



兰州大学

本科毕业论文

论文题目 (中文) 基于强化学习的多无人机
草原修复方法研究

论文题目 (英文) Research on Multi-UAV Grassland Restoration
Method Based on Reinforcement Learning

学生姓名 王贤义
指导教师 焦栋斌
学 院 信息科学与工程学院
专 业 计算机科学与技术
年 级 2021 级

兰州大学教务处

诚信责任书

本人郑重声明：本人所呈交的毕业论文（设计），是在导师的指导下独立进行研究所取得的成果。毕业论文（设计）中凡引用他人已经发表或未发表的成果、数据、观点等，均已明确注明出处。除文中已经注明引用的内容外，不包含任何其他个人、集体已经发表或未发表的论文。

本声明的法律责任由本人承担。

论文作者签名： 签名 日 期： 签名

关于毕业论文（设计）使用授权的声明

本人在导师指导下所完成的论文及相关的职务作品，知识产权归属兰州大学。本人完全了解兰州大学有关保存、使用毕业论文（设计）的规定，同意学校保存或向国家有关部门或机构送交论文的纸质版和电子版，允许论文被查阅和借阅；本人授权兰州大学可以将本毕业论文（设计）的全部或部分内容编入有关数据库进行检索，可以采用任何复制手段保存和汇编本毕业论文（设计）。本人离校后发表、使用毕业论文（设计）或与该毕业论文（设计）直接相关的学术论文或成果时，第一署名单位仍然为兰州大学。

本毕业论文（设计）研究内容：

☒ 可以公开

☐ 不宜公开，已在学位办公室办理保密申请，解密后适用本授权书。

（请在以上选项内选择其中一项打“√”）

论文作者签名： 签名

导师签名： 签名

日 期： 签名

日 期： 签名

基于强化学习的多无人机草原修复方法研究

中文摘要

本文系统研究了多无人机协同草原修复面积最大化问题。草原生态系统作为维持全球生物多样性和生态平衡的重要组成部分，近年来却因人类活动和气候变化等多重因素面临严重退化，表现为植被覆盖率下降、土壤沙化加剧等多方面生态失衡。传统的草原修复方法主要依赖人工作业和机械化手段，存在效率低、成本高、难以适应大范围复杂地形等局限性。随着无人机技术的快速发展，多无人机协同作业为草原生态修复提供了新的解决思路，但如何在能量受限、任务复杂的环境下实现无人机协同与修复面积最大化，仍是亟需突破的难题。

针对上述问题，本文提出了一种基于深度强化学习的多无人机协同草原修复优化方法。首先，将待修复草原建模为一个完全无向图，节点代表待修复区域，每个区域包含位置坐标、退化程度和面积等信息。综合考虑无人机在飞行、播种和航拍过程中的能量消耗模型，以及草种重量动态变化对能耗的影响，建立了以修复面积最大化为目标的优化模型。本文创新性地采用深度强化学习框架，利用 Transformer 作为编码器提取静态环境特征和动态无人机状态特征，结合指针网络作为自回归解码器，动态构建无人机修复路径和任务分配方案。通过设计兼顾能量约束与修复效率的奖励函数，并采用 Actor-Critic 算法进行训练，使模型能够自主学习并优化修复策略。

在此基础上，本文进一步设计了多无人机协同调度算法，通过地面信息中心实时整合各无人机的状态信息，动态调整修复地图，实现多无人机间的高效协同与任务再分配，从而提升整体修复效率。实验部分在多种规模和分布的草原实例上进行了仿真验证，结果表明，本文提出的算法在修复面积、路径优化和能量利用率等方面均优于传统优化方法，修复面积提升幅度最高可达 40%。本文提出的算法为大规模草原生态修复任务提供了一种高效、智能的技术方案，对无人机协同作业和生态环境治理具有重要的理论意义和应用价值。

关键词：多无人机协同；草原修复；组合优化；深度强化学习；Transformer；指针网络；Actor-Critic 算法

Research on Multi-UAV Grassland Restoration Method Based on Reinforcement Learning

Abstract

This paper investigates the problem of maximizing restoration area through multi-UAV coordination in grassland rehabilitation. Grassland ecosystems play a crucial role in maintaining biodiversity and ecological balance, but grassland degradation has become increasingly serious in recent years. To address the inefficiency of traditional restoration methods in large-scale environments, this research proposes a deep reinforcement learning-based optimization method for multi-UAV cooperative grassland restoration.

We first model the grassland to be restored as a complete undirected graph, where each node represents a restoration area with information on location coordinates, degradation degree, and area size. Considering the energy consumption models of UAVs during flight, seeding, and aerial photography, as well as the dynamic impact of seed weight on energy consumption, we formulate an optimization model aimed at maximizing restoration area. The proposed method is built on a deep reinforcement learning framework that employs Transformer as an encoder to extract static environmental features and dynamic UAV status features, combined with a pointer network as an autoregressive decoder to construct solution schemes. By designing a reward function that accounts for energy constraints and applying the Actor-Critic algorithm for training, the model can autonomously learn optimal restoration strategies.

On this basis, we also design a multi-UAV coordination scheduling algorithm that dynamically adjusts restoration maps through a ground information center that integrates information from all UAVs, maximizing restoration efficiency. Experimental results show that compared with traditional optimization methods, our proposed algorithm demonstrates superior performance in grassland restoration scenarios of various scales, with restoration area improvements of up to 40%, providing an efficient and intelligent technical solution for grassland ecological restoration.

Keywords: UAV coordination; grassland restoration; area maximization; deep reinforcement learning; Transformer; pointer network; Actor-Critic algorithm

目 录

中文摘要	I
英文摘要	II
第一章 绪 论	1
1.1 研究背景与意义	1
1.2 国内外研究现状	1
1.3 研究内容	3
第二章 多无人机协同的草原修复面积最大化模型	4
2.1 各项能量约束	5
2.1.1 无人机飞行能耗	5
2.1.2 无人机播种能耗	5
2.1.3 无人机拍照能耗	6
2.2 优化目标	6
2.3 问题描述	6
2.4 本章小结	7
第三章 基于强化学习的最优修复方案	8
3.1 马尔可夫过程建模	8
3.1.1 状态	8
3.1.2 动作	8
3.1.3 状态	8
3.1.4 策略函数	8
3.1.5 奖励函数	9
3.2 神经网络模型构建	9
3.2.1 特征提取与模型架构	9
3.2.2 编码器结构	9

3.2.3	解码器结构	10
3.2.4	训练流程	11
3.3	本章小结.....	13
第四章	多无人机协同调度算法.....	14
4.1	协同修复问题描述	14
4.2	协同调度机制	14
4.3	动态地图更新	15
4.4	决策与执行流程	15
4.5	本章小结.....	16
第五章	实验与分析.....	17
5.1	仿真配置.....	17
5.2	实例设置.....	17
5.3	算法对比设置	18
5.4	参数设置.....	18
5.5	仿真结果.....	18
5.6	本章小结.....	22
第六章	总结与展望.....	23
6.1	总结	23
6.2	展望	23
参考文献	24
致 谢	28

图 目 录

图 2.1 无人机草原修复区域示意图。图中展示了待修复区域的分布情况，每个圆点代表一个待修复区域，区域大小和颜色深浅表示其退化程度和面积。.....	4
图 5.1 深度强化学习训练过程中的损失函数变化	19
图 5.2 深度强化学习训练过程中的修复面积变化	19
图 5.3 基于深度强化学习的多无人机草原修复路径规划	20
图 5.4 基于迭代局部搜索的多无人机草原修复路径规划	21

表 目 录

表 5.1 实验硬件配置 17

表 5.2 仿真实验实例设置参数表 17

表 5.3 实验软件配置 18

表 5.4 路径长度与修复面积对比（CHAPBILM 与 DRL） 22

第一章 绪 论

1.1 研究背景与意义

草原生态系统在全球环境保护和可持续发展中发挥着至关重要的作用。作为陆地生态系统的重要组成部分，草原面积约占全球陆地总面积的 26% 至 40%^[1]，承担着防风固沙、涵养水源、调节气候、美化环境和维持生物多样性等多重生态功能^[2]。然而，近年来，全球草原退化问题日益严重，主要受到人类活动和气候变化的双重影响，表现为植被覆盖率下降、土壤沙化、水资源减少等多方面的生态失衡^[3]。这些退化不仅削弱了草原的生态服务功能，还对畜牧业生产和区域经济发展造成了不利影响。因此，高效、精准地实施草原修复措施，以恢复其生态功能，已成为当前环境治理领域亟待解决的关键问题之一^[4]。

草原生态系统的健康状况直接关系到全球气候变化的缓解、生物多样性的维护以及区域经济的可持续发展。研究表明，草原作为重要的碳汇，在碳循环和气候调节方面具有不可替代的作用^[5-6]。同时，草原退化导致的生态系统服务功能下降，也对农牧业生产和当地居民生计产生了严重影响。因此，开展草原生态修复研究不仅具有重要的环境保护意义，也具有显著的社会经济价值。

在草原修复方法方面，传统手段主要依赖人工植被恢复、机械作业及定点喷洒草籽等方式。这些传统人工方法通常耗费大量人力物力，且难以在大范围草原环境下进行高效修复。机械化手段虽然提高了作业效率，但在复杂地形和偏远区域的应用受到诸多限制。近年来，无人机（Unmanned Aerial Vehicles, UAV）技术的快速发展为草原修复提供了一种全新的解决方案。无人机凭借其机动性强、作业成本低、可远程操作等优势，在草原生态监测、病虫害防治和生态修复等领域得到了广泛应用^[7]。然而，无人机在实际修复任务中仍面临诸多挑战，主要受到电池续航能力、载荷能力和复杂地形的限制。特别是如何在有限的能量和时间内覆盖更大的修复区域，以及如何优化无人机的路径规划以提高修复效率，成为亟需解决的技术难题。

1.2 国内外研究现状

在无人机辅助草原修复领域，国内外研究主要集中在两方面：无人机技术在生态监测与修复中的应用，以及路径优化与任务分配算法研究。国际上，Nex 等人^[8]综述了无人机在三维地图绘制和环境监测的应用优势，Mohan 等人^[9]和 Sun 等人^[10]将无人机技术应用于森林、草原等生态系统监测。Zeng 等人^[11]针对能量效率问题，研究了基于轨迹优化的无人机通信系统，为能源受限条件下的任务执行提供了新思路。

