arXiv:2301.06749v2 [cs.MA] 8 Mar 2023

基于强化学习的大规模无人机系统事件触发最优编队跟踪控制

Ziwei Yan[1], Liang Han[1*]、Xiaoduo Li[1],Jinjie Li[2] 和 Zhang Ren[2]

摘要—大规模无人机切换编队跟踪控制已广泛应用于搜救、协同运输和无人机灯光秀等诸多领域。为优化控制性能并降低系统的计算负担，本研究针对离散时间大规模无人机系统（UASs）提出了一种事件触发最优编队跟踪控制器。此外，通过引入匈牙利算法和演员-评论家神经网络（神经网络）的实现，构建了最优决策—最优控制框架。最后，搭建了一个大规模混合现实实验平台，以验证所提算法的有效性。该平台包括大量虚拟无人机节点和数量有限的实体无人机节点。这弥补了真实场景中实验场地和设备的局限，保障了实验安全，大幅降低了实验成本，并适用于实现大规模无人机编队灯光秀。

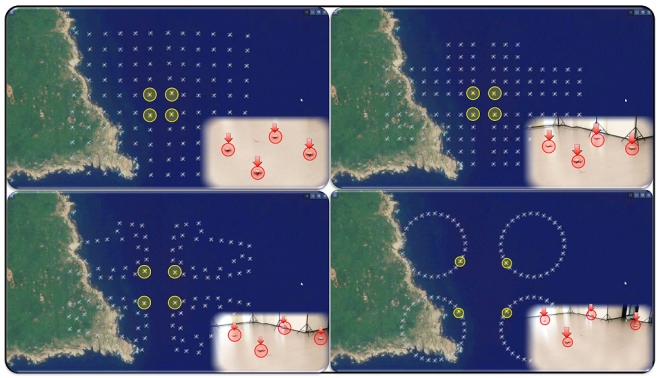


图1. 大规模混合现实实验平台中的数百架虚拟无人机和四架实体无人机，共同实现无人机编队灯光秀。

一、引言

近年来，关于多无人机系统协同控制的研究层出不穷。其中，多无人机编队跟踪控制是一项常见且基础的问题，编队规模也从以往的少数无人机群发展到如今的大规模编队。随着编队规模呈指数级增长，对系统性能的要求也日益严格。因此，研究人员提出了最优控制，以优化系统性能和编队成本，减轻系统负担。

最优控制是平衡系统性能与计算消耗的有效策略。对于离散时间系统，可通过求解贝尔曼方程 [1]–[4] 实现最优控制。然而，由于难以直接推导出贝尔曼非线性方程的解析解，以上研究均针对线性系统。为近似求解非线性系统的最优控制器，文献 [5]–[10]，引入了强化学习（RL），其中演员-评论家神经网络（NN）是常用的强化学习架构。评论家神经网络通过评估系统性能指标，为优化动作提供反馈；执行器神经网络则发出优化后的控制指令，以改善系统行为。

本研究部分得到“新一代人工智能”国家科技创新2030重大项目（项目编号：2018AAA0102305）资助，部分得到国家自然科学基金（项目编号：61803014和61873011）资助。

*通讯作者：L. Han，liang_han@buaa.edu.cn。

[1]Z. Yan、L. Han和X. Li就职于北京航空航天大学中法工程师学院，中国北京 100191。yanziwei@buaa.edu.cn; liang_han@buaa.edu.cn; lixiaoduo@sjtu.edu.cn。

[2]李进杰和任章就职于中国北京100191北京航空航天大学自动化科学与电气工程学院。lijinjie@buaa.edu.cn; renzhang@buaa.edu.cn.

近年来，强化学习架构已广泛应用于最优协同控制。在 [11]，中，采用一种基于学习的自适应动态规划算法，解决障碍环境下具有模型不确定性的多无人机系统编队跟踪问题。在 [12]，中，针对存在执行器故障的多智能体跟踪控制，通过结合强化学习算法与反步法，设计了一种自适应最优容错跟踪控制器。然而，以上研究均采用时间触发机制，持续或定期更新控制器和神经网络。当编队规模较大时，这必然会造成巨大的计算开销。

本研究提出了事件触发机制：仅当特定条件触发时，无人机才更新控制器和执行器神经网络。这使无人机系统能够按需、非周期性地更新控制指令，而非持续且无差别地更新。同时，与时间触发的情形相比，事件触发机制能够有效降低系统的计算负担。针对基于强化学习的最优协同问题，已有多项研究将事件触发机制作为一种改进方法，涵盖连续时间系统 [13]–[18] 和离散时间系统 [19]–[24]。然而，这些研究仅进行了数值仿真，并未验证算法在工程应用中的有效性。

为验证所提出算法的效率，本研究建立了一个大规模混合现实实验平台。利用有限数量的实体无人机和数百架虚拟无人机，实现了一场大型无人机灯光秀（图 1）。考虑到大型编队表演的主要局限，例如对实验场地要求高、实验设备昂贵，以及

