



# 面向飞行器避障绕飞任务的思维链推理大语言模型微调方法

王瑞涛<sup>1</sup>, 李博研<sup>2</sup>, 石晓荣<sup>2</sup>, 文永明<sup>2</sup>, 程荣<sup>1</sup>, 余跃<sup>2</sup>, 张宁宁<sup>2</sup>, 汤宏垚<sup>1</sup>, 郑岩<sup>1</sup>, 郝建业<sup>1\*</sup>

1. 天津大学智能与计算学部, 天津 300354

2. 北京控制与电子技术研究所信息系统工程重点实验室, 北京 100038

\* E-mail: [jianye.hao@tju.edu.cn](mailto:jianye.hao@tju.edu.cn)

收稿日期: 2025-08-14; 接受日期: 2025-12-18; 网络版发表日期: 2026-01-16

国家自然科学基金项目(批准号: 62422605, 62533021, 92370132)和国家科技部重点研发项目(编号: 2024YFE0210900)资助

**摘要** 飞行器在复杂环境中的自主路径规划是保障航行安全和任务成功的关键, 这要求规划系统必须同时满足实时避障与战略导航的双重要求。然而, 现有方法在应对动态障碍物分布、环境不确定性以及现实约束时, 通常在鲁棒性、泛化能力和计算效率等方面存在不足。针对这一问题, 本文提出了一种两阶段的大语言模型(large language model, LLM)驱动的路径规划方法, 充分利用了LLM的空间推理与泛化能力。在第一阶段, 无需依赖传统的建图与搜索模块, 直接通过结构化环境描述和任务提示模板引导模型生成可行的绕飞路径; 在第二阶段, 构建高质量的思维链绕飞推理路径样本, 并通过有监督微调提升模型对空间结构的理解和路径输出的稳定性, 实现从通用模型到任务专用模型的平滑过渡。实验结果表明, 在高复杂度场景下, 该方法在绕飞成功率、路径合理性以及未知环境适应性方面均表现出显著提升, 这不仅验证了LLM在结构化空间任务中的可行性与优势, 还为现实约束条件下的智能航空导航提供了新的解决方案。

**关键词** 避障绕飞, 大语言模型, 监督微调训练, 路径规划

## 1 引言

在飞行器自主导航领域, 避障与绕飞路径规划是确保系统安全性与任务成功率的核心问题。该问题广泛存在于各类应用中, 包括无人机、城市低空交通、地形复杂区域巡航等。在这些应用中, 飞行器通常需要在不确定、动态变化的环境中快速感知周围障碍物, 并自主规划一条避障且高效的飞行路径, 这就需要

系统具备实时性、鲁棒性和良好的泛化能力<sup>[1]</sup>。随着飞行器应用场景的不断扩展和环境复杂度的显著提升, 导航系统不仅需要处理多源异构的信息输入, 还要在不可预知的障碍物分布、环境扰动以及外部限制(如空域管制、通信中断等)条件下, 实时做出合理的决策和响应<sup>[2,3]</sup>。

上述挑战对传统路径规划算法提出了更高的要求, 使得单纯依赖规则驱动或经典启发式方法已难以

引用格式: 王瑞涛, 李博研, 石晓荣, 等. 面向飞行器避障绕飞任务的思维链推理大语言模型微调方法. 中国科学: 技术科学, 2026, 56: 342–355  
Wang R T, Li B Y, Shi X R, et al. Fine-tuning of large language models for obstacle circumvention in path planning via chain-of-thought reasoning (in Chinese). Sci Sin Tech, 2026, 56: 342–355, doi: [10.1360/SST-2025-0219](https://doi.org/10.1360/SST-2025-0219)

满足实际需求. 典型方法如A\*搜索、RRT\*以及Dijkstra等算法, 能够在静态环境下较快生成可行路径<sup>[4,5]</sup>. 然而, 这些算法通常基于栅格地图或图结构构建路径, 对环境结构缺乏深层理解. 在障碍物分布密集或动态变化的场景中, 算法容易陷入局部最优, 且计算代价显著上升, 导致规划效率下降, 难以满足高动态飞行的实时性需求<sup>[6,7]</sup>. 近年来基于深度强化学习的路径规划方法开始受到关注. 这类方法通过与环境交互学习避障策略, 理论上具有更强的灵活性和泛化能力. 但在实际应用中, 强化学习方法面临两个核心问题: (1) 依赖大量高质量仿真数据和长时间的训练过程, 导致开发成本高; (2) 在从仿真向真实系统迁移时, 模型容易遭遇性能退化, 缺乏稳健性和可靠性, 难以应用于高风险任务场景<sup>[8~10]</sup>.

与此同时, 以大语言模型(large language model, LLM)为代表的人工智能技术迅速发展, 展现出强大的认知推理能力, 已被广泛应用于任务规划、自动驾驶决策等需要高层推理的场景<sup>[11~13]</sup>. 然而, 在路径规划领域, 现有工作对LLM空间推理能力的利用仍处于浅层. 方法上或受限于栅格搜索等对分辨率敏感的环境建模<sup>[14]</sup>, 或依赖脆弱的提示工程<sup>[15]</sup>, 极易导致规划出现幻觉与路径矛盾. 因此, 如何将LLM的空间推理能力系统性地嵌入路径规划中, 仍是一个值得深入探索的课题.

为解决上述问题, 本文提出了一种两阶段的大语言模型驱动的路径规划方法, 面向飞行器在复杂环境中的绕飞任务, 充分挖掘LLM的空间推理与泛化潜力<sup>[16,17]</sup>. 在第一阶段, 我们设计了一套基于大语言模型的路径规划方法, 探索如何利用现有语言模型的推理与表达能力, 引导其生成可行的飞行路径. 该阶段重点在于任务建模与Prompt工程, 通过设计结构化的环境描述方式和规划任务的提示模板, 使大模型能够理解障碍环境与任务目标, 从而完成路径生成<sup>[18,19]</sup>. 这一阶段不依赖传统的导航模块, 而是充分挖掘大模型在无需显式空间建图前提下的通用推理能力. 在此基础上, 为进一步提升路径规划的质量、稳定性以及对复杂环境的适应能力, 第二阶段引入了有监督训练机制. 我们构建了一个自动化的训练数据生成流程, 结合多样化场景生成、路径样本构建与标签提取, 产出高质量的路径数据. 随后, 采用有监督微调的方法对大模型进行能力增强, 使其对空间结构的理解更加

深入、路径输出更加稳定, 从而提升系统在复杂环境下的表现<sup>[20~22]</sup>. 该两阶段设计实现了从通用模型向任务专用模型的平滑过渡, 同时保留了大语言模型在泛化性与灵活性方面的优势.

为验证方法的有效性, 我们在多个高复杂度的场景中开展了系统性的实验. 实验结果表明, 我们的模型在绕飞成功率、路径合理性以及对未知场景的适应能力方面均取得了显著提升.

