

基于深度强化学习的 RIS 辅助 UAV 通信轨迹与 相移联合优化研究综述

王恩良^{1,2}, 李 琨^{1,2}, 张炜荣^{1,2}, 孙知信^{1,2*}

(1. 南京邮电大学江苏省邮政大数据技术与应用工程研究中心, 江苏南京 210023;
2. 南京邮电大学国家邮政局邮政行业技术研发中心(物联网技术), 江苏南京 210023)

摘 要: 可重构智能表面(Reconfigurable Intelligent Surface, RIS)与无人机(Unmanned Aerial Vehicle, UAV)的协同部署为提升未来无线网络的覆盖范围、传输容量和服务灵活性提供了新的技术路径,其核心在于 UAV 轨迹、RIS 相移矩阵、主动波束成形及资源分配变量的联合优化。该问题具有跨时隙耦合、动作维度高、约束非凸和环境动态变化等特征,传统交替优化、逐次凸近似和半正定松弛方法在大规模或快速时变场景中面临迭代开销较高、在线适应性不足以及对精确信道状态信息依赖较强等问题。深度强化学习(Deep Reinforcement Learning, DRL)能够将上述联合优化表述为序贯决策过程,为复杂动态环境下的在线策略生成提供新的求解思路。本文围绕基于 DRL 的 RIS 辅助 UAV 通信轨迹—相移联合优化开展系统综述。首先,建立包含轨迹、相移和波束成形变量的统一优化框架,并从动作空间、状态空间和奖励函数三个维度分析马尔可夫决策过程建模的关键选择。随后,依据联合动作的表示与分解方式,将现有研究归纳为离散决策、连续决策和分布式协同三类技术路线,比较其算法机制、适用场景、样本效率、收敛稳定性、可扩展性与工程代价。进一步,从可扩展性、信道状态信息假设、仿真—部署迁移和公平比较基准四个方面提炼该领域的共性瓶颈。综合分析表明,离散方法适用于中小规模 RIS 和低成本相移控制;连续方法具有较高控制精度,但易受高维策略学习与样本效率限制;多智能体方法与分布式系统架构较为匹配,但会引入环境非平稳性和信用分配问题。在大规模场景下,现有方法普遍依赖结构化降维或 DRL 与传统优化相结合的混合分解,严格意义上的端到端联合求解仍较少。最后,从结构化 DRL、离线强化学习与数字孪生、标准化评测平台、近场与太赫兹通信以及通感算一体化等方面讨论未来研究方向,为相关算法设计、性能评估与系统部署提供参考。

关键词: 可重构智能表面;无人机通信;深度强化学习;轨迹优化;相移优化

基金项目: 国家自然科学基金(No.62272239);贵州省重点技术研发计划(No.[2023]272)

中图分类号: TN929.5;TP18

文献标识码: A

文章编号: 0372-2112(2026)06-2879-20

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20260377

A Review of Joint Trajectory and Phase-Shift Optimization in RIS-Assisted UAV Communication Based on Deep Reinforcement Learning

WANG Enliang^{1,2}, LI Kun^{1,2}, ZHANG Weiyong^{1,2}, SUN Zhixin^{1,2*}

(1. Jiangsu Provincial Engineering Research Center for Postal Big Data Technology and Application, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China; 2. State Post Bureau Technical Research and Development Center for Postal Industry (IoT Technology), Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China)

Abstract: The collaborative deployment of reconfigurable intelligent surfaces (RISs) and unmanned aerial vehicles (UAVs) provides a promising technological approach to improving the coverage, transmission capacity, and service flexibility of future wireless networks. Its central challenge lies in the joint optimization of UAV trajectories, RIS phase-shift matrices, active beamforming, and resource allocation variables. This problem is characterized by inter-slot coupling, high-dimensional action spaces, nonconvex constraints, and dynamically varying environments. In large-scale or rapidly time-varying scenarios, conventional methods based on alternating optimization, successive convex approximation, and semidefinite relaxation generally suffer from high iterative complexity, limited online adaptability, and a strong dependence on accurate channel state information. Deep reinforcement learning (DRL) can formulate the joint optimization problem as a sequential decision-making process, thereby providing a new approach to online policy generation in complex dynamic environments. This paper presents a systematic review of DRL-based joint trajectory and phase-shift optimization in RIS-assisted UAV communications. First, a unified optimization framework incorporating trajectory, phase-shift, and beamforming variables is

established, and the key choices in Markov decision process (MDP) modeling are analyzed from the perspectives of action-space design, state-space construction, and reward-function formulation. Existing studies are then classified into three technical paradigms according to the representation and decomposition of joint actions: discrete decision-making, continuous decision-making, and distributed cooperation. Their algorithmic mechanisms, applicable scenarios, sample efficiency, convergence stability, scalability, and implementation costs are comparatively examined. Four common bottlenecks are further identified, namely limited scalability, idealized channel state information (CSI) assumptions, the simulation-to-deployment gap, and the lack of fair benchmarking protocols. The comparative analysis shows that discrete methods are suitable for small- and medium-scale RISs with low-bit phase-shift control, whereas continuous methods provide higher control precision but are constrained by high-dimensional policy learning and limited sample efficiency. Multi-agent methods are well matched to distributed system architectures but introduce environmental nonstationarity and credit-assignment difficulties. In large-scale scenarios, most existing approaches rely on structured dimensionality reduction or hybrid decomposition combining DRL with conventional optimization, while strictly end-to-end joint optimization remains relatively uncommon. Finally, future research directions are discussed, including structured DRL architectures, offline reinforcement learning and digital twins, standardized benchmarking platforms, near-field and terahertz communications, and integrated communication, sensing, and computing. This review provides a reference for subsequent algorithm design, performance evaluation, and system deployment.

Keywords: reconfigurable intelligent surface; UAV communication; deep reinforcement learning; trajectory optimization; phase-shift optimization

Foundation Item(s): National Natural Science Foundation of China (No.62272239); Guizhou Provincial Key Technology Research and Development Program (No.[2023]272)

0 引言

在第六代(Sixth Generation, 6G)移动通信系统演进中,如何以低成本、低功耗方式实现传播环境的可编程调控,已成为无线网络设计的重要议题。可重构智能表面(Reconfigurable Intelligent Surface, RIS)通过协同控制大量低成本反射单元的相移,在无需额外射频链路的条件下重塑传播环境^[1]。与此同时,无人机(Unmanned Aerial Vehicle, UAV)凭借高机动性和按需部署能力,已广泛应用于空中基站、中继通信和数据采集等场景^[2-3]。RIS与UAV因而具有天然互补性:前者提升受遮挡环境下的链路可控性,后者提供空间位置上的动态自由度,二者协同被认为是提升未来无线网络覆盖与容量的重要路径。

图1给出了典型的RIS辅助UAV通信架构,地面基站(Base Station, BS)经由部署在建筑物立面的RIS为用户设备(User Equipment, UE)提供下行服务,UAV作为空中接入节点沿轨迹 $\mathbf{q}(t)$ 飞行以扩展覆盖范围。当BS与UE之间的直射链路受阻时,RIS通过调节相移矩阵 $\boldsymbol{\theta}$ 构造有效的非视距(Non-Line-of-Sight, NLoS)反射链路;同时,UAV与BS、RIS之间的空地链路为系统提供额外的空间自由度。因而,系统性能本质上取决于UAV轨迹 $\mathbf{q}(t)$ 、RIS相移矩阵 $\boldsymbol{\theta}$ 以及发射波束成形向量 $\mathbf{w}$ 的联合设计。

该联合优化问题的困难不在于变量数量本身,而在于变量之间的强耦合:UAV位置的变化会改变BS-RIS-UE级联信道,从而使RIS的最优相移随之改变;

RIS相移的调整又会改变不同空间位置处的有效信道增益分布,进而影响轨迹决策。由此,问题同时呈现三重特征:其一,目标函数关于轨迹、相移与波束成形变量高度非凸;其二,RIS单元的单位模约束使可行域本身即为非凸集合;其三,在多用户、多天线或多RIS设定下,决策维度会随系统规模迅速扩张。

围绕该问题,早期研究主要采用模型驱动的数学优化框架。Li等人^[4]较早研究了RIS辅助UAV通信中轨迹与被动波束成形的联合设计,通过“给定轨迹求相移—给定相移优轨迹”的分解思路求解原问题;此后,交替优化(Alternating Optimization, AO)、逐次凸近似(Successive Convex Approximation, SCA)和半正定松弛(SemiDefinite Relaxation, SDR)等方法被广泛用于类似问题^[5-7]。这类方法具有可解释性强、分析路径清晰等优点,但也存在共性约束:一是变量分解往往只能得到次优解,且随着子问题数量增加,迭代复杂度和收敛代价显著上升;二是SCA、SDR等方法依赖问题松弛与局部逼近,当系统规模增大或约束增强时,近似误差会累积放大;三是多数方法依赖较为理想的信道状态信息(Channel State Information, CSI)和离线求解流程,难以直接应对UAV移动引发的信道快时变与拓扑动态变化。

深度强化学习(Deep Reinforcement Learning, DRL)为该问题提供了另一类求解范式。与显式分解原问题的传统方法不同,DRL将联合优化过程表述为马尔可夫决策过程(Markov Decision Process, MDP),通过

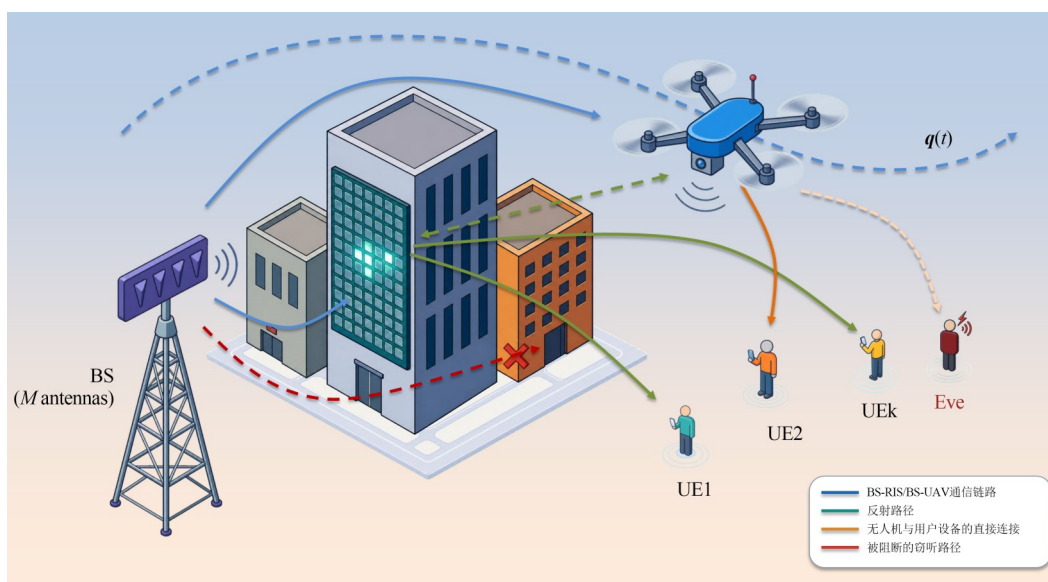


图1 RIS辅助UAV通信系统典型架构示意图

Figure 1 Typical architecture of an RIS-assisted UAV communication system

智能体与环境交互学习状态到动作的映射策略,更适合处理轨迹规划这类序贯决策问题^[8]。在方法层面,DRL通过数据驱动方式逼近高质量策略,无需对优化步骤进行手工展开;在应用层面,策略网络训练完成后即可支持在线快速推理。沿这一思路,相关研究自2018年前后在两个分支上独立起步。在UAV分支上,Bayerlein等人^[9]较早将强化学习用于自主飞行基地的轨迹优化,Liu等人^[10]进一步把RL扩展至多UAV网络的部署与移动联合设计,验证了强化学习(Reinforcement Learning, RL)在UAV序贯决策中的可行性。在RIS分支上,Huang等人^[11]率先将深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)用于RIS辅助多用户多输入单输出(Multiple-Input Single-Output, MISO)系统的被动波束成形优化,证明DRL能够在不依赖完整CSI的条件下学习RIS相移策略;Feng等人^[12]则在MISO场景下进一步验证了DRL用于RIS配置的可扩展性。在两条早期分支汇合的基础上,前述Li等人的工作虽仍属模型驱动范畴,却奠定了“轨迹—相移联合优化”这一具体问题的研究模板;Liu等人^[13]随后将机器学习方法引入该问题,开启了由模型驱动向学习驱动过渡的方向。自2022年前后开始,深度Q网络(Deep Q-Network, DQN)及其变体与DDPG、双延迟深度确定性策略梯度(Twin Delayed Deep Deterministic Policy Gradient, TD3)、软演员-评论家(Soft Actor-Critic, SAC)等Actor-Critic方法被持续引入RIS-UAV联合轨迹—相移优化,相关研究进入快速增长阶段^[14-15]。

