doi:10.6041/j.issn.1000-1298.2026.13.004

基于深度强化学习的茶芽采摘三维路径规划方法

桂志勇[1]　梅相贞[2]　陈建能[1,3]　俞焘杰[1]　贾江鸣[1,3]　贺磊盈[1]　武传宇[1]

(1.浙江理工大学机械工程学院,杭州 310018;2.建德市峰鼎茶业有限公司,杭州 311614;
3.全省农业智能感知与机器人重点实验室,杭州 310018)

摘要:路径规划是决定自动化茶芽采摘机器人采摘效率的核心决策环节。针对自动茶芽采摘路径规划中三维空间建模困难、实时响应要求高以及对动态节点规模适应性差等问题,将茶芽采摘任务建模为三维旅行商问题,并提出了一种基于深度强化学习的茶芽采摘三维路径规划方法。本研究采用零填充策略构建了参数化的三维茶芽数据集,设计了动态掩码机制以确保路径的有效性,引入了长短期记忆模块来动态融合历史路径信息,开发了约束多头自注意力机制以提高决策准确性,并基于演员-评论家(Actor-Critic)框架稳定模型训练。对比试验表明,所提出的算法与精确的 Concorde 算法之间的相对最优性差距控制在 2% 以内,而在处理 80 个茶叶芽的情况下计算时间仅为 Concorde 算法的 1.7%。该算法优于经典的启发式算法和其他基于深度强化学习的算法,在实际茶园场景中表现稳定,为采茶机器人路径规划提供了高效方案,可迁移至其他精准采摘任务,也为三维路径规划提供了理论基础。

关键词:茶芽采摘;三维路径规划;深度强化学习;注意力机制;动态掩码机制

中图分类号:S126　　**文献标识码**:A　　**文章编号**:1000-1298(2026)13-0068-11　　**OSID**:

Deep Reinforcement Learning-based Three-dimensional Path Planning Algorithm for Tea Bud Picking

Gui Zhiyong[1]　Mei Xiangzhen[2]　Chen Jianneng[1,3]　Yu Taojie[1]　Jia Jiangming[1,3]　He Leiying[1]　Wu Chuanyu[1]

(1. *School of Mechanical Engineering*, *Zhejiang Sci-Tech University*, *Hangzhou* 310018, *China*
2. *Fengding Tea Industry Co.*, *Ltd.*, *Hangzhou* 311614, *China*
3. *Zhejiang Key Laboratory of Intelligent Sensing and Robotics for Agriculture*, *Hangzhou* 310018, *China*)

Abstract: Picking path planning is the core decision-making link that determines the harvesting efficiency of automated tea bud picking robots. Aiming at the problems of difficult three-dimensional spatial modeling, high real-time response requirements and poor adaptability to dynamic node scales in automatic tea bud picking path planning, the tea bud picking task was modelled as a three-dimensional traveling salesman problem (3D-TSP), and a deep reinforcement learning-based 3D path planning algorithm was proposed for tea bud picking. A parametric 3D tea bud dataset with a zero-padding strategy was constructed, a dynamic masking mechanism was designed to ensure path validity, a long short-term memory (LSTM) module was introduced to dynamically fuse historical path information, a constrained multi-head self-attention mechanism was developed to improve decision-making accuracy, and the stability model training was done based on the Actor-Critic framework. Comparative experiments showed that the relative optimality gap between the proposed algorithm and the exact Concorde algorithm was controlled within 2%, and its computation time in the scenario with 80 tea buds was only 1.7% of that of Concorde. The algorithm outperformed classical heuristic algorithms and other deep reinforcement learning-based algorithms, maintained stable performance in real tea garden scenarios, and provided an efficient solution for tea picking robot path planning.

Key words: tea bud picking; three-dimensional path planning; deep reinforcement learning; attention mechanism; dynamic masking mechanism

收稿日期:2026-03-09　修回日期:2026-04-10

基金项目:国家自然科学区域创新联合基金重点项目(U23A20175)和国家自然科学基金项目(32472009)

作者简介:桂志勇(2000—),男,博士生,主要从事智能采摘关键技术研究,E-mail:zhiyonggui_zstu@163.com

通信作者:武传宇(1976—),男,教授,博士生导师,主要从事农业装备及其智能化研究,E-mail:cywu@zstu.edu.cn

0　引言

自动化茶芽采摘路径规划是决定采收效率的核心环节，其目标是为采摘末端执行器生成访问离散茶芽节点的最优空间序列，该问题从组合优化视角可建模为三维旅行商问题（3D－TSP），求解质量直接影响采茶机器人的作业效率与能耗水平[1]。随着茶园规模化生产对采收自动化的需求日益迫切，开发兼具高效性与自适应性的 3D－TSP 求解方法，已成为突破采茶机器人作业效能瓶颈的关键技术方向。

现有 3D－TSP 求解方法难以适配茶芽采摘场景的实际需求，各类型算法均存在显著局限性。经典精确算法如 Held－Karp 动态规划法[2]、分支定界法[3]虽能保证全局最优解，但计算复杂度随节点数量呈超多项式级激增，计算耗时过长，无法满足机器人动态作业的实时性要求；启发式算法[4－7]通过牺牲部分最优性降低了计算成本，却高度依赖手工设计参数。尽管部分研究尝试优化参数适应性[8－9]，例如 Xu 等[10]利用蚁群算法规划二维茶芽采摘顺序，需多组试验筛选最优参数，李涛等[11]基于遗传算法优化采摘路径，仍难以应对茶芽数量的动态波动，且这类算法处理 50 个节点仍需秒级耗时，动态适配性与效率均不足。

近年来，神经网络为组合优化问题求解提供了新方向[12]，但现有相关研究应用于茶芽采摘路径规划仍面临多重关键瓶颈，且多数聚焦二维 TSP 场景，难以直接迁移至三维任务。传统神经网络方面，Vinyals 等[13]提出的 RNN 指针模型虽提升了二维 TSP 求解效率，却受限于高质量标注数据集，泛化能力受限；Kool 等[14]基于 Transformer 构建的注意力模型，采用 REINFORCE 算法优化路径，耗时问题仍未得到解决。随着图神经网络的发展[15－19]，部分研究者将二维欧几里得 TSP 转化为图特征输入，结合强化学习训练神经网络实现路径规划[20－21]。Ma 等[22]提出的图指针网络虽提升了二维 TSP 解质量，但处理 50 个节点仍需 1.2 s 以上；Joshi 等[23]的深度图卷积网络虽实现中小规模二维 TSP 高优化率，却存在高耗时、依赖固定输入节点规模的缺陷，与实际茶园的动态节点需求严重脱节。整体而言，现有启发式算法与神经网络方法均无法同时满足茶芽采摘的毫秒级实时响应、动态节点规模适配与三维空间建模核心需求，难以有效应用于实际采摘场景。

针对上述问题，本研究提出一种基于深度强化学习的自适应茶芽采摘三维路径规划方法。该方法构建含 30～80 个茶芽的参数化三维模拟数据集，通过零填充实现动态节点规模适配；设计动态掩码机制确保路径合法性并缩小决策空间，结合特征编码-解码器与 LSTM 门控模块提取茶芽全局与局部三维特征，设计带约束多头自注意力机制提升节点决策精准性；基于 Actor－Critic 框架优化模型训练，降低策略梯度方差以保障训练稳定性。通过与精确算法、经典启发式算法及其他深度强化学习算法的多组对比试验，验证所提方法在解质量、实时响应能力与场景适应性上的优势，为采茶机器人路径规划提供高效解决方案。

