



南京农业大学学报
Journal of Nanjing Agricultural University
ISSN 1000-2030, CN 32-1148/S

《南京农业大学学报》网络首发论文

题目： 种子动态萌发表型分析算法的研究和软件实现
作者： 赵建华，周洁，丁国辉，孙港，徐凌翔，关雪莹，周济
收稿日期： 2021-11-25
网络首发日期： 2022-03-01
引用格式： 赵建华，周洁，丁国辉，孙港，徐凌翔，关雪莹，周济. 种子动态萌发表型分析算法的研究和软件实现[J/OL]. 南京农业大学学报.
<https://kns.cnki.net/kcms/detail/32.1148.S.20220228.2138.002.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

种子动态萌发表型分析算法的研究和软件实现

赵建华^{1,2}, 周洁¹, 丁国辉¹, 孙港¹, 徐凌翔¹, 关雪莹³, 周济^{1,4*}

(1. 南京农业大学前沿交叉研究院/作物表型组学交叉研究中心, 江苏 南京 210095; 2. 南京农业大学人工智能学院, 江苏 南京 210031; 3. 浙江大学农业与生物技术学院, 浙江 杭州 310029; 4. 英国国立农业植物研究所/英国剑桥作物研究中心, 剑桥 CB3 0LE, 英国)

摘要: [目的]种子是植物研究中重要的对象之一, 对种子萌发表型的动态监测为了解不同植物生存、生长和繁衍提供了重要依据。本研究提出了一种结合自动化图像处理、图论和监督式机器学习的方法, 实现了对种子关键萌发表型的量化分析。[方法]以禾本科植物小麦 (*Triticum aestivum*) 为研究对象, 利用监督式机器学习算法 (如 K 近邻、支持向量机、随机森林), 在不同颜色空间上对三种弱筋小麦品种的种子萌发图像序列进行前、背景对象训练及背景分割, 然后通过构建自动化图像处理算法进行目标提取, 再结合图论和二维骨架动态分析幼根和根尖点的位置变化, 实现关键萌发性状的高通量数字化提取。[结果]本研究可获得大量人工难以计量的萌发性状, 包括种子长、宽、面积、周长, 幼根和幼芽长度及生长速率等。通过与人工统计数据的线性回归分析, 关键动态性状如幼根长、根生长速率、芽长的决定系数 R^2 值分别为 0.922 ($n=188$, $P<0.001$, RMSE=1.727)、0.719 ($n=191$, $P<0.001$, RMSE=0.406)、0.897 ($n=115$, $P<0.001$, RMSE=2.726)。[结论]结果表明, 本研究提出的算法能有效获取种子萌发动态表型组, 并可扩展至棉花 (*Gossypium barbadense*) 和油菜 (*Brassica napus*), 为遗传育种和植物研究提供表型分型依据和智能化解析技术。

关键词: 种子萌发; 动态表型分析; 自动化图像处理; 监督式机器学习; 小麦

中图分类号: TP317.4

文献标志码: A

Developing vision-based analytic algorithms and software to dynamically measure key traits in seed germination

ZHAO Jianhua^{1, 2}, ZHOU Jie¹, DING Guohui¹, SUN Gang¹, XU Lingxiang¹, GUAN Xueying³, ZHOU Ji^{1, 4*}

(1. Academy for Advanced Interdisciplinary Studies/Plant Phenomics Research Center, Nanjing Agricultural University, Nanjing 210095, China; 2. College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210031, China; 3. College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310029, China; 4. National Institute of Agricultural Botany/Cambridge Crop Science, Cambridge CB3 0LE, United Kingdom)

Abstract: [Objectives] Seed is one of the most important research topics in plant research. The ability of dynamically detecting key seed germination traits provides important phenotypic evidence for researchers to understand plant survival, growth, development, and reproduction. Here, we proposed a set of algorithms for quantifying germination-related traits by combining automated image analysis, graph theory and supervised machine learning techniques. [Methods] Utilizing Poaceae such as wheat (*Triticum aestivum*) as a model plant, we applied automated image analysis together with machine learning algorithms (e.g. K-Nearest Neighbors, Support Vector Machine, Random forests) to train foreground and background objects, followed by background segmentation and object extraction based on image series collected from three weak gluten wheat varieties. Then, graph theory and two-dimensional skeletonization were employed to dynamically analyze changes of radicles and radicle tip positions to measure key germination-related traits in a high-throughput manner. [Results] We have collected a range of phenotypic traits in this study that were difficult to obtain through traditional approaches, including seed length, width, area, perimeter, radicle and seedling length, and their growth rates. We applied a linear regression analysis to validate the computational results with manual scoring, the square of the correlation coefficient, R^2 , computed for traits such as radical length, radical growth rate and seedling length are 0.922 ($n=188$, $P<0.001$, , RMSE=1.727), 0.719 ($n=191$, $P<0.001$, RMSE=0.406), 0.897 ($n=115$, $P<0.001$, RMSE=2.726), respectively. [Conclusions] The results suggest that the algorithm

收稿日期: 2021-11-25

基金项目: 江苏省基础研究计划 (BK20191311); 中央高校基本科研专项资金 (JCQY201902)

作者简介: 赵建华, 硕士研究生。通信作者: 周济, 教授, 博导, 研究方向为表型组学、系统开发、机器学习、图像分析、小麦育种, E-mail: Ji.Zhou@njau.edu.cn (中), Ji.Zhou@NIAB.com

and open-source software presented here can reliably obtain dynamic seed germination traits, which can also be extended to other crop species such as cotton (*Gossypium barbadense*) and oilseed rape (*Brassica napus*), providing phenotypic evidence and smart analytic solutions to enable studies in plant genetics and crop breeding.

Keywords: seed germination; dynamic phenotypic analysis; automated image processing; supervised machine learning; wheat

种子是高等植物生存、传播和繁衍的一种基本形式，既具有抵御不利环境的形态结构和生理特征^[1]，也是维护种质资源的重要形式^[2]。种子保存了植物关键遗传信息，亦被称为“农业芯片”，对衡量作物品质和农业生产都具有重要意义^[3]。种子在合适的环境下萌发和出苗，根据其萌发特征可分为吸胀（imbibition）、萌动（radicle emergence）和发芽（seedling）3个阶段，包括种子因吸水体积膨胀的吸胀阶段，胚根、胚芽向外生长突破种皮的萌动阶段（protrusion，俗称“露白”），胚根长与种子等长或胚芽长为种子1/2长的发芽阶段（germination）^[4]。种子发芽是种子生活力（viability）和活力（vigor）的基本表征，量化分析其萌发过程中的关键性状，为了解不同作物的生存、生长、产量形成及遗传完整性提供了重要依据^[5]。特别是在全球环境变化加剧的今天，胚根突破种皮的时间和胚芽生长速率等性状^[6]以及相关趋性^[7]往往体现了植物的抗逆性和适应环境的能力。因此，通过对幼根（radicle）、胚芽（embryonic shoot）和子叶（cotyledon）的表型分析，可为判断种子在不同环境下的活力提供依据，进而为研究作物早期萌发、出苗和品质等重要农艺性状提供表型组信息^[8]。

