doi:10.6041/j.issn.1000-1298.2026.12.024

基于最优视图选择和 VGGT 架构的兰州百合种球质量、体积估测

张晶晶[1]　尚永强[2]　杜清国[2]　景　旺[3]　黄海博[3]　刘　燕[3]　丁永军[3]

（1. 兰州城市学院信息工程学院，兰州 730070；2. 甘肃爽口源生态科技股份有限公司，兰州 730059；
3. 西北师范大学人工智能与计算机学院，兰州 730070）

摘要：兰州百合是我国唯一的甜百合，其种球质量、体积是影响成活率、产量及商品价值的重要因素。为解决种球形态不规则、人工表型测量费时费力的问题，本文提出了一种融合最优视图选择和 VGGT 架构的三维表型估测方法。利用信息熵、新点率、颜色差异、全局特征、视角冗余构建多维评价指标，筛选出由 3 幅图像构成的最优视图集，作为 VGGT 输入，端到端的生成相机参数及三维点云，并结合 Solov2 分割掩码进行三维投影，实现种球点云分割；进而提取 14 个 3D 形状特征，通过逐步回归消除共线性影响，构建种球质量、体积线性预测模型，而且在随机森林特征重要性筛选的基础上又构建了 SVR、KNN、GBR、BPNN 等非线性模型。结果表明，3 个最优视图可实现 90% 以上的几何与颜色覆盖；BPNN 质量、体积预测模型的 R^2 均超过 0.95，MAPE 低于 10%。另外，在 GeForce RTX 3090 平台，本研究提出的种球三维重建仅耗时约 3 250 ms，4 s 内便可完成 1 个样本的三维测量。研究结果为兰州百合种球自动化分选与品质评估提供了技术支撑。

关键词：兰州百合；最优视图；VGGT；质量估测；体积估测

中图分类号：S24　　**文献标识码**：A　　**文章编号**：1000-1298(2026)12-0254-10　　**OSID**：

Weight and Volume Estimation of Lanzhou Lily Bulbs Based on Next Best View and VGGT Framework

ZHANG Jingjing[1]　SHANG Yongqiang[2]　DU Qingguo[2]　JING Wang[3]　HUANG Haibo[3]　LIU Yan[3]　DING Yongjun[3]

(1. *School of Information Engineering, Lanzhou City University, Lanzhou* 730070, *China*
2. *Gansu Shuang kouyuan Ecological Technology Co., Ltd., Lanzhou* 730059, *China*
3. *College of Artificial Intelligence and Computer Science, Northwest Normal University, Lanzhou* 730070, *China*)

Abstract: Lanzhou lily is the only sweet lily variety in China, and the weight and volume of its bulbs are key indicators affecting survival rate, yield, and commercial value. To address the challenges of irregular bulb morphology and the low efficiency of manual phenotypic measurements, a three-dimensional phenotypic estimation method that integrated best view selection with the VGGT architecture was proposed. A multidimensional evaluation metric system was constructed based on information entropy, new-point rate, color difference, global features, and view redundancy to select an optimal set of three images as VGGT inputs. The model then performed end-to-end estimation of camera parameters and 3D point clouds. By incorporating Solov2 segmentation masks for 3D projection, precise bulb point cloud segmentation was achieved. Subsequently, totally 14 3D shape features were extracted, and stepwise regression was employed to eliminate multicollinearity effects and build linear prediction models for bulb weight and volume. Furthermore, based on feature importance ranking from a random forest model, nonlinear models, including SVR, KNN, GBR, and BPNN were developed. Experimental results demonstrated that the optimal three-view set achieved over 90% geometric and color coverage. The BPNN models for weight and volume prediction achieved R^2 values above 0.95 and MAPE values below 10%.

收稿日期：2025-10-18　修回日期：2026-01-26

基金项目：国家自然科学基金项目(32360434)、甘肃省高校产业支撑计划项目(2023CYZC-10)、甘肃省自然科学基金项目(23JRRA705)和兰州市哲学社会科学规划项目(25-B88)

作者简介：张晶晶(1981—)，女，讲师，主要从事计算机技术与应用研究，E-mail：lzcuzhangjj@163.com

通信作者：丁永军(1978—)，男，教授，博士，主要从事农业信息化技术研究，E-mail：dingyj@nwnu.edu.cn

In addition, on a GeForce RTX 3090 platform, the proposed 3D reconstruction process required only took about 3 250 ms, enabling complete 3D measurement of a single sample within 4 s. The research result can provide an efficient and practical technical approach for the automated sorting and quality assessment of Lanzhou lily bulbs.

Key words: Lanzhou lily; next best view; VGGT; weight estimation; volume estimation

0　引言

兰州百合是我国唯一的甜百合，种植面积约 1.58×10^4 hm^2，年产量达 1.116×10^5 t，产值 10.06 亿元，是兰州重要经济作物[1]。种球质量及体积是决定兰州百合成活率、产量与商品价值的核心指标。然而，种球形态不规则、根须分布复杂且易受损伤，传统人工测量费时费力，难以无损、精准地获取种球质量及体积。因此，开展兰州百合种球质量、体积无损估测研究，对百合标准化种植、自动化分选与品质评估具有重要意义。

种子的表型参数估测是满足高通量育种筛选的关键技术，而机器视觉具备自动化采集、批量处理特点，已成为该领域研究热点。早期，众多研究对俯视图像进行分割、形态学处理，提取长度、宽度、等效直径等 2D 特征[2-3]。但百合种球为众多鳞片聚合而成的类球体，二维图像缺乏深度信息，难以获得准确的表型参数估计。

三维测量技术可以准确生成坐标、方位和距离估计，更易估计植物形态、生长和生物量[4-6]。激光三角测量[7]、地面激光扫描[8]、飞行时间相机[9]、结构光[10]等主动三维测量，受专用设备和辐射安全限制，应用成本高。被动测量包括双目立体视觉和运动恢复结构/多视图重建（Structure from motion, SfM; Multi-view stereo, MVS）[11]等，双目立体视觉虽可实时计算，但依赖于相机参数的精确标定，易受环境影响；运动恢复结构则需要通过连续图像估计相机参数，计算量大、特征匹配易导致累积误差。

