技术交底书（实用新型或发明）

申请人		**所有发明人名字及第一发明人的身份证号**	
地址及邮编			
组织机构代码（企业）或个人身份证号码（个人）			
专利名称（建议为两个备选）/专利类型（发明或实用新型）/细分领域（可按技术部门或小分类填写）		一种基于分布式 A2C 的多无人机追逃博弈方法	

1. 产品使用领域

在复杂场景、非完整信息下的，多无人机围捕的追逃博弈问题。

2.目前市场上同类方案的缺陷

1) 当前无人机追逃博弈场景的设计存在较为简化的考虑，仅仅依靠追逐无人机和逃逸无人机位置的重合和拉开来判断追逐和逃逸的成功与否，无法充分体现追逐无人机之间的合作关系和策略协同。这种简化设计忽略了实际场景中的复杂性和多变性，导致模型无法全面捕捉追逃博弈过程中的各种协同行为。

2) 当前的多无人机追逃博弈方法存在一些缺陷，其中一个主要问题是所设计的无人机系统矩阵通常是已知的，并且没有考虑到外部干扰的存在。这些限制性因素限制了追逃博弈方法法的适用范围和求解精度，从而降低了追逃博弈的效果。

3) 目前基于强化学习的追逃博弈方法存在的主要困扰是如何克服过大方差的博弈性能指标函数高估问题对控制策略的不利影响。特别是在无人机系统存在系统矩阵不确定性和外部干扰的情况下，算法所使用的状态信息会受到过大方差的影响，从而导致无人机的控制决策不稳定，甚至可能出现无法获得有效控制策略的情况，进而阻碍多追逐无人机的围捕控制效率和逃逸无人机的逃离控制效率的提高。

4) 为了保证无人机之间能够有效地使用局部信息进行交流和合作，目前采用的多智能体强化学习算法面临着一些挑战，包括训练困难、难以收敛以及高计算开销的问题。此外，不同无人机之间的协作围捕与逃逸的竞争关系可能导致系统性能的不稳定性和不一致性。同时，这种竞争关系容易使系统陷入局部最优解或不稳定的状态，从而难以收敛到全局最优解，进而影响系统的性能和效果。

3.发明本方案的目的（解决了什么问题，有什么优点）：

1) 解决问题：解决多追逐无人机编队围捕逃逸无人机的追逃博弈场景中存在的问题，包括无人机系统矩阵未知和外部干扰的影响，以及过大方差的博弈性能指标函数高估问题对无人机控制决策的稳定性影响。

2) 涉及一种将多无人机追逃博弈扩展到多追逐无人机协作围捕逃逸无人机博弈的方法：本发明利用图论中的节点表示无人机，并建立相应的连接关系，基于每个无人机及其邻居的状态信息，计算局部误差变量，以评估当前状态与理想编队状态之间的差异。该方法使得追逐无人机能够在保持多边形编队的同时将逃逸无人机包围在编队中心，逃逸无人机能够在跟踪逃逸目标的同时远离多追逐无人机的编队中心，从而实现多追逐无人机之间协作围捕逃逸无人机的目标，增强了团队的合作能力和整体效率。

3) 涉及一种解决无人机系统存在系统矩阵未知和外部干扰问题的方法：本发明通过利用无人机及其邻居的局部状态信息来代替包含系统矩阵和外部干扰的相关信息，从而消除对无人机系统模型以及环境无扰动的依赖。通过这种方式，扩展了方法的适用范围和求解准确度，从而获得理想的多无人机追逃博弈效果。

4) 涉及一种提高多无人机追逃博弈的最优控制策略稳定性和收敛速度的方法：本发明将优势函数与决策-评判架构相结合，即采用 Advantage Actor-Critic（A2C）方法，并将其扩展至多无人机追逃博弈问题，即分布式 A2C 方法。通过采用分布式 A2C 方法，本发明能够克服由无人机系统存在系统矩阵未知和外部干扰所导致过大方差的博弈性能指标函数高估问题对控制策略平稳性的影响。优势函数用于评估当前控制策略的优越性并指导策略更新，而决策-评价架构则用于进行实时决策和策略的更新。通过不断优化控制策略，本发明能够提高多追逐无人机和逃逸无人机控制策略的优化速度和稳定性。

4.本方案的分成几个大的部件/部分，各部件/部分的专业名称叫什么？请附图说明（可编辑的 CAD 图或标准机械制图或者框图）； 对于纯方法类发明：请列出各方法步骤，并着重描述改进点所涉及的核心步骤；配方类的详细说明配方及其组分。

为了实现多无人机保持编队围捕逃逸无人机的追逃博弈对抗，本发明提出了一种基于分布式 A2C 的多无人机追逃博弈方法，具体步骤如下：

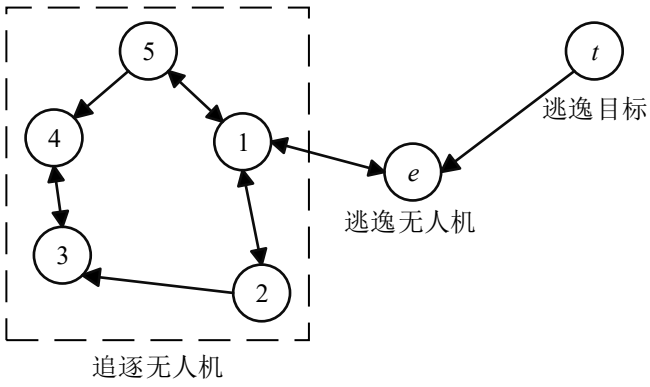

图 1 追逃博弈拓扑关系图

步骤 1：通过建立如图 1 所示包含追逐无人机i，逃逸无人机e，以及逃逸目标t的图论G，构建多无人机追逃博弈模型。多追逐无人机系统的交互为一个固定的强连接子图$G_p=(V_p,E_p)$，定义追逐无人机节点$V_p=\{n_1,...,n_N\}$为有限集，N为追逐者无人机的个数，追逐无人机之间的信息交互连接表示为$E_p\subseteq V_p\times V_p$。节点$n_i$的邻居集被定义为具有进入$n_i$的边的节点集，并表示为$N_i=\{n_j:(n_j,n_i)\in E_p\}$。图论的连通性矩阵被定义$N\times N$矩阵为$A_p=[\alpha_{ij}]_{N\times N}$，其中$\alpha_{ij}$表示为

$$\alpha_{ij}=\begin{cases}1, & \text{if } (n_j,n_i)\in E_p\\ 0, & \text{otherwise}\end{cases} \tag{1}$$

