基于扩散强化学习的UAV能量导向轨迹规划与物联网服务质量优化

张子天，葛天豪，诸葛斌，郑运强，董黎刚，蒋献
（浙江工商大学，浙江 杭州 310018）

摘　要：为解决物联网（Internet of Things，IoT）设备产生的异构计算任务需要高效调度的问题，提出基于扩散强化学习（Diffusion Reinforcement Learning，DiffRL）的无人机（Unmanned Aerial Vehicle，UAV）通信网络与任务卸载系统，主要创新点包括：（1）扩散强化学习卸载决策框架，通过去噪扩散隐式模型（Denoising Diffusion Implicit Model，DDIM）采样技术将采样步数从 50 步减少到 15 步，加速 70% 的采样过程，同时保持 98% 的决策质量；（2）能量导向无人机轨迹规划算法，降低系统总能耗 15.3%；（3） 实现 DiffRL 决策与轨迹规划的紧耦合，解决动态环境下多目标优化问题。实验表明，本系统在能耗和任务延迟方面较传统方法均有提升，在任务卸载决策中比传统深度 Q 网络（Deep Q-Network，DQN）和深度确定性策略梯度（Deep Deterministic Policy Gradient，DDPG）算法分别降低任务延迟 30.2% 和 9.2%。

关键词：UAV通信网络；扩散强化学习；移动边缘计算；任务卸载决策
中图分类号：TN929.5
文献标志码：A
doi：10.11959/j.issn.2096−3750.XXXX.

UAV Energy-Oriented Trajectory Planning and IoT Service Quality Optimization Based on Diffusion Reinforcement Learning

ZAHNG Zitian, GE Tianhao, ZHUGE Bin, ZHENG Yunqiang, DONG Ligang, JAING Xian
Zhejiang Gongshang University, Hangzhou 310018, China

Abstract: To address the high-frequency heterogeneous computing task requirements generated by IoT devices, this paper proposes a Unmanned Aerial Vehicle(UAV) communication network and task offloading system based on diffusion reinforcement learning(DiffRL).The main innovations include: (1) A diffusion reinforcement learning offloading decision framework that reduces sampling steps from 50 to 15 steps through denoising diffusion implicit model(DDIM)sampling technology, accelerating the sampling process by 70% while maintaining 98% of decision quality; (2) An energy-oriented UAV trajectory planning algorithm that reduces total system energy consumption by 15.3%;(3)Achieving tight coupling between DiffRL decision-making and trajectory planning to solve multi-objective optimization problems in dynamic environments. Experiments show that this system achieves improvements in both energy consumption and task delay compared to traditional methods, reducing task delay by 30.2% and 9.2% respectively when compared to traditional Deep Q-network(DQN)and deep deterministic policy gradient algorithms (DDPG) in task offloading decisions.

Key words: UAV Communication Network, Diffusion Reinforcement Learning, Mobile Edge Computing, Task Offloading Decision

收稿日期：XXXX−XX−XX；**修回日期：**XXXX−XX−XX
通信作者：张子天，邮件：zitian.zhang@mail.zjgsu.edu.cn
基金项目：国家自然科学基金资助项目（No.62301488,W2421086）
Foundation Items: the National Science Foundation of China(No.62301488, W2421086)

1 引言

随着物联网技术的快速发展，移动边缘计算（Mobile Edge Computing，MEC）成为解决分布式设备计算需求的关键技术。无人机作为移动MEC节点，因其灵活性和广阔覆盖范围而备受关注。Zeng和Zhang[1]的研究表明，通过轨迹优化可以显著提高UAV通信的能源效率。然而，UAV网络面临着能源受限、通信环境动态变化等挑战，这对任务卸载决策提出了更高要求。Zhou等人[2]在研究UAV无线供电的移动边缘计算系统时，提出了计算率最大化的优化框架，但其静态优化假设和能量模型简化导致与实际系统需求存在显著差距。

近年来，深度强化学习在UAV控制和任务卸载决策中展现出巨大潜力。Mnih等人[3]提出的深度Q网络在离散动作空间中取得了突破性成功，但在处理高维连续状态和动作空间时，仍面临着采样效率低、收敛速度慢等问题。Schulman等人[4]提出的近端策略优化（Proximal Policy Optimization，PPO）算法通过引入信任域约束提高了训练稳定性，但在复杂动态环境中仍难以保证策略的一致性。扩散模型因其强大的生成能力和稳定的训练过程而在图像生成等领域取得突破。Janner等人[5]的研究表明，扩散模型在连续控制任务中也具有潜力，其提出的扩散规划方法为行为合成提供了新的思路。 在面向空天地一体化网络的复杂资源分配场景中，AI驱动的方法，特别是深度强化学习与扩散模型结合正成为解决动态资源约束、实现多目标优化的关键范式[6]。Hansen-Estruch等人[7]证明扩散策略在离线强化学习中优于传统策略。Du等人[8]将扩散模型与深度强化学习深度融合，提出深度扩散软演员-评论家算法，通过扩散模型生成最优离散决策方案，验证了DiffRL离散动作空间优化中的突破性优势。受此启发，本文提出了一种基于DiffRL的UAV通信网络与任务卸载系统，通过创新的DiffRL框架和能量导向轨迹规划算法，实现高效的任务处理和能源优化。

1.1 研究背景与挑战

随着物联网设备的快速增长，UAV作为移动边缘计算节点在通信网络中发挥着越来越重要的作用。然而，UAV网络面临着多重挑战：首先，UAV具有有限的能源容量$E_{\max}$和计算能力$F_{\max}$，这严重限制了其服务时长和计算能力；其次，UAV与地面设备间的通信信道质量$h_{ij}(t)$随时间和位置动态变化，这增加了网络管理的复杂性，UAV飞行速度与信道衰落的动态交互，导致传统静态优化方法难以适应实时通信需求；最重要的是，任务卸载决策需要同时权衡能耗E_{total}、延迟T_{total}和通信质量Q_{comm}等多个目标，这使得决策优化变得极其困难。传统的确定性算法难以适应如此复杂的动态环境，而传统深度强化学习（如DQN、DDPG）在UAV任务卸载中面临三大局限：（1）高维连续动作空间探索效率低，导致收敛速度慢；（2）策略分布建模能力弱，难以捕捉多模态优化解；（3）动态通信环境下策略波动大。DiffRL通过构建马尔可夫链式生成过程，利用去噪扩散隐式模型（DDIM）的非参数分布建模能力，实现对高维动作空间的精确采样，显著提升策略表达的丰富性和稳定性。

当前基于扩散模型的规划方法虽在连续控制任务中展现出强大的生成能力，但现有方法假设通信信道静态稳定，未考虑UAV高速移动导致的信道时变特性以及UAV网络的核心约束。因此，如何在有限资源约束下实现高效的任务处理和能源优化，成为当前研究的重要挑战。

1.2 本文创新点

本文提出了一种基于DiffRL的UAV通信网络与任务卸载系统，现有扩散模型未考虑UAV信道时变特性。本文创新在于：

● 将DDIM采样与通信感知扩散过程结合，动态适应信道变化，利用Song等人[9]提出的去噪扩散隐式模型的非参数分布建模能力，实现对高维连续动作空间的精确建模。将DDIM采样与通信感知扩散过程结合，动态适应信道变化，将采样步数从50步减少到15步，在保持98%决策质量的同时，实现了约70%的采样加速；

● 基于Wu等人[10]提出的联合轨迹和通信设计方法，设计了一种综合考虑通信质量和能源消耗的轨迹优化方法。实现DiffRL决策与能量导向轨迹的联合优化，突破静态优化假设，通过自适应航点生成机制，实现UAV能源效率与服务质量的动态平衡。该算法在降低系统总能耗15.3%的同时，保持了任务处理性能。作为辅助技术，借鉴Liu等人[11]在多波束UAV通信中提出的协作干扰消除方法，提出了一种基于能源感知的多UAV协调系统，

解决密集部署干扰问题，进一步增强系统整体性能。

2　相关工作

2.1　UAV通信网络

在UAV通信网络领域，早期研究主要集中在单UAV场景下的轨迹优化和资源分配。Zhang等人[12]提出了一种基于无线电地图的3D路径规划方法，通过考虑通信质量和空间约束进行轨迹优化。Liu等人[13]研究了多UAV网络中的部署和移动设计问题，提出了一种基于强化学习的优化方法。Khan等人[14]验证了多无人机协作在覆盖扩展与资源弹性供给中的优势。然而，这些方法主要基于确定性优化，难以应对动态环境。