国内无人机技术在草原修复中也取得一定进展，特别是在播种、喷洒^[12]和生态监测等

应用场景。路径优化方面,传统方法主要基于旅行商问题和定向问题^[13],通过最优路径搜索减少飞行距离,提高执行效率。针对无人机巡逻、覆盖等任务,研究者提出了多种优化算法^[14-16]。人工智能方法虽在制造系统优化等领域有广泛应用^[17],但在草原修复领域的应用尚待深入探索。

在无人机播种系统方面,Elliott^[18]探索了低成本无人机在热带森林生态系统中通过空中播种实现加速自然再生的可能性。Pedersen 等人^[19]则利用无人机创建再播种地图,而非直接用于作物生产。Huang 等人^[20]提出了一种特殊的油菜籽无人机播种系统,采用微型气助式集中计量装置,提高了播种效率与精度。Faical 等人^[12]开发了一种能自适应规划飞行路线的系统,确保精确将农药沉积在目标区域。Guo 等人^[16]提出了喷洒无人机的喷雾分布模型和覆盖路径规划方法,仿真结果显示了其有效性。然而,这些研究主要基于规则田地地图,且未充分考虑无人机的能量消耗,也很少将路径规划与播种任务结合考虑,可能与实际应用场景存在差异。

在无人机路径规划研究中,研究者主要关注高效区域覆盖算法以延长无人机作业时间。Vasisht 等人^[21]设计了一种新颖的路径规划算法,充分利用风力辅助加速和减速,延长无人机操作时间并最小化覆盖给定区域所需的时间。Palomino 等人^[22]提出了与路径规划过程相关的自动化工作流程算法,该算法不仅能识别工作空间,还能在农业监督活动中处理路径规划。Shivgan 等人^[13]研究了无人机在环境传感和测量应用中完成任务的路径规划问题,将问题建模为旅行商问题 (TSP),通过优化无人机能量来处理无人机的飞行时间限制。Rossello 等人^[23]为精准农业开发了一种新型路径规划算法,覆盖大规模区域,考虑了飞行时间约束并最大化系统状态的估计质量,将问题建模为特殊的定向问题 (OP),即混合整数半定规划 (MISDP),通过启发式算法求解。

智能优化算法方面,Feng 等人探索了进化优化与迁移学习在复杂路径问题中的应用^[24],Rossello 等人提出了基于信息驱动的路径规划方法^[23]。多层次优化方法被用于任务分配与路径优化^[25],协同优化与分布式智能算法则应用于大规模组合优化问题^[26]。

在两阶段优化方法研究方面,Shen 等人^[27]考虑了大规模生物恐怖主义紧急情况下的两阶段车辆路径问题。Li 等人^[28]开发了一个不精确两阶段水管理模型,用于在不确定条件下规划农业灌溉。Azadeh 等人^[29]提出了一种两阶段路径优化方法,限制飞行边界并减少轻型飞机运输系统中的变量和约束数量。Maini 等人^[30]设计了一种两阶段策略,为覆盖应用中燃料约束下的协同空中-地面车辆路径规划问题寻找高效解决方案。Rajan 等人^[14]为数据收集任务场景中的无人机路径规划问题制定了两阶段随机规划模型。虽然上述研究主要采用两阶段方法处理路径规划问题,但大多数研究仅根据问题过程将其机械分解为两个单独解决的阶段,而没有考虑它们之间的耦合关系。

此外,双层规划问题^[25]是另一种两阶段优化问题。无人机草原修复问题与双层规划问题相关,它考虑了优化问题两个阶段之间的耦合关系。Angelo 等人^[31]研究了双层生产-分配规划问题。Sinha 等人^[32]全面回顾了从基本原理到解决方案的双层优化,包括经典和进

化方法。然而，这些方法难以直接解决本研究中的问题，因为它们未考虑无人机草原修复问题的特定特性。

然而，传统方法在面对动态环境、复杂地形和多目标任务时仍存局限性^[33]，通常基于静态环境优化，而实际应用常受风速、温度、地形等因素影响^[34]。单架无人机作业能力有限，难以满足大规模修复需求，因此多无人机协同作业成为提高效率的重要研究方向^[30,35]。虽已有研究探索多无人机系统的协同控制^[36]，但专门针对草原修复的多无人机协同优化研究仍较缺乏，特别是考虑能量约束、播种动态负载变化及修复面积最大化等多重目标的综合优化问题。

1.3 研究内容

本文旨在解决草原生态修复过程中面临的高效率覆盖与资源有限的矛盾问题。针对传统修复方法在大规模、复杂地形下的局限性，提出了一种基于深度强化学习的多无人机协同草原修复优化模型。研究采用"单机智能、多机协同"的分层设计思路：首先，利用强化学习算法训练单架无人机在复杂环境中的智能决策能力，使其能够自主学习并优化修复策略；其次，在此基础上构建多无人机协同作业的优化框架，实现在能源有限约束下的任务动态分配与路径协同优化。这种层次化的智能决策架构不仅能够根据草原退化状况、地形特征和无人机状态自适应调整修复策略，还能在多无人机间建立高效的信息交互与任务协调机制，从而在最大化修复面积的同时，提高了系统对复杂动态环境的适应性和鲁棒性。

第二章 多无人机协同的草原修复面积最大化模型

我们将待修复草原实例定义为一个完全无向图 $G = (V, E)$, 其中 $V = \{v_0, v_1, \dots, v_N\}$ 表示所有带修复区域的集合, 每个区域 v_i 具有位置坐标、退化程度和待修复区域面积等信息, 其中区域 v_0 定义为地面信息融合中心, 退化程度和待修复面积均为 0。 $E = (e_{ij} | i, j \in V, i \neq j)$ 表示所有边的集合, 每条边 e_{ij} 仅有边长这一属性。无人机 (多旋翼) 起飞时电池能量定义为 E_{max} , 携带的草种重量为 Q 。根据国际生态修复实践原则和标准^[33], 我们将待修复草原的退化程度范围定为 $[0.3, 0.8]$ 。

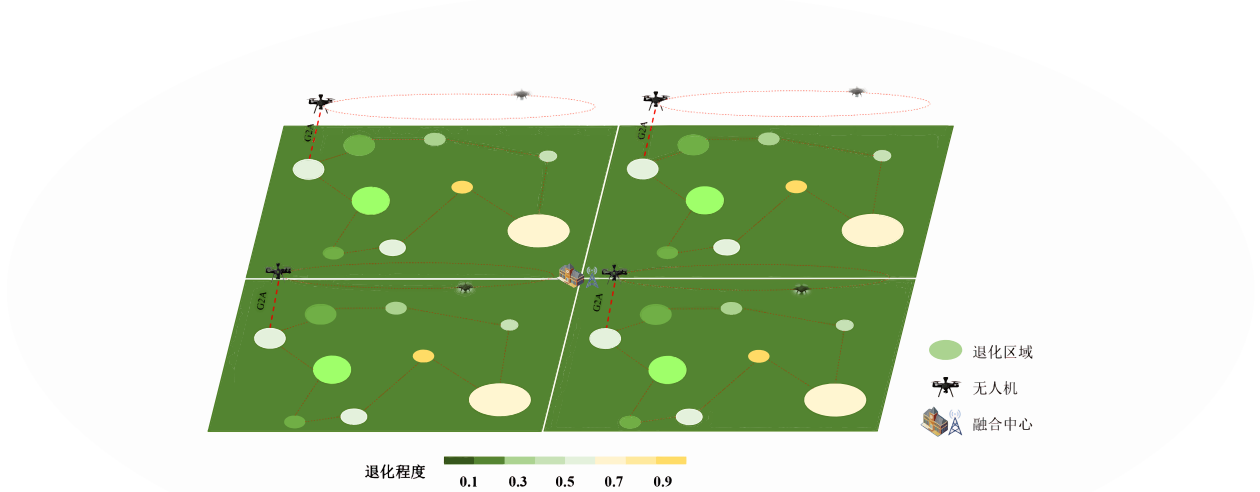


图 2.1 无人机草原修复区域示意图。图中展示了待修复区域的分布情况, 每个圆点代表一个待修复区域, 区域大小和颜色深浅表示其退化程度和面积。

针对待修复草原实例, 我们的目标是在无人机从基站出发并在为每个待修复区域提供服务前耗尽能量的情况下, 最大化无人机修复的总面积 (假设仅当无人机飞行至待修复区域时才能进行播种)。由于无人机在播撒草籽修复过程中自身重量不断变化, 其飞行能耗又与之直接相关, 如何权衡无人机飞行能耗与修复能耗之间的关系问题的核心所在。在修复过程中, 无人机的能量消耗主要包括三个部分: 无人机在待修复区域中播种的能量消耗、无人机进行航拍的能量消耗以及无人机在飞行过程中的能量消耗。相应地, 这三部分能量消耗可以分别表示为 E_s , E_{ap} 和 E_f 。此外, 无人机的能量消耗恒定速度下的载荷重量成正比, Dorling^[35] 等推导出 h 转子无人机的功耗方程, 表示如下:

$$P(\bar{q}_{ij}) = (M + \bar{q}_{ij})^{\frac{3}{2}} \sqrt{\frac{g^3}{2\rho\zeta h}}, \quad (1)$$

其中, $M = W + m$, W 表示无人机框架重量, m 表示其电池重量。 q_{ij} 表示无人机当前载荷重量 (草籽重量), g 为标准重力加速度, ρ 表示空气密度, ζ 表示无人机旋转叶片盘面积, h 表示无人机旋翼数量, 上述参数单位均参考国际单位制标准。为简化问题, 我们假设无人机在待修复区域间以固定高度、恒定速度飞行, 同时忽略不同天气条件对无人机飞行的影响, 如温度、风力、雨量和沙尘暴等。此外, 考虑到无人机能量的限制, 我们假设所有草原上的退化区域均可在一次修复过程中由一架无人机修复 (无论是否完全修复)。

2.1 各项能量约束

2.1.1 无人机飞行能耗

无人机飞行期间的能耗, 即单架无人机从基站出发, 根据修复方案遍历带修复区域最后返回基站的飞行路径能耗。我们设 $x_{ij} \in \{0, 1\}$ 为 0-1 决策变量, 定义如下:

$$x_{ij} = \begin{cases} 1, & (v_i, v_j) \text{ is covered in the tour,} \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