其中α_{ij}是与边$(n_j,n_i)\in E_p$相关联的权重，图G_p的度矩阵D是一个$N\times N$矩阵，其主对角线的第i个条目是加权度$d_i=\sum_{j\in N_i}\alpha_{ij}$，追逐无人机$n_i$的邻居数量由$|N_i|$表示，其等于$d_i$。

逃跑无人机的节点表示为n_e，邻居集由$N_e=\{n_1,n_t\}$表示。其中，g_{ei}和g_{et}分别是与边$(n_i,n_e),i\in N_p:=\{1,...,N\}$和$(n_t,n_e)$相关联的权重，当且仅当$g_{e1}=g_{et}=1$，其余为$0$。显然逃逸目标$t$的节点表示为$n_t$，节点$n_t$没有邻居集。

步骤 2：追逐无人机i和逃逸无人机e以及其邻居无人机的系统状态被用来构造局部误差变量，其中考虑了位置偏移。考虑由N个追逐无人机组成的互联系统G_p，追逐无人机i被建模为

$$\dot{x}_i=A_ix_i+B_iu_i+d_i(x_i,t) \tag{2}$$

其中$x_i\in\mathbb{R}^n,i\in N_p$为追逐无人机$i$的位置向量，$u_i\in\mathbb{R}^n,i\in N_p$是追逐无人机$i$的控制输入向量；$A_i\in\mathbb{R}^{n\times n}$是追逐无人机$i$的未知系统矩阵，$B_i\in\mathbb{R}^{n\times n},i\in N_p$是已知追逐无人机$i$的输入增益矩阵，$d_i(x_i,t)$为状态相关的未知非线性扰动。图论$G$中的多追逐无人机去围捕逃逸无人机$e$，逃逸无人机$e$系统被定义为

$$\dot{x}_e=A_ex_e+B_eu_e+d_e(x_e,t) \tag{3}$$

其中$x_e\in\mathbb{R}^n$和分别是逃逸无人机e的位置和控制输入向量，$A_e\in\mathbb{R}^{n\times n}$和$B_e\in\mathbb{R}^{n\times n}$分别是逃逸无人

机 e 的未知系统矩阵和已知输入增益矩阵，$d_e(x_e,t)$ 状态相关的未知非线性扰动。图论 G 中逃逸无人机 e 寻求在逃逸的过程中渐近跟踪具有动力学 $\dot{x}_t = Ax_t$ 的移动目标 t 系统状态，即状态 x_e 收敛至 x_t 。

为了保证多追逐无人机以正 N 边形编队围捕逃逸无人机 e，追逐无人机 i 相对于其邻居无人机和逃逸无人机 e 的局部误差变量 δ_i 被定义为

$$\delta_i = g_{ei}(x_i - x_e + \Delta_{ei}) + \sum_{j\in N_i} \alpha_{ij}(x_i - x_j + \Delta_{ij}) \tag{4}$$

其中 $\Delta_{ei} \in \mathbb{R}^n$ 为逃逸无人机和追逐无人机 i 的位置偏移量，其作用是确保逃逸无人机在多追逐无人机编队的中心，$\Delta_{ij} \in \mathbb{R}^n$ 为追逐无人机 i 和追逐无人机 j 的位置偏移量，其作用是确保多追逐无人机编队维持正 N 边形编队。引入逃逸无人机和追逐无人机之间的位置偏移量对于多追逐无人机编队围捕逃逸无人机任务至关重要。这些偏移量的优化和控制可以提高编队的协同行动能力，保持编队形状的稳定性，并最大限度地提高围捕任务的成功率。

使用动力学系统(2)-(3),可以从局部误差变量 δ_i 推导出

$$\dot{\delta}_i = A\delta_i + (d_i + g_{ei})B_i u_i - g_{ei}(u_e + \Delta_{ei}) - \sum_{j\in N_i} \alpha_{ij}(B_j u_j + A\Delta_{ij}) + \Delta_i(x_i,t) \tag{5}$$

其中 $\delta_i \in \mathbb{R}^n$， $\Delta_i(x_i,t)$ 为追逐无人机 i 的非线性扰动相关项。

同样，逃逸无人机 e 跟踪目标位置的同时，躲避多追逐无人机编队围捕。为此，逃逸无人机的局部误差变量 δ_e 被表示为

$$\delta_e = \kappa g_{et}(x_e - x_t) - g_{e1}(x_e - x_1 - \Delta_{e1}) \tag{6}$$

其中 $\kappa > 0$ 为标量增益，通过使用动力学系统(3)和逃逸目标动力学系统,局部误差变量 δ_e 可以被推导出

$$\dot{\delta}_e = A\delta_e + (\kappa g_{et} - g_{e1})B_e u_e + g_{e1}(Bu_1 - A\Delta_{e1}) + \Delta_e(x_e,t) \tag{7}$$

其中 $\delta_e \in \mathbb{R}^n$， $\Delta_e(x_e,t)$ 为逃逸无人机 e 的非线性扰动相关项。

步骤 3：构建多无人机追逃博弈性能指标函数和最优控制策略，以实现高效的追逃任务。逃逸无人机的目标是以最小的能量消耗尽可能的远离多追逐无人机编队，并同时追踪逃逸目标。与此同时，追逐无人机希望以最小的能量消耗保持正多边形编队形状，并紧密协作以拦截逃逸无人机，同时与队友之间保持适当的距离。为此，需要定义博弈性能指标函数，用于衡量多追逐无人机编队与逃逸无人机之间的竞争关系。根据步骤二中局部误差变量 δ_i，追逐无人机 i 的博弈性能指标函数可以被定义为

$$J_i = \int_0^\infty \frac{1}{2}\left(\delta_i^T Q_i \delta_i + u_i^T R_{ii} u_i + \sum_{j\in N_i} u_j^T R_{ij} u_j\right) d_t \tag{8}$$