Zheng与Chen[15]提出了地理感知的UAV三维最优部署方法，证明了仅需在二维中间垂直平面内搜索即可获得最优三维位置，为动态环境中的UAV部署提供了理论保证。然而，现有方法难以在能源约束、信道动态性及多目标冲突下实现高效联合优化，在实时资源分配场景中仍存在局限性。

2.2　任务卸载决策

任务卸载决策是MEC系统中的关键问题。Chen等人[16]提出了一种基于缓存的UAV部署策略，通过主动缓存和预测用户需求来优化服务质量。Liu等人[17]研究了能效UAV控制问题，提出了一种基于深度强化学习的方法来实现有效和公平的通信覆盖。Haarnoja等人[18]提出的软演员-评论家算法为连续控制问题提供了新的解决思路。然而在6G赋能的物联网高密度场景中，传统方法仍面临动态资源分配效率与多目标优化的双重挑战[19]。本文提出的DiffRL框架不仅能够处理高维连续动作空间，还通过DDIM采样技术显著提高了决策效率。

2.3　扩散模型与强化学习

扩散模型作为一种生成模型，近年来在多个领域取得了显著进展。Sohl-Dickstein等人[20]首次提出基于非平衡热力学的扩散模型框架，定义前向加噪和反向去噪的马尔可夫链过程，为后续DDPM、DDIM等研究奠定核心理论基础。Ho等人[21]首次提出了去噪扩散概率模型（DDPM），通过逐步添加和移除高斯噪声来生成高质量样本。在强化学习领域，Wang等人[22]探索了将扩散模型与策略优化相结合的方法，提出了一种基于扩散的策略学习框架，能够有效处理连续动作空间中的复杂决策问题。Du等人[23]深入分析了扩散模型在通信资源分配、任务卸载等网络优化问题中的方法论优势，特别是DiffRL通过结合生成能力与决策优化，显著提升了高维连续动作空间中的策略表达能力。这些研究为本文提出的DiffRL框架奠定了重要的理论基础。与传统的强化学习方法相比，基于扩散模型的方法能够更好地捕捉动作分布的多模态特性，并通过去噪过程实现更稳定的策略优化。

综上所述，现有研究虽已证明扩散模型在通用连续控制任务中的潜力，并提供了基础理论框架，但尚未深入探索其在资源高度受限、环境动态复杂且需同时优化多目标的UAV通信网络任务卸载决策中的具体应用与优化。本文的创新在于将DiffRL模型针对性应用于该特定场景，将DiffRL生成的高质量卸载决策与能量导向轨迹规划算法进行紧耦合联合优化并提出了包括DDIM加速采样、通信感知扩散过程等关键理论改进。

3　系统模型

3.1　系统架构

本系统包含三层网络架构：IoT设备层、UAV边缘计算通信层和卫星计算层。IoT设备生成计算任务后，可选择本地执行、卸载至UAV或卸载至卫星。系统优化目标是综合最小化网络能耗、通信延迟和任务处理时间。如图1所示，系统中的通信链路可分为控制链路和数据链路，控制链路用于UAV之间的状态同步和任务协调，支持分布式决策和资源调度，数据链路承载任务数据的上下行传输。能源状态用于监控和报告UAV的能源状态。该架构通过多层次的协同机制和灵活的任务处理策略，实现了高效、可靠的通信服务。

本文的优化目标是在能源、延迟和通信质量约束下，最小化加权目标函数：

$$\begin{aligned} \min_{\pi} \quad & w_1 E_{\text{total}} + w_2 T_{\text{total}} - w_3 Q_{\text{comm}} \\ s.t. \quad & E_{\text{total}} \leqslant E_{\max} \\ & T_{\text{total}} \leqslant T_{\text{deadline}} \\ & Q_{\text{comm}} \geqslant Q_{\min} \end{aligned} \tag{1}$$

其中w_1、w_2和w_3为权重系数，T_{deadline}为任务截止时间，$Q_{\min}$为最低通信质量要求。

3.2　无人机通信网络模型

本文的异构指计算资源请求（CPU/GPU）的

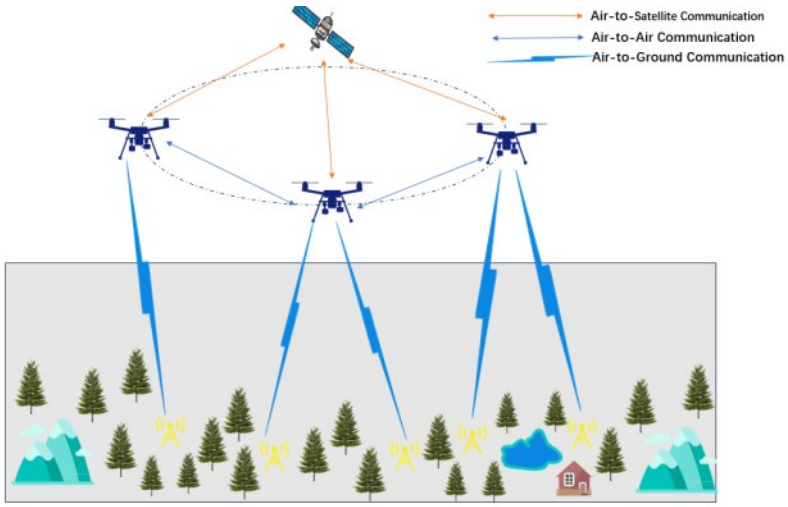


图1　UAV通信网络与任务卸载系统架构

差异，具体体现为任务计算复杂度和数据大小的多样性。考虑一个包含b架UAV的通信网络系统，覆盖区域$A=[0,X]\times[0,Y]$，区域内分布a个IoT设备。UAV i的位置表示为$p_i=(x_i,y_i,z_i)$，其中z_i为飞行高度。UAV的通信和移动受最大速度$v_{\max}$、信道带宽B和能量约束$E_{\max}$限制。

3.2.1　通信信道模型

UAV与地面设备间的通信采用高斯衰减模型。在理论分析中，信道增益$h_{i,d}$可表示为：

$$h_{i,d}=\exp\left(-\frac{1}{2}\left(\frac{\mathrm{dist}_{i,d}}{R_{\mathrm{comm}}}\right)^2\right) \tag{2}$$

其中$\mathrm{dist}_{i,d}$为UAV i与设备d之间的欧氏距离，R_{comm}为通信范围。当存在障碍物时，信道增益会进一步衰减：

$$h_{i,d}^{\mathrm{obs}}=h_{i,d}\cdot\prod\nolimits_{k\in O}\eta_k \tag{3}$$

其中O为障碍物集合，$\eta_k\in[0,1]$为障碍物k导致的衰减系数。在实际系统中，设置$\eta_k=0.5$表示障碍物造成50%的信号衰减。

基于此信道模型，UAV i与设备d之间的通信容量$C_{i,d}$可表示为：

$$C_{i,d}=B\cdot\log_2\left(1+\frac{P_t\cdot h_{i,d}}{N_0+I_{i,d}}\right) \tag{4}$$

其中B为信道带宽，P_t为发射功率，N_0为背景噪声功率，$I_{i,d}$为干扰功率，来自其他UAV的同频干扰。

3.2.2　能量消耗模型

UAV的能量消耗模型包含四部分：

$$E_{\mathrm{total}}=E_{\mathrm{hover}}+E_{\mathrm{movement}}+E_{\mathrm{altitude}}+E_{\mathrm{comm}} \tag{5}$$

其中，E_{hover}为悬停能耗，E_{movement}为移动能耗，E_{altitude}为高度相关能耗，E_{comm}为通信能耗。在系统实现中，我们对各部分能耗进行了优化建模：

$$E_{\mathrm{hover}}=P_{\mathrm{hover}}\cdot t \tag{6}$$

其中，P_{hover}为UAV悬停功率，通过优化设置为85W，t为悬停时间。悬停能耗与悬停时间成正比关系。

$$E_{\mathrm{movement}}=P_{\mathrm{move}}\cdot v^2\cdot t\cdot\eta_{\mathrm{speed}} \tag{7}$$