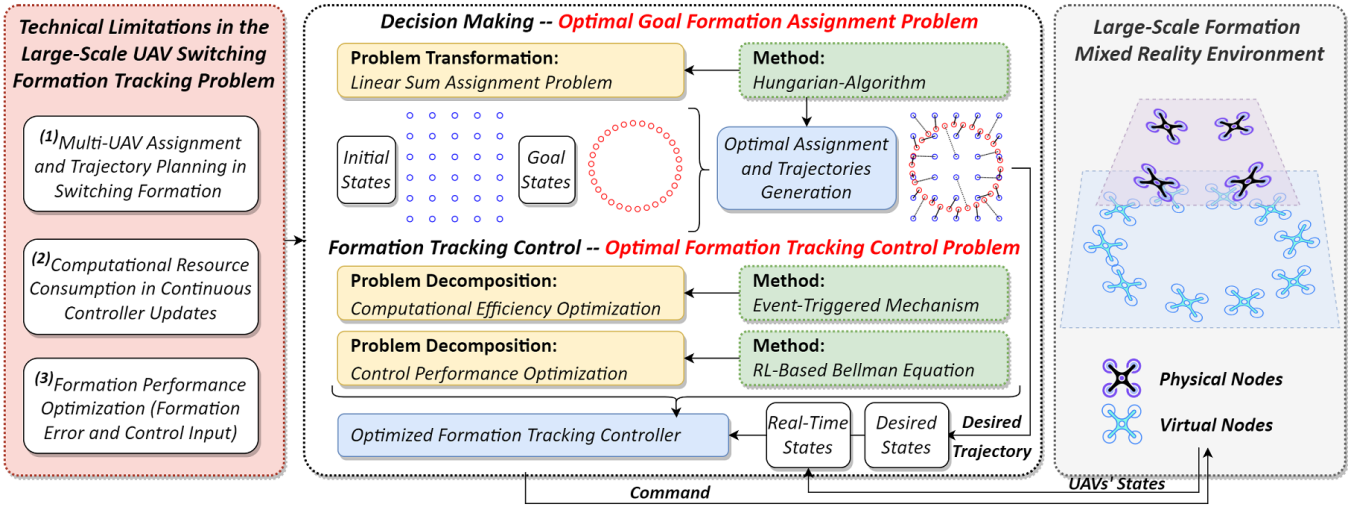


图 2. 本研究的算法概述。为解决大规模无人机切换编队跟踪问题，在决策层采用匈牙利算法求解目标编队分配问题；在控制层，分别采用事件触发机制和基于强化学习的贝尔曼方程来优化系统的计算效率和控制性能。

安全系数低，该实验平台具有实验成本低、安全系数高的优点。实验结果表明，该平台仅需一间房间和四架无人机，即可在广阔空域中实现数百架无人机的预期效果。

本文针对离散时间大规模无人机系统提出了一种事件触发最优编队跟踪控制器，并建立了一种在线演员-评论家神经网络结构来实现控制。本研究的主要贡献包括：

1）针对大规模切换编队在决策层和控制层面临的主要困难，建立了一种集成最优编队分配与最优控制的算法框架。2）提出了事件触发在线演员-评论家神经网络，用于逼近局部性能指标并学习控制器。与 [5]–[12],相比，演员神经网络权重和控制器仅在触发时刻更新。3）本研究强调所提算法的工程应用，建立了一个大规模无人机编队混合现实实验平台，并在该平台上验证了所提算法。

II. 算法概述

如图 2 所示，本文设计了一种算法框架，用于解决大规模无人机最优编队跟踪问题。最优分配和最优控制算法分别总结于算法 1 和算法 2 中。

算法 1 决策层 - 最优分配

输入： 1) 初始位置集合 $\vec{\mathbf{G}}$；2) 期望编队为 $\vec{\mathbf{S}}$；3) 无人机数量 N。

输出： 1) 目标位置集合 $\vec{\mathbf{F}}$；2）最优分配列表 χ^*；3）期望切换轨迹$h(k)$。

1: $[\vec{\mathbf{F}}, \chi^*, h(k)]$=最优_分配($\vec{\mathbf{G}}, \vec{\mathbf{S}}, N$)。

算法 2 控制层 - 最优控制

输入：1）期望轨迹 $\mathrm{h(k)}$；2）实时状态 $\mathrm{x(k)}$。输出：1）优化控制器 $\mathrm{u(k)}$；2）触发时刻序列 $\{\mathrm{k}_l^i\}_{l\in\mathbb{Z}^+}$。1：初始化演员-评论家神经网络；2：根据 $\mathrm{h}_i\mathrm{(k)}$ 和 $\mathrm{x}_i\mathrm{(k)}$ 计算触发条件 $\mathrm{f}_i\mathrm{(k)}$；3：如果 $\mathrm{f}_i\mathrm{(k)} > 0$ 则 4：使用执行器神经网络学习控制器 $\mathrm{u}_i\mathrm{(k)}$；5：计算并更新执行器神经网络；6：将对应时刻 k 记录在 $\{\mathrm{k}_l^i\}_{l\in\mathbb{Z}^+}$ 中。7：否则 8：$\mathrm{u}_i\mathrm{(k+1)} = \mathrm{u}_i\mathrm{(k)}$；9：执行器神经网络保持不变。10：结束 if；11：计算并更新评论家神经网络；12：近似系统性能指标。

III. 预备知识

A. 图论

对于一个包含虚拟领导者和 N 个跟随者的大规模无人机系统，其通信网络可描述为图$\mathscr{G} = \{\mathscr{A}, \mathscr{E}, \mathscr{W}\}$，其中 $\mathscr{A} = \{\mathrm{a1},\ldots,\mathrm{a}_N\}$表示节点集，$\mathscr{E}$表示边集，且 $\mathrm{e_ij} = (\mathrm{a}_i, \mathrm{a\,j}) \in \mathscr{E}$；$\mathscr{W}$为邻接矩阵，其元素为 $\mathrm{w}ij$，对于 $\mathrm{e}_{ij} \in \mathscr{E}$，$\mathrm{i} \neq \mathrm{j}$，$\mathrm{w}i\mathrm{j} = 1$，否则 $\mathrm{w}_{ij} = 0$。定义 N_i 为无人机 a_i的邻居集。于是，入度矩阵为 $\mathscr{D} = \mathrm{diag}_{i\in[1,N]}\{\sum_{j\in N_i}\mathrm{w}i\mathrm{j}\}$。拉普拉斯矩阵为 $\mathrm{L} = \mathscr{D} - \mathscr{W}$。此外，对角矩阵 $\mathrm{B} = \mathrm{diag}_{i\in[1,N]}\{\mathrm{b}_i\}$ 描述跟随者与虚拟领导者之间的通信。假设领导者至少能够与一名跟随者通信，且在有向群 $\hat{\mathscr{G}} = \{\hat{\mathscr{A}}$中存在一棵以领导者为根的生成树，其中 $\hat{\mathscr{E}}\}$ 包含节点 $\mathrm{a0},\ldots,\mathrm{a}_N\}$，且 $\hat{\mathscr{E}} \in \hat{\mathscr{A}} \times \hat{\mathscr{A}}$。