随着相关成果迅速增长,已有综述开始从不同侧面梳理该方向。在RIS场景中的强化学习应用方面,

Puspitasari等人^[16]总结了DQN、DDPG、TD3、近端策略优化(Proximal Policy Optimization, PPO)等方法在RIS无线通信中的使用,但讨论重点仍是RIS本体优化。在UAV场景中的强化学习方法方面,Chen等人^[17]从价值函数、策略梯度与Actor-Critic三类框架综述了RL在UAV系统中的应用;王侃等人^[18]则聚焦UAV辅助移动边缘计算(Mobile Edge Computing, MEC)中的轨迹规划与资源分配问题,但均未将RIS相移控制作为核心分析对象。在RIS信号处理与系统综述方面,Pan等人^[7]和张海霞等人^[19]系统梳理了波束成形、信道估计与架构设计等问题,但仍以模型驱动方法为主线,对DRL在RIS-UAV联合轨迹与相移优化中的适用边界、性能上限与共性瓶颈缺乏专门比较。总体而言,现有综述尚未从方法论层面对该问题形成系统整合,本文旨在对此加以补充。

基于上述分析,本文的主要贡献如下:

(1)面向RIS辅助UAV通信中的轨迹—相移联合优化问题,建立“离散决策、连续决策、分布式协同”三路线分析框架,系统比较现有DRL方法的建模逻辑、适用场景与性能边界。

(2)围绕动作空间、状态空间与奖励函数三类关键设计,重构联合优化问题的MDP建模过程,揭示不同建模选择对样本效率、收敛稳定性和最终性能的影响机制。

(3)从可扩展性、CSI假设、仿真与部署差距以及比较基准四个层面提炼该领域的共性瓶颈,并据此分析DRL在RIS-UAV联合优化中尚未实现真正端到端联合求解的根本原因及后续突破方向。

1 问题框架与方法论基础

尽管现有 RIS 辅助 UAV 研究在网络拓扑、业务场景和性能指标上各不相同,其核心都可归结为轨迹变量、RIS 相移与主动波束成形变量的耦合优化。本节先给出统一问题框架,再从动作空间、状态空间和奖励函数三个维度分析该问题向马尔可夫决策过程(MDP)映射时的关键建模选择,最后建立传统优化与 DRL 的比较框架,为第 2 节的方法分类提供共同参照。

1.1 联合优化问题的统一形式

考虑图 1 所示 RIS 辅助 UAV 通信系统,系统中包含地面基站(BS)、由 N 个无源反射单元构成的 RIS、若干 UAV 以及若干地面用户。将总通信时段 T 离散为 L 个等长时隙,索引集为 $\mathcal{L} = \{1, 2, \dots, L\}$ 。

为后续表述清晰,本文采用如下符号约定:小写斜体(如 r)表示标量,小写粗体(如 $\mathbf{q}, \mathbf{w}$)表示向量,大写粗体(如 $\mathbf{\Theta}$)表示矩阵;上标 T 与 H 分别表示转置与共轭转置。在第 l 个时隙中, UAV 的三维位置记为 $\mathbf{q}[l] \in \mathbb{R}^3$, RIS 相移矩阵记为对角阵 $\mathbf{\Theta}[l] = \text{diag}(e^{j\theta_1[l]}, e^{j\theta_2[l]}, \dots, e^{j\theta_N[l]})$, 其中 $\theta_n[l] \in [0, 2\pi)$, BS 或 UAV 侧的发射波束成形向量统一记为 $\mathbf{w}[l]$ 。BS 通过直射链路和经 RIS 的级联反射链路向地面用户提供服务, UAV 则通过位置调度进一步塑造有效传播环境。因而,系统在时隙 l 的瞬时性能可统一表示为 $\mathbf{q}[l], \mathbf{\Theta}[l]$ 与 $\mathbf{w}[l]$ 的联合函数。

由于 UAV 运动学约束和能耗预算在时间维度上彼此耦合,相邻时隙的决策不可独立处理,因而本文所讨论的联合优化问题始终面向整个 L 时隙序列进行整体求解,其统一形式可写为

$$\max_{\{\mathbf{q}[l], \mathbf{\Theta}[l], \mathbf{w}[l]\}_{l=1}^L} \sum_{l=1}^L U(\mathbf{q}[l], \mathbf{\Theta}[l], \mathbf{w}[l]) \quad (1)$$

$$\text{s.t. } \mathbf{q}[l] \in \mathcal{Q}, \forall l \in \mathcal{L} \quad (2)$$

$$\left| [\mathbf{\Theta}[l]]_{n,n} \right| = 1, n=1, 2, \dots, N, \forall l \in \mathcal{L} \quad (3)$$

$$\mathbf{w}[l] \in \mathcal{W}, \forall l \in \mathcal{L} \quad (4)$$

其中,效用函数 U 随场景而变化,常见形式包括和速率、能量效率、保密速率以及无线供能场景下的吞吐量等^[20-21]。式(2)中 $\mathcal{Q}$ 表示由 UAV 运动学、能耗预算、禁飞区等约束所确定的可行轨迹集合;式(3)为 RIS 单位模约束;式(4)中 $\mathcal{W}$ 表示由发射功率、最低服务质量(Quality of Service, QoS)和公平性等通信约束所确定的可行波束成形集合。在多用户、多 UAV、多 RIS、同时透射与反射可重构智能表面(Simultaneously Transmitting And Reflecting Reconfigurable Intelligent Surface, STAR-RIS)或上行/通感一体化(Integrated Sens-

ing And Communication, ISAC)等扩展场景中, $\mathbf{w}[l]$ 与约束集 $\mathcal{W}$ 的具体结构会相应变化,但式(1)~(4)的数学骨架保持一致。

该问题的困难并非来自单一子任务,而是来自三类耦合:首先,轨迹、相移与波束成形变量共同作用于级联信道,目标函数关于 $\mathbf{q}[l], \mathbf{\Theta}[l]$ 与 $\mathbf{w}[l]$ 通常呈现强非凸耦合^[7];其次, RIS 单位模约束式(3)本身构成非凸可行域;最后,决策维度会随 RIS 规模、天线数和时隙数同步增长。以最常见的单 UAV、单 RIS、 K 用户下行 MISO 场景为例,设 BS 配备 M 根天线,则单时隙内轨迹贡献 3 维实变量,相移贡献 N 维实变量,波束成形贡献 $2KM$ 维实变量(复变量按 2 倍折算),合计 $(3+N+2KM)L$ 维;当场景拓展至多 UAV 或多 RIS 时,该维度还会按相应实体数进一步线性放大。系统规模稍大即进入高维优化区间。

除通信约束外,轨迹还必须服从 UAV 自身的动力学和能耗规律。Zeng 等人^[22]针对旋翼 UAV 给出了闭合形式的推进功率模型,表明飞行能耗与速度、加速度之间存在显著非线性关系,轨迹设计因而不能只追求链路增益,而必须同时平衡推进功耗。这意味着 RIS 辅助 UAV 优化从一开始就不是单纯的“通信参数配置”问题,而是通信、控制与能耗约束交织的联合决策问题。

1.2 从优化问题到 MDP:三个关键建模选择

DRL 并不是对原始优化问题的直接替代,而是先将其重写为 MDP,再通过策略学习近似求解。因此,决定算法最终表现的关键,往往在于动作空间、状态空间和奖励函数的设计。图 2 总结了这三个建模维度及其相应权衡。

1.2.1 动作空间设计

动作空间设计决定了联合优化是被表述为有限动作选择,还是高维连续控制。

离散动作方案将 UAV 位移与 RIS 相移同时量化为有限集合,从而可直接套用 DQN、双深度 Q 网络(Double Deep Q-Network, DDQN)等值函数方法。Zhang 等人^[15]在该框架下联合优化轨迹方向和量化相移,实现 RIS-UAV 容量提升。其优点是训练过程相对稳定、实现简单,但代价同样明确:离散化误差与动作空间规模之间存在结构性冲突。对于 N 个 RIS 单元、每个单元 B 比特量化的设定,仅相移子空间规模即为 2^{BN} 。Wu 等人^[23]已表明,低比特量化会带来明显的波束成形增益损失;而一旦提高 B 以追求精度,动作空间又会迅速超出值函数方法的可处理范围。

连续动作方案则保持变量的原始连续属性,由

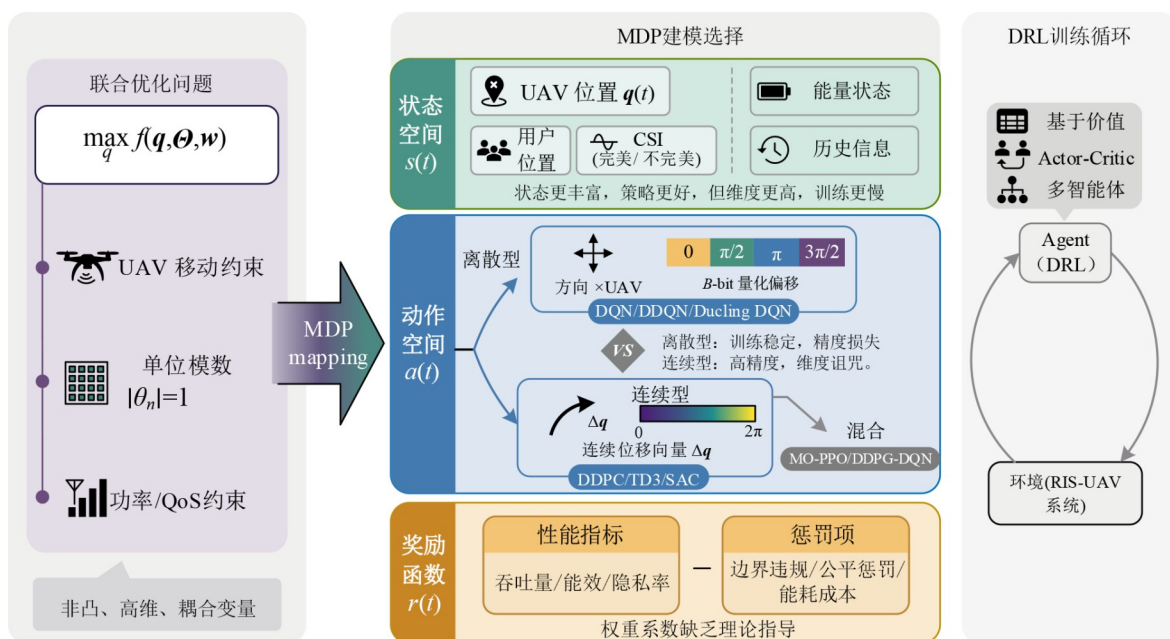


图2 联合优化问题到MDP的映射框架与三个关键建模选择

Figure 2 Mapping from the joint optimization problem to an MDP and three key modeling choices

Actor-Critic 结构直接输出 UAV 位移和相移向量。Mei 等人^[14]对 DDQN 与 DDPG 进行了对比,显示连续控制在相移精度上更具优势。Lee 等人^[24]进一步指出,随着 RIS 规模增大,Actor 输出维度升高会显著恶化收敛性。这种现象并非 RIS 场景所独有,在一般 UAV 轨迹与功率联合控制问题中,连续动作同样面临样本效率与训练稳定性的双重压力^[25]。因此,连续方案避免了离散化损失,却把困难转移为高维策略学习。

介于二者之间的是混合动作方案。Xu 等人^[26]在 RIS 辅助高铁通信中对连续波束成形与离散 RIS 控制分别建模,Umer 等人^[27]在空中 RIS 辅助协作多点-非正交多址 (Coordinated Multi-Point and Non-Orthogonal Multiple Access, CoMP-NOMA) 网络中利用多目标近端策略优化 (Multi-Objective Proximal Policy Optimization, MO-PPO) 同时输出连续位移与离散功率等级。其共同逻辑是:并非所有变量都必须被同一种动作表示方式统一处理。对 RIS-UAV 问题而言,混合动作空间本质上是一种工程化折中,用以在控制精度与训练复杂度之间寻找可接受平衡。

1.2.2 状态空间设计

状态空间决定智能体可观测的信息范围,也决定策略所能学习的决策依据。现有研究通常将 UAV 三维位置、用户坐标、瞬时 CSI、剩余能量以及当前性能指标拼接为状态向量;部分工作还会补充前序轨迹点、上一时隙相移或历史速率等时序信息,以增强状态的马尔可夫性^[28]。然而,状态向量的维度并非越

高越好。对 RIS 辅助 UAV 系统而言,最关键也最具争议的状态变量始终是 CSI。

大多数工作默认智能体能够获得完美瞬时 CSI,这在算法验证层面简化了建模,却弱化了对真实系统代价的刻画。RIS 作为被动器件无法自主发射导频,BS-RIS 和 RIS-用户级联合信道的获取本身就伴随额外训练开销^[5]。Liu 等人^[13]将机器学习用于 RIS 辅助 UAV 网络中的轨迹与被动波束成形联合设计,并考虑了级联信道估计过程,但其结果仍建立在较理想的估计精度之上。换言之,现有 DRL 策略往往是在“高质量 CSI 已知”的前提下学习决策,而非在“CSI 如何获得”这一更现实的问题上共同优化。

