引

k——嵌入向量维度索引，用于控制正弦余弦函数的周期性

将位置编码向量与块嵌入向量叠加，为金耳形态特征注入空间信息，公式为

$$\boldsymbol{z}_i = \boldsymbol{x}_i + \boldsymbol{p}_i \tag{8}$$

式中 $\boldsymbol{x}_i$——第 i 个图像块的原始嵌入向量

$\boldsymbol{z}_i$——ViT 编码器的输入，确保模型在学习金耳全局形态时不丢失局部空间关联

1.3.3 局部感知增强模块

ViT 模型基于 Transformer 结构，编码器由多个 Transformer 块组成，每个块包含多头自注意力机制（Multi-head self-attention，MHSA）和前馈网络（Feed-forward network，FFN），用于捕捉全局特征[15]。为增强模型对金耳细微形态差异的敏感性，在标准 Transformer 编码器块中引入了局部感知增强模块。该模块通过残差连接的方式，将计算得到的局部上下文特征融合到自注意力输出中，其改进后的结构如图 3 中局部感知增强模块所示。

（1）局部上下文特征计算：针对每个图像块 X_i，计算其在展平序列中邻域内的局部上下文特征 $\boldsymbol{L}_{Pi}$，$\boldsymbol{L}_{Pi} \in \mathbf{R}^D$ 聚焦金耳局部细节。该特征的计算式为

$$\boldsymbol{L}_{Pi} = \sum_{j=i-L;\ j\neq i}^{i+L} \boldsymbol{\alpha}_{ij}\,\mathrm{RelPosEmbed}(j-i) \tag{9}$$

式中　L——求和范围，覆盖了当前块 i 前后共 6 个邻域块（即 $j \in \{i-3, i-2, i-1, i+1, i+2, i+3\}$，当索引超出序列范围时进行边界处理，如截断），取 3

$\boldsymbol{\alpha}_{ij}$——标准 ViT 自注意力机制中，查询块 i 对键块 j 计算得到的注意力权重（经过 Softmax 归一化后的标量）

$\mathrm{RelPosEmbed}(j-i)$——可学习的、维度为 D（256）的相对位置嵌入向量，它捕获了邻域块 j 相对于当前块 i 的空间偏移信息

$\boldsymbol{L}_{Pi}$——计算得到的局部上下文特征向量，它聚合了邻域内所有块的注意力权重所加权的相对位置信息

（2）多头自注意力与局部增强特征融合：首先，多头自注意力（MHSA）模块接收 Transformer 块的输入特征 $\boldsymbol{Z}_i$，并按照标准 Transformer 流程计算得到自注意力输出 $\boldsymbol{A}_{tti} \in \mathbf{R}^D$。

随后，将上述计算得到的局部上下文特征 $\boldsymbol{L}_{Pi}$ 与自注意力输出 $\boldsymbol{A}_{tti}$ 进行元素级加法融合，形成局部增强特征 $\boldsymbol{L}_{EFi}$。这种融合以残差连接的形式突出金耳局部细节，更新公式为

$$\boldsymbol{L}_{EFi} = \boldsymbol{A}_{tti} + \boldsymbol{L}_{Pi} \quad (\boldsymbol{L}_{EFi} \in \mathbf{R}^D) \tag{10}$$

（3）Transformer 块内部残差与前馈网络：局部增强后的特征 $\boldsymbol{L}_{EFi}$ 将首先与 Transformer 块的原始输入 $\boldsymbol{Z}_i$ 进行第 1 次残差连接，形成融合特征 $\boldsymbol{F}_1$，公式为

$$\boldsymbol{F}_1 = \boldsymbol{Z}_i + \boldsymbol{L}_{EFi} \tag{11}$$

随后，这个融合特征 $\boldsymbol{F}_1$ 将作为输入传递给前馈网络（FFN）。该 FFN 由 2 个线性层和 1 个激活函数组成，负责对每个 token 进行独立的非线性变换，其输出记为 $\boldsymbol{F}_2$。

（4）最终残差连接与层归一化：为了稳定训练并保持特征表达能力，在 Transformer 块的末端引入第 2 次残差连接。该连接将 FFN 的输出 $\boldsymbol{F}_2$ 与 FFN 的输入 $\boldsymbol{F}_1$ 相加，然后进行层归一化（Layer Normalization）。这个过程生成了 Transformer 块的最终输出 E_{final}，计算公式为

$$E_{\mathrm{final}} = \mathrm{LayerNorm}(\boldsymbol{F}_1 + \boldsymbol{F}_2) \tag{12}$$

这种双重残差和层归一化的结构有助于缓解深层网络中的梯度消失问题，确保金耳的细微形态特征在深层网络中得以保留和有效传递。

1.4 环境权重生成模块

环境分支以归一化后的温度、相对湿度、光照强度和 CO_2 浓度向量（记为 $\boldsymbol{x}_{\mathrm{norm}} \in \mathbf{R}^4$）为输入，其目标不是直接生成一个融合特征，而是动态地为图像特征生成通道级的调制权重。首先，归一化后的环境因子 $\boldsymbol{x}_{\mathrm{norm}}$ 经过一个多层感知机（MLP Encoder），映射到一个中间环境特征 $\boldsymbol{f}_{\mathrm{env_raw}} \in \mathbf{R}^{D'}$。这个中间特征捕捉了环境因子之间的复杂非线性关系。随后，将中间环境特征 $\boldsymbol{f}_{\mathrm{env_raw}}$ 送入一个权重生成模块（Generating channel-wise weights）。该模块通常由一个或多个全连接层构成，并最终通过 Sigmoid 激活函数输出一个与图像特征维度相同的环境权重向量 $\boldsymbol{w}_{\mathrm{env}} \in \mathbf{R}^D$。Sigmoid 激活函数将权重限制在 [0,1] 之间，表示每个通道的贡献度。这个过程可表示为

$$\boldsymbol{w}_{\mathrm{env}} = \mathrm{Sigmoid}(MLP_{\mathrm{weight}}(\boldsymbol{f}_{\mathrm{env_raw}})) \tag{13}$$