同样，根据步骤二中局部误差变量 δ_e，逃逸无人机 e 的博弈性能指标函数可以被定义为

$$J_e = \int_0^\infty \frac{1}{2}\left(\delta_e^T Q_e \delta_e + u_e^T R_e u_e\right) d_t \tag{9}$$

其中，$Q_i > 0, Q_e > 0, R_{ii} > 0, R_{ij} > 0, R_e > 0$，均为适当维数的自定义矩阵。

利用上述定义的博弈性能指标函数，多无人机追逃博弈问题可以被表示为以下形式的耦合分布最小化问题：

$$\begin{aligned} V_i^*(\delta_i) &= \min_{u_i} J_i(\delta_i(0), u_i, u_{N_i}, u_e) \\ V_e^*(\delta_e) &= \min_{u_e} J_e(\delta_e(0), u_e, u_1) \end{aligned} \tag{10}$$

其中V_i^*和V_e^*分别是追逐无人机i和逃逸无人机e的最优代价函数，$\min_a\{\cdot\}$表示关于a的最小元素。

由于图论G中连接的无人机的知识有限，因此有必要考虑让无人机为其邻居行为中的最坏情况做好准备。所得解的概念被视为最小-最大策略，具体而言，追逐无人机i和逃逸无人机e的最大最小化策略被定义为以下形式：

$$\begin{aligned} u_i^* &= \arg\min_{u_i} \max_{u_{N_i}, u_e} J_i \\ u_e^* &= \arg\min_{u_e} \max_{u_1} J_e \end{aligned} \tag{11}$$

其中$\max_b\{\cdot\}$表示关于b的最大元素$\arg\min_a\{\cdot\}$表示关于a的最小元素对应变量的取值。

通过利用每个追逐无人机i的局部误差变量δ_i和博弈性能指标函数J_i，追逐无人机i的哈密顿量被定义为

$$\begin{aligned} H_i = \nabla V_i^* \Big[& A\delta_i + (d_i + g_{ei}) B_i u_i - g_{ei}(u_e + \Delta_{ei}) - \textstyle\sum_{j \in N_i} \alpha_{ij}(B_j u_j + A\Delta_{ij}) + \Delta_i(x_i, t) \Big] \\ & + \delta_i^T Q_i \delta_i + u_i^T R_{ii} u_i + \textstyle\sum_{j \in N_i} u_j^T R_{ij} u_j \end{aligned} \tag{12}$$

其中$\nabla V_i^* = \partial V_i^* / \partial \delta_i$表示为追逐无人机$i$的最优代价函数$V_i^*$对局部误差变量$\delta_i$的偏导。

采用哈密顿量中的平稳性条件-哈密顿量H_i对控制输入u_i的偏导等于零，即$\partial H_i / \partial u_i = 0$，追逐无人机$i$的最优控制输入可以表示为

$$u_i^*(\delta_i) = -\frac{1}{2}(d_i + g_{ei}) R_{ii}^{-1} B_i^T \nabla V_i^* \tag{13}$$

同样，逃逸无人机e的哈密顿量和最优控制输入可以表示为如下形式：

$$\begin{aligned} H_e = \nabla V_e^* \big[& A\delta_e + (\kappa g_{et} - g_{e1}) B_e u_e + g_{e1}(Bu_1 - A\Delta_{e1}) + \Delta_e(x_e, t) \big] \\ & + \delta_e^T Q_e \delta_e + u_e^T R_e u_e \\ u_e^*(\delta_e) = -\frac{1}{2} & (\kappa g_{et} - g_{e1}) R_e^{-1} B_e^T \nabla V_e^* \end{aligned} \tag{14}$$

其中$\nabla V_e^* = \partial V_e^* / \partial \delta_e$表示为逃逸无人机$e$的最优代价函数$V_e^*$对局部误差变量$\delta_e$的偏导。

步骤 4：在步骤三中，追逐无人机i和逃逸无人机e的最优控制输入，即最优控制策略，被认为是多无人机追逃博弈问题的最优解。然而，追逐无人机和逃逸无人机的最优代价函数和其梯度是未知的，因此实际上无法使用最优控制策略。为了解决这个问题，设计出分布式 A2C 方法去获取追逃控制策略，该方法被划分为以下步骤 4.1 和 4.2 进行设计。

步骤 4.1：由于追逐无人机i和逃逸无人机的最优代价函数V_i^*和V_e^*包含了无法使用的未来信息，为了解决这个问题，近似最优代价函数的评判网络及其对局部误差的偏导可以被构造为

$$\begin{aligned} \hat{V}_i(\delta_i) &= \hat{W}_{ic}^T \phi_i(\delta_i), \frac{\partial \hat{V}_i}{\partial \delta_i} = \frac{\partial \phi_i^T(\delta_i)}{\partial \delta_i} \hat{W}_{ic} \\ \hat{V}_e(\delta_e) &= \hat{W}_{ec}^T \phi_e(\delta_e), \frac{\partial \hat{V}_e}{\partial \delta_e} = \frac{\partial \phi_e^T(\delta_e)}{\partial \delta_e} \hat{W}_{ec} \end{aligned} \tag{15}$$

其中$\phi_e(\delta_e): \mathbb{R}^n \to \mathbb{R}^N$和$\phi_i(\delta_i): \mathbb{R}^n \to \mathbb{R}^N$均为基函数集向量，$\hat{W}_{ic}$和$\hat{W}_{ec}$分别为追逐无人机$i$和逃逸无人机的最优代价函数$V_i^*$和$V_e^*$的当前权重估计。

