Aerospace Science and Technology 168 (2026) 111133

Contents lists available at ScienceDirect

Aerospace Science and Technology

journal homepage: www.elsevier.com/locate/aescte

Original article

Fixed-time adaptive consensus control for multi-quadrotor subject to external disturbances via deep reinforcement learning

Yefeng Yang [a,b], Kang Liu [a,*], Li-Yu Lo [a], Tao Huang [a,b], Yanming Fu [b], Chih-Yung Wen [a]

[a] Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, Hong Kong, China
[b] Center for Control Theory and Guidance Technology, School of Aeronautics, Harbin Institute of Technology, Harbin, Heilongjiang, 154000, China

ARTICLE INFO

Editor: Tsourdos Antonios

Keywords:
Quadrotor consensus control
Fixed-time control
Deep reinforcement learning
Disturbance observer
Sliding mode control

ABSTRACT

Formation control of quadrotors is particularly challenging under external disturbances and dynamic mission requirements. This paper introduces a hybrid control framework that combines fixed-time control with deep reinforcement learning (DRL) to achieve adaptive and robust multi-quadrotor formation control. A fixed-time disturbance observer (FTDO) is designed to accurately estimate disturbances, while a fully distributed fast non-singular terminal sliding mode controller ensures fixed-time convergence of both translational and rotational dynamics without singularities. To enhance adaptivity, a DRL-based mechanism enables online parameter tuning, thereby improving flight performance without compromising system stability. Both simulations and real-world experiments validate the effectiveness of the proposed framework, showing an average 50 % reduction in consensus tracking error compared with non-adaptive baselines.

1. Introduction

In recent years, multi-quadrotor technology has attracted growing attention in applications such as search and rescue [1], post-disaster relief [2], environmental monitoring [3], and logistics delivery [4]. Despite rapid progress, enabling autonomous flight for multiple quadrotors still faces several critical technical and theoretical challenges, including cooperative localization, 3D map reconstruction, obstacle avoidance, and formation control. Cooperative localization allows multiple agents to estimate their positions by sharing relative measurements and sensor data, offering greater accuracy and robustness than individual localization, particularly in GPS-denied or cluttered environments [5,6]. Similarly, cooperative 3D reconstruction enables agents to collaborate in sensing, exchanging, and fusing spatial information to generate a comprehensive environmental model, thereby improving accuracy, spatial coverage, and robustness in large-scale or complex scenarios [7,8]. After building a 3D map, clustering algorithms can group sensor-detected objects into obstacles, making cooperative obstacle avoidance essential for safe navigation. By sharing perception and coordinating maneuvers, multiple agents can reduce collision risks and improve efficiency in dynamic environments compared with independent avoidance strategies [9,10]. Obstacle avoidance ensures collision-free trajectories, while cooperative control guarantees that each quadrotor tracks its assigned trajectory within the specified time to accomplish the mission. Compared with single-quadrotor systems, multi-quadrotor systems exhibit superior capabilities in executing complex and large-scale tasks. Consequently, consensus-based formation control is vital for maintaining coordinated behavior among agents and enabling them to preserve a desired formation under dynamic environments and task requirements [11,12].

Among the challenges associated with multi-quadrotor formation flight, this paper focuses on the problem of consensus control. Considerable research has been conducted in this field. For example, Wang et al. [13] investigated a fully distributed dynamic event-triggered control scheme for quadrotors subject to unknown perturbations and input saturation, while Khodaverdian et al. [14] proposed a predictor-based sliding mode controller (SMC) for formation flight. To address the impact of uncertainties and external disturbances, Liu et al. [15] developed a finite-time adaptive control protocol, Zhang et al. [16] designed a prescribed finite-time distributed controller with aperiodic updates based on fuzzy logic, and Nie et al. [17] introduced a distributed asynchronous SMC method for multi-agent systems. Among these approaches, sliding mode control has attracted particular attention due to its robustness, fast response, and relatively simple design. Several studies have extended SMC-based frameworks for multi-quadrotor applications. Liang et al. [18] combined backstepping with SMC for multi-aircraft aerial transportation, while Hou et al. [19] developed an adaptive SMC method for trajectory tracking under environmental uncertainties. Liang et al. [20] further enhanced quadrotor func-