百合种球表面较光滑均匀，颜色单一，小范围内特征变化不显著，往往导致传统特征匹配方法失效。深度学习则具有高度抽象和复杂特征表示能力，为多视图三维重建提供了新的技术路径，包括 MVSNet[12]、Cascade Cost Volume[13]、GeoMVSNet[14]等基于代价体正则化的深度图估计方法；PatchmatchNet[15]、DUSt3R[16]、MASt3R[17]等基于体素或点云的直接回归法；Geo-Neus[18]、NerfingMVS[19]等隐式函数法。但上述方法多要求通过传统运动恢复结构获取相机内参、外参，而百合种球表面纹理特征稀疏、光学特性均匀，特征匹配方法易失效，相机参数估计偏差严重。另外，模型训练依赖于大量标注样本，需要通过高精度三维扫描设备获取 3D 真值，操作复杂、成本昂贵，模型泛化能力有限。VGGT（Visual geometry grounded transformer）三维重建模型是在 Co3Dv2、ScanNet 等大规模数据集上预训练的 Transformer 架构，同时引入“帧内-全局”交替注意力机制，能够从多视图图像直接预测相机内参、外参、深度图、3D 点云，有效解决上述问题[20]。

在作物三维重建的视角数量与表型参数反演研究方面，陈允琳等[21]使用 6 个视角，在 60 s 内基于 Kinect V3 型传感器实现了叶菜类作物三维重建，点云间平均距离误差为 0.296 5 cm；李添天等[22]则通过 300 个视角图像，结合改进神经辐射场实现植物表型参数测量，重建时间为 211 s，平均标准差为 0.001 8。可以看出，三维重建耗时是研究成果实用化的主要瓶颈，有必要针对百合种球和 VGGT 模型开展最优视图数目研究，提高时空效率。同时，相关研究也证实，长、宽、高、点云表面积、点云投影体积、最小凸包体积[23]以及横径、纵径、球形度、饱满度[24]等特征均具备表型参数反演能力，为百合种球表型参数估测提供了可行方向。

针对上述问题，本研究以兰州百合种球为研究对象，基于 VGGT 架构确定三维重建最优视角，提取等效球直径、长高比等 14 个形状特征，进而建立种球质量与体积估测模型，以期为种球自动化分选与品质评估提供技术支持。

1　材料与方法

1.1　试验材料与图像采集

于 2025 年 3 月在甘肃爽口源生态科技股份有限公司日光温室，收集形状各异、大小不等的兰州百合种球 500 枚作为试验材料。为便于后续开发面向农业种植户的种球表型系统，选用 600 mL 百事可乐瓶盖（直径为 32.2 mm，高度为 12.8 mm）作为标定物，用于标准化图像尺度。图像采集则使用自行搭建的拍摄平台进行，该平台由载物台和电控旋转圆环组成。载物台位于圆环中心，固定不动；电控旋转圆环上安置 2 台 Intel RealSense D435 型相机。拍摄时，将种球置于载物台上，一台相机与种球中部水平，另一台呈 25°俯角，驱动圆环匀速旋转 360°，每

10°触发双相机同步拍摄,最终为每颗种球拍摄 72 幅多视角图像,图像索引号为 1 ~ 72,作为三维重建最优视角选择基础。

1.2 性状测量与数据处理

图像采集完成后,使用经校准的 YHM 型电子秤(英衡电子科技有限公司,精度为 0.01 g)逐个测量种球质量,再使用量筒排水法测量体积。为了获取图像分割掩码,实现点云分割,构建百合实例分割数据集。将采集到的图像导入 MVTec Deep Learning Tool 进行智能标注,生成 Halcon 字典文件,再编写 Python 脚本递归解析 Halcon 字典,提取图像基础信息、种球类别、区域掩模等数据,并转换为 COCO 数据集通用的 JSON 格式,供迁移训练 Solov2[25] 分割模型使用。

1.3 基于 VGGT 的三维重建

VGGT 是一种端到端的前馈式 Transformer 模型,如图 1 所示,能直接从单幅或多幅图像中推理出场景三维属性,包括相机参数、深度图、点云。设输入为一组图像序列 $\{I_i\}_{i=1}^{N}$,$I_i \in \mathbb{R}^{3\times H\times W}$,VGGT 定义了一个映射函数,即

$$f(\{I_i\}_{i=1}^{N}) = \{(g_i, D_i, P_i)\}_{i=1}^{N} \tag{1}$$

其中 $g_i \in \mathbb{R}^9$ $D_i \in \mathbb{R}^{H\times W}$ $P_i \in \mathbb{R}^{3\times H\times W}$

式中 g_i——相机参数

D_i——深度图 P_i——点云

点云坐标采用第一帧相机 g_1 的坐标系作为世界参考系,从而保证视角不变性。在特征提取阶段,每幅输入图像先被划分为不重叠的 patch,并通过 DINO 编码器映射为 tokens[26],即

$$t_i = \phi(I_i) \in \mathbb{R}^{K\times C} \tag{2}$$

其中 $K = HW/p^2$

式中 K——patch 数量 p——patch 大小

C——特征维度

为了预测相机参数,每帧额外引入 1 个相机 token $t_i^g \in \mathbb{R}^{1\times C}$,最终拼接得到 $z_i = [t_i; t_i^g] \in \mathbb{R}^{(K+1)\times C}$。所有帧的 tokens 合并后输入由全局注意力(Global attention)和逐帧注意力(Frame attention)组成的 Transformer 结构,即

$$z^{(l+1)} = \mathrm{FrameAttn}(\mathrm{GlobalAttn}(z^{(l)})) \tag{3}$$

式中 FrameAttn 提取帧内特征,GlobalAttn 则用于特征的跨帧融合。最终,经 L 层变换,将特征向量分别送入相机预测头和 DPT 预测头,输出相机位姿、深度图、点云。

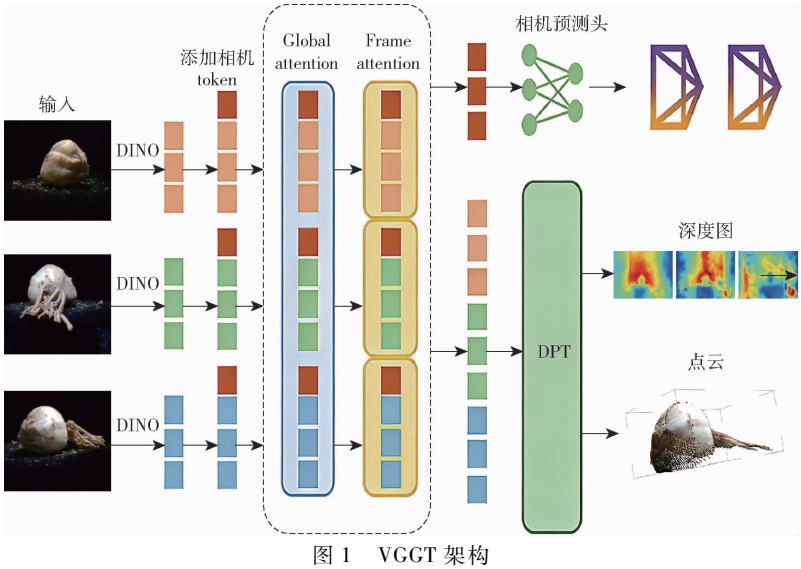

图 1 VGGT 架构

Fig. 1 Architecture of VGGT

1.4 最优视图选择

每个百合种球被密集采样 72 次,如果使用所有图像进行 VGGT 重建虽能覆盖种球 360°外观信息,但计算成本高、受噪声累积影响易出现重影断层现象,很难应用于生产实际。为此,利用信息熵、新点率、颜色差异、全局特征、视角冗余构建多维评价指标,进行最优视图选择。首先选择信息熵最大的图像 v_{init},加入最优视图集 $S = \{v_{init}\}$,并将 S 送入 VGGT 模型获取相机外参 $\boldsymbol{T}_{init}$、内参 $\boldsymbol{K}_{init}$、深度图 d_{init},再通过逆投影公式将像素坐标转换至三维空间,构建初始点云,公式为

