arXiv:2601.00207v1 [cs.CV] 1 Jan 2026

CropNeRF：基于神经辐射场的作物计数框架

Md Ahmed Al Muzaddid 和 William J. Beksi

摘要——**精准的作物计数对于有效的农业管理和制定明智的干预策略至关重要。然而，在户外田间环境中，部分遮挡与从单个视角难以区分成簇作物和单个作物这一固有歧义相结合，对基于图像的分割方法构成了巨大挑战。为解决这些问题，我们提出了一种新颖的作物计数框架，通过三维实例分割实现精确计数。我们利用从多个视角拍摄的二维图像，并为神经辐射场（NeRF）视图合成关联独立的实例掩码。我们提出了作物可见性和掩码一致性分数，并将其与 NeRF模型提供的三维信息结合使用。由此可实现对三维作物实例的有效分割，并获得极高精度的作物计数。此外，我们的方法无需依赖作物专属参数调优。我们在三个农业数据集上验证了该框架，这些数据集分别包含棉铃、苹果和梨；结果表明，尽管作物颜色、形状和大小存在显著差异，该方法仍能保持稳定的计数性能。与最先进水平的对比分析表明，该方法在作物计数任务中表现更优。最后，我们贡献了一个棉花植株数据集，以推动该主题的后续研究。**

索引词——**农业自动化；自动化计算机视觉；目标检测、分割、分类**

一、引言

随着移民劳动力不断减少、全球人口持续增长，农业部门面临着亟待创新解决方案的重大挑战。此外，气候变化带来的不利影响，例如难以预测的天气模式和水资源短缺，使得制定更具适应性和效率的农业实践成为必要。通过利用先进技术（如机器人自动化、计算机视觉、传感器网络等）来提高资源利用效率并优化生产力，精准农业为应对这些挑战提供了一种切实有效的方案。

精准农业中的作物计数对于优化资源、估算产量和采后管理至关重要。然而，在田间条件下进行作物计数仍面临诸多问题。例如，枝叶、树枝和邻近植株造成的遮挡会使作物检测和再识别更加困难，从而增加重复计数的可能性。此外，作物重叠和聚类会妨碍对单个实例的精确区分，而作物在不同生长阶段的颜色、形状和大小差异也进一步增加了计数难度。

人们提出了各种方法来应对这些挑战 [1], , 但很少有方法利用

作者就职于美国德克萨斯州阿灵顿的德克萨斯大学阿灵顿分校计算机科学与工程系。电子邮箱：mdahmedal.muzaddid@mavs.uta.edu、william.beksi@uta.edu。

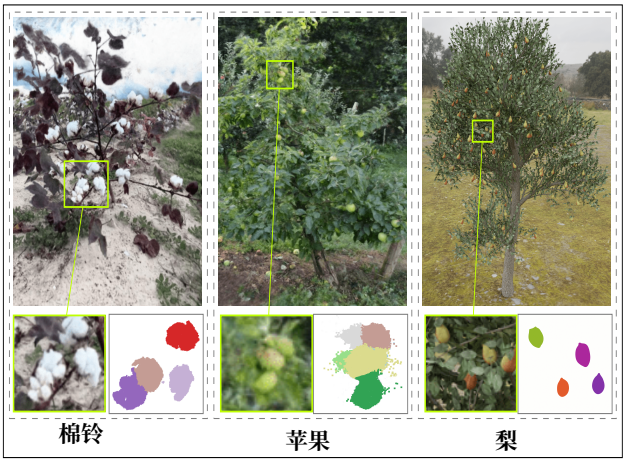

图 1：基于三维实例分割的作物计数概览。放大视图分别展示了棉花植株、苹果树和梨树上的棉铃、苹果和梨簇。使用 CropNeRF 得到的这些簇的实例分割结果以彩色显示在放大视图的右侧。

这一问题。本文提出了一种新颖的实例分割方法，能够在重建的三维环境中有效区分各个实例，从而准确统计目标作物的数量。我们的方法能够抵御其所依赖的二维实例分割掩码中的误差。此外，该方法利用体积信息处理遮挡和透视引起的歧义等问题，并且无需重新校准即可用于物理属性各异的作物。

本文提出了一种作物神经辐射场（CropNeRF）三维实例分割框架，可准确区分目标作物的各个实例，见图1。通过结合可见性和实例掩码一致性分数，CropNeRF 能有效处理遮挡和视角变化导致的歧义，同时适用于尺寸各异的作物，无需进行敏感的参数调优。与现有的实例分割方法（例如 [2], [3]）不同，后者依赖于建立多张图像中实例掩码之间的对应关系，而 CropNeRF 旨在独立处理每个掩码。这省去了显式的图像间映射，从而简化三维分割过程并提高准确性。