另一方面,状态维度过高同样会削弱学习效率。Wang 等人^[29]在 RIS 辅助 MISO 系统中指出,当状态中直接纳入所有用户位置和完整级联信道系数时,训练所需交互显著增加。因此,状态空间设计的关键在于在保留决策所必需信息的同时,控制冗余维度与 CSI 获取开销。

1.2.3 奖励函数设计

在单目标场景中,奖励通常直接取为即时和速率、总吞吐量、能量效率或保密速率等性能指标。在约束较多的场景下,常见做法是将越界飞行、功率超限或服务不公平等行为转化为惩罚项并入回报。Tang 等人^[30]在 Dueling DQN-PER (Dueling Deep Q-Network with Prioritized Experience Replay) 框架中将用户公平性嵌入奖励设计,Adhikari 等人^[31]则通过对越

界行为施加强惩罚来隐式编码飞行约束。这类设计提高了任务表述能力,但也使奖励函数本身成为一个需要精细调参的复合对象。

更值得注意的是,RIS-UAV 研究正在从单一速率指标扩展到能量感知和信息新鲜度等更复杂目标。张涛等人^[32]在 RIS 辅助无人机时效数据收集场景中同时考虑能量消耗与任务时效,陈宫紫钰等人^[33]则将信息年龄(Age of Information, AoI)直接纳入 RIS 辅助 UAV 系统优化。随着目标维度增加,奖励函数通常被写成加权和形式,但不同权重之间缺乏统一理论依据,往往依赖经验调参。这一现象的直接后果是:不同文献间的性能差异,未必完全来自算法优劣,也可能部分源自奖励塑形方式的不同。

1.3 传统优化方法与 DRL 方法的对比框架

在具体比较各类 DRL 路线之前,有必要先明确传统优化与 DRL 分别解决什么问题。

以 AO、SCA 和 SDR 为代表的模型驱动方法在理论分析上更成熟,对子问题收敛性、局部最优性及约

束处理方式具有更明确的解释^[6]。但其代价在 RIS-UAV 场景中同样突出:一方面,这类方法通常依赖精确 CSI,对级联信道误差敏感^[34];另一方面,当 RIS 规模增大或环境快速变化时,反复求解带来的时间开销会迅速累积。相比之下,DRL 的优势不在于提供更强的最优性保证,而在于学习完成后只需一次前向传播即可输出动作,对动态拓扑和序贯控制更具适配性^[8]。二者的多维度差异汇总见表 1。

然而,DRL 也并非“无代价地更优”。其训练过程高度依赖样本规模、网络结构和超参数设置,学习结果通常缺乏严格理论保证,且策略行为的可解释性弱于模型驱动方法。正因如此,当前更有现实意义的趋势不是用 DRL 完全替代传统优化,而是在二者之间建立分工互补。Feng 等人^[12]已表明 DRL 可有效处理 RIS 配置问题,而在 RIS 辅助 UAV 研究中,更常见的做法是令 DRL 负责轨迹或高层控制决策,再由 SCA、SDR 或闭式相位对齐方法求解低层相移、波束成形子问题。后续第 2 节将进一步说明这种混合范式实际上已成为现阶段多数工作的默认选择。

表 1 传统优化方法与 DRL 方法在 RIS 辅助 UAV 联合优化中的多维度对比

评估维度	传统优化方法(AO+SCA/SDR)	DRL 方法	文献
全局最优性保证	仅保证局部最优或 KKT 点;交替优化的子问题分解丢失全局信息	无理论最优性保证;但可通过充分探索逼近高质量解	文献[8]
计算复杂度	每个时隙需迭代求解凸子问题,复杂度随 N 和 K 多项式增长;运行时间为秒级	训练阶段耗时长;推理阶段仅需一次前向传播,运行时间为毫秒级	文献[11,14]
实时适应性	离线求解,无法适应信道的动态变化;需在每个相干时间块内重新求解	在线决策,可实时响应环境变化;训练后的策略具有泛化能力	文献[12,14]
CSI 依赖性	严格依赖精确 CSI;CSI 误差直接导致约束违反或性能退化	训练阶段隐式依赖 CSI;策略对 CSI 噪声具有一定鲁棒性,但完美 CSI 仍是主流假设	文献[13,35]
可扩展性	SDR 复杂度为 $O(N^3.5)$; $N > 128$ 时计算负担极重	离散方法:动作空间指数爆炸;连续方法:Actor 输出维度 $3 + N$; $N > 64$ 时收敛显著恶化	文献[23-24,31]
多目标处理	通过加权和或 ϵ 约束法处理;权重需手工设定但理论框架完善	通过奖励函数加权实现;权重缺乏理论指导,依赖实验调参	文献[30,32]
部署可行性	成熟的凸优化求解器(CVX、MOSEK)广泛可用	需要图形处理器(Graphics Processing Unit,GPU)资源进行训练;缺乏标准化仿真平台;无实测验证	文献[36-37]

2 DRL 方法的三条技术路线

面对第 1 节建立的统一问题框架,现有 DRL 方法围绕“如何表示和分解联合动作空间”逐渐分化为三类技术路线:离散决策框架、连续决策框架和分布式协同框架。三类路线的差异主要体现在它们对联合动作的处理方式:将联合动作量化后转化为有限动作选择、保留连续控制自由度后直接学习高维策略,以及将原本集中的联合决策拆解为多个协同智能体分别求解。

关于近年来出现的混合优化方法,本文将其作为

连续决策框架的延伸形式而非独立路线加以讨论,原因有二。其一,从问题表述上看,混合方法仍以连续控制为基础,所引入的 SCA、块坐标下降(Block Coordinate Descent, BCD)或启发式算法仅承担部分子问题(通常是高维相移或波束成形子问题)的精细求解,DRL 智能体本身仍输出连续轨迹动作,并未改变连续动作空间的基本设定。其二,从演进逻辑上看,混合方法是在连续决策框架遭遇高维相移训练困难后出现的工程化应对策略,其动机和求解架构均依附于连续路线,而非提出新的方法论回答。基于上述考

虑,本文将在 2.2.2 节中展开混合框架,但不将其计入路线分类。表 2 在汇总代表性工作时,对采用连续-混合架构的工作以单独标注的形式注明,以便读者识别其在连续路线内部所占的比例。

2.1 路线一:离散决策框架

2.1.1 核心思想与算法演进

离散决策框架将 UAV 轨迹与 RIS 相移同时量化为有限动作集合,使联合优化问题转化为值函数学习问题,其基本优势在于训练稳定、实现简单、超参数敏感性相对较低,因此最早进入 RIS-UAV 联合优化的主流视野。Zhang 等人^[15]和 Mei 等人^[14]分别以 DDQN 求解容量最大化与能效最大化问题,验证了“轨迹方向离散化与相移量化”这一建模路径在中小规模 RIS 场景中的可行性。随后,Tang 等人^[30]在此基础上引入 Dueling 结构和优先经验回放机制,通过 Dueling DQN-PER 提升值函数估计精度,并将公平性显式嵌入奖励设计,代表着离散路线从“能用”走向“更稳、更细”的一次典型演进。

离散路线的第二阶段并非算法范式上的根本更替,而是应用场景的持续外扩。Wang 等人^[38]将该框架扩展到 RIS 辅助全双工 UAV 混合上下行系统,联合优化轨迹与相移以提升能效;Nguyen 等人^[39]将其用于 RIS 辅助多 UAV 网络中的资源分配问题,说明值函数方法并不局限于单 UAV 场景;Lin 等人^[40]进一步将对象拓展至固定翼 UAV,在最小转弯半径与持续飞行约束下验证了离散化轨迹控制的适用性;Yu 等人^[41]则在 RIS 辅助 UAV 认知无线电系统中联合优化轨迹与资源配置,表明离散路线可迁移至频谱共享等更复杂场景。由此可见,离散决策并非只是一种“早期过渡方案”,而是在中等规模、低比特量化和较强稳定性需求下长期具有方法匹配性。

2.1.2 固有限制

离散路线的优势与局限来源于同一机制:它通过量化降低了控制难度,却也因此引入了不可消除的表示误差。其性能上限首先受到相移量化精度约束。若每个 RIS 单元仅采用低比特量化,则即便训练收敛,也只能在离散相位集合内逼近连续最优解;而一旦提高量化精度,动作空间规模又会以组合方式急剧膨胀^[23]。这意味着该路线面临“精度提升—可训练性下降”的结构性张力,而不是简单地实现细节问题。

现有代表性工作也印证了这一点。无论是 Zhang 等人^[15]、Mei 等人^[14],还是 Tang 等人^[30]、Wang 等人^[38]与 Yu 等人^[41],其有效结果大多建立在中小规模 RIS 和有限动作设计基础上。换言之,离散路线的适用区间为中小规模 RIS、较粗量化精度且对训练稳定性要

求较高的联合优化问题,而难以胜任大规模精细相移控制任务。因此,它更适合作为强基线方案或受限硬件条件下的实用方案,而难以承担大规模 RIS 场景下高精度端到端联合优化的主力角色。

2.2 路线二:连续决策框架

2.2.1 核心思想与算法演进

连续决策框架保留 UAV 位移、RIS 相移乃至功率控制等变量的连续属性,通过 Actor-Critic 结构直接输出连续动作向量,从根本上避免了离散化误差。该路线已成为现阶段文献的主体,其根本原因在于 RIS 相移本身属于连续控制对象,UAV 轨迹优化也更适合以连续状态—动作映射加以表达。

DDPG 构成了这一路线的起点,Huang 等人^[11]首先在 RIS 辅助多用户 MISO 系统中验证了 DDPG 相较传统 SDR 方法的在线推理优势;Mei 等人^[14]将 DDPG 引入 RIS-UAV 联合轨迹与相移优化,显示其在相移控制精度上优于离散 DDQN 方案。中文核心文献中,万诗晴等人^[42]进一步将深度强化学习用于 RIS 辅助 UAV 联合波束成形与轨迹优化,说明连续控制框架已开始形成较为明确的中文研究链条。与此同时,Jiao 等人^[43]将 DDPG 拓展到 RIS-UAV-NOMA 下行网络,使连续决策不再局限于“轨迹+相移”二元控制,而是逐步并入功率分配等更复杂变量。

随着问题维度上升,连续路线很快从 DDPG 过渡到更稳定的 TD3 与 SAC 系算法。Adhikari 等人^[31]在空中 RIS 场景中表明,TD3 在收敛速度和方差控制上普遍优于 DDPG;Chen 等人^[44]通过并行 TD3 加速轨迹策略探索;Dong 等人^[45]和 Yang 等人^[46]则分别在安全速率和可靠安全传输场景下使用 TD3 类框架处理轨迹、波束与相移的联合设计。与 TD3 并行发展的另一条分支是最大熵 Actor-Critic。SAC 为连续控制提供了更强探索能力^[47],Adam 等人^[48]进一步提出自适应 SAC 框架,将其用于 RIS 辅助、UAV 参与的通信系统中。由此,连续路线已形成从 DDPG 到 TD3、SAC 的内部演进链:前者解决“能否直接控制连续相移”,后者解决“在高维连续动作下如何保持训练稳定”。

在应用层面,连续路线也是当前场景扩展最充分的一类方法。Nguyen 等人^[49]将其用于无线能量传输(Wireless Power Transfer, WPT)/物联网(Internet of Things, IoT)融合场景,Hong 等人^[50]、Pang 等人^[51]和胡浪涛等人^[52]将其扩展到物理层安全问题,Xu 等人^[53]关注 RIS 辅助工业物联网绿色通信,Feng 等人^[54]研究多 UAV-NOMA 网络下的功率最小化,Moon 等人^[55]则将 RIS-UAV 系统推进到 ISAC 场景,Bansal 等人^[35]进一步把不完美和过时 CSI 引入多 RIS 选择问题。相较于离散路线主要依靠动作量化来换取稳定

性,连续路线的代表性优势在于,它更容易容纳新变量和新场景,因此成为多目标、跨层和多功能融合研究的主要载体。

2.2.2 混合优化框架

如第2节开头所述,混合优化框架在本文分类中属于连续决策路线的延伸形式,其出现是连续路线遭遇高维相移训练困难后的直接产物。当连续动作维度继续升高时,纯粹的端到端 Actor-Critic 往往难以直接处理全部相移、轨迹和资源分配变量。由此产生的一个务实方向,是将 DRL 与传统优化方法组合起来构成混合框架:由 DRL 处理轨迹规划、高层控制或部分连续变量,而将波束成形、相移细调等更具解析结构的子问题交由 SCA、BCD、启发式算法或低复杂度闭式方法求解。