鉴于追逐无人机i和逃逸无人机e最优代价函数V_i^*和V_e^*的梯度未知，追逐无人机i和逃逸无人机e所支配的理想最优控制策略实际上不可用。因此，决策网络被设计，以近似最优控制策略，同时实现评判网络和决策网络之间的信息交互更新。追逐无人机i和逃逸无人机e的决策网络的当前权重估计$\hat{W}_{ia}$和$\hat{W}_{ea}$，并使用近似的追逐无人机i和逃逸无人机控制输入表示为

$$\begin{aligned} \hat{u}_i(\delta_i) &= -\frac{1}{2}(d_i + g_{ei}) R_{ii}^{-1} B_i^T \frac{\partial \phi_i^T(\delta_i)}{\partial \delta_i} \hat{W}_{ia} \\ \hat{u}_e(\delta_e) &= -\frac{1}{2}(\kappa g_{et} - g_{e1}) R_e^{-1} B_e^T \frac{\partial \phi_e^T(\delta_e)}{\partial \delta_e} \hat{W}_{ea} \end{aligned} \tag{16}$$

此外，Hamilton-Jacobi-Bellman (HJB)方程的误差表示为

$$\begin{aligned} \varepsilon_i^{HJB} &= \delta_i^T Q_i \delta_i + \hat{u}_i^T R_{ii} \hat{u}_i + \sum_{j \in N_i} \hat{u}_j^T R_{ij} \hat{u}_j + \hat{W}_{ic}^T \chi_i^* \\ \varepsilon_e^{HJB} &= \delta_e^T Q_e \delta_e + \hat{u}_e^T R_e \hat{u}_e + \hat{W}_{ic}^T \chi_e^* \end{aligned} \tag{17}$$

其中追逐无人机i和逃逸无人机e的学习回归变量χ_i和χ_e被分别定义为

$$\begin{aligned} \chi_i^* &= \nabla \phi_i^T \left(A\delta_i + (d_i + g_{ei}) B_i \hat{u}_i - g_{ei}(\hat{u}_e + \Delta_{ei}) - \sum\nolimits_{j \in N_i} \alpha_{ij}(B_j \hat{u}_j + A\Delta_{ij}) + \Delta_i(x_i, t) \right) \\ \chi_e^* &= \partial \phi_e^T(\delta_e) / \partial \delta_e \left(\delta_e + (\kappa g_{et} - g_{e1}) B_e \hat{u}_e + g_{e1}(B\hat{u}_1 - A\Delta_{e1}) + \Delta_i(x_i, t) \right) \end{aligned} \tag{18}$$

其中$\nabla \phi_i^T = \partial \phi_i^T(\delta_i) / \partial \delta_i$表示为追逐无人机$i$的基函数集向量$\phi_i(\delta_i)$对局部误差变量$\delta_i$偏导的转置，$\nabla \phi_e^T = \partial \phi_e^T(\delta_e) / \partial \delta_e$表示为逃逸无人机$e$的基函数集向量$\phi_e(\delta_e)$对局部误差变量$\delta_e$偏导的转置。

由于 χ_i^* 和 χ_e^* 包含未知系统矩阵和非线性扰动相关信息，即 $A_i,\Delta_i(x_i,t)$ 和 $A_e,\Delta_e(x_e,t)$。为了解决这个难题，在分布式 A2C 的在线学习过程中，引入追逐无人机 i 和逃逸无人机 e 的在线学习回归变量为 $\chi_i=\Delta\phi_i(\delta_i)/T,\chi_e=\Delta\phi_e(\delta_e)/T$ 去替代学习回归变量 χ_i^* 和 χ_e^*，其中 T 为任意时间间隔，$\Delta\phi_i(\delta_i)=\phi_i(\delta_i(t))-\phi_i(\delta_i(t-T))$ 以及 $\Delta\phi_e(\delta_e)=\phi_e(\delta_e(t))-\phi_e(\delta_e(t-T))$。为了通过分别调整临界神经网络权重 $\hat{W}_{ic}$ 和 $\hat{W}_{ec}$ 来确定实现最小 ε_i^{HJB} 和 ε_e^{HJB} 的指标函数，目标函数被考虑为 $E_{ic}=1/2(\varepsilon_i^{HJB})^2$ 和 $E_{ec}=1/2(\varepsilon_e^{HJB})^2$，为了能够对临界权重进行细化和调整以实现其最佳估计，当前权重 $\hat{W}_{ic}$ 和 $\hat{W}_{ec}$ 的更新律设计为

$$\begin{aligned}\dot{\hat{W}}_{ic}&=-\Gamma_{ic}\frac{\chi_i}{(\chi_i^T\chi_i+1)^2}\left(\delta_i^TQ_i\delta_i+\hat{u}_i^TR_{ii}\hat{u}_i+\sum\nolimits_{j\in N_i}\hat{u}_j^TR_{ij}\hat{u}_j+\hat{W}_{ic}^T\chi_i\right)\\\dot{\hat{W}}_{ec}&=-\Gamma_{ec}\frac{\chi_e}{(\chi_e^T\chi_e+1)^2}\left(\delta_e^TQ_e\delta_e+\hat{u}_e^TR_e\hat{u}_e+\hat{W}_{ic}^T\chi_e\right)\end{aligned}\tag{19}$$