我们假定无人机在固定高度以恒定速度 v 飞行, 每单位距离的能耗相同。因此修复过程中无人机飞行的总能耗可以表示

$$E_f = \sum_{i=0}^N \sum_{j \neq i}^N e_{ij}^f d_{ij} x_{ij}, \quad (3)$$

其中, e_{ij}^f 表示无人机的飞行单位距离的能耗, 与无人机在待修复区域 v_i 和 v_j 之间携带的种子重量 \bar{q}_{ij} 相关, 可由公式 (1) 实时计算得出。 \bar{q}_{ij} 表示从待修复区域 v_i 和 v_j 中无人机携带的草种重量, 满足以下条件:

$$\sum_{j=0, i \neq j}^N \bar{q}_{ji} - \sum_{j=0, i \neq j}^N \bar{q}_{ij} = Q_i, \forall i \in V_a, \quad (4)$$

$$\bar{q}_{ij} \leq Q x_{ij}, \forall (i, j) \in A. \quad (5)$$

2.1.2 无人机播种能耗

我们将每个待修复区域离散为 $c_i (i = 1, \dots, n)$ 个单位圆面积。无人机播种的能量消耗可视为与草地退化程度和其携带种子的重量相关的函数。无人机每单位播种面积的能耗可以定义为: $e_i = \eta q_i$ 。其中 η 是一个与能量消耗相关的正参数, q_i 是第 i 个待修复区域每单位

圆修复所需种子的重量。进一步地，修复每个区域单位圆面积所需的种子重量可视为该区域的草地退化程度的函数^[37]： $q_i = (1 + l_i)\gamma$ 。其中， $l_i \in [0.3, 0.8]$ 表示第 i 个待修复区域的草地退化程度， γ 是与草地环境有关的正参数。综上所述，无人机 u 在第 i 个待修复区域中修复 σ_i 个单位圆所需的草种重量可以表示为 $Q_i^u = \sigma_i q_i$ ， σ_i 表示无人机修复的单位圆数量 ($1 \leq \sigma_i \leq c_i$)。因此，无人机 u 在待修复区域内播种的总能量消耗可以表示为：

$$E_s = \sum_{i=1}^N \sum_{j \neq i}^N \sigma_i e_i x_{ij}, \quad (6)$$

其中，二元 0-1 变量 x_{ij} 用于确定无人机是否将对待修复区域 v_j 进行修复。

2.1.3 无人机拍照能耗

除飞行能耗以外，无人机还在各待修复区域收集修复信息。无人机搭载高光谱相机进行航拍的总能耗可表示为：

$$E_{ap} = e_{ap} \sum_{i=1}^N \sum_{j \neq i}^N x_{ij} \sigma_i, \quad (7)$$

其中， e_{ap} 表示无人机在 σ_i 个被修复的单位圆中航拍所消耗的能量。

2.2 优化目标

由于能量有限，无人机在一次修复过程中无法修复草原上所有的退化区域。因此，我们的目标是通过无人机技术尽可能多地修复草原的退化面积，即优化目标为：

$$C = \sum_{i=1}^N \sigma_i \quad (8)$$

2.3 问题描述

综上所述，我们考虑在无人机的最大能量限制 E_{max} 下，对待修复区域的播种修复和航拍能耗 E_s 和 E_{ap} 、待修复区域间的飞行能耗 E_f 以及无人机的飞行轨迹 x_{ij} 进行组合优化。无人机通过将自身决策与彼此信息实时交流从而实现协同 (修复地图再分配等)。多无人机协同的草原修复模型可被描述为：

$$\max_{x_{ij}, \sigma_i} C = \sum_{i=1}^N \sum_{j \neq i}^N x_{ij} \sigma_i \quad (9a)$$

$$\begin{aligned} s.t. \quad & \sum_{i=1}^N \sum_{j \neq i}^N \sigma_i e_i x_{ij} + e_{ap} \sum_{i=1}^N \sum_{j \neq i}^N x_{ij} \sigma_i \\ & + \sum_{i=0}^N \sum_{j \neq i}^N (M + \bar{q}_{ij})^{\frac{3}{2}} \sqrt{\frac{g^3}{2\rho\zeta h}} x_{ij} d_{ij} \leq E_{max}, \end{aligned} \quad (9b)$$

$$\sum_{i=1, i \neq j}^N \sigma_i q_i x_{ij} \leq Q, \forall j \in V_a, \quad (9c)$$

$$\sum_{j=0, j \neq i}^N \bar{q}_{ji} - \sum_{j=0, j \neq i}^N \bar{q}_{ij} = \sigma_i q_i, \quad \forall i \in V_a, \quad (9d)$$

$$\bar{q}_{ij} \leq Q x_{ij}, \quad \forall (i, j) \in A, \quad (9e)$$

$$\sum_{i=0, i \neq j}^N x_{ij} = \sum_{j=0, j \neq i}^N x_{ij} = 1, \quad \forall i, j \in V_a, \quad (9f)$$

$$\sum_{j=1}^N x_{0j} = \sum_{j=1}^N x_{j0} = 1, \quad (9g)$$

$$x_{ij} \in \{0, 1\}, \quad (9h)$$

$$1 \leq \sigma_i \leq c_i, \quad (9i)$$

$$\bar{q}_{ij} \geq 0. \quad (9j)$$

上述优化模型中，约束条件(9b)确保无人机在一次调度周期内的能量消耗不超过其最大能量容量 E_{max} 。约束(9c)保证无人机携带的草种重量 Q 必须在无人机返回基站前全部播撒完毕。约束(9d)表示无人机在服务完一个修复区域后草种重量的减少，并与该区域的需求修复面积相匹配，同时消除任何非法的子回路。约束(9e)保证修复区域 v_j 的草种需求不超过无人机携带的剩余草种重量。约束(9f)确保无人机最多进入每个修复区域一次，并在播种后离开该区域。约束(9g)保证无人机的路径始于并终于基站。约束(9h)确保二元变量取整数值。约束(9i)确保待修复的单位圆数量不超过其最大区域。约束(9j)为非负性约束。综合而言，该问题是一个多变量组合优化问题，变量包括二元变量 x_{ij} 和整数变量 σ_i ，因此直接使用传统优化方法求解较为困难。

2.4 本章小结

本章提出并形式化了无人机修复模型，重点分析了飞行、播种、航拍三种能耗，并以修复面积最大化为目标，形成了多变量组合优化问题，为后续强化学习求解奠定基础。

第三章 基于强化学习的最优修复方案

3.1 马尔可夫过程建模

我们将修复过程中的每架无人机视为一个智能体，将待修复草原视为以欧几里得距离为边权值的无向全联通图。对单架无人机而言，其部分修复过程可视为从某点出发，遍历修复地图并完成相应修复任务，最后返回起点。

在无人机的修复过程中，由于其在两块相邻修复区域间的飞行能耗正比于此时无人机自身重量与路径长度，而自身重量又随着播散种子修复区域的过程动态变化，因而其总飞行能耗与总飞行长度、修复区域的顺序及修复面积均有复杂关系。换言之，作为优化目标的总修复面积由于与无人机耗能而与其飞行路径间接相关，因而二者之间存在复杂的耦合关系，如何解决修复面积的最优化，毫无疑问是一个困难的 NP-hard 问题。

针对该难题，我们拟采用强化学习方法学习这一复杂的函数关系，具体建模如下：

3.1.1 状态

本文将一个待修复草原实例设为 $V = \{v_i\}_{i=0}^n$ ，其中包含了各个待修复区域的位置、退化程度等各项信息。各点以 v_0 为起始点按照某种策略 π ，逐渐完善部分解 $\{(v_0, a_0), (v_i, a_i)\}_{i=1}^{uav_{now}}$ 。其中，编号 i 表示无人机访问各点的次序， uav_{now} 为无人机已经过的点数量， a_i 代表按照策略 π 在点 v_i 修复的面积。我们将该部分解作为智能体在强化学习中的状态，记作 $s(< i), i \in [1, n+1]$ 。初始状态记作 $s(< 1) = v_0$ ，表示无人机处于地面中心；结束状态记作 $s(< n+1) = s$ ，表示无人机已访问全部点。

3.1.2 动作

如上所述，无人机状态是修复过程的部分解，因而我们将动作设置为无人机修复过程某一步的解，即将无人机在状态 $s(< i)$ 时的动作记作 $\pi_i = s(i), i \in [1, n+1]$ ，即可实现强化学习中动作-状态对的自然更新。

3.1.3 状态

转移函数如上所述，在动作选取完毕后无人机新的状态也被唯一确定，因而状态转移函数具有完全确定性，记其为 $S(s(< i)|s(i)) = s(< i+1), i \in [1, n]$ 。

3.1.4 策略函数

如上所述，草原修复过程被创建为一个马尔可夫决策过程，因而其策略函数可被链式分解为： $p(\pi|s) = \prod_{i=1}^n p(s(i)|s(< i))$ (10) 其中， $p(s(i)|s(< i))$ 表示智能体在状态 $s(< i)$ 下选

择动作 $s(i)$ 的概率。

3.1.5 奖励函数

作为强化学习方法的核心之一，奖励函数的设计必须考虑全面，以防止模型难以收敛。经我们实验证明，单纯以最基本的修复面积作为模型的奖励函数会使得模型收敛速度大大减缓，同时考虑到无人机自身能量限制引入的惩罚项，我们最终将一组解的奖励函数设置为：

$$R(\pi|V) = \alpha_p * Pel + \alpha_r * \sum_{i=1}^n (l_i - 0.3) * a_i \quad (10)$$

$$Pel = \begin{cases} d_n^0 + \sum_{i=1}^n d_i^{i+1} & E_{rest} < 0 \\ 0 & E_{rest} \geq 0 \end{cases} \quad (11)$$

上式中， α_p, α_r 为修正系数， l_i 为第 i 个区域的退化程度， Pel 为控制无人机能量约束的惩罚项。通过引入 $(l_i - 0.3)$ 作为权重因子，模型可以优先修复退化程度更高的区域，从而实现对草原生态健康状况的精准干预。同时为了加快模型收敛速度，我们将不符合无人机能量约束时的惩罚项设置为其从起点到返回地面中心经历的路径长度，以确保在模型收敛的前期尽可能获得的解路径长度较短、符合能量约束，并在此前提下进行修复面积的决策。

3.2 神经网络模型构建

对于序列决策问题，编码器-解码器模型^[38]作为最经典的神经网络模型结构之一，取得了优异的效果。我们采用稍加修改的经典的 Tranformer 作为编码器，指针网络^[39]作为自回归解码器以实现构造式求解。