当前大规模种子萌发表型鉴定仍主要依靠人工观察，限制了萌发实验的监测频率、规模和准确性，因此植物研究领域对自动化表型获取和高通量鉴定的需求日益增长^[9]。近年来，国内外学者结合计算机视觉和机器学习对种子萌发表型分析进行了很多有益尝试。例如，Halcro等^[10]利用图像处理、k均值聚类算法和高斯混合方法采集种子的颜色、大小和形状特征数据；SeedExtractor^[11]和PhenoSeeder^[12]可以在单一背景下自动化获取单个种子的二维、三维形态数据，但分析速度受限于数据采集和处理算法；Vlaminck等^[13]利用ImageJ来测量水稻幼芽特征数据，但分析准确性和通量受限于软件平台；Germinator^[14]通过胚根和种皮之间颜色的对比度快速鉴定发芽，但只能应用于相对简单的萌发图像；RGSES^[15]通过图像处理获得种子颜色、形状、纹理等特征，并结合人工神经网络技术来预测水稻种子发芽趋势，但准确性相对较低；SeedGerm^[16]表型平台利用计算机视觉分析，根据幼根生长鉴定发芽和种子发芽率，进而通过表型性状鉴定油菜品种的遗传差异和对萌发调节蛋白磷酸酶（*HAI3*）在染色体上完成定位，但不能对交错幼根进行分割和动态分析；SeedQuant^[17]结合深度学习和计算机视觉统计发芽种子，但该算法不能对发芽后的种子继续动态监测；GrowScreen-Agar^[18]利用匹配滤波器对未粘连的拟南芥幼苗根冠进行定量分析，并获得根系结构数据，但在幼根互相遮挡情况下的分析准确率较低；Gong等^[19]基于语义分割算法分割水稻幼苗根系，但只能对单个种子进行分析，且对种子排列要求较高；RootNav2.0^[20]利用卷积神经网络从图像中获取复杂根系结构，精确分割根系，但算法计算复杂度高，很难应用于对种子幼根根系的大规模分析。

上述研究对种子萌发的动态表型分析进行了一定的探索，但目前尚无以分割交叉幼根为目标的种子动态萌发表型检测的相关研究，也未发现针对从萌发到胚根出现生长的整个过程来构建量化监测平台的报道。针对现有分析软件算法在这一研究方向上的不足之处，本研究借鉴了SeedGerm的开源软件包^[16]，创新地提出了基于幼根分割、根尖点追踪的种子萌发动态表型分析算法。同时，本研究还利用Python编程语言和图像处理、机器学习等开源软件库设计了数字化性状提取软件包。对幼根动态分割和根尖点追踪等表型分析难题开展了有益的研究，为自动化检测关键的种子萌发表型提供参考。

1 材料与方 法

1.1 种子萌发实验设置

种子萌发实验图像数据于2020年5月至8月完成采集。以小麦（*Triticum aestivum*）萌发实验为例，在发芽盒底部放置双层蓝色滤纸，并在滤纸上方放置A3尺寸的深蓝色或深红色发芽纸。研究采用3个弱筋小麦品种，分别为‘扬麦25’、‘宁麦26’和‘淮麦30’，于2019年收获，保存期为1年。如图1所示，每个品种选取18粒（6行3列）麦粒放置于蓝色发芽纸上，并将固定体积的水加入到滤纸中，静置2h，

待滤纸均匀吸收水分后开始种子萌发实验，发芽盒内温度保持为 22 ~ 24 °C，湿度为 99%。拍摄和人工实验数据采集周期为 7 d。

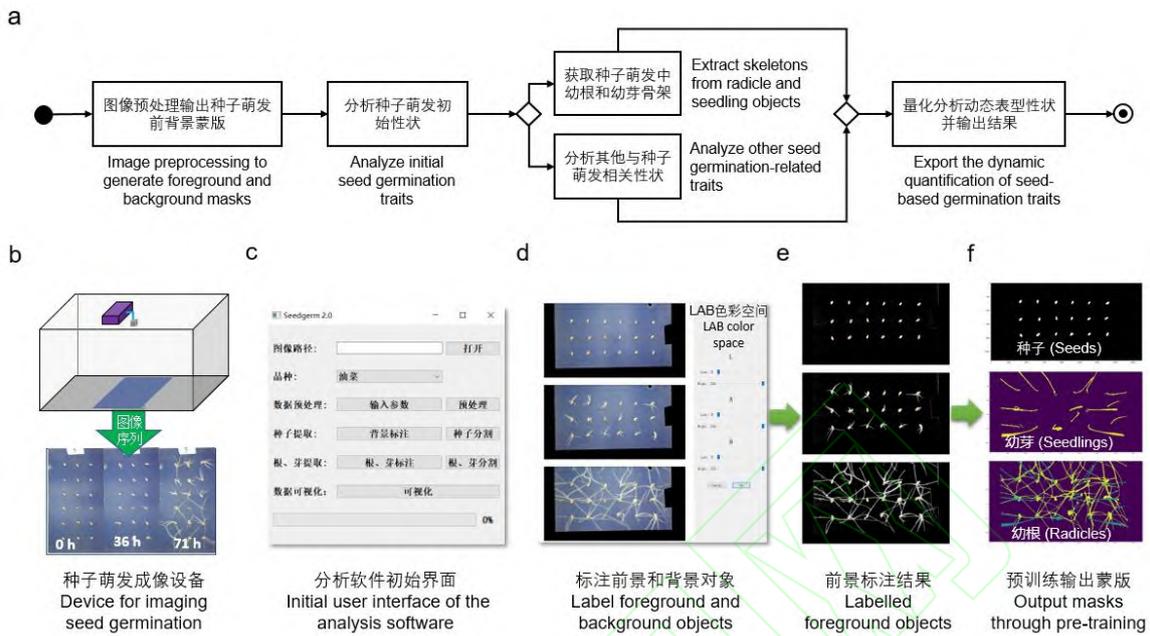


图 1 技术路线、数据获取、分析软件界面和前背景标注、预训练结果

Fig.1 Technical roadmap of research, data collection, graphic user interface (GUI) of the analysis software, data labelling of foreground and background objects, and pre-training results

1.2 算法技术路线

在获取图像序列后，本研究构建了自动化分析流程，其技术路线如图 1-a 所示。首先，对输入图像序列（图 1-b）完成预处理，即通过图形用户界面（graphic user interface, GUI; 图 1-c）中的颜色空间（color space）获得种子的前景和背景蒙版。然后，根据种子蒙版获得种子萌发初始性状，如种粒长度、宽度等；再通过种子二维骨架追踪种子尖点，获得种子萌发早、后期的相关性状，如幼根根长、幼根生长速率等。最后输出种子萌发性状的量化分析结果，并存储在纯文本 CSV 文件中。

1.3 萌发时间序列图像采集

为高质量提取种子萌发性状，完整记录整个萌发实验过程，实验对置于半透明塑料发芽盒中的多个种子进行连续拍摄，拍摄时间间隔为 1h，以时序图像方式记录种子萌发过程。本研究使用了 *Pi Camera* 高清摄像头（Sony IMX477 sensor, Japan; 5M 像素），拍摄图像最大分辨率为 2 592×1 944，摄像头由运行于树莓派微型单板计算机（*Pi Foundation*, UK）的 Python 软件控制（图 1-b）。基于时序的小麦种子图像分辨率为 2 130×1 130 像素（垂直×水平）。实验采集 140~160 张图像，可记录 7 d 的小麦种子萌发过程。

