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

张 涛[①②] 张 迁*[①] 朱颖雯[①] 代 陈[③]

[①](江苏开放大学信息工程学院 南京 210000)

[②](南京航空航天大学计算机科学与技术学院 南京 210016)

[③](南京邮电大学计算机学院 南京 210023)

摘 要：为了应对智能反射面(RIS)辅助的无人机(UAV)在物联网数据收集过程中能量高效利用与信息收集时效性之间的均衡问题，该文提出一种基于深度强化学习的数据收集优化策略。针对无人机在数据采集过程中的飞行能耗、通信复杂性及采集信息时效性(AoI)约束，设计了一种基于双深度Q网络(DDQN)的联合优化方案，涵盖无人机轨迹规划、物联网设备调度以及智能反射面相位调整。该方案有效缓解了传统Q学习方法中Q值过估计的问题，使无人机能够根据实时环境动态调整飞行轨迹和通信策略，从而在提升数据传输效率的同时降低能量消耗。仿真结果表明，与传统方法相比，所提方案能够显著提高数据收集效率。此外，通过合理分配能量与通信资源，所提方案能够动态适应不同通信环境参数变化，确保系统在能耗与AoI之间达到最佳均衡。

关键词：无人机辅助通信；时效性；深度强化学习；智能反射面

中图分类号：TN929.5 **文献标识码**：A **文章编号**：1009-5896(2025)02-0427-12

DOI: 10.11999/JEIT240866

1 引言

物联网(Internet of Things, IoT)设备凭借其广泛部署和持续感知优势，在智能城市、工业控制和智能交通系统中得到了广泛应用，产生了大量对实时管理至关重要的数据[1]。然而，由于IoT设备能量和通信能力受限，如何高效、及时地收集这些数据仍然是一个重要挑战。无人机(Unmanned Aerial Vehicle, UAV)因其灵活的部署和移动能力，可有效执行数据的收集任务，通过合理规划无人机的飞行轨迹，能够与地面物联网设备建立视距通信链路，从而提升数据收集的效率。然而，在城市等复杂环境中，建筑物的遮挡使得UAV的视距通信受到限制，难以实现全面覆盖[2]。智能反射面(Reconfigurable Intelligent Surface, RIS) 可以安装在城市建筑物表面，通过调整反射信号的相位，显著改善无线通信质量，为应对这一挑战提供了新的解决方案[3]。RIS可以通过动态调节信号反射方向，优化通信信号的覆盖范围和质量，这不仅可以增强无人机在复杂环境中的适应性，使UAV能够在不同的飞行高度和角度下实现稳定的通信连接，而且可以有效缓解遮挡物或信号干扰导致的通信中断问题，提高数据采集效率和可靠性。针对此问题，已有研究致力于探索RIS辅助下的UAV数据采集技术[4]。文献[5]针对多UAV路径规划和复杂地形通信问题，提出了一种联合轨迹规划和传输调度算法，并采用模型预测控制方法，优化多UAV轨迹和能量消耗，以完成边缘计算任务。文献[6]解决了IoT中UAV队列在非平坦地形场景中考虑碰撞规避的3维导航和数据收集问题。该方案通过模型预测控制方法优化多UAV轨迹和数据传输，同时使用RIS来提高通信效率。文献[7]针对工业IoT中RIS辅助UAV通信安全问题，提出联合优化RIS被动波束成形、用户关联、UAV轨迹和发射功率的方案，使用了块坐标下降法和逐次凸逼近法来解决非凸优化问题。文献[8]研究了在IoT中使用UAV和RIS辅助的无线携能通信系统。该研究采用分层Stackelberg博弈模型来优化UAV的轨迹、资源分配和定价机制，增强能量传输和信息传输的效率，目标在于最大化IoT设备的能效和公平性。文献[9]通过结合UAV和RIS，优化波束成形、资源分配和UAV位置，以提高短数据包的传输可靠性，数值结果表明其在超高可靠性通信中的有效性。文献[10]针对UAV搭载RIS进行地面智能车辆的视距无线通信问题，提出基于模型预测控制的3维导航算法，并利用动态规划和逐次凸逼近方法优化UAV的轨迹和通信性能。文献[11]针对UAV辅助RIS在移动边缘计算场景下的轨迹规划、相位偏移设计以及IoT设备关联问题，提出了一种基于深度确定性策略梯度的优化

收稿日期：2024-10-14；改回日期：2025-01-07；网络出版：2025-01-11

*通信作者： 张迁 zhangqian@jsou.edu.cn

基金项目：国家自然科学基金(62402232)，江苏省高等学校自然科学研究项目(23KJB520024)

Foundation Items: The National Natural Science Foundation of China (62402232), The Natural Science Foundation of the Jiangsu Higher Education Institutions of China (23KJB520024)

算法。然而，上述算法没有考虑UAV在数据采集过程的能量消耗问题。

为此，一些学者对RIS辅助UAV数据采集场景下的能耗问题展开研究。文献[12]针对IoT中UAV的数据收集能效问题，提出了利用深度强化学习(Deep Reinforcement Learning, DRL)和块坐标下降算法的联合优化方案，以提升UAV数据采集能效。文献[13]研究了IoT中UAV辅助RIS进行数据收集的能效优化问题，通过构建基于RIS级联特性的空对地信道模型，并使用凸优化技术和1阶泰勒展开，优化了UAV轨迹和传感器节点的唤醒调度。文献[14]研究了多RIS元素辅助的能效数据收集问题，其中考虑了一个高度固定且飞行时间受限的UAV作为收集器，建立了一个复杂空对地信道模型，考虑了级联RIS信道和UAV的时变轨迹，目标在于最小化IoT设备的最大能量消耗。文献[15]针对UAV在物联网数据收集过程中面临的干扰问题，提出了一种结合RIS和UAV的调度优化方法。通过引入蚁群算法，该方法对UAV的飞行轨迹和IoT设备的调度进行了联合优化，不仅有效地增强了数据传输的稳定性，还大幅减少了数据收集过程中的延迟，提升了整体的运行效率。然而，上述方法并没有考虑UAV数据采集过程中的时效问题。

对于UAV任务，尤其是数据采集过程中，保持较低的AoI至关重要。低AoI意味着信息是最新的，能够确保控制系统或决策算法基于最新数据做出响应，从而提高任务执行效率和准确性。为此，一些研究考虑了RIS辅助下UAV数据收集AoI问题。文献[16]提出了通过RIS辅助UAV通信系统来优化IoT中数据收集的能效问题，提出基于AoI的深度Q网络算法，通过解耦UAV轨迹和RIS相位的优化问题，减少UAV的飞行距离和能量消耗，提高数据收集效率并保证系统的AoI要求。文献[17]探讨了通过结合UAV和RIS来优化IoT中AoI的问题，采用了近端策略优化算法解决了UAV高度、通信调度和RIS相位调整的联合优化问题。数值结果表明，该方案显著减少了AoI，并提升了系统通信性能。文献[18]提出了一种基于UAV和RIS的联合轨迹规划和资源优化方法，以降低AoI，提高数据收集效率。文献[19]通过结合多UAV和RIS技术，提出一种优化IoT数据收集AoI的方法，利用DRL来动态规划UAV轨迹并调整RIS的相位，以实现较低的AoI和较高的能量效率。文献[20] 通过联合优化UAV位置、RIS相位偏移、设备传输功率和数据上传时间，以减少IoT中数据传输的平均AoI，提升数据传输的实时性和可靠性。

与现有研究只考虑能耗或者数据AoI不同，本文联合考虑了3维UAV移动轨迹、UAV移动能量、数据采集能量和AoI优化问题。从上述研究可知，现有的UAV辅助数据收集方法多使用确定性优化算法，通常仅寻求次优解，且复杂度较高，难以适应UAV运动轨迹和通信环境动态变化的复杂场景。此外，现有基于DRL的方案大多采用基于Q学习或者动态规划的数据收集方法，导致算法复杂度高和Q值过估计问题。本文针对RIS辅助UAV数据采集场景，构建了一个综合考虑UAV移动能量消耗、数据采集能量、IoT设备调度和RIS相位偏移的时效数据收集优化模型，并将该问题建模为马尔可夫决策过程。针对现有方案存在的不足，提出基于双深度Q网络 (Double Deep Q Network, DDQN)的数据采集方案，优化了UAV的3维轨迹规划和IoT设备调度，从而在保证数据AoI需求的同时最小化UAV数据采集能量消耗。