假设 1. 图 $\mathscr{G}$ 是连通的，且有向图 $\hat{\mathscr{G}}$ 包含一棵生成树。

B. 径向基函数（RBF）神经网络

未知非线性函数 $\zeta(\Omega)$ 可由径向基函数神经网络近似表示：$\zeta(\Omega) = W^{*\mathrm{T}}\psi(\Omega)$，其中 $W^* \in \mathbb{R}^s$是权重向量，$\psi(\Omega) \in \mathbb{R}^s$是基函数向量，其中 $\psi(\Omega) = [\psi_1(\Omega)\ldots,\psi_s(\Omega)]^{\mathrm{T}}$，且 $s \in \mathbb{Z}^+$表示隐藏层中的神经网络节点数。本文中，$\psi_i(\Omega)$ 选为高斯基函数

$$\psi_i(\Omega) = \exp\left(-(\Omega-\nu_i)^{\mathrm{T}}(\Omega-\nu_i)/2\iota_i^2\right),\ i=1,\ldots,s. \quad (1)$$

其中，ν_i 是高斯核函数的中心，ι_i是该函数的宽度参数。对于径向基函数神经网络的逼近，总是可以找到一个理想权重$W^* = \arg\min_W\{\sup_\Omega ||\zeta(\Omega) - W^{\mathrm{T}}\psi(\Omega)||\}$。因此，未知函数 $\zeta(\Omega)$可以表示为

$$\zeta(\Omega) = W^{*\mathrm{T}}\psi(\Omega) + \sigma(\Omega). \quad (2)$$

其中，$\sigma(\Omega)$是逼近误差，且 $|\sigma(\Omega)| \le \sigma_T$，$\sigma_T$ 是一个常数。

C. 问题描述

考虑一个包含N 个跟随者和一个虚拟领导者的大规模UAS，其中每架无人机$a_i(i \in \{0,\ldots,N\})$的模型如下：

$$\begin{cases} p_i(k+1) = p_i(k) + v_i(k)T \\ v_i(k+1) = v_i(k) + u_i(k)T + f_i(p_i(k), v_i(k)). \end{cases} \quad (3)$$

其中，$p_i(k) \in \mathbb{R}^m$ 和$v_i(k) \in \mathbb{R}^m$ 分别是无人机a_i的位置和速度，$u_i(k) \in \mathbb{R}^m$ 是控制输入，m 表示无人机的运动维度；$f_i(\cdot)$是未知的有界非线性函数。对于虚拟领导者a_0，，$u_0(k)$是已知的，且$f_0(\cdot) = 0$。

然后，对于每架跟随无人机 a_i，位置跟踪误差$\xi_{pi} \in \mathbb{R}^m$ 和速度跟踪误差 $\xi_{vi} \in \mathbb{R}^m$ 定义为

$$\begin{cases} \xi_{pi}(k) = p_i(k) - p_0(k) - \eta_{pi}(k) \\ \xi_{vi}(k) = v_i(k) - v_0(k) - \eta_{vi}(k). \end{cases} \quad (4)$$

其中，$\eta_i(k) = [\eta_{pi}^{\mathrm{T}}(k)\ \eta_{vi}^{\mathrm{T}}(k)]^{\mathrm{T}} \in \mathbb{R}^{2m}$ 是领航者a_0 与跟随者 a_i 之间的期望状态差。然后，将无人机a_i 的分歧误差定义为

$$\begin{cases} \varepsilon_{pi}(k) = \sum_{j\in N_i} w_{ij}(\xi_{pi}(k) - \xi_{pj}(k)) + b_i\xi_{pi}(k) \\ \varepsilon_{vi}(k) = \sum_{j\in N_i} w_{ij}(\xi_{vi}(k) - \xi_{vj}(k)) + b_i\xi_{vi}(k). \end{cases} \quad (5)$$

令 $\varepsilon(k) = [\varepsilon_1^{\mathrm{T}}(k)\ldots,\varepsilon_N^{\mathrm{T}}(k)]^{\mathrm{T}}$，其中 $\varepsilon_i(k) = [\varepsilon_{pi}^{\mathrm{T}}(k)\ \varepsilon_{vi}^{\mathrm{T}}(k)]^{\mathrm{T}}$。将局部性能指标函数（PIF）定义为

$$V_i(\varepsilon_i(k), u_i(k)) = \sum_{l-k}^{\infty} \beta^{l-k} R_i(\varepsilon_i(k), u_i(\varepsilon_i)), \quad (6)$$

其中，$R_i(\varepsilon_i(k),u_i(\varepsilon_i)) = \varepsilon_i^{\mathrm{T}}(k)\varepsilon_i(k) + u_i^{\mathrm{T}}(\varepsilon_i)u_i(\varepsilon_i)$ 是效用函数。

IV. 最优分配与最优控制

本节将图 2 所示的算法概述分为三个部分展开并分别详细介绍：最优编队分配、事件触发式最优控制器设计，以及在线执行器-评价器神经网络的实现。

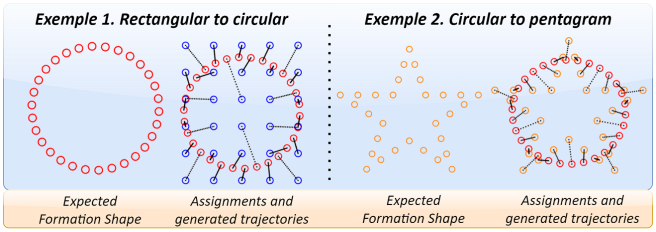


图 3. 切换编队的两个示例：左侧为期望编队形状，右侧为分配结果和生成的轨迹。

A. 切换编队的最优分配

对于由N架半径为r的无人机构成的系统，假设其初始位置集合为 $\vec{\mathbf{G}} = \{\vec{\mathbf{g}}_i^{\mathrm{T}}\} \in \mathbb{R}^{N\times m}$，期望编队为 $\vec{} = \{\vec{}_j^{\mathrm{T}}\} \in \mathbb{R}^{N\times m}$ $\vec{} = \{\vec{}_j^{\mathrm{T}}\} \in \mathbf{Ss}$，目标位置集合为$\mathbf{Ff}$，且$\mathbf{fs}_j d$，其中$d$是缩放因子，$\vec{}$，而$d$ 是平移向量。 $\mathbb{R}^{N\times m}$ $\vec{}_j = \rho\vec{} + \vec{}$ $\rho \in \mathbb{R}^+$