* Corresponding author.
E-mail address: kang1.liu@polyu.edu.hk (K. Liu).

https://doi.org/10.1016/j.ast.2025.111133
Received 15 February 2025; Received in revised form 23 September 2025; Accepted 21 October 2025
Available online 28 October 2025

tionality by integrating a robotic manipulator with an adaptive prescribed performance controller, and Ijaz et al. [21] proposed an integral SMC scheme with fault estimation for quadrotor systems. Although these methods have advanced the autonomous control of quadrotor formations, they still exhibit limitations in time-critical missions. Specifically, many approaches suffer from slow response, weak robustness, and strong dependence on accurate models in strongly coupled nonlinear systems. Moreover, conventional finite-time control strategies may not guarantee sufficiently fast convergence when quadrotor trajectories are explicitly time-parameterized, as is typical in practical planning modules.

To address the consensus control problem, various fixed-time control algorithms have been proposed. Zhao et al. [22] developed a fixed-time event-triggered SMC for multi-agent systems with unknown dynamics. Miao et al. [23] designed a fixed-time fault-tolerant controller for multi-quadrotor systems, while Su et al. [24] proposed a fixed-time formation-containment scheme. However, these methods typically assume that both the disturbances and their derivatives are bounded, an assumption that is often unrealistic in practice. In real-world operations, disturbances such as wind gusts or sudden changes in quadrotor mass may vary abruptly, making it essential to design controllers that guarantee fixed-time convergence while remaining robust against such uncertainties. Fixed-time sliding mode control with disturbance observers can ensure convergence, but selecting appropriate control gains remains a critical challenge. Insufficient gains may lead to slow convergence and poor disturbance rejection, whereas overly large gains can cause aggressive responses, instability, or even divergence. Consequently, it is highly desirable to develop adaptive, time-varying gain mechanisms that can balance convergence speed and robustness under diverse operating conditions.

Numerous adaptive sliding mode methods have been proposed to address the problem of gain tuning. For instance, Rodriguez et al. [25] developed an adaptive mechanism based on chattering detection. However, this approach adjusts the control gain only after chattering occurs, which is inadequate for quadrotor applications requiring strict real-time performance. Smith et al. introduced an adaptive tuning law where the adaptation rate was proportional to the error magnitude, but the optimality of such proportional adaptation remains unverified. Beyond classical adaptive methods, artificial intelligence (AI)-based approaches have been increasingly explored to enhance control performance. Deep reinforcement learning (DRL) is particularly attractive due to its independence from precise system models, adaptability, and powerful nonlinear approximation capability. Yan et al. [26] integrated DRL with SMC for multi-agent systems with time delays, while Wang et al. [27] investigated a data-driven framework combining SMC, DRL, and event-triggered control for unknown nonlinear systems. Similarly, Wang et al. [28] proposed a DRL-based SMC scheme for decentralized event-triggered control. Although DRL enhances adaptivity, it also introduces theoretical challenges that may undermine stability. Alternative learning paradigms such as Adaptive Dynamic Programming (ADP) have also been applied [29,30], but these often degrade fixed-time stability to uniformly ultimately bounded (UUB) stability, which is insufficient for time-critical applications. Therefore, it is crucial to develop a systematic adaptive gain tuning strategy that not only leverages learning-based adaptability but also guarantees rigorous theoretical stability, thereby bridging the gap between advanced control theory and practical quadrotor applications.