本文的贡献主要有以下四个方面:

(1) 提出了一种基于思维链微调的LLM路径规划框架, 克服了传统算法对栅格分辨率等环境建模参数的依赖性, 大幅提升了泛化能力;

(2) 设计了一套融合多样化场景生成与规则法引导的监督数据构建方案, 实现了高质量数据的高效制备;

(3) 训练出具备强空间认知与推理能力的路径规划大模型, 所引入的思维链技术有效抑制了模型幻觉与输出不一致问题;

(4) 通过大量实验验证, 我们的模型在多个核心评估指标上都有全面的提升.

## 2 面向飞行器避障绕飞的规划大模型

本文提出了一种面向飞行器避障绕飞的规划大模型全流程实现方法(图1), 实现了场景数据构建、大模型绕飞规划构建以及模型训练的完整闭环. 上述方法主要涉及3个部分: (1) 自动化构建标准化避障场景数据; (2) 基于大语言模型的绕飞规划系统; (3) 大语言模型的监督微调训练. 下面将逐步阐述上述部分.

### 2.1 自动化场景构建

在该小节中, 我们首先将飞行器任务参数和环境约束等参数统一整理为标准化结构化场景格式. 之后通过“先定线后定障”的摆放法和目标导向的A\*搜索法自动生成避障场景数据, 其中摆放法基于航路点几何特征动态部署规避区从而生成场景, 目标导向法则通过路径搜索生成最优路径标签.

**场景标准化定义** 针对当前飞行器避障绕飞场景中存在的要素混杂、描述不一致问题, 本文提出标准化场景定义格式, 明确以下核心要素: 起始位置、起始速度、目标位置、规避区(圆形区域, 包含编号、坐

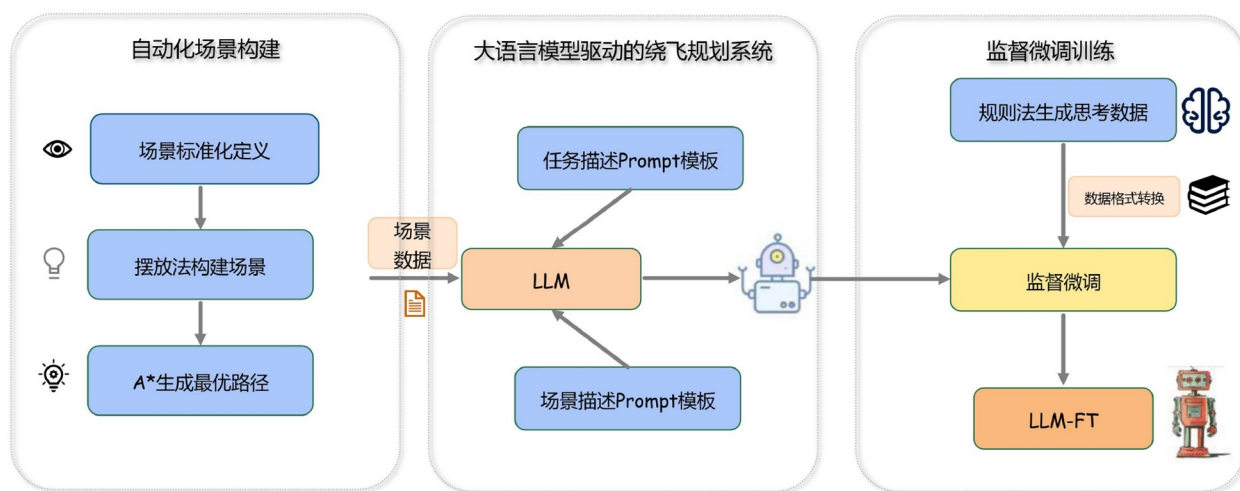


图1 本文提出的框架概览

Figure 1 The overview of our proposed framework.

标和半径)、绕飞的战略级规划以及绕飞的航路级规划,并且以JSON格式存储原始坐标.其中,绕飞的战略级规划指宏观的绕飞策略,例如确定飞行器从规避区的左侧或右侧绕行;航路级规划则指从起点到终点实现绕避障碍的关键航路点序列。“strategy”对应战略级规划,“trajectory”对应航路级规划.

**摆放法构建场景** 在完成场景要素标准化定义后,我们可以使用多种方法来构建场景,但是为了使得构建的场景更加多样化、真实化,我们提出了“先定线后定障”的摆放法来构造场景数据.首先生成三个航路点:从起始点开始,逐个采样航路点,先随机采样 $y$ 坐标,再采样 $x$ 坐标,且需满足后一个点与前一个点的 $x$ 坐标差值小于阈值,若目标点位置不满足条件,则重复采样.之后进行航路线的生成:基于采样的航路点,使用scipy库插入生成密集平滑的路径.最后是规避区的摆放:对每个航路点判断是否放置规避区,若三个航路点中有明显拐点(即航路点 $p$ 到直线两端点的距离大于阈值),则放置规避区.摆放方式有两种思路:思路1(图2)是使规避区的 $y$ 与航路点 $y$ 接近, $x$ 与与航路点 $x$ 符号相反的值( $\text{offset} = -1.2x$ ),并基于到路径上所有点的最短距离确定半径 $r$ ;思路2(图3)是计算特定直线交点( $y_1 = p - y$ 与 $y_2 = p_0 - p_2$ ,  $p$ 为航路点,  $p_0$ 为航路点 $p$ 前一个点,  $p_2$ 为航路点 $p$ 后一个点)并加随机扰动确定规避区坐标,再基于最短距离确定半径 $r$ .

**A\*生成最优路径** 通过“先定线后定障”的摆放

法可以得到包含多个避障区的场景数据,但是摆放法中的避障路径不一定是最优路径,为此我们需要用A\*启发式算法来搜寻最优路径(图4).首先进行输入处理,从场景数据中提取起点、终点和规避区;之后利用A\*算法搜索路径,使用8邻域网格在二维平面中搜索路径,计算每个候选点的实际代价和启发代价(欧几里得

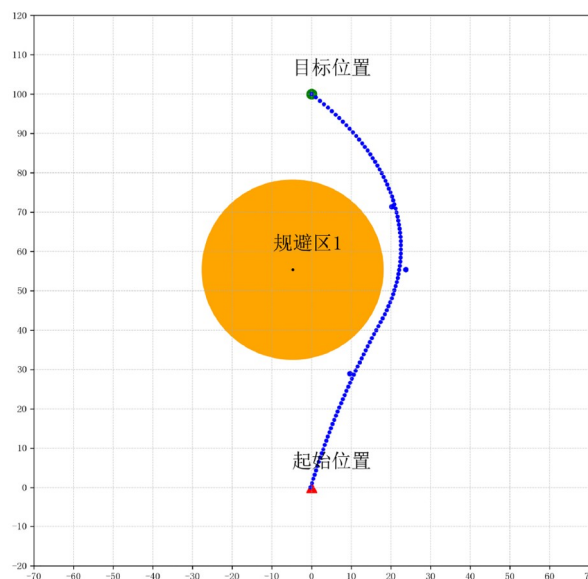


图2 基于航路点镜像偏移与最短距离半径的规避区摆放示意图(思路1)

Figure 2 Schematic diagram of avoidance zone placement based on waypoint mirroring offset and shortest-distance radius (approach 1).