引理1. *[25]* 最优缩放因子 ρ^* 和位移 $\vec{}$向量 $\vec{d}^*$ 可计算如下

$$\rho^* = \frac{\sum_{i=1}^N \vec{\mathbf{g}}_i(\sum_{j=1}^N \vec{\mathbf{s}}_j)^{\mathrm{T}} + N\kappa^*}{\sum_{j=1}^N \vec{\mathbf{s}}_j(\sum_{j=1}^N \vec{\mathbf{s}}_j)^{\mathrm{T}} - N\sum_{j=1}^N \vec{\mathbf{s}}_j\vec{\mathbf{s}}_j^{\mathrm{T}}},$$
$$\vec{d}^* = \frac{\sum_{i=1}^N \vec{\mathbf{g}}_i - \rho^*\sum_{j=1}^N \vec{\mathbf{s}}_j}{N}. \quad (7)$$

算法 3 [25]伪代价分配算法

$\vec{}$**输入：** 1) 初始位置集合 $\mathbf{G}$；2) 期望编队为 $\mathbf{S}$。**输出：** 1) 最优分配列表 $\chi^* \in \mathbb{Z}^N$；2) 最小代价矩阵 $\kappa^* \in \mathbb{R}^{N\times N}$。

1 $\kappa_i = -\vec{}_i\vec{}_j^{\mathrm{T}}.:_j\mathbf{gs}$：($\chi^*$,) 匈牙利线性和

2 $\kappa^* =$ 分配问题算法(κ)。

引入算法3，使无人机能够同时到达目标位置且不发生碰撞。将编队切换持续时间定义为 $t_s = \max_{i\in[1,N]} ||\vec{}\mathbf{g}_i - \vec{}\mathbf{f}_{\chi(i)}||_2/v_{\max}$，其中 $\chi(i)$ 是分配给无人机a_i的位置索引，且 $v_{\max}$ 是最大允许速度。于是，无人机a_i 的期望切换轨迹为

$$h_i(k) = \vec{\mathbf{g}}_i + \frac{\vec{\mathbf{f}}_{\chi(i)} - \vec{\mathbf{g}}_i}{t_s}kT, \qquad k \in [0, \frac{t_s}{T}]. \quad (8)$$

无人机的避碰条件记为$||\vec{}\mathbf{g}_i - \vec{}\mathbf{g}_j||_2 \ge 2\sqrt{2}r$和 $||\vec{}\mathbf{s}_i - \vec{}\mathbf{s}_j||_2 \ge 2\sqrt{2}r$。

B. 事件触发最优控制器设计

定义无人机a_i 的触发时刻序列为$\{k_l^i\}_{l\in\mathbb{Z}^+}$，其由以下条件确定，

$$f_i(k) = ||e_l^i(k)||^2 - \frac{(1-2\kappa^2)}{2\kappa^2}||\varepsilon_i(k)||^2, \quad (9)$$

其中，$\kappa \in (\frac{1}{2}, \frac{\sqrt{2}}{2})$，$\varepsilon_i(k)$ 是当前分歧误差，控制器 $u_i(k)$ 仅在触发时刻$k_l^i = \inf\{k > k_{l-1}^i, f_i(k) > 0\}$ 更新，即 $\hat{}u_i(k) = u_i(\varepsilon_i(k_l^i))$，其中

$\varepsilon_i(k_l^i)$是触发误差。直到下一次触发时刻到来前，采用零阶保持器（ZOH），于是有

$$u_i(k)=\hat{u}_i(k)=u_i(k_l^i),\quad k\in[k_l^i,k_l^{i+1}). \tag{10}$$

引理 2. *[20]* 存在一个正常数 κ 满足如下不等式，

$$||\varepsilon_i(k+1)||=||g_i(\varepsilon_i(k),u_i(k_l^i))||\le\kappa||\varepsilon_i(k)||+\kappa||e_l^i(k)|| \tag{11}$$

其中 $e_l^i(k)=\varepsilon_i(k_l^i)-\varepsilon_i(k)$，$k\in[k_l^i,\ k_l^{i+1})$。当事件触发时，$e_l^i(k)=0$。

定理 1. 考虑满足假设 *2* 和触发条件 *(9)* 的大规模*UAS* (3)，其分歧误差渐近稳定，并实现了大规模*UAS* (3)的编队跟踪控制。

证明。 **(a)** 对于 $\forall k\in[k_l^i,k_{l+1}^i)$，控制器 $u_i(k)$ 保持为常数，$u_i(k)=u_i(k+1)$。设计李雅普诺夫函数

$$\mathscr{V}_{ei}(k)=\varepsilon_i^{\mathrm{T}}(k)\varepsilon_i(k)+u_i^{\mathrm{T}}(k)u_i(k). \tag{12}$$

结合引理 2 和触发条件$\Delta\mathscr{V}_{ei}(k)=\mathscr{V}_{ei}(k+1)-\mathscr{V}_{ei}(k)$，(9)，可推导得

$$\Delta\mathscr{V}_{ei}(k)\le2\kappa^2||\varepsilon_i(k)||^2+2\kappa^2||e_l^i(k)||^2-||\varepsilon_i(k)||^2\le0. \tag{13}$$

(b) 对于 $k\in\{k_l^i\}_{l\in\mathbb{Z}^+}$，控制器 $u_i(k)$ 将更新，李雅普诺夫函数设计为

$$\mathscr{V}_{ei}(k)=\beta^kV_i(\varepsilon_i(k)). \tag{14}$$

那么，$\mathscr{V}_{ei}$ 的一阶差分为

$$\Delta\mathscr{V}_{ei}=-\beta^kR_i(\varepsilon_i(k),u_i(k_l^i))\le0. \tag{15}$$

总之，$\varepsilon_i(k)$ 渐近收敛，这意味着实现了事件触发编队跟踪控制。 □

C. 在线执行器-评价器神经网络的实现

最优控制器 $u_i^*(k)$ 是方程 (6) 的唯一解，且 $u_i^*(k)$ 可由方程 $\partial V_i(\varepsilon_i(k))\,\partial u_i(k)=0$ 得到，

$$u_i^*(k)=-\frac{\beta\gamma_iT}{2}\frac{\partial V_i^*(\varepsilon_{vi}(k+1))}{\partial\varepsilon_{vi}(k+1)}. \tag{16}$$

where $\gamma_i=\sum_{j\in N_i}w_{ij}+b_i$. Let $\dfrac{\partial V_i^*(\varepsilon_{vi}(k+1))}{\partial\varepsilon_{vi}(k+1)}=\dfrac{2\alpha_p\varepsilon_{pi}(k)}{\beta\gamma_iT}+\dfrac{2\alpha_v\varepsilon_{vi}(k)}{\beta\gamma_iT}+\dfrac{V_i^0(\varepsilon_i)}{\beta\gamma_iT}$, where constants $\alpha_p>0$, $\alpha_v>0$, and $V_i^0(\varepsilon_i)=-2\alpha_p\varepsilon_{pi}(k)-2\alpha_v\varepsilon_{vi}(k)+\beta\gamma_iT\dfrac{\partial V_i^*(\varepsilon_{vi}(k+1))}{\partial\varepsilon_{vi}(k+1)}$. Then one has

$$u_i^*(k)=-\alpha\varepsilon_i(k)-\frac{V_i^0(\varepsilon_i)}{2}, \tag{17}$$

其中 $\alpha=[\alpha_p,\alpha_v]\otimes I_m$。为了实现对最优控制器和优化后 PIF 的近似，引入了执行器-评价器神经网络。