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \boldsymbol{T}_{init}\boldsymbol{K}_{init}^{-1}\begin{bmatrix} ud_{init}(u,v) \\ vd_{init}(u,v) \\ d_{init}(u,v) \\ 1 \end{bmatrix} \tag{4}$$

在迭代过程中,引入综合评估函数对未选视图进行价值排序,逐步扩充最优视图集 S,即

$$F(v_i) = w_{new}R_{new}(v_i) + w_{rgb}\Delta E_{00}(v_i) + w_{glob}D_{glob}(v_i) + w_{angle}P_{angle}(v_i) \tag{5}$$

其中　$P_{angle}(v_i) = (\alpha_{\min}/180°)^{0.1}$

式中　$R_{new}(v_i)$——新点率

$\Delta E_{00}(v_i)$——颜色差异均值

$D_{glob}(v_i)$——余弦距离

$P_{angle}(v_i)$——视角惩罚项

新点率表征新视图为点云带来的增益；$\Delta E_{00}(v_i)$为CIEDE2000 颜色空间中，候选视图与已选视图间的颜色差异均值，差异值越大说明候选视图越能补充颜色细节；$D_{glob}(v_i)$为候选视图与已选视图全局特征间的余弦距离，距离越大说明视觉互补性越强；根据相机外参确定每个视图的相机指向，计算候选视图与所有已选视图指向向量间夹角，取其中最小项为 $\alpha_{\min}$，求得视角惩罚项，即视角越近，惩罚越强，有效减少视角冗余，保证所选视图尽可能从不同角度覆盖目标。试验设置 $w_{new}=0.5$、$w_{rgb}=0.3$、$w_{glob}=0.2$、$w_{angle}=0.4$ 以平衡各维度贡献，每次迭代选择 $F(v_i)$最大的视图加入 S，更新点云。当相邻两次的综合评估差小于 0.03 时，停止迭代。

1.5　点云分割

采用“二维分割 + 三维投影裁剪”的两阶段点云分割算法进行种球提取。第 1 阶段采用 Solov2 模型对图像进行实例分割，获取高质量掩膜。第 2 阶段，利用 VGGT 生成的深度估计 $D(x,y)$和相机内外参数$(\boldsymbol{K},\boldsymbol{R},t)$，对掩膜信息进行三维反投影。给定像素点$(x,y)$，其对应的三维点可表示为

$$P(x,y)=\boldsymbol{R}^{-1}(D(x,y)\boldsymbol{K}^{-1}[x,y,1]^{\mathrm{T}}-t) \quad (6)$$

从而得到点云集合 $p=\{P(x,y)\}$。结合 Solov2 分割掩膜 $M(x,y)$，仅保留目标区域，获得种球点云，即

$$p'=\{P(x,y)\in p \mid M(x,y)=1\} \quad (7)$$

最后，对点云应用离群点去除，进一步提升点云的质量。

1.6　特征提取

对目标点云进行 3D 形状特征提取。使用 Alpha Shape 算法进行表面重建，即基于 Delaunay 三角化从点云中提取具有拓扑一致性的封闭曲面，再根据 Alpha 值控制外接球半径，筛选出有效三角形面元[27]。随后，采用主成分分析（PCA）进行坐标标准化，即对网格顶点集合进行协方差矩阵的特征值分解，获得 3 个主轴方向，并将顶点坐标变换至 PCA 坐标系，同时构造轴对齐包围盒（AABB）。最后，提取 14 个 3D 形状特征，如表 1 所示。

1.7　建模方法

采用线性和非线性 2 类方法，建立百合种球质量及体积预测模型。线性建模先剔除方差膨胀因子（VIF）大于10的高共线性特征，再以赤池信息准则（AIC）最小为标准，执行前向逐步回归。非线性建模，先用随机森林（RF）节点分裂时的均方误差减少量计算各特征重要性，再逐步消除重要性最低的特征，每次剔除后重新训练模型，保留模型精度突变前的特征子集，随后使用支持向量回归（Support vector regression，SVR）、K-近邻（K-nearest neighbors，KNN）回归、随机森林（Random forest，RF）回归、梯度提升（Gradient boosting，GB）回归、反向传播神经网络（Backpropagation neural network，BPNN）构建预测模型，模型超参数设置如表 2 所示。

表 1　3D 形状特征定义

Tab. 1　Definition of 3D shape features

特征名称	计算公式	含义
长度	$L=\max(x_{align})-\min(x_{align})$	包围盒（AABB）长度
宽度	$W=\max(y_{align})-\min(y_{align})$	包围盒（AABB）宽度
高度	$H=\max(z_{align})-\min(z_{align})$	包围盒（AABB）高度
表面积	$S=\sum Area(T_i)$	三角形面元的面积之和
切片体积	$V=\sum_{k=1}^{n}A_k\frac{z_{\max}-z_{\min}}{n}$	切片面积与切片间距乘积之和
长宽比	$r_{LW}=L/W$	包围盒（AABB）长宽之比
长高比	$r_{LH}=L/H$	包围盒（AABB）长高之比
等效球直径	$d_s=\sqrt[3]{\frac{6v}{\pi}}$	与切片体积相等的球体直径
球形度	$\psi=\frac{\pi^{1/3}(6V)^{2/3}}{S}$	等效球体表面积与实际表面积之比
面积体积比	$\gamma=\frac{S}{V}$	反映形状紧凑性
片状度	$\phi_f=\frac{H}{W}$	反映形状扁平程度
饱满度	$\phi_p=\sqrt[3]{\frac{WH}{L^2}}$	反映形状饱满程度
形状因子	$F=\frac{LH}{W^2}$	反映整体形状比例
矩形度	$R=\frac{S}{LW}$	反映与矩形的相似性
紧实度	$J=\frac{V}{LWD}$	反映在包围盒内的空间填充率