3.2.1 特征提取与模型架构

我们对待求解的待修复草原模型提取两部分特征：待修复区域的坐标、退化程度等静态特征，及无人机自身剩余能量、剩余重量等动态特征。模型参数用 $\theta \in \Theta$ 表示，包含编码器和解码器的所有可学习参数。

模型初始化过程如算法1的第 1-2 行所示：

$$M_\theta \leftarrow \text{InitializeModel}(d_e, d_h, n_{layers}, n_{heads}, \theta), \quad (12)$$

其中， d_e 为嵌入维度， d_h 为隐藏层维度， n_{layers} 为编码器层数， n_{heads} 为注意力头数量。同时，我们初始化基线模型 B_{model} 用于估计值函数，支持 Actor-Critic 算法的训练。

3.2.2 编码器结构

参考了 Bresson^[40]等人的方法，编码器由 n_{layers} 层 Transformer 层堆叠而成，采用 Batch-Norm 归一化，表示为：

$$h^{l=0} = h^{in} W^{in} \in \mathbb{R}^{n \times d_e} \quad (13)$$

$$h_{rc}^{l+1} = \text{BN} \left(\text{MHA}^{l+1}(h^l) + h^l \right) \quad (14)$$

$$h^{l+1} = \text{BN} \left(\text{Relu} \left(h_{rc}^{l+1} W_1^{l+1} \right) W_2^{l+1} + h_{rc}^{l+1} \right) \quad (15)$$

编码器同时处理静态环境特征和动态无人机状态特征,生成用于解码器的上下文表示。

3.2.3 解码器结构

自回归解码器通过循环解码逐步构造解决方案。解码器接收编码器的输出,其工作流程为:

$$rm^i, h_{gru}^i = \begin{cases} \text{GRU}(h_0^i, 0) & i = 0 \\ \text{GRU}(h_{i-1}^i, h_{gru}^{i-1}) & i > 0, \end{cases} \quad (16)$$

其中, 0 为全零向量, GRU(gated recurrent unit) 为循环神经网络更新解嵌入。解码器通过注意力机制选择下一个待修复区域及修复面积:

$$h_p^{l=i} = \text{Attention}(\text{static}, \text{dynamic}^i, mn^i) \quad (17)$$

$$\mu_i = \begin{cases} C \times \text{Tanh} \left((W_q h_p^{l=i})^T (W_k h_p^{l=i}) \right) & x_i \notin \pi_i \\ -\infty & \text{otherwise} \end{cases} \quad (18)$$

$$p(i|s_i) = \text{Softmax}(\mu_i) \quad (19)$$

在模型前向传播中,如算法2第4行所示,解码器生成成本值 C 和对数概率 $\log p$:

$$(C, \log p) \leftarrow M_\theta(x), \quad (20)$$

其中 C 表示当前策略的成本(负奖励), $\log p$ 表示选择各动作的对数概率。

3.2.4 训练流程

算法 1 Main Training Procedure for GRP

输入: $\theta \in \Theta$ (model parameters), $d_e \in \mathbb{N}$ (embedding dimension), $d_h \in \mathbb{N}$ (hidden dimension), $n_{layers} \in \mathbb{N}$ (encoder layers), $n_{heads} \in \mathbb{N}$ (attention heads), $\alpha \in \mathbb{R}^+$ (learning rate), $N_{epoch} \in \mathbb{N}$ (training epochs), $B \in \mathbb{N}$ (batch size), $G_{size} \in \mathbb{N}$ (graph size)

输出: $\theta^* \in \Theta$ (optimal parameters)

```

1:  $M_\theta \leftarrow \text{InitializeModel}(d_e, d_h, n_{layers}, n_{heads}, \theta)$ 
2:  $B_{model} \leftarrow \text{InitializeBaseline}()$ 
3:  $O \leftarrow \text{Adam}(\{M_\theta, B_{model}\}, \alpha)$ 
4:  $\mathcal{S} \leftarrow \text{LRScheduler}(\alpha)$ 
5:  $\mathcal{D}_{val} \leftarrow \{x_i\}_{i=1}^{val\_size}$  ▷ 生成验证数据集
6:  $R_{best} \leftarrow -\infty$  ▷ 最佳奖励追踪器
7: for  $e = 0$  to  $N_{epoch} - 1$  do
8:    $\mathcal{D}_{train} \leftarrow \{x_i\}_{i=1}^{epoch\_size}$  ▷ 生成训练实例
9:    $\{\mathcal{B}_j\}_{j=1}^{n_B} \leftarrow \text{BatchSplit}(\mathcal{D}_{train}, B)$ 
10:   $M_\theta \leftarrow M_\theta.\text{train}()$  ▷ 设置模型为训练模式
11:  for  $\mathcal{B} \in \{\mathcal{B}_j\}_{j=1}^{n_B}$  do
12:     $\theta \leftarrow \text{TrainBatch}(M_\theta, B_{model}, O, \mathcal{B})$  ▷ 算法 2
13:  end for
14:   $M_\theta.\text{decode\_type} \leftarrow \text{"greedy"}$ 
15:   $M_\theta \leftarrow M_\theta.\text{eval}()$ 
16:   $C_{val} \leftarrow \text{Evaluate}(M_\theta, \mathcal{D}_{val})$ 
17:   $\bar{r} \leftarrow -\frac{1}{|\mathcal{D}_{val}|} \sum_{i=1}^{|\mathcal{D}_{val}|} C_{val,i}$  ▷ 平均奖励
18:   $B_{model} \leftarrow B_{model}.\text{epoch\_callback}(M_\theta, e)$ 
19:   $\alpha \leftarrow \mathcal{S}(\alpha)$  ▷ 更新学习率
20:  if  $e \bmod k_{save} = 0 \vee e = N_{epoch} - 1$  then
21:     $\text{SaveCheckpoint}(\theta, O, B_{model}, e)$ 
22:  end if
23:  if  $e = 0 \vee \bar{r} > R_{best}$  then
24:     $\theta^* \leftarrow \theta$  ▷ 保存最佳模型参数
25:     $R_{best} \leftarrow \bar{r}$ 
26:  end if
27: end for
28: return  $\theta^*$ 

```

算法 2 TrainBatch Subroutine for GRP**输入:** M_θ (model), B_{model} (baseline model), O (optimizer), \mathcal{B} (batch data)**输出:** Updated model parameters θ

```

1:  $(x, b_v) \leftarrow B_{model}.unwrap\_batch(\mathcal{B})$ 
2:  $x \leftarrow x.to(device)$ ,  $b_v \leftarrow b_v.to(device)$  if  $b_v \neq \text{None}$  else None
3:  $(C, \log p) \leftarrow M_\theta(x)$  ▷  $C$ : 成本,  $\log p$ : 对数概率
4:  $r \leftarrow -C$  ▷  $r \in \mathbb{R}^B$ : 奖励
5: if problem_type = "grp" then
6:    $A_{repair} \leftarrow \begin{cases} x.total\_repair\_area & \text{if exists} \\ -C/\alpha_r & \text{否则 (近似值)} \end{cases}$ 
7: end if
8:  $(b_v, \mathcal{L}_b) \leftarrow \begin{cases} B_{model}.eval(x, C) & \text{if } b_v = \text{None} \\ (b_v, 0) & \text{otherwise} \end{cases}$ 
9:  $\mathcal{L}_R \leftarrow \frac{1}{B} \sum_{i=1}^B (C_i - b_{v,i}) \cdot \log p_i$ 
10:  $\mathcal{L} \leftarrow \mathcal{L}_R + \mathcal{L}_b$  ▷ 总损失
11:  $\nabla \theta \leftarrow \mathbf{0}$  ▷ 梯度清零
12:  $\nabla \theta \leftarrow \nabla_\theta \mathcal{L}$  ▷ 计算梯度
13:  $\|\nabla \theta\|_2 \leftarrow \min(\|\nabla \theta\|_2, g_{max})$  ▷ 裁剪梯度范数
14:  $\theta \leftarrow O(\theta, \nabla \theta)$  ▷ 更新参数
15: return  $\theta$ 

```

我们采用算法1和2所示的训练流程，将 Actor-Critic 方法与梯度下降相结合。训练过程中使用 Adam 优化器，并结合学习率调度器动态调整学习率：

$$O \leftarrow \text{Adam}(\{M_\theta, B_{model}\}, \alpha) \quad (21)$$

每个训练周期中，我们生成训练和验证数据集，并进行批次训练。在算法2中，策略梯度损失计算如下：

$$\mathcal{L}_R \leftarrow \frac{1}{B} \sum_{i=1}^B (C_i - b_{v,i}) \cdot \log p_i, \quad (22)$$

其中， $b_{v,i}$ 为基线模型对状态价值的估计， $(C_i - b_{v,i})$ 构成优势函数，用于指导策略更新方向。总损失函数包括策略梯度损失和基线损失：

$$\mathcal{L} \leftarrow \mathcal{L}_R + \mathcal{L}_b \quad (23)$$

参数更新通过梯度下降完成，并应用梯度裁剪以确保训练稳定性：

$$\|\nabla\theta\|_2 \leftarrow \min(\|\nabla\theta\|_2, g_{max}) \quad (24)$$

$$\theta \leftarrow \mathcal{O}(\theta, \nabla\theta) \quad (25)$$

训练过程中，我们保存性能最佳的模型参数 θ^* ，以便后续使用。模型评估采用贪婪解码策略，计算验证集上的平均奖励作为评估指标：

$$\bar{r} \leftarrow -\frac{1}{|\mathcal{D}_{val}|} \sum_{i=1}^{|\mathcal{D}_{val}|} C_{val,i} \quad (26)$$