1.4 图像处理和机器学习的软件实现

本研究的性状分析算法结合了自动化图像处理、图论和监督式机器学习，通过编程语言 Python (3.70) 完成算法和各子功能的软件实现。其中使用的开发程序库包括：图像处理库 *scikit-image*(0.17.2)和 *OpenCV-Python* (4.4.0.46)；机器学习库 *scikit-learn* (0.23.2)；科学数据处理库 *NumPy* (1.19.4), *Pandas* (1.2.3) 和 *SciPy* (1.5.4)；图形界面库 *PyQt5* (5.15.2)。以上开发均通过 *PyCharm* 集成开发环境（*Integrated Development Environment*, IDE）完成源代码编写和测试，以下提及的各类函数均基于上述软件开发库完成相应软件实现（software implementation）。

针对本研究在背景分割时使用的机器学习部分，算法首先自动选择图像序列中位于前、中、后 3 个时间点的图像作为训练样张，然后通过用户拖动 GUI 界面（图 1-d）的滑杆对不同颜色空间中的前背景对象进行初始分割（图 1-e）。在得到前景和背景蒙版（mask）后，将不同萌发阶段的前景和背景蒙版所包含的像素通过红绿蓝（Red-Green-Blue, RGB）和色相、饱和度、明度（Hue-Saturation-Value, HSV）两个颜色

空间（共六个通道）作为机器学习的训练样本数据集，再通过用户选择的监督式机器学习模型进行学习。最后将训练后的前、背景分割模型应用到整个图像序列，为后续的性状分析提供动态蒙版（图 1-f）。

1.5 人工检测和算法评估

在建立人工真实数据集（ground truth dataset）时，本研究使用了 ImageJ 软件对 140 张随机选取的萌发图像进行人工测量获得种粒长、宽、周长、面积、种子露白时间、幼根根长、幼根生长速率、种子发芽时间、幼芽芽长、幼芽面积等性状。将算法获取的关键性状数据如幼根根长、幼根生长速率和幼芽芽长与人工真实数据对比完成数理统计，并使用分析结果中的决定系数 R^2 和均方根误差（Root Mean Square Error, RMSE）作为对算法获取性状的评价指标。根据决定系数 R^2 来确定算法数据和人工数据的线性回归关系，若 R^2 越高表示两种数据间的关联性越高，RMSE 越低表示模型误差越小。相关工作通过微软 Excel 的数理统计函数集完成。通过 IDE 的性能分析（Python profiler），本算法在处理单张小麦图像（5 MB）的平均用时为 90~100s，处理一个小麦图像序列（140~150 张）用时为 210~220min，占用内存为 667 MB，程序运行时间和内存消耗与图像序列大小呈线性关系。

2 结合监督式机器学习和自动化图像处理的种子萌发表型分析

2.1 图形界面（GUI）

鉴于开源软件包 SeedGerm 界面交互能力不强的问题，本研究使用功能更齐全的 PyQt5 开发包^[21]搭建了种子萌发表型分析软件的 GUI。GUI 集成了本研究的分析算法，可对多物种种子萌发图像序列进行批量处理，并输出关键萌发性状的量化分析和数字化提取结果，包括种子的长、宽、面积、周长，种子露白时间、萌发时间，幼根长度、幼根生长速度，幼芽长度等。通过初始界面（图 1-c），用户可在“图像路径”选择图像序列，在“品种”下拉列表选择作物种类（如油菜、小麦或棉花等）。在“数据预处理”定义参数，如滤纸面板数量、发芽实验面板中种子的行列数、开始和结束图像序号（默认值为所选图像序列的第一和最后一张图像）。在“种子提取”的“背景标注”下拉列表中可选择颜色空间（例如，LAB 颜色空间，L 代表亮度，A 和 B 代表颜色分量；HSV 颜色空间，H 代表色调，S 代表饱和度，V 代表亮度），并通过调整滑杆位置（图 1-d）对背景进行高效人工标注（图 1-e），获得基于像素的前背景训练集。

在“种子分割”下拉列表中，可选择三种为分割幼根幼芽重新训练的监督式机器学习算法，如 K 近邻^[22]（k-Nearest Neighbor, KNN）、支持向量机^[23]（Support Vector Machine, SVM）和随机森林^[24]（Random Forest, RF）。不同机器学习算法可对不同质量图像的前景（种子、幼根和幼芽）、背景（发芽纸）的像素簇分别学习，进而得到不同时期的种子萌发像素特征及其蒙版（图 1-e）。若需对幼根、幼芽特征信息单独提取，用户还可在“根、芽提取”部分对获取的种子蒙版进行操作，提高建立训练集的效率。

2.2 种子萌发初始性状分析

以下算法解析以小麦种子萌发性状的量化分析和数字化提取为例展开介绍。对小麦种子萌发初始阶段的性状分析从第 0 h 开始，种子经历吸胀到露白阶段初期。如图 2 所示，算法通过一系列图像处理算法完成了对初始萌发性状的量化分析。首先，算法从 GUI 端口获取输入的 RGB（red-green-blue，红绿蓝）图像（图 2-a）和用机器学习算法处理后得到的前景（即种子）、背景（即发芽纸和发芽盒）蒙版（mask）。然后，算法利用 SciPy 中的 label 函数和 regionprops 函数对种子蒙版（图 2-b）进行筛选，测量并记录每粒种子的质心坐标（橙色）及其编号（黄色）。虽然种子在吸胀过程中会发生滚动，质心位置随之变化，但在一小时内（拍摄间隔），种子质心位置变化较小，可通过对比两张相邻图像中种子的质心坐标完成种子的定位（图 2-c）。

基于获取的种子质心，使用 thin 算法提取种子的二维骨架（图 2-d）。若二维骨架的长度变长（超过种子长度 1/3）且监测区域中出现白色像素，则判断种子进入露白阶段。对种子的初始性状分析还包括种子面积、种子轮廓长度（即种子周长）、种子宽度和长度（图 2-e）。这些初始性状测量基于像素，算法也可通过测量放在发芽盒中的标尺（如图 2-a 白色标签）完成从像素到公制（如 mm）的转换。

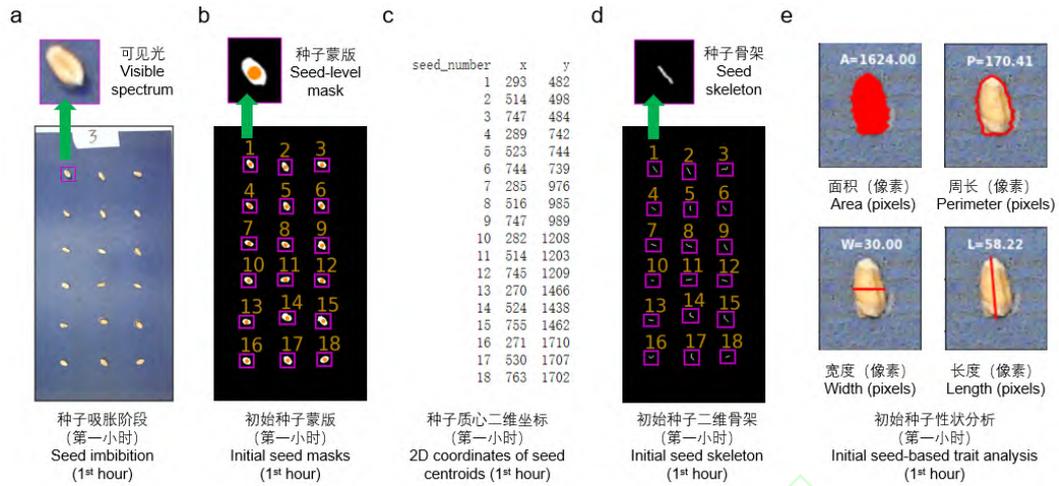


图 2 种子萌发初始性状（如穗粒和吸胀特征等）分析（第 1 小时）

Fig. 2 Phenotypic analysis of initial seed germination traits such as grain-level phenotypes and imbibition in the first hour