这一思路在近两年尤为明显。Xu 等人^[26]在 RIS 辅助高铁通信中采用 DRL 与传统算法结合的干扰抑制方案;Umer 等人^[27]用 MO-PPO 处理空中 RIS 辅助 CoMP-NOMA 网络中的混合动作输出;Zhao 等人^[56]在多 UAV 辅助 ISAC 系统中采用“DRL 学轨迹、SCA 求波束”的分层模式;Wang 等人^[57]将 DDQN 与贪心算法结合,用列式启发方式应对大规模 RIS 的高维相移配置;Tang 等人^[58]在 UAV 搭载有源 RIS 的安全通信系统中进一步采用“外层 DRL 部署+内层 BCD 求解”的双层结构。严格说,这类方法并未维持“一个智能体统一生成全部最优动作”的端到端承诺,但它们明显提高了大规模场景下的可训练性和工程可落地性。

因此,混合框架在方法论上的意义不在于证明 DRL 可以完全替代传统优化,而在于表明二者已经开始形成稳定分工:DRL 擅长处理序贯决策与在线适应,传统优化仍然更适合承接高维相移和精细波束控制。这也是当前连续路线最具现实色彩的发展方向。

2.2.3 核心瓶颈

连续决策框架并非无代价地优于离散路线。其核心困难在于,Actor 网络的输出维度与 RIS 元件数近似线性耦合:当仅考虑三维轨迹与 RIS 相移时,输出维度已达 $3+N$;若再并入功率、波束或资源分配变量,动作空间还会进一步扩张。输出维度增长会放大策略梯度估计方差,并显著提高训练所需样本量^[24]。这也是为什么大量连续控制工作虽然声称实现了联合优化,但实际验证大多仍停留在中等规模 RIS 设定。

面对这一瓶颈,现有工作主要采取三种规避策略。第一,直接回到混合优化框架,将相移子问题交由传统方法接管^[24,54,56];第二,对相移空间进行结构化降维,只让 DRL 控制较低维度的宏观参数^[57];第三,在保持连续动作表达的同时,将研究范围限定于中小

规模 RIS 和较理想的 CSI 条件^[42,43,47,53]。这说明连续路线的真正挑战并不只是“换一个更强的 Actor-Critic 算法”,而是如何在高维连续控制下引入问题结构,使 DRL 不再被迫充当通用黑箱求解器。

2.3 路线三:分布式协同框架

2.3.1 动机与核心思想

当前两类单智能体路线面对多 UAV、多 RIS 或通信一感知一计算深度耦合场景时,动作空间和状态空间都会迅速膨胀。分布式协同框架的出发点正是将原本集中的联合决策按实体职责拆分:UAV 负责轨迹与发射控制,RIS 负责反射系数配置,或将多架 UAV、多面 RIS、多用户设备分别作为独立智能体,通过协同学习实现系统级优化。该路线的核心在于将难以直接求解的高维联合动作改写为多智能体马尔可夫博弈下的协同行为学习问题,而非简单地增加智能体数量。

现有工作大多采用集中训练一分布式执行(Centralized Training with Decentralized Execution, CTDE)范式。训练阶段各智能体共享全局信息,由集中式 Critic 提升学习稳定性;部署阶段则只保留本地观测,以满足分布式执行需求。CTDE 之所以适合 RIS-UAV 问题,是因为这类系统天然具有“局部控制、全局耦合”的结构:单个 RIS 控制器只操纵本地相移,单架 UAV 只控制本机轨迹,但它们的行为又共同决定系统总速率、总能耗或安全性能。

2.3.2 代表性工作与技术路径

在具体算法上,多智能体双延迟深度确定性策略梯度(Multi-Agent Twin Delayed Deep Deterministic policy gradient, MATD3)、多智能体深度确定性策略梯度(Multi-Agent Deep Deterministic Policy Gradient, MADDPG)及其变体构成了当前分布式路线的主干。Wang 等人^[59]针对 STAR-RIS 辅助 UAV-ISAC 安全传输提出 MATD3 框架,将 UAV 侧和 STAR-RIS 侧建模为两个独立智能体,分别处理轨迹、波束与反射/透射控制;Tesfaw 等人^[60]使用多智能体双深度 Q 网络(Multi-Agent Double Deep Q-Network, MADDPQN)联合优化 3D 轨迹与相移,在多用户通信场景下验证了多智能体值函数方法的可行性;Huroon 等人^[61]将 MATD3 用于 UAV-RIS 辅助 ISAC 系统,通过 UAV 智能体与 RIS 智能体协同平衡感知与通信性能。由此可见,分布式路线并未降低联合动作空间的整体复杂度,而是重新划分了各智能体在动作空间中的承担范围。

这一思路还向更广义的邻近场景扩展。Wang 等人^[62]在多 UAV 辅助 MEC 系统中提出通信增强型多智能体深度确定性策略梯度(Communication-enhanced Multi-Agent Deep Deterministic Policy Gradi-

ent, C-MADDPG),通过消息交互机制增强多智能体间协调;Zhao等人^[63]在RIS辅助空地NOMA通信中引入分布式鲁棒DRL以应对不确定性资源分配;Qin等人^[64]在多UAV-BS协同轨迹规划中通过多智能体DRL处理用户级公平性约束;Li等人^[65]则在飞行RIS辅助MEC网络中联合优化轨迹、相移和IoT设备关联,展示了多智能体方法在通信—计算耦合问题中的扩展能力。虽然其中部分工作并非严格意义上的“RIS-UAV轨迹与相移联合优化”核心样本,但它们为分布式路线提供了直接的方法学外推证据。

分布式协同框架表面上通过职责拆分缓解了单智能体的高维动作压力,但并未消除联合优化的复杂性,而是将其转化为新的协调难题。首要问题是环境非平稳性:任一智能体的策略更新都会改变其他智能体所面对的“环境”,从而破坏单智能体MDP中的平稳性假设。第二个问题是信用分配:系统总收益提升后,很难清楚区分增益究竟主要来自轨迹优化、相移调节还是多智能体间的配合机制。第三个问题是可扩展性并未如预期那样线性改善。现有代表性文献大多仍停留在“双智能体”或少数智能体设定^[59-61];一旦进入多UAV多RIS场景,Critic的输入维度、智能体间通信负担以及训练样本复杂度都会显著上升^[60-63]。

因此,分布式协同的真实价值不在于自动消解大规模问题的复杂度,而应理解为:当系统本身要求分布式控制时,多智能体DRL提供了与之匹配的建模范式。它所解决的是架构匹配性问题,而非无条件提升性能上界的问题。

2.4 横向比较

表2汇总了三条技术路线的代表性工作。在代表性工作总结之外,三条路线在收敛性、样本复杂度、泛化能力等理论层面也存在本质差异,表3从七个理论维度对其进行系统对比。可以看出,离散路线的优势在于训练稳定与可解释性强,但样本复杂度随RIS规模指数增长;连续路线在样本效率与泛化能力上更具优势,但缺乏全局收敛保证;分布式路线在多智能体非平稳性与信用分配上面临根本性挑战,其理论保证最弱。三类路线在理论层面的差异与第2.1~2.3节所讨论的工程层面差异相互印证;理论层面的局限往往是工程层面适用边界的根本来源。

综上提炼出三点较稳定的规律:第一,RIS规模仍是区分各路线能力边界的首要指标。当前大多数工作仍集中在 $N=16\sim 40$ 的中小规模设定;当 N 提升至64及以上时,更多研究转向混合框架或结构化降维方法^[24,51,55];第二,被动固定RIS仍是主流对象,但RIS形态正在向更复杂硬件演进。除传统固定RIS外,空中RIS、有源RIS、STAR-RIS和双RIS等新形态

已开始进入研究视野。其中,常宽等人^[66]给出了双RIS辅助车辆通信中的相移与轨迹协同控制,周毅等人^[67]则讨论了空中RIS辅助无线供能通信中的轨迹优化。这些研究工作虽然未必都采用纯DRL路线,但它们清楚表明:RIS-UAV联合优化所面对的控制自由度仍在持续扩展;第三,优化目标正从单一吞吐量或能效向复合业务目标扩展。安全通信、WPT、ISAC、MEC、数据收集和IoT等方向显著增多。其中,贾向东等人^[68]已将IRS辅助UAV系统扩展到无线传感网络数据采集场景,说明RIS-UAV联合优化的应用对象正在从传统下行通信转向更广义的空地信息获取与任务执行问题。

因此,当系统为单UAV、单RIS、中小规模 N 设定,且相移可接受低比特量化时,离散决策框架仍是最稳健的基线方案。当问题需要连续相移控制,或轨迹、功率、波束等变量同时连续可调时,连续决策框架通常更具上限;在这一类方法中,TD3和SAC相较基础DDPG更适合承担中等维度的复杂控制任务。当 N 进一步扩大、相移空间维度过高,或系统本身具有明显分层结构时,混合优化往往是更现实的选择。而当场景涉及多UAV、多RIS或明确的分布式执行需求时,多智能体框架才真正显示出建模优势。

因此,对RIS辅助UAV联合优化而言,并不存在一条对所有问题都占优的DRL路线。更准确的结论是:离散路线强调稳定性,连续路线追求控制精度,分布式路线服务于架构可分解性,而混合框架则是当前面向复杂系统最具工程可行性的折中方案。

3 共性瓶颈与批判性分析

前两节已经表明,现有RIS辅助UAV联合轨迹—相移优化方法的差异,主要来自动作空间表达、状态信息组织与决策架构设计,而不只是算法名称的更替。本节进一步从方法边界而非方法类型出发,聚焦四个反复出现且尚未真正解决的共性瓶颈:可扩展性、CSI假设、仿真—部署迁移以及比较基准。与第2节回答“哪类方法更适合哪类问题”不同,本节关心的是另一个更关键的问题:为什么当前文献中的优异结果,尚不足以支持面向真实RIS-UAV系统的强结论。

3.1 可扩展性困境

从表2可以看出,现有代表性工作中的RIS规模仍高度集中于中小规模设定,真正推进到大规模的工作仍然很少。这个分布并不反映RIS硬件本身的物理极限。相反,已有原型系统与实验平台早已表明,RIS硬件可以容纳远高于当前仿真常用规模的反射单元数^[69]。因此,这一“隐形天花板”首先是算法可训练性边界,而不是硬件可制造性边界。

表 2 三条 DRL 技术路线的代表性工作对比

Table 2 Comparison of representative studies under three DRL technical paradigms

技术路线	文献	DRL 算法	优化目标	RIS 类型	动作空间	N	融合技术	期刊/会议
离散决策	Zhang 等人 ^[15]	DDQN	容量最大化	被动固定	离散	32	—	IEEE TWC 2023
	Mei 等人 ^[14]	DDQN	能效最大化	被动固定	离散	20	—	IEEE TVT 2022
	Tang 等人 ^[30]	Dueling DQN-PER	能效最大化	被动固定	离散	30	—	Phys.Commun.2024
	Wang 等人 ^[38]	DDQN	能效最大化	被动固定	离散	20	全双工	Comp.Netw.2024
	Lin 等人 ^[40]	DRL	覆盖优化	被动固定	离散	40	固定翼	IEEE IoT J.2025
	Yu 等人 ^[41]	DRL	平均速率最大化	被动固定	离散	32	认知无线电	IEEE TVT 2023
	Nguyen 等人 ^[39]	DRL	资源分配	被动固定	离散	32	多 UAV	IEEE JSTSP 2022
	Mei 等人 ^[14]	DDPG	能效最大化	被动固定	连续	20	—	IEEE TVT 2022
连续决策	Adhikari 等人 ^[31]	TD3	能效最大化	被动 UAV 搭载	连续	36	空中 RIS	IEEE TWC 2024
	Jiao 等人 ^[43]	DDPG	和速率最大化	被动 UAV 搭载	连续	16	NOMA	Front.Sig.Proc.2022
	Hong 等人 ^[50]	TD3	安全能效最大化	被动固定	连续	32	NOMA/安全	Phys.Commun.2025
	Adam 等人 ^[48]	ASAC	最小速率最大化	被动固定	连续	40	Transformer	arXiv 2024
	Nguyen 等人 ^[49]	DDPG/TD3	和速率最大化	被动固定	连续	32	WPT/IoT	IEEE JSTSP 2022
	Pang 等人 ^[51]	DDPG	保密速率最大化	被动固定	连续	40	安全	IEEE TCOM 2022
	Xu 等人 ^[53]	DRL	能效最大化	被动固定	连续	32	D2D/IIoT	IEEE IOTJ 2024
	Dong 等人 ^[45]	TD3	安全速率最大化	被动固定	连续	40	安全	IEEE TVT 2024
	Yang 等人 ^[46]	DRL	安全传输	被动固定	连续	32	mmWave	IEEE TWC 2024
	Feng 等人 ^[54]	DRL	功率最小化	被动固定	连续	64	多 UAV/NOMA	IEEE TCOM 2023
	Bansal 等人 ^[35]	DRL	RIS 选择	被动固定	连续	48	不完美 CSI	IEEE TCOM 2023
	Moon 等人 ^[55]	DRL	和速率最大化	被动固定	连续	30	ISAC	ICT Express 2024
连续—混合	Umer 等人 ^[27]	MO-PPO	和速率最大化	被动 UAV 搭载	混合	16	CoMP-NOMA	IEEE GLOBECOM 2024
	Xu 等人 ^[26]	DRL	干扰抑制	被动固定	混合	64	高铁	IEEE TWC 2023
	Wang 等人 ^[57]	DDQN+GA	和速率最大化	被动固定	混合	128	启发式	IEEE CL 2025
	Tang 等人 ^[58]	DRL+BCD	保密速率最大化	有源 UAV 搭载	混合	24	有源 RIS	CJA 2025
	Zhao 等人 ^[56]	DRL+SCA	速率比最大化	被动固定	混合	32	ISAC	Comp.Netw.2025
分布式协同	Wang 等人 ^[59]	MATD3	安全速率最大化	STAR-RIS	连续多智能体	30	ISAC	MDPI Sensors 2025
	Tesfaw 等人 ^[60]	MADDQN	覆盖+速率	被动固定	离散多智能体	20	IoT	IEEE OJVT 2024
	Huroon 等人 ^[61]	MATD3	感知+通信	被动固定	连续多智能体	32	ISAC	IEEE IoTJ 2025
	Wang 等人 ^[62]	C-MADDPG	能耗最小化	—	连续多智能体	—	MEC	IEEE TCCN 2021
	Zhao 等人 ^[63]	DR-DRL	资源分配	被动固定	连续多智能体	40	NOMA	IEEE JSAC 2022
	Li 等人 ^[65]	MADDPG	计算位最大化	被动 UAV 搭载	连续多智能体	32	MEC/NOMA	IEEE JIOT 2024