式中　$\boldsymbol{f}_{\mathrm{env_raw}}$——经 MLP Encoder 编码后的中间环境特征

MLP_{weight}——权重生成模块中的多层感知机

$\boldsymbol{w}_{\mathrm{env}}$——最终生成的环境权重向量，其维度 D 与图像分支的输出特征 $\boldsymbol{f}_{\mathrm{img}}$ 的维度相同

Sigmoid 激活函数确保权重在 [0,1] 范围内，实现对图像特征通道的软门控（soft-gating）调制。通过这种方式，环境分支能够根据实时的环境条件，自适应地调整图像特征中每个通道的重要性，而非简单地对环境因子本身进行加权。

1.5 环境权重调制与分类

模型采用双分支结构实现金耳生长阶段的预测，具体结构如图 4 所示：图像分支以金耳 RGB 图像为输入，经预处理、多尺度特征嵌入、位置增强模块、改进 Transformer 编码器等流程，输出全局特征向量 $\boldsymbol{f}_{\mathrm{img}} \in \mathbf{R}^D$；环境分支则基于环境传感器数据，依次经标准化、MLP 编码，最终通过权重生成模块得到通道级的环境权重向量 $\boldsymbol{w}_{\mathrm{env}}$。

随后，图像特征 $\boldsymbol{f}_{\mathrm{img}}$ 与环境权重 $\boldsymbol{w}_{\mathrm{env}}$ 在特征层进

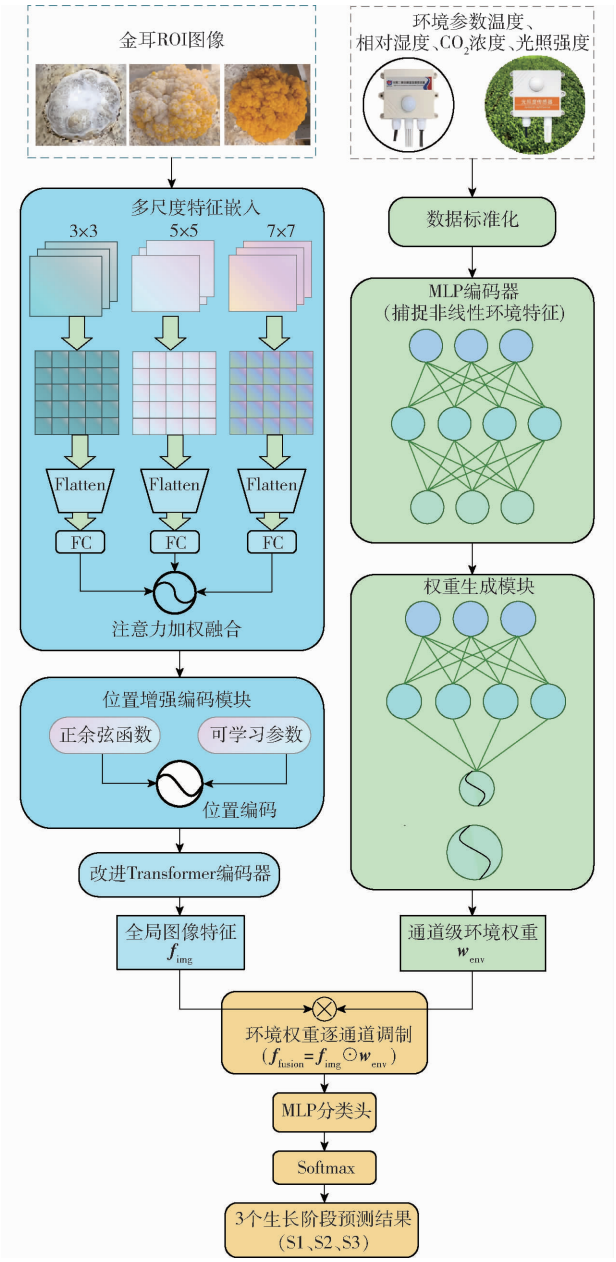

图 4 多模态模型结构图

Fig. 4 Structure diagram of multimodal model

行环境加权调制(Environmental weighted modulation)。这种调制机制通过逐元素乘法实现,动态地调整图像特征中每个通道的强度,以响应当前的环境条件,计算公式为

$$\boldsymbol{f}_{\text{fusion}}=\boldsymbol{f}_{\text{img}}\odot\boldsymbol{w}_{\text{env}} \tag{14}$$