2.3 种子萌发早期性状分析和幼根根尖追踪算法

在种子萌发早期，种子幼根还未发生交叉。幼根根尖追踪算法对该时期的种子幼根进行的性状分析主要集中于露白时间、萌发时间、幼根长度变化等性状。算法流程如图 3-a 所示，首先对种子萌发图像（图 3-b）进行前背景分割，获得每粒种子的质心坐标和编号（图 3-c），然后提取每粒种子及其幼根的二维骨架（绿色；图 3-d），分析其形态特征。再通过 SciPy 中的 `binary_hit_or_miss` 算法获得骨架端点，将距离质心远的端点作为幼根根尖点，并在追踪算法中记录根尖点坐标（红色；图 3-e）。同时，将根骨架沿法线方向到根轮廓的二倍距离作为幼根直径。最后，对本图获得的种子根尖点（淡黄色；图 3-f）与上一张图像中对应种子的根尖点（红色；图 3-f）坐标进行比较，并根据不同时间点的根尖点坐标重建幼根骨架，并根据重建的骨架长度计算幼根生长速率（图 3-f）。

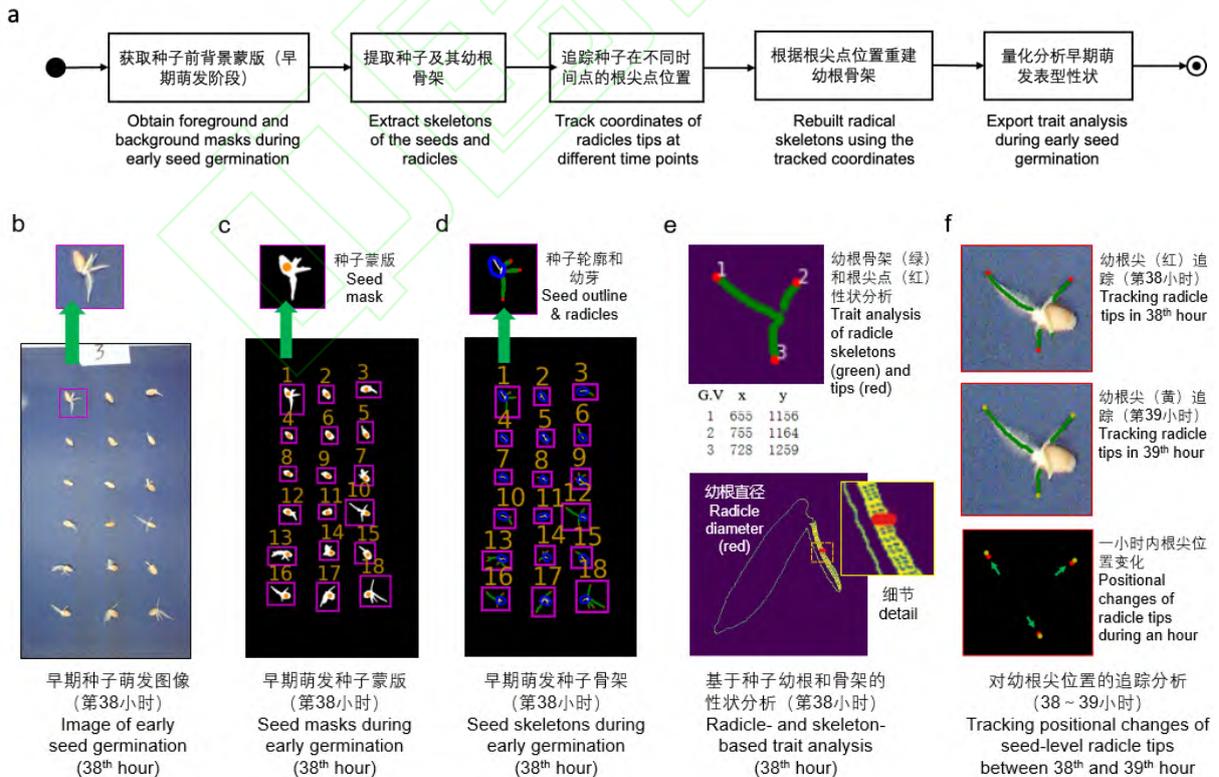


图 3 分析流程和早期萌发（即激活新陈代谢阶段）的种子性状分析（第 38 小时）

Fig. 3 Analysis workflow and phenotypic analysis of seed-based traits at the reactivation of metabolism phase (38th hour)

2.3.1 幼根根尖点追踪算法 本研究原创的根尖点追踪算法是为了重建幼根骨架，并获得幼根根长和幼根生长速率。算法会对未露白种子 i 用初始化有向图 $G_i(V,E)$ 为空集来表示，当种子露白后以第一次检测到的根尖点坐标 $C_{i,j}$ 来初始化 G_i 的顶点 $G_i.V$ ，并把 $G_i.E$ 的边集合设置为空集，即目前没有根骨架。其公式如式 (1) 所示：

$$G_i(V,E) = \{G_i(V,E) | G_i.V = C_{i,j}, G_i.E = \emptyset, \text{ if } C_{i,j} \neq \emptyset \text{ and } G_i(V,E) = \emptyset\} \quad (1)$$

式中： i ($1 \sim n$, n 为发芽实验种子总数) 为种子序号， j ($1 \sim m$, m 为序列图像的总数) 表示该张图像在图像序列中的序号； $G.V$ 表示有向图的顶点集合，存放的是幼根骨架上的当前根尖点坐标，其中“.”表示对有向图 G 中成员数据的调用； $G.E$ 表示有向图的边集合，即幼根二维骨架集合； $C_{i,j}$ 为第 j 张图像中检测到的第 i 粒种子的根尖点集合。

2.3.2 幼根重建算法 根据当前图像 j 检测到的第 i 粒种子及其对应的根尖点集合 $C_{i,j}$ ，与图像 $j-1$ (即前一张图像) 中相对应的第 i 粒种子及其对应的根尖点集合 $C_{i,j-1}$ ，遍历同一个种子 i 在前后两张图像中的根尖点集合里的根尖点元素 $T_{k,i,j} \in C_{i,j}$ 和 $T_{k,i,j-1} \in C_{i,j-1}$ ，其中 k ($1 \sim K$; K 表示此根尖点集合 $C_{i,j}$ 中的根尖点总数) 为根尖点 k 在种子 i 在图像 j 里的根尖点集合中的序号。计算 $T_{k,i,j}$ 和 $T_{k,i,j-1}$ 的距离是否小于 20 像素 (根据图像分辨率获取的经验阈值)。同时，计算相对应的根尖点 $T_{k,i,j}$ 和 $T_{k,i,j-1}$ 构成的向量和重建骨架的夹角是否大于 120° 。若以上两项条件均符合，算法将连接 2 个尖点 $T_{k,i,j}$ 和 $T_{k,i,j-1}$ ，以重建线段作为两根尖点间的幼根骨架，并添加到有向图的重建根骨架集合 $G_i.E$ 中。在第 i 个种子在图像 j 上检测到的根尖点及对应幼根骨架重建完毕后，用 $C_{i,j}$ 替代并更新该种子的有向图集合 $G_i.V$ 。具体公式如 (2) 所示：