给定来自多个视角的一组图像，以及相应的相机位姿和目标作物的二维实例掩码，CropNeRF 能够通过三维实例分割准确地统计作物

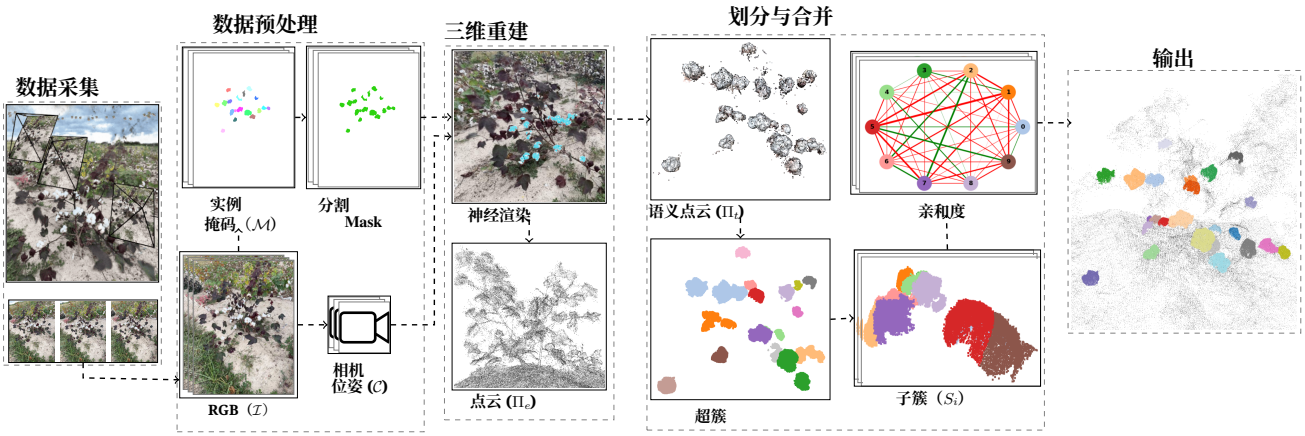

图 2：作物计数流程。在数据采集步骤中，从不同视角拍摄目标作物的多张图像。在数据预处理阶段，从采集的图像（$\mathcal{I}$）中提取相机位姿（$\mathcal{C}$），并将作物分割为实例掩码（$\mathcal{M}$）。随后，根据实例掩码生成语义分割掩码。在三维重建阶段，利用提取的相机位姿、采集的图像和语义掩码训练语义 NeRF 模型。训练后的模型会生成表示作物的语义点云（Π_t）以及表示环境的点云（Π_e）。在划分与合并步骤中，首先将语义点云划分为超簇，再划分为子簇（S_i）。根据子簇可见性和掩码一致性分数计算子簇之间的亲和度。最后，根据子聚类亲和度合并子簇，生成最终输出。

实例。为此，我们在建模语义场的同时，学习高保真度的三维重建。随后，从语义场中提取仅表示目标作物的点云，并将其划分为子簇。根据相机位姿、对应的点云和相关实例掩码，计算子簇的可见性和掩码一致性。然后将这些分数在多个视图间汇总，把子簇合并为目标作物的不同实例，从而实现精确的作物计数。

总而言之，我们的贡献如下。

- 我们开发了一种三维实例分割方法，能够区分颜色、形状和大小各异的作物。
- 我们提出了一种掩码可靠性分数，综合考虑作物可见性和掩码一致性，从而能够稳健应对遮挡和标注差异。
- 我们发布了一个公开的田间棉花植株数据集，专为三维渲染和棉铃计数任务设计。

与本项目相关的源代码、数据集和多媒体材料可在以下网址获取：https://robotic-vision-lab.github.io/cropnerf。

II. 相关工作

A. 基于图像的技术

基于图像的方法通常采用目标检测来识别图像中的作物。例如，Chen 等人 [4]利用多个卷积神经网络（CNNs），将输入图像映射为水果总数。类似地，H¨ani 等人[5] 将作物计数定义为一个多类分类问题，其中卷积神经网络会为每个检测到的作物簇预测一个计数。Tedesco 等人 [6] 采用检测模型（例如 Faster R-CNN [7], SSD [8], 和 MobileNetV2 [9]），并根据计算限制选择架构。同样，James 等人 [10], [11] 评估了目标检测模型以及迁移学习在对树上和地面上的柑橘类水果进行分类、以估算产量方面的应用。尽管现代目标检测模型在户外光照条件下能够达到较高的准确率，但由于视角有限且缺乏深度信息，其性能往往会受到遮挡的不利影响。

B. 基于视频的跟踪方法

为应对遮挡并实现农场级作物计数，以往研究采用了基于检测式跟踪范式的多帧技术。例如，Smitt 等人 [12] 通过使用 Mask R-CNN [13] 进行分割，并结合轮式里程计和深度信息提高跟踪精度，从而对 RGBD视频中的甜椒进行计数。Zhang 等人 [14] 改进了 YOLOv3 [15]，以提升小型水果的检测效果，并采用基于区域的计数策略进行跟踪。Al Muzaddid 和 Beksi [16]创建了一种棉铃计数框架，利用相邻棉铃之间的空间关系重新识别被遮挡的实例。在后续工作中，Al Muzaddid 等人 [17]开发了一种结合外观与运动信息的作物跟踪方法。尽管这些方法有效，但其性能取决于目标检测器和跟踪器的准确性。

C. 基于三维重建的解决方案

运动恢复结构（SfM）已被用于生成果园的三维点云重建，以实现精确的作物定位。Sun 等人提出的一种棉铃计数方法[18] 利用 SfM 从多视角田间图像重建的三维点云。Matos 等人 [19] 提出了一种帧间跟踪技术，利用从 SfM 获取的三维信息对图像序列中的水果进行计数。然而，这些方法通常需要大量的

相机标定，并且在枝叶茂密的环境中可能产生带噪重建。

NeRF 是一种广受欢迎的三维重建方法。它能够提供高保真度的体积表示，而无需进行大量相机标定。例如，FruitNeRF 是一种基于 NeRF 的水果计数框架。它依赖于基于聚类的分割，该方法假设作物尺寸一致，并且需要手动进行参数调优。这限制了 FruitNeRF 在不同作物生长阶段的泛化能力。相比之下，我们的方法无需根据作物尺寸和分布进行精细的参数调优。

三维高斯泼溅（3DGS）已成为 NeRF 的一种替代方法。Jiang 等人利用 3DGS 重建了高保真的棉花植株三维模型。然而，数据集是在室内、受控光照条件下采集的，且仅包含脱叶棉花植株。Zhang 等人将 3DGS 应用于重建小麦地块，并通过分割提取形态特征（例如穗长、宽度和体积）。然而，该分割技术依赖于一种迭代过程：根据投影三维高斯与二维分割掩码之间的重叠进行处理，并使用交并比进行量化，再通过经验阈值进行筛选。与之不同的是，我们的方法在评估掩码重叠时会考虑作物可见性，并且无需采用任意阈值处理。最终得到的是一种更稳健、与作物种类无关且适用于田间条件的分割框架。

III. 基于神经辐射场的作物计数框架

设 $\mathcal{I} = \{I_1, I_2, \ldots, I_n\}$ 为从多个视角捕获目标环境的一组RGB图像，其中 $I_i \in \mathbb{R}^{h\times w\times 3}$，且每幅图像的分辨率为 $h \times w$。此外，假设每幅图像都有目标作物的实例掩码$\mathcal{M} = \{M_1, M_2, \ldots, M_n\}$，并且 $M_j \in \mathbb{R}^{h\times w}$。接下来，我们从这组图像中提取相机位姿 $\mathcal{C} = \{C_1, C_2, \ldots, C_n\} \in \mathbb{R}^{3\times 4}$。为保持一致性，我们使用 C_j 和 M_j 分别表示与图像 I_j相关联的相机位姿和实例掩码。请注意， C_j 可以交替用于指代j^{th} 相机、其位姿或相关的相机视图，具体含义由上下文确定。作物计数流程的概览如图2所示。