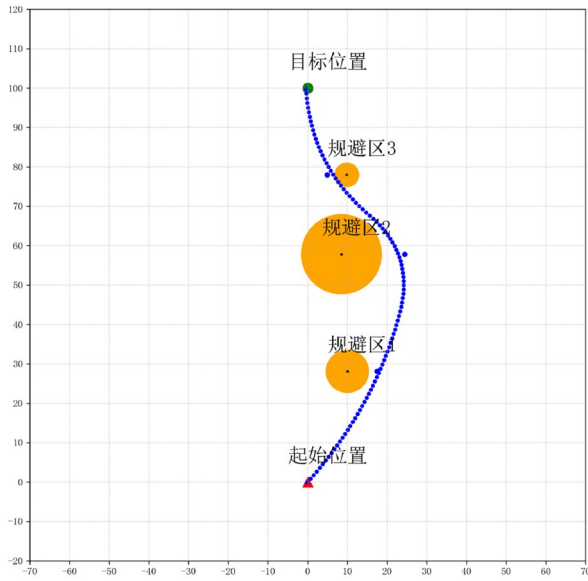


图 3 (网络版彩图)基于直线交点计算与随机扰动的规避区摆放示意图(思路2)

Figure 3 (Color online) Schematic diagram of avoidance zone placement based on line intersection calculation and random perturbation (approach 2).

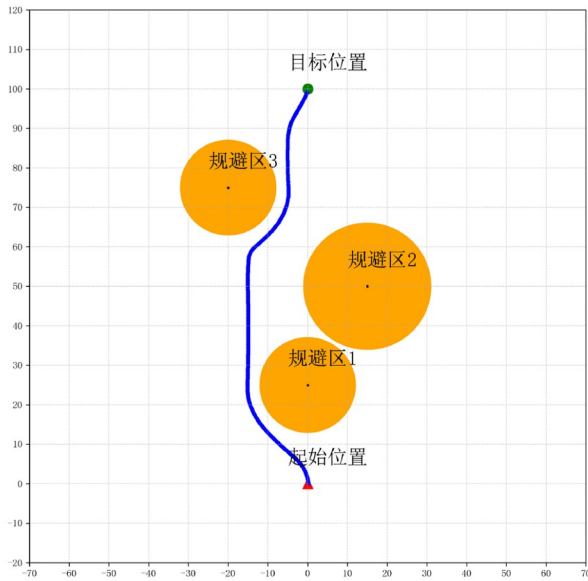


图 4 (网络版彩图)场景A\*生成最优路径效果图

Figure 4 (Color online) Rendering of the optimal path generated by A\*.

距离), 若路径穿越规避区则跳过该点, 靠近目标即认为到达并回溯生成原始路径; 最后还需要进行路径平滑处理, 使用B样条插值对A\*生成的折线路径进行平滑处理, 以便获得自然的绕飞路径。

## 2.2 基于大语言模型的绕飞规划系统

在本小节中, 我们构造了任务描述Prompt模板以及实时填入场景数据的Prompt模板, 继而开发出了基于大语言模型的绕飞规划系统, 该系统将利用LLM的推理与规划能力, 生成飞行器避障绕飞的战略级规划和航路点序列。

**任务描述Prompt模板** 任务描述模板用于定义LLM的推理逻辑和输出规范, 确保生成的绕飞规划方案符合系统要求(图5)。该模板包含以下核心要素: (1) 任务背景, 明确飞行器需从起始点飞至目标点, 并以第一视角进行路径规划; (2) 分析要求, 要求模型先进行战略级路径分析, 再输出精确航迹点; (3) 输出规范, 强制以JSON格式返回绕飞策略和航迹点序列, 其中航迹点数量严格限制为5个。

**场景描述Prompt模板** 场景描述为系统提供实时环境信息, 通过动态注入方式与任务描述模板结合形成完整输入(图6)。场景描述模板包含两个核心要素: 首先是飞行任务的起止位置坐标, 其将作为航迹点的起点和终点; 其次是各规避区的详细参数, 包括中心坐标和半径, 用于计算安全绕飞路径。

**输出格式** 正如在任务描述模板中规定的那样, 我们要求LLM以JSON格式输出绕飞结果, 其中strategy字段记录规避区编号及绕飞方向(左/右), trajectory字段按顺序列出航路点坐标(包含起始点), 形成从起点到终点的关键点坐标序列。详细输出如图7所示。

## 2.3 监督微调训练

我们发现, 现有的LLM不擅长空间决策能力, 在绕飞规划任务上表现较差, 因此如何构建训练数据从而引导LLM学会绕飞规划成为核心突破点。在该小节中, 我们首先通过规则法构建了数据的思考过程, 之后对数据格式进行了特定格式上的转换, 最后利用构造的监督微调数据在Base模型上进行了监督微调训练。

**基于规则法的Think内容生成** LLM的思维链(chain of thought, CoT)机制能够引导模型在复杂任务中生成详细的推理过程, 从而提高模型的准确性和可解释性。因此, 我们需要在训练数据中加入“Think”过程, 以充分激发并利用模型的推理能力。具体而言, 我们将使用规则法来生成Think内容。首先复述飞行器与

## 任务描述Prompt模板

```
### 任务背景
你正面临一个飞行器决策规划的场景。飞行器需要从起始位置飞到目标位置，但是中途可能存在若干个规避区，因此需要你帮助飞行器
绕过这些规避区。注意：在任何时候，你的分析和推理都应该基于飞行器的第一视角进行判断。

### 任务描述
现在，你开始你的分析和策略选择。先进行分析，给出大致路径描述，然后用坐标的形式表示路径。在分析最后，你应以 json 格式总结最
终规划，并用 ```json 和 ``` 包裹，json 具体格式如下：

```json
{
  "strategy": [ # 绕飞策略
    [str, str], # 例如: ["1", "1"] 表示第1个规避区从左侧绕行
  ],
  "trajectory": [ # 航迹点
    [int, int], # 起始点
    ...
    [int, int] # 目标点
  ]
}
```

为了提高规划效率，请控制最终方案的航迹点数量为 5(包括起始点和目标点)。避免不必要的冗长路径规划，确保路径简洁且规避规避。
```

图 5 任务描述Prompt模板

Figure 5 Prompt template for the task description.

## 场景描述Prompt 模板(注入后)

```
###当前场景
整个场景图的详细描述如下（从俯视图的角度描述）：
起始位置坐标：(0,0)
目标位置坐标：(0,100)
规避区范围：(x, y, r)，其中 x 和 y 是规避区的中心坐标，r 是规避区的半径：
规避区 1: (3, 24, 18)
规避区 2: (0, 55, 0)
规避区 3: (3, 77, 15)
```

图 6 场景描述Prompt模板(注入后)

Figure 6 Prompt template for the scene description (after injection).