式中 $\odot$——逐元素乘法符号

融合特征$\boldsymbol{f}_{\text{fusion}}$经多层感知机(MLP)和 Dropout、LayerNorm 等结构进一步压缩与重构,最终送入三分类输出层。该输出层通过线性变换和 Softmax 激活函数,得到 3 个生长阶段的预测概率,计算公式为

$$\hat{\boldsymbol{y}}=\text{Softmax}(\boldsymbol{W}\boldsymbol{f}_{\text{fusion}}+\boldsymbol{b})$$
$$(\boldsymbol{W}\in\mathbf{R}^{3\times D},\boldsymbol{b}\in\mathbf{R}^{3},\hat{\boldsymbol{y}}\in\mathbf{R}^{3}) \tag{15}$$

式中 $\boldsymbol{W}$——分类层权重矩阵

$\boldsymbol{b}$——分类层偏置向量

$\hat{\boldsymbol{y}}$——3 个阶段(幼耳期、转色期、成熟期)的预测概率分布

模型选择概率最高的阶段作为最终预测结果。

1.6 模型训练与评价指标

1.6.1 损失函数与优化器

采用交叉熵损失函数作为训练目标,损失函数$\mathscr{L}$公式为

$$\mathscr{L}=-\sum_{k=1}^{3}y_k\ln(\hat{\boldsymbol{y}}_k) \tag{16}$$

式中 y_k——真实标签的 one-hot 编码

$\hat{\boldsymbol{y}}_k$——模型预测概率

优化器采用 AdamW,初始学习率设为1×10^{-4},权重衰减系数为1×10^{-5},用以抑制过拟合;动量参数设置为$\beta_1=0.9$、$\beta_2=0.999$。采用余弦退火学习率衰减策略,训练轮数为 100 轮,批量大小为 32。

1.6.2 实验平台与软件环境

实验平台搭载 AMD Ryzen 7 5800H CPU 和 NVIDIA RTX 3060 GPU,显存容量 6 GB,内存 16 GB,操作系统为 Windows 11。深度学习部分基于 Python 3.9 和 PyTorch 1.10 实现,图像处理采用 OpenCV 4.5,传统机器学习与评价指标计算使用 Scikit-learn 0.24.2。

1.6.3 评价指标

为客观评价模型在金耳生长阶段识别任务中的有效性,采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1 值等指标综合衡量分类性能;同时记录模型在独立泛化测试集上的分类准确率,以评估其对不同批次和不同条件下数据的适应能力。

另外,采用浮点运算量(FLOPs)和参数量(Parameters)来衡量模型运算复杂度与资源需求,采用图像处理帧率(FPS)来衡量模型推理速度。

2 实验

2.1 生长阶段识别结果

2.1.1 不同生长阶段识别性能

改进后的模型在金耳生长阶段识别任务中取得了优异且稳定的性能。5 次独立重复实验统计结果如表 4 所示,模型整体宏观平均 F1 值达到(93.2±0.3)%,各指标标准差均小于 0.5 个百分点,证明了模型具备良好的鲁棒性和可重复性。

分阶段来看:幼耳期识别效果最优(F1 值(98.8±0.2)%),这与该阶段金耳原基初步形成、菌丝结构细密、颜色特征鲜明且形态单一密切相关,模型可快速捕捉到稳定且特异的视觉特征;成熟期表现良好(F1 值(91.1±0.5)%),表明模型对金耳子实体

表 4　金耳各生长阶段识别结果对比

Tab. 4　Comparison of recognition results for *Tremella aurantialba* at different growth stages　%

生长阶段	精确率	召回率	F1 值	准确率
幼耳期	99.2 ± 0.3	98.5 ± 0.4	98.8 ± 0.2	
转色期	80.1 ± 1.5	90.5 ± 1.2	85.0 ± 0.8	
成熟期	94.5 ± 0.6	88.1 ± 1.1	91.1 ± 0.5	
所有类别	93.4 ± 0.4	93.2 ± 0.5	93.2 ± 0.3	93.2 ± 0.4

膨大、色泽加深等典型成熟特征学习充分；转色期识别难度相对较大（F1 值（85.0 ± 0.8）%），是金耳生长阶段识别中的主要难点。转色期是金耳营养生长向生殖生长过渡的关键时期，菌丝不断分化、色素逐步合成、子实体形态快速变化，使其同时携带幼耳期与成熟期的双重特征[24]，类间边界模糊，这使其成为典型的细粒度生长过渡态识别问题，对模型的精确区分能力提出了更高要求。

2.1.2　混淆矩阵分析

模型对金耳各生长阶段的分类结果呈现出明显的阶段差异，图 5 为模型识别混淆矩阵。

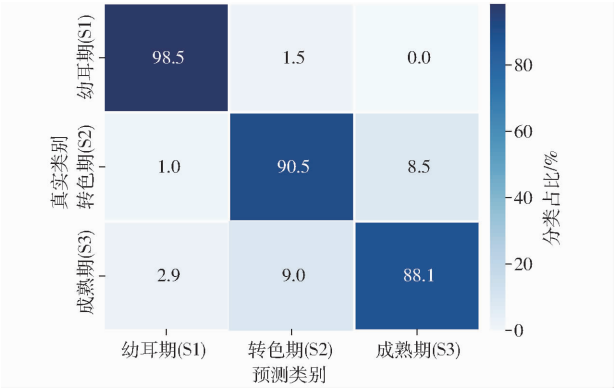

图 5　模型识别混淆矩阵

Fig. 5　Model recognition confusion matrix

对于幼耳期（S1），其召回率高达 98.5%。从混淆矩阵可见，仅有 1.5% 的样本被误分为转色期（S2），且无样本被误分为成熟期（S3），这进一步验证了幼耳期形态特征独特、发育状态稳定、类间差异显著，模型对其具有极高的识别精度。