A. 体积渲染

给定图像 $\mathcal{I}$ 及其对应的相机位姿$\mathcal{C}$，我们训练NeRF模型，以重建高保真的环境三维表示。该模型通过将位置$x = (x, y, z)$和观察方向 $d = (\phi, \theta)$映射为体积密度 σ和RGB辐射 $c = (r, g, b)$，将场景编码于多层感知机中。密度场$\mathcal{F}_\sigma : x \to \sigma$ 被定义为位置的函数，而外观场 $\mathcal{F}_c : (x, d) \to c$ 同时取决于位置和观察方向。在语义场 [24]，概念的基础上，我们扩展NeRF框架，以整合作物特定信息。具体而言，我们利用一组掩码 $\mathcal{M}$ 将作物信息嵌入三维体积中，通过语义场 $\mathcal{F}_s : x \to s$ 预测空间逻辑值。

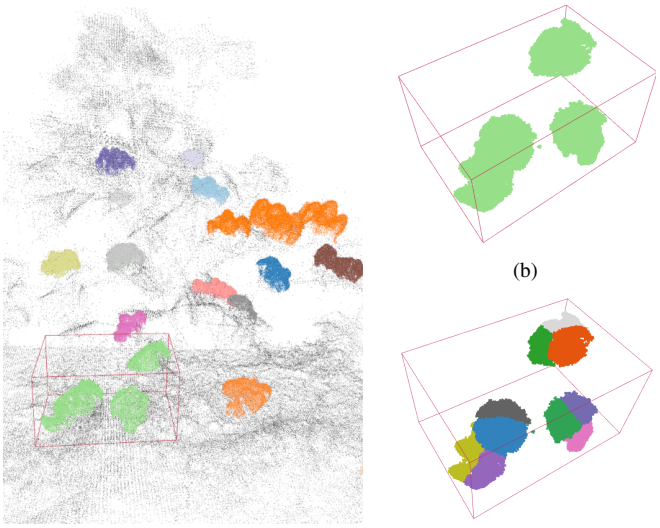

图3：点云聚类过程：(a) 表示作物的点云 Π_t 被分割为多个超簇，每个超簇以一种独特的颜色标识；(b) 三维视图中的单个超簇；(c) 该超簇进一步划分为多个子簇。

B. 点云生成

训练好的 NeRF模型在密度场 $\mathcal{F}_\sigma$中编码了整个场景（即土壤、树干、茎、枝叶和作物）的空间信息。相反，语义场 $\mathcal{F}_s$ 仅表示三维体积内目标作物的空间分布。通过对这两个场进行采样 $\mathcal{F}_\sigma$ 和$\mathcal{F}_s$，我们分别生成两个独立的点云：一个表示环境（包括目标作物），另一个仅表示目标作物。为了提取表示环境 Π_e的点云，我们从 $\mathcal{F}_\sigma$中均匀采样点。然而，通过对 $\mathcal{F}_s$ 采样来提取仅表示目标作物 Π_t 的点，会产生稀疏的点云。为了解决这一问题，我们采用 FruitNeRF 使用的方法，根据采样语义点对应的密度值对其进行筛选，从而获得更稠密的表示。

C. 点云预处理

我们采用两阶段方法对点云进行预处理： Π_t 在第一阶段， Π_t 被划分为超簇，

$$\psi_1 \ \psi_2 \ \ldots \ \psi_\kappa \qquad \Pi_t = \dot{\bigcup}_{k=1}^{\kappa} \psi_k ,,,,$$

使得其中 $\bigcup$ 表示不相交集合的并（图 3a）。此步骤旨在分离空间上相距较远且相互独立的作物实例，从而将问题分解为更小、更易处理的子问题，以便高效解决。我们采用 DBSCAN [25] ，根据点密度识别超簇，而无需预先假设簇的数量或大小。

我们根据聚类结果移除离群点，以细化提取出的超簇，确保其更加连贯

分割（图 3b）。值得注意的是，我们的方法的输出不受超簇数量的影响。例如， Π_t 可以视为一个超簇，而不会影响最终的分割结果。不过，为了提高计算效率，并尽量减少点云中彼此独立区域之间不必要的交互，我们将第一阶段聚类作为预处理步骤。每个超簇都独立进行分割，然后汇总所有超簇的结果以生成输出。

在第二阶段，我们进一步将一个超簇划分 ψ_i为 K 子簇。这通过图中所示的K均值聚类算法实现。3c，使得 $\psi_i = \bigcup_{k\in K} S_k$ 。此步骤的目标是将一个超簇划分为更小的子簇，并确保每个作物实例至少包含一个或多个子簇。我们通过选取更大的 K 值来减小每个簇的大小。请注意，当 K适度较大时，分割结果对 K 的变化仍然稳健，正如评估部分（第IV-G节）所讨论的那样。

D. 子簇可见性

由于遮挡，从某些视角只能观察到部分子簇。为量化其可见性，我们为从相机 $\mathcal{C}_j$观察到的每个子簇 S_i 赋予一个分数 $v_{ij} \in [0,1]$ 。更正式地说，

$$v_{ij} = \frac{Area\ of\ S_i\ visible\ from\ \mathcal{C}_j}{Area\ of\ S_i\ visible\ from\ \mathcal{C}_j\ without\ occlusions} = \frac{Area\left(\mathcal{V}_j\left(S_i\right)\right)}{Area\left(\mathcal{P}_j\left(S_i\right)\right)}. \tag{1}$$