其中 $\Gamma_{ic}>0$ 和 $\Gamma_{ec}>0$ 分别为追逐无人机 i 和逃跑无人机 e 适当维数的评判网络学习率。

步骤 4.2：考虑到在自学习过程中，外部扰动和未知系统矩阵的不确定性导致过大方差的博弈性能指标函数高估问题，从而在控制策略中引入不稳定性，阻碍追逐无人机 i 和逃跑无人机 e 的控制效率，甚至导致多无人机追逃博弈出现的求解不稳定甚至无解的情况。为此，结合优势函数的分布式 A2C 方法，旨在减轻方差带来的不利影响。具体而言，追逐无人机 i 和逃跑无人机 e 的优势函数设计如下

$$\begin{aligned}\delta_t^i&=\int_{t-T}^{t}\left(\delta_i^TQ_i\delta_i+\hat{u}_i^TR_{ii}\hat{u}_i+\sum_{j\in N_i}\hat{u}_j^TR_{ij}\hat{u}_j\right)ds+\hat{W}_{ic}^T\Delta\phi_i(\delta_i)\\\delta_e^i&=\int_{t-T}^{t}\left(\delta_e^TQ_e\delta_e+\hat{u}_e^TR_e\hat{u}_e\right)ds+\hat{W}_{ec}^T\Delta\phi_e(\delta_e)\end{aligned}\tag{20}$$

优势函数在优化追逐无人机的包围效率方面起着关键作用，它不仅能够提供更优的决策，还能够排除劣质的决策。其工作原理如下：$\delta_t^i<0$ 表示当前控制策略利于最小化博弈性能指标，从而促使追逐无人机 i 的决策网络更新；$\delta_t^i>0$ 表示当前控制策略不利于最小化性能指标，从而促使追逐无人机 i 的决策网络反向更新；$\delta_t^i=0$ 表示当前控制策略处于最优状态，追逐无人机 i 的决策网络不会更新。其优势函数对于逃逸无人机 e 作用机制也是相同的，能够优化逃逸无人机 e 的逃跑效率。

通过这样的机制，优势函数能够指导追逐无人机 i 和逃逸无人机 e 的决策网络调整控制策略，以获得最优的追逃博弈控制效率。为此，追逐无人机 i 和逃逸无人机 e 当前权重 $\hat{W}_{ia}$ 和 $\hat{W}_{ea}$ 的更新律设计为

$$\dot{\hat{W}}_{ia}=-\Gamma_{ia}\left[k_{ia}+\frac{\hat{u}_i^T R_{ii}\hat{u}_i}{4(\chi_i^T\chi_i+1)^2}\right]\hat{W}_{ia}+\Gamma_{ia}\frac{(d_i+g_{ei})\nabla\phi_i^T(\delta_i)B_i\delta_t^i\hat{u}_i}{2T(\chi_i^T\chi_i+1)^2}$$
$$-\Gamma_{ia}\frac{(d_i+g_{ei})\nabla\phi_i^T(\delta_i)B_i\hat{u}_i\hat{u}_j^T R_{ij}\hat{u}_j}{2(\chi_i^T\chi_i+1)^2} \tag{21}$$
$$\dot{\hat{W}}_{ea}=-\Gamma_{ea}\left[k_{ea}+\frac{\hat{u}_e^T R_e\hat{u}_e}{4(\chi_e^T\chi_e+1)^2}\right]\hat{W}_{ea}+\Gamma_{ea}\frac{(\kappa g_{et}-g_{e1})\nabla\phi_e^T(\delta_e)B_e\delta_t^e\hat{u}_e}{2T(\chi_e^T\chi_e+1)^2}$$

其中 $\Gamma_{ia}>0$ 和 $\Gamma_{ea}>0$ 分别为追逐无人机 i 和逃跑无人机 e 具有适当维数的决策网络学习率，此外 $k_{ia}>0$ 和 $k_{ea}>0$ 为具有适当维数的反馈增益。

步骤 5：根据步骤 4 中的基于分布 A2C 方法的多无人机追逃博弈控制方法，其算法训练过程如下所述：

首先，对于多无人机追逃博弈控制方法，算法训练将按照以下步骤进行：初始化追逐无人机 i 和逃逸无人机 e 位置 $x_i(0),x_e(0)$，控制策略 $\hat{u}_i(0),\hat{u}_e(0)$；初始化逃逸目标的位置 $x_t(0)$。接下来，计算 $e_{ic}=\left\|\hat{V}_i^{u_i(k)}(\delta_i)-\hat{V}_i^{u_i(k-1)}(\delta_i)\right\|$ 和 $e_{ec}=\left\|\hat{V}_e^{u_e(k)}(\delta_e)-\hat{V}_e^{u_e(k-1)}(\delta_e)\right\|$，并判断是否满足同时满足 $e_{ic}\le\xi_{ic}$ 和 $e_{ec}\le\xi_{ec}$，其中，$\xi_{ic}>0$ 和 $\xi_{ec}>0$ 是自定义的较小常数。如果满足条件，则当前追逐无人机 i 和逃逸无人机 e 的控制策略为最优控制策略，反之根据步骤 4.2 中的当前权重 $\hat{W}_{ic},\hat{W}_{ec},\hat{W}_{ia}$ 和 $\hat{W}_{ea}$ 的更新律进行学习训练，直至满足上述条件为止。