通过这种训练流程，模型能够逐步学习优化多无人机草原修复的策略，实现修复面积最大化的目标。

3.3 本章小结

本章介绍了强化学习方法如何应用于草原修复决策，阐述了模型结构与 Actor-Critic 训练流程，为端到端自动决策奠定了技术基础。

第四章 多无人机协同调度算法

4.1 协同修复问题描述

在多无人机协同草原修复问题中，假设各无人机每轮巡航所携带的种子数量相同，表示为：

$$Q_u = Q, \quad \forall u \in U, \quad (27)$$

其中， U 表示无人机集合， Q 为初始携带种子重量。在修复过程中，无人机自身重量因种子播撒而减轻，导致飞行能耗动态变化。每架无人机已知自身修复地图的详细信息，包括退化区域数量、位置坐标、退化程度及待修复面积等信息。任意两点间距离采用二维平面内的欧式距离计算：

$$d(v_i, v_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (28)$$

4.2 协同调度机制

本文提出的多无人机协同调度算法通过局部信息与全局信息相结合的方式实现协同。对于每架无人机，我们定义：

$$M_u = \{v_1, v_2, \dots, v_n\}, \quad \forall u \in U, \quad (29)$$

其中， M_u 为无人机 u 的修复地图，包含 n 个待修复区域。每个区域 v_i 表示为四元组：

$$v_i = (id_i, (x_i, y_i), l_i, a_i), \quad i \in \{1, 2, \dots, n\}, \quad (30)$$

其中， id_i 为区域编号， (x_i, y_i) 为坐标位置， l_i 为退化程度， a_i 为可修复面积。

无人机状态集 S_u 定义为：

$$S_u = (p_u, E_u^{rem}, V_u^{vis}, A_u^{rep}), \quad \forall u \in U, \quad (31)$$

其中， p_u 表示无人机当前位置， E_u^{rem} 为剩余能量， V_u^{vis} 为已访问序列， A_u^{rep} 为已修复面积。

为模拟无人机与地面中心交互顺序，引入信号量 P_u 表示无人机完成首个待修复区域的时间：

$$P_u = \frac{d(p_u, v_i^{next})}{v} + \frac{a_i^{rep}}{r}, \quad (32)$$

其中, v 为无人机飞行速度, r 为单位时间内可修复面积, v_i^{next} 为下一个待修复区域。

4.3 动态地图更新

本文设计的多无人机协同算法核心在于动态调整修复地图, 通过适应度函数 $\Theta(u, v_i)$ 评估无人机 u 与待修复区域 v_i 的匹配度:

$$\Theta(u, v_i) = \alpha \cdot d(p_u, v_i) + \beta \cdot \frac{1}{l_i} + \gamma \cdot \frac{1}{a_i}, \quad (33)$$

其中, α 、 β 、 γ 为权重系数, 分别调节距离、退化程度和修复面积对适应度的影响。

多无人机协同调度过程可分为以下几个关键步骤:

1) 初始地图分配: 使用 K-means 等聚类方法将待修复区域分配给各无人机, 形成初始修复地图 M_u^0 。

2) 第一次路径规划: 各无人机以当前位置为起点, 计算最优路径及修复面积:

$$(P_u^1, A_u^1) = \arg \max_{P \in \mathcal{P}(M_u)} \sum_{v_i \in P} a_i^{rep}, \quad (34)$$

其中, $\mathcal{P}(M_u)$ 表示满足能量约束的所有可行路径集合, a_i^{rep} 为在点 v_i 的修复面积。

3) 地图动态更新: 地面中心根据各无人机当前状态, 更新全局修复地图:

$$M_u^{tmp} = \{v_i | v_i \in M^{global}, u = \arg \min_{u' \in U} \Theta(u', v_i)\} \quad (35)$$

4) 第二次路径规划: 无人机基于更新后的地图再次规划路径, 并计算修复面积 A_u^2 。

5) 地图选择决策: 比较两次规划的总修复面积, 选择更优方案:

$$M_u = \begin{cases} M_u^{tmp}, & \text{if } \sum_{u \in U} A_u^2 \geq \sum_{u \in U} A_u^1 \\ M_u, & \text{otherwise} \end{cases} \quad (36)$$

4.4 决策与执行流程

无人机根据信号量优先级依次执行决策。对于信号量最小的无人机 u^* :

$$u^* = \arg \min_{u \in U} P_u \quad (37)$$

该无人机在当前位置完成修复任务后, 前往下一个待修复区域:

$$v_i^{next} = \arg \min_{v_i \in M_{u^*}} \Theta(u^*, v_i) \quad (38)$$

同时更新状态信息:

$$S_{u^*} = (v_i^{next}, E_{u^*}^{rem} - E_{flight} - E_{repair}, V_{u^*}^{vis} \cup \{v_i^{curr}\}, A_{u^*}^{rep} + a_i^{rep}), \quad (39)$$

其中, E_{flight} 为飞行能耗, E_{repair} 为修复能耗, v_i^{curr} 为当前位置。

当某架无人机的修复地图为空时, 该无人机返回地面中心; 当全局修复地图为空时, 算法结束。完整的多无人机协同调度算法如算法 3 所示。

算法 3 多无人机协同调度算法

输入: 参数序列 $Parms$, 无人机修复地图集合 M_u , 无人机状态集合 S_u

输出: 无人机访问的节点序列 O_p , 无人机修复的面积 O_a , 无人机剩余能量 O_e

- 1: $M_u^i \leftarrow \text{Initial}(M_u)$ ▷ 无人机根据初始化方法 (如 K-means) 分配初始地图
 - 2: $P_u^i \leftarrow \text{Initial}(P_u)$ ▷ 初始化无人机信号量以决定决策优先级
 - 3: **while** $M_u \neq \emptyset$ **do**
 - 4: $E_u^{rel} \leftarrow \text{PlaningPath}(M_u, P_u^{self})$ ▷ 无人机群第一次路径规划
 - 5: $\text{SendToCenter}(S_u, M_u, E_u^{rel})$ ▷ 无人机第一次上报中心
 - 6: $M_u^{tmp} \leftarrow \text{updateMap}(M^{global}, P_u^{self})$ ▷ 中心更新地图
 - 7: $\text{RecvToUAV}(M_u^{tmp})$ ▷ 中心给无人机下发新地图
 - 8: $E_u^{r2} \leftarrow \text{PlaningPath}(M_u^{tmp}, P_u^{self})$ ▷ 无人机群第二次路径规划
 - 9: $\text{SendToCenter}(E_u^{r2}, Area_u^{r2})$ ▷ 无人机第二次上报中心
 - 10: **if** $\sum_{u=1}^U Area_u^{r2} \geq \sum_{u=1}^U Area_u^{r1}$ **then**
 - 11: $\text{RecvToUAV}(M_u^{tmp})$ ▷ 无人机选择修复面积更多的地图
 - 12: $M_u \leftarrow M_u^{tmp}$
 - 13: **end if**
 - 14: $\sigma_u^{max-p} \leftarrow \text{DecideArea}(E_u, M_u)$ ▷ 信号量最优先的无人机决策修复面积
 - 15: $\text{ActionUAV}(\sigma_u^{max-p}, C_{max-p}, P_u^{max-p})$ ▷ 无人机执行修复和信息采集
 - 16: $\text{DropPointFromMap}(M_u, P_u^{max-p})$
 - 17: $P_u^{max-p} \leftarrow \text{FlyToPoint}(P_u^{benefit})$ ▷ 无人机飞往下个最优点
 - 18: $\text{Update}(P_u^{max-p})$ ▷ 更新无人机信号量
 - 19: **end while**
 - 20: $\text{FlyToPoint}(P_u^0)$ ▷ 无人机返回起点
-

4.5 本章小结

本章给出了多无人机协同算法, 通过两次路径规划与地图更新实现任务分配动态调整, 兼顾局部优化与全局效益, 为后续实验提供了多无人机协同调度框架。

第五章 实验与分析

本部分我们展示了一些仿真实验的具体配置与结果。

5.1 仿真配置

表 5.1 实验硬件配置

配置	描述
CPU	AMD Ryzen Threadripper 3970X 32-Core Processor, 3.79 GHz
GPU	NVIDIA GeForce RTX 3090 Ti
RAM	ADATA 192GB-DDR4
OS	Ubuntu Server 22.04.3 LTS
Python	Python 3.9
PyTorch	version 1.12.1+cu113

5.2 实例设置

为验证所提出模型与算法的有效性，本文在六种不同规模的草原实例上进行了仿真，草原区域边长分别为 500 到 1000，步长为 100。所有实例均在二维空间内定义，任意两个待修复区域 v_i 与 v_j 之间的距离采用欧氏距离计算。每组实例中，待修复区域数量 N 分别设置为 60, 80, 100, 120, 140, 160，无人机数量 U 设置为 4, 6, 8，单架无人机的活动范围 L 与草原边长一致。所有待修复区域的坐标均在对应草原区域内随机生成，退化程度 l_i 服从 $[0.3, 0.8]$ 的均匀分布，基站均位于 $(0, 0)$ 。每个待修复区域的最大可修复单位圆数量 σ_i 随区域面积变化，范围为 10 到 35，步长为 5。无人机初始能量 E_{max} 与草原规模对应，具体参数如表 5.2 所示。

表 5.2 仿真实验实例设置参数表

参数	描述	值
N	待修复区域数量	60, 80, 100, 120, 140, 160
U	无人机数量	4, 6, 8
L	单无人机活动范围	500, 600, 700, 800, 900, 1000

5.3 算法对比设置

为了与研究热点中的最优化路径规划策略^[41]进行比较，我们设置了如下对比算法：

协同启发式与概率模型学习算法 (CHAPBILM)^[42]：该方法将问题分解为两阶段，第一阶段采用启发式算法（如 2-opt、or-opt 等）优化无人机的访问顺序，第二阶段利用概率模型学习（PBIL）动态分配各区域的修复面积，并结合最大剩余能量局部搜索（MRELS）进一步提升解的质量。该方法通过协同优化路径和修复面积，充分考虑了两者的耦合关系，提升了整体修复效果。

该算法可与多无人机协同调度算法结合，进一步验证我们提出方法的有效性和优越性。

5.4 参数设置

实验仿真具体参数如下。

表 5.3 实验软件配置

参数	描述	值
M	无人机重量	1.5
g	重力加速度	9.8
ρ	空气流体密度	1.024
ζ	无人机旋转螺旋桨面积	0.2
h	无人机旋翼数	6
γ	修复能耗正系数	2
η	草原环境相关的正系数	1×10^4
e_{ap}	无人机收集一个单位圆信息所需的能耗	2×10^4
E_{max}	无人机所携带的最大能量	3×10^6
B	强化学习模型训练时的图批次大小	512
N_{epoch}	强化学习模型训练时的回合数	512