表 2　预测模型超参数设置

Tab. 2　Settings of prediction models

模型	超参数设置
SVR	径向基函数、正则化参数 1.0
KNN	邻居数 5、均匀权重、闵可夫斯基距离
RF	决策树数量 100、叶节点最小样本数 1
GB	弱学习器数量 100、学习率 1.0、最大深度 1
BPNN	隐藏层包含 100 个神经元、ReLU、Adam 优化器，学习率 0.01

1.8　评价指标

对于图像实例分割，使用平均精度均值（Mean average precision，mAP）作为核心评价指标，同时对比模型的浮点运算量和参数量；对于三维重建任务，采用相机标定时间、重建时间、重建精度、完整度及

综合评分进行多维度评估,从效率与质量两方面验证模型性能。百合种球质量、体积预测模型评价则使用使用决定系数 R^2、均方根误差(Root-mean-square error, RMSE)、平均绝对百分比误差(Mean absolute percentage error, MAPE)。

2 结果与分析

2.1 样本基础统计量

表 3 为样本质量与体积基础统计量。由表 3 可知,样本质量均值为 14.82 g,体积均值为 15.97 mL,二者数值相近,反映出样本质量与体积具有一定关联。且质量变异系数为 0.63,体积变异系数为 0.60,二者极差分别为 55.89 g 和 59.00 mL,质量标准差为 9.32 g,体积标准差为 9.59 mL,说明数据波动范围较大。另外,从峰度和偏度可知,样本质量和体积均呈右偏分布,即低质量、小体积的样品数量相对更多,但整体仍覆盖了不同规格区间,所用样本具有较好的代表性,为构建稳健预测模型奠定了基础。

表 3 样本质量与体积基础统计量

Tab.3 Basic statistics of sample weight and volume

统计量	质量/g	体积/mL
均值	14.82	15.97
最大值	59.30	62.00
最小值	3.41	3.00
极差	55.89	59.00
标准差	9.32	9.59
变异系数	0.63	0.60
峰度	1.68	1.65
偏度	1.34	1.30

2.2 Solo 实例分割

基于 OpenMMLab 计算机视觉库的 MMDetection,在预训练 Solov2 基础上进行迁移学习,实现百合种球与瓶盖的双类别像素级分割。模型训练设置为 EpochBasedTrainLoop,优化器为 OptimWrapper 封装的 SGD,初始学习率为 0.01、动量为 0.9、权重衰减为 0.000 1,并使用 CocoMetric 评估器进行模型评价。经过 15 轮训练,模型收敛,分割平均精度均值达到 97.0%。为进一步说明 Solov2 模型对百合实例分割的有效性,选择 Mask R-CNN、Yolact、Scnet、Solov1 在相同输入尺寸下开展试验,结果如表 4 所示。由表 4 可看出,Solov2 在保持高分割精度的同时,其浮点运算量仅为 6.860×10^{10},参数量仅为 3.11×10^{7},精度与效率间的平衡优势最突出。

表 4 不同实例分割模型测试结果

Tab.4 Test results of different instance segmentation models

模型	图像尺寸	浮点运算量	参数量	mAP/%
Mask R-CNN[28]	(448,768)	1.300×10^{11}	4.39×10^{7}	98.3
Yolact[29]	(448,768)	6.850×10^{10}	3.47×10^{7}	95.1
Scnet[30]	(448,768)	1.588×10^{12}	9.44×10^{7}	97.7
Solov1[31]	(448,768)	1.040×10^{11}	3.25×10^{7}	96.8
Solov2[25]	(448,768)	6.860×10^{10}	3.11×10^{7}	97.0

2.3 视图选择及三维重建

利用信息熵、新点率、颜色差异、全局特征、视角冗余构建多维评价指标,在 72 个视角中进行最优视图迭代筛选。试验发现索引号为 22 的图像具有最大熵值,由其构建的初始点云具有 3 941 个有效点。如图 2 所示,第 2 步迭代选中的视图索引号为 10,与初始视图夹角为 130°,视角冗余极低,点云新增有效点 2 453 个,新点率达 0.602,补全 62.2% 的几何信息,综合得分达到 0.692。第 3 步迭代选中的视图索引号为 5,新点率降至 0.125 5,但 CIEDE2000 颜色差异达 10.285,全局特征间的余弦距离为 0.184,全局点云增至 10 272 个有效点,是初始点云的 2.6 倍,种球几何和颜色覆盖度超 90%。第 4 步迭代开始,视图价值骤降,对重建精度提升微乎其微。因此,为了平衡信息覆盖与计算成本,选用第 22、10、5 号 3 个视图构造最优视图子集。

图 3 为视图增加过程中新点变化情况。各视图

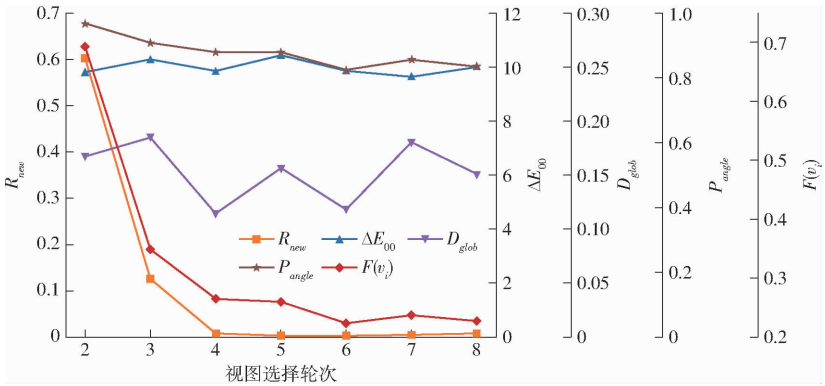

图 2 最优视图评价指标变化趋势

Fig.2 Trend of best view evaluation metrics

生成点云分别用红色、绿色、蓝色、黑色表示。使用 4 个视图重建结果，由图 3d 可看出，新增黑色点云所占比例很低。

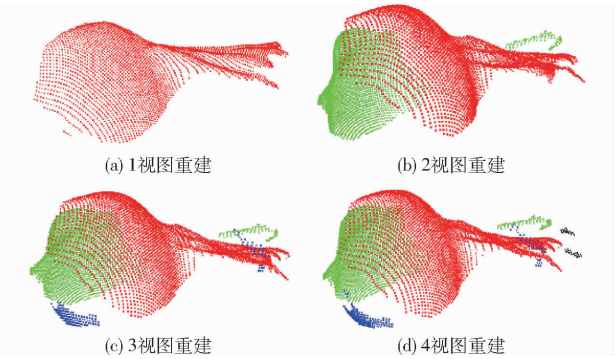

图 3　视图增加过程中的新点变化

Fig. 3　Changes of new points during view addition process

为验证 VGGT 的有效性，选择 Colmap、MVSNet、MASt3R、DUSt3R 等主流三维重建方法进行对比试验，结果如表 5 所示。DUSt3R、VGGT 无需依赖真实相机参数，但 VGGT 显著优于 DUSt3R，在 DTU 数据集综合评分从 1.741 降至 0.382；另外，VGGT 与 MASt3R 的重建精度相当，但重建时间由 62.5 s 降至 3.2 s。MVSNet 虽在重建精度、完整性、重建时间等方面表现良好，但需要花费 29.35 s 进行相机标定。