1　茶芽采摘三维路径规划方法

1.1　茶芽采摘 3D－TSP 问题建模与环境构建

1.1.1　三维茶芽采摘路径优化问题的定义

本研究将茶芽采摘路径序列优化问题设定为三维空间路径规划问题，通过构建三维旅行商问题（3D－TSP）数学模型实现对实际采摘作业的精准建模。设采摘场景包含 N 个待采茶芽（$N\geqslant 1$），每个茶芽由三维坐标 (x_i,y_i,z_i)（$i\in(1,2,\cdots,N)$）作为特征标识。采摘机械臂需执行闭环路径规划任务：从固定原点 $O(0,0,1)$ 出发，访问全部 N 个茶芽节点（过程中每个茶芽节点仅可被访问一次），最终返回起始原点 O，形成完整采摘回路。

该优化问题的目标函数表述为：在三维空间内，寻求使总行程距离最小化的节点访问序列，其中总行程距离 L_{em} 计算公式为

$$L_{em}=\min\sum_{k=0}^{N}d((x_k,y_k,z_k),(x_{k+1},y_{k+1},z_{k+1})) \tag{1}$$

其中　$(x_0,y_0,z_0)=(x_{N+1},y_{N+1},z_{N+1})=(0,0,1)$

式中 $d((x_k,y_k,z_k),(x_{k+1},y_{k+1},z_{k+1}))$ 为路径序列中第 k 个与第 $k+1$ 个节点间的空间距离。

1.1.2　三维模拟数据集构建

本研究采用深度强化学习框架解决茶芽采摘路径序列规划问题，模型性能依赖模拟场景的质量。针对实际茶园中茶芽空间分布的随机性特征，本研究构建参数化三维模拟场景数据，以增强模型的现实场景适应能力；通过对实际茶园茶芽空间分布的数据收集与分析，确保数据生成的高仿真性。

本研究场景数据生成规则基于真实采摘场景的空间统计特性，在单位立方体空间内对茶芽进行三维随机采样，其中 z 轴坐标限定于 0.60～0.95 m 区间，以模拟茶芽顶端聚集分布规律；设置固定起止点 (0,0,1)，构成采摘路径闭环约束；每个样本包含 30～80 个随机生成的茶芽坐标点，当单一样本中茶芽实际生成数量不足 80 个时，采用零填充策略对剩

余特征维度进行标准化处理,以保证模型输入的固定维度要求。图 1 为数据集生成的场景示例。通过参数化生成机制,大批量生成兼具场景真实性与算法适应性的训练样本。

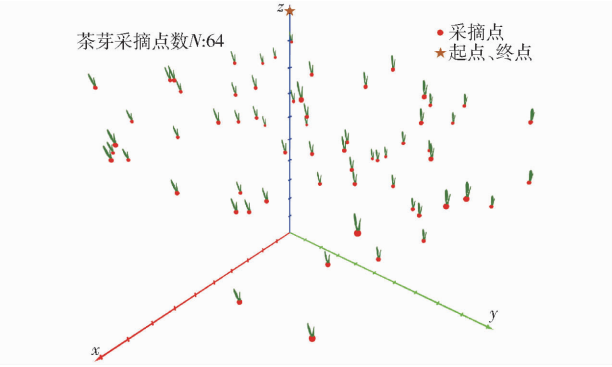

图 1 参数化模拟场景生成示例

Fig. 1 Demonstration of parametric simulation scenario generation

1.1.3 动态掩码机制与状态-动作空间设计

为确保茶芽采摘路径序列的合法性,规避节点重复访问、遗漏及误选无效坐标问题,本研究设计动态掩码机制约束节点访问条件,并耦合该机制定义适配 3D-TSP 的状态空间与动作空间。

如图 2 所示,动态掩码机制通过一维二进制张量编码实现节点访问状态的实时更新。初始化时构建与节点总数匹配的掩码张量 M,将起点掩码值置 1、其余茶芽节点、终点及零填充无效坐标置 0;路径规划过程中采用递归更新策略,已访问节点掩码值即时置 0,剩余未访问茶芽节点掩码值置 1,以形成动态可选集合;所有茶芽访问完成后,将终点掩码值置 1,强制形成闭环作业约束。该机制同时缩小模型决策空间,提升训练与决策效率。

状态空间表征为采摘场景中所有节点的三维坐标集合 $X=\{(x_i,y_i,z_i)\}(i=1,2,\cdots,N)$,涵盖待采

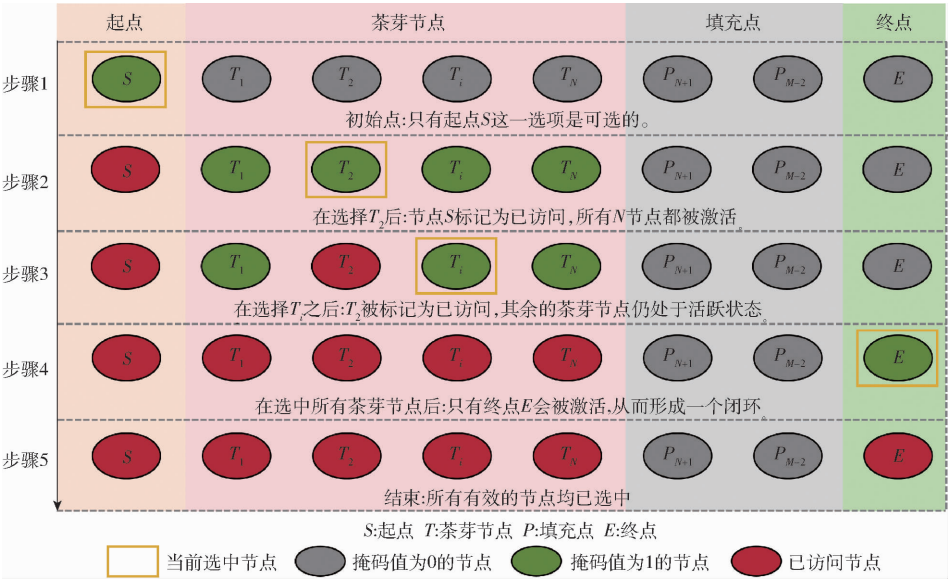

图 2 动态掩码机制执行逻辑图

Fig. 2 Diagram of execution logic for dynamic masking mechanism

茶芽、起点、终点及填充节点,为策略网络提供全局空间位置特征输入。动作空间遵循动态收缩机制,与动态掩码机制深度绑定,其维度随决策步骤实时变化并与当前掩码值为 1 的可选节点数一致:初始阶段仅起点可访问,动作空间维度为 1;每访问一个茶芽节点,动作空间维度同步递减 1;所有茶芽访问完成后,动作空间维度收缩至 1,仅保留终点为唯一可选动作,确保规划序列满足 3D-TSP 闭环路径要求。

1.2 基于 Actor-Critic 的深度强化学习路径规划模型

三维 TSP 属于非确定性多项式时间困难问题,传统算法难以兼顾求解质量与效率,深度强化学习具备端到端学习能力,可自动提取复杂环境特征、学习高效决策策略。本研究基于 Actor-Critic 框架构建路径规划模型,缓解策略梯度高方差问题,提升模型收敛速度与稳定性;同时融合特征编码-解码器、LSTM 门控模块与带约束多头自注意力机制,实现茶芽特征的高效提取与精准节点决策。