2 系统模型

本文构建的RIS辅助UAV数据收集系统模型如图1所示，其中，地面IoT设备采集环境信息，并将数据及时上传到信息收集中心。假设由于IoT设备能量受限，无法直接将数据发送给数据收集中心，因此派遣一架UAV完成数据收集任务并在某栋建筑上安装RIS辅助建立视距传播(Line of Sight, LoS)通信链路。这里，本文只考虑单天线UAV，而一些多天线技术(如多输入多输出(Multi-Input Multi-Output, MIMO))能够通过多根天线实现同时传输和接收多个信号流，增加了信号的空间分集[21,22]，从而提升了数据传输速率和可靠性，在未来研究中考虑结合MIMO技术以进一步优化UAV数据收集的效率和鲁棒性。假设有N个IoT设备，RIS由一组均匀矩形阵列(Uniform Rectangular Array, URA)组成，包含$M = M_{\mathrm{r}} \times M_{\mathrm{c}}$个被动反射单元(Passive Reflecting Units, PRUs)，M_{r}和M_{c}分别代表PRU的行数和列数。假设每个PRU可以使用

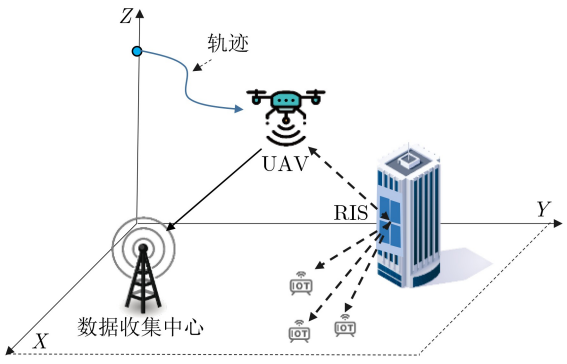


图 1 RIS辅助UAV数据收集模型

一个反射系数$r_{m_{\mathrm{c}},m_{\mathrm{r}},t}=\omega \mathrm{e}^{\mathrm{j}\theta_{m_{\mathrm{c}},m_{\mathrm{r}},t}}$控制相位偏移，其中$\omega\in[0,1]$表示反射损失，$\theta_{m_{\mathrm{c}},m_{\mathrm{r}},t}$表示第$m_{\mathrm{r}}$行第$m_{\mathrm{c}}$列反射元素在时隙$t$的相位偏移。为了便于分析，本文将整个数据采集过程分为T个时隙，并假设每个时隙占用很小的时间段，信道传输特性在该时间段近似不变。相应地，IoT设备集合、反射阵列集合和时隙集合分别表示为$\mathcal{N}=\{1,2,\cdots,N\}$，$\mathcal{M}=\{1,2,\cdots,M\}$和$\mathcal{T}=\{1,2,\cdots,T\}$。

2.1 移动模型

本文将UAV移动空间均匀划分为3维正方形网格，网格边长为L，UAV 可以在相邻正方体中心间移动，T个时隙内的轨迹可以表示为$\{\boldsymbol{q}_1^{\mathrm{U}},\boldsymbol{q}_2^{\mathrm{U}},\cdots,\boldsymbol{q}_T^{\mathrm{U}}\}$，其中$\boldsymbol{q}_t^{\mathrm{U}}=(x_t^{\mathrm{U}},y_t^{\mathrm{U}},z_t^{\mathrm{U}})$表示UAV在时隙$t$时的位置，$\boldsymbol{L}_t^{\mathrm{U}}=(l_t^X,l_t^Y,l_t^Z)$表示UAV在时隙$t$的移动向量，$l_t^X$, l_t^Y和l_t^Z分别表示移动的网格数目。为了避免UAV移动出给定区域，给出限制为

$$0\le x_t^{\mathrm{U}}\le x^{\mathrm{MAX}} \tag{1}$$

$$0\le y_t^{\mathrm{U}}\le y^{\mathrm{MAX}} \tag{2}$$

$$z^{\mathrm{MIN}}\le z_t^{\mathrm{U}}\le z^{\mathrm{MAX}} \tag{3}$$

2.2 传输模型

本文假设UAV通过直接链路和间接链路收集地面IoT设备的数据。对于直接链路，本文采用概率路损模型来描述UAV和IoT设备的通信路损[23]。值得注意的是，在城市场景中，RIS和无人机的通信链路可能受到建筑物遮挡的显著影响，RIS布置受限于建筑位置和可视范围，而无线信道的不稳定性则可能导致信号衰减和反射路径不稳定。为应对这些挑战，提出的算法通过动态调整无人机轨迹和RIS相位优化策略，使系统能够灵活适应环境变化，从而在建筑物密集和信道动态变化的场景中维持较高的通信效率和数据收集时效性。根据概率路损模型，该UAV与IoT设备n之间的LoS概率可以表示为

$$\xi_{n,t}^{\mathrm{LoS}}=\frac{1}{1+c_1\mathrm{e}^{(-c_2(\vartheta_{n,t}-c_1))}} \tag{4}$$

其中，c_1和c_2代表2个固定的环境参数。$\vartheta_{n,t}$表示UAV与IoT设备n的传播俯仰角，即

$$\vartheta_{n,t}=\frac{180}{\pi}\arctan\left(\frac{z_t^{\mathrm{U}}}{d_{n,t}^{\mathrm{UI}}}\right) \tag{5}$$

此外，该UAV与IoT设备n之间的非视距传播(Non-Line of Sight, NLoS)概率表示为$\xi_{n,t}^{\mathrm{NLoS}}=1-\xi_{n,t}^{\mathrm{LoS}}$。根据该路损模型[24]，UAV与IoT设备$n$在时隙$t$的平均信道功率增益可以表示为

$$W_{n,t}^{\mathrm{UI}}=K_0^{-1}(d_{n,t}^{\mathrm{UI}})^{-2}\left[\xi_{n,t}^{\mathrm{LoS}}\mu^{\mathrm{LoS}}+\xi_{n,t}^{\mathrm{NLoS}}\mu^{\mathrm{NLoS}}\right]^{-1} \tag{6}$$

其中，$K_0=(4\pi f_{\mathrm{c}}/\mathrm{C})^2$，$f_{\mathrm{c}}$代表载波频率。C代表光速。$\mu^{\mathrm{LoS}}$和$\mu^{\mathrm{NLoS}}$分别代表视距和非视距传播下的衰减因子。

对于非直接链路，UAV与RIS以及RIS与IoT设备间的通信信道可以用莱斯衰落来描述[25,26]。因此，对于UAV与RIS m之间的信道增益可以表示为

$$\boldsymbol{W}_{m,t}^{\mathrm{UR}}=\sqrt{\rho\left(d_{m,t}^{\mathrm{UR}}\right)^{-\alpha}}\sqrt{\frac{K_1}{1+K_1}}\bar{\boldsymbol{W}}_{m,t}^{\mathrm{UR}} \tag{7}$$

其中，$d_{m,t}^{\mathrm{UR}}$代表UAV到RIS m的距离，ρ代表平均功率损失增益，K_1代表莱斯因子。$\bar{\boldsymbol{W}}_{m,t}^{\mathrm{UR}}$代表LoS成分，即

$$\begin{aligned}\bar{\boldsymbol{W}}_{m,t}^{\mathrm{UR}}&=\left[1,\mathrm{e}^{-\frac{2\pi\mathrm{j}}{\lambda}d_{\mathrm{r}}\phi_{m,t}^{\mathrm{UR}}\varphi_{m,t}^{\mathrm{UR}}},\cdots,\mathrm{e}^{-\frac{2\pi\mathrm{j}}{\lambda}(M_{\mathrm{r}}-1)d_{\mathrm{r}}\phi_{m,t}^{\mathrm{UR}}\varphi_{m,t}^{\mathrm{UR}}}\right]^{\mathrm{T}}\\&\otimes\left[1,\mathrm{e}^{-\frac{2\pi\mathrm{j}}{\lambda}d_{\mathrm{c}}\psi_{m,t}^{\mathrm{UR}}\varphi_{m,t}^{\mathrm{UR}}},\cdots,\mathrm{e}^{-\frac{2\pi\mathrm{j}}{\lambda}(M_{\mathrm{c}}-1)d_{\mathrm{c}}\psi_{m,t}^{\mathrm{UR}}\varphi_{m,t}^{\mathrm{UR}}}\right]^{\mathrm{T}}\end{aligned} \tag{8}$$