5.5 仿真结果

下面我们展示本文所提出方法的仿真实验结果，包括训练过程中的损失函数和修复面积变化曲线，以及不同方法的路径规划可视化对比结果。

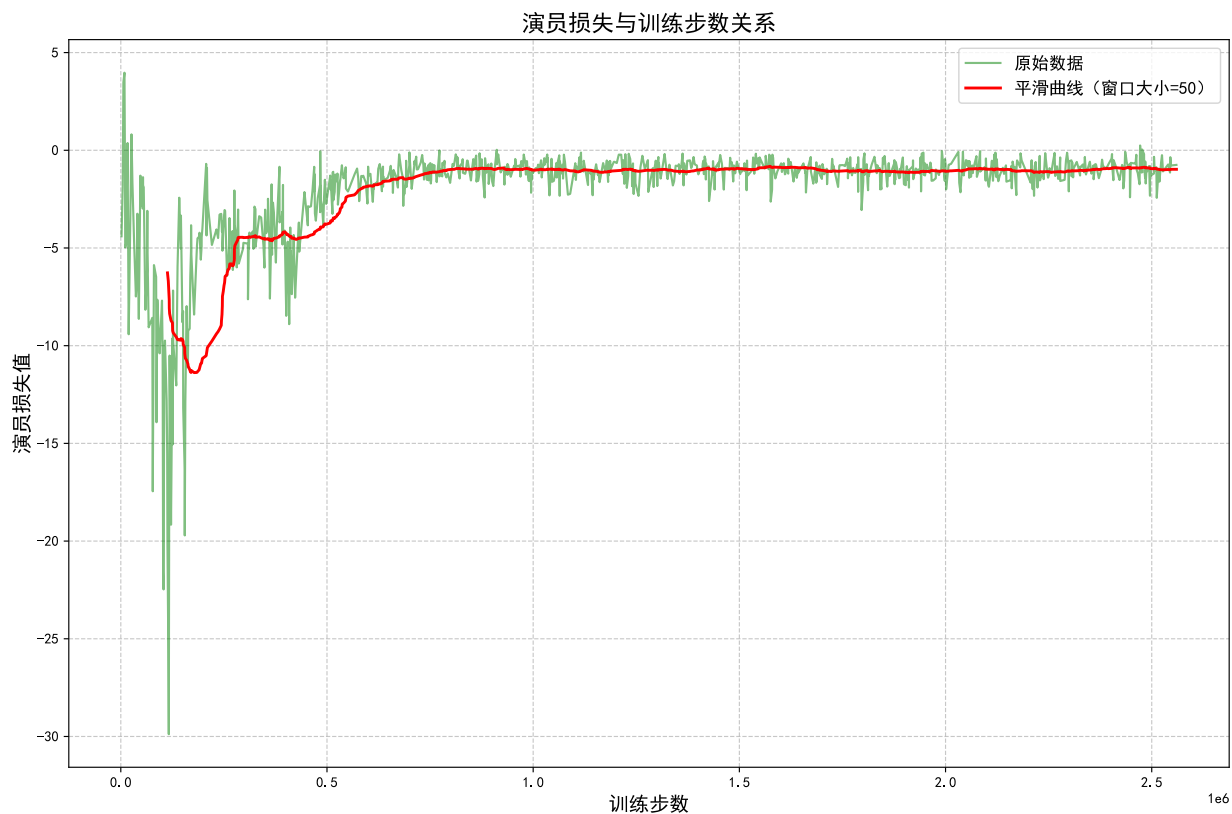


图 5.1 深度强化学习训练过程中的损失函数变化

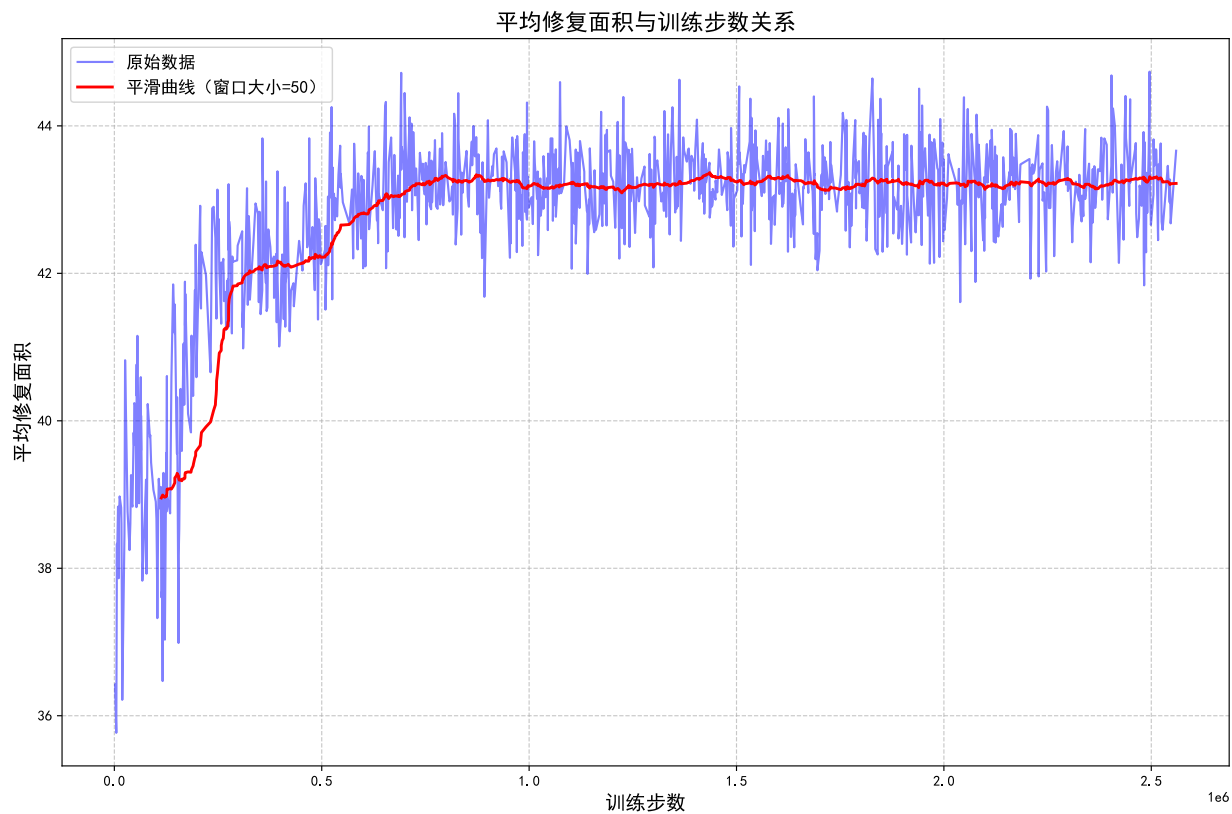


图 5.2 深度强化学习训练过程中的修复面积变化

如图5.1所示,该图展示了深度强化学习模型在训练过程中的损失函数随迭代轮数的变化趋势,可以看出损失值整体呈下降趋势,表明模型逐步收敛。从曲线可观察到,当训练步数达到约 500000 步后,损失函数趋于稳定,波动范围显著减小,表明模型已基本收敛。图5.2展示了训练过程中平均修复面积的变化曲线,随着训练的进行,平均修复面积逐步提升,说明模型的决策能力不断增强,能够获得更优的修复方案。同样,在约 500000 步后,修复面积也达到相对稳定状态,进一步证实了模型的有效收敛。这两个图共同验证了所提出深度强化学习方法在多无人机草原修复任务中的有效性和收敛性。

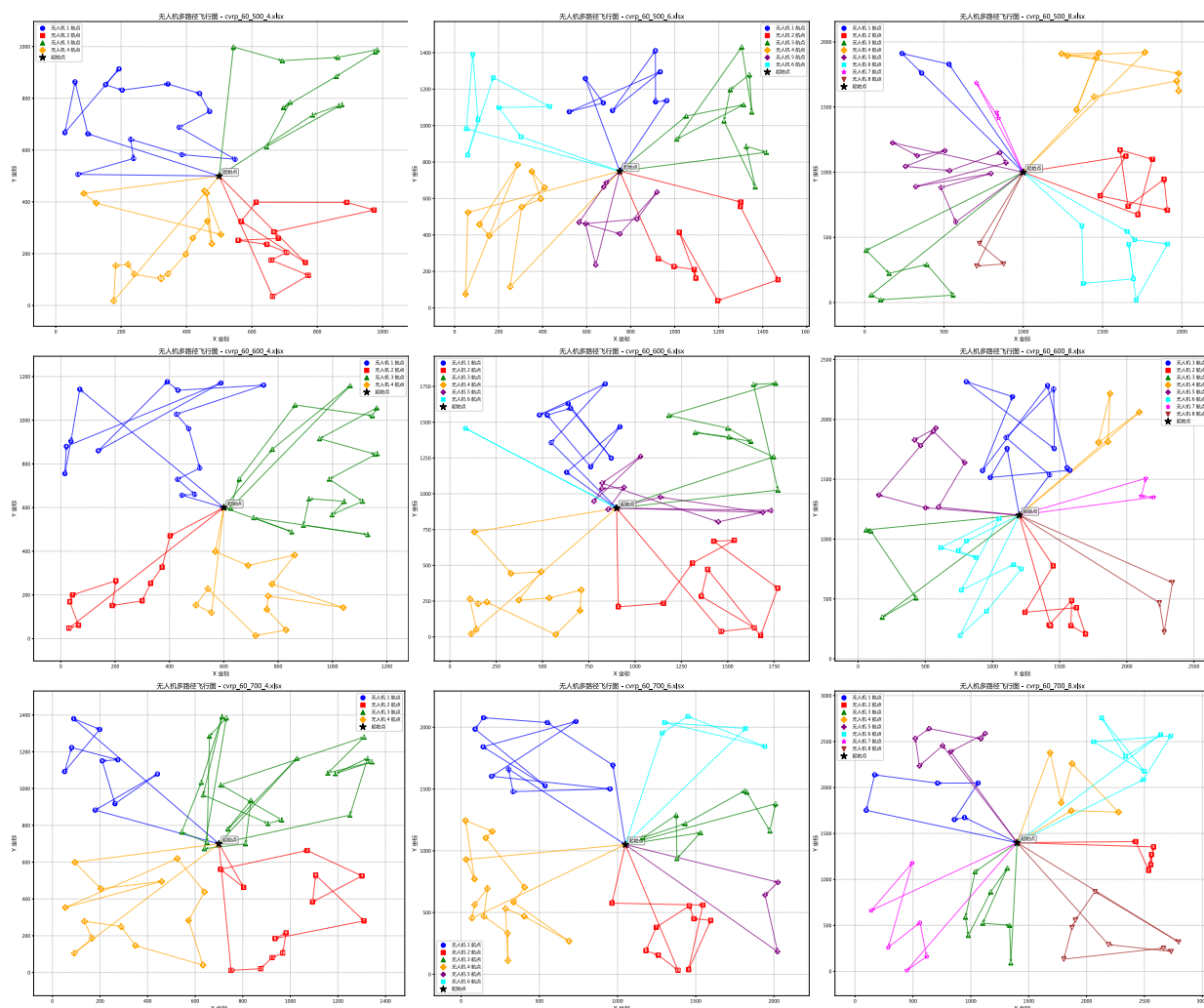


图 5.3 基于深度强化学习的多无人机草原修复路径规划

图5.3展示了基于深度强化学习方法的多无人机草原修复路径规划结果。可以看到,深度强化学习方法能够为每架无人机分配合理的修复区域,路径交叉较少,任务分配均衡,有效减少了无人机间的干扰。尤其在待修复区域分布密集时,算法能够根据能量约束和区域分布动态优化任务分配,使各无人机负载更为均衡。