1.2.1 模型整体架构设计

如图 3 所示,模型整体架构分为 Actor 网络与 Critic 网络两大核心模块,二者协同实现路径规划与训练优化。

Actor 网络为策略网络,负责生成茶芽节点访问序列,通过特征编码-解码器提取茶芽全局空间特征与局部时序特征,LSTM 门控模块动态整合历史路

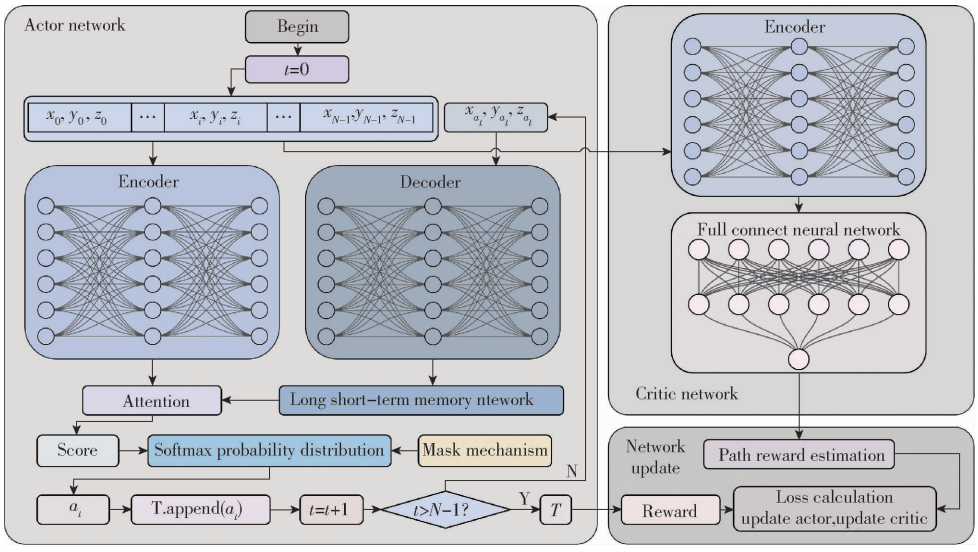

图 3　茶芽采摘路径规划的深度强化学习方法原理图

Fig. 3　Schematic of deep reinforcement learning method for tea bud picking path planning

径信息，带约束多头自注意力机制计算候选节点相关性分数，结合动态掩码机制筛选合法节点并输出选择概率分布，最终通过采样生成采摘路径。

Critic[24]网络为价值网络，负责状态价值估计，通过特征编码器提取全局特征，经全连接网络输出当前环境的路径奖励估计，作为 Actor 网络损失函数的基线，为 Actor 网络提供优化方向。

模型通过计算实际回报与 Critic 网络价值估计的偏差，更新 Actor 网络策略参数；以均方误差为损失，更新 Critic 网络价值估计参数，实现双网络的协同训练。

1.2.2　茶芽空间特征编码-解码器

茶芽的空间分布特征是路径规划的依据，高效的特征编码-解码器能够增强模型对三维空间结构的感知能力，为决策过程提供充足的空间关联信息。卷积神经网络（Convolutional neural network，CNN）凭借局部感受野与权值共享特性，在空间特征提取任务中表现出显著优势，相比全连接网络，其参数规模更小、训练效率更高，且对空间模式的泛化能力与拟合能力更强。

本研究设计 3 层一维卷积网络构建特征编码-解码器，用于捕捉茶芽在三维空间中的分布特征。特征编码器以所有茶芽的三维坐标集合 $\{(x_i, y_i, z_i)\}(i=1,2,\cdots,N)$ 为输入，通过多层卷积操作映射至高维特征空间，为每一步的选择提供全局视野。其公式为

$$\boldsymbol{H}=\mathrm{ReLU}(\mathrm{Conv1D}_3(\mathrm{ReLU}(\mathrm{Conv1D}_2(\mathrm{ReLU}(\mathrm{Conv1D}_1(\boldsymbol{X})))))) \tag{2}$$

式中　ReLU——激活函数

$\boldsymbol{X}$——输入特征，$\boldsymbol{X}\in\mathbf{R}^{B\times 3\times N}$

B——单批次数据数量

$\mathrm{Conv1D}_i$——卷积核大小为 3 的一维卷积

$\boldsymbol{H}$——编码后的特征

特征解码器采用与编码器完全一致的网络结构，其输入是单个节点的特征。第一步输入为零向量，后续为上一步选中节点的特征。解码器主要负责基于上一步选择的节点生成局部时序状态，提供当前决策的历史依赖。

该设计利用全局特征获取节点间的关联，利用局部时序状态捕捉路径的连续性，最终实现基于全局信息的局部决策特征。特征编码器与解码器通过协同作用，分别从全局和局部层面解析茶芽的空间分布特征，为节点选择提供历史上下文特征。

1.2.3　LSTM 门控模块

茶芽采摘路径规划为连续决策过程，模型需动态整合历史路径信息与当前节点局部特征，传统循环神经网络（Recurrent neural network，RNN）易出现梯度消失和梯度爆炸问题，难以捕捉长期依赖。本研究采用长短期记忆网络[25]（LSTM）作为时序建模核心，通过门控机制与记忆单元的协同调控，实现路径历史信息的有效存储、筛选与整合。

如图 4 所示，LSTM 单元由遗忘门（Forget gate）、输入门（Input gate）、输出门（Output gate）、记忆单元（Cell state）构成，其输入为特征解码器输出的单个节点高维局部特征，输出为融合历史信息与当前特征的隐藏状态，为后续节点选择提供上下文支撑。

遗忘门负责筛选需从长期记忆中丢弃的无效历史信息（如已访问节点的冗余特征），输出 0 ~ 1 区间的概率值以决定历史信息的保留比例，计算公式为

$$\boldsymbol{F}_t=\sigma(\boldsymbol{W}_f[\boldsymbol{h}_{t-1};\boldsymbol{x}_t]+\boldsymbol{b}_f) \tag{3}$$

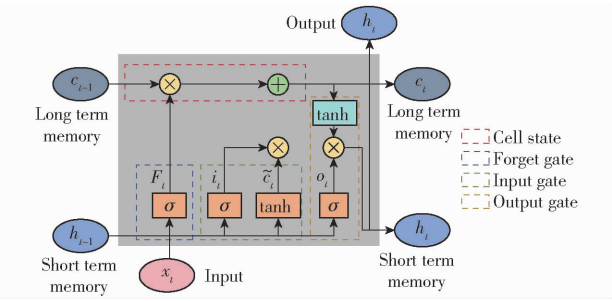

图 4 LSTM 门控模块结构图

Fig. 4 Structure diagram of LSTM gating module

式中 $\boldsymbol{h}_{t-1}$——$t-1$ 时刻 LSTM 的短期隐藏状态

$\boldsymbol{x}_t$——t 时刻当前节点的高维局部特征

$\boldsymbol{F}_t$——t 时刻遗忘门的输出值

$\boldsymbol{W}_f$——遗忘门的权重矩阵

$\boldsymbol{b}_f$——遗忘门的偏置项

σ——Sigmoid 激活函数

遗忘门的输出越接近 1 表示对应历史信息需保留，越接近 0 表示需丢弃。

输入门用于筛选需注入长期记忆的新信息，分为信息筛选与信息更新两步。先通过 $\boldsymbol{i}_t$ 确定新信息的保留比例，再通过 $\tilde{\boldsymbol{c}}_t$ 生成候选记忆单元，tanh 函数将新信息映射至 $[-1,1]$ 区间，最终通过逐元素乘法得到需注入的新信息，计算公式为

$$\boldsymbol{I}_t = \boldsymbol{i}_t \odot \tilde{\boldsymbol{c}}_t = [\sigma(\boldsymbol{W}_i[\boldsymbol{h}_{t-1};\boldsymbol{x}_t] + \boldsymbol{b}_i)][\tanh(\boldsymbol{W}_c[\boldsymbol{h}_{t-1};\boldsymbol{x}_t] + \boldsymbol{b}_c)] \tag{4}$$