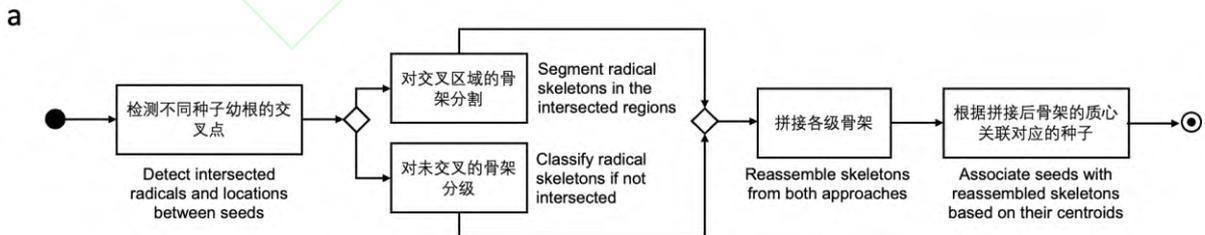
$$G_i(V,E) = \left\{ G_i(V,E) \left| \begin{array}{l} G_i.V = C_{i,j} \text{ and } G_i.E = G_i.E + L(T_{k,i,j}, T_{k,i,j-1}), \\ \text{if } E_D(T_{k,i,j}, T_{k,i,j-1}) < 20 \text{ and } \cos(\overrightarrow{T_{k,i,j} T_{k,i,j-1}}, \overrightarrow{T_{k,i,j-1} P}) < -1/2 \end{array} \right. \right\} \quad (2)$$

式中： $L(T_{k,i,j}, T_{k,i,j-1})$ 表示两个点之间的线段像素集合； $E_D(T_{k,i,j}, T_{k,i,j-1})$ 表示顶点 $T_{k,i,j}$ 和 $T_{k,i,j-1}$ 的欧式距离； P 为 $G_i.E$ 中与 $T_{k,i,j-1}$ 距离为 10 像素的坐标； $\overrightarrow{T_{k,i,j-1} P}$ 表示根尖点 $T_{k,i,j-1}$ 所在的幼根重建骨架的向量， $\cos(\overrightarrow{T_{k,i,j} T_{k,i,j-1}}, \overrightarrow{T_{k,i,j-1} P})$ ，表示 $T_{k,i,j}$ 和 $T_{k,i,j-1}$ 构成的向量和重建骨架夹角的余弦，余弦值小于 $-1/2$ 表示夹角大于 120° 。

算法对序列中的所有图像中的根尖点都重复上述操作，可对图像序列中所有可测种子的根尖点完成追踪和幼根二维骨架的重建。通过 IDE 的性能分析，此部分算法对每张图像的处理时间介于 25~35s，运行时间随图像数量增加呈线性增加。

2.4 对复杂种子幼根的分割和追踪

在种子萌发后期，种子幼根将出现较复杂的交叉遮掩现象。针对这一问题，算法首先分割交叉的幼根骨架，再根据每粒种子的根尖点坐标完成动态追踪，按时序重建每粒种子所对应的幼根骨架。算法流程如图 4-a 所示，其中对交叉遮掩幼根的分割算法主要分为二维骨架分解和重建两部分。



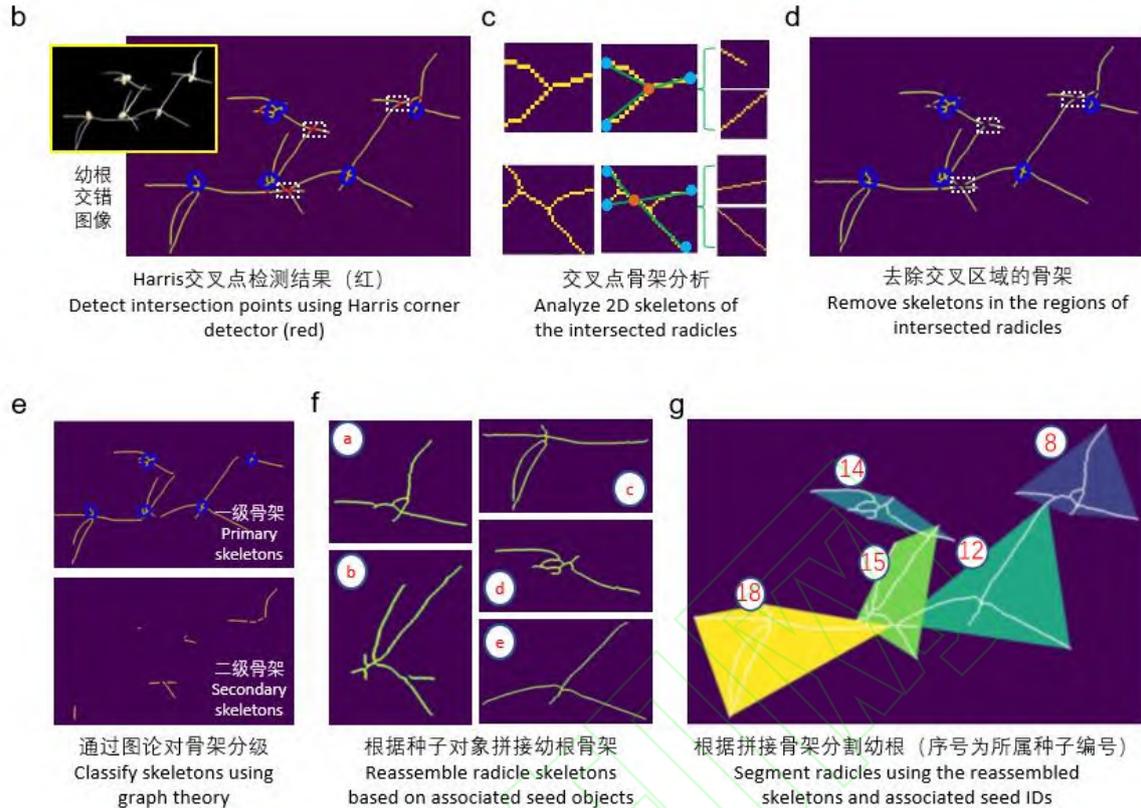


图 4 分析流程和基于图论分割、重建粘连种子骨架

Fig. 4 Analysis workflow and segment overlapped radicle skeletons using graph theory and skeleton reconstruction techniques