在（1）中， $\mathcal{P}_j$（S_i）是一个函数，它使用相机矩阵将点云 S_i投影到相机平面上，其中相机矩阵是由 $\mathcal{C}_j$导出的内参矩阵与外参矩阵的乘积。我们将此称为S_i的无遮挡投影。遮挡感知投影函数表示为 $\mathcal{V}_j$（S_i），并考虑茎、枝叶和作物等环境元素（即相机 $\mathcal{C}_j$ 与 S_i 之间可能遮挡 S_i的任何物体）。我们通过将 $\mathcal{V}_j$（S_i）和 S_i 都 Π_e 投影到相机平面上，并结合深度信息来处理遮挡，从而得到该结果。

E. 掩码一致性

实例掩码 $\mathcal{M}$ 基于二维图像，因此无论是由机器学习模型还是人工标注者标注，都不可避免地容易出现不一致。这些偏差源于固定的相机视点以及缺少深度信息，妨碍了对彼此间距较近的实例进行精确区分。为考虑掩码中的差异，我们定义了一致性分数 $c_{ij} \in [0,1]$。该分数定量评估与 $\mathcal{V}_j(S_i)$重叠的掩码 M_j 的一致性，其中 $\mathcal{V}_j(S_i)$是对 S_i的遮挡感知投影。一致性分数的计算方式如下：

$$c_{ij} = \frac{\max\limits_{l} Area\left(\mathcal{V}_j\left(S_i\right) \cap M_{jl}\right)}{Area\left(\mathcal{V}_j\left(S_i\right)\right)}, \tag{2}$$

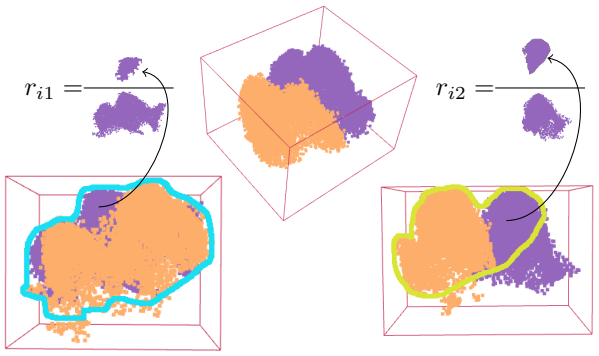

图 4：计算掩码可靠性分数的示意图。紫色子簇的分数根据两个不同的相机视图计算。分子仅考虑子簇可见投影区域中与掩码重叠的部分（由青色和黄色边界表示）。

其中， M_{jl} 表示 M_j 中标记为实例 l的区域。max 操作可确保当M_j 中多个不同标签与 $\mathcal{V}_j(S_i)$重叠时，仅考虑与 $\mathcal{V}_j(S_i)$重叠最多的标签 l所对应的区域。我们还将标签赋为，

$$\lambda_{ij} = \underset{l}{\operatorname{argmax}}\ Area\left(\mathcal{V}_j\left(S_i\right) \cap M_{jl}\right), \tag{3}$$

在视图 M_j 中的子簇 S_i 。

F. 掩码可靠性

为了量化与子簇相关联的实例掩码的置信度 S_i 在视图 M_j中的可靠性分数 $r_{ij} \in [0,1]$。如图4所示， r_{ij} 通过结合子簇可见性分数 v_{ij}和掩码一致性分数 c_{ij}计算得出。具体而言，

$$\begin{aligned} r_{ij} &= v_{ij} c_{ij} \\ &= \frac{Area\left(\mathcal{V}_j\left(S_i\right)\right)}{Area\left(\mathcal{P}_j\left(S_i\right)\right)} \cdot \frac{\max\limits_{l} Area\left(\mathcal{V}_j\left(S_i\right) \cap M_{jl}\right)}{Area\left(\mathcal{V}_j\left(S_i\right)\right)} \\ &= \frac{\max\limits_{l} Area\left(\mathcal{V}_j\left(S_i\right) \cap M_{jl}\right)}{Area\left(\mathcal{P}_j\left(S_i\right)\right)}. \end{aligned} \tag{4}$$

G. 合并过程

合并步骤旨在将构成同一作物实例的子簇合并。理想情况下，对于给定的视图 C_j，如果一对子簇 S_i 和 $S_{i'}$ 在 M_j中具有相同的实例标签，我们就会将它们合并。然而，我们从 n个对应视图中获得了 n此类观测。因此，为了纳入所有观测，我们引入 S_i 与 $S_{i'}$之间的子簇亲和度分数 $\alpha_{ii'}$ 。该分数根据 n 个视图中的掩码可靠性分数，通过以下方式确定：

$$\alpha_{ii'} = \sum_{j=1}^{n} r_{ij} r_{i'j} \cdot (-1)^{\mathbb{1}\{\lambda_{ij} \neq \lambda_{i'j}\}}, \tag{5}$$

其中 $\mathbb{1}\{\cdot\}$是一个指示函数，条件满足时取值为 1，否则取值为 0。如图 5所示，对于超簇中的每对子簇，我们都会计算其亲和度分数，并构建一个加权完全

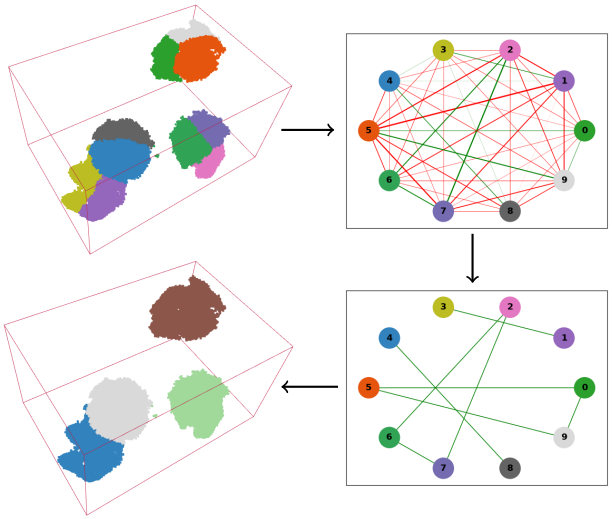

图 5：子簇合并过程。左上：一个超簇被划分为 10 个子簇，每个子簇以一种独特的颜色表示。右上：加权完全图展示了子簇之间的亲和度，其中正亲和度和负亲和度分别以绿色和红色表示。每条边的宽度与其亲和度值成正比。右下：通过标签传播，根据亲和度分数将图划分为更小的子图。左下：属于同一子图的相应子簇被合并。



