arXiv:2303.00460v1 [cs.RO] 1 Mar 2023

基于多智能体强化学习的水果采摘多臂机器人任务规划

Tao Li[1], Feng Xie[2],Ya Xiong[1] 和 Qingchun Feng[1]*

摘要—采摘机器人技术的发展为农业劳动力资源有限以及水果需求不断增长的问题提供了颇具前景的解决方案。尽管采摘机器人技术取得了显著进展，但此类技术在果园中的应用仍然有限。关键挑战在于提高作业效率。本文考虑机械臂间冲突、自由度耦合和动态任务，提出了一种适用于四臂采摘机器人的任务规划策略。所提方法采用马尔可夫博弈框架对四臂机器人采摘任务进行建模，从而避免求解 NP 难调度问题所带来的计算复杂度。此外，我们采用具有完全中心化协作协议的多智能体强化学习（多智能体强化学习）结构，训练基于多智能体强化学习的任务规划网络。我们通过若干次仿真和果园实验，将所提方法与现有方法进行比较，以验证其在多臂收获机器人上的有效性。

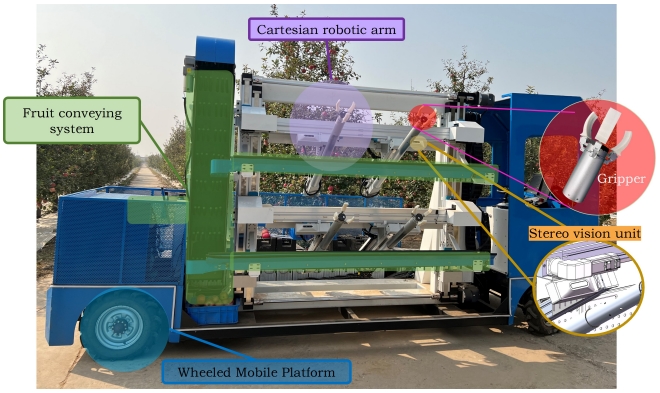


图1. 本文提出的多臂苹果采摘机器人。

I. 引言

水果生产中的采摘过程劳动密集且成本高昂 [1]。为降低劳动力成本，解决水果产业瓶颈的迫切需求是实现自主无人采摘。过去十年中，机器人采摘机的应用不断增加，被视为一种前景广阔的技术。

随着计算机和传感器技术的最新发展，果园中的采收机器人受到了广泛关注，并且在水果检测与定位 [2], 路径规划 [3], 机械设计 [4], 等方面取得了显著进展。近年来，采收机器人已应用于苹果 [5],橙子 [6],番茄 [7], 菠萝 [8], 等水果。

近年来，关于多臂采摘机器人的研究日益增多 [9]–[11] ，旨在提高作业效率，并取得了显著进展，例如双臂草莓采摘机器人 [10], 12、多臂苹果采摘机器人 [12], 4、多臂猕猴桃采收机器人 [9],等。一些多臂机器人已成功实现商业化。多臂收获机器人通常由多个机械臂、夹爪和视觉单元组成，用于同时进行采摘作业，从而提高效率。

尽管多臂机器人具有较大的工作空间和感知范围等优势，但仍面临诸多技术挑战。一般而言，与单臂机器人相比，多臂机器人需要更复杂的任务规划策略，才能实现更好的协作与效率 [13]。多臂采摘机器人的任务规划包括两个方面：采摘顺序和果实目标分配。当机器人需要采摘大量水果时，采摘顺序会导致机械臂在目标之间移动不同的距离 [14]。如果距离较大，机器人在从一个水果移动到另一个水果的过程中可能会产生冗余路径。其次，果实目标分配意味着每个机械臂都明确要采摘哪个目标，尤其是对于某些其他机械臂也能够到达的目标 [15]。避免通常会导致额外等待时间的臂间冲突，对于多臂协作具有重要意义。只有当各机械臂在其工作区域内拥有各自的目标时，才都有任务可执行。如果水果分布不均匀 [16], , 一个或多个机械臂可能会处于空闲状态，直到其他机械臂完成任务。因此，任务规划对于多臂采摘机器人而言是必要的，可减少冗余运动和机械臂之间的冲突，从而最大限度地提高工作性能。

然而，多臂机器人的任务规划仍面临诸多技术挑战。在实际操作中，果实采摘往往无法一次成功 [17],, 可能会失败。为提高采摘完成率，机器人需要具备重新抓取采摘失败的果实的能力，而不是机械地执行原有的任务规划策略。然而，现有的启发式规划方法计算复杂度较高，

*本研究由中国北京市科学技术计划项目（项目编号：Z201100008020009）、北京市农林科学院青年科研基金（QNJJ202318）和北京新星计划（20220484023）资助[1]T. Li、Y. Xiong 和 Q. Feng 就职于北京市农林科学院智能装备研究中心，中国北京 100097。Q. Feng 为通信作者。电子邮箱： {lit, yaxiong and fengqc}@nercita.org.cn [2]F. Xie 就职于江苏大学农业工程学院，中国镇江 212013。电子邮箱：2111916009@stmail.ujs.edu.cn

难以满足实时控制要求。另一项挑战是为多臂采摘多个水果建立并求解时间优化模型。本研究中的机器人涉及机械结构干涉、关节自由度耦合和多臂交互等多重约束。目前，针对阐明决策过程对整体作业时间的具体影响，以及建立并求解能够有效纳入多重约束的优化问题，仍缺乏可行方案。

为解决这些问题，我们为多臂采摘机器人设计了一种任务规划方案。本文的贡献如下：

1) 我们提出了一种采摘机器人系统，如图 1 所示，该系统由四个笛卡尔机械臂、相机、输送系统、照明设备和移动平台组成。2) 为尽可能降低多臂作业的总体采摘时间成本，我们采用马尔可夫博弈对序贯决策问题进行建模。3) 为解决多臂协作问题，我们提出了一种完全中心化的多智能体强化学习（MARL）框架协作策略。