其中，P_{move}为移动功率系数，v为UAV飞行速度，η_{speed}为速度效率因子，表示不同速度下的能效变化。此公式体现了UAV移动能耗与速度平方成正比的物理特性。

$$E_{\mathrm{altitude}}=E_{\mathrm{hover}}\cdot(\alpha_{\mathrm{alt}}\cdot\frac{z}{10}-1) \tag{8}$$

其中，α_{alt}为高度影响系数，z为飞行高度。高度相关能耗反映了高度对UAV能耗的影响，高度每增加10m，能耗增加α_{alt}倍。

$$E_{\mathrm{comm}}=\sum\nolimits_{d\in D_i}P_{\mathrm{trans}}\cdot t_{\mathrm{comm}}\cdot f(q_d)\cdot\eta_{\mathrm{coord}} \tag{9}$$

其中，D_i为UAV i服务的设备集合，P_{trans}为通信传输功率，t_{comm}为通信时长，$f(q_d)$为通信质量修正因子，η_{coord}为协调效率系数，表示多UAV协调对能耗的影响。

通信能耗E_{comm}与通信质量q_d、通信时长t_{comm}和协调效率η_{coord}相关，其中通信质量q_d定义为：

$$q_d=\min\{1,\frac{C_{i,d}}{C_{\mathrm{req}}}\} \tag{10}$$

其中，$C_{i,d}$为UAV i与设备d之间的实际通信容量（单位：Mbps），C_{req}为满足通信需求的最低容量阈值（单位：Mbps）。通信质量修正因子$f(q_d)=1/q_d^2$

表示通信质量越低，为保证传输成功率所需的能耗越高。

3.2.3 干扰协调模型

多UAV系统中，同频干扰是影响通信质量的关键因素。UAV j 对UAV i 服务的设备 d 的干扰功率可表示为：

$$I_{j\to i,d} = P_t \cdot h_{j,d} \cdot \psi(f_i,f_j) \tag{11}$$

其中，$I_{j\to i,d}$ 表示UAV j 对UAV i 服务的设备 d 造成的干扰功率，P_t 为发射功率，$h_{j,d}$ 为UAV j 与设备 d 之间的信道增益，$\psi(f_i,f_j)$ 为频率重叠因子。当UAV i 和 j 使用相同频率时 $\psi(f_i,f_j)=1$，使用正交频率时 $\psi(f_i,f_j)=0$，部分重叠时取0-1之间的值。

通过协调机制，系统可优化频率分配和空间部署，使总干扰降至最低：

$$I_{i,d} = \sum_{j\neq i} I_{j\to i,d} \tag{12}$$

其中，$I_{i,d}$ 表示设备 d 接收到的来自所有其他UAV的总干扰功率。通过协调机制，系统旨在最小化这一总干扰值，提高通信质量。

3.3 任务模型

IoT设备生成的任务由一个六元组表示：$\boldsymbol{\tau} = (id,d,\lambda,c,\rho,t_{\text{crt}})$。其中，$id$ 为任务的唯一标识符，用于在系统中追踪和管理任务；d 为生成该任务的设备标识符，用于定位任务来源；λ 表示任务数据大小，单位为千字节(KB)，反映了数据传输需求；c 为任务的计算复杂度，以百万指令(MI)为单位，表示处理该任务所需的计算资源；ρ 表示任务优先级，采用1-4的整数值分别对应低、中、高、紧急四个等级，用于任务调度决策；t_{crt} 记录任务的创建时间，用于计算任务处理延迟和评估系统性能。

任务执行目标包括本地执行、UAV执行和卫星执行。任务处理总时延包括传输时延和计算时延：

$$T_{\text{total}} = T_{\text{trans}} + T_{\text{comp}} \tag{13}$$

其中，T_{total} 为任务处理总时延，T_{trans} 为数据传输时延，T_{comp} 为计算处理时延。

传输时延与通信容量直接相关：

$$T_{\text{trans}} = \frac{\lambda \cdot 8}{C_{i,d}} \tag{14}$$

其中，$\lambda \cdot 8$ 表示数据大小（单位：kb），表示将数据大小从千字节（KB）转换为千比特（Kb），$C_{i,d}$ 为通信容量（单位：Mbps）。传输时延与数据大小成正比，与通信容量成反比。

计算时延与计算复杂度和计算能力相关：

$$T_{\text{comp}} = \frac{c}{F} \tag{15}$$

其中，c 为任务的计算复杂度，F 为计算设备的计算能力。计算时延与计算复杂度成正比，与计算能力成反比。

任务处理的总能耗包括传输能耗和计算能耗：

$$E_{\text{total}} = E_{\text{trans}} + E_{\text{comp}} \tag{16}$$

其中，E_{total} 为任务处理总能耗，E_{trans} 为传输能耗，E_{comp} 为计算能耗。

传输能耗与通信质量和数据大小相关：

$$E_{\text{trans}} = \lambda \cdot e_{\text{trans}} \cdot \frac{1}{q_d} \tag{17}$$

其中，λ 为数据大小，e_{trans} 为单位传输能耗基准值，q_d 为通信质量因子。通信质量越低，为保证数据可靠传输所需的能耗越高，体现了低质量通信下的重传和错误恢复开销。

4 方法

4.1 DiffRL模型

本文优化问题（公式1）属于高维、非线性、多目标随机规划。传统强化学习（如DDPG）因策略分布建模能力弱，难以收敛至全局最优。DiffRL通过扩散过程捕捉动作空间多模态特性，结合DDIM采样实现高效决策，适用于动态通信环境

DiffRL模型通过构建马尔可夫链来实现从噪声到动作的生成过程。该过程包含以下关键步骤：

● 通过逐步添加高斯噪声将初始状态 $\boldsymbol{x}_0 \in R^n$ 映射到噪声分布，n 为状态空间的维度：

$$q_{\text{comm}}(\boldsymbol{x}_t|\boldsymbol{x}_0,\boldsymbol{h}) = \mathrm{N}(\boldsymbol{x}_t;\mu(\boldsymbol{x}_0,\boldsymbol{h},t),\sigma^2(t)\mathbf{I}) \tag{18}$$

其中 $\boldsymbol{x}_t$ 为 t 时刻的状态向量，$\boldsymbol{h} = [h_{1,1},\cdots,h_{a,b}]^\top \in R^{a\times b}$ 为信道状态矩阵，a 为物联网设备数量，b 为UAV数量，$\mu(\cdot)$ 为均值函数，$\sigma^2(t)$ 为时变方差函数，$\mathbf{I}$ 为单位矩阵。

● 为提高采样效率，采用DDIM技术，将采样步数从50步减少到15步。DDIM采样过程可表示为：

$$\boldsymbol{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\left(\frac{\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(\boldsymbol{x}_t)}{\sqrt{\bar{\alpha}_t}}\right) + \sqrt{1-\bar{\alpha}_{t-1}}\,\epsilon_\theta(\boldsymbol{x}_t) \tag{19}$$

其中 $\boldsymbol{x}_t \in R^d$ 为 t 时刻的状态向量，$\bar{\alpha}_t$ 为累积噪声系数，$\epsilon_\theta(\cdot)$ 为参数化的噪声预测函数，θ 为模型参数。

● 策略优化目标是最大化期望累积奖励：

$$J(\theta) = \mathrm{E}_{s\sim\rho_\pi}[V^\pi(s)] \tag{20}$$

其中$J(\theta)$为目标函数，E[·]为期望算子，s为系统状态，ρ_π为策略π下的状态分布，$V^\pi(s)$为状态价值函。

4.2 能量导向轨迹规划算法

能量导向轨迹规划算法的核心是将通信质量与运动能耗转化为航点评分，通过优化UAV路径最小化能源消耗。对于每个设备d，其评分函数为：

$$\begin{aligned}\mathrm{score}(d) = {} & w_d \cdot \mathrm{distance\_score}(d) + w_{\mathrm{dir}} \cdot \\ & \mathrm{direction\_score}(d) - w_o \cdot \mathrm{overlap\_penalty}(d)\end{aligned} \tag{21}$$