表 5　不同三维重建模型测试结果

Tab. 5　Test results of different 3D reconstruction models

方法	相机参数	兰州百合种球数据集		DTU		
		相机标定时间/s	重建时间/s	精度	完整度	综合评分
Colmap[32]	√	29.35	223.34	0.400	0.664	0.532
MVSNet[12]	√	29.35	1.80	0.396	0.527	0.462
MASt3R[17]	√	29.35	62.50	0.403	0.344	0.374
DUSt3R[16]	×	0	48.40	2.677	0.805	1.741
VGGT[20]	×	0	3.20	0.389	0.374	0.382

注：√表示采用相机参数，×表示未采用相机参数。

2.4　种球点云分割

如图 4a 所示，选择 3 种不同视角的图像输入 VGGT 模型，生成深度估计和相机内外参数（图 4c），由此构建包含背景信息的三维点云，但受图像背景影响，导致点云分布稀疏，边界模糊，且与环境噪声混杂严重，如图 4d 所示。为解决这一问题，使用 Solov2 模型对图像进行实例分割，获得前景掩膜，如图 4b 所示，再根据掩码进行三维反投影，对原始点云进行分割，提取种球点云，其轮廓清晰可见，说明掩码三维反投影是一种有效的点云分割方法，如图 4e 所示。另外，掩码三维反投影分割的点云还可能存在少量噪声，统计每个点周围 20 个邻近点到该点的平均距离和标准差，设定阈值删除离群噪声，如图 4f 所示。

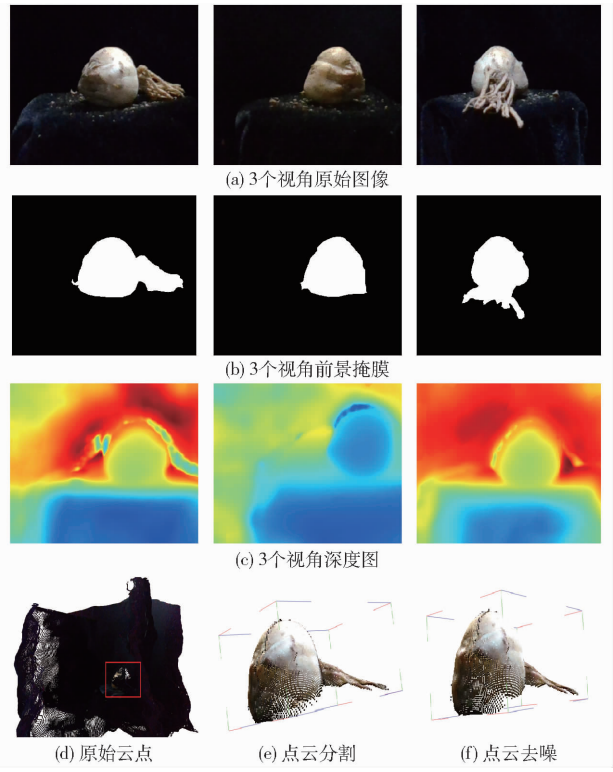

图 4　三维重建及点云分割

Fig. 4　Results of 3D reconstruction and point cloud segmentation

2.5　3D 形状特征提取

为了构建百合种球质量、体积预测模型，提取 14 个 3D 形状特征。采用主成分分析建立与物体几何特征对齐的规范坐标系，并构建轴对齐包围盒（AABB），为形状参数提取提供统一的参考框架，如图 5a 所示。图 5b、5c 分别为点云 Delaunay 三角化和切片体积，试验结果表明 Delaunay 三角化表面积及切片体积存在显著系统误差，不能直接用于种球表型。其原因在于：①多视角拍摄无法获得百合种球与载物台接触的底面。导致重建模型底部缺失，不能形成闭合的三维结构。②表面重建过程中 Alpha 参数的选取对三角形网格生成质量具有决定性影响，不同参数设置可能引入额外的计算偏差。③计算体积所用的垂直切片法对形态复杂区域刻画粗糙，如图 5d 所示，第一主轴与第二主轴比、第一主轴与第三主轴比均偏离 1.0，表明种球具有明显的各向异性特征；球形度为 0.804，反映出百合种球与标准球体的形状差异；紧实度为 0.357 0，说明实际体积仅占包围盒体积的 35.7%，以上指标均说明百合种球形态上的不规则性，特别是根须等纤细结构密集分布的区域，切片轮廓的提取误差显著。因此，提取 14 个形状参数，从不同维度捕捉种球几何特