II. 相关工作

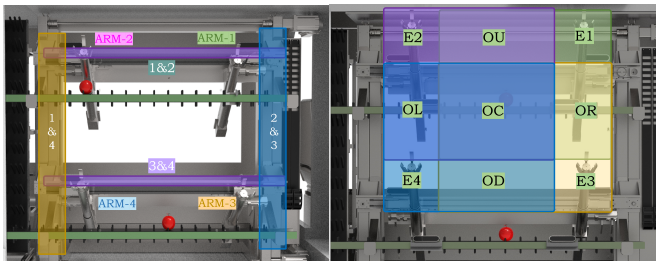
已有多种方法旨在实现多臂机器人协同采摘水果。根据相关文献，多臂机器人协作可分为三类。第一类是类人双臂模式，其中一条机械臂执行采摘动作，另一条机械臂则提供辅助 [18], [19]。这种模式适用于需要高超技巧的采摘任务，例如采摘茄子和甜椒等具有细长茎秆的作物；在这类任务中，采摘时不宜拉扯并分离果实，而应由末端执行器准确定位果梗以便切割。在这种模式下，机器人可以移除障碍物，使果实茎秆露出，并摘下果实，从而促进采摘过程。值得庆幸的是，种植者正越来越多地通过调整植物形态来促进作物生产。例如，修剪成平面冠层可以提高果实的可及性；美国的苹果种植者也一直在移除靠近棚架铁丝和树干的果实，以优化适合机器人采摘的果实分布 [20]。

为了进一步提高机器人收割效率，许多多机械臂机器人采用并行协作模式（第二类），例如 FFR [12], AGROBOT [21], HarvestCROO [22], andtherobots in [10], [23]。在这种模式下，机械臂以空间隔离的方式采摘水果，从而缩短单个水果采摘的平均循环时间。例如，Harvest CROO 的采摘机声称每十秒可采摘三个水果。值得注意的是，循环时间是衡量收割效率的关键指标，但不同研究者对其定义各不相同 [6]。在一些研究中，为缩短循环时间，会忽略感知或果间移动所需的时间；但在大规模收割作业中，整体采摘速率可能反而较低 [24]。原因有两点。首先，移动时间可能很长，这意味着机器人在机械臂从一个水果移动到另一个水果的过程中可能会经过冗余路径。其次，机械臂空闲也会产生影响。只有当目标位于各自的工作区域内时，所有机械臂才都有任务可做。如果水果分布不均，一个或多个机械臂可能会一直空闲，直到其他机械臂完成任务。在这种情况下，循环时间取决于果实分布的均匀程度。

为进一步提高机器人采摘机的生产率，需要采用第三类协作模式。在这种协作模式中，各机械臂的工作空间存在重叠，需要在重叠区域内进行协作采摘，以避免相互干扰并提高效率。这种模式既不同于多臂并行作业，也不同于双臂协作方法，而是结合了上述两种模式的特点。其重点在于如何合理安排随机分布的多个目标果实的遍历顺序，以最大限度地提高采摘效率。Barnett等通过考虑任务划分和可达性，研究了一种多臂猕猴桃采摘机器人，以实现果实分布均匀并尽量缩短任务完成时间 [25], [26]。Mann等从机械臂数量、机械臂能力和机器人速度等方面研究了最优机器人，并为每个机械臂分配任务，以实现最大采收量 [27]。Li等提出了一种多臂苹果采摘机器人的任务规划方法，为每个机械臂规划采摘顺序，以减少机械臂作业空闲和碰撞，并声称采摘效率最高可提高4.25倍 [28]。

III. 多臂苹果采摘机器人

如图1所示，本文研究的多臂收获机器人面向具有简单、狭窄、可接近且高产（SNAP）结果墙树冠结构的标准苹果园而设计。该机器人由轮式移动平台、四个笛卡尔机器人手臂、立体视觉单元、果实输送系统、控制系统和电源系统组成。视觉感知系统包括4个RGBD相机，用于采集目标果实的视觉信息，随后将这些信息传输至边缘计算模块进行果实定位。控制系统用于计算、图像处理和人机交互，包括一台计算机、两个边缘计算设备和一个屏幕。操作员可以监控机器人的运行状态，例如已采摘水果的数量、作业时间、电池使用情况和故障诊断结果。输送系统包括果实储存箱和三台传送带，用于将上、下方输送机输送的水果汇集至果实储存箱。多臂系统的整体结构如图2所示。机械臂的工作空间在共享线性导轨处重叠，形成图2(a)所示的分区。图2(a)展示了四臂示意图（编号为1至4）以及共享导轨，其编号分别为1&4、2&3、1&2和3&4。图2(b)展示了划分为九个区域的工作空间轮廓，其中包括专属区域E1至E4，这些区域仅允许单个机械臂进入；以及公共区域OU至OC，这些区域允许多个机械臂进入。



(a) 线性导轨和机械臂　　(b) 工作空间划分

图2. 所提出的采摘机器人多臂机械结构与工作空间示意图。

为便于分析，我们将关节-1定义为垂直运动关节，使机器人能够上下移动。关节2定义为水平运动关节，使机器人能够左右移动。最后，将关节3定义为纵向运动关节，使机器人能够前后移动。这些定义为各个关节提供了清晰且准确的名称，可在整个分析过程中保持一致。根据机器人的机械结构，No.1和No.2归为组-U，No.3和No.4归为组-D。关节-1是耦合关节，用于同步内组机械臂的操作。本文中，机械臂的收获动作按顺序进行，如图3所示，包括接近（A）、伸展（E）、抓取（G）、回缩（R）和放置（P）。在收获过程中，接近（A）、伸展（E）和抓取（G）动作占用关节-1，而回缩（R）和放置（P）动作释放该关节。当某个机械臂执行R和P时，其内组相邻机械臂可以同时执行A-E-G。