对于转色期（S2），其召回率为 90.5%。误分样本中，有 8.5% 被错误归为成熟期（S3），另有 1.0% 被错误归为幼耳期（S1）。这直观反映出转色期作为生长过渡阶段，其形态特征的渐变性和类间界限的模糊性导致了与相邻阶段的混淆问题，尤其是与形态发展接近的成熟期。

对于成熟期（S3），其召回率为 88.1%。误分样本中，有 9.0% 被误分为转色期（S2），另有 2.9% 被误分为幼耳期（S1）。这说明在金耳子实体接近成熟但尚未完全转色完成时，模型仍易出现阶段混淆。这进一步验证了金耳在成熟过程中形态与色泽渐进式变化的生理特性，使得在过渡边界处的精确判断更具挑战性。

2.2　消融实验

为验证所提模型各模块的有效性，通过消融实验分析了移除多阶段嵌入、增强位置编码、局部感知增强及环境参数后的性能变化。实验结果如表 5 所示，完整实验组的精确率、召回率、F1 值及准确率均达到 93.0% 以上，移除任一模块均导致模型性能发生显著下降（$p < 0.01$）。

其中，移除环境参数后性能降幅最大（F1 值降至（87.7 ± 0.9）%，下降 5.5 个百分点），说明在金耳生长识别中，环境信息与视觉信息具有强互补性，金耳属于变温结实性、环境敏感型食用菌，温度、相对湿度、CO_2 浓度、光照强度直接调控其菌丝生长、子实体分化与色素合成[6]，因此多模态信息融合更符合食用菌生长识别的内在规律。移除局部感知增强模块后 F1 值降至（89.6 ± 0.7）%，表明金耳菌丝纹理、表面致密程度等细微视觉特征对阶段判别至关重要，局部感知结构能够有效捕捉这类与发育阶段密切相关的精细形态。同时，多尺度特征嵌入与增强位置编码也带来稳定性能增益，证明层次化特征提取与空间结构建模更适应金耳复杂的形态变化过程。各模块协同作用，使模型更贴合农业场景下的生长监测需求。

表 5　消融实验结果对比

Tab. 5　Comparison of ablation experiment results　%

模型	精确率	召回率	F1 值	准确率
完整实验组	93.4 ± 0.4	93.2 ± 0.4	93.2 ± 0.4	93.2 ± 0.4
移除多尺度特征嵌入	91.7 ± 0.5	90.4 ± 0.6	91.1 ± 0.5	90.6 ± 0.5
移除增强位置编码	92.2 ± 0.4	90.6 ± 0.5	91.5 ± 0.4	91.1 ± 0.5
移除局部感知增强	90.2 ± 0.6	89.0 ± 0.7	89.6 ± 0.7	89.2 ± 0.6
移除环境参数（仅图像）	88.4 ± 0.8	87.1 ± 0.9	87.7 ± 0.9	87.3 ± 0.8

2.3　环境因子重要性与可解释性分析

2.3.1　环境因子重要性分析

金耳不同生长阶段环境因子权重如图 6 所示，由图 6 结果可以看出 3 个阶段环境权重分布与金耳生理需求和栽培机理高度一致。幼耳期，温度（0.38）与 CO_2 浓度（0.35）权重最高，光照强度

(0.07)最低。这是因为幼耳期以菌丝快速增殖、原基分化为主,适宜温度与较高 CO_2 浓度有利于菌丝生长,而强光易抑制原基形成[6,25]。转色期,相对湿度(0.40)权重显著升高,光照强度(0.25)权重达到峰值。从栽培生理来看,转色期是色素沉积与子实体形态建成的关键期,高湿可维持细胞膨压、促进色素合成,适宜光照能诱导类胡萝卜素等色素生成[6,26],直接决定金耳商品品质。成熟期,相对湿度(0.42)仍保持高位,CO_2 浓度与温度权重趋于平衡,此时金耳子实体快速膨大,需稳定相对湿度防止菌肉失水、开裂,同时适度降低 CO_2 浓度以避免徒长、畸形。上述结果不仅量化了各阶段环境需求差异,更从农业机理层面解释了环境参数为何能显著提升识别精度,为工厂化金耳栽培的环境调控与智能监测提供了理论依据。