其中，λ代表载波波长，$\phi_{m,t}^{\mathrm{RU}}=(x_t^{\mathrm{U}}-x_{m,t}^{\mathrm{R}})/d_{m,t}^{\mathrm{RU}}$，$\psi_{m,t}^{\mathrm{RU}}=(y_t^{\mathrm{U}}-y_{m,t}^{\mathrm{R}})/d_{m,t}^{\mathrm{RU}}$，$\varphi_{m,t}^{\mathrm{RU}}=(z_t^{\mathrm{U}}-z_{m,t}^{\mathrm{R}})/d_{m,t}^{\mathrm{RU}}$。$(x_{m,t}^{\mathrm{R}},y_{m,t}^{\mathrm{R}},z_{m,t}^{\mathrm{R}})$代表RIS$m$的坐标。

类似地，RIS与IoT设备n间的信道增益和LoS成分表示为

$$\boldsymbol{W}_{n,t}^{\mathrm{RI}}=\sqrt{\rho\left(d_{n,t}^{\mathrm{RI}}\right)^{-\alpha}}\sqrt{\frac{K_2}{1+K_2}}\bar{\boldsymbol{W}}_{n,t}^{\mathrm{RI}} \tag{9}$$

$$\begin{aligned}\bar{\boldsymbol{W}}_{n,t}^{\mathrm{RI}}&=\left[1,\mathrm{e}^{-\frac{2\pi j}{\lambda}d_{\mathrm{r}}\phi_{n,t}^{\mathrm{RI}}\varphi_{n,t}^{\mathrm{RI}}},\cdots,\mathrm{e}^{-\frac{2\pi j}{\lambda}(M_{\mathrm{r}}-1)d_{\mathrm{r}}\phi_{n,t}^{\mathrm{RI}}\varphi_{n,t}^{\mathrm{RI}}}\right]^{\mathrm{T}}\\&\otimes\left[1,\mathrm{e}^{-\frac{2\pi\mathrm{j}}{\lambda}d_{\mathrm{c}}\psi_{n,t}^{\mathrm{RI}}\varphi_{n,t}^{\mathrm{RI}}},\cdots,\mathrm{e}^{-\frac{2\pi\mathrm{j}}{\lambda}(M_{\mathrm{c}}-1)d_{\mathrm{c}}\psi_{n,t}^{\mathrm{RI}}\varphi_{n,t}^{\mathrm{RI}}}\right]^{\mathrm{T}}\end{aligned} \tag{10}$$

其中，$d_{n,t}^{\mathrm{RI}}$代表RIS与IoT设备n间的距离，K_2代表莱斯因子。$\phi_{n,t}^{\mathrm{RI}}=(x_{n,t}^{\mathrm{I}}-x_{m,t}^{\mathrm{R}})/d_{m,n,t}^{\mathrm{RI}}$，$\psi_{n,t}^{\mathrm{RI}}=(y_{n,t}^{\mathrm{I}}-y_{m,t}^{\mathrm{R}})/d_{m,n,t}^{\mathrm{RI}}$，$\varphi_{n,t}^{\mathrm{RI}}=(z_{n,t}^{\mathrm{I}}-z_{m,t}^{\mathrm{R}})/d_{n,t}^{\mathrm{RI}}$。$(x_{n,t}^{\mathrm{I}},y_{n,t}^{\mathrm{I}},z_{n,t}^{\mathrm{I}})$代表IoT设备$n$的坐标。

这时，UAV经过RIS到 IoT设备n这条间接链路获得数据率可以表示为

$$\begin{aligned}R_{n,t}^{\mathrm{URI}}&=B_1\eta_{n,t}\\&\cdot\log_2\left(1+\frac{P_1|\boldsymbol{W}_{n,t}^{\mathrm{UI}}+\left(\boldsymbol{W}_{m,t}^{\mathrm{RI}}\right)^{\mathrm{T}}\boldsymbol{Q}_{m,t}\boldsymbol{W}_{n,t}^{\mathrm{UR}}|^2}{B_1\sigma_1^2}\right)\end{aligned} \tag{11}$$

其中，P_1和B_1分别表示UAV数据收集所用功率和带宽。σ_1表示噪声功率。$\boldsymbol{Q}_{m,t}=\mathrm{diag}\left(\mathrm{e}^{\mathrm{j}\theta_{1,1,t}},\cdots,\mathrm{e}^{\mathrm{j}\theta_{m_{\mathrm{r}},m_{\mathrm{c}},t}},\cdots,\mathrm{e}^{\mathrm{j}\theta_{M_{\mathrm{r}},M_{\mathrm{c}},t}}\right)\in\mathbb{C}^{M_{\mathrm{r}}M_{\mathrm{c}}\times M_{\mathrm{r}}M_{\mathrm{c}}}$表示相位偏移对角矩阵。$\eta_{n,t}\in\{0,1\}$表示IoT设备在时隙$t$是否被调度。

除了数据收集，UAV还需要将收集到的数据上传到一个数据收集中心，该链路的信道增益可以表示为[27]

$$W_t^{\mathrm{UP}} = \sqrt{\beta (d_t^{\mathrm{UP}})^{-\alpha_0}} g_0 \tag{12}$$

其中，β表示在参考距离1 m范围内的路径损失。α_0代表相应的路损因子。d_t^{UP}代表时隙t时UAV到数据收集中心的距离。$g_0 \sim \mathcal{CN}(0,1)$表示归一化小尺度衰落。 同样，该上传过程的数据速率可以表示为

$$R_t^{\mathrm{UP}} = B_2 \log_2 \left(1 + \frac{P_2 |W_t^{\mathrm{UP}}|^2}{B_2 \sigma_2^2}\right) \tag{13}$$

其中，P_2和B_2分别表示UAV数据上传的功率和带宽。σ_2表示噪声功率。

2.3 能耗模型

UAV的数据传输过程中，涉及的能量消耗有UAV飞行能量，数据收集能量，以及数据上传能量。因为IoT设备收集的数据量大小会直接影响通信能量消耗，并且UAV与数据收集中心较远或者数据上传经过较为复杂地理环境(如山地、城市建筑和丛林)导致能量消耗可能会非常显著，因此相比于其他相关研究，本文考虑UAV的飞行能量、数据收集能量和上传能量。对于UAV飞行能量，本文参考UAV推进能量模型[24]，定义为

$$\begin{aligned} E_t^{\mathrm{F}} = {} & P_{\mathrm{B}} \left(1 + \frac{3\left(v_t^{\mathrm{H}}\right)^2}{U_{\mathrm{tip}}^2}\right) + \frac{1}{2} \Lambda \Upsilon \Xi G \left(v_t^{\mathrm{H}}\right)^3 \\ & + P_{\mathrm{I}} \left(\sqrt{1 + \frac{\left(v_t^{\mathrm{H}}\right)^4}{4 v_0^4}} - \frac{\left(v_t^{\mathrm{H}}\right)^2}{2 v_0^2}\right)^{\frac{1}{2}} + P_{\mathrm{V}} v_t^{\mathrm{V}} \end{aligned} \tag{14}$$

其中，P_{B}和P_{I} 分别表示UAV恒定桨叶型功率和诱导功率。P_{V}是恒定的垂直移动功率。U_{tip}表示是旋翼叶尖速度。v_0表示平均旋翼诱导速度。v_t^{H}和v_t^{V}分别表示水平和垂直移动速度。Λ和Υ分别表示机身阻力比和旋翼盘面比。Ξ和G分别表示空气密度和旋翼盘面积[28]。

此外，当时隙t调度IoT设备n，数据收集能量定义为

$$E_t^{\mathrm{I}} = P_1 \frac{\chi_{n,t}}{R_{n,t}^{\mathrm{URI}}} \tag{15}$$

其中，$\chi_{n,t}$表示IoT设备n在时隙t的数据需求。类似地，数据上传能量可以表示为

$$E_t^{\mathrm{UP}} = P_2 \frac{\sum\limits_n \eta_{n,t} \chi_{n,t}}{R_t^{\mathrm{UP}}} \tag{16}$$

2.4 AoI模型

对于UAV数据采集任务，AoI是衡量采集信息及时性的重要指标。它反映了从信息生成到信息被接收之间的时间延迟。在本文中，数据采集系统采用了单数据包队列机制，即老的状态更新数据包会被丢弃，取而代之的是最新的样本数据包[17]。因此，在每个时隙的开始，UAV会采集IoT设备最新更新的一个数据包并传给基站。在这个过程中，UAV需要完成IoT设备的选择并且要满足最小的数据传输速率，因此数据采集AoI表示为[29]

$$A_{n,t} = \begin{cases} 1 \ \ \eta_{n,t} = 1, \ R_{n,t}^{\mathrm{URI}} \geq R^{\mathrm{MIN}} \\ A_{n,t-1} + 1 \ \ \text{其他} \end{cases} \tag{17}$$