此外，在采摘过程中，为确保采摘率，可能会多次尝试采摘水果目标；由于机械限制，相邻机械臂无法采摘彼此距离过近的目标。在上述约束条件下，任务规划的目标是为每个机械臂确定采摘顺序和水果目标分配方案，以优化整体作业时间。整体作业时间由最后完成任务的机械臂决定。综上所述，本研究旨在基于上述约束条件，为四臂确定采摘顺序和分配方案，以优化作业时间。

IV. 基于多智能体强化学习的多臂任务规划

A. 马尔可夫博弈

马尔可夫决策过程（MDP） [29] 是序贯决策的经典表示。在 MDP 的每个步骤 t中，系统的下一状态可以根据当前状态和动作计算得到，由一个元组（$\mathbf{S}$、$\mathbf{A}$ f、g、γ）定义。其中，$\mathbf{S}$ 表示系统状态的集合，$\mathbf{A}$ 表示允许动作的集合，f：$s_k \times a_k \rightarrow s_{k+1}$ 是状态转移函数，g：$s_k \times a_k \rightarrow r_k$ 是奖励函数，并且

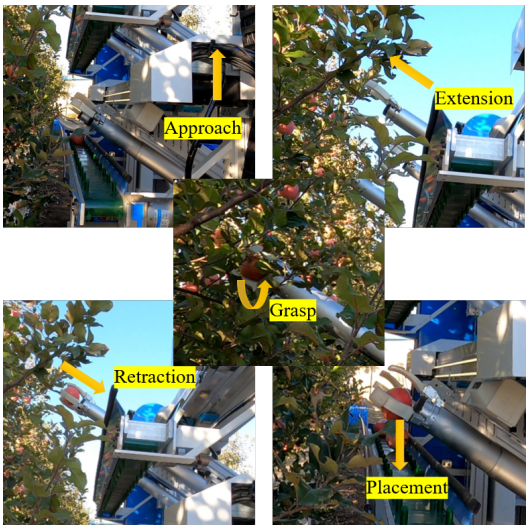


图 3. 采收作业中机械臂的五个连续动作。

$\gamma \in [0,1]$ 是折扣因子，用于补偿即时奖励和未来奖励的影响。

然而，当涉及多个智能体时，环境的状态动态会受到其他智能体的影响，因此，马尔可夫决策过程已不足以对此类问题进行建模。在这种情况下，可使用马尔可夫博弈（MG）框架，该框架由一个元组（N，$\mathbf{S}$ $\{\mathbf{A}^i\}_{i\in N}$，$f$，$\{g^i\}_{i\in N}$，$\gamma$）定义，其中 $N = 1, 2, 3, ..., N$为 $N > 1$ 个智能体的集合；$\mathbf{S}$ f 和 γ 的定义与马尔可夫决策过程中相同；$\{\mathbf{A}^i\}_{i\in N}$ 表示第 i个智能体的动作空间；$\mathbf{A} := A^1 \times A^2 \times A^3 \times \cdots A^N$为联合动作空间。奖励函数 $\{g^i\}_{i\in N}$ 表示第 i个智能体获得的即时奖励。

在本研究中，我们考虑一个由两个智能体（即D组和U组）组成的多智能体博弈，它们控制分为两组的四个机械臂。系统状态 $\mathbf{S}$、动作集 $\{\mathbf{A}^i\}_{i\in N}$ 以及奖励函数 $\{g^i\}_{i\in N}$ 将在以下小节中介绍。

B. 系统状态

在时刻 k，所提出的任务规划决策模型的状态可表示为集合$\mathbf{s}_k$ $=\mathbf{D}_k,\mathbf{P}_k,\mathbf{\Psi}_k,\mathbf{\Phi}_k,\mathbf{O}_k$。其中，$\mathbf{s}_k \in \mathbf{S}$，$\mathbf{D}_k \in \mathbb{R}^{N\times 3}$ 表示果实分布，$\mathbf{P}_k \in \mathbb{R}^M$ 表示机械臂位置和状态，其中机械臂的坐标相对于基坐标系给出，机械臂状态定义为 0（位于 AEG）或 1（位于 RP）。$\mathbf{\Psi}_k \in \mathbb{N}^{N\times M}$ 表示分配状态，其中第 n行的第 m个元素表示第 n个水果是否分配给第 m个机械臂（0：“否”，1：“是”）。$\mathbf{\Phi}_k \in \mathbb{N}^{N\times M}$ 表示尝试次数，其中第 n行的第 m个元素表示第 m个机械臂尝试采摘第 n个水果的次数（0：“尚未采摘”，1-3：尝试次数）。最后，$\mathbf{O}_k \in \mathbb{N}^{N\times M}$ 表示目标水果是否已采摘，其中第n行的第 m个元素表示第 n个水果是否已由第 m个机械臂采摘（0：“否”，1：“是”）。

C. 动作定义

在时刻 k，智能体-i 的动作可表示为向量 $\mathbf{a}_k^i = \Psi_k^i, B_{k,1}^i, B_{k,2}^i$。其中，$\mathbf{a}_k^i \in \mathbf{A}^i$，$\Psi_k^i \in 0, 1, \ldots, N$ 表示智能体-i 将采摘的水果数量，其中 0 表示没有新的目标，1∼N 表示与$\mathbf{D}_k$一致的待采摘水果数量。$B_{k,1}^i, B_{k,2}^i \in 0,1$ 分别表示组内左臂和右臂的动作。0 和 1 分别对应 AEG 和 RP 动作，1 也表示在放置状态停止。

如第三节所述，关节-1是一个耦合关节，因此有必要将组内两条机械臂的动作分为 AEG 和 RP，以避免空闲时间。为了更好地理解不同动作下的状态转移，我们在表 I 中列出了三种动作组合（0, 1）、（1,0）和（1,1）对应的 9 种状态转移。为便于分析，我们省略 Ψ_k^i。