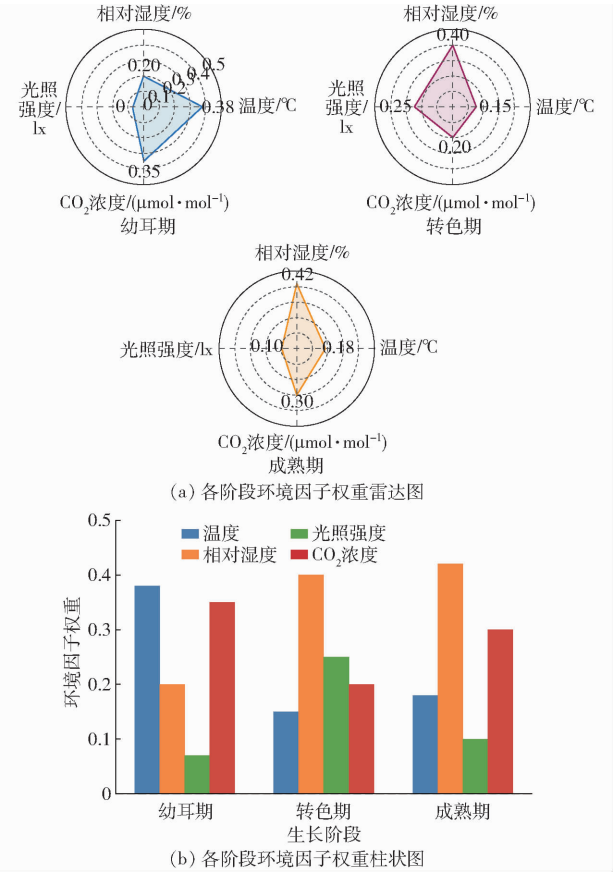

图 6 环境因子权重图

Fig. 6 Environmental factor weight map

2.3.2 不同生长阶段识别 Grad－CAM 热力图分析

为验证模型对金耳生长阶段识别的可解释性,采用 Grad－CAM 方法生成热力图,可视化模型在不同阶段的关注区域,结果如图 7 所示(图中每行对应一个生长阶段,从上至下依次为:幼耳期、转色期、成熟期);热力图中红色区域代表模型高关注区域,蓝色区域代表低关注区域。

幼耳期,模型的高关注区域(热力图中红色部分)精准聚焦于金耳菌体的中央致密区域与边缘菌丝团块,这与幼耳期“原基分化、菌丝快速增殖、菌体呈白绿色块状”的形态特征高度吻合[2],表明模型有效捕捉到了幼耳期最具辨识度的核心形态标志,也为该阶段识别精度最高提供了可解释性依据。转色期,模型重点关注金耳菌体中黄白渐变的过渡区域与局部致密凸起结构,暖色激活区域恰好对应色素沉积与子实体分化的关键部位,这与转色期“局部开始转黄、菌丝致密化、子实体逐步成型”的生理过程高度一致[2,25],说明模型能够有效区分转色期与其他阶段的模糊特征,为解决过渡阶段识别难点提供了可靠支撑。成熟期,模型的高关注区域集中在金耳菌体的饱满瓣状结构与深橙黄色凸起部位,与生产中“子实体耳片完全展开、色泽均匀橙黄、商品品质形成”[1]的采收标准完全契合,验证了模型对成熟期关键特征的精准捕捉能力,可满足金耳采收期智能判别需求。

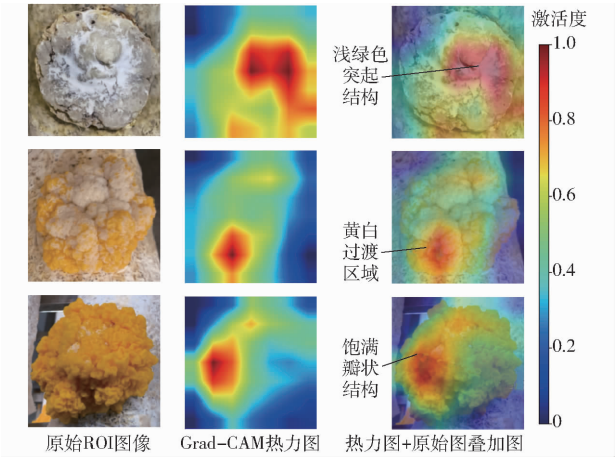

图 7 金耳三阶段识别 Grad－CAM 热力图

Fig. 7 Grad－CAM heatmap for three-stage recognition of *Tremella aurantialba*

整体来看,模型在各阶段的关注区域均聚焦于金耳主体,未对背景菌架等无关区域产生明显激活,识别结果具有科学性与可解释性,为金耳生长阶段的智能识别提供了可靠依据。

2.4 不同模型实验结果对比

为进一步验证所提金耳生长阶段识别模型在方舱化生产、边缘端部署及小农户简易设备应用场景下的综合性能,在统一实验环境下,将改进 ViT＋环境权重融合模型与 ResNet50、原始 ViT、MobileNetV3－Large、EfficientNet－B0 等经典网络进行对比实验。其中,MobileNetV3－Large 与 EfficientNet－B0 为当前移动端与嵌入式设备主流的轻量级骨干网络,更贴合金耳实际生产中低成本、轻量化、可现场部署的应用需求。

由表 6 可以看出，本文所提改进 ViT + 环境权重融合模型在精确率、召回率、F1 值等核心识别指标上均显著优于其他对比模型。该模型 F1 值达到 93.3%，与 ResNet50（CNN 基线）相比提升 8.7 个百分点，与原始 ViT 相比提升 7.2 个百分点，与 MobileNetV3 - Large、EfficientNet - B0 2 种轻量级模型相比分别提升 12.2 个百分点和 10.7 个百分点。传统轻量级 CNN 模型虽结构简单、推理速度快，但在金耳生长阶段这一细粒度、弱差异、过渡特征模糊的识别任务中，特征表达能力存在明显瓶颈，尤其对转色期样本的识别误差较大，难以满足高品质金耳精准识别与智能监测需求。