目标点信息，梳理规避区数量、坐标分布等场景态势；随后基于贪心策略按序处理每个规避区：以当前位置为基准，优先选择最近的未处理规避区，通过分析距离和方位信息来判断绕飞的必要性。若需绕飞，则先通过向量运算确定绕飞方向，接着计算两个绕飞点。第一个绕飞点从规避区圆心垂直于路径的方向上取安全的点；第二个绕飞点将从第一个绕飞点出发，沿着当前路径方向飞行一定距离，确保绕过该规避区后进行选取。当前规避区绕飞后，再以新位置为起点迭代处

理后续规避区。

**SFT数据格式转换** 经过规则法构建监督微调数据后，我们需要对数据格式进行转换，最终的数据格式模板如图8所示，其中{query}为2.2节中任务描述Prompt和场景描述Prompt相结合的内容，{thinking\_content}为本小节中通过规则法构建的Think内容，{response}为2.1节通过A\*方法所生成的最优路径标签(战略级和航路点)。最终用于SFT训练和测试的数据量如表1所示。

```

'''json
{
  "strategy": [
    [str, str], # 例如: ["1", "1"]
    ...
  ],
  "trajectory": [ # 轨迹点
    [int, int], # 起点
    [int, int], # 中间点1
    [int, int], # 中间点2
    [int, int], # 中间点3
    [int, int] # 目标点
  ]
}
'''

```

图 7 规范的输出格式

Figure 7 Standardized output format.

```

<|im start|>user
{query} /think<|im end|>
<|im start|>assistant
<think>
{thinking_content}
</think>
{response} <|im end|>

```

图 8 (网络版彩图)监督微调数据格式模板

Figure 8 (Color online) Supervise fine-tuning of data format templates.

表 1 用于LLM训练和测试的数据集用量

Table 1 Number of training and testing data

| 规避区数量 | 训练集 | 测试集 |
|-------|-----|-----|
| 1     | 300 | 100 |
| 2     | 500 | 100 |
| 3     | 400 | 100 |

**SFT训练参数与损失函数** 我们选择在基础模型上进行监督微调训练(supervised fine-tuning training, SFT), 使用batch\_size = 128和学习率 $lr = 3 \times 10^{-6}$ 的训练配置. 模型的监督微调损失函数采用标准的语言建模损失, 其核心公式为

$$\mathcal{L}_{\text{SFT}} = - \sum_{t=1}^T \log P(y_t | y_{<t}, x), \quad (1)$$

其中 $y_t$ 表示第 $t$ 个目标token,  $y_{<t}$ 表示当前token之前的所有token,  $x$ 表示输入上下文,  $T$ 表示序列长度.

### 3 实验

本节基于合成的绕飞避障任务对所提方法进行性能评估. 首先介绍实验使用的数据集和评估指标, 随后详细说明实验配置与训练设置. 最后基于实验结果从格式正确性、策略规划、路径效率、平均响应时间、平均路径长度以及路径曲率六个维度对模型性能进行分析与讨论.

#### 3.1 数据集和指标

为评估所提出方法的有效性, 我们基于2.1节中描述的自动化数据生成流程, 构建了一个用于飞行器绕飞路径规划任务的合成数据集. 该数据集覆盖了多种不同的环境布局、障碍物密度以及任务目标类型, 能够全面模拟飞行器在复杂场景下的路径规划需求. 我们总计生成了1500个样本, 其中80%用于模型训练, 其余20%用于测试和性能评估. 每个样本均包含结构化的环境描述、任务目标信息以及由传统路径规划算法生成的最优路径作为参考标签.

在评估过程中, 我们采用以下六项指标, 从格式规范性、战略规划合理性、路径效率、实时性与路径曲率等维度综合衡量模型的输出效果.

(1) 格式准确率. 该指标用于衡量模型输出结果是否符合预期的结构化格式要求. 具体而言, 输出需包含<think>...</think>的推理内容段, 以及一段符合规范的JSON格式路径描述. 若输出不符合格式规范, 则该样本得分为0.

(2) 战略级规划准确率. 该指标评估模型输出路径在高层次规划策略上的正确性, 主要关注路径在每个障碍区或威胁区域附近的绕飞方向是否与最优路径一致. 若在任意规避点出现绕行方向不一致的情况, 则判定为错误, 得分为0.

(3) 路径长度分数. 该指标用于衡量生成路径的效率, 通过计算最优路径长度与模型生成路径长度之比来评估. 该值越接近1, 说明模型生成的路径越接近最优路径; 值越小则表示路径存在不必要的绕行, 效率较低.

(4) 平均响应时间. 鉴于路径规划的决策速度直接

影响系统的实时性能, 我们特别增加了对平均响应时间的统计. 我们重点测量了模型的规划时延, 并将其与A\*, RRT\*, Dijkstra等传统算法的平均耗时进行横向对比, 以直观揭示不同方法在时间效率上的差异与瓶颈.

(5) 平均路径长度. 平均路径长度是衡量路径规划算法优劣的核心指标. 为量化评估模型性能, 我们在相同测试集上对比了微调模型与A\*, RRT\*, Dijkstra等传统算法的规划路径长度, 从而全面评估模型的规划能力.

(6) 最大路径曲率. 路径曲率是评估规划结果是否平滑、易于无人机跟踪执行的关键指标. 为量化分析路径的平滑性与可行性, 我们统计了模型与A\*, RRT\*, Dijkstra等传统算法所生成路径的最大曲率, 旨在揭示不同方法在生成符合无人机动力学约束的平滑路径方面的能力差异.

以上六个指标共同构成对路径规划模型的全面评估体系, 既保证输出格式的规范性, 又兼顾策略合理性与执行效率, 为后续性能分析提供可靠依据.

## 3.2 训练细节

本研究选用Qwen3-1.7B作为基础语言模型, 并在此基础上进行有监督微调, 以适配绕飞路径规划任务. 选择该模型主要基于以下三点考虑. 首先, Qwen3系列模型定义了特殊的推理格式(chat template), 可将CoT推理过程封装在<think>...</think>标签内, 而将最终答案置于“response”内容中. 这一特性天然支持推理过程与答案输出的分离, 与本文旨在构建的具备显式推理能力的飞行器避障绕飞模型高度契合, 可通过“response.reasoning\_content”和“response.content”分别获取推理链和输出回复. 其次, Qwen3模型对中文具有良好的支持能力, 确保了其在处理本场景相关文本时的理解与生成质量. 最后, 1.7 B的参数规模相对较小, 使得模型能够在有限的微调数据(本研究仅使用1200条样本)下快速收敛, 证明了该方法在轻量化模型上的高效性. 尽管模型选择具有特定性, 但通过将监督微调数据格式构建为与其他基础模型的chat template相匹配, 还可以顺利适配到Llama3, ChatGLM等主流开源模型.

在训练框架方面, 我们采用LLaMAFactory进行模型训练, 并引入DeepSpeed ZeRO Stage 3优化策略, 以

实现高效的全参数微调. 学习率设置为 $3 \times 10^{-6}$ , 并使用Cosine学习率调度器进行动态调整.

全部实验在4张NVIDIA A100-PCIE-80GXB GPU上完成, 总训练时长约为10 h.