其中，R^{MIN}表示IoT设备的最小数据率需求；相应地，对所有IoT设备的平均数据采集AoI可以表示为

$$A_t = \frac{1}{N} \sum_n A_{n,t} \tag{18}$$

2.5 优化问题建模

在本文中，UAV在每个时隙内移动并调度收集地面IoT设备的数据，并将数据上传到信息收集中心。考虑到该过程中UAV的能量消耗和收集信息的新鲜度，本文优化目标系统代价函数定义为

$$\mathrm{cost}_t = \zeta_1 A_t + \zeta_2 E_t^{\mathrm{F}} + \zeta_3 E_t^{\mathrm{I}} + \zeta_4 E_t^{\mathrm{UP}} \tag{19}$$

其中，ζ_1, ζ_2, ζ_3, ζ_4分别表示权重，取值可以反映信息新鲜度和能耗之间的偏好选择。根据该代价函数的定义，本文的优化目标在于在T个时隙内找到最佳UAV运动轨迹，调度地面IoT设备，采集并且上传数据，最小化长期平均系统代价，即

$$\min_{\boldsymbol{q}_1^{\mathrm{U}}, \boldsymbol{q}_2^{\mathrm{U}}, \cdots, \boldsymbol{q}_T^{\mathrm{U}}, \eta} \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathrm{cost}_t \tag{20}$$

在UAV数据采集过程中，由于信道的时变性、UAV飞行的随机性等不确定性因素，使得难以建立精确的数学模型来求解相关的优化问题。因此，传统的凸优化等确定性优化工具难以有效应对这些挑战。为了解决这一问题，本文尝试采用基于DRL的方法，通过设计UAV移动规划和IoT设备调度策略，来实现上述UAV数据采集优化目标。

3 算法设计

在强化学习算法设计过程中，通常需要设计智能体与环境不断交互，以获得最大的长期回报。对于本文提出的问题来说，设定UAV携带的控制器为智能体，将数据采集区域以及对应的通信场景作为智能体的交互环境。该智能体仅能通过观察环境的局部状态信息，学习UAV移动和IoT设备调度策略，根据该策略执行相应的动作，并从环境中获得相应的即时奖励。上述过程不断迭代执行，直到智

能体可以获得最优策略。具体来说，对于问题式(20)，对应的强化学习要素定义如下。

3.1 状态空间设计

为了设计易于实现的强化学习算法，本文假设智能体可以获得连续的状态信息，并根据上一时隙UAV的移动信息、IoT设备的调度信息，学习有效的策略，然后转到下一状态。具体来说，智能体在时隙t观察到的环境状态可以表示为

$$\boldsymbol{s}_t=[x_{t-1}^{\mathrm{U}},y_{t-1}^{\mathrm{U}},z_{t-1}^{\mathrm{U}},l_{t-1}^{X},l_{t-1}^{Y},l_{t-1}^{Z},A_t] \tag{21}$$

其中，x_{t-1}^{U}, y_{t-1}^{U}, z_{t-1}^{U}表示UAV上一时隙的位置，l_{t-1}^{X}, l_{t-1}^{Y}, l_{t-1}^{Z}表示UAV在上一时隙在X, Y, Z轴上移动的移动信息。

3.2 动作空间设计

动作空间刻画了智能体在环境中的行为。对应于状态空间设计，本文假设智能体可以根据策略做出连续的动作。对于本文定义的UAV数据采集环境来说，控制变量包括：UAV的方向，在该方向上移动的距离以及地面IoT设备的调度，据此，智能体的动作空间可以表示为

$$\boldsymbol{a}_t=[x_t^{\mathrm{U}},y_t^{\mathrm{U}},z_t^{\mathrm{U}},l_t^{X},l_t^{Y},l_t^{Z},\eta_t] \tag{22}$$

其中，$\eta_t\in\{1,2,\cdots,N\}$表示智能体在第$t$个时隙对IoT设备的调度动作。

3.3 奖励函数设计

奖励函数衡量了智能体在环境中执行动作后产生的系统增益或者价值。本文的奖励函数包括两部分，一部分是智能体获得的系统增益，包括UAV在数据采集过程的能量消耗和收集信息的新鲜度；另一部分是对于智能体产生的不合规行为给予惩罚，包括UAV越过指定3维区域获得的惩罚。具体定义为

$$r_t=\zeta_1A_t+\zeta_2E_t^{\mathrm{F}}+\zeta_3E_t^{\mathrm{I}}+\zeta_4E_t^{\mathrm{UP}}-\Delta_t \tag{23}$$

其中，Δ_t表示UAV越过指定3维区域获得的惩罚。至此，本文的优化目标转变为最大化智能体的长期回报，即

$$G_t=\sum_{t=1}^{T}\gamma^{t-1}r_t \tag{24}$$

其中，γ表示折扣因子。

3.4 学习算法设计

如上节所述，本文设计的状态和动作空间均为离散变化值。因此，拟采用适用于离散空间的双深度Q网络(Double Deep Q-Network, DDQN)学习有效的数据采集策略。该方法可以有效解决传统DQN(Deep Q-Network)中存在的过估计问题。通过引入双重估计的方式，DDQN能够更加稳定地学习并提高策略的表现。基于DDQN的RIS辅助UAV时效数据收集策略结构如图2所示。

具体来说，对于优化目标式(24)，由于智能体从某个状态出发到达最终状态一般需要经历不同的学习路径，因此对于给定的状态和动作，需要求解不同学习路径下的累积回报期望，即状态动作值函数

$$Q_\pi(\boldsymbol{s},\boldsymbol{a})=\mathbb{E}[G_t\mid\boldsymbol{s}_t=\boldsymbol{s},\boldsymbol{a}_t=\boldsymbol{a}] \tag{25}$$

最优策略的发现过程往往伴随着状态值函数的求解过程，常见的求解方法有动态规划、Sarsa和Q学习等算法[30]。然而，动态规划算法通常需要完整的环境信息，这对于本文动态的无线环境是无法实现的。Sarsa和Q学习虽然不需要完整的环境信息，但是由于需要存储Q表，只能处理状态和动作空间较小的问题。为此，DQN借鉴了函数逼近的思想，采用深度神经网络(Deep Neural Network, DNN)来逼近Q函数，降低了Q表的搜索难度。对于式(25)中的Q值函数，基于的DNN的近似过程可以表示为

$$Q_\pi(\boldsymbol{s},\boldsymbol{a};\theta)\approx Q_\pi(\boldsymbol{s},\boldsymbol{a}) \tag{26}$$

其中，θ代表的DNN的参数。

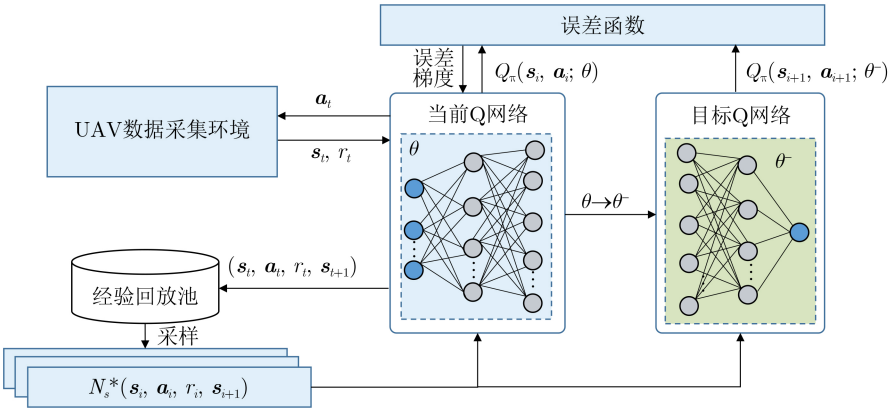


图 2 RIS辅助UAV时效数据收集策略结构

值得注意的是，DQN算法通常使用2个DNN来近似Q函数，分别称之为当前Q网络和目标Q网络[31]。DQN算法使用一个误差函数来优化这2个网络的参数，即

$$L^{\mathrm{DQL}}(\theta)=\mathbb{E}\left[\left(y^{\mathrm{DQL}}-Q_{\pi}(\boldsymbol{s},\boldsymbol{a};\theta)\right)^2\right] \tag{27}$$