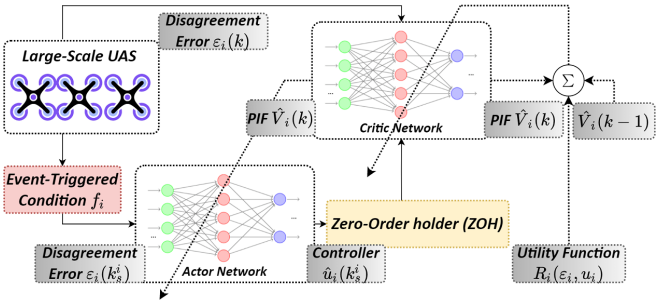


图 4. 用于实现事件触发在线执行器-评论家神经网络的框架，以实现编队跟踪控制。

1) 评论家神经网络设计： 评论家神经网络用于近似优化后的 PIF，

$$\frac{\partial\hat{V}_i(\varepsilon_{vi}(k+1))}{\partial\varepsilon_{vi}(k+1)}=\frac{2\alpha\varepsilon_i(k)}{\beta\gamma_iT}+\frac{1}{\beta\gamma_iT}\hat{W}_{ci}^{\mathrm{T}}(k)\psi_i(\varepsilon_i), \tag{18}$$

其中，$\hat{}W_{ci}\in\mathbb{R}^{s\times m}$ 是W_i^* 的估计值，通过评论家神经网络更新律进行训练

$$\hat{W}_{ci}(k+1)=\hat{W}_{ci}(k)-\mu_{ci}T\psi_i(\varepsilon_i)\psi_i^{\mathrm{T}}(\varepsilon_i)\hat{W}_{ci}(k), \tag{19}$$

其中，μ_{ci} 是评论家神经网络的常数增益。

2) 行动者神经网络设计：用于近似编队跟踪控制器（17）的行动者神经网络，

$$u_i^*(k)=\begin{cases}-\alpha\varepsilon_i(k)-\dfrac{1}{2}\hat{W}_{ai}^{\mathrm{T}}(k)\psi_i(\varepsilon_i(k)), & k\in\{k_l^i\}_{l\in\mathbb{Z}^+},\\ u_i^*(k_l^i), & k\in[k_l^i,k_{l+1}^i).\end{cases} \tag{20}$$

其中，$\hat{}W_{ai}\in\mathbb{R}^{s\times m}$ 是W_{ai}^*的估计值，由行动者神经网络更新律进行训练

$$\hat{W}_{ai}(k+1)=\begin{cases}\hat{W}_{ai}(k)-\mu_{ai}T\psi_i(\varepsilon_i)\psi_i^{\mathrm{T}}(\varepsilon_i)\\ \quad\times\left(\hat{W}_{ai}(k)-\hat{W}_{ci}(k)\right), & k\in\{k_l^i\}_{l\in\mathbb{Z}^+},\\ \hat{W}_{ai}(k), & k\in[k_l^i,k_{l+1}^i).\end{cases} \tag{21}$$

其中，μ_{ai} 是行动者神经网络的常数增益。

定义1。如果存在正常数 δ 和 ρ，$\forall x(k_0)<\delta$、$\exists T(\rho,\delta)\ge0$使得.....则称系统是一致最终有界（*UUB*）的

$$\forall k>k_0+T,\qquad ||x(k)||<\rho.$$

定理2。对于大规模无人机系统(3)中的任意无人机 a_i，若其初始状态有界，则采用优化后的事件触发编队跟踪控制器(20)，其中演员-评论家神经网络的迭代更新律为(19)和(21)，且所涉及的常数增益满足以下条件，

$$\mu_{ci}>\mu_{ai}>0,\quad 0<T<\frac{\mu_{ci}-\mu_{ai}}{\mu_{ci}^2\lambda_{\max}^{\psi_i\psi_i^{\mathrm{T}}}}, \tag{22}$$

其中，$\lambda^{\psi_i\psi}Timax$ 是 $\psi_i(k)\psi_i^{\mathrm{T}}(k)$ 的最大特征值。于是，分歧误差 $\varepsilon_i(k)$ 以及演员-评论家神经网络权重的估计误差$\tilde{}W_{ai}(k)$、$\tilde{}W_{ci}(k)$ 均一致最终有界，其中$\tilde{}W_{ai}(k)=\hat{}W_{ai}(k)-W_{ai}^*$且 $\tilde{}W_{ci}(k)=\hat{}W_{ci}(k)-W_{ci}^*$。

证明。稳定性分析过程将根据触发时刻$k \in \{k_l^i\}_{l\in\mathbb{Z}^+}$和触发时间间隔$k \in [k_l^i,\ k_{l+1}^i)$分为两部分。

(a) 对于 $\forall k \in [k_l^i, k_{l+1}^i)$，满足触发条件$f_i(k) \le 0$时，李雅普诺夫函数设计如下，

$$\begin{aligned}\mathcal{V}(k) = &\sum_{i=1}^{N} \varepsilon_i^{\mathrm{T}}(k)\varepsilon_i(k) + \frac{1}{2}\sum_{i=1}^{N} \mathrm{Tr}\left\{\tilde{W}_{ai}^{\mathrm{T}}(k)\tilde{W}_{ai}(k)\right\} \\ &+ \frac{1}{2}\sum_{i=1}^{N} \mathrm{Tr}\left\{\tilde{W}_{ci}^{\mathrm{T}}(k)\tilde{W}_{ci}(k)\right\} > 0.\end{aligned} \tag{23}$$