## 3.3 实验结果

从训练过程中的损失函数变化趋势可以看出, 模型在初始阶段表现出较快的收敛速度(图9和图10). 具体而言, 在前50步内, 训练损失显著下降, 反映出模型对任务基本模式的快速适应能力, 包括路径格式、环境要素识别与基础推理能力的学习. 随着训练的逐步深入, 损失下降趋于平缓, 进入稳定优化阶段, 模型逐渐增强了对复杂场景结构与高层次策略的理解与建模能力, 整体训练过程表现出良好的收敛性与稳定性.

在评估阶段, 我们首先从三个维度对模型的路径规划效果进行系统分析, 分别为输出格式准确率、战略级规划准确率以及路径长度分数. 下文将结合不同训练步数与场景复杂度, 详细展开评估结果的对比与

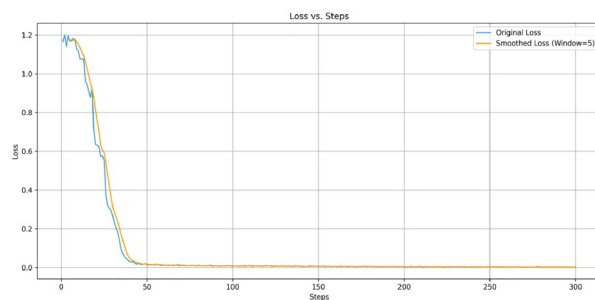


图9 训练过程中的损失曲线  
Figure 9 Loss curve during training.

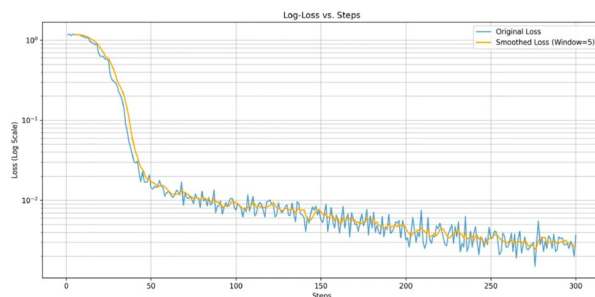


图10 训练过程中的损失曲线(对数尺度). 当模型训练到第50步时, 训练损失显著下降, 在对数尺度上接近0.01

Figure 10 Loss curve during training (log scale). When the model reaches the 50th step of training, the training loss significantly decreases, approaching 0.01 on a log scale.

分析.

首先, 在输出格式准确率方面(表2), 初始未训练模型在格式控制能力上表现不佳, 平均准确率仅为0.5, 说明其缺乏对任务特定输出结构(如JSON格式嵌套在<think>标签中)的掌握能力. 然而, 随着训练步数的增加, 该指标迅速提升, 在约60步后趋于稳定, 并在100步左右达到接近100%的准确率, 表明模型在结构化表达学习上具有较强的适应性和稳定性. 这一结果验证了监督训练对输出格式控制能力的显著促进作用.

其次, 在战略级规划准确率方面(表3), 实验结果显示, 未经训练的大模型难以进行有效的战略推理, 准确率几乎为零; 而在训练40步之后, 该指标迅速上升, 在80步左右趋于稳定, 其中最优组别准确率达到0.98, 表现出模型在理解规避区布局与方向选择方面的显著提升. 这一趋势表明, 大模型在获得少量有监督指导后, 能够有效学习任务所需的高层次空间策略.

最后, 在路径长度分数指标上(表4), 我们评估模

型生成路径的整体效率, 即其在满足规避策略的同时, 路径是否接近最短路径. 初始阶段, 模型生成的路径长度偏离理想解较远, 但在训练60步后已显著改善, 多个组别的分数超过0.9, 表明模型能够有效地在复杂约束下生成质量较高的规划路径. 该结果说明, 经过监督微调后, 模型不仅学会规避威胁, 还能优化路径代价, 具备较强的综合规划能力.

在评估的第二阶段, 我们重点对比了训练300步的LLM与A\*, RRT\*, Dijkstra等传统算法在平均响应时间、平均路径长度及路径曲率三项关键指标上的性能, 相应结果展示于表5~7.

平均响应时间分析. 各算法的平均响应时间对比结果如表5所示. 可以看出, A\*算法凭借其在离散栅格地图上的高效搜索策略, 在分别包含1~3个避障区的测试场景中均保持了最快的响应速度(0.96~1.28 s). 值得注意的是, 我们微调的LLM响应时间(2.84~5.13 s)稳定地仅次于A\*, 显著优于RRT\*(11.46~11.83 s)和Dijkstra算法(3.14~5.34 s), 表现出了优异的效率. 这一结果凸显了不同方法的内在差异: A\*, RRT\*和Dijkstra算法的

表 2 格式分数: 原始模型在输出JSON格式方面存在困难; 训练可稳定输出格式

Table 2 Format score: raw models struggle with JSON format; training stabilizes formatting

| cnt\step | 0     | 20    | 40   | 60 | 80 | 100 | 120 | 140 | 160 | 180 | 200 | 220 | 240 | 260 | 280 | 300 |
|----------|-------|-------|------|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1        | 0.50  | 0.705 | 1    | 1  | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
| 2        | 0.505 | 0.68  | 0.99 | 1  | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
| 3        | 0.505 | 0.71  | 0.96 | 1  | 1  | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |

表 3 准确率得分(战略级): 训练40步后推理能力有所提升; 训练80步时趋于稳定

Table 3 Acc score (strategic level): reasoning improves after 40 steps; stabilizes by 80 steps

| cnt\step | 0    | 20   | 40   | 60   | 80   | 100  | 120  | 140  | 160  | 180  | 200  | 220  | 240  | 260  | 280  | 300  |
|----------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1        | 0    | 0.25 | 0.93 | 0.98 | 0.96 | 0.96 | 0.97 | 0.97 | 0.96 | 0.96 | 0.97 | 0.97 | 0.97 | 0.98 | 0.97 | 0.98 |
| 2        | 0.01 | 0.01 | 0.57 | 0.62 | 0.66 | 0.63 | 0.58 | 0.66 | 0.62 | 0.60 | 0.61 | 0.62 | 0.63 | 0.63 | 0.63 | 0.62 |
| 3        | 0    | 0.04 | 0.44 | 0.69 | 0.73 | 0.70 | 0.70 | 0.73 | 0.71 | 0.74 | 0.73 | 0.68 | 0.66 | 0.73 | 0.68 | 0.68 |

表 4 准确率得分(航路点级别): 通过标签与输出长度比衡量的准确率

Table 4 Acc score (waypoint level): accuracy measured by label-to-output length ratio

| cnt\step | 0 | 20   | 40   | 60   | 80   | 100  | 120  | 140  | 160  | 180  | 200  | 220  | 240  | 260  | 280  | 300  |
|----------|---|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1        | 0 | 0.15 | 0.90 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| 2        | 0 | 0.05 | 0.54 | 0.82 | 0.90 | 0.91 | 0.93 | 0.93 | 0.93 | 0.94 | 0.94 | 0.94 | 0.95 | 0.95 | 0.95 | 0.96 |
| 3        | 0 | 0.03 | 0.60 | 0.84 | 0.91 | 0.91 | 0.91 | 0.93 | 0.91 | 0.91 | 0.93 | 0.91 | 0.90 | 0.90 | 0.90 | 0.89 |

性能严重依赖于环境建模的参数(如栅格分辨率), 参数设置不当会直接影响其搜索效率与路径质量. 而我们方法的高效率源于其无需进行复杂的图搜索或随机采样, 避免了参数敏感性问题, 鲁棒性更强. 其当前的耗时主要来自于LLM生成可解释决策思维链的过程. 考虑到LLM的推理速度可通过提升计算硬件的显存带宽来优化, 本方法的响应时间具备进一步提升到1 s甚至更短的理论潜力, 实现了在规划效率与决策透明度之间的良好平衡.