例如，如果 U组的动作为 10，0，1，这意味着 U组将使用左臂通过 A-E-G 动作采摘 10 号水果，而右臂将使用 R-P 动作将水果放置到传送带上，或保持在放置状态。如果 U组的动作为 0，1，1，这意味着 U组不会采摘任何水果，左右臂都将保持在放置状态或执行 R-P 动作。

表 I 组内两条机械臂的状态转移。

转移	s_k	s_{k+1}	动作	描述
$\mathcal{T}_1$	AEG，RP	RP，RP	1, 1	暂停
$\mathcal{T}_2$	AEG、RP	RP、AEG	1, 0	交替
$\mathcal{T}_3$	AEG、RP	AEG、RP	0, 1	非交替
$\mathcal{T}_4$	RP, AEG	AEG, RP	0, 1	交替
$\mathcal{T}_5$	RP, AEG	RP, AEG	1, 0	非交替
$\mathcal{T}_6$	RP，AEG	RP，RP	1, 1	暂停
$\mathcal{T}_7$	RP，RP	RP，RP	1, 1	暂停
$\mathcal{T}_8$	RP，RP	AEG，RP	0, 1	重新启动
$\mathcal{T}_9$	RP，RP	RP，AEG	1, 0	重新开始

D. 奖励定义

该模型中的奖励由三个方面组成：

(1) 探索奖励，记为 r_{exp}^i。当智能体探索一个新的目标时，它会获得+0.05的正向奖励，该奖励从$\mathbf{D}_t$中观测得到。如果智能体没有探索新的目标，则 $r_{\text{exp}}^i = 0$。

(2) 冲突惩罚，记为 r_{confl}^i。当两个智能体发生碰撞时，它们都会获得 -0.1 的负奖励。否则，$r_{\text{confl}}^i = 0$。请注意，我们只考虑组间冲突，因为同一组内的冲突不会影响到达目标所需的时间。

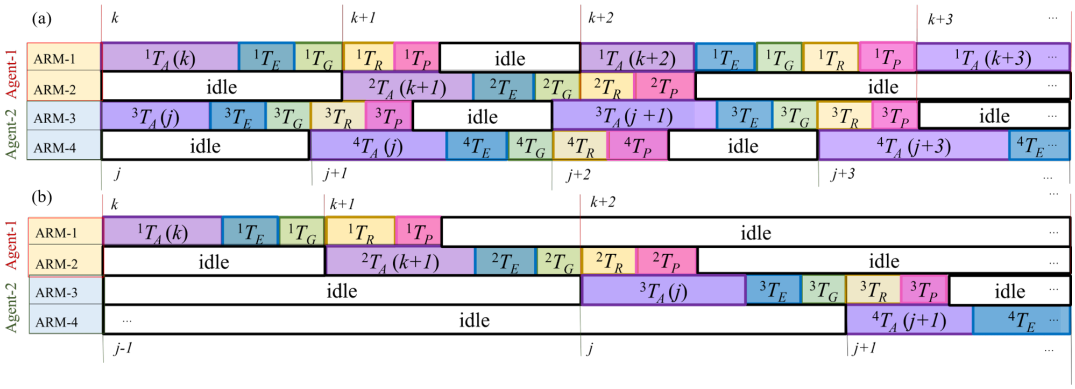


图4。 多臂协作的状态转移 作。(a) 展示了并行操作情况下的转移；(b) 说明了交替 on.

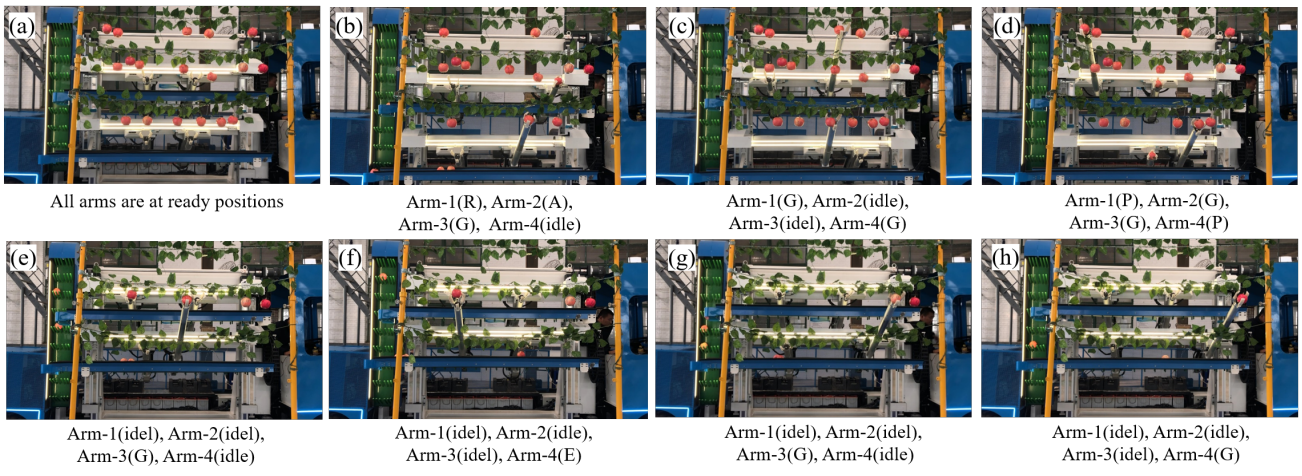


图5。多臂收获机器人状态转移序列的示例。

(3) 时间，包括移动和执行动作所需的时间。对于智能体 i，从状态 s_k 到状态s_{k+1} 的时间成本由以下公式确定：

$$ {}^{i}t_k^{k+1} = \begin{cases} {}^{i}t_{\text{AEG}}, & \text{if } \mathcal{T}_k^{k+1} = \mathcal{T}_2, \mathcal{T}_4, \mathcal{T}_8, \mathcal{T}_9 \\ {}^{i}t_{\text{all}}, & \text{if } \mathcal{T}_k^{k+1} = \mathcal{T}_3, \mathcal{T}_5 \\ {}^{i}t_{\text{idle}}, & \text{if } \mathcal{T}_k^{k+1} = \mathcal{T}_1, \mathcal{T}_6, \mathcal{T}_7 \end{cases} \tag{1} $$