注: 本文将“连续—混合”作为连续决策路线的延伸形式, 该类工作的共同特征是 DRL 智能体仍输出连续动作, 仅将部分子问题(高维相移或波束成形)交由 SCA、BCD 或启发式算法承担。

在离散决策框架中, 这一瓶颈表现为动作空间的组合爆炸。若每个 RIS 单元采用 B 比特量化, 则仅相移子空间的规模就达到 2^{BN} 。为保证训练可行, 大多数工作只能降低量化精度或缩小 RIS 规模; 但量化比特降低又会带来不可忽略的波束成形损失^[23]。在连续决策框架中, 该问题转化为高维策略学习困难: Actor 网络输出维度与 RIS 规模近似线性耦合, 随着 N 增长, 策略梯度估计方差和训练样本需求同步上升。Lee 等人^[24]已证明, 连续控制框架在 RIS 规模增大时会出现明显的收敛恶化; Adhikari 等人^[31]虽然用 TD3 改善了稳定性, 但验证范围仍主要停留在中等规模

设定。
更值得注意的是, 多智能体拆分并未根本消除这一问题。将 RIS 相移子空间分配给独立智能体, 只是把单一高维动作改写为“局部高维动作+协调开销”, 而没有改变每个 RIS 控制器仍需面对高维连续配置的事实。因此, 当前大多数结果更准确地说是在刻画可训练规模下的性能边界, 而不是可部署规模下的系统边界。
现有缓解思路主要有两类: 第一类是显式牺牲端到端性, 以启发式或传统方法辅助 DRL 完成大规模相移控制。Wang 等人^[57]通过将 DDQN 与贪心策略集

表 3 三条 DRL 技术路线的理论层面对比

Table 3 Theoretical comparison of three DRL technical paradigms

对比维度	离散决策框架	连续决策框架	分布式协同框架
典型算法基础	DQN, DDQN, Dueling DQN	DDPG, TD3, SAC, PPO	MADDPG, MATD3, QMIX
收敛性理论	表格型 RL 有收敛保证;函数逼近下保证较弱但训练相对稳定	Actor-Critic 在连续动作下无全局收敛保证;高维动作下易陷入局部最优	多智能体环境非平稳,违反单智能体 MDP 平稳假设;CTDE 可缓解但无理论保证
样本复杂度	与离散动作空间规模成正比;RIS 规模增大时呈指数增长	与状态—动作空间维度近似线性—多项式关系;高维相移下样本需求显著上升	与智能体数量及其动作空间维度的乘积相关;信用分配进一步加剧样本需求
泛化能力	受限于离散化粒度;跨场景迁移能力较弱	连续动作输出便于跨场景插值;策略对状态扰动具有一定鲁棒性	局部观测下的策略可在相似拓扑间复用;但全局协调机制对场景变化敏感
超参数敏感性	较低;学习率、探索率等少量参数主导	较高;Actor/Critic 学习率、目标网络更新频率、策略噪声等多参数耦合	最高;除单智能体超参外还需调控通信协议、Critic 结构与同步机制
可解释性	较强; Q 值可显式比较	较弱;策略以神经网络形式隐式表达	最弱;多策略协同行为难以单独归因
典型适用规模	中小规模 RIS($N \leq 32$)+低比特量化	中等规模 RIS($N \leq 64$)+连续相移控制	多 UAV/多 RIS 分布式系统
代表文献	文献[14-15,30,38-41]	文献[35,42-43,46,48-51,53-55]	文献[59,60-62,65]

成,在 $N=128$ 的近场场景下实现了有效配置,但其代价是 DRL 只负责宏观控制,而非元件级联合学习。第二类是利用 RIS 物理结构进行降维,例如通过码本设计和波束训练将高维相移映射为较低维参数空间^[70]。这两类方法表明,当前的核心瓶颈并非 DRL 算法自身的表达能力不足,而是现有方法尚未充分利用 RIS 的物理结构信息。

3.2 CSI 假设与现实鸿沟

如果说可扩展性暴露的是动作空间问题,那么 CSI 假设暴露的则是状态空间问题。从表 1 可以看出,完美瞬时 CSI 仍然是当前研究中的主流设定。这一假设在模型驱动优化中常被作为标准化简化处理,但在 DRL 语境下,它带来了更为本质的矛盾:DRL 虽常被描述为“无模型”方法,但其策略学习仍直接依赖于观测质量。只要状态向量中包含由 CSI 计算得到的信道增益、SINR 或级联链路指标,策略学习质量就仍然被 CSI 获取精度所上限约束。

RIS 系统中的 CSI 获取难度本就高于传统有源 MIMO 系统。由于被动 RIS 无法自主发射或接收导频,BS-RIS-用户级联信道的估计需要额外训练开销。相关综述与基础研究均指出,这类开销会随 RIS 规模和用户数增长而迅速上升^[5,71]。在多用户场景下,这种开销不仅侵占相干时间,还直接压缩了可用于真实数据传输的资源。Dai 等人^[72]虽然提出了基于分布式机器学习的下行信道估计方案,以降低估计负担,但这一方向本身仍需要额外训练过程,并且在用户移动或拓扑快速变化条件下性能会进一步下降。

因此,当前多数 DRL 工作实际上学习的不是“在不完美观测下如何决策”,而是“在高质量 CSI 已知时如何决策”。这两者的差别并非实现细节,而是问题定义层面的差别。少数工作已开始正视该问题:

Bansal 等人^[35]在多 RIS 下行场景中纳入过时 CSI 模型,分析其对 RIS 选择策略的影响;Zhou 等人^[6]在 RIS-MISO 系统中以有界误差模型刻画级联 CSI 不确定性,并给出鲁棒波束成形设计。但这两类工作仍沿用模型驱动思路,将 CSI 误差视为外部扰动,而未把 CSI 获取过程本身纳入策略学习;DRL 智能体在其中仍假设观测向量已经过预处理。换言之,CSI 问题在现有 RIS-UAV-DRL 文献中已从联合优化的核心议题降格为辅助性鲁棒约束,而真实系统中它恰恰仍是开放性挑战。

3.3 仿真与实际部署差距

就 DRL 驱动的风险-UAV 轨迹—相移联合优化而言,公开研究结果至今仍几乎全部基于数值仿真验证。需要指出的是,这并不意味着 RIS-UAV 方向完全没有任何硬件探索。张馨予等人^[36]已经对加载 RIS 的无人机通信系统进行了设计与验证研究,说明机载 RIS 系统层面的实验工作正在出现。但就 DRL 参与的联合轨迹—相移优化这一具体问题而言,真实平台验证仍明显滞后于相关算法研究的发展。

这一差距首先来自信道与硬件建模的理想化。现有多数仿真仍采用几何统计模型、自由空间损耗模型或简化的级联信道表达,而实际 RIS 的电磁行为受到近场/远场转换、元件互耦、频率相关响应以及相移—幅度耦合等多重因素影响^[73-75]。特别是 Abeywickrama 等人^[75]所揭示的非理想相移响应表明,“单位模独立相移”这一常见假设本身就与真实硬件存在偏差。只要策略训练建立在理想响应模型上,部署时就不可避免地面临性能迁移损失。

第二个差距来自 UAV 平台本身的物理约束。当前仿真中常见的匀速飞行、平滑转向和理想定位假设,往往忽略了风扰、姿态误差、飞控延迟、禁飞区与

空域监管等真实因素。Zeng 等人^[22]给出的推进功耗模型已经说明,飞行能耗与速度、加速度之间存在显著非线性;徐俊杰等人^[76]进一步把禁飞区约束显式纳入 RIS 辅助 UAV 通信网络优化。这说明真实部署中的轨迹问题并不只是“路径更复杂”,而是“可行域本身更受限”。同样,RIS 相移更新也并非无代价瞬时完成,其硬件响应时延、控制信令传输和机载计算预算都会影响在线决策闭环。

第三个差距来自 DRL 训练范式与真实部署条件之间的不匹配。仿真允许智能体进行大规模试错、收集海量交互样本,而真实 UAV 系统在飞行时间、能量储备和运行安全上都难以承受这种探索成本。因此,仿真与部署之间的差距不仅源于仿真模型的精度不足,更源于训练范式本身难以直接迁移到真实系统。离线强化学习^[37]和数字孪生辅助训练^[77]因而具有现实意义:前者尝试用预采集数据替代在线试错,后者尝试在高保真虚拟环境中完成大部分训练,再通过少量真实交互微调策略。但到目前为止,这两类思路在 RIS-UAV 联合轨迹-相移优化中的实质性应用仍然很少。

3.4 公平比较基准缺失

即便暂时不考虑部署问题,当前文献之间的横向比较本身也缺乏足够公平的共同基准。最直接的原因是仿真设定高度异质:RIS 规模、飞行区域大小、用户数量、功率预算、信道模型、训练轮次、奖励塑形方式乃至停训准则都存在显著差异。在这种情况下,

“某算法优于另一算法”往往只在特定参数组合下成立,而难以转化为具有一般性的路线判断。

第二个问题是基准方案的选择偏弱。大量工作仍以“随机轨迹+随机相移”或“固定轨迹+固定相移”作为主要下界。相对于此类弱基准,任何经过训练的策略都容易取得夸张的性能提升,但这并不能说明所提方法相对于合理传统方法或强 DRL 基线的真实优势。少数工作如 Mei 等人^[14]和 Adhikari 等人^[31]在统一场景下比较了多种 DRL 算法,因而更具解释力;但这类“同环境、同指标、同训练预算”的严谨比较在现有文献中仍属少数。

第三个问题来自奖励定义的不统一。即便都以“能效”或“频谱-能耗折中”为目标,不同工作对速率、功耗、惩罚项和公平性指标的加权方式往往相差很大。类似问题在一般 RIS-DRL 研究中已经出现^[78],当其迁移到 RIS-UAV 联合优化场景后,轨迹约束、越界惩罚和多用户公平性又会进一步放大奖励设计差异。于是,不同论文之间的性能差异,往往混杂了算法能力差异与任务定义差异两种因素。

最后,当前领域仍缺乏可复现实验平台。通用强化学习之所以能够实现较成熟的横向比较,一个重要前提是存在共享环境、共享协议和相对稳定的评测指标;而 RIS-UAV 方向至今大多依赖各自实现的私有仿真代码。这不仅削弱了结果可复现性,也提高了新方法进入比较体系的门槛,表 4 对上述四个瓶颈进行了归纳。

表 4 基于 DRL 的 RIS 辅助 UAV 联合优化研究的四个共性瓶颈总结

Table 4 Summary of four common bottlenecks in DRL-based joint optimization for RIS-assisted UAV systems

瓶颈维度	现状概括	典型现象	代表性尝试	核心难点
可扩展性	研究仍以中小规模 RIS 为主	多数工作 $N \leq 48$, 超过 64 者较少	Wang 等人 ^[57] 在大规模 RIS 中引入启发式集成 DRL	通用 DRL 未利用单位模、空间相关性和稀疏性等结构先验
CSI 假设	完美 CSI 仍是默认设定	训练与测试通常都在理想 CSI 下完成	Bansal 等人 ^[35] 考虑过时 CSI; Zhou 等人 ^[6] 讨论鲁棒设计	级联信道估计开销随 RIS 规模 and 用户数同步增长
仿真一部署差距	DRL-RIS-UAV 联合优化仍几乎停留在仿真层面	硬件非理想、空域约束和在线训练成本很少被完整建模	张馨予等人 ^[36] 已开始机载 RIS 系统设计与验证探索	仿真环境与真实平台之间存在显著系统级偏差
公平比较基准	文献间参数设定、奖励设计和 baseline 差异大	“性能提升”常难以跨论文横向解释	Mei 等人 ^[14] 、Adhikari 等人 ^[31] 进行了同场景多算法比较	缺乏标准环境、统一协议和开源实现