其中，$Q_{\pi}(\boldsymbol{s},\boldsymbol{a};\theta)$代表当前Q网络，$y^{\mathrm{DQL}}$代表目标Q网络，即

$$y^{\mathrm{DQL}}=r+\gamma\max_{\boldsymbol{a}'}Q(\boldsymbol{s}',\boldsymbol{a}';\theta^{-}) \tag{28}$$

其中，θ^{-}表示目标Q网络的参数。

然而，传统DQN算法在更新目标网络时采用最大化Q值来选择和评估动作，会导致对Q值的过高估计[32]。DDQN引入了2个Q网络，分别用于动作选择和动作评估，从而缓解了目标网络的过估计的问题[32]。经过改进后，目标Q值表示为

$$y^{\mathrm{DDQL}}=r+\gamma Q_{\pi}(\boldsymbol{s}',\arg\max_{\boldsymbol{a}}Q_{\pi}(\boldsymbol{s}',\boldsymbol{a};\theta);\theta^{-}) \tag{29}$$

可以看到，式(29)中动作选择和评估是分离的。结合经验回放机制，DDQN算法的更新过程可以表示为

$$L^{\mathrm{DDQL}}(\theta)=\frac{1}{N_{\mathrm{s}}}\sum_{i=1}^{N_{\mathrm{s}}}\left(y_i^{\mathrm{DDQL}}-Q_{\pi}(\boldsymbol{s}_i,\boldsymbol{a}_i;\theta)\right)^2 \tag{30}$$

其中，N_{s}表示采样数目。

3.5 RIS相位偏移优化

当给定UAV的位置以及IoT设备的调度时，RIS的相位偏移需要获得进一步的优化以使UAV获得最大的数据收集速率。通过分析信号传播路径的几何关系，可以推导出IRS需要补偿的相位偏移，这提供了一种闭式解形式，使得相位调整仅依赖于位置参数，而无需实时复杂的信道状态信息[33,34]，即

$$\begin{aligned}\theta_{m_{\mathrm{c}},m_{\mathrm{r}},t}=\frac{2\pi}{\lambda}\{&(m_{\mathrm{r}}-1)d_{\mathrm{r}}\phi_{m,t}^{\mathrm{UR}}\varphi_{m,t}^{\mathrm{UR}}+(m_{\mathrm{c}}-1)\\&\cdot d_{\mathrm{c}}\psi_{m,t}^{\mathrm{UR}}\varphi_{m,t}^{\mathrm{UR}}+(m_{\mathrm{r}}-1)d_{\mathrm{r}}\phi_{n,t}^{\mathrm{RI}}\varphi_{n,t}^{\mathrm{RI}}\\&+(m_{\mathrm{c}}-1)d_{\mathrm{c}}\psi_{n,t}^{\mathrm{RI}}\varphi_{n,t}^{\mathrm{RI}}\}\end{aligned} \tag{31}$$

该公式旨在通过精确控制RIS的反射相位，使信号能够在空间中朝向特定方向反射，从而增强接收端的信号强度。具体而言，相位调整通过改变信号的传播路径，使得信号与接收端在空间中高效对齐，最大化能量传递效率。通优化策略设计，可动态调整RIS的相位，使无人机与IoT设备间的通信链路稳定，尤其在受到建筑遮挡的复杂环境中，可提高信号覆盖范围和通信质量。

3.6 算法整体设计

该算法基于DDQN优化UAV在数据采集任务中的轨迹选择和IoT设备的调度策略。整体流程如算法1所示。首先，随机初始化神经网络参数，然后在多个训练回合中迭代。对于每一回合，算法初始化仿真环境参数，并在每个时间步内，根据当前策略选择最优动作，通过公式计算RIS的相位偏移，并执行该动作控制UAV移动和IoT设备调度，同时计算获得的奖励并更新环境状态。如果UAV超出边界，则执行状态回滚操作并对该动作施加惩罚。接着，算法将当前状态、动作、奖励和新状态存储到经验池中，并从经验池中采样，更新目标网络以最小化损失。最后，每隔一定步长更新策略参数，并持续优化UAV的移动和IoT设备的调度，以实现最佳的UAV轨迹和IoT设备调度策略。

4 仿真结果分析

本文通过仿真实验验证提出算法性能，主要仿真参数如表1所示。本文通过仿真实验与以下3种算法做了对比分析：(1)固定相位偏移算法(FIXed Reconfigurable Intelligent Surface, FIX-RIS)[29]：固定RIS相位，使其偏向用户所在范围；(2)DQN算法：保持提出网络结构不变，使用式(27)、式(28)来训练数据收集策略；(3)随机策略：使用随机策略训练网络。

4.1 收敛性能

为了验证算法的收敛性能，本文将算法设置运行1 000回合，每回合运行1 000步，图3显示了所

算法 1 基于DDQN的UAV数据采集算法

输入：UAV观测到的环境状态

输出：UAV轨迹和IoT设备调度策略

(1) 随机初始化神经网络参数

(2) **for** episode = 1,2,$\cdots$,N^{EP} **do**

(3) 初始化仿真环境参数

(4) **for** t = 1,2,$\cdots$,T **do**

(5) 根据$Q_{\pi}(\boldsymbol{s},\boldsymbol{a};\theta)$选取状态$\boldsymbol{s}$对应的动作$\boldsymbol{a}$;

(6) 根据式(31)获得RIS的相位偏移;

(7) 执行动作控制UAV飞行和IoT调度后，使用式(23)计算瞬时奖励r并获得下一时刻的状态$\boldsymbol{s}'$;

(8) **if** UAV移动跃出边界 **do**

(9) 状态回滚$\boldsymbol{s}'\leftarrow\boldsymbol{s}$，为$r$添加惩罚项;

(10) 将环境数据$(\boldsymbol{s},\boldsymbol{a},r,\boldsymbol{s}')$存入经验池;

(11) **if** 经验池存满数据 **do**

(12) 从经验池取出N_{s}个样本;

(13) 对于每个样本，用目标值网络计算式(29);

(14) 更新当前Q网络以最小化损失式(30);

(15) 每隔一定步长$\theta^{-}\leftarrow\theta$;

(16) **end**

(17) **end**

表 1　主要仿真参数

参数	数值
x^{MAX}, y^{MAX}(m)	1 000, 1 000
L,z^{MIN}, z^{MAX}(m)	10, 60, 80
l_t(m)	{0, 1, 2}
c_1, c_2	12.081, 0.113 95[7]
μ^{LoS} , μ^{NLoS}	1.445 44, 199.526[7]
K_1, K_2	3, 4
ρ, β (dBm)	–30, –50[8]
δ_1, δ_2(dBm/Hz)	–174, –174
B_1, B_2(MHz)	2, 2
P_1, P_2(W)	0.5, 0.8
P_{B}, P_{I}, P_{V}	88.63, 79.85, 11.46
ζ_1, ζ_2, ζ_3, ζ_4	100, 1, 10, 10
Λ, Ξ	0.6, 0.05
U_{tip}(m/s)	120[29]
Υ(kg/m^3)	1.225[29]
G(m^2)	0.503[29]
χ(bit)	1 000