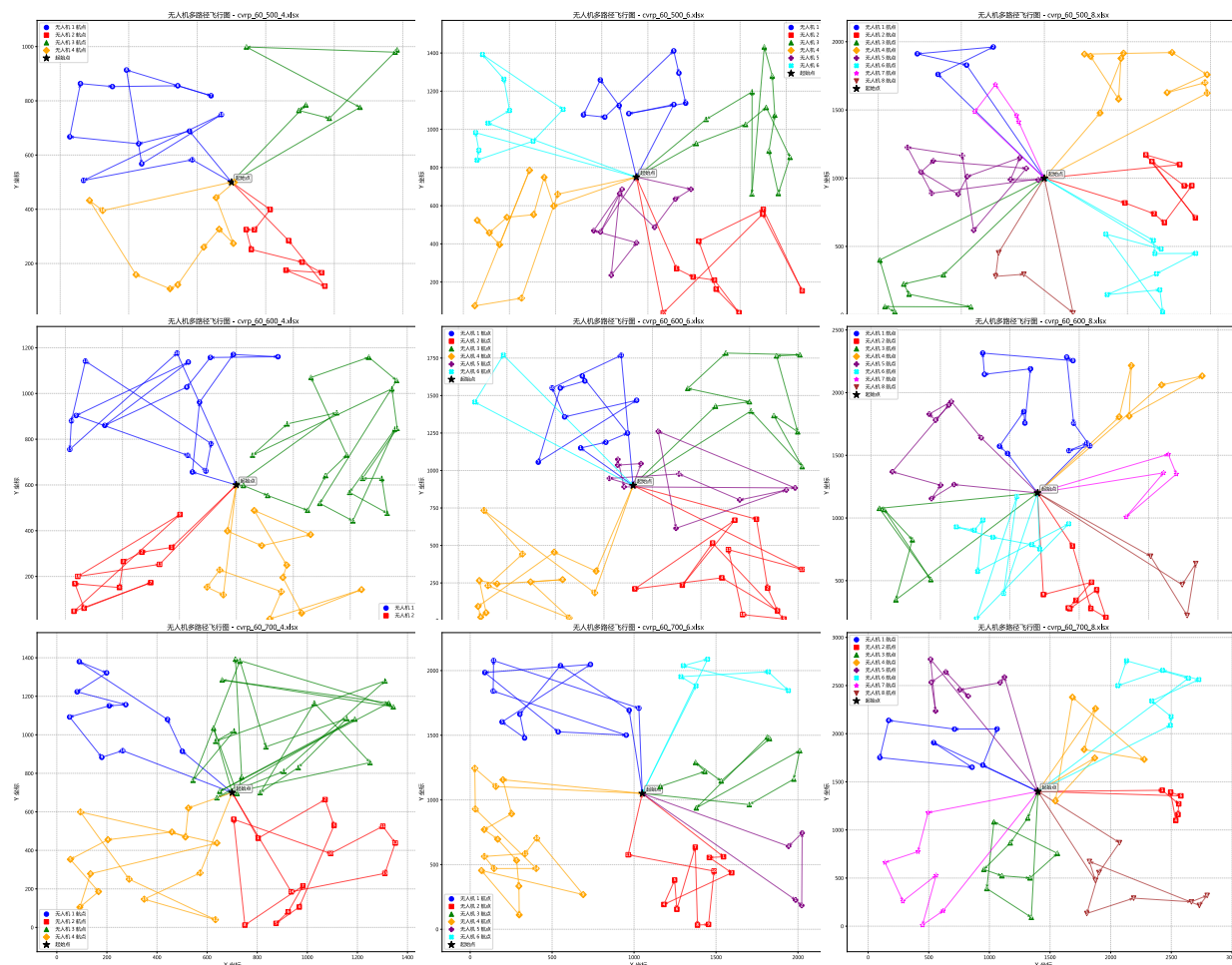


图 5.4 基于迭代局部搜索的多无人机草原修复路径规划

图5.4为迭代局部搜索方法的多无人机路径规划结果。与深度强化学习方法相比，迭代局部搜索在路径分配上表现出一定局限性，部分无人机路径存在较多长距离飞行和交叉，导致能量消耗增加，任务分配不够灵活。该方法在区域分配上较为机械，难以充分适应区域分布和能耗的动态变化。

通过对比两种方法的路径规划结果可以发现，深度强化学习方法在多无人机协同草原修复任务中表现出更强的自适应能力和优化效果，能够更好地协调任务分配和路径规划，提高修复效率和能量利用率。表5.4进一步量化了两种方法在路径长度和修复面积上的性能差异。

表 5.4 路径长度与修复面积对比 (CHAPBILM 与 DRL)

区域数	草原边长	无人机数	路径长度			修复面积		
			DRL	CHAPBILM	Gap(%)	DRL	CHAPBILM	Gap(%)
60	500	4	9396.35	12648.21	-25.68	267.00	194.00	37.63
		6	21348.69	19110.48	11.71	258.00	205.00	25.85
		8	26118.19	27145.66	-3.79	304.00	218.00	39.45
	600	4	13339.68	16255.96	-17.92	271.00	223.00	21.52
		6	26523.73	27174.68	-2.39	294.00	206.00	42.72
		8	31186.17	31494.13	-0.98	257.00	224.00	14.73
	700	4	17622.78	22103.46	-20.27	282.00	291.00	-3.09
		6	28252.23	28539.91	-1.01	270.00	218.00	23.85
		8	37585.42	36859.35	1.97	238.00	254.00	-6.30
	500	4	10221.88	17217.03	-40.60	370.00	315.00	17.46
		6	22135.76	25252.45	-12.37	375.00	240.00	56.25
		8	30491.72	31922.23	-4.48	306.00	267.00	14.61
80	600	4	15973.03	22825.91	-30.02	363.00	314.00	15.61
		6	29742.35	28044.73	6.05	406.00	282.00	44.00
		8	39943.98	40845.86	-2.21	399.00	269.00	48.33
	700	4	22577.68	26113.54	-13.56	335.00	267.00	25.47
		6	33566.63	34738.58	-3.38	380.00	298.00	27.52
		8	44308.24	42493.45	4.28	355.00	311.00	14.14

由表 5.4 可以看出, 深度强化学习 (DRL) 方法在大多数测试场景下均优于模拟退火 (SA) 算法。在路径长度方面, DRL 在多数情况下实现了更短的总路径, 部分场景下路径长度缩短超过 20%。在修复面积方面, DRL 也表现出显著提升, 提升幅度普遍在 15% 以上, 最高可达 56%。尤其在无人机数量较少或区域分布密集时, DRL 能更好地协调多无人机任务分配, 实现更高的修复效率。总体来看, DRL 在路径优化和面积最大化两个方面均优于传统启发式算法, 显示出其在复杂多无人机协同任务中的有效性和优势。

5.6 本章小结

本章通过实验验证了方法的有效性。结果表明, 深度强化学习在路径规划和修复面积方面优于传统算法, 展示了应用潜力。

第六章 总结与展望

6.1 总结

我们提出了结合深度强化学习方法求解的多无人机协同草原修复方法，利用深度强化学习的序列决策能力，构建编码器-解码器架构深度神经网络模型，使用强化学习 Actor-Critic 方法对模型进行训练，无需人为设计即可自动学习出优秀的策略。深度神经网络模型在编码器中的多头注意力层对待修复区域个体特征和待修复区域位置特征进行信息交互，通过自回归解码器输出下一步可能的修复方案。此外，本文还提出了多无人机协同的修复算法，在充分利用已有强化学习方法的基础上进一步优化求解效果。

实验结果表明，本文提出的算法优于基于贪婪策略的双层决策算法，且随着问题规模的增大优势更为明显；相较于贪婪策略，本文提出的单无人机训练、多无人机协同的策略具有更强寻优能力，在规定时间内进一步提升了求解效果，展示出神经网络算法强大的泛化与寻优能力，同时为求解大规模草原修复问题提供了新的思路和方法。

6.2 展望

在仿真结果与分析部分，由于项目时间限制，其中仍有众多方面可以值得扩展研究。例如：在给定的参数范围内，选择更多的启发式算法作为对比算法，从而能对控制算法的性能有更公正的评估。或在本文提出的控制算法框架下，通过控制变量法改变草场规模、无人机初始能量 E_{max} 等参数多次训练神经网络，而后通过大量的随机模拟取最佳、最差及平均等指标，来探讨算法的稳定性。

参考文献

- [1] Chapin F S, Sala O E, Huber-Sannwald E. Global biodiversity in a changing environment: scenarios for the 21st century: vol. 152[M]. Springer Science & Business Media, 2013.
- [2] Gibson D J. Grasses and grassland ecology[M]. Oxford University Press, 2009.
- [3] Bai Y, Ma L, Degen A A, et al. Long-term active restoration of extremely degraded alpine grassland accelerated turnover and increased stability of soil carbon[J]. Global Change Biology, 2020, 26(12): 7217-7228.
- [4] Freitag M, Klaus V H, Bolliger R, et al. Restoration of plant diversity in permanent grassland by seeding: Assessing the limiting factors along land-use gradients[J]. Journal of Applied Ecology, 2021, 58(8): 1681-1692.
- [5] Reinermann S, Asam S, Kuenzer C. Remote Sensing of Grassland Production and Management—A Review[J]. Remote Sensing, 2020, 12(12).
- [6] Dass P, Houlton B Z, Wang Y, et al. Grasslands may be more reliable carbon sinks than forests in California[J]. Environmental Research Letters, 2018, 13(7): 074027.
- [7] Steffen W, Richardson K, Rockström J, et al. Planetary boundaries: Guiding human development on a changing planet[J]. Science, 2015, 347(6223): 1259855. eprint: <https://www.science.org/doi/pdf/10.1126/science.1259855>.
- [8] Nex F, Remondino F. UAV for 3D mapping applications : a review[J]. Applied geomatics, 2014, 6(1): 1–1-2015.
- [9] Mohan M, Richardson G, Gopan G, et al. UAV-supported forest regeneration: Current trends, challenges and implications[J]. Remote Sensing, 2021, 13(13): 2596.
- [10] Sun Y, Yi S, Hou F. Unmanned aerial vehicle methods makes species composition monitoring easier in grasslands[J]. Ecological Indicators, 2018, 95: 825-830.
- [11] Zeng Y, Zhang R. Energy-Efficient UAV Communication With Trajectory Optimization[J]. IEEE Transactions on Wireless Communications, 2017, 16(6): 3747-3760.
- [12] Faıçal B S, Freitas H, Gomes P H, et al. An adaptive approach for UAV-based pesticide spraying in dynamic environments[J]. Computers and Electronics in Agriculture, 2017, 138: 210-223.