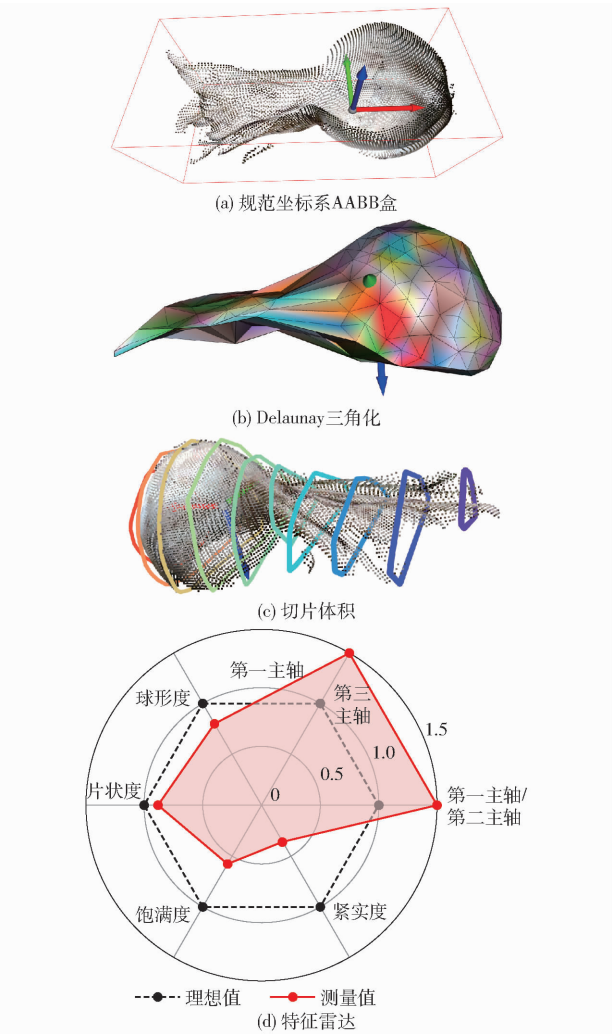

图 5 百合种球 3D 形状特征

Fig. 5 3D shape features of lily bulbs

性,有效弥补单一形状参数的不足。

2.6 多元线性回归模型预测结果与分析

为构建稳健的百合种球质量与体积预测模型,先进行方差膨胀因子 VIF 诊断,发现 14 个几何特征存在严重的多重共线性,其中等效球直径、片状度、球形度的 VIF 值均超过 1 000,形状因子的 VIF 值达到 205.56。为消除多重共线性对预测模型的不利影响,采用基于 AIC 准则的逐步回归方法进行特征选择,最终预测质量和体积的最优特征子集均包含等效球直径与长高比,分别反映种球的密实度和各向异性,模型表现出较好的预测能力。质量预测模型 R^2 为 0.705,RMSE 为 5.88 g;体积预测模型 R^2 为 0.730,RMSE 为 5.67 mL。但是,模型相对预测误差 MAPE 接近 30%,其分布呈现明显的异方差性,对于尺寸居中的样本,MAPE 较小,但对于极小(质量小于 5 g)或极大(质量大于 50 g)样本,其 MAPE 急剧增大,部分样本甚至超过 100%,说明线性模型捕捉三维空间中复杂的非线性关系存在局限。

2.7 非线性回归模型预测结果与分析

为识别影响种球质量、体积估测的关键特征并构建非线性模型,先根据 RF 产生的特征重要性得分,评估各输入特征的贡献度,如图 6 所示,发现影响种球质量与体积估测的核心特征存在显著共性,其中等效球直径、切片体积、长高比、面积体积比和片状度等几何特征对 2 个目标变量的贡献度均位居前列,表明这些特征在表征百合种球三维形态与物理属性方面具有关键作用。

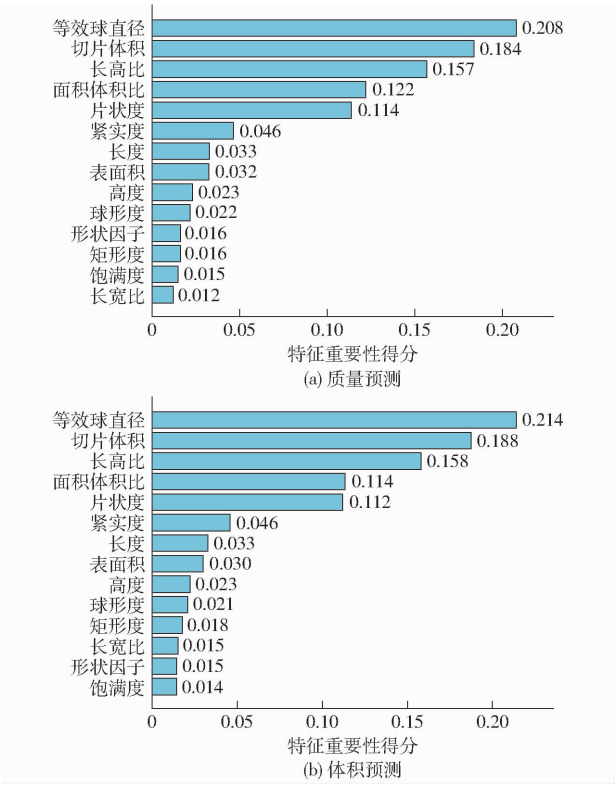

图 6 特征重要性得分

Fig. 6 Feature importance scores

在维持模型性能不显著降低的前提下,采用特征逐步消除策略,获得最优特征组合。如图 7 所示,对于质量预测,当特征数量从初始集合逐步减少至 8 个时,RF 模型的 R^2 仍维持在较高水平(大于 0.88),RMSE 与 MAPE 也保持在相对较低区间,若继续移除特征将导致各项评价指标的大幅下降,因此质量预测的最优变量组合中含 8 个特征。同理,10 个特征构成体积预测的最优变量组合。

使用最优特征组合,分别构建 5 种非线性模型,结果如表 6 所示。由表 6 可以看出,BPNN 在质量与体积预测中均最优,R^2 分别达到 0.957 6 和 0.957 3,RMSE 分别为 1.759 8 g 和 1.827 2 mL,MAPE 均低于 10%,显示出强大的非线性拟合能力。随机森林模型表现次之,R^2 分别为 0.883 7 和 0.901 6,在保持较好精度的同时提供了更好的模型可解释性。相比之下,SVR 和KNN模型的预测精度相对较低,而

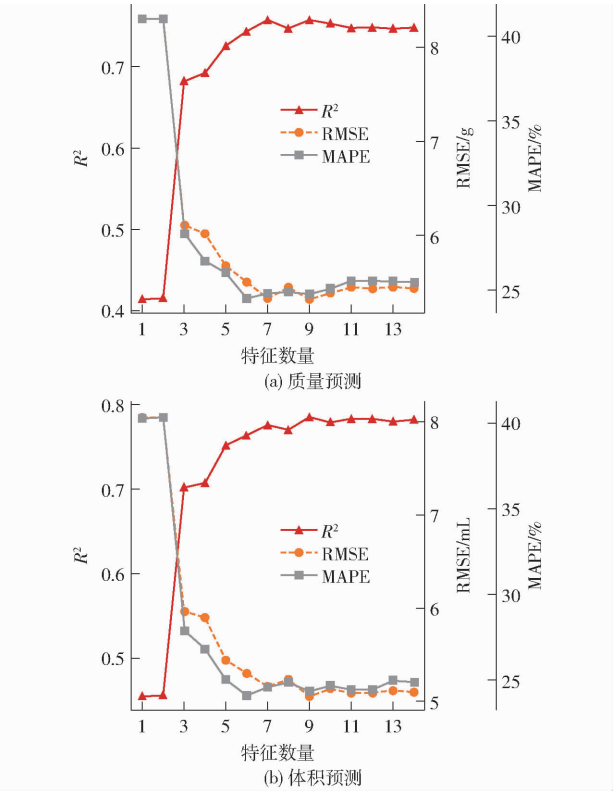

图 7　RF 模型性能演变

Fig. 7　Evolution of RF model performance

表 6　5 种非线性回归模型对种球质量、体积估测性能

Tab. 6　Estimation performance of five nonlinear regression models for bulb weight and volume