3.5 深层矛盾

综合以上四个瓶颈,可以识别出贯穿该领域的一个方法论矛盾: DRL 被引入 RIS-UAV 联合优化的初衷是摆脱传统交替优化对耦合变量的显式分解,但在当前主流实践中,研究者又不得不以新的形式重新引入分解机制。在离散决策框架中,这种拆分表现为量化与裁剪,研究者通过降低相移精度(如 Tang 等人^[30]、Wang 等人^[38]所采用的 2-3 比特量化)将连续相位空间收缩为有限码本。在连续决策框架中,这种拆分表

现为分层求解: DRL 负责轨迹或高层部署,SCA、BCD 或闭式方法承担高维相移与波束控制。混合框架将该分工直接制度化,例如 Wang 等人^[57]在 $N=128$ 场景中由 DDQN 输出宏观控制方案、由贪心算法完成元件级配置; Tang 等人^[58]在 UAV 搭载有源 RIS 的安全通信中采用“外层 DRL 部署+内层 BCD 求解”的双层结构。多智能体框架则将其转写为智能体职责切分,如 Wang 等人^[59]在 STAR-RIS-UAV-ISAC 中将 UAV 侧与 STAR-RIS 侧分别建模为独立智能体,所谓“联合”通

过集中式 Critic 间接耦合而非由单一策略统一输出。值得注意的是,表 2 中 $N \geq 48$ 的高维场景几乎全部落入后两类分解模式,这并非偶然,而是当前可训练性边界的客观结果。因此,文献中频繁出现的“联合优化”在很多情形下更准确的描述应为分层协同求解;所谓“端到端”也往往仅指策略推理链条的端到端,而非变量求解过程的端到端。

Shang 等人^[79]在讨论空中 RIS 与 MEC 结合时已经指出,当前 DRL 更多表现为传统优化流程中的智能化替代,而非方法论意义上的彻底改写。Ning 等人^[80]进一步强调,RIS 与 UAV 深度融合后的真正突破,未必来自单纯更换一种更强的 RL 算法,而更可能来自面向系统结构重新设计学习架构。沿着这一判断回看全文,可以得到一个更稳妥的结论:当前局面意味着通用 DRL 尚未真正匹配 RIS-UAV 这类结构化高维优化问题。只有把单位模约束、空间相关性、信道稀疏性和硬件可实现性真正嵌入学习框架之中,DRL 才可能从“传统优化的智能求解器”演进为“面向 RIS-UAV 联合控制的专用方法论”。

4 未来研究方向

前述分析表明,当前 RIS 辅助 UAV 联合轨迹—相移优化的主要障碍已不再是网络结构本身的容量是否

足够,而在于通用 DRL 与问题结构之间尚未建立稳定匹配关系。第 3 节识别的可扩展性、CSI 假设、仿真—部署差距与公平比较基准四个共性瓶颈,分别指向算法架构、状态观测、训练范式与评价体系四个维度的失配,本节据此围绕结构先验注入、训练范式重构、场景边界扩展与系统接口落地四条主线展开未来研究方向。

四条主线进一步按时间尺度分为近期、中期与远期,划分依据为方法成熟度、与 RIS-UAV 场景的契合度以及对外部硬件或标准化进展的依赖程度。近期方向所依赖的核心技术(如结构化学习架构、离线 RL、数字孪生与平台工程)在通用领域已较为成熟,向 RIS-UAV 场景迁移的工作量可控且无强外部依赖;部分核心技术已被提出,例如多跳 RIS-太赫兹(Terahertz, THz)混合波束成形方案早在 2021 年即由 Huang 等人^[81]给出,但其向 RIS-UAV 场景的实质迁移仍依赖近场信道建模^[82-83]与 UAV 近场移动性等基础工作的进一步成熟;远期方向则进一步依赖超越对角线可重构智能表面(Beyond-Diagonal Reconfigurable Intelligent Surface, BD-RIS)等新型硬件形态与 6G 空地控制接口的标准化进展,超出单一研究团队可独立推动的范围。基于上述时间尺度划分,图 3 进一步归纳了四类问题来源、近期至远期研究任务及其面向 RIS-UAV 结构化联合优化目标之间的对应关系。

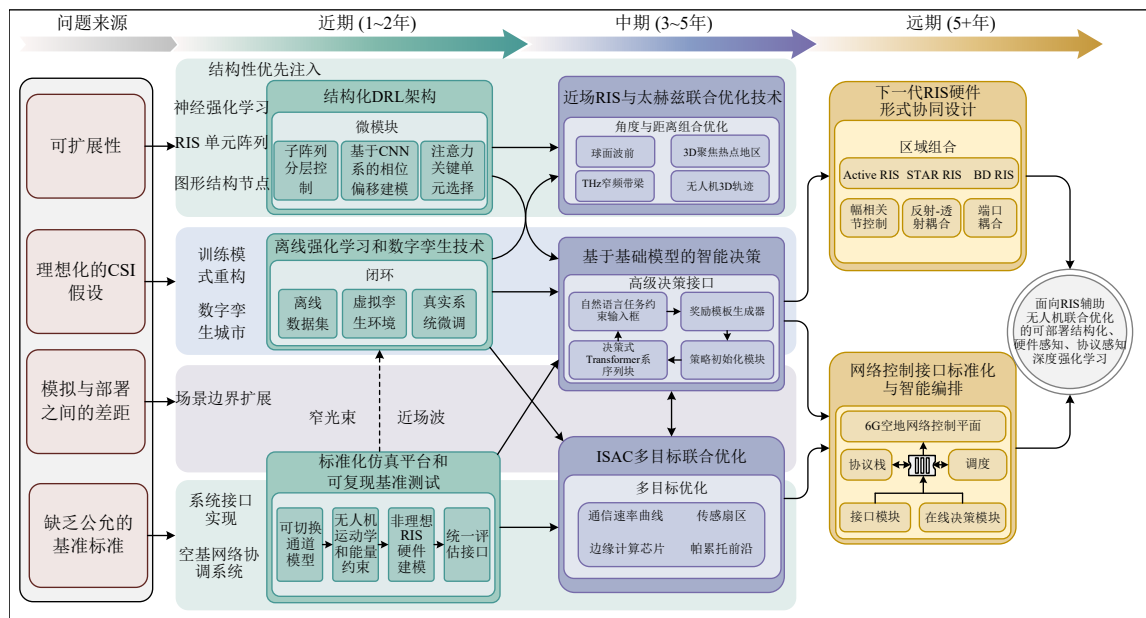


图 3 基于 DRL 的 RIS 辅助 UAV 联合优化未来研究方向路线图

Figure 3 Roadmap of future research directions for DRL-based joint optimization in RIS-assisted UAV systems

4.1 近期可突破的方向

4.1.1 面向大规模 RIS 的结构化 DRL 架构

面向大规模 RIS 的首要任务是显式利用 RIS 问题的结构信息,避免继续将高维相移向量整体交由通用

Actor 网络处理。更有前景的方向至少包括三类。其一,是基于子阵列或列式控制的分层决策:先由 DRL 输出较低维的组级宏观动作,再由低复杂度局部算法完成组内精调,以此降低输出维度并缓解训练不稳定

性^[55,68]。其二,是基于图结构的相移建模:将 RIS 元件及其空间邻接关系表示为图,通过图神经网络编码元件间耦合,而不是把所有相移视为彼此独立的标量变量^[84]。其三,是基于注意力或序列建模的选择性控制:对贡献更大的关键单元分配更高的建模分辨率,而不是对全部单元进行等权处理^[85]。

该方向的核心判断在于:大规模 RIS 问题所欠缺的是能够将单位模约束、空间相关性与传播稀疏性转化为网络归纳偏置的专用学习架构,而非通用 DRL 算法本身的容量。

4.1.2 离线强化学习与数字孪生辅助训练

真实 UAV 平台难以承受大规模在线试错,因此训练范式必须从“纯在线交互”转向“离线数据驱动+少量在线微调”。离线强化学习为此提供了直接路径:研究者可以先利用传统优化器、启发式算法或既有飞行日志生成轨迹—相移—性能数据集,再用离线 RL 从静态样本中学习可部署策略^[37]。其优势在于显著降低探索成本,但其瓶颈同样明确:一旦学习策略偏离数据分布,值函数就可能产生严重外推误差。

数字孪生则为这一问题提供了中间层。通过构建包含信道演化、UAV 运动学、能耗模型与 RIS 硬件响应的高保真虚拟环境,可以在虚拟空间中完成大部分策略训练,再通过少量真实交互完成校准^[77]。对 RIS-UAV 系统而言,最具可行性的训练范式很可能是“离线数据预训练—数字孪生迭代—真实环境微调”的三级流程。

4.1.3 标准化仿真平台与可重复性基准

如果缺乏统一的实验环境,算法优劣就很难被公平比较。因此,建立标准化仿真平台是该方向进一步积累知识的基础设施。一个有意义的平台至少应同时包含三类要素:其一,支持概率 LoS、莱斯衰落、毫米波/近场等多类可切换信道模型;其二,内置旋翼与固定翼 UAV 的运动学和能耗约束;其三,显式建模实际 RIS 的相移—幅度耦合、响应时延和硬件非理想性。在此基础上,还应预设标准场景、统一指标接口,并预留硬件在环或实测数据接入能力,以便逐步与机载 RIS 设计验证结果衔接。该方向的目标是形成可复现实验协议、可比较评测环境和可迁移部署接口的统一基准,而非仅仅开源代码。

4.2 中期探索方向

4.2.1 近场 RIS 与太赫兹频段下的联合优化

随着 RIS 规模继续增大,RIS 与用户之间的传播条件将越来越频繁地进入近场区间,远场平面波模型不再适用,信道建模必须转向球面波与距离相关聚焦机制^[82-83]。这意味着 RIS 相移设计不再只依赖角度信息,还必须同时感知距离维度;相应地,UAV 轨迹

优化也不再只是“飞向更优角度”,而是“飞向更优三维聚焦位置”。对于 DRL 而言,这不是参数微调问题,而是状态表征、动作设计和奖励定义都需要随物理模型一起改变。

与之相关的,是太赫兹频段下的 RIS-UAV 协同。THz 链路虽可提供更大带宽,但其路径损耗、大气吸收和波束偏移更为突出,因而对波束控制与环境感知提出更高要求^[81]。因此,近场与 THz 将在更深层次上迫使现有 DRL 框架重新适配新的传播规律,其影响远超一般意义上的应用场景扩展。

4.2.2 基础模型赋能的智能决策

基础模型可能为 RIS-UAV 决策带来新的接口层,但这一方向需要保持克制。更现实的切入点不是让大模型直接生成轨迹与相移,而是让其承担任务规格解析、奖励函数辅助设计和策略初始化等高层角色。例如,可以先用自然语言描述系统目标与约束,再由基础模型将其转写为更规范的奖励或约束模板;也可以利用预训练模型从历史优化数据中抽取跨场景决策模式,为后续 DRL 训练提供更好的初始化。

在模型架构层面,Decision Transformer 一类序列决策模型提供了另一条思路:将强化学习改写为条件序列建模,以处理长时域轨迹依赖和延迟回报问题^[85]。但必须指出,基础模型并不会自动提升无线控制性能。若其生成过程不能与物理可行域、QoS 约束和实时预算严格对齐,反而可能放大奖励失配与策略失真。因此,该路径需要重点研究的问题并非是否引入大模型,而是如何将基础模型限制在与其能力匹配的决策层级上。

4.2.3 通感算一体化中的多目标联合优化

RIS-UAV 系统正逐步从“单一通信优化”扩展为“通信—感知—计算”联合系统。在 ISAC 场景中,RIS 配置与 UAV 轨迹必须同时兼顾通信速率、感知精度和安全性^[86];在 MEC 场景中,联合决策变量又扩展到任务卸载、计算资源分配、链路调度以及轨迹—相移协同^[87]。这意味着现有大量工作所依赖的单一标量奖励,越来越难以准确表达系统目标。

因此,中期更值得推进的方向是约束强化学习、多目标强化学习和分层策略设计。前者适合显式处理最低速率、最大时延和能耗预算等硬约束;后者适合刻画通信性能、感知性能与计算代价之间的帕累托关系;分层策略则有助于把高层任务分配与低层轨迹—相移控制分开处理。对 RIS-UAV 领域而言,这类研究的意义在于将系统从链路级优化推进到任务级优化。

4.3 远期展望

从更长周期看,DRL 在 RIS-UAV 系统中的演进将

主要受两条外部主线牵引。第一条主线是RIS硬件形态持续扩展。从有源RIS、STAR-RIS到超越对角线RIS,控制自由度已从单纯相位调节逐步扩展到幅相联合控制、反射—透射耦合控制与更复杂的端口耦合结构^[57, 88-89]。这意味着未来的学习算法不能再假定“相移向量就是全部决策变量”,而必须面向更复杂的硬件约束作协同设计。