To address the challenges discussed above, this paper proposes a DRL-based fast nonsingular terminal sliding mode control (FNTSMC) framework. Specifically, a FNTSMC approach integrated with a fixed-time disturbance observer (FTDO) is developed to achieve fixed-time robust control for multi-quadrotor systems, with the FTDO serving as a model compensation strategy for formation control. Building on this, a hybrid scheme combining DRL with fixed-time SMC is introduced, which guarantees fixed-time convergence while enabling adaptive tuning and optimization of control parameters. Compared with prior studies, the main contributions of this work are summarized as follows:

1) A fully distributed FNTSMC is developed to address the consensus control problem in multi-quadrotor systems, with rigorous fixed-time stability guaranteed in the Lyapunov sense. Unlike the methods in [18–21], which only ensure finite-time stability, the proposed FNTSMC achieves fixed-time formation maintenance, offering faster and more predictable convergence.
2) A fixed-time disturbance observer (FTDO) is developed to estimate unknown external disturbances, with the estimation error guaranteed to converge to a small neighborhood of the origin within a fixed time. In contrast to the observers in [22–24], the proposed FTDO removes the restrictive assumption that disturbances must vary slowly with near-zero time derivatives, enabling more robust performance under abrupt disturbances.
3) Unlike the methods in [27–30], the proposed framework employs DRL to optimize FNTSMCs rather than directly replacing the controller with a neural network. This hybrid approach preserves fixed-time stability while significantly enhancing the robustness and flight performance of the quadrotor formation. The effectiveness and superiority of the proposed control framework are further validated through extensive simulations and real-world experiments.

The remainder of the paper is organized as follows: Section 2 introduces some preliminaries and outlines the problem addressed in this study. The controller design is presented in Section 3. In Section 4, a reinforcement learning-based parameter optimization framework is introduced to further improve the hyperparameters of the FNTSMCs. Simulations and experiments are conducted in Sections 5 and 6, respectively. Finally, Section 7 concludes the paper.

Notations: Given the extensive use of abbreviations and mathematical notations in this paper, we provide a comprehensive list of their definitions in Appendix A to enhance the clarity and reproducibility of our findings.

2. Preliminaries and problem formulation

2.1. Fundamental mathematics

The quadrotor group, consisting of N quadrotors, can be modeled as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ represents the set of nodes, with v_i denoting the ith quadrotor, and $\mathcal{E} = \{(v_i, v_j)\}$ denotes the set of edges, with (v_i, v_j) indicating that information can be transmitted from v_i to v_j. The neighbor set of v_i is defined as $\mathcal{N}_i = \{v_j \mid (v_j, v_i) \in \mathcal{E}\}$. The adjacency matrix is denoted by $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{N\times N}$, where $a_{ij} = 1$ if $(v_j, v_i) \in \mathcal{E}$ and $a_{ij} = 0$ otherwise. The graph is undirected if $a_{ij} = a_{ji}$ for all i, j, and directed if there exists at least one pair (i, j) such that $a_{ij} \neq a_{ji}$. The in-degree matrix is defined as $\mathcal{D} = \mathrm{diag}(d_1, d_2, \dots, d_N)$ with $d_i = \sum_{j=1}^{N} a_{ji}$, and the Laplacian matrix is given by $\mathcal{L} = \mathcal{D} - \mathcal{A}$. For leader-follower control, the leader adjacency matrix is defined as $\mathcal{B} = \mathrm{diag}(b_1, b_2, \dots, b_N)$, where $b_i = 1$ if v_i receives information from the leader and $b_i = 0$ otherwise. The augmented graph including the leader node v_b is denoted by $\mathcal{G}_{\mathcal{B}} = (\mathcal{V}, \mathcal{E}, v_b, \{b_i\})$. Without loss of generality, the following assumptions are made. Moreover, for clarity and convenience, several useful lemmas and definitions are provided below.

Assumption 1. [31] For the graph theory used in the study, the following standard conditions are required: (1) The graph $\mathcal{G}$ is undirected. (2) There are no self-loops in the graph $\mathcal{G}$. Namely, $a_{ii} = 0, i = 1, 2, \dots, N$. (3) There exists at least one spanning tree with the leader node v_b as the root of graph $\mathcal{G}_{\mathcal{B}}$.