模型	质量			体积		
	R^2	RMSE/g	MAPE/%	R^2	RMSE/mL	MAPE/%
SVR	0.650 3	5.054 0	19.786 6	0.641 4	5.292 4	20.678 7
KNN	0.806 9	3.755 4	20.439 7	0.795 9	3.993 0	21.036 5
RF	0.883 7	2.914 2	16.434 0	0.901 6	2.772 9	14.602 7
GB	0.839 2	3.426 7	20.854 3	0.852 1	3.398 8	18.803 8
BPNN	0.957 6	1.759 8	9.397 0	0.957 3	1.827 2	9.212 0

梯度提升回归在 2 项任务的表现介于中间水平。

3　讨论

针对兰州百合种球形态不规则、根须分布复杂，传统表型测量效率低的问题，提出了一套融合最优视图选择的三维测量方法，无需依赖相机标定，就能从图像序列中端到端地预测出相机参数与三维结构。实验结果表明，经过“初始熵值筛选—多特征评估—视角惩罚修正”最优视图选择框架，选择的仅包含 3 个视图的图像序列，就能实现高精度三维重建，点云有效点数能达到初始视图的 2.6 倍，几何与颜色信息覆盖度超过 90%，效率显著优于文献[21－22]所采用的重建方法。

为进一步说明最优视图选择方法可有效平衡重建精度和计算耗时。在 GeForce RTX 3090 GPU 平台，对 VGGT 架构进行时间线分析，结果如表 7 所示。可以看出，特征聚合是 VGGT 三维重建的核心模块，占总耗时的 87% 左右；“全局-逐帧”注意力机制耗时约为 392 ms，是 DINO 图像编码的 16.7%。而相机参数预测、深度图生成、点云生成耗时仅占总耗时的 5% 左右。

表 7　VGGT 各核心模块耗时

Tab. 7　Time consumption of core modules of VGGT

模块	含义	耗时/ms
Vggt_3d_reconstruction	VGGT 三维重建	3 250.002
Vggt_camera_head	相机参数预测	26.066
Vggt_dpt_depth_head	深度图生成	102.171
Vggt_dpt_point_head	点云生成	20.568
Vggt_aggregator_full	特征聚合	2 828.656
Aggregator_dino_patch_embed	DINO 图像编码	2 353.429
Aggregator_transformer_layers	全局-逐帧注意力	392.903

而且，使用不同视图数目进行三维重建，耗时情况如图 8 所示。VGGT 三维重建与特征聚合高度重合。“全局-逐帧”注意力层耗时呈线性上升，从 3 个视角的 392.903 ms 增至 25 个视角的 6 500.999 ms，增加15.5 倍，说明全局注意力模块的计算量随视图数平方增长。而 DINO 图像编码层耗时始终稳定在 2 500 ms 附近，说明 GPU 并行优化使得批量的分块、线性投影、位置编码操作与单幅图像的相应操作耗时相当。

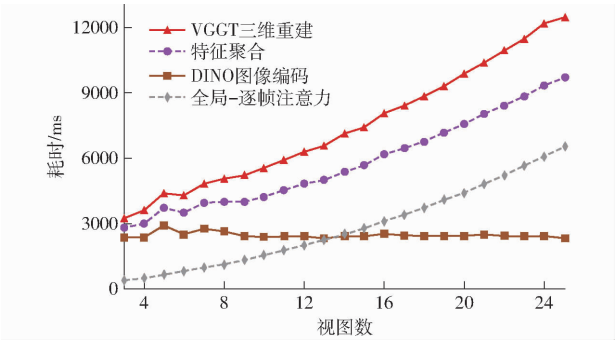

图 8　不同视图数目下各模块耗时

Fig. 8　Time consumption of each module under different numbers of views

另外，对于逐步回归模型预测种球质量、体积 R^2 达到 0.72 以上，但 MAPE 高达 30%，说明线性模型捕捉三维空间中复杂的非线性关系存在局限。而 BPNN 能够学习到等效球直径、切片体积、长高比、片状度等形态特征与预测目标间的复杂非线性关系，其 R^2 大于 0.95，MAPE 小于 10%，等效球直径和切片体积作为“总量”特征，其重要性最高，符合直觉，而长高比与片状度的高重要性，表明百合种球各向异性程度是决定其密实度和质量的关键因素。

本研究为百合选种提供了切实可行的技术方案，系统由 RGB 相机构建，成本低廉，易于在广大农户和合作社中推广，而且基于最优视图的 3D 重建时空效率高，使基于三维测量的自动化分选从实验室走向生产线成为可能。尽管研究取得了积极成果，但仍存在一些局限性。首先，受拍摄平台限制，重建模型存在底部缺失问题，不可避免地引入系统误差，导致衍生形状参数的偏差。其次，表面重建中的 Alpha Shape 算法对参数设置较为敏感，体积计算所采用的垂直切片法对根须等复杂结构的刻画仍显粗糙。

4 结论

(1)提出了一种融合最优视图选择与 VGGT 架构的三维表型参数估测方法。在最优视图筛选方面，基于“初始熵值筛选—多特征评估—视角惩罚修正”的最优视图选择框架，从 72 幅密集采样图像中筛选出 3 个最优视图，可实现种球 90% 以上的几何与颜色信息覆盖，点云有效点数达初始视图的 2.6 倍，在保证重建精度的同时，显著降低了计算成本。

(2)在三维重建与点云处理环节，VGGT 架构可端到端生成相机参数与三维点云，再通过 Solov2 分割掩码进行三维投影裁剪，能有效提取种球点云；在 3090 GPU 平台下，种球点云提取仅耗时约 3 250 ms，4 s 内可完成三维表型测量，基本满足自动化分选的及时性需求。

(3)在预测模型性能方面，对比了线性与非线性建模方法。逐步回归线性模型 R^2 大于 0.72，但 MAPE 接近 30%，难以捕捉复杂非线性关系；而经随机森林特征重要性筛选后构建的 BPNN 模型表现最优，质量与体积预测的 R^2 均超过 0.95，RMSE 分别为 1.759 8 g、1.827 2 mL，MAPE 均低于 10%，且等效球直径、长高比、片状度等特征对预测结果贡献度最高。

(4)在技术实用性方面，三维重建仅使用 RGB 图像，无需进行相机标定，易于在农户与合作社推广；同时，最优视图选择平衡了重建精度与计算耗时的矛盾，使三维测量驱动的种球自动化分选从实验室走向生产线成为可能。

参 考 文 献

[1] 厚毅清，王立光，张敏敏，等. 兰州百合脱毒组培苗炼苗技术规程[J]. 寒旱农业科学，2025，4(1)：84－86.
HOU Yiqing, WANG Liguang, ZHANG Minmin, et al. Technical regulation for the seedling hardening in virus-free tissue culture seedlings of Lanzhou lily[J]. Journal of Cold-Arid Agricultural Sciences, 2025, 4(1): 84－86. (in Chinese)

[2] TANABATA T, SHIBAYA T, HORI K, et al. SmartGrain: high-throughput phenotyping software for measuring seed shape through image analysis[J]. Plant Physiology, 2012, 160(4): 1871－1880.