2.4.1 幼根骨架分解部分 对粘连种子骨架进行 Harris^[25]角点检测，筛选合适的角点（2种，分别为“十”字交叉点、“人”字交叉点）作为幼根交叉点（图 4-b 红色像素区域），选择交叉点上下左右各 8 像素作为交叉区域。分析交叉区域骨架，分为两种情况：如图 4-c 所示，第一种是包含一个交叉点和 3 个端点的“人”字交叉点；第二种是包含 2 个交叉点和 4 个端点的“十”字交叉点。对于第一种幼根交叉方式，分割方法为：连接交叉点（橙色）和 3 个端点（天蓝色），构建 3 个向量；分别计算向量间夹角 $[0^\circ, 180^\circ]$ 的余弦，余弦越小表示夹角越大，连接夹角最大的两个端点构建为一条幼根骨架，连接剩下的端点和交叉点构建另一条幼根骨架。对于第二种交叉方式，算法分割方法为：取 2 个交叉点连线的中点作为新交叉点（橙色），从新交叉点到四个端点（天蓝色）连线作为 4 个向量；计算向量之间的夹角余弦，连接夹角最大的 2 个端点作为一条幼根骨架，连接剩下的 2 个端点作为另一条幼根骨架，并分别保存分割后的 2 条幼根。

去除交叉区域后剩余幼根骨架不存在粘连（图 4-d），因此可根据其是否与种皮部分（图 4-d，深蓝色）连接分类为一级骨架和二级骨架（图 4-e）。将交叉点区域骨架处理后结果加入二级骨架集合。

2.4.2 基于种子的幼根骨架重建 遍历并计算二级骨架和一级骨架之间距离最近的两个端点，计算 2 个端点之间线段的距离和 2 个端点之间线段与一级骨架的夹角。若距离小于距离阈值（即 20 像素）且夹角大于角度阈值（即 120° ），则将该二级骨架拼接对应的一级骨架中。通过以上方法可将全部的一、二级骨架重新拼接（图 4-f），并根据不同种子的质心和所在骨架区域对应种子序号，完成对粘连幼根的分割和种子归属分析（图 4-g）。根据种子序号和分割粘连后的种子根骨架，再运用 2.3 节中的幼根根尖追踪和骨架重建算法，可追踪某粒种子对应的所有幼根根尖点，并基于时序完成对幼根骨架的重建。

2.5 种子萌发后期性状分析

在萌发后期，种子扎根扶叶，胚芽鞘向上竖起。此时对种子发芽性状的量化分析主要集中于幼根长度、根生长速率和幼芽长度等性状。算法流程如图 5-a 所示，首先根据萌发图像（图 5-b）进行前背景分割，对每粒种子的质心坐标和二维骨架完成初始分析（图 5-c）。针对种子粘连问题，利用 2.4 节算法分割幼根骨架，和从机器学习中获取的幼根、幼芽蒙版（图 5-d）提取幼根（绿色）和幼芽（蓝绿色）的二维骨架，基

于骨架尖点坐标获得幼根、幼芽的长度和生长速率（图 5-e）。利用 `convex_hull_image` 算法获得幼根的凸包（convex hull），以凸包中的像素总数作为凸包面积，即为根面积。利用种子幼芽蒙版的像素总数作为幼芽的面积。并通过图像中标尺完成从像素到毫米的转换。

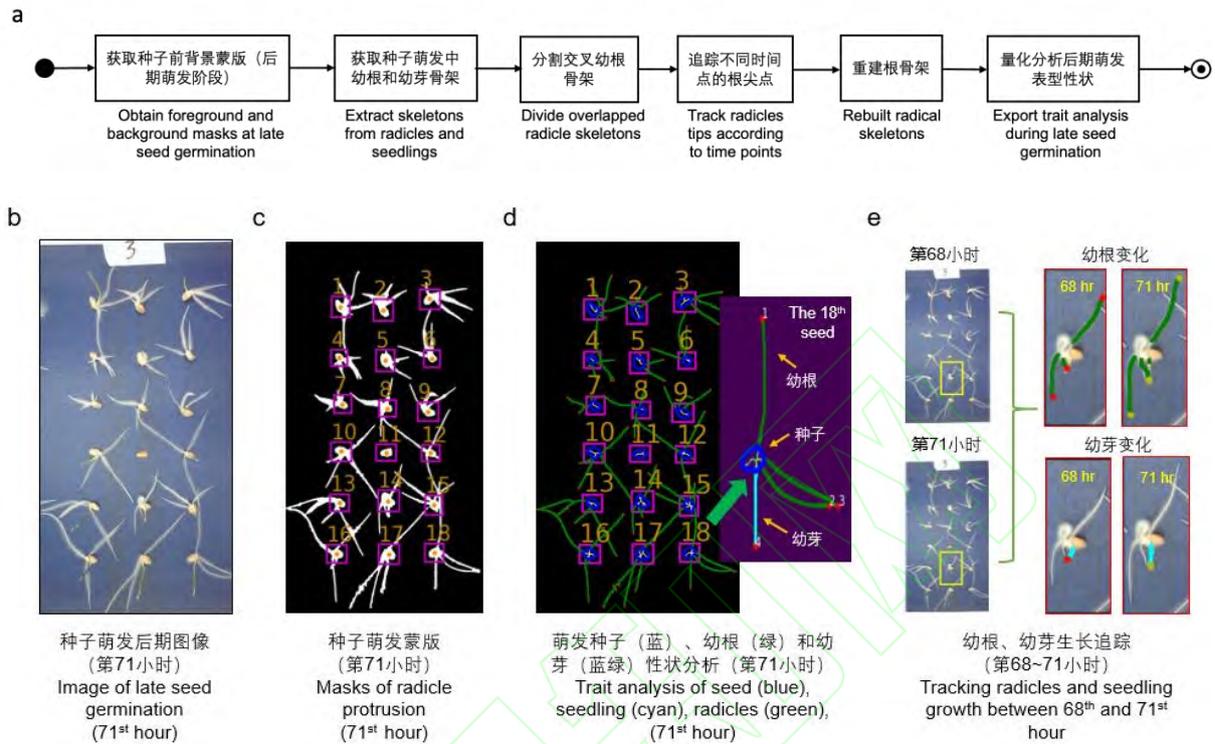


图 5 分析流程和后期萌发（即种子萌发至幼苗发育阶段）的种子性状分析

Fig. 5 Analysis workflow and phenotypic analysis of seed-based traits during the seed germination and seedling development phases

3 结果与分析

软件处理完图像序列后，会输出相关性状的 CSV 文件，初始性状包括：种粒的长度、宽度、周长、面积。萌发早期性状包括：露白时间、根长、根直径、根面积、根生长速率、发芽时间。萌发后期包括：芽长、芽面积、芽直径以及对幼根持续追踪后的根长、根直径、根面积、根生长速率。若图像中包含标尺，算法会针对标尺分析，将长度单位从像素转换为毫米。