对于 $\mathcal{V}(k)$ 的一阶差分，

$$\begin{aligned}\Delta\mathcal{V}(k) \le &-\sum_{i=1}^{N} \frac{\mu_{ai}T}{2} \mathrm{Tr}\left(\tilde{W}_{ai}^{\mathrm{T}}(k)\psi_i\psi_i^{\mathrm{T}}\tilde{W}_{ai}(k)\right) \\ &-\sum_{i=1}^{N} \frac{(\mu_{ci}-\mu_{ai})T}{2} \mathrm{Tr}\left(\tilde{W}_{ci}^{\mathrm{T}}(k)\psi_i\psi_i^{\mathrm{T}}\tilde{W}_{ci}(k)\right) \\ &+\sum_{i=1}^{N} \frac{\mu_{ci}T}{2} \mathrm{Tr}\left(W_{ci}^{*}\psi_i\psi_i^{\mathrm{T}}W_{ci}^{*}\right),\end{aligned} \tag{24}$$

其中 $\sum_{i=}^{N} \mu_{ci}T2\ \mathrm{Tr}\ (W^{*}{}_{ci}\psi_i\psi_i^{\mathrm{T}}W^{*}{}_{ci}) \le \tau,\ \tau > 0$。并且 $\varsigma = \min_{i\in[1,N]}\left\{\mu_{ai}\overline{T\,\lambda}^{\psi}{}_{i\psi_{Ti}}\max\ ,(\mu_{ci}-\mu_{ai})T\,\lambda^{\psi}{}_{i\psi_{Ti}}\max\right\}$。因此，有 $\Delta\mathcal{V}(k) \le -\varsigma\mathcal{V}_l(k) + \tau$。然后，经过$k$次迭代，可得

$$\begin{aligned}\mathcal{V}(k) \le &(1-\varsigma)^k\mathcal{V}_l(0) + \frac{\tau}{\varsigma}\left(1-(1-\varsigma)^k\right) \\ &-\mathcal{V}_l(0) + \sum_{i=1}^{N} \varepsilon_i^{\mathrm{T}}(0)\varepsilon_i(0).\end{aligned} \tag{25}$$

(b) 对于 $\forall k \in \{k_l^i\}_{l\in\mathbb{Z}^+}$，更新控制器 $u_i^*(k_l^i)$，然后将李雅普诺夫函数设计如下，

$$\begin{aligned}\mathcal{V}'(k) = &\sum_{i=1}^{N} \beta^k V_i(\varepsilon_i(k)) + \frac{1}{2}\sum_{i=1}^{N} \mathrm{Tr}\left\{\tilde{W}_{ai}^{\mathrm{T}}(k)\tilde{W}_{ai}(k)\right\} \\ &+ \frac{1}{2}\sum_{i=1}^{N} \mathrm{Tr}\left\{\tilde{W}_{ci}^{\mathrm{T}}(k)\tilde{W}_{ci}(k)\right\} > 0.\end{aligned} \tag{26}$$

与前一种情况类似，$\mathcal{V}'(k)$ 有界。对于这两种情况，系统的分歧误差 $\varepsilon(k)$以及所有估计误差 $\tilde{}W_{ai}(k)$、$\tilde{}W_{ci}(k)$ 都是一致最终有界的。 □

V. 混合现实实验

本节构建了一个混合现实实验平台，用于验证所提出的大规模无人机编队最优分配和最优控制算法。

A. 混合现实实验平台

为验证该算法的有效性，搭建了图5所示的混合现实大规模无人机实验平台，其中包括真实实验场景和虚拟实验场景。真实场景中仅包含受实验场地限制的若干实体无人机节点，其状态信息由运动捕捉系统 OptiTrack 获取，并通过路由器传输至编队控制器。相比之下，虚拟场景可包含数百个虚拟无人机节点，其动力学模型设计为二阶模型。编队控制器利用获取的所有节点状态信息计算各节点的控制输入，并将控制输入发送回各节点以进行编队控制。为直观呈现编队效果，实体节点的状态信息也会发送至虚拟场景进行实时显示。

本实验使用4架实体无人机和116架虚拟无人机，共计120架无人机，以实现大规模编队。系统中使用的拉普拉斯矩阵为

$$L = \begin{bmatrix} 4 & -1 & -1 & 0 & \cdots & 0 & -1 & -1 \\ -1 & 4 & -1 & -1 & 0 & \cdots & 0 & -1 \\ -1 & -1 & 4 & -1 & -1 & 0 & \cdots & 0 \\ \cdots & \ddots & \ddots & \ddots & \ddots & \ddots & \cdots & \cdots \\ -1 & -1 & 0 & \cdots & 0 & -1 & -1 & 4 \end{bmatrix} \in \mathbb{R}^{120\times 120}.$$