基于协作 A2C 方法的多无人机追逃博弈控制方法的算法流程图如图 3 所示。

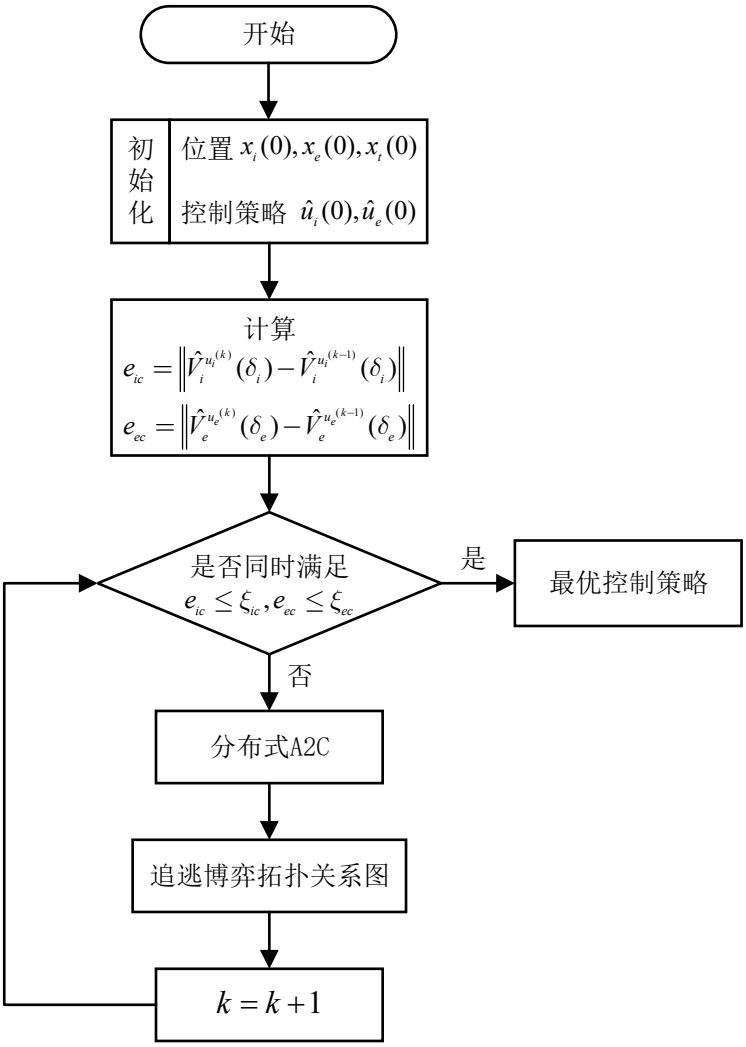

图 3 基于分布式 A2C 的多无人机追逃博弈的算法流程图

5.请说明上面各个部件/部分之间的连接/逻辑关系和工作原理，请着重从结构上来说明，而不能只强调功能；对于纯方法类发明：请阐述改进点所涉及的核心步骤所基于的原理，即它们为什么能够产生改善；配方类的说明配方起作用的原理。

为了实现多无人机对逃逸无人机进行编队围捕的追逃博弈对抗，本发明首先利用图论构建多无人机追逃博弈模型。然后，根据每个追逐无人机（逃逸无人机）及其邻居无人机的系统状态，构造带有位置偏移的局部误差变量，以确保多追逐无人机能够维持围捕编队并使逃逸无人机远离编队中心，与此同时，逃逸无人机能够在跟踪逃逸目标的同时远离多追逐无人机的编队中心。然后，利用局部误差变量定义博弈性能指标函数，将问题转化为追逃博弈的纳什均衡问题进行求解。最后，设计分布式 A2C 算法来求解纳什均衡，并得到多追逐无人机和逃逸无人机的最优控制策略，从而实现多无人机对逃逸无人机的编队围捕。图 2 展示了各模块之间的关系。

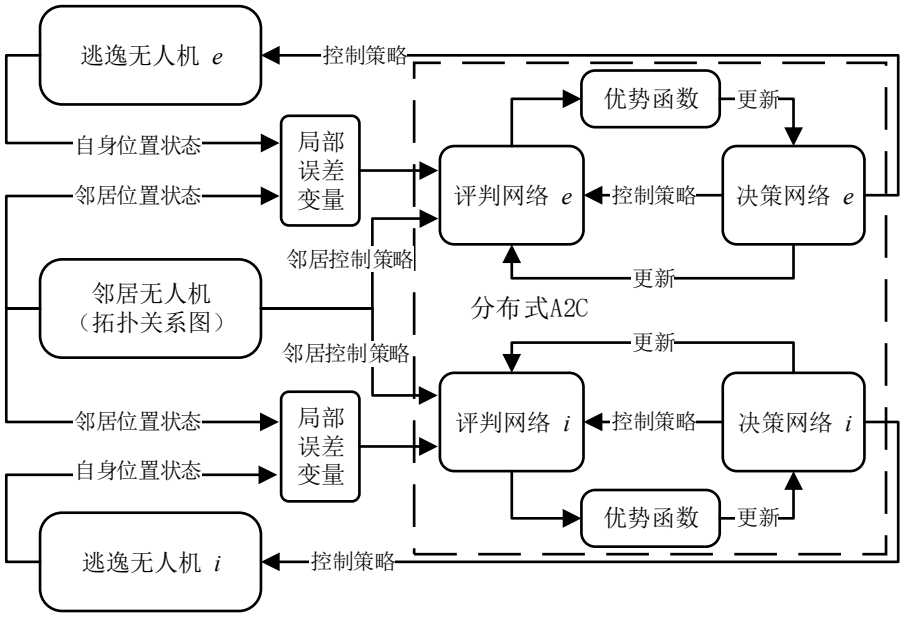

图 2 模块关系图

6.请说明该方案的优点，请从结构和功能上都说明。如该产品的优点有多条，请分别列出，结构改进上的优点请用图和文字一起描述出来。

改进点 1：将多无人机追逃博弈扩展到协作围捕逃逸博弈

为了保证多追逐无人机以正 N 边形编队围捕逃逸无人机 e，追逐无人机 i 相对于其邻居无人机和逃逸无人机 e 的局部误差变量 δ_i 被定义为

$$\delta_i = g_{ei}(x_i - x_e + \Delta_{ei}) + \sum_{j\in N_i} \alpha_{ij}(x_i - x_j + \Delta_{ij}) \tag{22}$$

其中 $\Delta_{ei} \in \mathbb{R}^n$ 为逃逸无人机和追逐无人机 i 的位置偏移量，其作用是确保逃逸无人机在多追逐无人机编队的中心，$\Delta_{ij} \in \mathbb{R}^n$ 为追逐无人机 i 和追逐无人机 j 的位置偏移量，其作用是确保多追逐无人机编队维持正 N 边形编队。引入逃逸无人机和追逐无人机之间的位置偏移量对于多追逐无人机编队围捕逃逸无人机任务至关重要。这些偏移量的优化和控制可以提高编队的协同行动能力，保持编队形状的稳定性，并最大限度地提高围捕任务的成功率。