式中 $\boldsymbol{i}_t$——输入门控系数

$\tilde{\boldsymbol{c}}_t$——候选记忆单元

$\boldsymbol{I}_t$——t 时刻输入门输出值

$\boldsymbol{W}_i$——输入门信息筛选的权重矩阵

$\boldsymbol{b}_i$——输入门信息筛选的偏置项

$\boldsymbol{W}_c$——输入门候选记忆单元的权重矩阵

$\boldsymbol{b}_c$——输入门候选记忆单元的偏置项

输出门作为连接长期记忆与短期决策的关键组件，控制着细胞状态中哪些信息应传递至当前隐藏状态，其核心作用是过滤长期记忆中的冗余信息，确保短期隐藏状态仅保留与当前决策相关的时序特征。输出门首先通过 Sigmoid 函数生成 0～1 区间的筛选向量 $\boldsymbol{o}_t$，然后将筛选向量 $\boldsymbol{o}_t$ 与经过 tanh 函数归一化的细胞状态进行逐元素乘法，最终生成当前时刻的短期隐藏状态 $\boldsymbol{h}_t$，计算公式为

$$\boldsymbol{h}_t = \boldsymbol{o}_t \tanh(\boldsymbol{c}_t) = [\sigma(\boldsymbol{W}_o[\boldsymbol{h}_{t-1};\boldsymbol{x}_t] + \boldsymbol{b}_o)]\tanh(\boldsymbol{c}_t) \tag{5}$$

式中 $\boldsymbol{o}_t$——t 时刻的输出门控系数

$\boldsymbol{c}_t$——t 时刻的细胞状态

$\boldsymbol{W}_o$——输出门的权重矩阵

$\boldsymbol{b}_o$——输出门的偏置项

记忆单元作为 LSTM 的长期记忆核心，先借助遗忘门丢弃无效历史，再注入筛选后的新信息，完成长期记忆的动态迭代，计算公式为

$$\boldsymbol{c}_t = \boldsymbol{F}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \tilde{\boldsymbol{c}}_t \tag{6}$$

1.2.4 带约束的多头自注意力机制设计

为实现全局路径优化与局部节点聚焦的协同调控，解决传统注意力机制在路径规划中存在的特征单一化与约束软化问题，本研究设计带约束多头自注意力机制[26]（Constrained multi-head attention）。如图 5 所示，该机制通过多子空间并行学习捕捉茶芽特征多维度关联，嵌入动态掩码硬约束确保决策合法性，提升复杂茶芽分布场景下的决策精度。

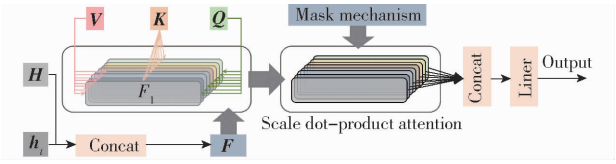

图 5 带约束的多头自注意力机制结构图

Fig. 5 Structure diagram of constrained multi-head self-attention mechanism

该机制将输入特征映射到多个子空间并行处理，首先将特征编码器输出的全局空间特征矩阵 $\boldsymbol{H}$ 与 LSTM 输出的当前隐藏状态 $\boldsymbol{h}_t$ 进行特征拼接，得到融合特征矩阵 $\boldsymbol{F}$

$$\boldsymbol{F} = \mathrm{concat}(\boldsymbol{H}, \boldsymbol{h}_t) \tag{7}$$

通过线性变换将 $\boldsymbol{F}$ 投影到 k 个相互独立的子空间，生成查询矩阵 $\boldsymbol{Q}$、键矩阵 $\boldsymbol{K}$ 和值矩阵 $\boldsymbol{V}$

$$\begin{cases} \boldsymbol{Q}_k = \boldsymbol{F}\boldsymbol{W}_q^{(k)} \\ \boldsymbol{K}_k = \boldsymbol{F}\boldsymbol{W}_k^{(k)} \\ \boldsymbol{V}_k = \boldsymbol{F}\boldsymbol{W}_v^{(k)} \end{cases} \tag{8}$$

式中 k——注意力头数

$\boldsymbol{W}_q^{(k)}$——第 k 个注意力头的 Query 线性投影权重矩阵

$\boldsymbol{W}_k^{(k)}$——第 k 个注意力头的 Key 线性投影权重矩阵

$\boldsymbol{W}_v^{(k)}$——第 k 个注意力头的 Value 线性投影权重矩阵

每个注意力头并行探索学习不同维度的特征，通过独立的参数学习，各分支可自主关注对路径优化有价值的关联特征，模型将根据训练过程中的奖励反馈，自适应调整各头的关注重点。

为严格遵守无重复访问的路径约束，每个注意力头的分数计算直接嵌入动态掩码机制。对于第 k 个头，首先计算未归一化注意力分数 $\boldsymbol{S}_k$，随后将分数矩阵与动态掩码向量 $\boldsymbol{M}'$ 进行逐元素乘法，获取约

束后的注意力分数矩阵 $\hat{\boldsymbol{S}}_k$

$$\hat{\boldsymbol{S}}_k = \boldsymbol{S}_k \odot \boldsymbol{M}' = \left(\frac{\boldsymbol{Q}_k \boldsymbol{K}_k^{\mathrm{T}}}{\sqrt{d_k}} \right) \odot \boldsymbol{M}' \tag{9}$$

式中　d_k——每个头的特征维度，通过平方根防止高维度下点积数值过大导致的梯度消失

$\boldsymbol{M}'$——动态掩码向量，已访问节点置 $-\infty$，未访问节点置 1

经约束后的分数通过 Softmax 归一化得到注意力权重，再与值矩阵加权求和生成第 k 头的输出 $\boldsymbol{h}_k$

$$\boldsymbol{h}_k = \mathrm{Softmax}(\hat{\boldsymbol{S}}_k) \boldsymbol{V}_k \tag{10}$$

最后将所有头的输出拼接并通过线性变换整合，得到最终注意力特征 $\boldsymbol{MH}$

$$\boldsymbol{MH}(\boldsymbol{H}, \boldsymbol{h}_t) = \boldsymbol{W}_{MH} \mathrm{Concat}(\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_k) \tag{11}$$

式中　$\boldsymbol{W}_{MH}$——输出投影矩阵

$\boldsymbol{h}_i$——第 i 个头的输出

基于注意力特征生成每个节点的选择概率

$$\boldsymbol{p}_t = \mathrm{Softmax}(\boldsymbol{MH}) \tag{12}$$

式中　$\boldsymbol{p}_t$——第 t 步的节点选择概率分布，Softmax 函数确保概率和为 1

模型采用随机策略采样（Stochastic policy sampling），即根据概率分布 $\boldsymbol{p}_t$ 中随机采样，而非直接选择概率最高的节点，该方式可增强模型探索性，避免陷入局部最优路径，同时符合强化学习中策略梯度方法的随机性要求，提升模型在不同茶芽分布场景下的泛化能力。当所有有效茶芽节点访问完成后，动态掩码仅保留终点节点的掩码值为 1，模型通过概率采样强制选择终点，最终形成闭环采摘路径。