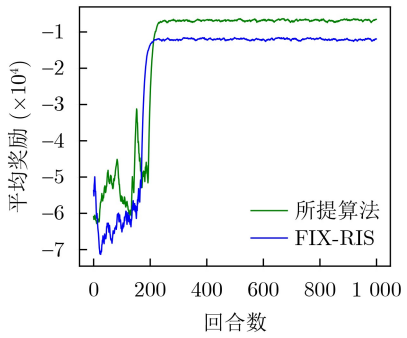


图 3 收敛曲线

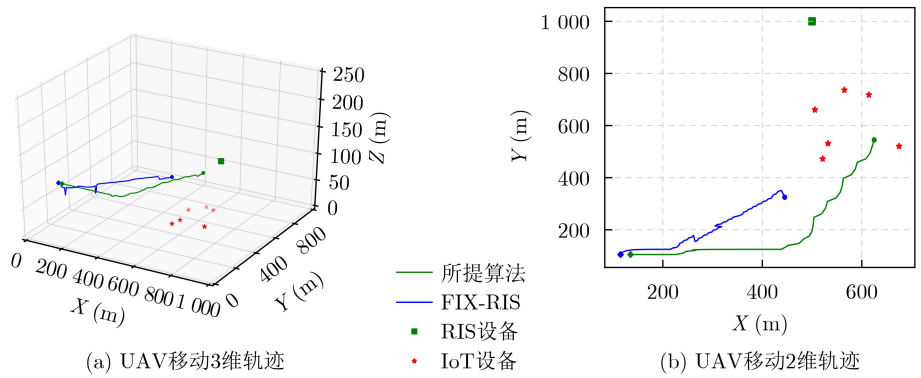
提算法和FIX-RIS收敛性能。其中，累计奖励值表示算法在每次迭代过程中所获得的总奖励之和，它衡量了无人机在特定策略下的整体任务表现。迭代回合数指算法训练过程中进行的迭代次数。在每个回合中，算法会不断学习并更新策略，以优化无人机的轨迹规划和数据收集效率。随着迭代次数增加，算法的学习效果逐步提升，最终在高回合数下达到稳定的累计奖励值。可以看到，当运行回合数超过200时，所提算法和FIX-RIS算法均可以收敛到一个稳定的值，这说明所提算法有较好的收敛性。还可以看到，随着回合数的增加，FIX-RIS算法收敛到一个较小的稳定值，这是因为FIX-RIS算法只使用了固定的RIS设置，导致算法收敛性较差。

4.2　UAV轨迹验证

为了验证提出UAV移动策略的有效性，图4分别展示了UAV在数据采集过程中产生的3维轨迹和2维轨迹。其中，红色五角星代表地面IoT设备，绿色方块代表了RIS设备，UAV统一从左下角出发，菱形和圆点分别代表UAV的起点和终点。从图4整体可以看到，UAV从出发后，通过不断调整在3维空间下的位置，最后到达IoT设备所在区域。因为在这些区域，UAV不但可以获得较好的直连数据率，还可以通过与RIS配合获得较好的通信性能。此外，通过比较提出算法和FIX-RIS算法可以看到，所提算法可以逐步运动到IoT设备所在区域从而可以获得较大的平均奖励，这一结论也可以从图3看出。对比文献[29]中使用的固定RIS朝向IoT设备的策略，本文提出的联合优化策略使UAV接近IoT设备所在区域，从而在RIS的辅助下可以获得更好的性能，而FIX-RIS对UAV移动适应性差，最终只能移动到距离IoT设备较远的位置，因而获得较差的性能。

4.3　数据采集速率需求对优化性能的影响

为了分析数据采集速率需求对算法优化性能的影响，图5分别展示了平均数据采集数据率需求对平均系统代价、平均AoI和平均系统能耗3个性能指标的影响。其中，平均AoI指数据从生成到被无人机接收到的时间延迟，反映了数据的时效性。能量消耗指无人机在数据收集过程中所消耗的总能量，包括飞行能量、数据收集和上传能量。系统代价衡



(a) UAV移动3维轨迹　　(b) UAV移动2维轨迹

图 4 UAV移动轨迹图

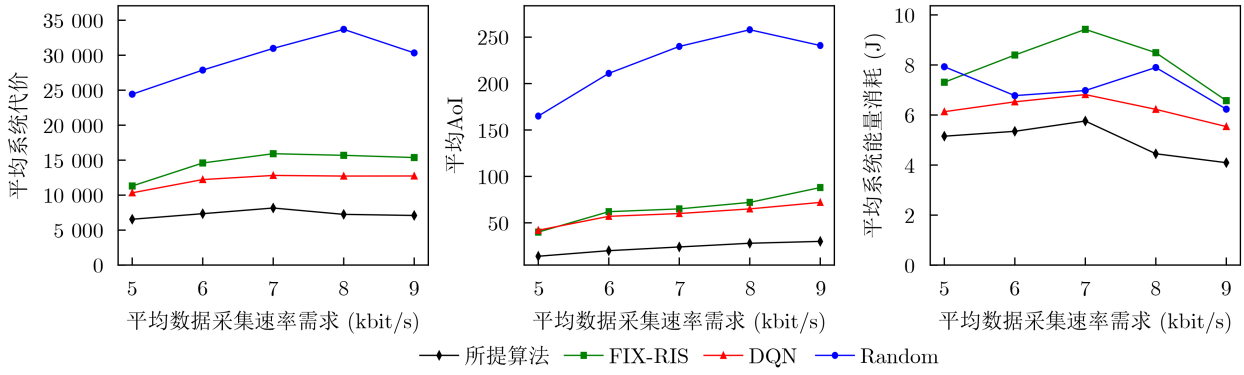
量了整个无人机数据收集任务的综合性能，反映了在能量消耗和数据时效性之间的平衡。首先，整体来看，相比于其他算法，随着平均数据采集速率需求的增大，所提算法的获得了较低的平均代价，这说明所提算法对IoT设备的数据采集速率需求有较好的适应性。其次，从图5(a)可以看到，随着平均数据采集速率需求的增大，所有算法获得平均代价都是先增大后缓慢降低，这是因为平均AoI随着数据采集速率需求的增加而增加，而系统能量随着数据需求的增加先增加后减小。具体来说，图5(b)展示的随着数据采集需求的增大，平均AoI逐步上升，这是因为随着需求增大，UAV越来越难满足数据采集需求，从而导致数据的时效性降低。而从图5(c)可以看出，系统能量消耗随着数据采集需求的增加而减小，这是因为当需求较小时UAV可以较容易完成该需求，从而获得较好的性能。而随着需求的增加，UAV在距离IoT设备较远的地方一直很难满足该数据传输需求，从而导致强化学习智能体采取较为节能的方式来减少移动。

4.4 IoT设备数对优化性能的影响

为了分析IoT设备数对算法优化性能的影响，图6分别展示了IoT设备数对平均系统代价、平均AoI和平均系统能耗的影响。首先，如图6(a)所示，随着IoT设备数目的增多，提出算法获得了较低的平均代价，这说明所提算法在IoT设备数的变化下获得了较好的性能。其次，从图6(b)可以看到，随着IoT设备数的增加，平均AoI曲线逐步上升，这是因为在IoT设备增加过程中，UAV完成数据采集需求的难度逐渐增加，导致AoI的上升。最后，从图6(c)可以看到随着IoT设备数的增加，UAV能量消耗逐步增加，但和其他方案想比，所提算法可以获得最低的能耗水平，因此提出算法可以获得较好的能量管理性能。

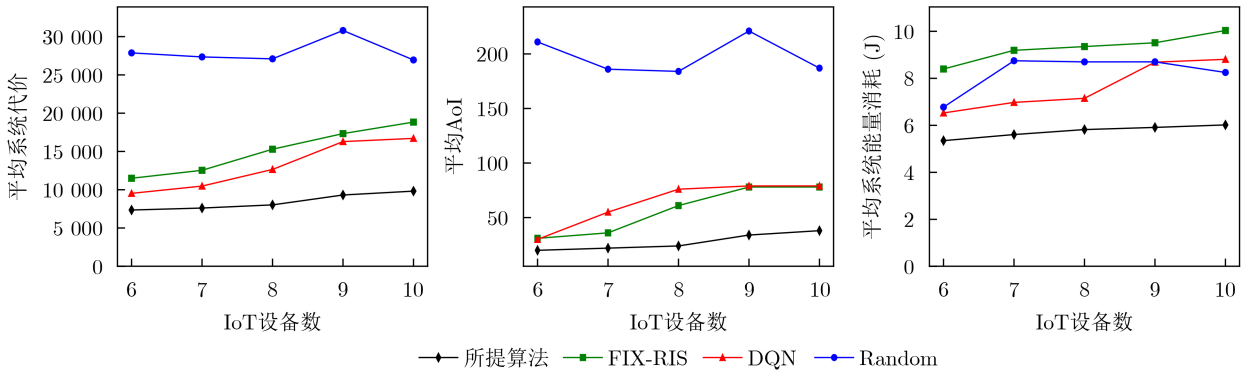
5 结束语

针对RIS辅助的UAV数据收集场景，本文构建了一个联合优化UAV轨迹规划、IoT设备调度和RIS相位调整的模型，综合考虑了数据收集时效性和能量消耗。通过引入DRL算法，本文提出一种智能优化方案，使得UAV能够根据实时环境动态调整飞行路径及通信策略，从而在数据收集任务中实现信息时效性和能量效率性能均衡。仿真结果表明，与



(a) 平均数据采集速率需求相对于平均系统代价　(b) 平均数据采集速率需求相对于平均AoI　(c) 平均数据采集速率需求相对于平均系统能耗

图 5 平均数据采集速率需求对优化性能的影响



(a) IoT设备数相对于平均系统代价　(b) IoT设备数相对于平均AoI　(c) IoT设备数相对于平均系统能耗

图 6 IoT设备数对优化性能的影响

传统方法相比，本文方法可以获得更高的平均奖励，并且在降低数据AoI和提升系统能效方面表现出较好的性能，未来的研究可以进一步扩展至多维度场景，包括多UAV协同工作、空域限制以及安全约束等复杂资源优化问题，为智慧城市和IoT应用提供新的技术支持。

参 考 文 献

[1] 段洁, 胡显静, 林欢, 等. 面向物联网数据特征的信息中心网络缓存方案[J]. 电子与信息学报, 2021, 43(8): 2240–2248. doi: 10.11999/JEIT200631.
DUAN Jie, HU Xianjing, LIN Huan, *et al*. Information-centric networking caching scheme for data characteristics of internet of things[J]. *Journal of Electronics & Information Technology*, 2021, 43(8): 2240–2248. doi: 10.11999/JEIT200631.