- [13] Shivgan R, Dong Z. Energy-Efficient Drone Coverage Path Planning using Genetic Algorithm[C]//2020 IEEE 21st International Conference on High Performance Switching and Routing (HPSR). 2020: 1-6.
- [14] Rajan S, Sundar K, Gautam N. Routing problem for unmanned aerial vehicle patrolling missions—a progressive hedging algorithm[J]. *Computers & Operations Research*, 2022: 105702.
- [15] Yang P, Tang K, Lozano J A, et al. Path planning for single unmanned aerial vehicle by separately evolving waypoints[J]. *IEEE Transactions on Robotics*, 2015, 31(5): 1130-1146.
- [16] Guo Y, Liu C, Coombes M. Spraying coverage path planning for agriculture unmanned aerial vehicles[C]//2021 26th International Conference on Automation and Computing (ICAC). 2021: 1-6.
- [17] Renzi C, Leali F, Cavazzuti M, et al. A review on artificial intelligence applications to the optimal design of dedicated and reconfigurable manufacturing systems[J]. *The International Journal of Advanced Manufacturing Technology*, 2014, 72(1): 403-418.
- [18] Elliott S. The potential for automating assisted natural regeneration of tropical forest ecosystems[J]. *Biotropica*, 2016, 48(6): 825-833.
- [19] Pedersen S M, Fountas S, Sørensen C G, et al. Robotic seeding: Economic perspectives[J]. *Precision Agriculture: Technology and Economic Perspectives*, 2017: 167-179.
- [20] Huang X, Zhang S, Luo C, et al. Design and Experimentation of an Aerial Seeding System for Rapeseed Based on an Air-Assisted Centralized Metering Device and a Multi-Rotor Crop Protection UAV[J]. *Applied Sciences*, 2020, 10(24).
- [21] Vasisht D, Kapetanovic Z, Won J, et al. Farmbeats: An IoT platform for data-driven agriculture[C]//14th USENIX Symposium on Networked Systems Design and Implementation NSDI 17. 2017: 515-529.
- [22] Palomino-Suarez D, Pérez-Ruiz A. Towards automatic UAV path planning in agriculture oversight activities[C]//Proceedings of the Latin American Congress on Automation and Robotics. Springer, 2019: 22-30.
- [23] Bono Rossello N, Carpio R F, Gasparri A, et al. Information-Driven Path Planning for UAV With Limited Autonomy in Large-Scale Field Monitoring[J]. *IEEE Transactions on Automation Science and Engineering*, 2022, 19(3): 2450-2460.
- [24] Feng L, Ong Y S, Tan A H, et al. Memes as building blocks: a case study on evolutionary optimization+ transfer learning for routing problems[J]. *Memetic Computing*, 2015, 7: 159-180.

- [25] Colson B, Marcotte P, Savard G. An overview of bilevel optimization[J]. *Annals of Operations Research*, 2007, 153(1): 235-256.
- [26] Huang X. Cooperative optimization for solving large scale combinatorial problems[G]// *Theory and Algorithms for Cooperative Systems*. World Scientific, 2004: 117-156.
- [27] Shen Z, Dessouky M M, Ordóñez F. A two-stage vehicle routing model for large-scale bioterrorism emergencies[J]. *Networks*, 2009, 54(4): 255-269. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/net.20337>.
- [28] Li W, Li Y, Li C, et al. An inexact two-stage water management model for planning agricultural irrigation under uncertainty[J]. *Agricultural Water Management*, 2010, 97(11): 1905-1914.
- [29] Sharif Azadeh S, Bierlaire M, Maknoon M. A two-stage route optimization algorithm for light aircraft transport systems[J]. *Transportation Research Part C: Emerging Technologies*, 2019, 100: 259-273.
- [30] Maini P, Sundar K, Singh M, et al. Cooperative aerial-ground vehicle route planning with fuel constraints for coverage applications[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2019, 55(6): 3016-3028.
- [31] Angelo J S, Barbosa H J C. A study on the use of heuristics to solve a bilevel programming problem[J]. *International Transactions in Operational Research*, 2015, 22(5): 861-882. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/itor.12153>.
- [32] Sinha A, Malo P, Deb K. A Review on Bilevel Optimization: From Classical to Evolutionary Approaches and Applications[J]. *IEEE Transactions on Evolutionary Computation*, 2018, 22(2): 276-295.
- [33] Gann G D, McDonald T, Walder B, et al. International principles and standards for the practice of ecological restoration. Second edition[J]. *Restoration Ecology*, 2019, 27(S1).
- [34] Blackburn R C, Barber N A, Farrell A K, et al. Monitoring ecological characteristics of a tallgrass prairie using an unmanned aerial vehicle[J]. *Restoration Ecology*, 2021, 29: e13339.
- [35] Dorling K, Heinrichs J, Messier G G, et al. Vehicle Routing Problems for Drone Delivery [J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017, 47(1): 70-85.
- [36] Buters T, Belton D, Cross A. Seed and seedling detection using unmanned aerial vehicles and automated image classification in the monitoring of ecological recovery[J]. *Drones*, 2019, 3(3): 53.

- [37] Klaus V H, Schäfer D, Kleinebecker T, et al. Enriching plant diversity in grasslands by large-scale experimental sward disturbance and seed addition along gradients of land-use intensity[J]. Journal of Plant Ecology, 2016, 10(4): 581-591. eprint: <https://academic.oup.com/jpe/article-pdf/10/4/581/19287948/rtw062.pdf>.
- [38] Vaswani A, Shazeer N, Parmar N, et al. Attention is All you Need[C]//Guyon I, Luxburg U V, Bengio S, et al. Advances in Neural Information Processing Systems: vol. 30. Curran Associates, Inc., 2017.
- [39] Vinyals O, Fortunato M, Jaitly N. Pointer Networks[C]//Cortes C, Lawrence N, Lee D, et al. Advances in Neural Information Processing Systems: vol. 28. Curran Associates, Inc., 2015.
- [40] Bresson X, Laurent T. The transformer network for the traveling salesman problem[J]. arXiv preprint arXiv:2103.03012, 2021.
- [41] Aggarwal S, Kumar N. Path planning techniques for unmanned aerial vehicles: A review, solutions, and challenges[J]. Computer communications, 2020, 149: 270-299.
- [42] Jiao D, Wang L, Yang P, et al. Unmanned Aerial Vehicle-enabled grassland restoration with energy-sensitive of trajectory design and restoration areas allocation via a cooperative memetic algorithm[J]. Engineering Applications of Artificial Intelligence, 2024, 133: 108084.

致 谢

在毕业论文完成之际，我想向所有在我学习和研究过程中给予帮助和支持的人表达我诚挚的感谢。

首先，我要衷心感谢我的导师焦栋斌教授。在整个研究过程中，焦老师以其渊博的学识、严谨的治学态度和敏锐的科研洞察力给予我悉心指导。他不仅在学术上为我提供了宝贵的建议和方向，而且在科研思维和方法论上也给了我深刻的启发。没有焦老师的悉心指导和鼓励，我难以完成这项研究工作。

特别感谢学长袁雨辰、刘波和王凌宇，本研究工作是在他们先前研究的基础上进行的改进和拓展。他们的开创性工作为我提供了重要的理论基础和技术支持，使我能够站在巨人的肩膀上继续探索。感谢他们在我研究过程中给予的宝贵建议、经验分享和无私帮助。

感谢课题组的其他成员。我们的头脑风暴和技术讨论不仅促进了研究的顺利进行，也让我对问题有了更加深入的理解。

感谢信息科学与工程学院的各位老师，在我本科学习期间传授知识、培养能力，为我打下了坚实的专业基础。特别感谢参与我论文评审和答辩的各位老师，你们的建议和意见使我的论文更加完善。

感谢实验室的师兄师姐和同学们，在日常学习和研究中的讨论和交流，让我受益匪浅。你们的陪伴和支持，使我的研究生活充满乐趣和动力。

感谢我的家人，是你们无条件的支持和理解，让我能够专心致志地投入学习和研究。你们是我坚强的后盾和不断前行的动力源泉。

最后，感谢所有在我成长道路上给予过帮助的人，正是因为有了你们的支持与鼓励，我才能顺利完成学业，迎接人生的新挑战。

毕业论文（设计）成绩表

导师评语

研究工作中，王贤义同学表现出了扎实的专业基础、良好的科研能力和创新思维。他能够独立思考问题，积极探索解决方案，并在算法设计和实验验证中付出了大量努力。论文写作规范，结构清晰，内容翔实，理论与实践相结合，充分体现了作者的学术素养。

本论文的研究成果对于提升草原生态修复效率具有重要参考价值，也为无人机协同任务优化领域提供了新的思路。建议作者在未来研究中进一步探索算法在更复杂环境下的适应性，以及与实际应用场景的结合。

综上所述，该论文达到了本科毕业论文的要求。

建议成绩 签名

指导教师（签字） 签名

答辩委员会意见

答辩委员会负责人（签字） _____

成绩 _____

学院（盖章） _____

年 月 日