3.1 种子幼根追踪算法在动态表型分析中的应用

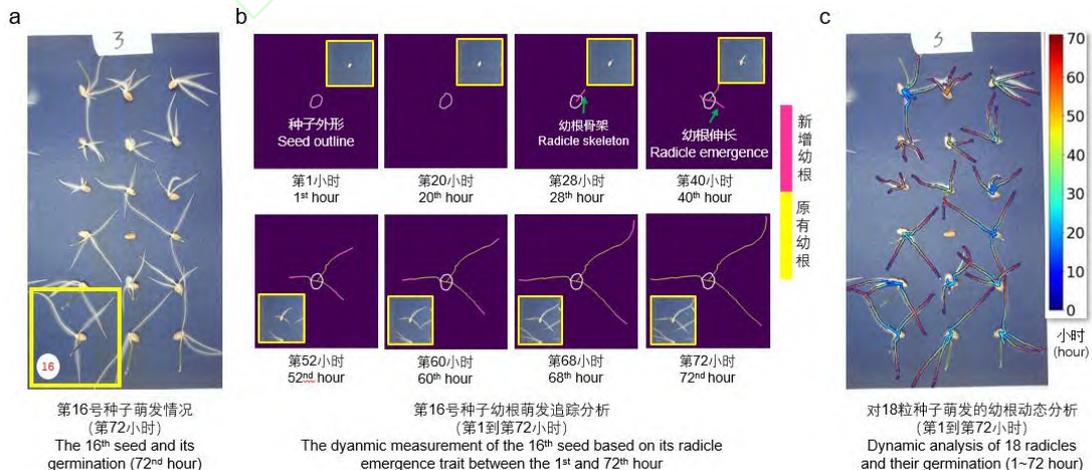


图 6 种子幼根萌发性状动态分析

Fig. 6 Dynamic analysis of the seed-based radical emergence trait

以图像序列 3 中的第 16 号种子为例 (图 6-a), 种子露白前持续分析种粒性状 (图 6-b 白色轮廓)。种子露白后, 按时序获得幼根长度和生长速率等性状 (图 6-b 粉色)。根据追踪结果可以发现, 随着种胚继续生长, 在第 28h 胚根长与种子等长, 标记种子发芽; 在 40 h, 种子出现 2 条次生胚根; 在 60 h, 出现根交叉。基于时序对所有种子的根骨架重建效果如图 6-c 所示, 表示从 1~72 h 幼根生长过程 (深蓝色为第 1h, 深红色为第 72h)。

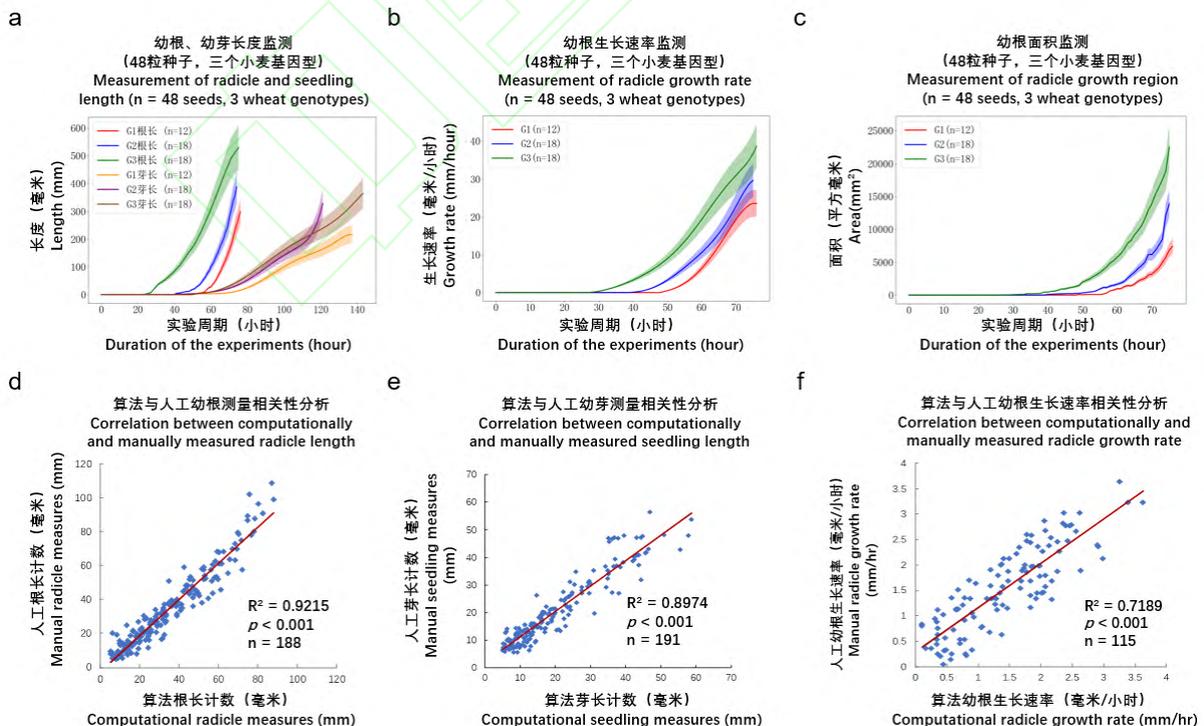
3.2 萌发性状量化分析与人工数据对比

以小麦为例, 在所选 3 种小麦基因型 (分别为: '扬麦 25' '宁麦 26' 和 '淮麦 30') 的图像序列中, 每个基因型均有 3 个重复。关键性状量化分析结果如图 7 (a-c) 所示。从动态表型分析可见, 三个小麦基因型分别在 56、40、23 h 开始露白, 此后根长、幼根生长速率和幼根面积稳定增长。其中, 第 3 基因型的幼根生长速率和幼根面积出现较快, 体现了较好的萌发性。在 120 h 左右, 第 2 基因型的幼芽生长速度加快, 并快速超越第 1、3 基因型, 体现了良好的出芽能力。

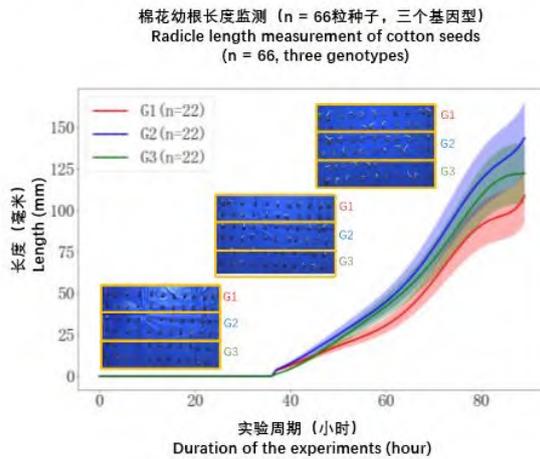
将不同重复中获得的根长、芽长、根生长速率的数据通过线性回归分析与人工测量数据进行对比, 获得的决定系数 R^2 值分别为 0.922 ($n=188$ 粒种子, $P < 0.001$, $RMSE=1.727$)、0.719 ($n=191$ 粒种子, $P < 0.001$, $RMSE=0.406$)、0.897 ($n=115$ 粒种子, $P < 0.001$, $RMSE=2.726$)。如图 7d-f 所示, 相关性分析结果表明, 算法计算的幼根和幼芽性状数据与实际性状数据具有较显著的相关性, 算法能较精确地提取小麦种子的萌发性状数据。