同样，逃逸无人机 e 跟踪目标位置的同时，躲避多追逐无人机编队围捕。为此，逃逸无人机 e 的局部误差变量 δ_e 被表示为

$$\delta_e = \kappa g_{et}(x_e - x_t) - g_{e1}(x_e - x_1 - \Delta_{e1}) \tag{23}$$

其中 $\kappa > 0$ 是适当维数的增益。

多追逐无人机协作围捕逃逸无人机博弈问题的建模，改进了现有技术仅考虑简单多人机追逃博弈场景问题，例如，依靠追逐者和逃逸者位置的重合和拉开来判断追逐和逃逸的成功与否。基于每个无人机及其邻居的状态信息，计算局部误差变量，以评估当前状态与理想编队状态之间的差异。使得追逐无人

机能够在保持多边形编队的同时将逃逸无人机包围在编队中心，逃逸无人机能够在跟踪逃逸目标的同时远离多追逐无人机的编队中心，从而实现多追逐无人机之间协作围捕逃逸无人机的目标，增强了团队的合作能力和整体效率。

改进点 2：考虑无人机系统存在系统矩阵未知和外部干扰

追逐无人机 i 和逃逸无人机 e 以及其邻居无人机的系统状态被用来构造局部误差变量，其中考虑了位置偏移。考虑由 N 个追逐无人机组成的互联系统 G_p，追逐无人机 i 被建模为

$$\dot{x}_i = A_i x_i + B_i u_i + d_i(x_i,t) \tag{24}$$

其中 $x_i \in \mathbb{R}^n, i \in N_p$ 为追逐无人机 i 的位置向量，$u_i \in \mathbb{R}^n, i \in N_p$ 是追逐无人机 i 的控制输入向量；$A_i \in \mathbb{R}^{n\times n}$ 是未知具有适当维度的追逐无人机 i 系统矩阵，$B_i \in \mathbb{R}^{n\times n}, i \in N_p$ 已知的具有适当维度的追逐无人机 i 输入增益矩阵，$d_i(x_i,t)$ 为状态相关的非线性扰动。图论 G 中的多追逐无人机去围捕逃逸无人机 e，逃逸无人机 e 系统被定义为

$$\dot{x}_e = A_e x_e + B_e u_e + d_e(x_e,t) \tag{25}$$

其中 $x_e \in \mathbb{R}^n$ 和分别是逃逸无人机 e 的位置和控制输入向量，$A_e \in \mathbb{R}^{n\times n}$ 和 $B_e \in \mathbb{R}^{n\times n}$ 分别是逃逸无人机 e 的未知系统矩阵和已知输入增益矩阵，$d_e(x_e,t)$ 状态相关的非线性扰动。图论 G 中逃逸无人机 e 寻求在逃逸的过程中渐进跟踪具有动力学 $\dot{x}_t = Ax_t$ 的移动目标 t 系统状态，即状态 x_e 收敛至 x_t。

追逐无人机 i 和逃逸无人机 e 的学习回归变量 χ_i 和 χ_e 被分别定义为

$$\begin{aligned}
\chi_i^* &= \nabla\phi_i^T\left(A\delta_i + (d_i + g_{ei})B_i\hat{u}_i - g_{ei}(\hat{u}_e + \Delta_{ei}) - \sum\nolimits_{j\in N_i}\alpha_{ij}(B_j\hat{u}_j + A\Delta_{ij}) + \Delta_i(x_i,t)\right) \\
\chi_e^* &= \partial\phi_e^T(\delta_e)/\partial\delta_e\left(\delta_e + (\kappa g_{et} - g_{e1})B_e\hat{u}_e + g_{e1}(B\hat{u}_1 - A\Delta_{e1}) + \Delta_i(x_i,t)\right)
\end{aligned} \tag{26}$$

其中 $\nabla\phi_i^T = \partial\phi_i^T(\delta_i)/\partial\delta_i$ 表示为追逐无人机 i 的基函数集向量 $\phi_i(\delta_i)$ 对局部误差变量 δ_i 偏导的转置，$\nabla\phi_e^T = \partial\phi_e^T(\delta_e)/\partial\delta_e$ 表示为逃逸无人机 e 的基函数集向量 $\phi_e(\delta_e)$ 对局部误差变量 δ_e 偏导的转置。

由于 χ_i^* 和 χ_e^* 包含未知系统矩阵和非线性扰动相关信息，即 $A_i, \Delta_i(x_i,t)$ 和 $A_e, \Delta_e(x_e,t)$。为了解决这个难题，在分布式 A2C 的在线学习过程中，引入追逐无人机 i 和逃逸无人机 e 的在线学习回归变量为 $\chi_i = \Delta\phi_i(\delta_i)/T, \chi_e = \Delta\phi_e(\delta_e)/T$ 去替代学习回归变量 χ_i^* 和 χ_e^*，其中 T 为任意时间间隔，$\Delta\phi_i(\delta_i) = \phi_i(\delta_i(t)) - \phi_i(\delta_i(t-T))$ 以及 $\Delta\phi_e(\delta_e) = \phi_e(\delta_e(t)) - \phi_e(\delta_e(t-T))$。

利用无人机及其邻居的局部状态信息代替包含系统矩阵和外部干扰的相关信息，改进了现有技术仅考虑无人机系统矩阵已知和不受外部干扰的理想情况，消除对无人机系统模型以及环境无扰动的依赖，扩展了追逃博弈方法的适用范围和求解准确度，从而获得理想的多无人机追逃博弈效果。