对角矩阵 $B = \mathrm{diag}(\mathrm{mod}(N,2))$。其他参数列于表 I。

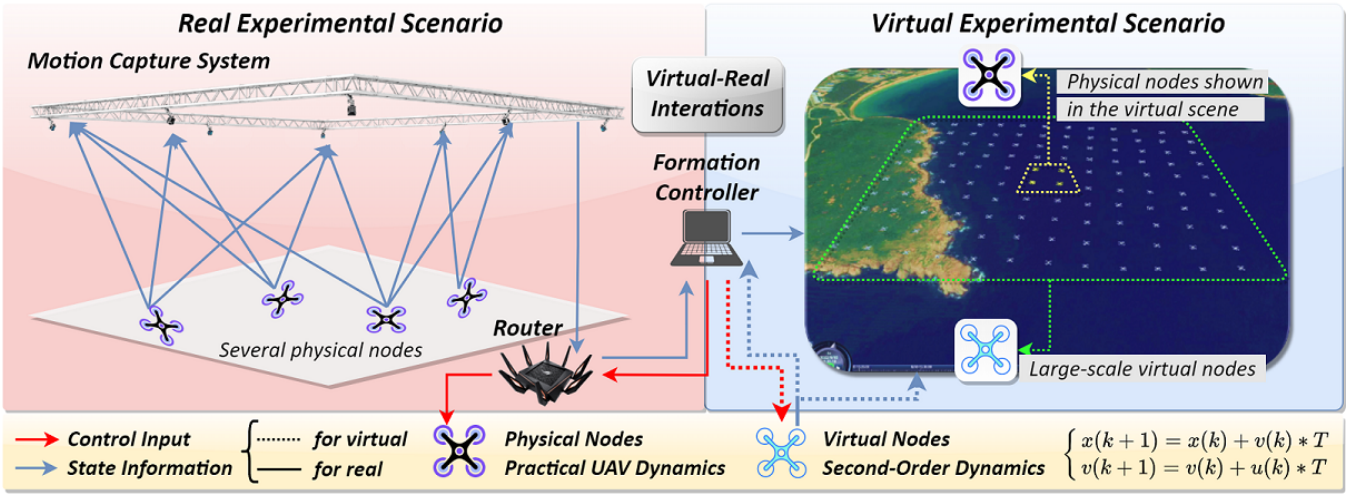


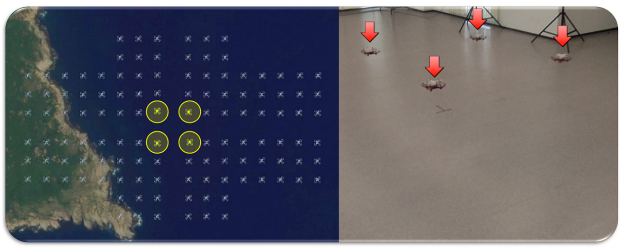
图 5. 混合现实实验平台框架。为实现大规模无人机切换编队控制，采用多个实体无人机节点和大规模虚拟无人机节点来实现混合现实交互。

表 I 实验设置

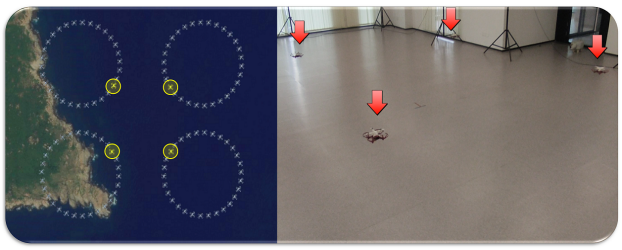
参数	数值	描述
r	0.14	无人机半径 (m)
s	60	径向基函数神经网络节点
ι_i	1	高斯基函数宽度
ν_i	[-3,3]	高斯基函数中心。
$\hat{W}_{ai}(0)$	$\{0.3\}_{60\times2}$	初始执行器神经网络权重
$\hat{W}_{ci}(0)$	$\{0.3\}_{60\times2}$	初始评价器神经网络权重
μ_{ai}	6	执行器神经网络增益
μ_{ci}	8	评价器神经网络增益
α	[6,4]	控制器中的参数$u(k)$
T	0.01	采样周期（秒）

B. 结果分析

如图 6 所示，大规模UAS使用最优指派算法，成功生成了多个编队之间无碰撞最优编队切换轨迹。通过优化后的编队跟踪控制器 (20)，大规模混合现实无人机节点能够无碰撞地跟随生成的期望轨迹。如图 7 所示，编队跟踪误差 $\xi(k)$ 有界。借助事件触发机制，每架无人机的控制器更新频率显著降低。如图 8 所示，与连续或定期更新控制器的时间触发算法相比，事件触发算法可将更新频率降至其 27%。图中还展示了最后 1000 次迭代中无人机 $1-30$ 的触发时刻序列，其中每架无人机的触发时刻均为非周期且按需触发。图 9 表明，执行器与评价器神经网络权重一致最终有界。最优编队跟踪控制问题得以解决。



(a) 十字形混合现实大规模编队



(b) 圆形混合现实大规模编队

图 6. 实验中的大规模无人机切换编队。

C. 总结

本研究构建的混合现实平台兼具物理实验平台在工程应用方面的优势，以及虚拟仿真平台的高灵活性和低成本。一方面，与完全虚拟仿真相比，该平台能够更准确地验证所提出的算法在实际无人机动力学和真实实验场景中的有效性。另一方面，与完全物理实验相比，该平台在实验成本和场地要求方面具有显著优势，能够借助轻量化实验设备实现大型无人机编队表演。

VI. 结论

针对大规模切换编队跟踪问题，本文引入了最优编队分配算法，并提出了一种最优控制器，从而打破了大规模无人机系统决策层和控制层面临的主要技术限制。此外，与时间触发控制器相比，所设计的事件触发算法将控制器和执行器神经网络的更新率降低至 27%。更值得一提的是，本文将实际实验的强工程适用性与虚拟平台的高度灵活性相结合，成功实现了大规模切换编队。然而，本研究忽略了实际场景中的空间约束以及虚拟无人机动力学的精确度。未来的工作将实现大规模切换编队中的自主避障，并进一步完善虚拟无人机动力学模型。

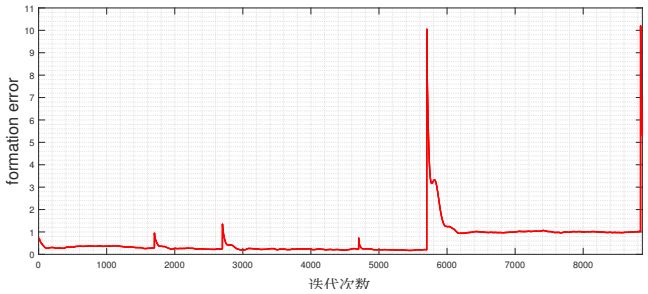


图 7. 编队跟踪误差的范数 $||\xi(k)||$。曲线中出现尖峰的时刻表示从一种编队切换到另一种编队。

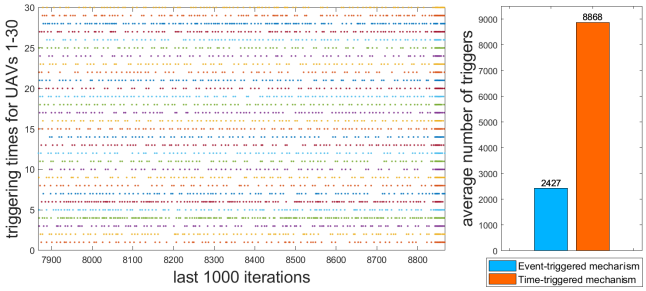


图 8. 事件触发机制与时间触发机制的触发时刻序列和平均触发次数。

参考文献

[1] H. Zhang, T. Feng, H. Liang 等，“基于LQR的线性离散时间多智能体系统最优分布式协同设计”，IEEE神经网络与学习系统汇刊，第28卷，第3期，第599–611页，2017年3月。