[2] JAVAID S, SAEED N, QADIR Z, *et al*. Communication and control in collaborative UAVs: Recent advances and future trends[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(6): 5719–5739. doi: 10.1109/TITS.2023.3248841.

[3] 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. doi: 10.11999/JEIT220195.
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. doi: 10.11999/JEIT220195.

[4] 张在琛, 江浩. 智能超表面使能无人机高能效通信信道建模与传输机理分析[J]. 电子学报, 2023, 51(10): 2623–2634. doi: 10.12263/DZXB.20221352.
ZHANG Zaichen and JIANG Hao. Channel modeling and characteristics analysis for high energy-efficient RIS-assisted UAV communications[J]. *Acta Electronica Sinica*, 2023, 51(10): 2623–2634. doi: 10.12263/DZXB.20221352.

[5] SAVKIN A V, HUANG Chao, and NI Wei. Joint multi-UAV path planning and LoS communication for mobile-edge computing in IoT networks with RISs[J]. *IEEE Internet of Things Journal*, 2023, 10(3): 2720–2727. doi: 10.1109/JIOT.2022.3215255.

[6] SAVKIN A V, HUANG Chao, and NI Wei. Collision-free 3-D navigation of a UAV team for optimal data collection in Internet-of-Things networks with reconfigurable intelligent surfaces[J]. *IEEE Systems Journal*, 2023, 17(3): 4070–4077. doi: 10.1109/JSYST.2023.3269095.

[7] LIN Xinzhong, XIE Cong, XIE Wenwu, *et al*. Security performance analysis of RIS-assisted UAV wireless communication in industrial IoT[J]. *The Journal of Supercomputing*, 2022, 78(4): 5957–5973. doi: 10.1007/s11227-021-04095-7.

[8] ZHAI Liangsen, ZOU Yulong, ZHU Jia, *et al*. RIS-assisted UAV-enabled wireless powered communications: System modeling and optimization[J]. *IEEE Transactions on Wireless Communications*, 2024, 23(5): 5094–5108. doi: 10.1109/TWC.2023.3324500.

[9] RANJHA A and KADDOUM G. URLLC facilitated by mobile UAV relay and RIS: A joint design of passive beamforming, blocklength, and UAV positioning[J]. *IEEE Internet of Things Journal*, 2021, 8(6): 4618–4627. doi: 10.1109/JIOT.2020.3027149.

[10] ESKANDARI M, HUANG Hailong, SAVKIN A V, *et al*. Model predictive control-based 3D navigation of a RIS-equipped UAV for LoS wireless communication with a ground intelligent vehicle[J]. *IEEE Transactions on Intelligent Vehicles*, 2023, 8(3): 2371–2384. doi: 10.1109/TIV.2022.3232890.

[11] 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. doi: 10.1109/JIOT.2023.3300700.

[12] AL-HILO A, SAMIR M, ELHATTAB M, *et al*. RIS-assisted UAV for timely data collection in IoT networks[J]. *IEEE Systems Journal*, 2023, 17(1): 431–442. doi: 10.1109/JSYST.2022.3215279.

[13] TYROVOLAS D, MEKIKIS P V, TEGOS S A, *et al*. Energy-aware design of UAV-mounted RIS networks for IoT data collection[J]. *IEEE Transactions on Communications*, 2023, 71(2): 1168–1178. doi: 10.1109/TCOMM.2022.3229672.

[14] LIU Jianghu and ZHANG Hongtao. Height-fixed UAV enabled energy-efficient data collection in RIS-aided wireless sensor networks[J]. *IEEE Transactions on Wireless Communications*, 2023, 22(11): 7452–7463. doi: 10.1109/TWC.2023.3250988.

[15] ALMASOUD A M. Robust anti-jamming technique for UAV data collection in IoT using landing platforms and RIS[J]. *IEEE Access*, 2023, 11: 70635–70651. doi: 10.1109/ACCESS.2023.3294596.

[16] YANG Bowen, YU Yao, LI Jianqi, *et al*. An AoI-guaranteed sensor data collection strategy for RIS-assisted UAV communication system[C]. 2023 IEEE/CIC International Conference on Communications in China, Dalian, China, 2023: 1–6. doi: 10.1109/ICCC57788.2023.10233285.

[17] SAMIR M, ELHATTAB M, ASSI C, *et al*. Optimizing age of information through aerial reconfigurable intelligent

surfaces: A deep reinforcement learning approach[J]. *IEEE Transactions on Vehicular Technology*, 2021, 70(4): 3978–3983. doi: 10.1109/TVT.2021.3063953.

[18] HUANG Hongli, LIU Juan, and XIE Lingfu. Intelligent reflecting surface-assisted fresh data collection in UAV communications[C]. The 11th International Conference in Communications, Signal Processing, and Systems, Singapore, 2022: 189–197. doi: 10.1007/978-981-99-2362-5_24.

[19] XIAO Xiongbing, WANG Xiumin, and LIN Weiwei. Joint AoI-aware UAVs trajectory planning and data collection in UAV-based IoT systems: A deep reinforcement learning approach[J]. *IEEE Transactions on Consumer Electronics*, 2024, 70(4): 6484–6495. doi: 10.1109/TCE.2024.3440406.

[20] JIANG Wenwen, AI Bo, LI Mushu, *et al*. Average age-of-information minimization in aerial IRS-assisted data delivery[J]. *IEEE Internet of Things Journal*, 2023, 10(17): 15133–15146. doi: 10.1109/JIOT.2023.3264618.

[21] CHEN Zhen, GUO Yeyong, ZHANG Peichang, *et al*. Physical layer security improvement for hybrid RIS-assisted MIMO communications[J]. *IEEE Communications Letters*, 2024, 28(11): 2493–2497. doi: 10.1109/LCOMM.2024.3427010.

[22] RUAN Chengyao, ZHANG Zaichen, JIANG Hao, *et al*. Wideband near-field channel covariance estimation for XL-MIMO systems in the face of beam split[J]. *IEEE Transactions on Vehicular Technology*. doi: 10.1109/TVT.2024.3471733.

[23] QIU Chen, WEI Zhiqing, YUAN Xin, *et al*. Multiple UAV-mounted base station placement and user association with joint fronthaul and backhaul optimization[J]. *IEEE Transactions on Communications*, 2020, 68(9): 5864–5877. doi: 10.1109/TCOMM.2020.3001136.

[24] SHI Wangqi, JIANG Hao, XIONG Baiping, *et al*. RIS-empowered V2V communications: Three-dimensional beam domain channel modeling and analysis[J]. *IEEE Transactions on Wireless Communications*, 2024, 23(11): 15844–15857. doi: 10.1109/TWC.2024.3434568.

[25] WEI Zhiqiang, CAI Yuanxin, SUN Zhuo, *et al*. Sum-rate maximization for IRS-assisted UAV OFDMA communication systems[J]. *IEEE Transactions on Wireless Communications*, 2021, 20(4): 2530–2550. doi: 10.1109/TWC.2020.3042977.

[26] XU Peng, NIU Wenqi, CHEN Gaojie, *et al*. Performance analysis of RIS-assisted systems with statistical channel state information[J]. *IEEE Transactions on Vehicular Technology*, 2022, 71(1): 1089–1094. doi: 10.1109/TVT.2021.3126374.