(a) 棉花植株。　　(b) 样本视图。

图 6：棉花植株数据集中的示例图像。(a) 展示了一株棉花植株，(b) 展示了从不同视角拍摄的样本视图。

图。每个子簇映射为图中的一个节点，一对子簇之间的亲和度分数由相应的边权重表示。最后，我们在图上应用标签传播算法 [26]，将其划分为代表各个作物实例的子图。

IV. 评估

A. 棉花植株数据集

我们收集了一个棉花植株数据集，其中包含在德克萨斯农工大学研究农场记录的 8 株棉花植株。图像使用 Apple iPhone 拍摄，分辨率为 1040×1920 像素。每株植株约拍摄 150 张图像，拍摄距离为 1 米，并通过记录多个视角。图 6 展示了数据采集过程中具有代表性的相机视点样例。为了估计图像位姿，我们使用了 Spectacular AI [27]移动应用程序。通过直接观察并手动计数所有棉铃，获得了每株植株的真实计数。为了生成实例掩码，我们采用了 Segment Anything Model（SAM） [28]，这一预训练的二维实例分割模型，并使用手动边界框监督作为输入提示。

B. 实验设置

除了在我们的棉花植株数据集上评估 CropNeRF 外，我们还使用 FruitNeRF 提供的苹果树和梨树数据评估了其性能。苹果树数据集包含 3 棵树，每棵树均使用配备 35 mm 镜头的数码单反相机拍摄，生成分辨率为 4000×6000 像素的图像。每棵树大约从 3 m 的距离拍摄了 350 张图像，涵盖不同的角度和高度。我们使用 COLMAP [29] 估计相机位姿。梨树数据集使用 Blender [30]合成生成。虚拟相机的焦距设置为 35 mm，图像分辨率设置为1024×1024 像素。在不同类型的作物数据中，棉花带来了更大的挑战。这是因为棉铃的大小和形状差异显著，而苹果和梨的尺寸则更为均匀。因此，我们的评估主要聚焦于棉花。

C. 实现细节

在三维重建和点云提取方面，我们使用了现有的 NeRF模型。具体而言，CropNeRF 基于 Nerfacto [31], 并结合语义场构建，其架构细节遵循 FruitNeRF。鉴于棉花植株与苹果树和梨树在尺寸上的差异，我们使用 FruitNeRF 网络渲染棉花植株，而使用 FruitNeRF-Big 网络渲染苹果树和梨树。模型分别训练了 30,000 和 100,000 个 epoch。为了使用语义掩码训练模型，我们将 SAM 生成的实例掩码中的所有实例赋予统一标签，从而得到语义掩码。CropNeRF 在一台运行 Ubuntu 22.04.5 LTS 的机器上训练，该机器配备 Intel Xeon Gold 6330 2.00 GHz CPU、64 GB 内存和 NVIDIA A100 GPU。训练每个棉花植株模型约需 12 分钟。

为了将导出的点云划分为超聚类，我们采用了 DBSCAN，并使用参数 $eps = 0.02$ 和 $min_points = 30$。在第二阶段，我们执行了 K均值聚类，将每个超聚类划分为 K 子簇，其中 $K = 10$。值得注意的是，两个数据集使用了相同的参数设置。在第IV-G节中，我们通过分析子簇数量与其对应计数之间的关系，展示了 CropNeRF 的鲁棒性。为了计算（ $\mathcal{V}_j$）S_i在（1）和（2）中的遮挡感知投影，我们采用了计算机图形学中常用的 Z 缓冲技术。树级计数

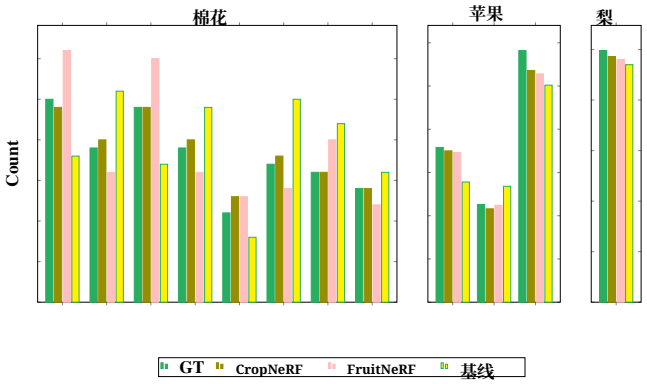

图 7：棉铃、苹果和梨的计数对比，包括真实标注（GT）、CropNeRF、FruitNeRF 和基线方法。

通过对所有超聚类中的各个实例计数求和得到。

D. 计数准确率

CropNeRF 使用相同的输入分割掩码，与基线方法及近期基于三维渲染的作物计数方法进行了比较。基线方法通过运动恢复结构算法使用 COLMAP 生成点云，随后采用 FruitNeRF 使用的聚类方法进行分割。图7 展示了单株/单树作物计数结果。对于 FruitNeRF 在苹果和梨数据集上的结果，我们报告了原始研究中发表的数值；该研究针对每棵树分别调整了参数设置。为了评估 FruitNeRF 在棉花数据集上的表现，我们将 $template\ size$参数设为 30 毫米，以与 Wallace 和 Fields 报告的平均棉铃直径保持一致 [32]。我们还使用多种计数误差指标评估了 CropNeRF，结果列于表I。

与 FruitNeRF 相比，CropNeRF 始终表现出更优的性能。具体而言，CropNeRF 对棉铃和苹果的平均绝对百分比误差（MAPE）分别为 4.9% 和 4.7%，表明其在不同作物类型上均保持了稳定的准确度。相比之下，FruitNeRF 对棉铃和苹果的 MAPE 分别为 18.1% 和 4.5%，显示其在这两种作物上的性能存在显著差异。这种差异源于基于聚类的点云分割方法依赖于预先定义的作物模板尺寸和形状。对于梨，两种框架的 MAPE 均相对较低（CropNeRF 为 2.4%，FruitNeRF 为 3.6%），低于苹果和棉铃；这可归因于梨通常单独生长，而非密集成簇。与 Cotton3DGaussians 相比， [22]，CropNeRF 尽管在真实户外环境中运行，仍实现了更低的计数误差。这些结果凸显了 CropNeRF 在不同农产品计数任务中的鲁棒性和通用性。