[3] WILLIAMS K, MUNKVOLD J, SORRELLS M. Comparison of digital image analysis using elliptic Fourier descriptors and major dimensions to phenotype seed shape in hexaploid wheat (*Triticum aestivum* L.)[J]. Euphytica, 2013, 190(1): 99－116.

[4] ZHAO G, CAI W, WANG Z, et al. Phenotypic parameters estimation of plants using deep learning-based 3-D reconstruction from single RGB image[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1－5.

[5] PAULUS S. Measuring crops in 3D: using geometry for plant phenotyping[J]. Plant Methods, 2019, 15(1): 103.

[6] KOCHI N, ISOBE S, HAYASHI A, et al. Introduction of all-around 3D modeling methods for investigation of plants[J]. International Journal of Automation Technology, 2021, 15(3): 301－312.

[7] DUPUIS J, KUHLMANN H. High-precision surface inspection: uncertainty evaluation within an accuracy range of 15μm with triangulation-based laser line scanners[J]. Journal of Applied Geodesy, 2014, 8(2): 109－118.

[8] DISNEY M. Terrestrial LiDAR: a three-dimensional revolution in how we look at trees[J]. New Phytologist, 2019, 222(4): 1736－1741.

[9] ZENNARO S, MUNARO M, MILANI S, et al. Performance evaluation of the 1st and 2nd generation Kinect for multimedia applications[C]//2015 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2015: 1－6.

[10] NGUYEN T T, SLAUGHTER D C, MAX N, et al. Structured light-based 3D reconstruction system for plants[J]. Sensors, 2015, 15(8): 18587－18612.

[11] 程向楠，马沁巍，马少鹏. 自标定多视角三维数字图像相关测量方法及其应用[J]. 北京理工大学学报，2024，44(12)：1297－1309.
CHENG Xiangnan, MA Qinwei, MA Shaopeng. Self-calibration based Multi-view 3D digital image correlation measurement method and its application[J]. Transactions of Beijing Institute of Technology, 2024, 44(12): 1297－1309. (in Chinese)

[12] YAO Y, LUO Z, LI S, et al. Mvsnet: depth inference for unstructured multi-view stereo[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 767－783.

[13] GU X, FAN Z, ZHU S, et al. Cascade cost volume for high-resolution multi-view stereo and stereo matching[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2495－2504.

[14] ZHANG Z, PENG R, HU Y, et al. Geomvsnet: learning multi-view stereo with geometry perception[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 21508－21518.

[15] WANG F, GALLIANI S, VOGEL C, et al. Patchmatchnet: learned multi-view patchmatch stereo[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 14194 – 14203.

[16] WANG S, LEROY V, CABON Y, et al. Dust3r: geometric 3D vision made easy[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 20697 – 20709.

[17] LEROY V, CABON Y, REVAUD J. Grounding image matching in 3D with mast3r[C]// European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 71 – 91.

[18] FU Q, XU Q, ONG Y S, et al. Geo-neus: geometry-consistent neural implicit surfaces learning for multi-view reconstruction[J]. Advances in Neural Information Processing Systems, 2022, 35: 3403 – 3416.

[19] WEI Y, LIU S, RAO Y, et al. Nerfingmvs: guided optimization of neural radiance fields for indoor multi-view stereo[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 5610 – 5619.

[20] WANG J, CHEN M, KARAEV N, et al. Vggt: visual geometry grounded transformer[C]// Proceedings of the Computer Vision and Pattern Recognition Conference, 2025: 5294 – 5306.

[21] 陈允琳,兰玉彬,韩鑫,等. 基于 Kinect V3 传感器的叶菜类作物三维重建与表型参数获取[J]. 农业机械学报,2025,56(3):101 – 110,197.
CHEN Yunlin, LAN Yubin, HAN Xin, et al. Three-dimensional reconstruction and phenotypic parameters acquisition of leafy vegetables based on Kinect V3 sensor[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(3): 101 – 110,197. (in Chinese)

[22] 李添天,戴倩,谢元澄. 基于改进神经辐射场的植物表型参数测量方法[J]. 农业工程学报,2024,40(17):188 – 196.
LI Tiantian, DAI Qian, XIE Yuancheng. Measuring plant phenotypic parameters using improved neural radiance fields[J]. Transactions of the CSAE, 2024, 40(17): 188 – 196. (in Chinese)

[23] 付函,余绍政,刘烽,等. 基于 RGB – D 的香蕉果梳及其果指质量估测[J]. 农业工程学报,2025,41(12):333 – 343.
FU Han, YU Shaozheng, LIU Feng, et al. Estimating the weight of banana hands and fingers using RGB – D[J]. Transactions of the CSAE, 2025, 41(12): 333 – 343. (in Chinese)

[24] 徐胜勇,易同舟,秦子轶,等. 基于单视角 RGBD 图像的柑橘果实三维重建与表型检测方法[J]. 农业机械学报,2025,56(3):80 – 90.
XU Shengyong, YI Tongzhou, QIN Ziyi, et al. Three-dimensional reconstruction and phenotype detection of citrus fruits based on single-view RGBD images[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(3): 80 – 90. (in Chinese)

[25] WANG X, ZHANG R, KONG T, et al. Solov2: dynamic and fast instance segmentation[J]. Advances in Neural Information Processing Systems, 2020, 33: 17721 – 17732.

[26] ZHANG H, LI F, LIU S, et al. Dino: Detr with improved denoising anchor boxes for end-to-end object detection[J]. arXiv preprint arXiv: 2203.03605,2022.

[27] 尹晓龙,沈恩亚,王建民. 典型三维模型表示方法转换技术综述[J]. 计算机辅助设计与图形学学报,2025,37(10): 1666 – 1680.
YIN Xiaolong, SHEN Enya, WANG Jianmin. A review of conversion techniques for typical 3D model representation methods[J]. Journal of Computer-Aided Design & Computer Graphics, 2025, 37(10): 1666 – 1680. (in Chinese)

[28] HE K, GKIOXARI G, DOLLAR P, et al. Mask R – CNN[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961 – 2969.

[29] BOLYA D, ZHOU C, XIAO F, et al. Yolact: real-time instance segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9157 – 9166.

[30] VU T, KANG H, YOO C D. Scnet: training inference sample consistency for instance segmentation[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2021: 2701 – 2709.

[31] WANG X, KONG T, SHEN C, et al. Solo: segmenting objects by locations[C]// European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 649 – 665.

[32] SCHONBERGER J L, FRAHM J M. Structure-from-motion revisited[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4104 – 4113.