其中w_d,w_{dir}和w_o分别为距离评分、方向评分和重叠惩罚的权重系数，$\mathrm{distance_score}(d)$为距离评分函数，$\mathrm{direction_score}(d)$为方向评分函数，$\mathrm{overlap_penalty}(d)$为重叠惩罚函数。

距离评分采用高斯衰减模型：

$$\mathrm{distance\_score}(d) = \exp\left(-\frac{1}{2}\left(\frac{\mathrm{dist}}{R_{\mathrm{ref}}}\right)^2\right) \tag{22}$$

其中dist为UAV与设备的欧氏距离，R_{ref}为参考距离。

对于每个IoT设备d，算法计算一个综合评分，包含两个主要组成部分：

● 考虑UAV当前运动方向与设备方向的一致性，方向评分表示为：

$$\mathrm{direction\_score}(d) = \frac{1+\cos(\theta)}{2} \tag{23}$$

其中θ为UAV当前速度向量与UAV到设备的向量之间的夹角。

● 为了避免多个UAV覆盖同一区域，设置重叠惩罚：

$$\mathrm{overlap\_penalty}(d) = \sum_{j\neq i}\exp\left(-\frac{\left\| p_j - p_d \right\|^2}{2\sigma^2}\right) \tag{24}$$

其中p_j为其他UAV的位置，p_d为设备位置，σ为重叠敏感度参数。

基于上述评分机制，能量导向轨迹算法执行以下步骤：

● 设置UAV初始位置、速度和能量状态，初始化设备列表和障碍物信息，计算初始通信覆盖范围。

● 每个时间步Δt更新UAV位置，计算所有可达设备的评分，选择最高评分的目标设备，根据能量约束调整飞行速度和高度。

● 实时监控能量消耗，当能量水平低于阈值E_{th}时启动节能模式，调整飞行参数以最小化能耗：

$$v_{\mathrm{opt}} = \min\left\{ v_{\max}, \sqrt{\frac{2E_{\mathrm{remain}}}{k_e \cdot t_{\mathrm{est}}}} \right\} \tag{25}$$

其中v_{opt}为优化后的速度，E_{remain}为剩余能量，k_e为能耗系数，t_{est}为预估任务时间。

● 与其他UAV交换位置信息，计算重叠区域并更新评分，动态调整覆盖范围以避免干扰。

● 实时评估信道质量，考虑障碍物影响，根据通信需求调整位置：

$$h_{\mathrm{opt}} = \arg\min_{h_{\min}\le h\le h_{\max}} \{ E_{\mathrm{hover}}(h) + \lambda \cdot \mathrm{PL}(h) \} \tag{26}$$

其中h_{opt}为最优高度，PL为路径损耗，λ为权重系数。

4.3 多UAV协调机制

为了优化整体性能，多UAV协调机制主要包括以下两点：

● 为了解决密集部署UAV间的通信干扰问题，确保网络通信质量，安全距离调整可表示为：

$$d_{\mathrm{safe}} = d_{\max} - (d_{\max} - d_{\min}) \cdot \min\left(1, \frac{N}{N_{\mathrm{ref}}}\right) \tag{27}$$

其中d_{safe}为安全距离，$d_{\max}$和$d_{\min}$分别为最大和最小安全距离，N为当前UAV数量，N_{ref}为参考UAV数量。

● 旨在应对UAV个体负载不均衡或能源不足的情况，防止因单一节点过载或能源耗尽导致的任务失败或服务中断，提升系统的整体可靠性和资源利用率，我们引入了任务转发机制，当UAV负载过高或能源不足时，任务转发概率计算如下：

$$p_{\mathrm{forward}} = \max\left(0, \min\left(1, \frac{Q_{\mathrm{current}} - Q_{\mathrm{threshold}}}{Q_{\max} - Q_{\mathrm{threshold}}}\right)\right) \tag{28}$$

其中p_{forward}为任务转发概率，Q_{current}为当前任务队列长度，$Q_{\mathrm{threshold}}$为任务队列阈值，$Q_{\max}$为最大任务队列长度。

4.4 DiffRL计算优化分析

4.4.1 扩散过程的数学基础

在DiffRL中，本文将策略学习问题转化为生成建模问题。令$\boldsymbol{x}_0$为初始状态，$\boldsymbol{x}_t$为t时刻的状态，$\bar{\alpha}_t = \prod_{s=1}^{t}\alpha_s$为累积噪声系数，$\alpha_t = 1 - \beta_t$为单步噪声系数，$\beta_t$为噪声调度参数，$\mathbf{I}$为单位矩阵。前向过程可以表示为：

$$q(\boldsymbol{x}_t|\boldsymbol{x}_0)=\mathrm{N}(\boldsymbol{x}_t;\sqrt{\bar{\alpha}_t}\,\boldsymbol{x}_0,(1-\bar{\alpha}_t)\mathbf{I}) \tag{29}$$

为了实现更好的训练效果，我们采用余弦噪声调度：

$$\beta_t=\beta_{\min}+(\beta_{\max}-\beta_{\min})\cdot\frac{1+\cos(\pi t/T)}{2} \tag{30}$$

其中$\beta_{\min}$和$\beta_{\max}$分别为噪声强度的下界和上界，T为总时间步数。这种调度方式能在训练初期保持较小的噪声以保留细节信息，在后期增加噪声以提高探索能力。

4.4.2　DiffRL的优势

DiffRL的核心思想是将策略学习问题转化为生成建模问题，通过扩散过程捕捉复杂的动作分布。与传统强化学习方法相比，DiffRL具有以下关键优势：

● DiffRL使用扩散过程建模复杂的非参数策略分布，相比传统RL的简单参数化分布，能更准确捕捉多模态动作空间，实现更精细的决策控制。

● 扩散采样机制天然实现了探索与利用的平衡，避免了传统RL中需要设计额外探索策略的复杂性，通过控制扩散步数和噪声调度实现灵活的温度调节。

● DIFFRL从有限样本中提取更丰富的分布信息，显著提高样本利用效率，特别适合资源受限的UAV场景。

在我们的UAV通信网络系统中，DiffRL实现了高质量的多目标决策优化，能同时平衡能耗、延迟等多个目标，并在动态变化的通信环境中展现出色的适应能力。

4.4.3　强化学习理论分析

在DiffRL框架中，我们将策略优化问题形式化为马尔可夫决策过程(Markov Decision Process，MDP)。令$\mathbf{S}\subseteq R^n$为状态空间，包含UAV位置、能源状态和任务信息，$\mathbf{A}\subseteq R^y$为动作空间，包含任务卸载决策，n和y分别为状态和动作空间的维度。值函数V^π:$\mathbf{S}\to R$表示评估策略π在状态s下的长期回报和策略函数π:$\mathbf{S}\to\mathrm{P}(\mathbf{A})$表示将状态映射到动作概率分布，定义如下：

$$V^\pi(s)=\mathrm{E}_\pi\left[\sum_{t=0}^{\infty}\gamma^t r_t|s_0=s\right] \tag{31}$$

其中$\gamma\in[0,1]$为折扣因子，用于平衡即时和长期回报，r_t为即时奖励，包含能耗和延迟的加权和：

$$r_t=-w_E\cdot E_{\text{total}}-w_T\cdot T_{\text{total}} \tag{32}$$

其中w_E和w_T分别为能耗和延迟的权重系数，且满足$w_E+w_T=1$。

策略优化目标是最大化期望累积奖励：

$$J(\theta)=\mathrm{E}_{s\sim\rho_\pi}[V^\pi(s)] \tag{33}$$

其中ρ_π为策略π下的状态分布，θ为策略网络参数。

4.5　扩散模型在通信系统中的理论创新

鉴于3.2节信道模型中动态信道对决策的显著影响，本节提出通信感知的扩散过程优化。

4.5.1　通信感知的扩散过程

在UAV通信网络中，提出了通信感知的扩散过程。不同于标准扩散模型，我们的扩散过程同时考虑了通信质量和系统状态,条件分布可表示为：

$$q_{\text{comm}}(\boldsymbol{x}_t|\boldsymbol{x}_0,\boldsymbol{h})=\mathrm{N}(\boldsymbol{x}_t;\mu(\boldsymbol{x}_0,\boldsymbol{h},t),\sigma^2(t)\mathbf{I}) \tag{34}$$