表 6　不同模型实验结果对比

Tab. 6　Comparison of experimental results for different models

模型	精确率/%	召回率/%	F1 值/%	浮点运算量	参数量	帧率/(f·s^{-1})
ResNet50（CNN 基线）	85.2	84.0	84.6	4.10×10^{9}	2.56×10^{7}	35.8
MobileNetV3 - Large（轻量级）	82.2	80.0	81.1	3.20×10^{8}	5.50×10^{6}	48.2
EfficientNet - B0（轻量级）	83.2	82.0	82.6	4.00×10^{8}	7.10×10^{6}	39.5
原始 ViT	86.8	85.5	86.1	6.73×10^{9}	8.64×10^{7}	18.5
改进 ViT	88.5	87.2	87.8	6.25×10^{9}	4.20×10^{7}	25.6
改进 ViT + 简单环境拼接	91.2	90.0	90.6	6.55×10^{9}	4.30×10^{7}	24.3
改进 ViT + 环境权重融合	93.5	93.2	93.3	6.92×10^{9}	4.40×10^{7}	22.2

从工程部署与实际应用角度分析，MobileNetV3 - Large 推理速度最快（48.2 f/s），参数量仅 5.5×10^{6}，但识别精度偏低；EfficientNet - B0 在精度与速度之间取得一定平衡（39.5 f/s，F1 值为 82.6%），但仍显著低于本文模型。本文所提改进 ViT + 环境权重融合模型参数量仅 4.40×10^{7}，仅为原始 ViT（8.64×10^{7}）的 52.1%，实现了明显轻量化；浮点运算量为 6.92×10^{9}，处于合理区间；推理速度为 22.2 f/s，为原始 ViT 的 1.20 倍，可在嵌入式设备与中端手机上实现实时识别，能够兼顾方舱工厂规模化监测与小农户低成本现场部署的双重需求。

总体而言，改进 ViT + 环境权重融合模型在识别精度、模型轻量化、推理速度与多模态信息融合之间实现了更优的平衡，不仅有效解决了金耳生长过渡阶段识别困难的问题，还具备良好的工程落地潜力，在金耳工厂化栽培与智慧化监测中具有明显优势。

3　结论

（1）针对金耳工厂化生产中生长阶段人工判别存在的诸多弊端，提出基于改进 Vision Transformer 与环境权重融合的多模态生长阶段识别方法，通过构建金耳多模态数据集、优化视觉特征提取网络、设计多模态特征融合策略，实现了金耳幼耳期、转色期、成熟期的精准智能识别。

（2）改进的 Vision Transformer 图像特征提取网络适配金耳菌包图像的特征识别需求，多尺度特征嵌入、动态位置编码与局部感知增强模块的协同作用，有效提升了模型对金耳菌包全局形态、局部纹理细节及空间关联特征的捕捉能力，大幅改善了特征提取的充分性，为高准确率识别奠定了核心基础。

（3）通过多传感器布设与精准匹配策略，有效克服了工厂化环境中环境因素的不均匀性。在此基础上，提出的环境因子自适应权重融合策略实现了视觉与环境信息的深度融合，不仅量化了金耳各生长阶段对不同环境因子的敏感性差异，契合其栽培生理规律（如幼耳期对温度、CO_2 浓度敏感，转色期对相对湿度、光照强度依赖度高，成熟期需稳定高湿），而且多模态信息的互补性显著提升了整体识别准确性，有效缓解了转色期等过渡阶段类间特征模糊的识别难点。

（4）消融实验证实了模型各核心模块的有效性，其中移除环境参数会导致模型 F1 值下降 5.5 个百分点，充分证明环境信息对金耳生长阶段识别的关键作用。局部感知增强模块的引入使金耳生长阶段识别的 F1 值提升了 4.2 个百分点，增强了模型对菌丝纹理等细微特征的敏感性，各模块的协同优化使模型更贴合农业场景下的生长监测需求。

（5）所提方法在识别性能与工程实用性上实现了良好平衡，总体识别准确率达 93.2%，显著优于 ResNet50、原始 ViT 及主流轻量级网络；模型经轻量化设计后参数量仅 4.40×10^{7}，推理速度达 22.2 f/s，可在嵌入式设备与中端手机上实现实时识别，同时适配工厂规模化监测与小农户低成本现场部署的双重需求。

（6）模型具备良好的可解释性，Grad - CAM 热力图显示其在各生长阶段均能精准聚焦金耳核心特征区域，识别逻辑与金耳的形态发育规律高度契合，为识别结果的可靠性提供了直观依据。

（7）本研究结果为金耳工厂化生产的智能监测与精准调控提供了技术支撑，也为食用菌生长阶段识别的多模态研究提供了参考范式。

参 考 文 献

[1] 张怡敏,李颜秘,任潇,等. 金耳多糖的营养价值及开发利用现状[J]. 食品研究与开发,2024,45(19):219 – 224.
Zhang Yimin, Li Yanmi, Ren Xiao, et al. Nutritional value and research progress on development and utilization of *Tremella aurantialba* polysaccharides[J]. Food Research and Development, 2024, 45(19): 219 – 224. (in Chinese)

[2] 王能爽,田爽,马庆芳,等. 金耳生物学特性、栽培技术和多糖功效的研究进展[J]. 中国瓜菜,2025,38(5):14 – 19.
Wang Nengshuang, Tian Shuang, Ma Qingfang, et al. Research progress on biological characteristics, cultivation techniques and polysaccharide efficacy of *Tremella aurantialba*[J]. China Cucurbits and Vegetables, 2025, 38(5): 14 – 19. (in Chinese)