E. 掩码差异处理

如图所示 8a，SAM 生成的实例掩码可能存在不一致之处，即它们未能与相应的实例完全对齐。此外，多个不同的

方法	棉铃		苹果		梨	
	R	**M**	**R**	**M**	**R**	**M**
基线	5.9	30.8	34.8	18.2	14.0	5.6
FruitNeRF [21]	3.9	18.1	15.9	**4.5**	9	3.6
Cotton3DGaussians [22]	1.7	9.2	-	-	-	-
CropNeRF（我们的方法）	**1.0**	**4.9**	**13.7**	4.7	**6.0**	**2.4**

表 I：不同方法的计数误差比较。**R** 表示均方根误差，**M** 表示平均绝对百分比误差。

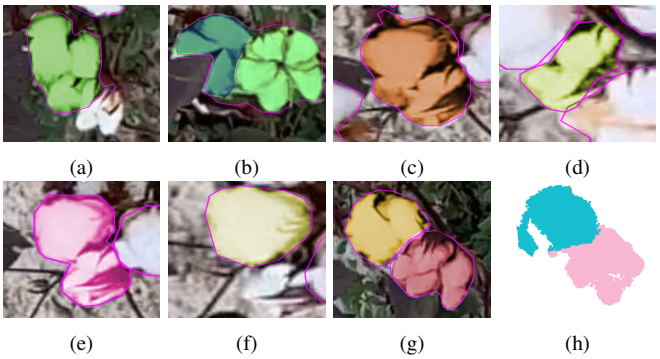

图8：掩码不一致的示例，其中两个棉铃实例在多个视图中被不一致地标注：（a）掩码未与相应的实例边界对齐；（b-e）两个棉铃被错误地标注为单个实例；（f）检测器漏检了其中一个棉铃；（g）棉铃被正确标注；（h）CropNeRF 的可靠性分数减弱了不一致性，从而能够成功识别棉铃。

实例可能被错误地标记为单个实例（例如，图8b-8e），而在其他情况下，某些实例可能因漏检而未被标记（例如，图 8f）。这些差异在棉花植株数据集中尤为明显，因为棉铃形状不规则，难以准确区分各个实例。

CropNeRF 能够缓解其中许多误差，并成功识别作物实例，如图8h所示。例如，在图8a和图8f所示的场景中，由于掩码一致性较低，可靠性分数也较低。同样，在图8c、图 8d和图8e所示的场景中，由于可见性分数降低，可靠性分数也会降低。由于可靠性分数较低，这些场景对后续亲和度计算的影响减弱，从而得到对掩码差异具有鲁棒性的分割结果。

F. 实例掩码有效性

我们评估了 CropNeRF 使用稀疏的实例掩码集合执行计数任务的效果。在这项实验中，我们改变了可用掩码的数量 m ，范围从 5 到 150，并分析了不同棉花植株的棉铃计数误差。对于每个 m值， m 均从包含 150 个掩码的完整集合中均匀抽取一组9掩码。图中展示的结果表明，在大多数情况下，仅需 15 到 30 个掩码即可实现准确计数。

G. 子簇数量的影响

设计 CropNeRF 的一个主要目标是尽量减少对特定作物参数优化的需求。因此，

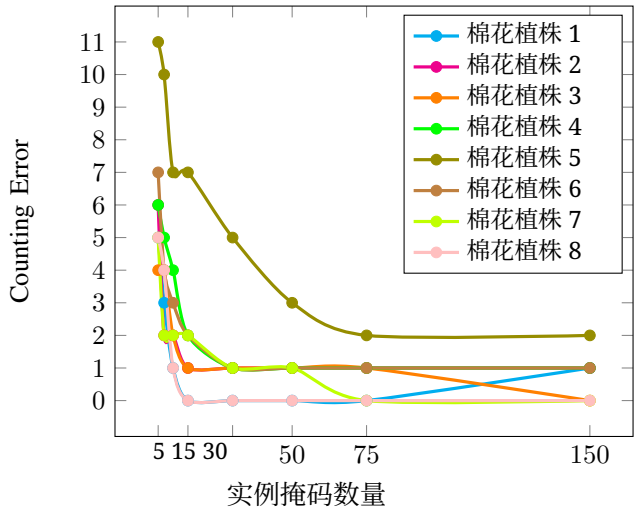

图 9：实例掩码数量与计数误差。

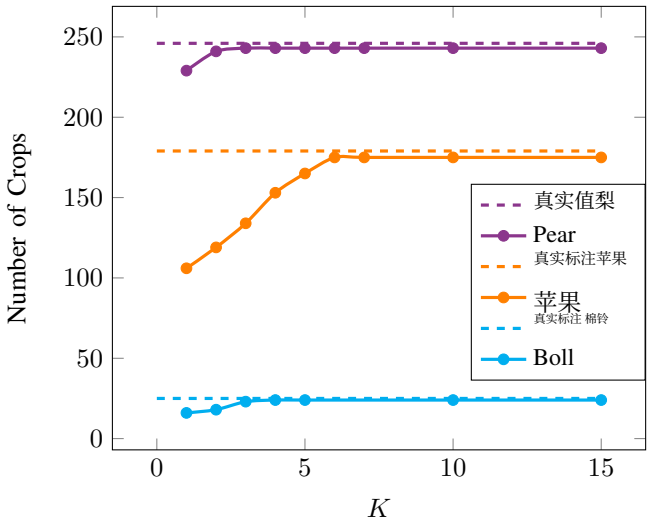

图10：子簇数量（K）与作物计数的关系。虚线表示真实计数（GT），实线表示 CropNeRF 识别出的作物数量。

我们分析了参数K与计数准确率之间的关系。该参数是k-means聚类算法的关键输入。我们选取苹果树、梨树和棉花植株数据集中的样本，以评估 K的影响。结果如图10所示。