Assumption 2. [32] The disturbances $\delta_{\rho,i}$, $\delta_{\eta,i}$ acted on the N quadrotors are bounded by unknown positive constants, namely, $||\delta_{\rho,i}|| \leq \overline{\delta}$ and $||\delta_{\eta,i}|| \leq \overline{\delta}$, where $\overline{\delta}$ is the unknown upper bound of the disturbances.

Assumption 3. [33] The yaw angle is bounded as $\psi_i \in [-\pi, \pi]$. To avoid singularities, the pitch and roll angles are bounded as $\phi_i, \theta_i \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

Lemma 1. *[34] For an undirected graph, the matrix $\mathcal{H} = \mathcal{L} + \mathcal{B}$ is symmetrical and positive definite if the graph $\mathcal{G}$ is connected and at least one follower can receive the leader's information.*

Lemma 2. *[35] For system $\dot{x} = f(x) \in \mathbb{R}^n$. If there exists a continuous radially unbounded function $V : \mathbb{R}^n \to \mathbb{R}_+\{0\}$ for the system such that $\dot{V} \le -\left[a_1 V(x)^{m_1} + a_2 V(x)^{m_2}\right]^k$ for some $a_1 > 0$, $a_2 > 0$, $m_1 > 0$, $m_2 > 0$, $k > 0$, $km_1 < 1$, and $km_2 > 1$. Then the system is said to be fixed-time stable and the settling time $T \le T_m = \frac{1}{a_1^k(1-km_1)} + \frac{1}{a_2^k(km_2-1)}$.*

Lemma 3. *[36] $\forall x_i \in \mathbb{R}$, $i = 1, 2, \dots, n$, $0 < p \le 1$, there is $\left(\sum_{i=1}^n |x_i|\right)^p \le \sum_{i=1}^n |x_i|^p \le n^{1-p}\left(\sum_{i=1}^n |x_i|\right)^p$.*

Definition 1. [37] For system $\dot{x} = f(x), f(0) = 0$, where $f : D \to \mathbb{R}^n$ is continuous on an open neighborhood $D \subseteq \mathbb{R}^n$ of the origin and $f(0) = 0$. The origin is said to be a finite-time-stable equilibrium of the system if there exists an open neighborhood $\mathcal{N} \subseteq D$ of the origin and a function $T : \mathcal{N}\backslash\{0\} \to (0, \infty)$, called the settling-time function, such that the following statements hold:
(1) Finite-time convergence: For every $x \in \mathcal{N}\backslash\{0\}$, ψ^x is defined on $[0, T(x)), \psi^x(t) \in \mathcal{N}\backslash\{0\}$ for all $t \in [0, T(x))$, and $\lim_{t\to T(x)} \psi^x(t) = 0$.
(2) Lyapunov stability: For every open neighborhood $\mathcal{U}_\varepsilon$ of 0 there exists an open subset $\mathcal{U}_\delta$ of $\mathcal{N}$ containing 0 such that, for every $x \in \mathcal{U}_\delta\backslash\{0\}$, $\psi^x(t) \in \mathcal{U}_\varepsilon, \forall t \in [0, T(x))$.
The origin is said to be a globally finite-time-stable equilibrium if it is a finite-time stable equilibrium with $D = \mathcal{N} = \mathbb{R}^n$.

Definition 2. [35] For system $\dot{x} = f(x), f(0) = 0, x_0 = x(0)$, where $f : D \to \mathbb{R}^n$ is continuous on an open neighborhood $D \subseteq \mathbb{R}^n$ of the origin and $f(0) = 0$. The system is said to be fixed-time stable if it is globally finite-time stable and the function $T(x)$ is bounded, i.e., $\exists T_{\max} > 0 : T(x_0) \le T_{\max}, \forall x_0 \in \mathbb{R}^n$.