**平均路径长度分析.** 平均路径长度是衡量路径规划质量的核心指标, 结果如表6所示. 在包含一个规避区的场景中, 我们微调的LLM规划出的路径长度(108.20)优于所有对比算法, 展现了其生成最优路径的潜力. 在更为复杂的包含两个和三个规避区的场景中, 微调LLM的路径长度(分别为108.38和107.43)均接近当前最优值(A\*算法的107.34和106.56), 稳定保持次优水平, 且显著优于RRT\*算法. 上述结果表明, 本方法在不同复杂度的环境中均能保持高度竞争性的规划精度, 可稳定地输出接近最短路径的可行解, 证明了其不

表 5 平均响应时间表

Table 5 Average response schedule

| 规避区数量\方法 | A*   | RRT*  | Dijkstra | Ours |
|----------|------|-------|----------|------|
| 1        | 0.96 | 11.46 | 3.14     | 2.84 |
| 2        | 1.20 | 11.83 | 4.26     | 3.86 |
| 3        | 1.28 | 11.82 | 5.34     | 5.13 |

表 6 平均路径长度表

Table 6 Table of average path length

| 规避区数量\方法 | A*     | RRT*   | Dijkstra | Ours   |
|----------|--------|--------|----------|--------|
| 1        | 108.53 | 110.99 | 109.10   | 108.20 |
| 2        | 107.34 | 110.29 | 107.96   | 108.38 |
| 3        | 105.56 | 109.33 | 107.46   | 107.43 |

表 7 最大路径曲率表

Table 7 Maximum path curvature table

| 规避区数量\方法 | A*     | RRT*   | Dijkstra | Ours   |
|----------|--------|--------|----------|--------|
| 1        | 0.2202 | 0.1575 | 0.1824   | 0.0953 |
| 2        | 0.2361 | 0.2085 | 0.1833   | 0.1051 |
| 3        | 0.2623 | 0.2545 | 0.2049   | 0.1022 |

错的泛化能力和可靠性.

**最大路径曲率分析.** 各算法的最大路径曲率对比结果如表7所示(所有算法的路径点均平滑拟合后重采样到200个点计算曲率). 在所有测试场景中, 微调的LLM规划路径的最大曲率值(0.095~0.105)低于所有对比算法(A\*: 0.220~0.262; RRT\*: 0.158~0.255; Dijkstra: 0.182~0.205). 实验结果表明, 微调的LLM生成的路径转弯变化平缓, 具有相当不错的平滑性, 可以大幅降低底层运动控制器的跟踪难度, 从而增强了仿真到现实(sim-to-real)迁移的可行性.

综上所述, 在训练过程中, 格式准确率、战略规划准确率与路径长度分数均随训练步数显著提升, 验证了本方法在结构理解、空间推理和路径生成等多方面的有效性. 更重要的是, 微调后的LLM在规划质量(如路径长度、平滑性)与响应效率上均展现出强大竞争力, 共同证明了本框架的综合优势与实际应用潜力.

### 3.4 案例分析

在本节中, 我们选取了一个典型测试用例进行案例分析(如图11所示), 通过进一步比较大语言模型在训练前后的行为差异, 展示出所提出方法在格式理解、空间推理与策略生成方面的提升效果.

首先, 在输出格式方面, 未训练模型未能按照预期生成<think>...</think>包裹的思考过程与标准JSON格

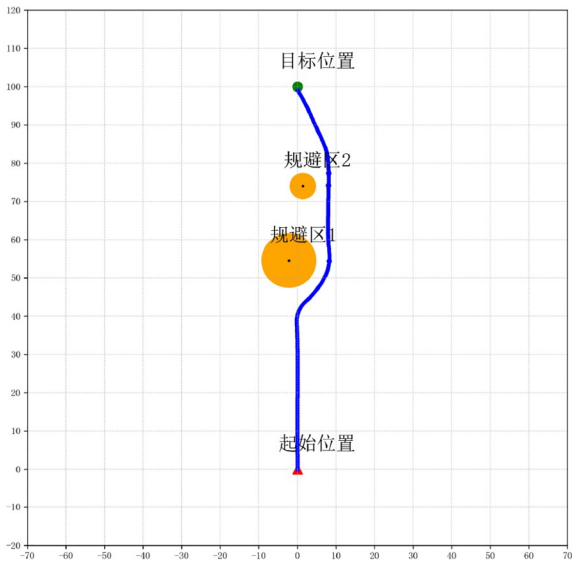


图 11 (网络版彩图)测试案例的场景示意图

Figure 11 (Color online) Rendering of test case scene.

式的策略输出(图12), 违反了任务设定的输出规范, 导致其在格式准确率评估中得分为零。而经过监督微调后, 模型能够稳定地产生符合规范的结构化输出, 并且标识信息清晰、格式正确, 与训练集中的模板保持一致。

其次, 在推理过程方面, 未训练模型的思维链(图12)显得混乱冗长, 缺乏清晰的空间结构理解。例如, 其错误地推断从(0,0)出发向右飞行将“经过 $x = -2$ 的位置”, 明显混淆了左右方向与障碍位置之间的相对关系。这

类错误表明, 初始模型在没有显式训练的情况下, 难以构建正确的几何空间模型, 其路径规划行为缺乏合理性与一致性, 整体策略也不具备可解释性。

相比之下, 训练后的模型在推理逻辑与策略组织上均表现出显著提升(图13)。其首先明确地对任务信息进行结构化理解(包括起点、终点及障碍物参数), 随后采用贪心策略分阶段处理各规避区域。在策略制定上, 模型能够准确判断路径是否与障碍物相交, 并基于障碍相对位置决定绕飞方向(如“从右侧绕飞”)。

### Qwen3-1.7B

<think>

好的, 我现在需要帮飞行器规划一条从(0,0)到(0,100)的路径, 同时避开两个规避区。首先, 我理解这些规避区的位置。规避区1在(-2,54)半径8, 也就是一个圆, 中心在左上, 半径8, 所以这个区域可能阻挡飞行器的路径。规避区2在(1,74)半径4, 是一个更小的圆, 位于右上区域。

飞行器的起始点是(0,0), 目标点是(0,100), 所以飞行器需要从左上到右上, 但中间可能有障碍。首先, 我得考虑如何绕过这两个规避区。

因为目标点在y轴上, 从0到100, 所以飞行器可能需要沿着 $x=0$ 的直线飞行, 但中间可能被规避区阻挡。比如, 规避区1在 $x=-2$ 的位置, 所以如果飞行器直接沿着y轴飞行, 可能会碰到规避区1的左侧。而规避区2在 $x=1$ 的位置, 可能在飞行器的右侧。