其中$\boldsymbol{x}_t$为t时刻的状态向量，$\mathrm{N}(\cdot;\mu,\boldsymbol{\Sigma})$表示均值为$\mu$、协方差矩阵为$\boldsymbol{\Sigma}$的高斯分布，$\mathbf{I}$为单位矩阵。

均值函数$\mu(\cdot)$和方差函数$\sigma^2(\cdot)$被设计为：

$$\mu(\boldsymbol{x}_0,\boldsymbol{h},t)=\sqrt{\bar{\alpha}_t}\,\boldsymbol{x}_0+\beta_t\boldsymbol{W}_h\boldsymbol{h} \tag{35}$$

其中μ为条件均值函数，$\boldsymbol{x}_0$为初始状态向量，t为时间步，$\bar{\alpha}_t$为累积噪声系数，β_t为噪声调度参数，$\boldsymbol{W}_h$为信道感知权重矩阵。该式描述了如何将信道状态信息整合到扩散过程的均值预测中。

信道质量对扩散过程噪声强度的调节作用可以表示为：

$$\sigma^2(t)=(1-\bar{\alpha}_t)(1+\gamma\,/\!/\,\boldsymbol{h}\,/\!/_2) \tag{36}$$

其中$\sigma^2(t)$为条件方差函数，$(1-\bar{\alpha}_t)$表示累积噪声的影响，$\gamma\in R$为信道影响因子，用于控制信道状态对噪声方差的影响程度，$/\!/\,\boldsymbol{h}\,/\!/_2$为信道状态矩阵的2范数。

在实现中，我们采用余弦调度来提高模型稳定性：

$$\beta_t=\beta_{\min}+(\beta_{\max}-\beta_{\min})\cdot\frac{1+\cos(\pi t/T)}{2} \tag{37}$$

其中$\beta_{\min}=10^{-4}$，$\beta_{\max}=0.02$为噪声强度的边界值，T为总时间步数。

4.5.2　通信约束下的策略生成

在反向过程中，提出了考虑通信约束的策略生成机制。给定当前状态s_t和信道状态$\boldsymbol{h}_t$，策略函数可表示为：

$$\pi(a|s_t,\boldsymbol{h}_t)=\int p_\theta(\boldsymbol{x}_0|\boldsymbol{x}_T,\boldsymbol{h}_t)\prod_{t=1}^{T}p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,s_t,\boldsymbol{h}_t)\,\mathrm{d}\boldsymbol{x}_{1:T} \tag{38}$$

其中p_θ为参数化的条件概率分布，$\boldsymbol{x}_0$为初始状态向量，$\boldsymbol{x}_T$为终止状态向量，$\boldsymbol{h}_t$为t时刻的信道状态矩

阵，s_t为系统状态，θ为模型参数。条件转移概率定义为：

$$p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,s_t,\boldsymbol{h}_t)=\mathrm{N}(\boldsymbol{x}_{t-1};\mu_\theta(\boldsymbol{x}_t,s_t,\boldsymbol{h}_t),\sigma_\theta^2(t)\mathbf{I}) \quad (39)$$

其中μ_θ为神经网络预测的均值，$\sigma_\theta^2(t)$为时间相关的方差。

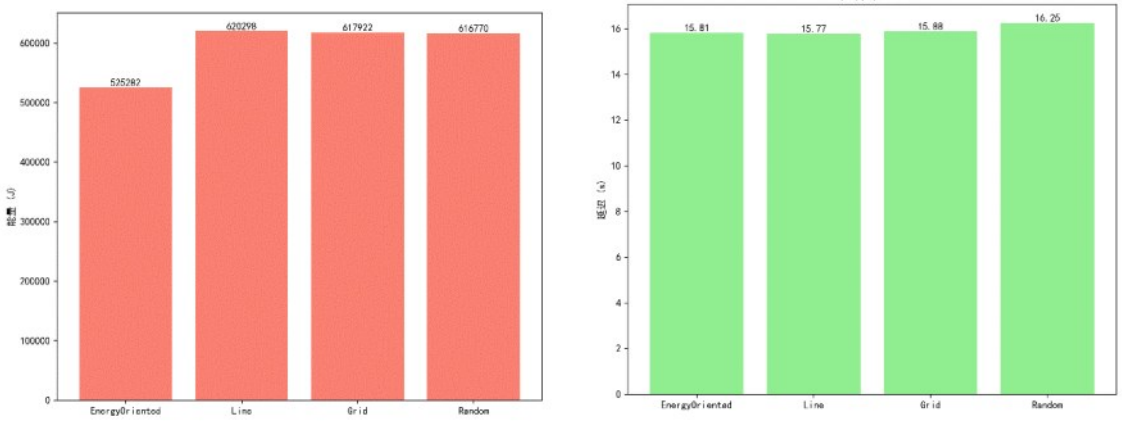
5　实验与分析

5.1　实验设置

本文实验环境为1000 m × 1000 m的区域，包含20个IoT设备和5个障碍物，部署3架UAV。

5.2　轨迹规划算法对比

为验证本文提出的能量导向轨迹规划算法的有效性，本文将其与直线轨迹、网格轨迹和随机轨迹三种基准算法进行对比。实验在相同的UAV网络环境下进行，对比指标包括系统总能耗和任务处理延迟。如图2所示，图（a）显示了能量导向轨迹算法在总能耗方面表现最优，仅需525282J，比直线轨迹、网格轨迹和随机轨迹分别节省约15.3%、14.9%和14.8%的能耗。图（b）显示在任务处理延迟方面，各算法表现相近，能量导向轨迹的平均延迟为15.81秒，与其他算法基本持平，表明能量优化并未影响系统的任务处理效率。



（a） 能耗对比　　（b） 延迟对比

图2　不同轨迹规划算法的性能对比

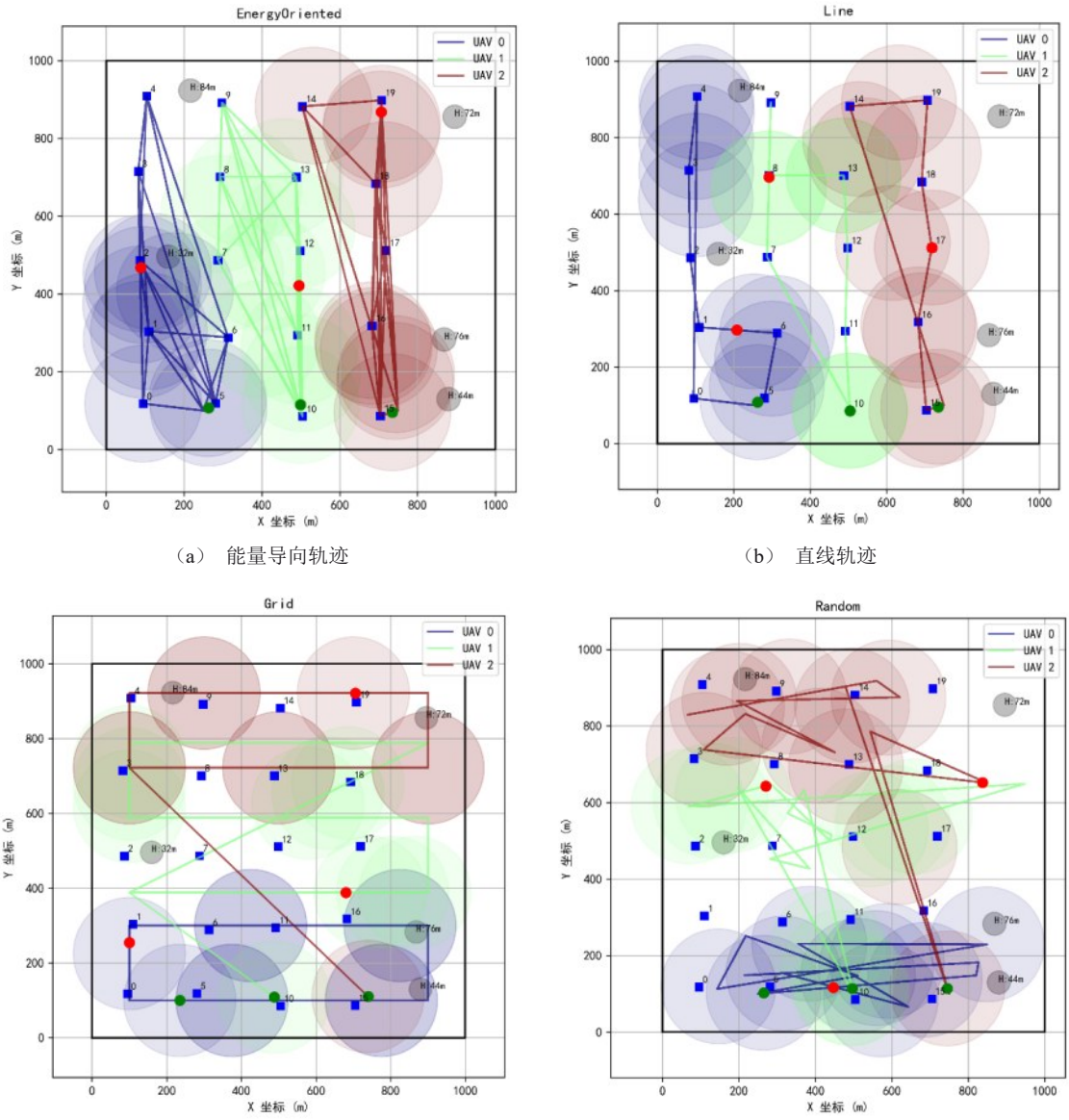
为了显示不同轨迹的区别，我们绘制了每一个轨迹的2D视图。图3展示了四种算法的典型轨迹模式。蓝色方块表示IoT设备，圆圈表示通信覆盖范围，彩色线条表示不同UAV的运动轨迹。能量导向轨迹算法(a)通过动态调整UAV位置，在保证设备覆盖的同时避免了不必要的移动，从而实现能源效率的优化。相比之下，直线轨迹(b)虽然路径简单但覆盖效率较低，网格轨迹(c)提供了系统化的区域覆盖但能耗较高，随机轨迹(d)则存在重复覆盖和能源浪费的问题。通过轨迹可视化可以清晰地看到，能量导向算法能够根据设备分布智能规划路径，在通信覆盖和能源消耗之间取得更好的平衡。