[3] 朱子墨,刘麟然,胡宇航,等. 金耳功能活性研究进展[J]. 生物技术进展,2025,15(6):944 – 951.
Zhu Zimo, Liu Linran, Hu Yuhang, et al. Research progress on functional activities of *Tremella aurantialba*[J]. Current Biotechnology, 2025, 15(6): 944 – 951. (in Chinese)

[4] 李书岭,刘光慧,王璐,等. 珍稀食药用菌金耳方舱式智能化周年栽培技术[J]. 食用菌,2025,47(4):74 – 77.
Li Shuling, Liu Guanghui, Wang Lu, et al. Intelligent annual cultivation technology of rare edible and medicinal fungus *Tremella aurantialba* based on shelter cultivation mode[J]. Edible Fungi, 2025, 47(4): 74 – 77. (in Chinese)

[5] 李海龙. 云仓种植模式在金耳食用菌高效培育中的应用研究[J]. 农业科技与装备,2024(6):49 – 50.
Li Hailong. Application research of cloud warehouse planting mode in efficient cultivation of *Tremella aurantialba*[J]. Agricultural Science & Technology and Equipment, 2024(6): 49 – 50. (in Chinese)

[6] 杨洁,左太强,顾鲁同,等. 环境因子对金耳和羊肚菌生长的调控解析[J]. 蔬菜,2023(10):53 – 56.
Yang Jie, Zuo Taiqiang, Gu Lutong, et al. Regulation analysis of environmental factors on the growth of *Tremella aurantialba* and *Morchella esculenta*[J]. Vegetables, 2023(10): 53 – 56. (in Chinese)

[7] 岳万松,李雪松,刘绍雄,等. 金耳栽培常见问题及对策探讨[J]. 中国食用菌,2022,41(8):80 – 85.
Yue Wansong, Li Xuesong, Liu Shaoxiong, et al. Common problems and countermeasures in the cultivation of *Tremella aurantialba*[J]. Edible Fungi of China, 2022, 41(8): 80 – 85. (in Chinese)

[8] 赵春江,文朝武,林森,等. 基于级联卷积神经网络的番茄花期识别检测方法[J]. 农业工程学报,2020,36(24):143 – 152.
Zhao Chunjiang, Wen Chaowu, Lin Sen, et al. Tomato florescence recognition and detection method based on cascaded neural network[J]. Transactions of the CSAE, 2020, 36(24): 143 – 152. (in Chinese)

[9] Farag M T, Alwaseela A, Gamal E, et al. Using deep convolutional neural network for image-based diagnosis of nutrient deficiencies in plants grown in aquaponics[J]. Chemosensors, 2022, 10(2): 45.

[10] 杨信廷,刘彤,韩佳伟,等. 基于 Swin Transformer 与 GRU 的低温贮藏番茄成熟度识别与时序预测研究[J]. 农业机械学报,2024,55(3):213 – 220.
Yang Xinting, Liu Tong, Han Jiawei, et al. Study on ripeness recognition and time series prediction of tomatoes under low-temperature storage based on Swin Transformer and GRU[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(3): 213 – 220. (in Chinese)

[11] 杨淑珍,朱浩宇,杨凯威. 基于改进 YOLO 11 的海鲜菇生长阶段精细化分类方法[J]. 农业机械学报,2026,57(3):342 – 352.
Yang Shuzhen, Zhu Haoyu, Yang Kaiwei. Fine classification method for *Hypsizygus marmoreus* growth stages based on improved YOLO 11[J]. Transactions of the Chinese Society for Agricultural Machinery, 2026, 57(3): 342 – 352. (in Chinese)

[12] Zhang Qiusi, Zhao Xiangyu, Han Yanyun, et al. Maize yield prediction using federated random forest[J]. Computers and Electronics in Agriculture, 2023, 210: 107930.

[13] Amin J, Anjum A M, Sharif M, et al. Fruits and vegetable diseases recognition using convolutional neural networks[J]. Computers, Materials & Continua, 2022, 70(1): 619 – 635.

[14] 王杨,李迎春,许佳炜,等. 基于改进 Vision Transformer 网络的农作物病害识别方法[J]. 小型微型计算机系统,2024,45(4):887 – 893.
Wag Yang, Li Yingchun, Xu Jiawei, et al. Crop disease identification method based on improved Vision Transformer network[J]. Journal of Chinese Computer Systems, 2024, 45(4): 887 – 893. (in Chinese)

[15] 黄林生,罗耀武,杨小冬,等. 基于注意力机制和多尺度残差网络的农作物病害识别[J]. 农业机械学报,2021,52(10):264 – 271.
Huang Linsheng, Luo Yaowu, Yang Xiaodong, et al. Crop disease recognition based on attention mechanism and multi-scale residual network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(10): 264 – 271. (in Chinese)

[16] 涂雪滢,张佳鹏,钱程,等. 融合 ViT 和通道注意力的水稻病害识别技术研究[J]. 农业与技术,2025,45(7):69 – 74.
Tu Xueying, Zhang Jiapeng, Qian Cheng, et al. Research on rice disease recognition technology integrating ViT and channel attention[J]. Agriculture and Technology, 2025, 45(7): 69 – 74. (in Chinese)