其中，${}^{i}t_{\text{AEG}}$ 和 ${}^{i}t_{\text{all}}$ 分别表示智能体 i 执行 A-E-G 动作和完整动作所需的时长。${}^{i}t_{\text{idle}}$ 表示智能体等待另一机械臂时的空闲时间；发生智能体间干扰时，该时间取决于另一智能体执行动作的时长。为进一步阐明状态和动作的转移，我们在图4中给出了一个示例。具体而言，图 4 (a) 展示了 $\mathcal{T}_2$，其中从状态 k 到状态$k+1$ 的转移表示机械臂-1已完成作业并切换至机械臂-2，时间段记为 ${}^{1}t_k^{k+1} = {}^{1}t_{\text{AEG}}$。类似地，图4 (a) 中从状态 $k+1$ 到状态 $k+2$ 的转移展示了 $\mathcal{T}_4$，而图4 (b) 中从状态 j 到状态 $j+1$ 的转移展示了 $\mathcal{T}_8$。

接下来，图4 (a) 展示了 $\mathcal{T}_3$，其中从状态 $k+2$ 到状态 $k+3$ 的转移表示机械臂-1已完成上一个作业，并继续执行下一个 AEG 动作，其时长由 ${}^{1}t_{\text{all}} = {}^{1}t_{\text{AEG}} + {}^{1}t_{\text{RP}}$计算得出。机械臂-2的情况相同。同时，图4 (b) 中从状态 $k+2$ 到状态 $k+3$ 的转移展示了 $\mathcal{T}_6$，这表示机械臂-2已完成上一个动作并处于空闲状态，其时间成本按另一智能体的作业时间计算，即 ${}^{1}t_{\text{idle}} = {}^{2}t_k^{k+1}$。类似地，$\mathcal{T}_1$ 表示机械臂-1已完成上一个动作并处于空闲状态。最后，图4 (b) 中从状态$j-1$ 到状态 $j+1$ 的转移展示了 $\mathcal{T}_7$，这表示两条机械臂均处于空闲状态，持续时间取决于另一智能体的作业时间。如果一个智能体没有目标可前往，而另一个智能体仍有目标，则该智能体将保持状态转移 $\mathcal{T}_7$，等待另一智能体完成任务。这样，两个智能体最终可以获得相同的累积总奖励。图5展示了收割作业中的状态转移示例。

根据式（1）定义的时间成本，动作 r_{time}^{i} 的奖励如下

$$ r_{\text{time}}^{i} = \alpha \left(e^{-{}^{i}t_k^{k+1}} - 1 \right) \tag{2} $$

其中，r_{time}^{i} 表示奖励；α 为比例因子。式（2）表示奖励与 ${}^{i}t_k^{k+1}$ 负相关。

基于上述定义，奖励函数定义如下

$$ r^{i}(s_k, a_k^{i}) = \begin{cases} -50, & \text{if } k = k_{\max}, \\ 100, & \text{if all targets picked}, \\ r_{\text{exp}}^{i} + r_{\text{confl}}^{i} + r_{\text{time}}^{i}, & \text{otherwise.} \end{cases} \tag{3} $$

E. 多智能体强化学习

协作式多智能体强化学习问题需要在智能体之间设计良好的协作策略。一种直接的方法是采用中央控制器，接收所有智能体的奖励并为每个智能体确定动作，这种方式称为完全集中式。由于控制器可以获取所有智能体的信息，因此本文中的问题可建模为马尔可夫决策过程，并使用单智能体强化学习算法求解。本文中，两组机械臂由中央控制器控制，且所有信息均可获取。因此，本文采用完全集中式策略，使单智能体强化学习算法能够轻松解决该问题。多机械臂采摘马尔可夫博弈环境涉及臂间耦合、组间机械臂的交替以及机械限制，使得转移动态变得复杂。考虑到本研究中状态空间的高维性以及环境中存在的不确定性，我们因此采用基于策略的学习算法 PPO [30]。

PPO 是一种无模型、在线、同策略的策略梯度强化学习算法，它使用多个随机梯度下降步骤的小批量数据来优化目标函数。在本研究中，我们使用 stable-baseline 3 [31] 实现 PPO。本研究中的 PPO 实现得益于一个设计完善且高效的框架，该框架能够轻松训练复杂的深度强化学习模型。为训练 PPO 网络，衰减因子设为 $\gamma = 0.95$，裁剪超参数设为$\epsilon = 0.20$。我们使用优化器 Adam [32]，学习率为 5e-4，并使用广义优势估计来估算优势，其中 $\lambda = 0.88$。按照算法要求，训练以批次为单位进行，因此我们使用同策略批量回放内存。训练轮数为 8，小批量大小为 512。为提高模型的泛化能力，我们采用循序渐进的训练方法。首先，我们在包含 60 个目标的环境中训练一个简单的预训练模型，共进行 100 万步训练。接着，我们让模型在布局和目标数量各不相同的 10 个环境中继续训练，额外进行 100 万步训练。最后，在训练期间每次重置时随机改变布局和目标数量，进一步提升模型的适应能力；经过 1 亿步训练后，得到最终的策略网络模型。