对于棉铃，结果显示，在 $K \geq 4$的情况下，计数保持稳定。这种稳定性源于大多数情况下，一个超簇中的实例数量不超过五个。在棉花植株数据集中，单个超簇中观察到的棉铃数量最多为四个。同样，对于苹果，在 $K \geq 6$的情况下，计数也大致保持不变，因为苹果簇通常最多包含六到七个实例。对于梨，簇通常仅包含一两个梨，因此在 $K \geq 2$的情况下计数保持稳定。

基于这些实证发现，我们在所有实验中均设置为 $K = 10$ 。一般而言，我们建议选择 K 使其大于一个簇中预期的实例数。不过， K 不应过大，因为子簇数量增加会导致计算复杂度上升。

方法	可见性	Mask	LPA	平均绝对百分比误差%
基线				7.1
+可见性	✓			6.3
+掩码		✓		6.5
+掩码，+可见性	✓	✓		5.4
CropNeRF	✓	✓	✓	**4.9**

表 II：在棉花植株数据集上进行的消融研究。第一行是基线方法，仅使用实例掩码合并子簇。在后续各行中，加入可见性分数和掩码一致性分数，逐步提高了计数准确率。最后一行代表 CropNeRF，除使用可见性分数和掩码一致性分数外，还应用了标签传播算法（LPA），因此误差最低。

H. 消融研究

为了了解 Crop- NeRF 中关键设计选择的影响，我们考察了以下组件：可见性分数、掩码一致性分数和标签传播算法。消融研究结果见表II，具体说明如下。基线方法不使用可见性分数和掩码一致性分数，而是仅在 $\alpha_{ii'} > 0$的条件下合并子簇对，其中 $\alpha_{ii'} = \sum_{j=1}^{n} (-1)^{\mathbb{1}\{\lambda_{ij} \neq \lambda_{i'j}\}}$。该 +可见性和 +掩码方法分别显示在第二行和第三行，根据可见性分数 $\alpha_{ii'} = \sum_{j=1}^{n} v_{ij} \cdot (-1)^{\mathbb{1}\{\lambda_{ij} \neq \lambda_{i'j}\}}$ 或掩码一致性分数$\alpha_{ii'} = \sum_{j=1}^{n} c_{ij} \cdot (-1)^{\mathbb{1}\{\lambda_{ij} \neq \lambda_{i'j}\}}$计算亲和度分数。仅在 $\alpha_{ii'} > 0$的条件下合并子簇对。

CropNeRF 综合考虑可见性分数和掩码一致性分数来计算可靠性分数，并采用标签传播策略合并子簇。消融研究结果表明，可见性分数是决定分割性能的主导因素，其次是掩码一致性分数。因此，当亲和度分数由这两个因素得出时，仅当 $\alpha_{ii'} > 0$ 时才进行合并，便足以实现精准分割。

V. 局限性与未来工作

尽管 NeRF 能够生成高保真度的三维表示，但它们需要大量计算资源。在 CropNeRF 流程中，NeRF 模型训练是计算量最大、耗时最长的阶段。此外，由于 RGB 相机的光谱局限，在茂密的冠层结构中准确检测某些作物仍然困难。未来的工作不仅将关注 NeRF 的替代方案，还将整合归一化差异植被指数和热成像等其他传感模态，以增强作物检测能力并提高框架的整体鲁棒性。

VI. 结论

本文介绍了 CropNeRF，这是一种遮挡感知计数框架，可通过三维实例分割准确统计单个作物实例。此外，我们创建了一个多视角棉花植株图像数据集，以推动该领域的未来研究。CropNeRF 具有很强的鲁棒性，且针对不同作物类型无需进行精细的参数调优。我们在三个不同的作物数据集（棉铃、苹果和梨）上评估了 CropNeRF 的有效性，这些作物在颜色、形状和大小方面差异显著。实验结果表明，该框架在所有三个数据集上的计数都很准确，超过了最先进水平，凸显了其在农业应用中的适应性和可靠性。

致谢

MdAhmed Al Muzaddid 获得了德克萨斯大学阿灵顿分校博士论文奖学金的资助。我们感谢 Aaron J. DeSalvio 协助在德克萨斯农工大学进行数据采集。我们感谢德克萨斯大学奥斯汀分校的德克萨斯高级计算中心（TACC）提供软件、计算和存储资源，为本研究成果作出了贡献。