[17] 周丽娟,毛嘉宁. 视觉 Transformer 识别任务研究综述[J]. 中国图象图形学报,2023,28(10):2969 – 3003.
Zhou Lijuan, Mao Jianing. Review on research of visual Transformer recognition tasks[J]. Journal of Image and Graphics, 2023, 28(10): 2969 – 3003. (in Chinese)

(下转第 311 页)

[25] 王守林. 水稻高产栽培技术[M]. 北京:中国盲文出版社,2000.
[26] 熊玉唐. 杂交水稻高产栽培技术[M]. 贵州:贵州科技出版社, 2007.
[27] Chen J, Xiao S, Zhang P, et al. M3-Embedding: multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation[PP/OL]. V4. arXiv (2024 - 06 - 28)[2025 - 09 - 02]. https://doi.org/10.48550/arXiv.2402.03216.
[28] 许浩翔,余敦辉,邓怡辰,等.基于分层强化学习的知识图谱约束问答模型[J].计算机应用,2025,45(12):3764 - 3770.
Xu Haoxiang, Yu Dunhui, Deng Yichen, et al. Knowledge graph constrained question answering model basedon hierarchical reinforcement learning[J]. Journal of Computer Applications, 2025, 45(12): 3764 - 3770. (in Chinese)
[29] 陈汇远.基于知识图谱的列控车载设备故障诊断的研究[J].机车电传动,2025(2):129 - 136.
Chen Huiyuan. Research on fault diagnosis of CTCS on-board equipment based on knowledge graph[J]. Electric Drive for Locomotives,2025(2): 129 - 136. (in Chinese)
[30] Anthropic. Claude 3.7 Sonnet[EB/OL]. [2025 - 02 - 19]. https://www.anthropic.com/claude.
[31] Guo D, Yang D, Zhang H, et al. Deepseek - R1: incentivizing reasoning capability in llms via reinforcement learning[PP/OL]. V1. arXiv (2025 - 01 - 22)[2025 - 09 - 02]. https://doi.org/10.48550/arXiv.2501.12948.
[32] Yang C, Wang R, Jiang J, et al. Nondeterministic polynomial-time problem challenge: an ever-scaling reasoning benchmark for LLMs[PP/OL]. V1. arXiv (2025 - 04 - 15)[2025 - 09 - 02]. https://doi.org/10.48550/arXiv.2504.11239.

(上接第 298 页)

[18] 赵朗月,吴一全.基于机器视觉的表面缺陷检测方法研究进展[J].仪器仪表学报,2022,43(1):198 - 219.
Zhao Langyue,Wu Yiquan. Research progress of surface defect detection methods based on machine vision[J]. Chinese Journal of Scientific Instrument,2022,43(1):198 - 219. (in Chinese)
[19] 李道亮,赵晔,杜壮壮.农业领域多模态融合技术方法与应用研究进展[J].农业机械学报,2025,56(1):1 - 15.
Li Daoliang,Zhao Ye,Du Zhuangzhuang. Research progress and application of multimodal fusion technology in agriculture[J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(1):1 - 15. (in Chinese)
[20] 秦韬,庄卫东.机器视觉技术在现代农业生产中的研究现状[J].现代化农业,2024(4):90 - 93.
Qin Tao,Zhuang Weidong. Research status of machine vision technology in modern agricultural production[J]. Modernizing Agriculture,2024(4):90 - 93. (in Chinese)
[21] 张智淇.基于表型检测与综合环境参数的香菇产量预测方法研究[D].沈阳:沈阳农业大学,2025.
Zhang Zhiqi. Research on lentinus edodes yield prediction methods based on phenotypic detection and comprehensive environmental parameters[D]. Shenyang: Shengyang Agricultural University, 2025. (in Chinese)
[22] Li Y, Che Y, Zhang H, et al. Wheat growth stage identification method based on multimodal data[J]. European Journal of Agronomy,2025,162:127423.
[23] Jia N, Zheng C. Multimodal rapid identification of growth stages and discrimination of growth status for Morchella[J]. Smart Agricultural Technology,2024,9:100507.
[24] 田爽,张云志,马庆芳,等.关于金耳的研究进展分析[J].中国林副特产,2024(4):101 - 102.
Tian Shuang,Zhang Yunzhi,Ma Qingfang,et al. Analysis of research progress on *Tremella aurantialba*[J]. Forest By-Product and Speciality in China,2024(4):101 - 102. (in Chinese)
[25] 刘源,王民乐,陈靖,等.冀北地区冷棚栽培金耳关键技术[J].中国食用菌,2021,40(12):89 - 92.
Liu Yuan,Wang Minle,Chen Jing,et al. Key cultivation technologies of *Tremella aurantialba* in cold greenhouses in northern Hebei Province[J]. Edible Fungi of China,2021,40(12):89 - 92. (in Chinese)
[26] 沈真辉,罗祥英,赵仙伟,等.利用转录组分析光照对金耳子实体转色的影响[J].食用菌学报,2023,30(6):12 - 27.
Shen Zhenhui, Luo Xiangying, Zhao Xianwei, et al. Transcriptome analysis of the effect of light on color transformation of *Tremella aurantialba* fruiting bodies[J]. Acta Edulis Fungi,2023,30(6):12 - 27. (in Chinese)