1.2.5　奖励机制设计和网络损失函数

为引导深度强化学习智能体学习最优采摘路径，本研究以路径总长度最小化的核心优化目标为导向，设计基于路径总长度负值的奖励信号机制。该机制将路径优化目标与奖励反馈进行直接量化映射，使智能体在训练过程中能直观感知决策质量，有效避免复杂奖励函数带来的训练偏移或梯度消失问题，同时简化模型训练流程，提升优化效率。奖励函数的具体定义为

$$R = -\sum_{i=0}^{N} d(\boldsymbol{\pi}_i, \boldsymbol{\pi}_{i+1}) = -\sum_{i=0}^{N} \sqrt{(x_{i+1}-x_i)^2 + (y_{i+1}-y_i)^2 + (z_{i+1}-z_i)^2} \tag{13}$$

式中　$\boldsymbol{\pi}_i$——路径序列中第 i 个节点的三维坐标

R——实际累计奖励

d——三维欧氏距离函数

本研究采用策略梯度方法计算 Actor 网络的损失，通过融合实际回报与价值估计的偏差指导参数优化，损失函数定义为

$$L(\theta) = E_{\tau \sim \pi_\theta} \left[\sum_{t=0}^{T} \ln \pi_\theta(a_t \mid s_t) (R - V(s_t)) \right] \tag{14}$$

式中　τ——采摘路径序列

T——路径规划中节点总数

$E_{\tau \sim \pi_\theta}$——策略 π_θ 下的期望值

$\pi_\theta(a_t \mid s_t)$——Actor 网络在状态 s_t 下选择动作 a_t 的概率分布

θ——Actor 网络参数

$L(\theta)$——Actor 网络的策略梯度损失函数

$V(s_t)$——Critic 网络对状态 s_t 的价值估计

网络损失计算首先通过当前轨迹的实际累计奖励 R 与 Critic 网络输出的状态价值 $V(s_t)$ 的差值得到优势函数 A_t，量化当前动作序列相对于模型估计的优劣，$A_t > 0$ 表示该序列优于预期，$A_t < 0$ 则表示劣于预期；随后将每个时刻动作的对数概率 $\log \pi_\theta(a_t \mid s_t)$ 与对应优势值 A_t 相乘并累加，得到最终损失值。

1.3　网络训练流程与参数更新策略

网络的训练流程如图 6 中的伪代码所示。

```
Algorithm 1  Training Procedure for Tea Bud Picking Sequence Planning
Initialize actor network π_θ, critic network V_ϕ
for each training episode do
  Sample a batch {X^batch} from dataset
  for each X in batch do
    Initialize mask M, LSTM state ((h,c)←(0,0))
    Encode nodes: H←Encode(X) T←[start]
    for t=0 to N−1 do
      if t>1: h,c ← LSTM(H[T[t−1]],(h,c))
      for each node j:
        x_j ← Decoder(X[j]); h_j,c_j ← LSTM(x_j,(h,c)); u_j ← Attention(H,h_j);
        p←softmax(u); a_t←Sample(p); T←T∪{a_t}; M[a_t]←0
    end for
    R←TourLength(T,X)
  end for
  for each (X) in batch do A ← R − V_ϕ(X)
    Update θ (actor) with A
    Update ϕ (critic) with (V_ϕ(X) − R)^2
  end for
end for
```

图 6　网络训练流程伪代码

Fig. 6　Pseudocode of network training process

为实现 Actor 与 Critic 网络的高效训练，采用独立的 Adam 优化器分别对两者参数进行更新。Actor 网络优化器负责调整策略参数 θ，Critic 网络优化器负责更新价值估计参数 Φ，通过分离优化避免参数更新相互干扰，提升训练稳定性。

参数更新流程遵循批量梯度下降的策略，每轮训练结束后，计算批次内所有样本的平均损失，即 Actor 的策略梯度损失与 Critic 的均方误差损失，通过损失

获取梯度信息,再由 Adam 优化器根据梯度方向更新网络参数。为防止梯度爆炸导致模型发散,训练中引入梯度裁剪机制,设置全局梯度范数阈值为 2.0,当梯度范数超过阈值时,按比例缩放梯度至阈值内,确保参数更新幅度可控,保障模型收敛稳定性。

模型训练阶段采用随机策略采样(Stochastic sampling)以增强探索性,而在模型评估或推理阶段则切换为贪婪策略采样(Greedy sampling),即选择注意力分数最高的节点作为下一个目标,通过确定性决策提升路径规划效率。

2 试验

2.1 测试集构建与评价指标

为验证本研究提出的融合带约束多头注意力机制的演员-评论家方法(Actor - Critic with constrained multi-head attention,AC - CMHA)的有效性,本研究通过与精确算法对比、与经典启发式和其他深度强化学习算法对比、真实茶园场景测试对比 3 组试验,从路径优化质量、计算效率验证方法的性能优势。

测试集生成规则与 1.1.2 节中环境构建一致。按茶芽数量将场景划分为稀疏、中等、密集 3 类子场景,分别包含 30、50、80 个茶芽,每类子场景设置 100 个独立的模拟案例,总计 300 个测试样本,通过大样本量消除茶芽随机分布带来的统计偏差。所有场景统一设置起始点坐标为(0,0,1.0) m。

为全面、客观衡量算法性能,选取 3 个核心评价指标:平均路径长度、平均计算时间、相对最优性差距,分别对应路径优化效果、实时响应能力以及输出解与最优解的偏差程度。平均路径长度为所有测试样本路径总长度的平均值,数值越小代表路径优化水平越高;平均计算时间为算法生成完整采摘路径的平均耗时,数值越小越能满足采摘机器人实时决策的作业需求;相对最优性差距用于量化算法输出解与理论最优解的偏差程度,计算公式为

$$G = \frac{L_{\text{AC-CMHA}} - L_{\text{Concorde}}}{L_{\text{Concorde}}} \times 100\% \tag{15}$$

式中 $L_{\text{AC-CMHA}}$——AC - CMHA 算法输出的路径长度

L_{Concorde}——Concorde 精确算法输出的路径长度

2.2 试验平台与超参数设置

试验硬件配置为 CPU Intel Core i7 - 12700K、GPU NVIDIA RTX 3090(24 GB)、内存 32 GB,确保满足深度强化学习模型的计算需求;软件框架采用 Python 3.8、PyTorch 1.12.1 及 CUDA 11.6,保障模型训练与测试的稳定性。

本研究提出的 AC - CMHA 模型超参数通过控制变量法迭代调试确定,关键超参数设置如下:批处理大小为 256,训练场景数量为 1×10^7,带约束的多头自注意力机制的头数为 8,Actor 网络学习率为 1×10^{-4},Critic 网络学习率为 1×10^{-4},学习率下降算法为余弦退火算法,优化器为 Adam,训练轮次为 200 轮。所有超参数均通过控制变量法调试确定,以兼顾模型收敛速度与最终性能。