V. 仿真与实验

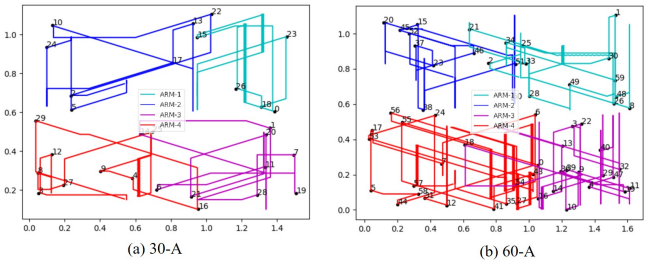
为验证所提出方法的有效性，我们结合仿真与实验进行了验证。仿真基于由 NVIDIA Omniverse 驱动的虚拟环境，实验则在室内实验室和苹果园中进行，如图 6 所示。

在作业过程中，仿真考虑了两种情形：一种是认为任务执行是确定的，即机器人按照规划依次采摘水果；另一种是由于采摘失败导致执行具有不确定性，因此部分目标需要尝试采摘多次。这两种情形对任务完成结果的影响截然不同。



图 6. 基于 NVIDIA Omniverse 的果园采摘仿真环境及多臂采摘机工作空间示意图。

考虑到作业点的水果数量和分布各不相同，我们采用两种不同布局（分别包含 30 个和 60 个水果）评估了性能。我们通过所提方法和文献 [28]提出的基于启发式的任务规划算法，开展了四组测试：分别针对 30 个和 60 个水果，以及分布不同的 A、B 两种情况。每组测试均重复五次，每次试验中的水果分布保持不变。在没有抓取失败的第一种情形下，表 II 展示了所提方法与参考方法 [28] 在五次重复实验中的比较结果。从平均总采摘时间来看，分别缩短了 3.68%（30-A）、17.56%（30-B）、5.35%（60-A）和 -1.59%（60-B）。尽管该方法与参考方法在平均采摘时长方面的优势并不显著，但所提方法显著缩短了计算时间，并提升了规划时间方面的实时性能。图 7 展示了 30-A 和 60-A 测试组中四臂的遍历路径。



(a) 30-A　　(b) 60-A

图 7. 四个机械臂在不同水果分布下的遍历路径。

在第二种抓取失败的情形中，我们考虑到有些水果无法一次抓取成功，需要尝试两次或三次。使用参考方法 [28], 进行仿真时，任务仅根据初始条件规划一次，所有经过的目标均被视为已成功抓取。执行过程中，抓取失败后不会重新规划任务。然而，使用所提方法时，任务规划会依据环境和机器人的当前状态进行，采摘失败的果实仍会继续尝试。在五次重复测试中，30-A 组中有 7 个水果需要采摘两次，另有 2 个水果需要采摘三次。30-B 组中有 8 个水果需要尝试 2 次，另有 4 个水果需要尝试 3 次。60-A 组中有 10 个水果需要尝试 2 次，另有 5 个需要尝试 3 次；60-B 组中有 12 个水果需要尝试 2 次，另有 3 个需要尝试 3 次。表 III 展示了抓取失败情况下的采摘实验结果。本文所提方法增加了总作业时间，但显著减少了剩余水果数量，提高了作业质量。

表 II 无抓取失败情况下的多臂采摘任务规划仿真结果。

Test 组	方法	最大值（秒）	最小值（秒）	平均值（秒）	规划时间（秒）
30-A	[28]	217.60	190.00	200.78	8.25
	ours	**201.40**	**177.35**	**193.40**	**1.21**
30-B	[28]	321.25	275.00	287.43	8.45
	ours	**241.05**	**231.30**	**236.93**	**1.18**
60-A	[28]	461.35	394.50	435.96	15.87
	ours	**434.45**	**387.78**	**412.64**	**1.24**
60-B	[28]	549.00	**436.65**	**490.54**	16.42
	ours	**536.20**	443.70	498.36	**1.14**

注：表中，最大值、最小值和平均值分别表示五次重复试验中采收作业时长的最大值、最小值和平均值。规划时间是指计算规划结果所耗费的时间。

在实际果园实验中，我们在一个标准苹果园内选取了三排树冠状况及水果数量相近的苹果树，如图 8(b) 和 (c) 所示，并开展了果园实验。在每排苹果树中，选取约 30 个作业点进行三组测试：第 1 条线、第 2 条线和第 3条线分别采用随机遍历方法、参考方法[28], 和所提方法。随机规划方法是指不使用任何任务规划方法，而是为每条机械臂随机分配目标。

如表 IV 所示，可达果实占总数的比例不高，约为 40%。这是因为该机器人能够采摘的最低位置为距地面 1.5 m，而实验中相当数量的水果都低于 1.5 m；此外，还有一些水果被

表 III 存在抓取失败时的多臂采摘任务规划仿真结果。

测试	方法	平均时间（秒）	剩余果实 I	II	III	IV	V
30-A	[28]	159.27	9	9	9	9	9
	ours	237.70	1	0	1	0	0
30-B	[28]	164.76	12	12	12	12	12
	ours	259.54	1	3	2	1	1
60-A	[28]	339.03	15	15	15	15	15
	ours	434.86	4	3	1	1	4
60-B	[28]	329.32	15	15	15	15	15
	ours	479.26	5	4	4	3	5

