

# 基于强化学习的无人机自主避障策略优化

陆俊凯

江南机电设计研究所 贵州 贵阳 550009

**【摘要】**：为了更好地解决无人机在复杂动态环境中自主避障的鲁棒性与效率问题，提出一种基于改进深度强化学习的避障策略。以双延迟深度确定性策略梯度算法为基础，优化状态空间表征与奖励函数设计，引入训练机制提升模型泛化能力。通过构建三维动态仿真环境，将所提策略与传统 A\* 算法、基础深度 Q 网络算法进行对比实验。结果表明，优化策略在避障成功率、路径平滑性及动态环境适应性方面均显著优于对比算法，避障成功率达 98.7%，平均路径曲率降低 32.1%，为无人机自主导航提供可靠技术支撑。

**【关键词】**：无人机；自主避障；强化学习；路径优化；深度强化学习

DOI:10.12417/3041-0630.26.03.022

## 1 引言

无人机凭借低成本、高机动性优势，已广泛应用于物流配送、电力巡检、应急救援等领域。自主避障作为无人机安全运行的核心技术，直接决定其作业范围与可靠性。传统避障方法多依赖预设地图或规则化路径规划，在动态未知环境中存在响应滞后、泛化能力不足等缺陷。随着低空经济快速发展，复杂城市空域、野外非结构化环境对无人机避障提出更高要求，亟需开发具备自适应能力的智能避障方案。强化学习通过智能体与环境的交互学习最优决策策略，无需精确环境模型，为无人机自主避障提供新路径。现有基于强化学习的避障研究虽取得一定进展，但仍存在训练效率低、动态障碍物规避能力弱、仿真与现实迁移差距大等问题。本文针对上述不足，从算法改进、状态与奖励函数设计、训练机制优化三方面入手，构建高效稳定的无人机自主避障策略，提升其在复杂环境中的适应能力。

## 2 优化避障策略设计

### 2.1 策略整体框架

本文以双延迟深度确定性策略梯度算法为基础构建核心框架，该算法通过演员-评论家网络结构处理连续动作空间，适合无人机姿态与速度的精细控制。优化框架主要包括三部分：状态感知模块、改进 DDPG 核心模块与 Curriculum Learning 训练模块。状态感知模块融合深度相机数据与无人机自身状态信息；核心模块通过优化网络结构与奖励函数提升决策精度；训练模块按难度梯度生成训练场景，加速模型收敛。

### 2.2 关键优化设计

状态空间设计直接影响策略感知能力。传统方法多依赖单一传感器数据，易受环境干扰。本文构建多维度状态空间，包含三类核心信息：深度相机采集的 144×256 像素深度图特征、无人机自身状态（位置、速度、姿态角）、环境动态特征（障碍物运动速度、距离）。通过卷积神经网络提取图像特征，与

数值状态特征融合后输入策略网络，提升环境感知的全面性。奖励函数设计采用多目标融合机制，解决传统稀疏奖励导致的训练效率低问题。奖励函数由安全奖励、效率奖励与平滑奖励三部分加权组成：

$$R = \omega_1 R_{\text{safe}} + \omega_2 R_{\text{eff}} + \omega_3 R_{\text{smooth}}$$

其中，安全奖励根据无人机与障碍物的最小距离设置，距离小于安全阈值时给予负奖励，无碰撞时给予正奖励；效率奖励与无人机到目标点的距离变化相关，距离减小给予正奖励；平滑奖励通过相邻路径点的曲率变化计算，曲率越小奖励越高。经实验调试，权重系数  $\omega_1$ 、 $\omega_2$ 、 $\omega_3$  分别设为 0.6、0.3、0.1，平衡安全与效率需求。训练机制引入 Curriculum Learning 策略，按场景难度梯度训练模型。初始阶段采用简单静态环境，仅包含少量障碍物；中期增加障碍物密度与静态复杂度；后期引入动态障碍物与传感器噪声，模拟真实环境干扰。该机制使模型从易到难逐步学习，避免初始阶段因环境复杂导致的训练崩溃，提升收敛速度与泛化能力。

## 3 实验验证与分析

### 3.1 实验环境搭建

构建三维仿真环境，场景尺寸设为 100m×100m×50m，包含静态障碍物（建筑物、电线杆）与动态障碍物（移动车辆、行人）。无人机选用四旋翼模型，搭载深度相机与 IMU 传感器，飞行速度范围 1-8m/s，安全避障距离设为 2m。实验硬件配置为 Intel i7-12700K 处理器、NVIDIA RTX 3090 显卡，软件基于 PyTorch 框架实现算法。实验设置三组对比，本文优化策略、传统 A\* 算法、基础 DDPG 算法。评价指标包括避障成功率、平均路径长度、最大曲率、计算时间，每组实验重复 30 次取平均值。

### 3.2 实验结果分析

在静态环境中，本文策略与 A\*算法均实现较高避障成功率，但本文策略的路径平滑性更优，平均路径曲率较 A\*算法降低 32.1%；在动态环境中，本文策略避障成功率达 98.7%，显著高于 A\*算法的 72.3%与基础 DDPG 算法的 89.5%。计算时间方面，本文策略因优化了网络结构，计算效率较基础 DDPG 算法提升 18.3%，满足实时避障需求。

表 1 实验结果

算法	环境类型	避障成功率	平均路径长度 (m)	最大曲率 (rad/m)	平均计算时间 (s)
A*算法	静态	96.2%	118.5	0.21	8.3
	动态	72.3%	142.7	0.25	10.5
基础 DDPG 算法	静态	95.8%	122.3	0.15	12.1
	动态	89.5%	129.8	0.18	13.4

### 参考文献:

- [1] 王丽明,辛朝阳.基于多传感器融合的无人机自主避障技术研究[J].科技创新与应用,2025,15(34):37-40.
- [2] 弭兆鹏,许富龙.城市低空环境的无人机自主避障算法分析[J].电子技术,2025,54(09):278-279.
- [3] 范弘悦.基于深度强化学习的无人机自主避障路径规划优化方法[J].中国新技术新产品,2025,(15):8-10.

本文优化策略	静态	99.2%	120.1	0.08	10.0
	动态	98.7%	125.3	0.11	11.0

### 3.3 鲁棒性测试

为验证策略鲁棒性，在传感器噪声干扰（添加 5%高斯噪声）与未知障碍物场景中进行测试。结果显示，本文策略避障成功率仍保持在 92.3%，而基础 DDPG 算法与 A\*算法成功率分别降至 78.1%与 65.4%。表明本文策略通过多维度状态感知与优化训练，具备较强的抗干扰能力与未知环境适应能力。

### 4 结论与展望

本文提出的基于改进强化学习的无人机自主避障策略，通过优化状态空间、设计多目标奖励函数与引入梯度训练机制，有效提升了无人机在复杂环境中的避障性能。实验验证表明，该策略在避障成功率、路径平滑性与动态环境适应性方面均优于传统算法与基础强化学习算法，具备实际应用价值。未来研究可结合数字孪生技术构建更真实的训练环境，缩小仿真与现实的差距。将单智能体策略扩展至多智能体场景，解决无人机集群避障问题，进一步拓展技术应用范围。