3.3 算法在不同作物种子萌发实验中的应用

虽然本算法基于单子叶植物须根系研发, 但为提高算法的泛化能力, 设计算法时也考虑到了双子叶植物直根系的幼根特征。以棉花 (*Gossypium barbadense*) 和油菜 (*Brassica napus*) 为例, 通过简单参数调整后, 本算法还可动态分析两种作物的幼根根长性状。对棉花幼根根长的动态分析显示 (图 7-g), 3 个棉花基因型的露白时间相差不大, 第 2 个基因型的幼根生长略快。对油菜幼根根长的分析显示 (图 7-h)。第 6 个基因型露白时间较晚, 幼根在萌发前期生长较慢, 在萌发后期生长加速。第 5 个基因型的幼根在萌发前期生长较快, 在后期生长速度减缓。



g



h

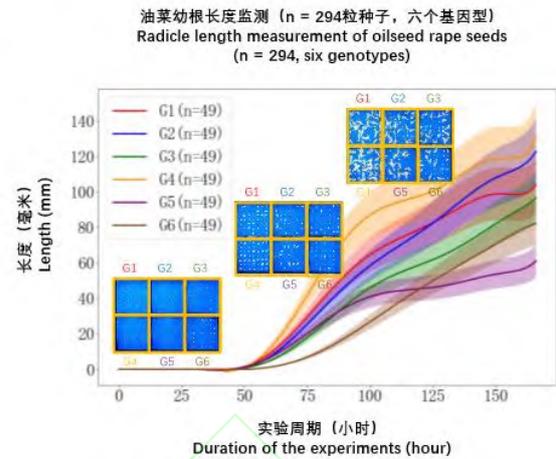


图7 性状量化分析结果、线性回归分析结果、算法应用

Fig. 7 The result of traits quantitative analysis and linear regression analysis and the application of dynamic detection algorithm

4 结论

1) 本研究提出的结合自动化图像处理、图论和机器学习的种子动态萌发表型检测算法, 即利用监督式机器学习算法进行图像前背景分割, 可用于不同品种不同背景的实验图像; 然后利用自动化图像处理算法可获取基于形状和颜色空间的性状, 进而通过图论算法解决了图像处理 and 机器学习算法无法解决的根尖点追踪和交叉根分割等问题, 提取种子萌发过程中的关键动态性状。使整个算法在种子遮挡或交叉情况下, 也能继续对关键表型如幼根和幼芽生长进行持续监测。为作物遗传育种和关键表型分析提供量化依据和智能解析技术, 为大规模种子萌发表型分析研究提出了新方法。

2) 虽然算法对关键动态性状如幼根长、幼根生长速率、幼芽长的分析可达到 92.2% ($n=188$)、71.9% ($n=191$) 和 89.7% ($n=115$) 的相关性, 但特征提取精度仍有提升空间, 后续研究也将进一步提高精度。

3) 目前已针对小麦、棉花和油菜三个作物完成了预实验, 后续将会研究更多物种, 扩展应用场景。经试验, 分割交叉幼根骨架算法对本文提到的两种交叉方式的分割效果较好, 还未针对更复杂的交叉场景进行研究。在今后的实验中, 还将尝试更多计算机视觉算法, 提高算法在复杂幼根交叉情况下的分割效果, 更准确地测量幼根生长速率, 并进一步提高算法的泛化性和鲁棒性。

参考文献 References:

- [1] Fenner M, Thompson K. The ecology of seeds [M]. Cambridge: Cambridge University Press, 2005.
- [2] FAO. The second report on the state of the world's plant genetic resources for food and agriculture[R]. Rome: FAO, 2010.
- [3] 高荣岐, 张春庆. 种子生物学[M]. 北京: 中国农业出版社, 2011: 203-209.
Gao R Q, Zhang C Q. Seed Biology[M]. Beijing: China Agriculture Press, 2011: 203-209 (in Chinese).
- [4] Yari L, Aghaalikani M, F Khazaei. Effect of seed priming duration and temperature on seed germination behavior of bread wheat (*Triticum aestivum* L.) [J]. Journal of Biomedical Materials Research Part B Applied Biomaterials, 2010, 5(1):1-6.
- [5] Fu YB. The vulnerability of plant genetic resources conserved ex situ [J]. Crop Sci, 2017, 57(5): 2314-2328.
- [6] Xie Q, Fernando K M C C, Mayes S, et al. Identifying seedling root architectural traits associated with yield and yield components in wheat[J]. Ann Bot, 2017, 119(7): 1115-1129.
- [7] Bao Y, Aggarwal P, Robbins N E, et al. Plant roots use a patterning mechanism to position lateral root branches toward available water[J]. Proc Natl Acad Sci U S A, 2014, 111(25): 9319-9324.
- [8] Colville L, Pritchard H W. Seed life span and food security[J]. New Phytologist, 2019, 224.
- [9] Elmasry G, Mandour N, Al-Rejaie S, et al. Recent Applications of Multispectral Imaging in Seed Phenotyping and Quality Monitoring—An Overview[J]. Sensors, 2019, 19(5):1090.

- [10] Halcro K, McNabb K, Lockinger A, et al. The BELT and phenoSEED platforms: shape and colour phenotyping of seed samples[J]. *Plant Methods*, 2020, 16(1).
- [11] Zhu F Y, Paul P, Hussain W, et al. SeedExtractor: an open-source GUI for seed image analysis[J]. *Frontiers in Plant Science*, 2021, 11:1-10.
- [12] Jahnke S, Roussel J, Hombach T, et al. phenoSeeder - A Robot System for Automated Handling and Phenotyping of Individual Seeds[J], *Plant Physiology*, 2016, 172(3):1358-1370.
- [13] Vlamincq L, Sang-Aram C, Botterman D, et al. Development of a novel and rapid phenotype-based screening method to assess rice seedling growth[J]. *Plant Methods*, 2020, 16(1):1-19.
- [14] Joosen R V L, Kodde J, Willems L A J, et al. GERMINATOR: a software package for high-throughput scoring and curve fitting of Arabidopsis seed germination[J]. *The Plant Journal*, 2010, 62(1):148-159.
- [15] Lurstwut B, Pornpanomchai C. Image analysis based on color, shape and texture for rice seed (*Oryza sativa* L.) germination evaluation[J]. *Agriculture and Natural Resources*, 2017, 51(5):383-389.
- [16] Colmer J, O'Neill C M, Wells R, et al. SeedGerm: a cost-effective phenotyping platform for automated seed imaging and machine-learning based phenotypic analysis of crop seed germination[J]. *New Phytologist*, 2020, 228(2):778-793.
- [17] Justine B, Merey R, Silvio G, et al. SeedQuant: a deep learning-based tool for assessing stimulant and inhibitor activity on root parasitic seeds[J]. *Plant Physiology*, 2021, 186(3):1632-1644.
- [18] Nagel K A, Lenz H, Kastenholz B, et al. The platform GrowScreen-Agar enables identification of phenotypic diversity in root and shoot growth traits of agar grown plants[J]. *Plant Methods*, 2020, 16(1):1-17.
- [19] Gong L, Du X F, Zhu K, et al. Pixel level segmentation of early-stage in-bag rice root for its architecture analysis[J]. *Computers and Electronics in Agriculture*, 2021,186.
- [20] Yasrab R, Atkinson J A, Wells D M, et al. RootNav 2.0: deep learning for automatic navigation of complex plant root architectures[J]. *Gigascience*, 2019, 8(11):1-16.
- [21] Summerfield M. Rapid GUI programming with Python and Qt: the definitive guide to PyQt programming[M]. Prentice Hall, 2008.
- [22] Peterson L E. K-nearest neighbor[J]. *Scholarpedia*, 2009, 4(2): 1883.
- [23] Hearst M A, Dumais S T, Osman E, et al. Support vector machines[J]. *IEEE Intelligent Systems & Their Applications*, 1998, 13(4):18-28.
- [24] Breiman L. Random forests[J]. *Machine Learning*, 2001,45(1):5-32.
- [25] Harris C G , Stephens M. A combined corner and edge detector[J]. *Computer Science*, 1988.