所有对比算法的超参数均参考其原始文献[4 - 7,21 - 22]最优配置,具体超参数配置如表 1 所示。

表 1 各算法超参数配置

Tab.1 Hyperparameter configuration of each algorithm

算法	参数设置
Concorde	最大求解时间限制:10 s
蚁群算法	信息素权重:1.0;启发式权重:2.0;信息素蒸发率:0.1;蚂蚁数量:50;迭代控制:200
模拟退火算法	初始温度:100;降温速率:0.95;最低温度:0.001;每温度迭代数:50
遗传算法	种群规模:50;交叉概率:0.8;变异概率:0.05;最大迭代数:100
L - K 算法	候选集大小:10;最大步数:1 000;惩罚系数:5;贪婪系数:0.5
S2V - DQN	优化器:Adam 优化器;学习率:0.001;训练批次大小:128;训练轮次:100 轮
GPN - HRL	优化器:Adam 优化器;学习率:0.001;训练批次大小:512;训练轮次:100 轮

2.3 与精确算法 Concorde 对比试验

Concorde 作为旅行商问题的经典精确算法,能够输出理论最优解。通过将本研究提出的 AC - CMHA 与 Concorde 对比,验证 AC - CMHA 在解质量和计算效率权衡中的优势,明确其是否在保证解质量接近最优的同时,具备实际应用所需的实时响应能力。

AC - CMHA 与 Concorde 在不同茶芽数量场景下的性能对比如表 2 所示。由表 2 可知,在解质量方面,AC - CMHA 的平均路径长度始终略高于 Concorde 的理论最优解,但相对最优性差距始终控制在 2% 以内,表明 AC - CMHA 已能通过 DRL 的端到端学习,有效捕捉三维空间中茶芽分布的全局优化规律;在计算效率方面,两者的差异随茶芽数量增

加逐渐扩大，30 个茶芽场景下，AC－CMHA 平均计算时间为 0.013 s，仅为 Concorde（0.095 s）的 13.7%；50 个茶芽场景下，AC－CMHA 耗时 0.023 s，降至Concorde（0.205 s）的 11%；80 个茶芽场景下，AC－CMHA 耗时 0.037 s，仅为 Concorde（2.125 s）的 1.7%，且 Concorde 的计算时间随茶芽数量增大呈急速增长，而 AC－CMHA 耗时增长平缓。

表 2　AC－CMHA 与 Concorde 算法性能对比

Tab.2　Performance comparison between AC－CMHA and Concorde algorithms

茶芽数量	算法	平均路径长度/mm	平均计算时间/s	相对最优性差距/%
30	Concorde	6 028	0.095	
	AC－CMHA	6 099	0.013	1.18
40	Concorde	7 227	0.142	
	AC－CMHA	7 302	0.018	1.04
50	Concorde	8 201	0.205	
	AC－CMHA	8 348	0.023	1.80
60	Concorde	9 354	0.350	
	AC－CMHA	9 506	0.026	1.62
70	Concorde	10 133	1.063	
	AC－CMHA	10 293	0.030	1.58
80	Concorde	11 043	2.125	
	AC－CMHA	11 250	0.037	1.87

从解质量角度看，AC－CMHA 的相对最优性差距始终控制在 2% 以内，表明其虽为近似算法，但解质量已接近理论最优，能够满足茶园采摘路径短程化的需求。从计算效率角度看，Concorde 的时间增长使其在茶芽数量超过 50 后，难以满足采摘机器人毫秒级响应的需求，而 AC－CMHA 在 80 个茶芽场景下仍保持 0.037 s 的快速决策，实时性优势显著。

从场景适应性看，AC－CMHA 耗时随茶芽数量增长平缓，核心原因在于其采用策略学习的端到端决策模式。模型经训练后，推理阶段无需像精确算法（Concorde）或传统启发式算法（如蚁群、遗传算法）依赖迭代搜索、枚举或路径片段优化，只需将茶芽节点特征输入模型，通过 Actor 网络直接输出节点选择概率分布并完成采样，整个过程的计算复杂度不随节点数量增长而急速增长，因此茶芽数量增加对模型计算时间的影响较小。

2.4　与其他算法对比试验

为进一步验证 AC－CMHA 在三维茶芽采摘路径规划任务中的综合优越性，选取两类主流路径规划算法作为对比对象：一类为经典启发式算法，包括蚁群算法[4]（Ant colony optimization，ACO）、模拟退火算法[5]（Simulated annealing，SA）、遗传算法[6]（Genetic algorithm，GA）及 L－K 局部搜索算法[7]（Lin-Kernighan，L－K），此类算法为组合优化领域求解 TSP 的常用方法，依赖手工设计的启发式规则搜索近似解。另一类为基于图网络的深度强化学习算法，包括 S2V－DQN[21]（Struct2Vec－Deep Q－Network）与 GPN－HRL[22]（Graph pointer network-hierarchical reinforcement learning），此类算法代表近年 DRL 求解 TSP 的优秀方案，通过图结构建模捕捉节点间关联。

通过多维度指标对比，全面验证 AC－CMHA 在不同节点规模下的适应性与优越性。所有对比算法在不同茶芽数量场景下的性能如表 3 所示。

表 3　AC－CMHA 与其他算法的性能对比

Tab.3　Performance comparison between AC－CMHA and other algorithms

茶芽数量	算法	平均路径长度/mm	平均计算时间/s	相对最优性差距/%
30	Concorde	6 028	0.095	
	ACO	6 426	0.782	6.60
	SA	6 412	0.114	6.37
	GA	7 957	0.987	32.00
	L－K	6 028	0.014	0
	S2V－DQN	6 123	1.183	1.57
	GPN－HRL	6 103	0.512	1.25
	AC－CMHA	6 099	0.013	1.18
50	Concorde	8 201	0.205	
	ACO	8 706	4.302	6.16
	SA	8 632	0.342	5.26
	GA	11 218	4.118	36.79
	L－K	8 294	0.067	1.13
	S2V－DQN	8 597	2.752	4.84
	GPN－HRL	8 578	1.215	4.60
	AC－CMHA	8 348	0.023	1.79
80	Concorde	11 043	2.125	
	ACO	11 527	24.010	4.38
	SA	11 347	0.864	2.75
	GA	15 500	16.325	40.36
	L－K	11 231	0.256	1.70
	S2V－DQN	11 995	6.291	8.62
	GPN－HRL	11 764	4.348	6.53
	AC－CMHA	11 250	0.037	1.87

AC－CMHA 在 30、50、80 个茶芽的不同规模场景中均表现出优异的综合性能。小规模相对最优性差距 1.18%、耗时 0.013 s，解质量仅略低于 L－K 算法，效率与 L－K 相当且远优于其他对比算法。中规模相对最优性差距增至 1.79%，仍优于多数算法，耗时 0.023 s，仅为 L－K 的 34.3%、GPN－HRL 的 1.9%。大规模综合性能最优，相对最优性差距 1.87% 接近 L－K，耗时 0.037 s，仅为 L－K 的 14.5%、GPN－HRL 的 0.85%。

经典启发式算法虽实现简单,但普遍存在计算效率低、解质量差的问题,L－K 算法虽中小规模解质量接近最优,却随节点数量增加耗时激增;S2V－DQN、GPN－HRL 等图网络强化学习算法,虽提升了节点关联捕捉能力,但计算效率低,且解质量随节点数量增加衰减明显,动态适配性不足。

AC－CMHA 通过提取茶芽空间特征与历史决策信息,结合带约束多头自注意力机制精准选点,既规避了启发式算法手工调参、效率低的局限,又克服了其他深度强化学习算法适配性差的短板,在各规模场景中均能兼顾稳定的解质量与高效的决策速度。

2.5 真实茶园场景试验对比

为进一步验证所提 AC－CMHA 方法在真实茶园环境中的有效性与工程适用性,本研究进行实际场景试验。基于真实茶园场景的茶芽三维坐标数据开展采摘路径规划测试,重点验证算法在实际茶芽分布场景下的路径优化质量、实时响应能力,确保算法从模拟场景到真实场景的迁移有效性。