第二条主线是网络控制接口与智能编排机制逐步规范化。一旦RIS控制与UAV网络编排进入更统一的系统框架,DRL的角色也会从“论文中的独立求解器”转向“协议感知的在线决策模块”^[80]。届时,算法优劣将不再只由仿真性能决定,还将取决于其是否能满足反馈开销、时延预算和控制接口约束。

5 结论

本文围绕RIS辅助UAV通信中基于DRL的轨迹—相移联合优化问题,建立了统一的问题框架,比较了离散决策、连续决策和分布式协同三类方法,并从可扩展性、CSI假设、仿真—部署差距与比较基准四个方面识别了该方向的核心瓶颈。全文的中心结论并非“哪一种DRL算法绝对更优”,而是通用DRL尚未真正匹配RIS-UAV这类结构化、高维、强约束的联合优化问题。

因此,后续工作的关键在于更深层的方法重构:在架构层面引入结构先验,在训练层面引入离线数据和数字孪生,在场景层面扩展到近场、THz、ISAC与MEC,在系统层面推动标准化环境与可部署接口。只有当DRL从“传统优化的智能求解器”演进为“面向RIS-UAV系统的结构化决策方法”,其端到端联合优化的潜力才可能真正兑现。

参考文献

- [1] Wu Qingqing, Zhang Rui. Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming[J]. IEEE Transactions on Wireless Communications, 2019, 18(11): 5394-5409.
- [2] Zeng Yong, Zhang Rui, Lim T J. Wireless communications with unmanned aerial vehicles: Opportunities and challenges[J]. IEEE Communications Magazine, 2016, 54(5): 36-42.
- [3] 陈新颖, 盛敏, 李博, 等. 面向6G的无人机通信综述[J]. 电子与信息学报, 2022, 44(3): 781-789.
Chen Xinying, Sheng Min, Li Bo, et al. Survey on unmanned aerial vehicle communications for 6G[J]. Journal of Electronics & Information Technology, 2022, 44(3): 781-789. (in Chinese)
- [4] Li Sixian, Duo Bin, Yuan Xiaojun, et al. Reconfigurable intelligent surface assisted UAV communication: Joint trajectory design and passive beamforming[J]. IEEE Wireless Communications Letters, 2020, 9(5): 716-720.
- [5] Wu Qingqing, Zhang Shuowen, Zheng Beixiong, et al. Intelligent reflecting surface-aided wireless communications: A tutorial[J]. IEEE Transactions on Communications, 2021, 69(5): 3313-3351.
- [6] Zhou Gui, Pan Cunhua, Ren Hong, et al. A framework of robust transmission design for IRS-aided MISO communications with imperfect cascaded channels[J]. IEEE Transactions on Signal Processing, 2020, 68: 5092-5106.
- [7] Pan Cunhua, Zhou Gui, Zhi Kangda, et al. An overview of signal processing techniques for RIS/IRS-aided wireless systems[J]. IEEE Journal of Selected Topics in Signal Processing, 2022, 16(5): 883-917.
- [8] Luong N C, Hoang D T, Gong Shimin, et al. Applications of deep reinforcement learning in communications and networking: A survey[J]. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3133-3174.
- [9] Bayerlein H, De Kerret P, Gesbert D. Trajectory optimization for autonomous flying base station via reinforcement learning[C]//2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). Piscataway: IEEE, 2018: 1-5.
- [10] Liu Xiao, Liu Yuanwei, Chen Yue. Reinforcement learning in multiple-UAV networks: Deployment and movement design[J]. IEEE Transactions on Vehicular Technology, 2019, 68(8): 8036-8049.
- [11] Huang Chongwen, Mo Ronghong, Yuen C. Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(8): 1839-1850.
- [12] Feng Keming, Wang Qisheng, Li Xiao, et al. Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems[J]. IEEE Wireless Communications Letters, 2020, 9(5): 745-749.
- [13] Liu Xiao, Liu Yuanwei, Chen Yue. Machine learning empowered trajectory and passive beamforming design in UAV-RIS wireless networks[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(7): 2042-2055.
- [14] Mei Haibo, Yang Kun, Liu Qiang, et al. 3D-trajectory and phase-shift design for RIS-assisted UAV systems using deep reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2022, 71(3): 3020-3029.

- [15] Zhang Haijun, Huang Miaolin, Zhou Huan, et al. Capacity maximization in RIS-UAV networks: A DDQN-based trajectory and phase shift optimization approach[J]. IEEE Transactions on Wireless Communications, 2023, 22(4): 2583-2591.
- [16] Puspitasari A A, Lee B M. A survey on reinforcement learning for reconfigurable intelligent surfaces in wireless communications[J]. Sensors, 2023, 23(5): 2554.
- [17] Chen Hengsheng, Lin Yuanguo, Fu Mingjian, et al. A survey on reinforcement learning methods for UAV systems[J]. ACM Computing Surveys, 2026, 58(4): 1-37.
- [18] 王侃, 曹铁林, 李旭杰, 等. 无人机辅助边缘计算网络轨迹规划与资源分配研究综述[J]. 电子与信息学报, 2025, 47(5): 1266-1281.
- Wang Kan, Cao Tielin, Li Xujie, et al. A survey on trajectory planning and resource allocation in unmanned aerial vehicle-assisted edge computing networks[J]. Journal of Electronics & Information Technology, 2025, 47(5): 1266-1281. (in Chinese)
- [19] 张海霞, 马晓燕. 可重构智能表面辅助的通信信号处理技术[J]. 信号处理, 2024, 40(6): 993-1006.
- Zhang Haixia, Ma Xiaoyan. Signal processing technologies for RIS-assisted wireless communications[J]. Journal of Signal Processing, 2024, 40(6): 993-1006. (in Chinese)
- [20] Adam A B M, Ouamri M A, Wan Xiaoyu, et al. Secure communication in UAV-RIS-empowered multiuser networks: Joint beamforming, phase shift, and UAV trajectory optimization[J]. IEEE Systems Journal, 2024, 18(2): 1009-1019.
- [21] 刘志新, 赵松晗, 杨毅, 等. 智能反射面辅助的无人机无线携能通信网络吞吐量最大化算法研究[J]. 电子与信息学报, 2022, 44(7): 2325-2331.
- Liu Zhixin, Zhao Songhan, Yang Yi, et al. Throughput maximization algorithm for intelligent reflecting surface-aided unmanned aerial vehicle communication networks with wireless energy transfer[J]. Journal of Electronics & Information Technology, 2022, 44(7): 2325-2331. (in Chinese)
- [22] Zeng Yong, Xu Jie, Zhang Rui. Energy minimization for wireless communication with rotary-wing UAV[J]. IEEE Transactions on Wireless Communications, 2019, 18(4): 2329-2345.
- [23] Wu Qingqing, Zhang Rui. Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts[J]. IEEE Transactions on Communications, 2020, 68(3): 1838-1851.
- [24] Lee G, Jung M, Kasgari A T Z, et al. Deep reinforcement learning for energy-efficient networking with reconfigurable intelligent surfaces[C]//Proceedings of IEEE International Conference on Communications (ICC). Piscataway: IEEE, 2020: 1-6.
- [25] 杨青青, 陈剑, 彭艺. 基于DDPG的无人机轨迹规划及功率控制算法[J]. 北京邮电大学学报, 2023, 46(3): 43-48.
- Yang Qingqing, Chen Jian, Peng Yi. Unmanned aerial vehicle trajectory planning and power control algorithm based on deep deterministic policy gradient[J]. Journal of Beijing University of Posts and Telecommunications, 2023, 46(3): 43-48. (in Chinese)
- [26] Xu Jianpeng, Ai Bo, Quek T Q S. Toward interference suppression: RIS-aided high-speed railway networks via deep reinforcement learning[J]. IEEE Transactions on Wireless Communications, 2023, 22(6): 4188-4201.
- [27] Umer M, Mohsin M A, Mahmood A, et al. Deep reinforcement learning for trajectory and phase shift optimization of aerial RIS in CoMP-NOMA networks[C]//Proceedings of the IEEE Global Communications Conference. Piscataway: IEEE, 2024: 79-84.
- [28] Tang Runzhi, Wang Junxuan, Zhang Yanyan, et al. Throughput maximization in NOMA enhanced RIS-assisted multi-UAV networks: A deep reinforcement learning approach[J]. IEEE Transactions on Vehicular Technology, 2025, 74(1): 730-745.
- [29] Wang Wei, Zhang Wei. Intelligent reflecting surface configurations for smart radio using deep reinforcement learning[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(8): 2335-2346.
- [30] Tang Runzhi, Wang Junxuan, Jiang Fan, et al. Joint 3D trajectory and phase shift optimization via deep reinforcement learning for RIS-assisted UAV communication systems[J]. Physical Communication, 2024, 66: 102456.
- [31] Adhikari B, Khwaja A S, Jaseemuddin M, et al. Energy efficient RIS-assisted UAV networks using twin delayed DDPG technique[J]. IEEE Transactions on Wireless Communications, 2024, 23(12): 18423-18439.
- [32] 张涛, 张迁, 朱颖雯, 等. 基于能量感知的智能反射面辅助无人机时效数据收集策略[J]. 电子与信息学报, 2025, 47(2): 427-438.
- Zhang Tao, Zhang Qian, Zhu Yingwen, et al. Energy aware reconfigurable intelligent surface assisted unmanned aerial vehicle age of information enabled data collection policies[J]. Journal of Electronics & Information Technology, 2025, 47(2): 427-438. (in Chinese)

- [33] 陈宫紫钰, 贾向东, 张鑫. 智能反射面辅助UAV通信系统的信息年龄优化方案[J]. 信号处理, 2024, 40(8): 1480-1488.
Chen Gongziyu, Jia Xiangdong, Zhang Xin. Age of information optimization scheme for an intelligent reflecting surface assisted UAV communication system[J]. Journal of Signal Processing, 2024, 40(8): 1480-1488. (in Chinese)
- [34] Zhou Gui, Pan Cunhua, Ren Hong, et al. Robust beamforming design for intelligent reflecting surface aided MISO communication systems[J]. IEEE Wireless Communications Letters, 2020, 9(10): 1658-1662.
- [35] Bansal A, Agrawal N, Singh K, et al. RIS selection scheme for UAV-based multi-RIS-aided multiuser downlink network with imperfect and outdated CSI[J]. IEEE Transactions on Communications, 2023, 71(8): 4650-4664.
- [36] 张馨予, 孔祥鲲, 刘子庆, 等. 加载可重构智能超表面的无人机通信系统设计与验证[J]. 电波科学学报, 2024, 39(4): 741-747.
Zhang Xinyu, Kong Xiangkun, Liu Ziqing, et al. Design and verification of a communication system based on UAV with airborne reconfigurable intelligent surface[J]. Chinese Journal of Radio Science, 2024, 39(4): 741-747. (in Chinese)
- [37] Levine S, Kumar A, Tucker G, et al. Offline reinforcement learning: Tutorial, review, and perspectives on open problems[PP/OL]. V3. arXiv (2020-11-01)[2026-04-24]. <https://arxiv.org/abs/2005.01643>.
- [38] Wang Yi, Deng Yu, Kang Ling, et al. Reinforcement learning-based energy efficiency optimization for RIS-Assisted UAV hybrid uplink and downlink system[J]. Computer Networks, 2024, 245: 110390.
- [39] Nguyen K K, Khosravirad S R, da Costa D B, et al. Reconfigurable intelligent surface-assisted multi-UAV networks: Efficient resource allocation with deep reinforcement learning[J]. IEEE Journal of Selected Topics in Signal Processing, 2022, 16(3): 358-368.
- [40] Lin Na, Liu Chunxiao, Wu Tianxiong, et al. Green communications: RIS-assisted fixed-wing UAV coverage scheme based on deep reinforcement learning[J]. IEEE Internet of Things Journal, 2025, 12(4): 4115-4127.
- [41] Yu Yingfeng, Liu Xin, Liu Zechen, et al. Joint trajectory and resource optimization for RIS assisted UAV cognitive radio[J]. IEEE Transactions on Vehicular Technology, 2023, 72(10): 13643-13648.
- [42] 万诗晴, 仲伟志, 何艺, 等. 基于深度强化学习的可重构智能超表面辅助无人机通信联合波束成形与轨迹优化[J]. 电波科学学报, 2024, 39(4): 722-731.
Wan Shiqing, Zhong Weizhi, He Yi, et al. The optimization of beamforming and trajectory for reconfigurable intelligent surface assisted UAV communication system based on deep reinforcement learning[J]. Chinese Journal of Radio Science, 2024, 39(4): 722-731. (in Chinese)
- [43] Jiao Shiyu, Xie Ximing, Ding Zhiguo. Deep reinforcement learning-based optimization for RIS-based UAV-NOMA downlink networks (invited paper) [J]. Frontiers in Signal Processing, 2022, 2: 915567.
- [44] Chen Shuo, Liu Zhongda, Wang Jianwei, et al. RIS assisted UAV trajectory planning based on deep reinforcement learning[C]//Proceedings of the 2nd International Conference on Networks, Communications and Intelligent Computing (NCIC 2024). Singapore: Springer Nature Singapore, 2025: 1079-1090.
- [45] Dong Runze, Wang Buhong, Cao Kunrui, et al. Secure transmission design of RIS enabled UAV communication networks exploiting deep reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2024, 73(6): 8404-8419.
- [46] Yang Helin, Liu Shuai, Xiao Liang, et al. Learning-based reliable and secure transmission for UAV-RIS-assisted communication systems[J]. IEEE Transactions on Wireless Communications, 2024, 23(7): 6954-6967.
- [47] Haarnoja T, Zhou A, Abbeel P, et al. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor[C]//Proceedings of the 35th International Conference on Machine Learning. PMLR, 2018, 80: 1861-1870.
- [48] Adam A B M, Diallo E M, Elhassan M A M. Adaptive soft actor-critic framework for RIS-assisted and UAV-aided communication[PP/OL]. arXiv (2024-11-16)[2026-04-24]. <https://arxiv.org/abs/2411.10882>.
- [49] Nguyen K K, Masaracchia A, Sharma V, et al. RIS-assisted UAV communications for IoT with wireless power transfer using deep reinforcement learning[J]. IEEE Journal of Selected Topics in Signal Processing, 2022, 16(5): 1086-1096.
- [50] Hong Zheng, Zhao Sai, Huang Gaofei, et al. RIS-assisted UAV NOMA secure communication based on deep reinforcement learning[J]. Physical Communication, 2025, 72: 102713.
- [51] Pang Xiaowei, Zhao Nan, Tang Jie, et al. IRS-assisted secure UAV transmission via joint trajectory and beamform-