5.3　DDIM采样加速与决策质量分析

为提高系统实时性能，本文对比分析了不同扩散采样方法对决策效率和质量的影响。表1展示了标准扩散(DDPM)与去噪扩散隐式模型(DDIM)在不同采样步数下的性能对比。

实验结果表明，DDIM采样技术能显著提高决策效率。采用15步DDIM可将决策时间从324.8ms降至97.5ms，提速接近70%，同时保持98%的决策质量。步数进一步减少至10步时，决策时间降至67.7ms，但质量下降至92%。系统中我们选择15步DDIM作为最佳平衡点，在保持决策质量的同时实现了显著的计算加速。

如图4所示，该图从理论预测与实际观察两个维度，对比了DDPM与DDIM两种采样方法在不同环境变化率下的归一化性能。理论预测（上图）表明，随着环境变化率的提升（从0.025增至0.200），两种模型的预期性能均呈下降趋势，但DDIM的性能衰减曲线显著平缓于DDPM。当环境变化率超过0.100时，DDPM的理论性能已趋近于零，而DDIM仍能维持约0.6的稳定性能表现。实际观察（下图）的实验结果与理论预测高度吻合。在实际部署中，DDIM（红色虚线）同样展现出优异的适



（a） 能量导向轨迹　　（b） 直线轨迹

（c） 网格轨迹　　（d） 随机轨迹

图3　不同轨迹规划算法的UAV运动轨迹示例

表1　DDPM与DDIM在DIFFRL中的性能对比

采样方法	采样步数	决策时间(ms)	决策质量	加速比例(%)
标准扩散(DDPM)	50	324.8	1.00	-
DDIM	20	129.1	0.98	60.2
DDIM	15	97.5	0.98	69.9
DDIM	10	67.7	0.92	79.1
DDIM	5	34.9	0.85	89.3

应性，其性能虽随环境动态性增强而略有降低，但整体幅度远小于DDPM（蓝色虚线）。这表明，DDIM通过其确定的逆扩散过程，有效降低了对历史采样路径的依赖，从而在面对高度动态、非平稳的UAV任务卸载场景时，能产生更稳定、更可靠的决策。本项仿真结果有力地支撑了本文的核心论点：与传统DDPM相比，DDIM采样技术凭借其更强的鲁棒性和更高的采样效率，更适用于动态变化的无线网络环境。这为本文在第4节中选择基于DiffRL框架解决UAV轨迹优化与资源分配问题提供了坚实的理论与实验依据。

5.4　算法收敛性分析

为了验证各算法的学习效率和收敛特性，本文对DiffRL、DDPG、PPO、自适应（Adaptive）算

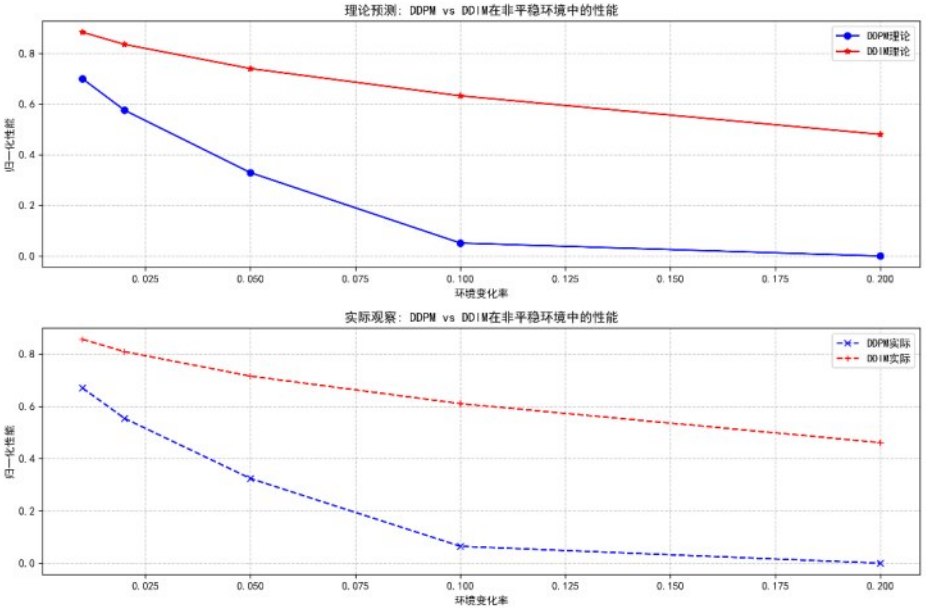


图4　DDPM与DDIM性能随环境变化率的演变趋势

法和DQN五种算法进行了收敛性分析。图5展示了不同强化学习算法的奖励值收敛过程。DiffRL与其他成熟算法（DDPG、DQN、PPO）达到了相似的性能水平，这一结果验证了扩散模型在强化学习中的有效性。特别是，DiffRL展现出稳定的收敛特性和较小的性能波动，同时在计算效率方面取得了显著提升。这些特点使得DiffRL特别适合于实际的UAV任务分配场景。DiffRL的收敛优势主要源于其结合了扩散模型的生成能力和强化学习的策略优化能力。扩散模型能够生成多样化的动作候选，覆盖更广泛的动作空间，减少了策略陷入局部最优的风险；而强化学习优化框架则指导这些候选动作向更高回报的方向优化。这种结合使得DiffRL在UAV任务卸载这类需要精细平衡能耗和延迟的复杂决策问题中表现出色。