[27] JIANG Hao, SHI Wangqi, ZHANG Zaichen, *et al*. Large-scale RIS enabled air-ground channels: Near-field modeling and analysis[J]. arXiv: 2403.12781, 2024.

[28] LIN Na, TANG Hailun, ZHAO Liang, *et al*. A PDDQNLP algorithm for energy efficient computation offloading in UAV-assisted MEC[J]. *IEEE Transactions on Wireless Communications*, 2023, 22(12): 8876–8890. doi: 10.1109/TWC.2023.3266497.

[29] CHEN Guqiao, CHENG Changjun, XU Xiaoli, *et al*. Minimizing the age of information for data collection by cellular-connected UAV[J]. *IEEE Transactions on Vehicular Technology*, 2023, 72(7): 9631–9635. doi: 10.1109/TVT.2023.3249747.

[30] WANG Xu, WANG Sen, LIANG Xingxing, *et al*. Deep reinforcement learning: A survey[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(4): 5064–5078. doi: 10.1109/TNNLS.2022.3207346.

[31] MNIH V, KAVUKCUOGLU K, SILVER D, *et al*. Human-level control through deep reinforcement learning[J]. *Nature*, 2015, 518(7540): 529–533. doi: 10.1038/nature14236.

[32] VAN HASSELT H, GUEZ A, and SILVER D. Deep reinforcement learning with double q-learning[C]. The 30th AAAI Conference on Artificial Intelligence, Phoenix, USA, 2016: 2094–2100. doi: 10.1609/aaai.v30i1.10295.

[33] WANG Liang, WANG Kezhi, PAN Cunhua, *et al*. Joint trajectory and passive beamforming design for intelligent reflecting surface-aided UAV communications: A deep reinforcement learning approach[J]. *IEEE Transactions on Mobile Computing*, 2023, 22(11): 6543–6553. doi: 10.1109/TMC.2022.3200998.

[34] 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. doi: 10.1109/TVT.2022.3143839.

张 涛：男，博士，研究方向为无线异构网络、通感一体低空网络、自主决策技术、智能化无线网络.

张 迁：男，博士，研究方向为计算机视觉、5gnr.

朱颖雯：女，副教授，研究方向为数据挖掘、人工智能、机器学习.

代 陈：男，博士，研究方向为全解耦异构网络、感传算一体网络.

责任编辑：余 蓉

Energy Aware Reconfigurable Intelligent Surface Assisted Unmanned Aerial Vehicle Age of Information Enabled Data Collection Policies

ZHANG Tao[①②]　ZHANG Qian[①]　ZHU Yingwen[①]　DAI Chen[③]

[①](*School of Information Technology, Jiangsu Open University, Nanjing* 210000, *China*)

[②](*College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing* 210016, *China*)

[③](*School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing* 210023, *China*)

Abstract:

Objective　This study aims to develop and implement an optimization framework that addresses the critical balance between energy consumption and information freshness in Unmanned Aerial Vehicle (UAV)-assisted Internet of Things (IoT) data collection systems, enhanced by Reconfigurable Intelligent Surfaces (RIS). In complex urban environments, traditional line-of-sight communication between UAVs and ground-based IoT devices is often obstructed by buildings and infrastructure, hindering comprehensive coverage and efficient data collection. While RIS technology offers promising solutions by dynamically adjusting signal reflection directions, optimizing communication signal coverage, and enhancing quality, it introduces additional complexity in system design and resource allocation, requiring sophisticated adaptive optimization techniques. The integration of RIS enables stable communication connections across various UAV flight heights and angles, mitigating disruptions caused by obstacles or signal interference, thus improving data collection efficiency and reliability. However, this integration must account for multiple factors, including UAV energy consumption, communication complexity, and Age of Information (AoI) constraints. These approaches must adapt to the dynamic nature of UAV operations and fluctuating communication conditions, ensuring optimal performance in terms of energy efficiency and data freshness. The research also addresses several key challenges, including real-time adaptation to environmental changes, optimal scheduling of IoT device interactions, dynamic adjustment of RIS phase configurations, efficient trajectory planning, and the maintenance of data freshness under various system constraints. The proposed framework establishes a robust foundation for next-generation IoT data collection systems that can adapt to diverse operational conditions while maintaining high performance standards. This is achieved through the implementation of advanced deep reinforcement learning techniques, specifically designed to manage the complex interplay between UAV mobility, RIS configuration, and IoT device scheduling, ensuring efficient and timely data collection while optimizing system resources.

Methods　A comprehensive data collection optimization strategy is proposed, based on deep reinforcement learning principles, specifically designed to address the complex challenges in UAV-assisted IoT data collection systems enhanced by RIS technology. The methodology employs a Double Deep Q-Network (DDQN) architecture, integrating UAV trajectory planning, IoT device scheduling, and RIS phase adjustment within a three-dimensional grid-based movement space. The system incorporates a channel model that accounts for both direct and RIS-assisted communication paths, including a probabilistic path loss model for direct links and Rician fading for RIS-assisted links. The optimization problem is formulated as a Markov Decision Process (MDP), where the state space includes the UAV position, previous movement information, and average AoI, while the action space involves 3D movement decisions and IoT device scheduling. The reward function is designed to balance multiple performance metrics, including system AoI, UAV flight energy consumption, data collection energy, data upload energy, and penalties for boundary violations. The DDQN implementation utilizes two Q-networks—the current and target networks—separating action selection from action evaluation, effectively addressing the issue of Q-value overestimation. The training process incorporates experience replay for sample storage and periodic updates to the target network to enhance learning stability. Additionally, the RIS phase shift optimization is derived through geometric relationships, considering both direct and RIS-

assisted communication paths. This comprehensive approach enables the joint optimization of UAV trajectory, IoT device scheduling, and RIS phase adjustment, while ensuring energy efficiency and timely data collection in complex communication environments.

Results and Discussions The proposed method enables the UAV to dynamically adjust its flight trajectory and communication strategy based on real-time environmental conditions, enhancing data transmission efficiency while reducing energy consumption. Extensive simulation experiments comprehensively evaluate the performance of the DDQN-based optimization framework. Convergence analysis demonstrates that the method achieves faster and more stable convergence compared to traditional DQN approaches. The average reward steadily increases and stabilizes after approximately 200 episodes, while baseline methods exhibit slower convergence and higher performance variance (Fig. 3). The optimized UAV trajectory visualization shows that the method effectively guides the UAV to collect data efficiently from all IoT devices while avoiding unnecessary detours. The trajectory strikes a balance between visiting high-priority devices (those with higher AoI) and maintaining energy-efficient flight paths, clearly illustrating the effectiveness of the joint optimization of movement and device scheduling decisions (Fig. 4). Energy consumption analysis reveals that the proposed method achieves superior energy efficiency, with a 15% reduction in total energy consumption while maintaining comparable data collection performance. This improvement results from the intelligent integration of RIS-assisted communication and optimal trajectory planning, which reduces the need for energy-intensive maneuvers and prolonged hovering periods (Fig. 5) (Fig. 6). The AoI performance evaluation further confirms the method's effectiveness in maintaining data freshness. The average AoI across all IoT devices remains consistently lower than in baseline methods, with a 20% improvement in worst-case AoI values. This demonstrates the method's ability to balance the trade-off between visiting different devices and maintaining acceptable AoI levels, even under challenging network conditions. The framework's adaptive nature is evident in its capacity to prioritize devices with critical AoI values while maintaining overall system efficiency, showing robust performance across varying network densities and device distributions (Fig. 5) (Fig. 6).

Conclusions The proposed deep reinforcement learning-based optimization policy effectively addresses the complex challenges in UAV-assisted IoT data collection systems enhanced by RIS technology, demonstrating significant improvements in both energy efficiency and information freshness. The integration of advanced learning techniques with RIS-assisted communication provides a robust and adaptive solution for practical deployment in urban IoT environments. The comprehensive evaluation framework and detailed performance analysis offer valuable insights for system designers and practitioners. The superior performance in terms of convergence speed, trajectory optimization, energy efficiency, and AoI management confirms the effectiveness of the proposed approach. Future research will focus on extending the framework to multi-UAV coordination scenarios, exploring the impact of dynamic environmental changes, and developing more sophisticated reward mechanisms to address additional operational constraints, such as security and airspace restrictions. The promising results also indicate potential applications in emergency response systems, smart city infrastructure, and environmental monitoring networks.

Key words: UAV-assisted communication; Age of Information (AoI); Deep reinforcement learning; Reconfigurable Intelligent Surfaces (RIS)