如图 7 所示,本次实测的茶芽三维坐标数据通过机器视觉和点云提取的方式获取。首先在茶园场景采集茶芽冠层 RGB 图像,人工标注实际可采摘茶芽节点;再将二维图像标注点映射至同步采集的茶芽冠层三维点云数据,提取每个采摘点的精准三维空间坐标;最后对三维点云坐标轴进行旋转平移处理,使茶芽分布至工作空间 $x\in(0,1)$,$y\in(0,1)$,$z\in(0.6,0.95)$,统一设置起点与终点为(0,0,1.0)(单位:m),与模拟场景的约束条件保持一致,确保路径规划结果满足闭环作业的实际要求。



图 7 茶芽标注采摘点与三维点云映射示意图

Fig. 7 Schematic of tea bud annotation points and 3D point cloud mapping

为突出 AC－CMHA 方法的实际性能优势,选取模拟场景中表现最优的精确算法 Concorde、经典启发式算法中解质量最优的 L－K 算法、深度强化学习算法 GPN－HRL 作为对比算法。选取真实茶园中茶芽数量为 31、52、80 个的 3 类典型场景开展可视化对比测试,3 类场景分别对应茶园实际作业中的稀疏、中等、密集采摘场景,贴合采茶机器人实际作业的节点规模特征。3 类真实场景的路径规划结果如图 8 所示。

AC－CMHA 在路径优化质量与计算效率上展现出显著的综合优势。在路径长度方面,AC－CMHA 规划的路径长度与 Concorde、L－K 算法接近,优于 GPN－HRL,无明显性能衰减,表明算法对真实茶芽空间分布的特征捕捉能力优异;在计算耗时方面,AC－CMHA 的耗时远低于其他对比算法,且随茶芽数量增加耗时增长幅度极小,31 个茶芽场景耗时仅 0.015 s,80 个茶芽场景耗时仅 0.039 s,完全满足采茶机器人毫秒级实时决策的作业需求,而 Concorde、GPN－HRL 在密集场景下的耗时增长,难以适配实际作业的实时性要求。为进一步量化算法在真实场景中的整体性能,扩大实测样本量,总计采集 50 个茶园真实采摘场景的茶芽三维坐标数据开展批量测试,测试场景茶芽数量覆盖 30～80 个,贴合实际作业的节点规模区间,批量测试结果如表 4 所示。

由表 4 数据可知,在 50 个真实茶园场景的批量测试中,AC－CMHA 展现出解质量优、计算效率高、鲁棒性强的综合优势。

AC－CMHA 的平均路径长度为 5 271 mm,仅比精确算法 Concorde 高出 95 mm,相对最优性差距为 1.84%,与经典启发式算法 L－K 的解质量接近,表明其在真实复杂环境中仍能精准捕捉茶芽空间分布特征,生成接近全局最优的采摘路径。

AC－CMHA 的平均计算时间仅为 0.029 s,远低于其他对比算法,分别是 L－K 算法的 26.85%、GPN－HRL 的 1.55%、Concorde 的 4.08%,毫秒级的决策速度完全满足采茶机器人实时作业的响应要求。AC－CMHA 在茶芽数量 30～80 个的动态真实场景中,路径长度与计算时间均无明显波动,解质量与效率始终保持稳定,未出现因茶芽分布密度变化导致的性能衰减,充分验证了其对真实茶园复杂环境的强适配性与可靠性。

3 结论

(1)本研究围绕采茶机器人三维路径规划的实际工程需求,将茶芽采摘路径优化问题转化为三维旅行商问题,提出了一种基于深度强化学习的茶芽采摘路径规划方法。该方法通过动态掩码机制确保路径合法性并缩小决策空间,结合特征编码-解码网络与 LSTM 门控模块,协同提取茶芽全局空间分布特征与历史上下文的决策特征,强化模型对三维空间关系的感知能力;引入带约束的多头自注意力机制,通过多子空间并行学习捕捉茶芽三维坐标多维度特征关联,同时嵌入动态掩码硬约束,提升决策精准性。在此基础上,使用 Actor－Critic 框架作为模型训练的优化架构,利用Critic网络提供状态价值

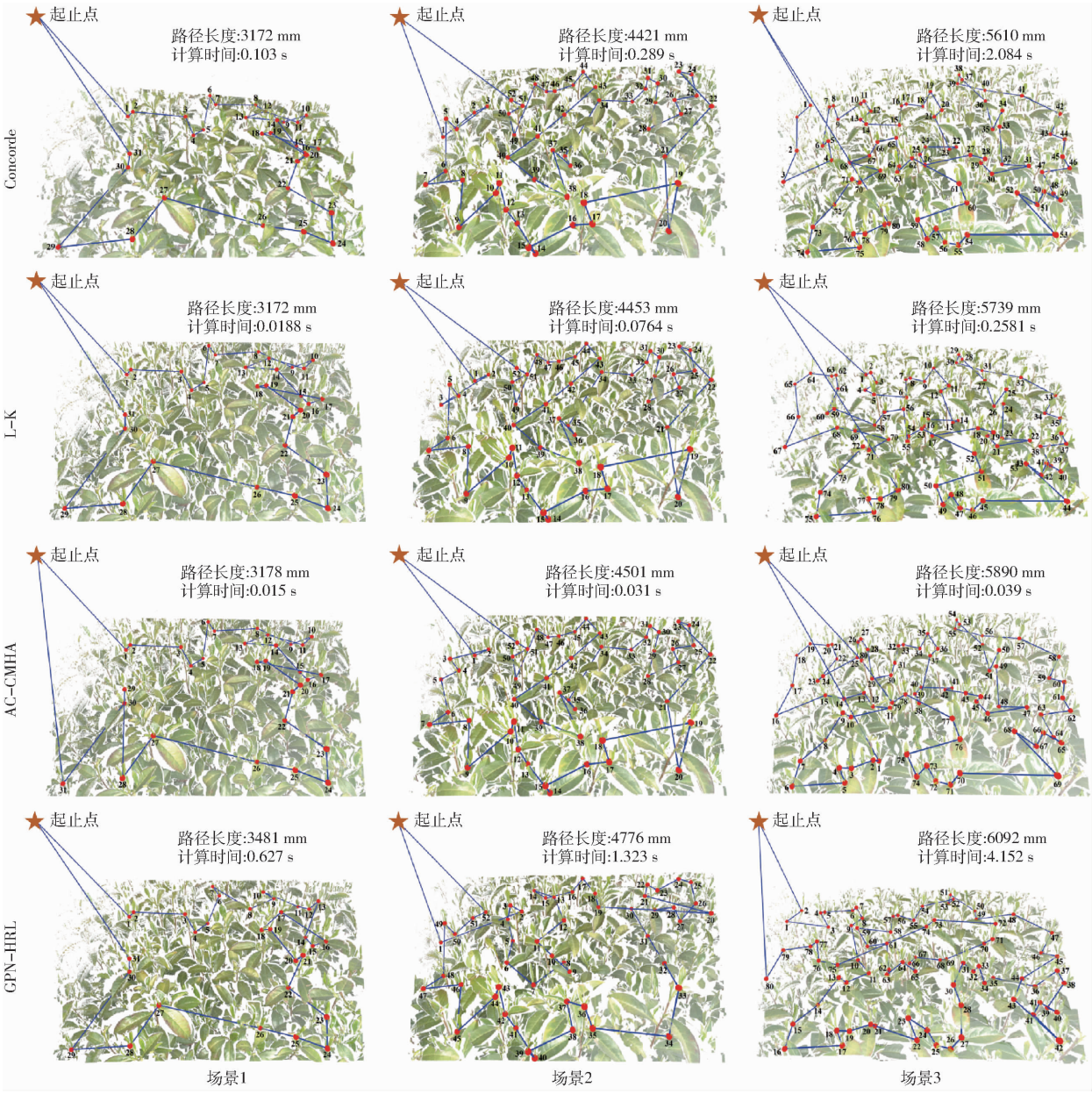

图 8　不同茶芽数量下真实场景路径规划对比结果

Fig. 8　Comparison of path planning results in real scenarios with different tea bud quantities