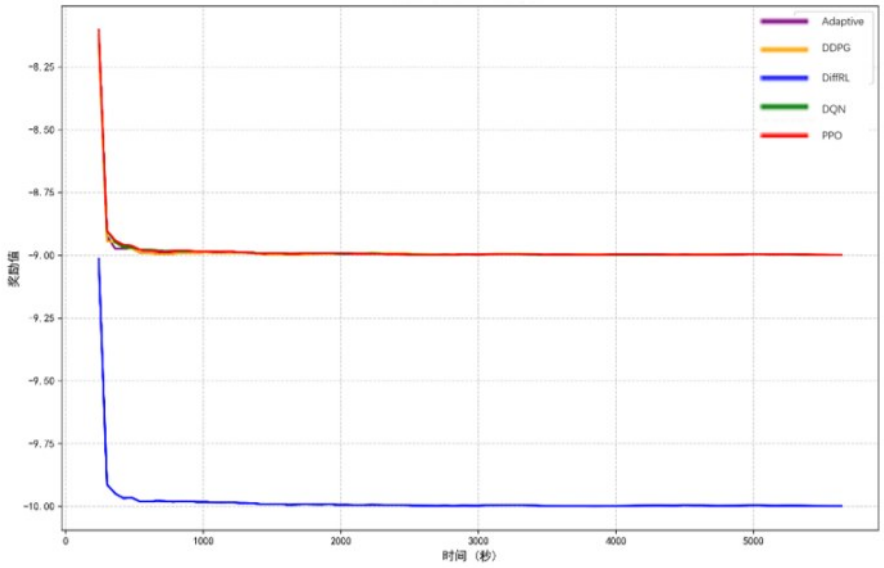


图5　不同算法的收敛性对比

5.5　任务卸载决策和性能分析

DiffRL在UAV任务卸载决策中的应用主要体现在三个方面：卸载目标选择、任务优先级排序和资源分配优化。如图6所示，测试场景为总任务量

500MI，紧急任务占比15%，卫星链路带宽20Mbps。其中Adaptive算法采用了极端的本地偏好策略，将80%的任务分配给本地执行，20%分配给卫星执行，这种保守策略减少了传输开销，但增加了本地设备负担。DDPG算法则表现出明显的卫星偏好，将60%的任务卸载到卫星执行，30%在本地执行，10%由UAV执行，这种过度依赖卫星的策略导致了较高的传输延迟。DiffRL本地执行约50%，卫星执行约45%，表现出对本地和卫星资源的均衡使用，能够智能平衡不同执行目标，在轻负载时倾向本地执行，在重负载时更多利用卫星资源，实现高效的任务分配。这些结果反映了不同算法在任务分配决策上的特点和权衡，特别是在考虑能源效率、资源利用和系统性能等多个目标时的表现差异。

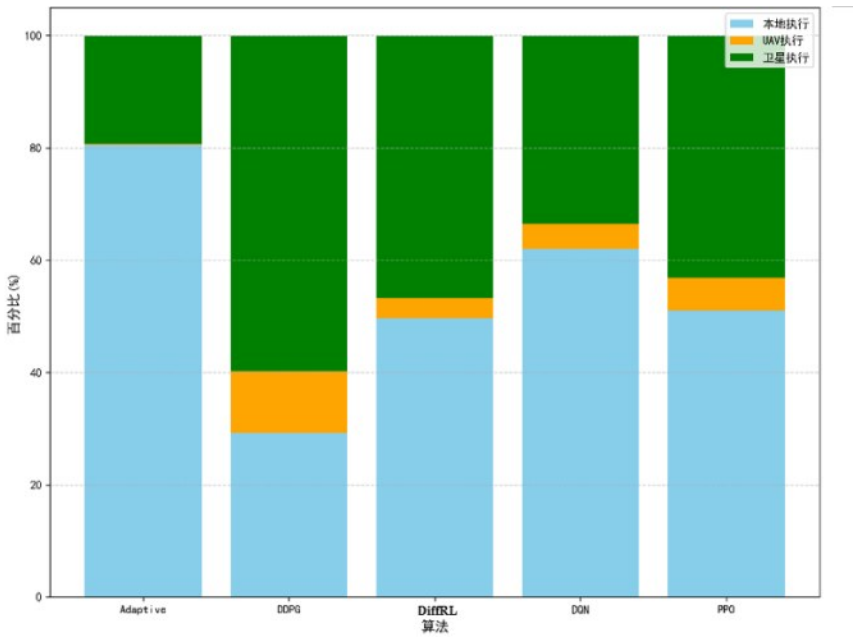


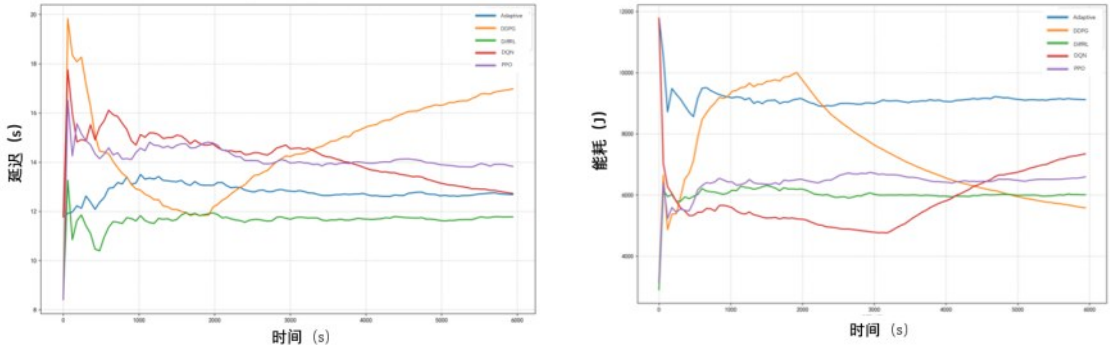
图6　DiffRL与其他算法在不同场景下的任务分配决策对比

为了研究DiffRL算法和其他算法在时间延迟和能耗上的区别，图7展示了不同算法在UAV通信网络任务卸载系统中的性能对比，如图7(a)所示，基于DiffRL的任务卸载算法显著降低了平均任务延迟。具体而言，相比DDPG算法，DiffRL通过提前预测高优先级任务的卸载决策，减少卫星链路的长距离传输延迟，将延迟降低了约30%。图7(b)进一步显示，DiffRL在系统能效方面也保持更稳定且更高的水平。DiffRL算法在整个实验周期内保持了较高且稳定的能效水平，尤其在系统稳定后的阶段，比其他算法平均高出约20-30%的能效。DiffRL算法在长期运行中表现出优秀的稳定性，避免了后期出现的性能衰退问题。

旨在系统分析DiffRL算法在归一化能耗与延迟指标上的相对优势，图8通过柱状图直观对比了包括DiffRL在内的五种算法在归一化能耗与归一化延迟两大关键性能指标上的表现，实验结果表明，DiffRL算法在归一化能耗指标上展现出了显著的相对优势。DDPG在动态信道环境下对动作噪声敏感，导致策略波动较大，但其整体趋势仍符合预期；其他算法（如DiffRL）因扩散过程的稳定性而表现更平滑，进一步验证了本文方法的鲁棒性。相较于Adaptive、DDPG、DQN和PPO等对比算法，DiffRL在实现超低能耗的同时，其归一化延迟亦维持在具有竞争力的水平。这表明DiffRL的决策机制能够更全面地评估和平衡不同执行目标的优缺点，从而做出更优的卸载决策。这种兼顾低能耗与可接受延迟的特点，凸显了DiffRL算法在资源受限或对能效要求严苛场景下的综合性能优势。