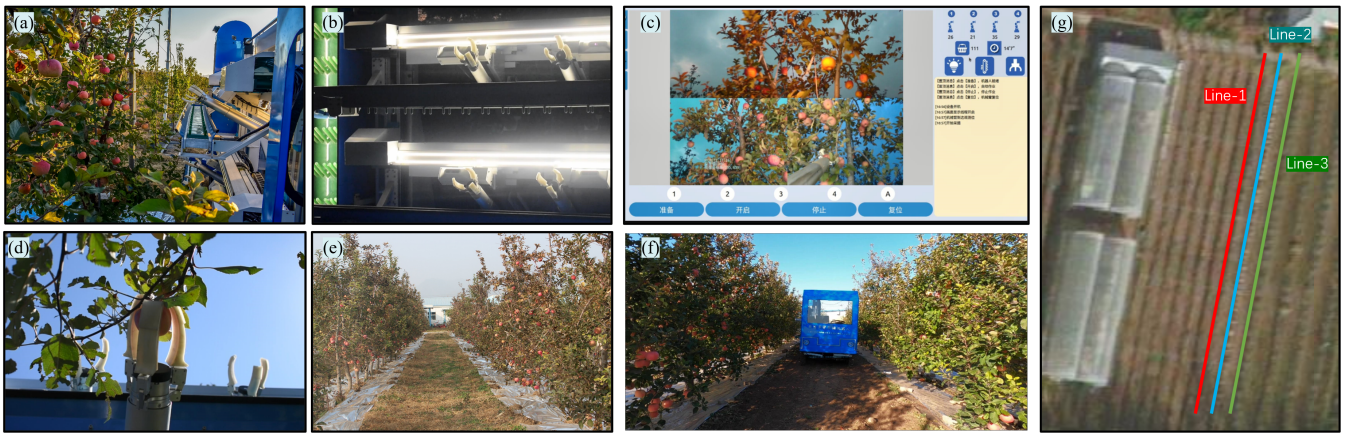


图 8. 实验中的四臂采摘机器人和苹果园。(a) 作业现场的多臂收获机器人；(b) 照明系统正在工作；(c) 机器人的软件平台；(d) 夹爪正在抓取一个苹果目标；(e) 果树行间的通道；(f) 机器人沿通道移动；(g) 果园实验中使用的树行。

表 IV 苹果园全田间试验中的采摘性能

组	作业点	苹果	水果到达率-能力（%）	已采摘	已采摘 相对于可达	Lost	平均值周期时间 (s)	方法
第1条线	30	453	186(41.06%)	149(32.89%)	80.11%	37	7	随机
第2条线	35	420	176（41.90%）	120（28.57%）	68.18%	56	6.1	[28]
第3条线	28	382	145(37.96%)	115(30.10%)	79.31%	30	5.8	Ours

树枝遮挡，难以接近。表 IV 中的实验结果表明：1）使用任务规划算法的实验组，即参考方法和所提方法，在各作业现场的平均作业时间显著短于随机遍历方法，说明任务规划能够有效缩短作业时间。2）对比方法的已采摘果实占可达果实的比例较低，仅为 68.18%，显著低于另外两种方法。这表明离线规划方法缺少任务动态更新，难以应对果实采摘失败的情况。3）所提方法的平均周期时间显著短于随机遍历方法，且略短于对比方法。在本次实验中，受可达范围限制，每个作业现场的水果目标实际平均约为 15 个；当水果数量较少时，该方法在作业时间方面的优势无法充分体现。不过，该方法在采收率方面展现出优势，损失的水果显著少于对比方法。总体而言，果园实验表明，采用本文所提方法的四臂采摘机器人采摘成功率约为 79.31%，平均每个水果的采摘时间为 5.8 s。

VI. 结论与未来工作

本研究提出了一种面向苹果园四臂采摘机器人的新型任务规划策略，以提升运行效率。为实现机械臂之间的有效协作，研究考虑了机器人的多种约束，包括机械臂间冲突、关节耦合、机械限制和抓取失败。研究采用马尔可夫博弈框架对四臂协同采摘问题进行建模。为支持多机械臂实时决策并降低计算复杂度，研究提出了一种基于多智能体强化学习的任务规划方法。通过使用四种不同水果布局进行仿真，验证了所提方法的有效性。结果表明，在规划效果方面，所提方法与启发式算法相当，同时规划效率显著提高，尤其是在水果数量较多时。相比之下，启发式方法需要约 8 秒（30 个水果）或 16 秒（60 个水果），而任务规划方案约 1 秒即可得出。此外，考虑到抓取失败的可能性，所提方法仍表现出令人满意的采摘完成率和实时规划性能。果园研究的实验结果证明了所提方法的实际适用性，其损失的水果更少，平均采摘周期也更短。

实验评估发现，机器人的可达工作范围有限，导致水果可达率约为 40%。为了提高机器人采收过程的效率，未来需要优化机器人的结构设计，以提高对树冠下部水果的可达性。这一局限将成为我们未来研究的重点。