改进点 3：结合优势函数的分布式 A2C 方法

考虑到在自学习过程中，外部扰动和未知系统矩阵的不确定性导致过大方差的博弈性能指标函数高估问题，从而在控制策略中引入不稳定性，阻碍追逐无人机 i 和逃跑无人机 e 的控制效率，甚至导致多无人机追逃博弈出现的求解不稳定甚至无解的情况。为此，结合优势函数的分布式 A2C 方法，旨在减轻方差带来的不利影响。具体而言，追逐无人机 i 和逃跑无人机 e 的优势函数设计如下

$$\begin{aligned}\delta_t^i &= \int_{t-T}^{t}\left(\delta_i^T Q_i \delta_i + \hat{u}_i^T R_{ii}\hat{u}_i + \sum_{j\in N_i}\hat{u}_j^T R_{ij}\hat{u}_j\right)ds + \hat{W}_{ic}^T \Delta\phi_i(\delta_i)\\ \delta_e^i &= \int_{t-T}^{t}\left(\delta_e^T Q_e \delta_e + \hat{u}_e^T R_e \hat{u}_e\right)ds + \hat{W}_{ec}^T \Delta\phi_e(\delta_e)\end{aligned} \tag{27}$$

优势函数在优化追逐无人机的包围效率方面起着关键作用，它不仅能够提供更优的决策，还能够排除劣质的决策。其工作原理如下：$\delta_t^i < 0$ 表示当前控制策略利于最小化博弈性能指标，从而促使追逐无人机 i 的决策网络更新；$\delta_t^i > 0$ 表示当前控制策略不利于最小化性能指标，从而促使追逐无人机 i 的决策网络反向更新；$\delta_t^i = 0$ 表示当前控制策略处于最优状态，追逐无人机 i 的决策网络不会更新。其优势函数对于逃逸无人机 e 作用机制也是相同的，能够优化逃逸无人机 e 的逃跑效率。

通过这样的机制，优势函数能够指导追逐无人机 i 和逃逸无人机 e 的决策网络调整控制策略，以获得最优的追逃博弈控制效率。为此，追逐无人机 i 和逃逸无人机 e 当前权重 $\hat{W}_{ia}$ 和 $\hat{W}_{ea}$ 的更新律设计为

$$\begin{aligned}\dot{\hat{W}}_{ia} &= -\Gamma_{ia}\left[k_{ia} + \frac{\hat{u}_i^T R_{ii}\hat{u}_i}{4(\chi_i^T\chi_i+1)^2}\right]\hat{W}_{ia} + \Gamma_{ia}\frac{(d_i+g_{ei})\nabla\phi_i^T(\delta_i)B_i\delta_t^i\hat{u}_i}{2T(\chi_i^T\chi_i+1)^2}\\ &\quad -\Gamma_{ia}\frac{(d_i+g_{ei})\nabla\phi_i^T(\delta_i)B_i\hat{u}_i\hat{u}_j^T R_{ij}\hat{u}_j}{2(\chi_i^T\chi_i+1)^2}\\ \dot{\hat{W}}_{ea} &= -\Gamma_{ea}\left[k_{ea} + \frac{\hat{u}_e^T R_e\hat{u}_e}{4(\chi_e^T\chi_e+1)^2}\right]\hat{W}_{ea} + \Gamma_{ea}\frac{(\kappa g_{et}-g_{e1})\nabla\phi_e^T(\delta_e)B_e\delta_t^e\hat{u}_e}{2T(\chi_e^T\chi_e+1)^2}\end{aligned} \tag{28}$$

其中 $\Gamma_{ia} > 0$ 和 $\Gamma_{ea} > 0$ 分别为追逐无人机 i 和逃跑无人机 e 具有适当维数的决策网络学习率，此外 $k_{ia} > 0$ 和 $k_{ea} > 0$ 为具有适当维数的反馈增益。

该改进点将优势函数与决策-评价架构相结合，即采用 Advantage Actor-Critic（A2C）方法，并将其扩展至多无人机追逃博弈问题，即分布式 A2C 方法，改进了现有技术所使用的状态信息会受到过大方差的博弈性能指标函数高估问题导致无人机的控制决策不稳定的问题，克服由无人机系统存在系统矩阵未知和外部干扰所导致的过大方差对控制策略平稳性的影响，能够提高多追逐无人机围捕控制效率和逃逸无人机逃跑控制策略的优化速度和稳定性。

第三部分阐述了解决的问题，具备的三个优点；第四部分针对三个优点，分别论述其优点的原理。

7.请用具体的例子说明一下该方案的最佳使用状态，如果有几个最佳状态请都说明。

要求：追逐无人机的数量被表示为N，表示至少包含 2 个无人机。追逃博弈拓扑关系图中的连接情况并不限于图 1 所示的特定情况，任意连接情况的通信图都适用于本方法。无人机的位置状态向量的维度被表示为n，并且不仅局限于平面二维空间，同样适用于三维空间。

最佳情况举例

场景：以在平面二维空间的图 1 通讯图进行举例说明：在该示例中，追逐无人机的数量为 5 个，它们需要形成一个正五边形编队，将逃逸无人机围捕至正五边形的中心位置。与此同时，逃逸无人机需要跟踪逃逸目标，并尽可能远离正五边形的中心位置。

效果图：本发明所提方法实现的效果为，5 个追逐无人机从任意初始位置出发，去围捕一个期望逃逸至具有运动轨迹的逃逸目标的逃逸无人机。如图 3 所示。使用本发明所提的方法后，理想的多无人机协作围捕博弈对抗的效果如图 4 所示。

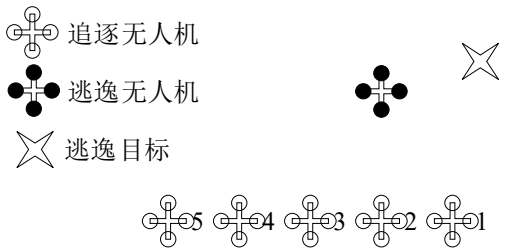

图 3 初始化追逐无人机、逃逸无人机、逃逸目标

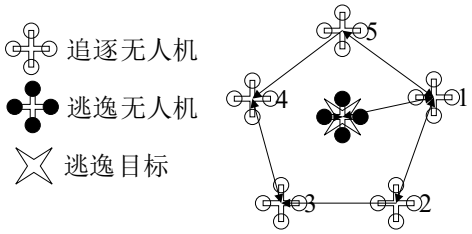

图 4 理想的追逃博弈效果