参考文献

[1] L. He、W. Fang、G. Zhao、Z. Wu、L. Fu、R. Li、Y. Majeed 和 J. Dhupia，《果园水果产量预测与估算：直接和间接方法的最新综合综述》，《农业中的计算机与电子学》，第 195 卷，第 106812 页，2022 年。[2] Y. Liu、B. Hu、J. Huang、Y.-W. Tai 和 C.-K. Tang，《实例神经辐射场》，载于《*IEEE/CVF*国际计算机视觉会议论文集》，2023 年，第 787–796 页。[3] M. Ye、M. Danelljan、F. Yu 和 L. Ke，《高斯分组：在三维场景中分割和编辑任意对象》，载于《欧洲计算机视觉会议论文集》。Springer，2024 年，第 162–179 页。[4] S. W. Chen、S. S. Shivakumar、S. Dcunha、J. Das、E. Okon、C. Qu、C. J. Taylor 和 V. Kumar，《利用深度学习计数苹果和橙子：一种数据驱动方法》，《*IEEE*机器人与自动化快报》，第 2 卷，第 2 期，第 781–788 页，2017 年。[5] N. H¨ani、P. Roy 和 V. Isler，《使用卷积神经网络进行苹果计数》，载于《*IEEE/RSJ* 智能机器人与系统国际会议论文集》，2018 年，第 2559–2565 页。[6] D. Tedesco-Oliveira、R. P. da Silva、W. Maldonado Jr 和 C. Zerbato，《利用商业农田图像中的卷积神经网络预测棉花产量》，《农业中的计算机与电子学》，第 171 卷，第 105307 页，2020 年。[7] S. Ren、K. He、R. Girshick 和 J. Sun，《Faster R-CNN：借助区域建议网络实现实时目标检测》，*IEEE*《模式分析与机器智能汇刊》，第 39 卷，第 6 期，第 1137–1149 页，2016 年。[8] W. Liu、D. Anguelov、D. Erhan、C. Szegedy、S. Reed、C.-Y. Fu 和 A. C. Berg，《SSD：单次多框检测器》，载于《欧洲计算机视觉会议论文集》。Springer，2016 年，第 21–37 页。[9] M. Sandler、A. Howard、M. Zhu、A. Zhmoginov 和 L.-C. Chen，《MobileNetV2：倒置残差与线性瓶颈》，载于《*IEEE/CVF*计算机视觉与模式识别会议论文集》，2018 年，第 4510–4520 页。[10] J. A. James、H. K. Manching、M. R. Mattia、K. D. Bowman、A. M. Hulse-Kemp 和 W. J. Beksi，《CITDET：柑橘类水果检测基准数据集》，《*IEEE*机器人与自动化快报》，第 9 卷，第 12 期，第 10 788–10 795 页，2024 年。[11] J. A. James、H. K. Manching、A. M. Hulse-Kemp 和 W. J. Beksi，《通过迁移学习实现少样本水果分割》，载于《*IEEE*国际机器人与自动化会议论文集》，2024 年，第 13 618–13 624 页。[12] C. Smitt、M. Halstead、T. Zaenker、M. Bennewitz 和 C. McCool，《Pathobot：用于温室作物表型分析与干预的机器人》，载于《*IEEE*国际机器人与自动化会议论文集》，2021 年，第 2324–2330 页。[13] K. He、G. Gkioxari、P. Doll´ar 和 R. Girshick，《Mask R-CNN》，载于《IEEE/CVF国际计算机视觉会议论文集》，2017 年，第 2961–2969 页。[14] W. Zhang、J. Wang、Y. Liu、K. Chen、H. Li、Y. Duan、W. Wu、Y. Shi 和 W. Guo，《基于深度学习的田间柑橘检测与跟踪》，《园艺研究》，第 9 卷，第 uhac003 页，2022 年。[15] J. Redmon，《YOLOv3：渐进式改进》，arXiv 预印本 arXiv:1804.02767，2018 年。[16] M. A. Al Muzaddid 和 W. J. Beksi，《Ntrack：用于田间棉铃计数的多目标跟踪器与数据集》，IEEE《自动化科学与工程汇刊》，第 21 卷，第 9 期，第 7452–7464 页，2024 年。[17] M. A. Al Muzaddid、J. A. James 和 W. J. Beksi，《Croptrack：面向精准农业的再识别跟踪框架》，arXiv 预印本 arXiv:2512.24838，2025 年。[18] S. Sun、C. Li、P. W. Chee、A. H. Paterson、Y. Jiang、R. Xu、J. S. Robertson、J. Adhikari 和 T. Shehzad，《基于点云分割与聚类的棉铃原位三维摄影测量制图》，《国际摄影测量与遥感杂志》，第 160 卷，第 195–207 页，2020 年。[19] G. P. Matos、C. Santiago、J. P. Costeira、R. L. Saldanha 和 E. M. Morgado，《在间歇性遮挡和低帧率条件下跟踪并计数果园中的苹果》，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2024 年，第 5413–5421 页。[20] B. Mildenhall、P. P. Srinivasan、M. Tancik、J. T. Barron、R. Ramamoorthi 和 R. Ng，《NeRF：将场景表示为用于视图合成的神经辐射场》，《ACM通讯》，第 65 卷，第 1 期，第 99–106 页，2021 年。[21] L. Meyer、A. Gilson、U. Schmid 和 M. Stamminger，《FruitNeRF：统一的神经辐射场水果计数框架》，载于《IEEE/RSJ智能机器人与系统国际会议论文集》，2024 年，第 1–8 页。[22] L. Jiang、J. Sun、P. W. Chee、C. Li 和 L. Fu，《Cotton3DGaussians：用于棉铃制图和植株结构分析的多视图三维高斯泼溅》，《农业中的计算机与电子学》，第 234 卷，第 110293 页，2025 年。[23] D. Zhang、J. Gajardo、T. Medic、I. Katircioglu、M. Boss、N. Kirchgessner、A. Walter 和 L. Roth，《Wheat3DGS：利用高斯泼溅进行小麦穗的田间三维重建、实例分割与表型分析》，载于《IEEE/CVF国际计算机视觉会议论文集》，2025 年，第 5360–5370 页。[24] S. Zhi、T. Laidlow、S. Leutenegger 和 A. J. Davison，《利用隐式场景表示进行原位场景标注与理解》，载于《IEEE/CVF国际计算机视觉会议论文集》，2021 年，第 15 838–15 847 页。[25] M. Ester、H.-P. Kriegel、J. Sander 和 X. Xu，《一种基于密度的算法：在含噪声的大型空间数据库中发现簇》，载于《知识发现与数据挖掘国际会议论文集》，第 96 卷，第 34 期，1996 年，第 226–231 页。[26] U. N. Raghavan、R. Albert 和 S. Kumara，《一种用于检测大规模网络社群结构的近线性时间算法》，《物理评论 E—统计、非线性与软物质物理》，第 76 卷，第 3 期，第 036106 页，2007 年。[27] (2026) Spectacular AI。 [在线]。网址：https://github.com/ spectacularai [28] A. Kirillov、E. Mintun、N. Ravi、H. Mao、C. Rolland、L. Gustafson、T. Xiao、S. Whitehead、A. C. Berg、W.-Y. Lo、P. Doll´ar 和 R. Girshick，《任意分割》，载于《IEEE/CVF国际计算机视觉会议论文集》，2023 年，第 4015–4026 页。[29] J. L. Sch¨onberger 和 J.-M. Frahm，《重新审视运动恢复结构》，载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2016 年，第 4104–4113 页。[30] (2026) Blender。 [在线]。网址：https://www.blender.org [31] M. Tancik、E. Weber、E. Ng、R. Li、B. Yi、T. Wang、A. Kristoffersen、J. Austin、K. Salahi、A. Ahuja、D. McAllister、J. Kerr 和 A. Kanazawa，《Nerfstudio：用于神经辐射场开发的模块化框架》，载于《ACM SIGGRAPH会议论文集》，2023 年，第 1–12 页。[32] B. W. Wallace 和 C. A. Fields，《棉铃直径和种子数量与单铃籽棉重量的关系》，载于《全美棉花会议论文集》，第 2 卷，1997 年，第 1444–1446 页。