努力。

参考文献

[1] B. Zou、F. Liu 和 Z. Zhang，“山地果园机械化：发展瓶颈与国外经验”，《农业机械化研究》，第 9 期，第 47 页，2019年。[2] A. Gongal、S. Amatya、M. Karkee、Q. Zhang 和 K. Lewis，“水果检测与定位的传感器和系统：综述”，《农业中的计算机与电子学》，第 116 卷，第 8–19 页，2015 年。[3] W. S. Kim、D. H. Lee、Y. J. Kim、T. Kim 和 H. J. Lee，“使用基于图像块的卷积神经网络检测果园中的自主行驶路径”，《农业中的计算机与电子学》，第 175 卷，第 105620 页，2020年。[4] Y. Jin、J. Liu、Z. Xu、S. Yuan、P. Li 和 J. Wang，“农业机器人技术的发展现状与趋势”，《国际农业与生物工程杂志》，第 14 卷，第 19 期，2021年4月，页码 1–。[5] G. Kootstra、X. Wang、P. M. Blok、J. Hemming 和 E. Van Henten，“选择性采收机器人技术：当前研究、趋势与未来方向”，《当前机器人学报告》，第 2 卷，第 1 期，第 95–104 页，2021年。[6] H. Zhou、X. Wang、W. Au、H. Kang 和 C. Chen，“用于水果采摘的智能机器人：最新进展与未来挑战”，《精准农业》，第 1–52 页，2022年。[7] J. Rong、P. Wang、T. Wang、L. Hu 和 T. Yuan，“番茄采收机器人的水果姿态识别与定向有序抓取策略”，《农业中的计算机与电子学》，第 202 卷，第 107430 页，2022年。[8] F. Kurbah、S. Marwein、T. Marngar 和 B. K. Sarkar，“ 菠萝采收机器人夹爪的设计与开发 ”，载于《机器人系统通信与控制》。施普林格，2022年，第 437–454 页。[9] H. A. Williams、M. H. Jones、M. Nejati、M. J. Seabright、J. Bell、N. D. Penhall、J. J. Barnett、M. D. Duke、A. J. Scarfe、H. S. Ahn 等，“利用机器视觉、卷积神经网络和机械臂进行猕猴桃机器人采收”，《生物系统工程》，第 181 卷，第 140–156 页，2019年。[10] Y. Xiong、Y. Ge、L. Grimstad 和 P. J. From，“自主草莓采收机器人的设计、开发、集成与田间评估”，《野外机器人学期刊》，第 37 卷，第 2 期，第 202–224 页，2020 年。[11] X. Ling 、 Y. Zhao 、 L. Gong 、 C. Liu 和 T. Wang，“利用双目视觉实现番茄机器人采收的双臂协作”，《机器人与自主系统》，2019年。[12] J. H. Tibbetts，“农业变革：新技术、整合或可带来产量增长与就业动荡”，《生物科学》，第 69 卷，第 4 期，第 237–243 页，2019年。[13] Y. Li、Q. Feng、T. Li、F. Xie、C. Liu 和 Z. Xiong，“鲜果机器人采收的目标视觉信息获取技术进展：综述”，《农学》，第 12 卷，第 6 期，第 1336 页，2022年。[14] P. Kurtser 和 Y. Edan，“采收机器人的任务序列规划”，《机器人与自主系统》，第 131 卷，第 103591 页，2020年。[15] C. W. Bac、E. J. Van Henten、J. Hemming 和 Y. Edan，“高价值作物采收机器人：最新进展综述与未来挑战”，《野外机器人学期刊》，第 31 卷，第 6 期，第 888–911 页，2014年。[16] J. Li、M. Karkee、Q. Zhang、K. Xiao 和 T. Feng，“机器人采收的苹果采摘模式特征分析”，《农业中的计算机与电子学》，第 127 卷，第 633–640 页，2016 年。

[17] X. Wang、H. Kang、H. Zhou、W. Au 和 C. Chen，《苹果园机器人采收的几何感知果实抓取估计》，《农业中的计算机与电子学》，第193卷，第106716页，2022年。[18] X. Ling、Y. Zhao、L. Gong、C. Liu 和 T. Wang，《利用双目视觉进行双臂协作与番茄机器人采收实施》，《机器人与自主系统》，第114卷，143，2019年，134页。–。[19] D. Sep´uLveda、R. Fern´andez、E. Navas、M. Armada 和 P. Gonz´alez-De-Santos，《使用双臂操作进行机器人茄子采收》，*IEEE Access*，第8卷，第121889–121904页，2020年。[20] C. J. Hohimer、H. Wang、S. Bhusal、J. Miller、C. Mo 和 M. Karkee，《采用3D打印软体机器人末端执行器的苹果机器人采收系统设计与田间评估》，*ASABE*学报，第62卷，第2期，第405–414页，2019年。[21] R. Bogue，《水果采摘机器人：它们的时代到来了吗？》，《工业机器人：机器人研究与应用国际期刊》，第47卷，第2期，第141–145页，2020年。[22] T. Delbridge 等，《机器人草莓采收前景良好，但需要改进技术并提高工资水平才能实现经济可行性》，《加利福尼亚农业》，第75卷，第2期，第57–63页，2021年。[23] H. Williams、C. Ting、M. Nejati、M. H. Jones、N. Penhall、J. Lim、M. Seabright、J. Bell、H. S. Ahn、A. Scarfe 等，《猕猴桃采收机器人的改进与大规模评估》，《野外机器人学期刊》，第37卷，第2期，第187–201页，2020年。[24] O. Ringdahl、P. Kurtser 和 Y. Edan，《采收机器人接近策略评估：甜椒采收案例研究：类别：（5）》，《智能与机器人系统期刊》，第95卷，第1期，第149–164页，2019年。[25] J. Barnett、M. Duke、C. K. Au 和 S. H. Lim，《用于猕猴桃采收的多笛卡尔机器人臂工作分配》，《农业中的计算机与电子学》，第169卷，第105202页，2020年。[26] C. Au、J. Barnett、S. H. Lim 和 M. Duke，《猕猴桃采收笛卡尔机器人系统的工作空间分析》，《工业机器人：机器人研究与应用国际期刊》，第47卷，第4期，第503–510页，2020年。[27] M. P. Mann、B. Zion、I. Shmulevich、D. Rubinstein 和 R. Linker，《多机械臂笛卡尔水果采收机器人组合优化与性能分析——图着色的扩展》，《智能与机器人系统期刊》，第82卷，411，2016年3月，第399页。第1期。[28] Q. Li Tao、Z. Quan、X. Chunjiang 和 Feng，《高密度矮化果园多机械臂采收机器人的任务规划》，《农业工程学报》，第37卷，第2期，第1–10页，2021年。[29] M. L. Puterman，《马尔可夫决策过程》，《运筹学与管理科学手册》，第2卷，第331–434页，1990年。[30] J. Schulman、F. Wolski、P. Dhariwal、A. Radford 和 O. Klimov，《近端策略优化算法》，*arXiv*，第abs/1707.06347卷，2017年。[31] A. Raffin、A. Hill、A. Gleave、A. Kanervisto、M. Ernestus 和 N. Dormann，《Stable-Baselines3：可靠的强化学习实现》，《机器学习研究期刊》，第22卷，第1期，第12348–12355页，2021年。[32] D. P. Kingma 和 J. Ba，《Adam：一种随机优化方法》，*arXiv*预印本 *arXiv:1412.6980*，2014年。