- ing design[J]. IEEE Transactions on Communications, 2022, 70(2): 1140-1152.
- [52] 胡浪涛, 毕松姣, 刘全金, 等. 基于强化学习的智能超表面辅助无人机通信系统物理层安全算法[J]. 电子与信息学报, 2022, 44(7): 2407-2415.
- Hu Langtao, Bi Songjiao, Liu Quanjia, et al. Physical layer security algorithm of reconfigurable intelligent surface-assisted unmanned aerial vehicle communication system based on reinforcement learning[J]. Journal of Electronics & Information Technology, 2022, 44(7): 2407-2415. (in Chinese)
- [53] Xu Qian, You Qian, Gong Yanyun, et al. RIS-assisted UAV-enabled green communications for industrial IoT exploiting deep learning[J]. IEEE Internet of Things Journal, 2024, 11(16): 26595-26609.
- [54] Feng Wanmei, Tang Jie, Wu Qingqing, et al. Resource allocation for power minimization in RIS-assisted multi-UAV networks with NOMA[J]. IEEE Transactions on Communications, 2023, 71(11): 6662-6676.
- [55] Moon S, Lee C G, Liu Huaping, et al. Deep reinforcement learning-based sum rate maximization for RIS-assisted ISAC-UAV network[J]. ICT Express, 2024, 10(5): 1174-1178.
- [56] Zhao Binglin, Zhai Linbo, Sun Jiande, et al. Beamforming design and trajectory optimization for learning-based multi-UAV-assisted integrated sensing and communication systems[J]. Computer Networks, 2025, 273: 111751.
- [57] Wang Wei, Li Peizheng, Doufexi A, et al. A heuristic-integrated DRL approach for phase optimization in large-scale RISs[J]. IEEE Communications Letters, 2025, 29(7): 1579-1583.
- [58] Tang Xiao, Xiong Zhihao, Dong Limeng, et al. UAV-enabled aerial active RIS with learning deployment for secured wireless communications[J]. Chinese Journal of Aeronautics, 2025, 38(10): 103383.
- [59] Wang Jianwei, Chen Shuo. Deep reinforcement learning-based secrecy rate optimization for simultaneously transmitting and reflecting reconfigurable intelligent surface-assisted unmanned aerial vehicle-integrated sensing and communication systems[J]. Sensors, 2025, 25(5): 1541.
- [60] Tesfaw B A, Juang R T, Lin H P, et al. Multi-agent deep reinforcement learning based optimizing joint 3D trajectories and phase shifts in RIS-assisted UAV-enabled wireless communications[J]. IEEE Open Journal of Vehicular Technology, 2024, 5: 1712-1726.
- [61] Huroon A M, Tarekegn G B, Abdo A M A, et al. Multi-agent deep reinforcement learning for AAV-RIS-assisted integrated sensing and communication[J]. IEEE Internet of Things Journal, 2025, 12(19): 40083-40097.
- [62] Wang Liang, Wang Kezhi, Pan Cunhua, et al. Multi-agent deep reinforcement learning-based trajectory planning for multi-UAV assisted mobile edge computing[J]. IEEE Transactions on Cognitive Communications and Networking, 2021, 7(1): 73-84.
- [63] Zhao Jingjing, Yu Lanchenhui, Cai Kaiquan, et al. RIS-aided ground-aerial NOMA communications: A distributionally robust DRL approach[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(4): 1287-1301.
- [64] Qin Zhenquan, Liu Zhonghao, Han Guangjie, et al. Distributed UAV-BSs trajectory optimization for user-level fair communication service with multi-agent deep reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2021, 70(12): 12290-12301.
- [65] Li Linpei, Guan Wanqing, Zhao Chuan, et al. Trajectory planning, phase shift design, and IoT devices association in flying-RIS-assisted mobile edge computing[J]. IEEE Internet of Things Journal, 2024, 11(1): 147-157.
- [66] 常宽, 张雷, 王玉, 等. 辅助车辆通信的双智能反射面相移优化与无人机轨迹控制[J]. 电子与信息学报, 2025, 47(10): 3670-3679.
- Chang Kuan, Zhang Lei, Wang Yu, et al. Dual-reconfigurable intelligent surface phase shift optimization and unmanned aerial vehicle trajectory control for vehicle communication[J]. Journal of Electronics & Information Technology, 2025, 47(10): 3670-3679. (in Chinese)
- [67] 周毅, 晋占齐, 石华光, 等. 空中智能反射面辅助的无线供能通信网络轨迹优化研究[J]. 电子与信息学报, 2024, 46(7): 2812-2820.
- Zhou Yi, Jin Zhanqi, Shi Huaguang, et al. Trajectory optimization research of wireless power communication networks assisted by aerial intelligent reflecting surface[J]. Journal of Electronics & Information Technology, 2024, 46(7): 2812-2820. (in Chinese)
- [68] 贾向东, 张鑫, 原帅前, 等. IRS辅助的UAV无线传感网络数据采集优化方案[J]. 信号处理, 2024, 40(6): 1041-1051.
- Jia Xiangdong, Zhang Xin, Yuan Shuaiqian, et al. IRS-assisted UAV wireless sensor network data collection optimization scheme[J]. Journal of Signal Processing, 2024, 40(6): 1041-1051. (in Chinese)
- [69] Dai Linglong, Wang Bichai, Wang Min, et al. Reconfigurable intelligent surface-based wireless communications:

- Antenna design, prototyping, and experimental results[J]. IEEE Access, 2020, 8: 45913-45923.
- [70] Wei Xiuhong, Dai Linglong, Zhao Yajun, et al. Codebook design and beam training for extremely large-scale RIS: Far-field or near-field? [J]. China Communications, 2022, 19(6): 193-204.
- [71] Wei Xiuhong, Shen Decai, Dai Linglong. Channel estimation for RIS assisted wireless communications - Part I: Fundamentals, solutions, and future opportunities[J]. IEEE Communications Letters, 2021, 25(5): 1398-1402.
- [72] Dai Linglong, Wei Xiuhong. Distributed machine learning based downlink channel estimation for RIS assisted wireless communications[J]. IEEE Transactions on Communications, 2022, 70(7): 4900-4909.
- [73] Di Renzo M, Zappone A, Debbah M, et al. Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(11): 2450-2525.
- [74] Liu Yuanwei, Liu Xiao, Mu Xidong, et al. Reconfigurable intelligent surfaces: Principles and opportunities[J]. IEEE Communications Surveys & Tutorials, 2021, 23(3): 1546-1577.
- [75] Abeywickrama S, Zhang Rui, Wu Qingqing, et al. Intelligent reflecting surface: Practical phase shift model and beamforming optimization[J]. IEEE Transactions on Communications, 2020, 68(9): 5849-5863.
- [76] 徐俊杰, 李斌, 杨敬松. 禁飞区约束下的无人机可重构智能表面辅助通信网络性能优化[J]. 电子与信息学报, 2026, 48(2): 743-751.
- Xu Junjie, Li Bin, Yang Jingsong. Performance optimization of UAV-RIS-assisted communication networks under No-fly zone constraints[J]. Journal of Electronics & Information Technology, 2026, 48(2): 743-751. (in Chinese)
- [77] Zhang Zhengming, Huang Yongming, Zhang Cheng, et al. Digital twin-enhanced deep reinforcement learning for resource management in networks slicing[J]. IEEE Transactions on Communications, 2024, 72(10): 6209-6224.
- [78] Kurma S, Singh K, Sharma P K, et al. DRL approach for spectral-energy trade-off in RIS-assisted full-duplex multi-user MIMO systems[C]//Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE, 2023: 1-6.
- [79] Shang Bodong, Poor H V, Liu Lingjia. Aerial reconfigurable intelligent surfaces meet mobile edge computing[J]. IEEE Wireless Communications, 2022, 29(6): 104-111.
- [80] Ning Zhaolong, Li Tengfeng, Wu Yu, et al. 6G communication new paradigm: The integration of autonomous aerial vehicles and intelligent reflecting surfaces[J]. IEEE Communications Surveys & Tutorials, 2025, 27(6): 3382-3416.
- [81] Huang Chongwen, Yang Zhaohui, Alexandropoulos G C, et al. Multi-hop RIS-empowered terahertz communications: A DRL-based hybrid beamforming design[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(6): 1663-1677.
- [82] Cui Mingyao, Wu Zidong, Lu Yu, et al. Near-field MIMO communications for 6G: Fundamentals, challenges, potentials, and future directions[J]. IEEE Communications Magazine, 2023, 61(1): 40-46.
- [83] Liu Yuanwei, Ouyang Chongjun, Wang Zhaolin, et al. Near-field communications: A comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2025, 27(3): 1687-1728.
- [84] Le Ha an, Van Chien T, Choi W. Graph neural network-based active and passive beamforming for distributed STAR-RIS-assisted multi-user MISO systems[J]. IEEE Transactions on Communications, 2025, 73(10): 9299-9312.
- [85] Chen Lili, Lu K, Rajeswaran A, et al. Decision transformer: Reinforcement learning via sequence modeling[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2021: 15084-15097.
- [86] 陈真, 杜晓宇, 唐杰, 等. 基于深度强化学习的RIS辅助通感融合网络: 挑战与机遇[J]. 电子与信息学报, 2024, 46(9): 3467-3473.
- Chen Zhen, Du Xiaoyu, Tang Jie, et al. DRL-based RIS-assisted ISAC network: Challenges and opportunities[J]. Journal of Electronics & Information Technology, 2024, 46(9): 3467-3473. (in Chinese)
- [87] 张广驰, 何梓楠, 崔苗. 基于深度强化学习的无人机辅助移动边缘计算系统能耗优化[J]. 电子与信息学报, 2023, 45(5): 1635-1643.
- Zhang Guangchi, He Zinan, Cui Miao. Energy consumption optimization of unmanned aerial vehicle assisted mobile edge computing systems based on deep reinforcement learning[J]. Journal of Electronics & Information Technology, 2023, 45(5): 1635-1643. (in Chinese)
- [88] Li Hongyu, Shen Shanpu, Nerini M, et al. Reconfigurable intelligent surfaces 2.0: Beyond diagonal phase shift matrices[J]. IEEE Communications Magazine, 2024, 62(3):

102-108.

[89] 李国权, 胡航, 王玥涛, 等. STAR-RIS辅助的CR-SWIPT系统安全波束成形算法[J]. 电子学报, 2024, 52(12): 4002-4008.

Li Guoquan, Hu Hang, Wang Yuetao, et al. Secure beam-forming algorithm for STAR-RIS assisted cognitive radio systems with SWIPT[J]. Acta Electronica Sinica, 2024, 52(12): 4002-4008. (in Chinese)

作者简介



王恩良 男, 1998年2月出生于江苏省南京市。现为南京邮电大学物联网学院博士研究生。主要研究方向为神经架构原理、强化学习、网络与通信技术。
E-mail: 3386204603@qq.com



李 琨 男, 2001年12月出生于山东省潍坊市。现为南京邮电大学现代邮政学院硕士研究生。主要研究方向为强化学习、低空技术。
E-mail: 19563132060@163.com



张炜荣 男, 2001年6月出生于福建省福州市。现为南京邮电大学现代邮政学院硕士研究生。主要研究方向为强化学习。
E-mail: 1224097333@njupt.edu.cn



孙知信 男, 1964年9月出生于江苏省南京市。现为南京邮电大学现代邮政学院教授、博士生导师。主要研究方向为物联网技术、计算机网络、通信技术。
E-mail: Sunzx@njupt.edu.cn