6　结论

本文提出了一种基于DiffRL的UAV通信网络与任务卸载系统，通过大量实验验证了该方法的有效性和优越性。实验结果表明：相比传统的



（a） 延迟变化　　（b） 能耗变化

图7　不同卸载决策算法性能对比

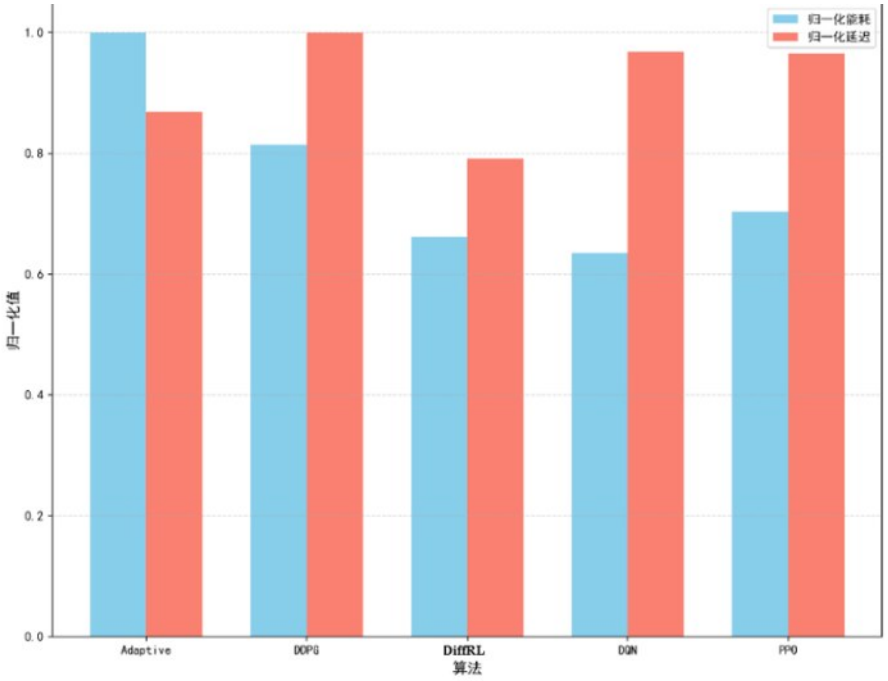


图8　各算法归一化延迟与能耗对比

DDPG、DQN和PPO算法，本系统在决策质量和执行效率上都展现出明显优势；提出的能量导向方案显著降低了系统总能耗，相比其他轨迹方案，节省了约15.3%的能源消耗；从时间序列分析来看，本系统在长期运行中表现出优秀的稳定性。

这些实验结果充分证明了本文提出的基于DiffRL的方法在UAV通信网络中的实用价值。该系统不仅显著提升了任务处理效率，还实现了更好的能源利用，为解决UAV通信网络中的能源受限和任务调度优化问题提供了一个有效的解决方案。当前方法在UAV密集部署场景（如>10架UAV）中，多节点协调的通信开销可能导致决策延迟增加10-15%，需进一步引入联邦学习降低通信成本。未来工作将计划引入分层强化学习架构，上层策略负责任务卸载决策（奖励间隔10s），下层策略优化UAV实时避障（奖励间隔100ms），解决稀疏奖励下的训练低效问题。

7 参考文献：

[1] Y. Zeng and R. Zhang, “Energy-efficient UAV communication with trajectory optimization,” IEEE Transactions on Wireless Communications, vol. 16, no. 6, pp. 3747-3760, Jun. 2017.

[2] F. Zhou, R. Q. Hu, and Y. Qian, “Computation rate maximization in UAV-enabled wireless-powered mobile-edge computing systems,” IEEE Journal on Selected Areas in Communications, vol. 36, no. 9, pp. 1927-1941, Sep. 2018.

[3] V. Mnih, K. Kavukcuoglu, D. Silver, et al., “Human-level control

through deep reinforcement learning,” Nature, vol. 518, no. 7540, pp. 529-533, Feb. 2015.

[4] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” arXiv preprint arXiv: 1707.06347, 2017.

[5] M. Janner, Y. Du, J. T. Tenenbaum, and S. Levine, “Planning with diffusion for flexible behavior synthesis,” in Proceedings of the 37th International Conference on Machine Learning (ICML), Baltimore, Maryland, USA: PMLR, 2022, pp. 9782 – 9793.

[6] H. Liang, Z. Yang, G. Zhang et al., “Resource allocation for space-air-ground integrated networks: A comprehensive review,” Journal of Communications and Information Networks, vol. 9, no. 1, pp. 1 – 23, 2024.

[7] Hansen-Estruch P., Zhang A. K., Vuong Q., et al. "Diffusion policies as an expressive policy class for offline reinforcement learning, " International Conference on Learning Representations (ICLR), 2023.

[8] H. Du, Z. Li, D. Niyato, J. Kang, Z. Xiong, H. Huang, and S. Mao, “Diffusion-based reinforcement learning for edge-enabled AI-generated content services,” IEEE Trans. Mobile Comput., to appear, 2024.

[9] J. Song, C. Meng, and S. Ermon, “Denoising diffusion implicit models,” in Proceedings of the 9th International Conference on Learning Representations (ICLR), Vienna, Austria (Online), 2021.

[10] Q. Wu, Y. Zeng, and R. Zhang, “Joint trajectory and communication design for multi-UAV enabled wireless networks,” IEEE Transactions on Wireless Communications, vol. 17, no. 3, pp. 2109-2121, Mar. 2018.

[11] L. Liu, S. Zhang, and R. Zhang, “Multi-beam UAV communication in cellular uplink: Cooperative interference cancellation and sum-rate maximization,” IEEE Transactions on Wireless Communications, vol. 18, no. 10, pp. 4679-4694, Oct. 2019.

[12] S. Zhang and R. Zhang, “Radio map-based 3D path planning for cellular-connected UAV,” IEEE Transactions on Wireless Communications, vol. 20, no. 3, pp. 1975-1989, Mar. 2021.

[13] X. Liu, Y. Liu, and Y. Chen, “Reinforcement learning in multiple-UAV networks: Deployment and movement design,” IEEE Transactions on Vehicular Technology, vol. 68, no. 8, pp. 8036-8049, Aug. 2019.

[14] N. Khan, A. Ahmad, A. Wakeel, Z. Kaleem, B. Rashid, and W. Khalid, ‘‘Efficient UAVs deployment and resource allocation in UAV-relay assisted public safety networks for video transmission,’’ IEEE Access, vol. 12, pp. 4561 – 4574, 2024.

[15] Y. Zheng and J. Chen, ‘‘Geography-aware optimal UAV 3D placement for LOS relaying: A geometry approach,’’ IEEE Trans. Wireless Commun., vol. 23, no. 8, pp. 9301 – 9314, Aug. 2024.

[16] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, and C. S. Hong, “Caching in the sky: Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience,” IEEE Journal on Selected Areas in Communications, vol. 35, no. 5, pp. 1046-1061, May 2017.

[17] C. H. Liu, Z. Chen, and Y. Zhan, “Energy-efficient UAV control for effective and fair communication coverage: A deep reinforcement learning approach,” IEEE Journal on Selected Areas in Communications, vol. 36, no. 9, pp. 2059-2070, Sep. 2018.

[18] T. Haarnoja, A. Zhou, and P. Abbeel, “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” in Proceedings of the 35th International Conference on Machine Learning (ICML), Stockholm, Sweden: PMLR, 2018, pp. 1861 – 1870.

[19] S. S. Sefati et al., “A comprehensive survey on resource management in 6G network based on Internet of Things,” IEEE Access, vol. 12, pp. 113741 – 113784, 2024.

[20] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep Unsupervised Learning using Nonequilibrium Thermodynamics," in Proceedings of the 32nd International Conference on Machine Learning (ICML), Lille, France, 2015, pp. 2256-2265.

[21] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” in Advances in Neural Information Processing Systems, vol. 33, pp. 6840-6851, 2020.

[22] P. Wang, Y. Zhou, and K. Xu, “Diffusion policies as an expressive policy class for offffine reinforcement learning,” in International Conference on Learning Representations, 2022.

[23] H. Du, R. Zhang, Y. Liu, J. Wang, Y. Lin, Z. Li, D. Niyato, J. Kang, Z. Xiong, S. Cui et al., “Enhancing deep reinforcement learning: A tutorial on generative diffusion models in network optimization,” IEEE Communications Surveys & Tutorials, 2024.

[作者简介]

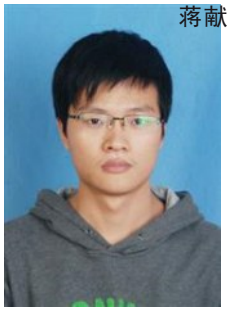
张子天（1988— ），男，博士，浙江工商大学信息与电子工程学院副研究员，主要研究方向为基于机器学习的网络流量预测与资源管理。

葛天豪（2000— ），男，浙江工商大学信息与电子工程学院硕士生，主要研究方向为空天地网络。

诸葛斌（1976— ），男，博士，浙江工商大学信息与电子工程学院教授，主要研究方向为网络和通信技术、互联网技术和网络安全。

郑运强（1998- ），男，浙江工商大学信息与电子工程学院硕士生，主要研究方向为基于强化学习的网络资源管理。

董黎刚（1972— ），男，博士，浙江工商大学信息与电子工程学院教授，主要研究方向为智能网络、在线教育。



蒋献（1988— ），男，浙江工商大学信息与电子工程学院讲师、实验员，主要研究方向为在线教育。