2.2. System description

The dynamics of the ith quadrotor can be described as

$$\begin{aligned}
\ddot{\eta}_i &= \frac{u_{f,i}}{m_i + \Delta m_i} A_i(\rho_i) - \mathbf{g} - \frac{k_t}{m_i + \Delta m_i}\dot{\eta}_i + \frac{\delta_{\eta,i}}{m_i + \Delta m_i},\\
\dot{\omega}_i &= J_i^{-1}[-k_r\omega_i - \omega_i \times (J_i\omega_i) + \delta_{\rho,i} + \tau_i],\\
\dot{\rho}_i &= W_i(\rho_i)\omega_i,
\end{aligned} \tag{1}$$

where η_i, $u_{f,i}$, k_t, and m_i respectively represent the position, throttle, drag coefficient of the translational loop, and mass of the i-th quadrotor, Δm_i is the mass uncertainty, $\mathbf{g} = [0, 0, g]^\top$ is the gravitational acceleration; ω_i, J, k_r, and τ_i respectively denote the angular rate, inertia tensor matrix, drag coefficient of the rotational loop, and torque of the i-th quadrotor; $A_i \triangleq A_i(\rho_i) = [C_{\varphi,i}C_{\psi,i}S_{\theta,i} + S_{\varphi,i}S_{\psi,i}, C_{\varphi,i}S_{\psi,i}S_{\theta,i} - S_{\varphi,i}C_{\psi,i}, C_{\theta,i}C_{\varphi,i}]^\top$ and $W_i \triangleq W_i(\rho_i) = \begin{bmatrix} 1 & S_{\varphi,i}T_{\theta,i} & C_{\varphi,i}T_{\theta,i} \\ 0 & C_{\varphi,i} & -S_{\varphi,i} \\ 0 & S_{\varphi,i}/C_{\theta,i} & C_{\varphi,i}/C_{\theta,i} \end{bmatrix}$. $\delta_{\eta,i}$ and $\delta_{\rho,i}$ respectively denote the disturbances acting on the translational and rotational subsystems of the ith quadrotor. $C_{(\cdot)}$, $S_{(\cdot)}$, and $T_{(\cdot)}$ denote the cosine, sine, and tangent functions, respectively.

2.2.1. Rotational subsystem

The tracking error $e_{\rho,i}$ and the 1st and 2nd order derivatives of $e_{\rho,i}$ are given by

$$e_{\rho,i} = \rho_i - \rho_{d,i},\ \dot{e}_{\rho,i} = W_i\omega_i - \dot{\rho}_{d,i},\ \ddot{e}_{\rho,i} = \dot{W}_i\omega_i + W_i\dot{\omega}_i - \ddot{\rho}_{d,i}, \tag{2}$$

where $\rho_{d,i} = \left[\phi_{d,i}, \theta_{d,i}, \psi_{d,i}\right]^\top$ is the reference attitude angle, and $\dot{W}_i = \begin{bmatrix} 0 & \dot{\varphi}_i T_{\theta_i}C_{\varphi_i} + \frac{\dot{\theta}_i S_{\varphi_i}}{C_{\theta_i}^2} & -\dot{\varphi}_i S_{\varphi_i}T_{\theta_i} + \frac{\dot{\theta}_i C_{\varphi_i}}{C_{\theta_i}^2} \\ 0 & -\dot{\varphi}_i S_{\varphi_i} & -\dot{\varphi}_i C_{\varphi_i} \\ 0 & \frac{\dot{\varphi}_i C_{\varphi_i}C_{\theta_i} + \dot{\theta}_i S_{\varphi_i}S_{\theta_i}}{C_{\theta_i}^2} & \frac{-\dot{\varphi}_i S_{\varphi_i}C_{\theta_i} + \dot{\theta}_i C_{\varphi_i}S_{\theta_i}}{C_{\theta_i}^2} \end{bmatrix}$. By defining $\Delta_{\rho,i} = J_i^{-1}\delta_{\rho,i} - \ddot{\rho}_{d,i}$, $f_{\rho,i} = -J_i^{-1}[k_r\omega_i + \omega_i \times (J_i\omega_i)]$, $A_{\rho,i} = \dot{W}_i\omega_i + W_i f_{\rho,i}$, $B_{\rho,i} = W_i J_i^{-1}$ and doing some manipulations, Eq. 2 can be finally simplified as