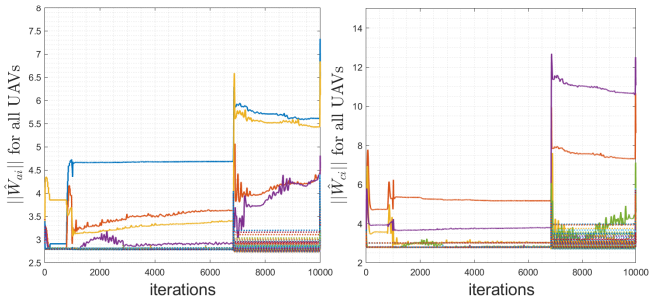


图9. 所有无人机的演员-评论家神经网络权重范数，其中实线表示实体无人机，点线表示虚拟无人机。

[2] C. O. Aguilar 和 A. J. Krener，“最优控制贝尔曼方程的数值解”，《优化理论与应用杂志》，第160卷，第2期，第527–552页，2014年2月。[3] B. Kiumarsi、F. L. Lewis、H. Modares 等，“强化学习在未知动力学线性离散时间系统最优跟踪控制中的应用”，《自动化》，第50卷，第4期，第1167–1175页，2014年4月。[4] H. J. Kappen，“最优控制理论与线性贝尔曼方程”，载于《贝叶斯时间序列模型》，第1版，D. Barber、A. T. Cemgil 和 S. Chiappa 编。剑桥大学出版社，2011年，第363–387页。[5] R. Kamalapurkar、J. A. Rosenfeld 和 W. E. Dixon，“用于近似在线最优控制的高效基于模型的强化学习”，《自动化》，第74卷，第247–258页，2016年12月。[6] B. Pang、Z. P. Jiang 和 I. Mareels，“强化学习在连续时间线性周期系统自适应最优控制中的应用”，《自动化》，第118卷，文章号109035，2020年8月。[7] A. Perrusquia 和 W. Yu，“利用循环神经网络和强化学习对非线性系统进行辨识与最优控制：综述”，《神经计算》，第438卷，第145–154页，2021年5月。[8] J. Zhang、Z. Wang 和 H. Zhang，“多智能体系统的基于数据的最优控制：一种强化学习设计方法”，《IEEE控制论汇刊》，第49卷，第12期，第4441–4449页，2019年12月。[9] H. Liu、Q. Meng、F. Peng 等，“通过强化学习实现输入受限领航者引导的多架无人机异构编队控制”，《神经计算》，第412卷，第63–71页，2020年10月。[10] W. Zhao、H. Liu 和 F. L. Lewis，“通过强化学习实现协作欠驱动四旋翼无人机的鲁棒编队控制”，IEEE神经网络与学习系统汇刊，第32卷，第10期，第4577–4587页，2021年10月。[11] Y. Guo、G. Chen 和 T. Zhao，“基于学习的多架不确定四旋翼无人机团队无碰撞协同”，《航空航天科学与技术》，第119卷，文章号107127，2021年12月。[12] H. Li、Y. Wu 和 M. Chen，“通过强化学习算法实现离散时间多智能体系统的自适应容错跟踪控制”，《IEEE控制论汇刊》，第51卷，第3期，第1163–1174页，2021年3月。

[13] M. Long、H. Su 和 Z. Zeng，“采用强化学习方法的多智能体系统无模型事件触发共识算法”，IEEE Trans. Syst. Man Cybern. Syst.，第52卷，第8期，第5212–5221页，2022年8月。[14] X. Guo、W. Yan 和 R. Cui，“完全未知连续时间非线性系统的事件触发强化学习自适应跟踪控制”，《IEEE控制论汇刊》，第50卷，第7期，第3231–3242页，2020年7月。[15] B. Yan、P. Shi 和 C. C. Lim，“基于自适应事件触发策略的非线性异构多智能体系统鲁棒编队控制”，IEEE Trans. Autom. Sci. Eng.，第1–13页，2021年8月。[16] V. Narayanan 和 S. Jagannathan，“采用探索的在线强化学习实现非线性交联系统的事件触发分布式控制”，《IEEE控制论汇刊》，第48卷，第9期，第2510–2519页，2018年9月。[17] X. Yang、H. He 和 D. Liu，“针对未知非线性系统的强化学习事件触发最优神经控制器设计”，IEEE Trans. Syst. Man Cybern. Syst.，第49卷，第9期，第1866–1878页，2019年9月。[18] X. Yang 和 H. He，“采用强化学习的某类非线性交联系统分散式事件触发控制”，《IEEE控制论汇刊》，第51卷，第2期，第635–648页，2021年2月。[19] H. Li、Y. Wu、M. Chen 等，“多智能体系统的自适应多梯度递归强化学习事件触发跟踪控制”，IEEE神经网络与学习系统汇刊，第1–13页，2021年7月。[20] Z. Peng、R. Luo、J. Hu 等，“通过事件触发强化学习实现离散时间多智能体系统的分布式最优跟踪控制”，IEEE Trans. Circuits Syst. Regul. Pap.，第3689–3700页，2022年6月。[21] W. Bai、T. Li、Y. Long 等，“多智能体系统的事件触发多梯度递归强化学习跟踪控制”，IEEE神经网络与学习系统汇刊，第1–14页，2021年7月。[22] J. Lu、Q. Wei、Y. Liu 等，“离散时间非线性系统的事件触发最优并行跟踪控制”，IEEE Trans. Syst. Man Cybern. Syst.，第52卷，第6期，第3772–3784页，2022年6月。[23] S. Zhao、J. Wang、H. Wang 等，“受输入约束的离散时间不确定系统的目标表示自适应评论家设计：事件触发情形”，《神经计算》，第492卷，第676–688页，2022年7月。[24] F. Tang、B. Niu、G. Zong 等，“通过强化学习实现非线性离散时间系统的周期性事件触发自适应跟踪控制设计”，Neural Netw.，第154卷，第43–55页，2022年10月。[25] S. Agarwal 和 S. Akella，“多机器人的任务分配与目标编队同步优化”，载于 IEEE 国际机器人与自动化会议（ICRA），布里斯班，昆士兰州，第6708–6715页，2018年5月。