不过, 飞行器的起始点是(0,0), 目标点是(0,100), 所以飞行器可能需要绕过这两个规避区。比如, 可能需要先向右飞行, 绕过规避区2, 然后向上飞行, 再绕过规避区1, 最后到达目标点。

但这里有个问题, 规避区1在 $x=-2$ , 所以如果飞行器从(0,0)出发, 向右飞行, 可能经过 $x=-2$ 的位置, 但规避区的中心是 $x=-2$ , 半径8, 所以 $x=-2$ 到 $x=0$ 的位置可能被阻挡。但飞行器的起始点是(0,0), 所以可能需要绕过这个区域。比如, 飞行器可能先向右飞, 绕过规避区1的左侧, 然后向上飞行, 再绕过规避区2, 最后到达目标点。

或者, 可能飞行器需要先绕过规避区2, 然后绕过规避区1。比如, 飞行器可能先向右飞, 绕过规避区2, 然后向上飞, 再绕过规避区1, 最后到达目标点。

不过, 由于目标点在 $x=0$ , 所以飞行器需要保持 $x=0$ 的路径, 但可能需要改变方向。比如, 飞行器可能需要先向右飞行, 绕过规避区2, 然后向上飞行, 再向左飞行, 绕过规避区1, 最后到达目标点。

.....

所以, 这样的路径是可行的, 绕过规避区2, 同时飞行器在 $x=0$ 的路径上不会碰到规避区1。因此, 这样的路径是正确的。

因此, 最终的轨迹点是:

[(0,0), (1,0), (1,50), (0,50), (0,100)]

而绕飞策略可能是绕过规避区2, 因为飞行器在 $x=1$ 的路径上绕过, 而规避区1可能不需要绕过, 因为飞行器在 $x=0$ 的路径上不会碰到规避区1。

但需要确认是否还有其他规避区需要绕过。比如, 是否有其他区域需要避免? 根据题目, 只有两个规避区, 所以可能只需要绕过规避区2, 因为规避区1的路径在 $x=0$ 的区域可能不会被阻挡。

所以, 最终的策略是绕过规避区2, 而规避区1的路径没有被阻挡。因此, 策略是["2"], 即绕过规避区2。

这样, 最终的JSON应该是:

```
{
  "strategy": ["2"],
  "trajectory": [[0, 0], [1, 0], [1, 50], [0, 50], [0, 100]]
}
```

但需要确认是否所有点都在规避区外。例如, 规避区2的范围是 $x=1 \pm 4$ ,  $y=74 \pm 4$ , 飞行器在 $x=1$ ,  $y=50$ 的点是否在规避区2内? 计算距离: 中心是(1,74), 半径4, 点(1,50)到中心的距离是 $|74-50|=24$ , 大于半径4, 所以不在规避区2内。因此, 飞行器的路径是安全的。

所以, 这样的路径是正确的。

</think>

```
{
  "strategy": ["2"],
  "trajectory": [[0, 0], [1, 0], [1, 50], [0, 50], [0, 100]]
}
```

```
1 {
2   "strategy": ["2"],
3   "trajectory": [[0, 0], [1, 0], [1, 50], [0, 50], [0, 100]]
4 }
```

图 12 (网络版彩图)该测试用例在未经训练的模型上的表现。输出格式不规范, 思维内容也不正确

Figure 12 (Color online) The performance of the test case on an untrained model. The output format is not standardized, and the thinking content is also incorrect.



图 13 (网络版彩图)该测试用例在训练好的模型上的表现. 输出格式和思考内容均正确  
Figure 13 (Color online) The performance of the test case on the trained model. The output format and thinking content are both correct.

整个路径的生成过程具有高度的目标导向性, 推理步骤简洁且具备明确的因果链条, 反映出训练后的模型已具备良好的空间理解与局部路径优化能力.

综上, 该案例充分体现了本方法通过监督微调实现从通用语言模型到结构化空间规划模型的有效转变, 显著提升了模型的格式输出能力、空间推理准确性与策略生成质量.

## 4 讨论

### 4.1 Sim-to-real核心挑战

本文在构造的标准化数据集上验证了LLM路径

规划方法的有效性, 但所有实验都基于规避区规则、无严重环境噪声、飞行参数恒定等理想假设, 这与真实的无人机运行环境存在差异. 该方法要真正投入现实中使用, 还需克服三大挑战: 首先, 环境建模差异. 仿真中障碍物规整、环境静止, 而现实世界障碍物不规则且动态变化, 这可能导致LLM在仿真中学到的规划策略“水土不服”. 如何让LLM适应真实传感器数据并与感知系统协同工作, 是一大难题. 其次, 飞行器物理限制. 测试实验仅考虑了曲率指标, 未充分考虑无人机的速度、侧向加速度等其他动力学约束, 生成的路径可能过于“几何化”而无法被实际跟踪, 导致控制误差累积. 最后, 系统的鲁棒性有待加强. 真实运行中可

能出现的通信中断等异常情况, 在测试中并未充分覆盖, 导致系统在面对此类扰动时缺乏有效的容错机制, 稳定性面临考验.

## 4.2 与PID控制器的接口设计

为实现模型输出与底层控制器的解耦, 我们采用分层接口: 首先对路径作弧长参数化与样条插值, 结合曲率在侧向加速度约束下生成速度曲线并时间化, 获得连续的参考 $(x_r, y_r, \psi_r, v_r, a_r, \kappa_r)$ , 其中 $(x_r, y_r, \psi_r)$ 为随时间变化的参考位姿,  $(v_r)$ 为速度,  $(a_r)$ 为加速度,  $(\kappa_r)$ 为参考曲率; 控制层采用“前馈+PID”的横纵向双环结构: 横向通过曲率给出前馈转向, 并对投影得到的横向/航向误差作PID补偿; 纵向以参考加速度为前馈, 对速度误差作PID补偿生成功率与制动指令. 接口层统一提供指令限幅、速率限制与防积分饱和, 并在定位信度或误差超阈时触发限速或停车等降级策略, 从而在保持可解释性的同时提升可标定性与复现性.

## 4.3 局限性分析

当前工作主要建立在规避区为静态且形状规则(如圆形)的假设之上, 这一简化模型虽有助于验证核心方法的有效性, 但确实限制了其在更复杂真实场景

中的直接应用. 例如, 模型尚未考虑动态移动障碍物、不规则形状威胁区域(如建筑物轮廓、禁飞区多边形)等复杂情况. 未来的研究将重点探索如何将本框架扩展至动态与非结构化环境, 例如通过引入时序位置信息感知、实时障碍物位置预测以及更灵活的环境表征方法来提升系统在真实任务中的实用性与鲁棒性.