表 4　真实场景批量测试性能对比

Tab. 4　Performance comparison of batch tests in real scenarios

算法	平均路径长度/mm	平均计算时间/s	相对最优性差距/%
Concorde	5 176	0. 710	
L－K	5 260	0. 108	1. 62
GPN－HRL	5 479	1. 875	5. 85
AC－CMHA	5 271	0. 029	1. 84

估计基线，通过优势函数降低策略梯度方差，保障模型稳定训练并向路径最短化目标优化，最终形成完整的路径规划解决方案。

（2）研究通过试验验证了本研究方法的优越性，在解质量上与精确算法 Concorde 相比相对最优性差距始终控制在 2% 以内，接近理论最优，实时性上全场景平均计算时间小于 0. 04 s，满足采摘机器人的实时性响应要求，动态适配性上随茶芽数量从 30 个增至 80 个耗时仅增长 2. 8 倍，显著低于其他算法，同时克服了经典启发式算法手工调参繁琐、图网络深度强化学习方法大规模场景效率低的局限，实现高质量解、毫秒级响应、适应节点数量变化的目标。

参　考　文　献

[1]　贾江鸣，王登艺，贺磊盈，等. 名优茶采摘机器人整机机械技术研究进展与展望[J]. 农业机械学报，2025，56(8)：193－206.

Jia Jiangming, Wang Dengyi, He Leiying, et al. Research progress and prospects of mechanical technology for whole machine of high-quality tea picking robot[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(8): 193 - 206. (in Chinese)

[2] Held M, Karp R M. A dynamic programming approach to sequencing problems [J]. Journal of the Society for Industrial and Applied Mathematics, 1962, 10(1): 196 - 210.

[3] Tunon M I C, Lopez M R. Design and use of the CPAN branch and bound for the solution of the travelling salesman problem (TSP) [C]// International Conference on Electronics, Communications and Computers (CONIELECOMP'05). Piscataway: IEEE, 2005: 262 - 267.

[4] Dorigo M, Birattari M, Stutzle T. Ant colony optimization [J]. IEEE Computational Intelligence Magazine, 2007, 1(4): 28 - 39.

[5] Kirkpatrick S, Gelatt J C D, Vecchi M P. Optimization by simulated annealing [J]. Science, 1983, 220 (4598): 671 - 680.

[6] Goldberg D E. Genetic algorithms in search, optimization, and machine learning [M]. Reading, MA: Addison Wesley, 1989: 36.

[7] Lin S, Kernighan B W. An effective heuristic algorithm for the traveling-salesman problem [J]. Operations Research, 1973, 21(2): 498 - 516.

[8] Yang H, Chen L, Ma Z, et al. Computer vision-based high-quality tea automatic plucking robot using delta parallel manipulator [J]. Computers and Electronics in Agriculture, 2021, 181: 105946.

[9] Tong J, Qiu Z, Zhou H, et al. Optimizing the path of seedling transplanting with multi-end effectors by using an improved greedy annealing algorithm [J]. Computers and Electronics in Agriculture, 2022, 201: 107276.

[10] Xu L, Xie Y, Chen X, et al. Design of an efficient combined multipoint picking scheme for tea buds [J]. Frontiers in Plant Science, 2022, 13: 1042035.

[11] 李涛,邱权,赵春江,等. 矮化密植果园多臂采摘机器人任务规划[J]. 农业工程学报,2021, 37(2): 1 - 10.
Li Tao, Qiu Quan, Zhao Chunjiang, et al. Task planning of multi-arm harvesting robots for high-density dwarf orchards[J]. Transactions of the CSAE, 2021, 37(2): 1 - 10. (in Chinese)

[12] Deudon M, Cournut P, Lacoste A, et al. Learning heuristics for the TSP by policy gradient [C]// International Conference on the Integration of Constraint Programming, Artificial Intelligence, and Operations Research. Cham: Springer International Publishing, 2018: 170 - 181.

[13] Vinyals O, Fortunato M, Jaitly N. Pointer networks [C]// Advances in Neural Information Processing Systems, 2015: 2692 - 2700.

[14] Kool W, Van Hoof H, Welling M. Attention, learn to solve routing problems! [PP/OL]. V3. arXiv(2019 - 02 - 07)[2026 - 04 - 10]. https://doi.org/10.48550/arXiv.1803.08475.

[15] Bruna J, Zaremba W, Szlam A, et al. Spectral networks and locally connected networks on graphs[PP/OL]. V3. arXiv(2014 - 05 - 21)[2026 - 04 - 10]. https://doi.org/10.48550/arXiv.1312.6203.

[16] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering [C]// Advances in Neural Information Processing Systems, 2016: 3844 - 3852.

[17] Sukhbaatar S, Szlam A, Fergus R. Learning multiagent communication with backpropagation [C]// Advances in Neural Information Processing Systems, 2016: 2244 - 2252.

[18] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[PP/OL]. V4. arXiv(2017 - 02 - 22)[2026 - 04 - 10]. https://doi.org/10.48550/arXiv.1609.02907.

[19] Hamilton W L, Ying R, Leskovec J. Inductive representation learning on large graphs [C]// Advances in Neural Information Processing Systems, 2017: 1024 - 1034.

[20] Bello I, Pham H, Le Q V, et al. Neural combinatorial optimization with reinforcement learning[PP/OL]. V3. arXiv(2017 - 01 -12)[2026 - 04 - 10]. https://doi.org/10.48550/arXiv.1611.09940.

[21] Dai H J, Khalil E B, Zhang Y Y, et al. Learning combinatorial optimization algorithms over graphs [C]// Advances in Neural Information Processing Systems, 2017: 6348 - 6358.

[22] Ma Q, Ge S, He D, et al. Combinatorial optimization by graph pointer networks and hierarchical reinforcement learning[PP/OL]. V1. arXiv(2019 - 11 - 12)[2026 - 04 - 10]. https://doi.org/10.48550/arXiv.1911.04936.

[23] Joshi C K, Laurent T, Bresson X. An efficient graph convolutional network technique for the travelling salesman problem[PP/OL]. V2. arXiv(2019 - 10 - 14)[2026 - 04 - 10]. https://doi.org/10.48550/arXiv.1906.01227.

[24] Barto A G, Sutton R S, Anderson C W. Neuronlike adaptive elements that can solve difficult learning control problems [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1983, 13(5): 834 - 846.

[25] Gers F A, Schmidhuber J, Cummins F. Learning to forget: continual prediction with LSTM [J]. Neural Computation, 2000, 12(10): 2451 - 2471.

[26] Vaswan A, Shazeer N, Parmar N, et al. Attention is all you need [C]// Advances in Neural Information Processing Systems, 2017: 5998 - 6008.