## 5 结论

本文针对飞行器在复杂环境中的路径规划任务, 提出了一种两阶段的大语言模型驱动方法, 系统地挖掘了LLM在空间推理与策略生成方面的能力. 在第一阶段, 我们通过结构化环境描述和任务提示模板引导模型生成绕飞路径, 使模型无需训练即可完成基本的路径规划任务; 在此基础上, 第二阶段引入自动化监督数据生成与微调机制, 显著提升了模型在格式规范性、策略合理性与路径质量等方面的表现. 实验结果表明, 该方法不仅能够稳定生成结构化输出, 还展现出清晰有序的推理流程与较强的泛化能力. 整体而言, 本研究验证了大语言模型在结构化空间任务中的可行性与潜力, 为其在智能体规划等场景中的进一步应用奠定了基础.

## 参考文献

- 1 Liu X B, Zhou S L, Xiao Z C, et al. Review on UAV obstacle avoidance methods (in Chinese). *J Ordn Equip Eng*, 2022, 43: 40–47 [刘玄冰, 周绍磊, 肖支才, 等. 无人机避障方法研究综述. *兵器装备工程学报*, 2022, 43: 40–47]
- 2 Wakabayashi T, Yukimasa Suzuki T, Suzuki S. Dynamic obstacle avoidance for Multi-rotor UAV using chance-constraints based on obstacle velocity. *Robotics Autonomous Syst*, 2023, 160: 104320
- 3 Vaidhun S, Guo Z, Bian J, et al. Dynamic path planning for unmanned aerial vehicles under deadline and sector capacity constraints. *IEEE Trans Emerg Top Comput Intell*, 2021, 6: 839–851
- 4 Mujumdar A, Padhi R. Evolving philosophies on autonomous obstacle/collision avoidance of unmanned aerial vehicles. *J Aerospace Computing Inf Communication*, 2011, 8: 17–41
- 5 Szczerba R J, Galkowski P, Glicktein I S, et al. Robust algorithm for real-time route planning. *IEEE Trans Aerosp Electron Syst*, 2000, 36: 869–878
- 6 Zhang Y L, Zhao C P, Yan H. Path planning for unmanned aerial vehicles using improved bidirectional A\* and vector field histogram algorithm (in Chinese). *Sci Technol Eng*, 2019, 19: 179–184 [张亚兰, 赵成萍, 严华. 基于改进双向A\*和向量场直方图算法的无人机航路规划. *科学技术与工程*, 2019, 19: 179–184]
- 7 Li X H, Miao M, Ran B J, et al. Obstacle avoidance path planning for UAV based on improved A\* algorithm (in Chinese). *Comput Syst Appl*, 2021, 30: 255–259 [李晓辉, 苗苗, 冉保健, 等. 基于改进A\*算法的无人机避障路径规划. *计算机系统应用*, 2021, 30: 255–259]
- 8 Dulac-Arnold G, Mankowitz D, Hester T. Challenges of real-world reinforcement learning, arXiv: 1904.12901
- 9 Hao J, Yang T, Tang H, et al. Exploration in deep reinforcement learning: From single-agent to multiagent domain. *IEEE Trans Neural Netw Learn Syst*, 2023, 35: 8762–8782

- 10 Da L, Turnau J, Kuttralingam T P, et al. A survey of sim-to-real methods in RL: Progress, prospects and challenges with foundation models, arXiv: [2502.13187](#)
- 11 Kannan S S, Venkatesh V L N, Min B C. SMART-LLM: Smart multi-agent robot task planning using large language models, arXiv: [2309.10062](#)
- 12 Xu J, Sun Q, Han Q L, et al. When embodied AI meets industry 5.0: Human-centered smart manufacturing. *IEEE CAA J Autom Sin*, 2025, 12: 485–501
- 13 Sun Q, Ji J, Mu J, et al. Vision-based hierarchical reinforcement learning for quadrotor UAV navigation. *IEEE ASME Trans Mechatron*, 2025, 30: 4154–4164
- 14 Meng S, Wang Y, Yang C F, et al. LLM-A\*: Large language model enhanced incremental heuristic search on path planning, arXiv: [2407.02511](#)
- 15 Tariq M T, Hussain Y, Wang C. Robust mobile robot path planning via LLM-based dynamic waypoint generation. *Expert Syst Appl*, 2025, 282: 127600
- 16 Xi Z, Chen W, Guo X, et al. The rise and potential of large language model based agents: A survey. *Sci China Inf Sci*, 2025, 68: 121101
- 17 Huang X, Liu W, Chen X, et al. Understanding the planning of LLM agents: A survey, arXiv: [2402.02716](#)
- 18 Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. *Adv Neural Inf Process Syst*, 2022, 35: 24824–24837
- 19 Sahoo P, Singh A K, Saha S, et al. A systematic survey of prompt engineering in large language models: Techniques and applications, arXiv: [2402.07927](#)
- 20 Aminabadi R Y, Rajbhandari S, Awan A A, et al. DeepSpeed-inference: Enabling efficient inference of transformer models at unprecedented scale. In: *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*. 2022. 1–15
- 21 Zheng Y, Zhang R, Zhang J, et al. LLaMA-Factory: Unified efficient fine-tuning of 100+ language models, arXiv: [2403.13372](#)
- 22 Ren J, Rajbhandari S, Aminabadi R Y, et al. ZeRO-Offload: Democratizing billion-scale model training, arXiv: [2101.06840](#)

## Fine-tuning of large language models for obstacle circumvention in path planning via chain-of-thought reasoning

WANG RuiTao<sup>1</sup>, LI BoYan<sup>2</sup>, SHI XiaoRong<sup>2</sup>, WEN YongMing<sup>2</sup>, CHENG Rong<sup>1</sup>, YU Yue<sup>2</sup>, ZHANG NingNing<sup>2</sup>, TANG HongYao<sup>1</sup>, ZHENG Yan<sup>1</sup> & HAO JianYe<sup>1</sup>

<sup>1</sup> College of Intelligence and Computing, Tianjin University, Tianjin 300354, China

<sup>2</sup> Science and Technology on Information Systems Engineering Laboratory, Beijing Institute of Control and Electronics Technology, Beijing 100038, China

Autonomous path planning for aerial vehicles in complex environments is essential for flight safety and mission success. Practical systems must achieve both real-time obstacle avoidance and long-range navigation. However, existing methods show limited robustness, weak generalization, and low computational efficiency when facing dynamic obstacles, environmental uncertainty, and physical constraints. In this paper, we propose a two-stage path planning framework driven by a large language model (LLM). The framework exploits the spatial reasoning and generalization capability of LLMs. In the first stage, the model generates feasible detour paths directly from structured environment descriptions and task prompts. This process does not rely on traditional mapping or search modules. In the second stage, we construct high-quality chain-of-thought reasoning samples for obstacle circumvention. We perform supervised fine-tuning to improve spatial understanding and to stabilize path generation. This stage enables a smooth transition from a general-purpose model to a task-specific planner. Experimental results show clear performance gains in highly complex scenarios. The proposed method significantly improves obstacle circumvention success rate, path rationality, and adaptability to unknown environments. These results demonstrate the feasibility of using LLMs for structured spatial planning tasks. The method also provides a practical solution for intelligent aerial navigation under real-world constraints.

**obstacle-circumvention, large language model, supervised fine-tuning, path planning**

doi: [10.1360/SST-2025-0219](https://doi.org/10.1360/SST-2025-0219)