$$\ddot{e}_{\rho,i} = A_{\rho,i} + B_{\rho,i}\tau_i + \Delta_{\rho,i}. \tag{3}$$

Remark 1. Note that the second-order derivative of $\rho_{d,i}$ is known for pure attitude control. However, in the case of position control, the desired attitude commands are generated by the translational subsystem. As a result, $\ddot{\rho}_{d,i}$ is absorbed into $\Delta_{\rho,i}$ and treated as part of the unknown disturbances.

2.2.2. Translational subsystem

The virtual expected acceleration of the ith quadrotor can be defined as

$$u_{\eta,i} = \left[a_{x,i}, a_{y,i}, a_{z,i}\right]^\top, \tag{4}$$

yielding

$$\ddot{\eta}_i = -\frac{k_t}{m_i}\dot{\eta}_i + u_{\eta,i} + \Delta_{\eta,i}, \tag{5}$$

where $\Delta_{\eta,i} = \frac{u_{f,i}}{m_i+\Delta m_i}A_i(\rho_i) - \mathbf{g} - \frac{k_t}{m_i+\Delta m_i}\dot{\eta}_i + \frac{k_t}{m_i}\dot{\eta}_i + \frac{\delta_{\eta,i}}{m_i+\Delta m_i} - u_{\eta,i}$ is the equivalent disturbance. Thereafter, it can be easily derived that

$$\begin{aligned}
u_{f,i} &= m_i\sqrt{a_{x,i}^2 + a_{y,i}^2 + (a_{z,i}+g)^2},\ \varphi_{d,i} = \arcsin\frac{m_i\left[a_{x,i}S_{\psi,i} - a_{y,i}C_{\psi,i}\right]}{u_{f,i}},\\
\theta_{d,i} &= \arctan\frac{a_{x,i}C_{\psi,i} + a_{y,i}S_{\psi,i}}{a_{z,i}+g}.
\end{aligned} \tag{6}$$

The consensus tracking error of the ith quadrotor can be defined as

$$e_{\eta,i} = \sum_{j=1}^N a_{ij}\left[(\eta_i - \nu_i) - (\eta_j - \nu_j)\right] + b_i(\eta_i - \eta_d - \nu_i), \tag{7}$$

where η_d is the reference trajectory of the geometric centre of the quadrotor formation and ν_i is the offset of the ith quadrotor to the geometric centre. Correspondingly, there is

$$\dot{e}_{\eta,i} = \sum_{j=1}^N a_{ij}\left[(\dot{\eta}_i - \dot{\nu}_i) - (\dot{e}a_j - \dot{n}u_j)\right] + b_i(\dot{\eta}_i - \dot{\eta}_d - \dot{\nu}_i), \tag{8}$$

For ease of theoretical derivation, a new variable can be defined as

$$\Lambda_i = b_i\eta_d + (b_i + d_i)\nu_i + \sum_{j=1}^N a_{ij}(\eta_j - \nu_j). \tag{9}$$

Correspondingly, one obtains

$$\begin{aligned}
\dot{\Lambda}_i =& b_i\dot{\eta}_d + (b_i + d_i)\dot{\nu}_i + \sum_{j=1}^N a_{ij}(\dot{\eta}_j - \dot{\nu}_j),\ \ddot{\Lambda}_i = b_i\ddot{\eta}_d + (b_i + d_i)\ddot{\nu}_i\\
&+ \sum_{j=1}^N a_{ij}(\ddot{\eta}_j - \ddot{\nu}_j).
\end{aligned} \tag{10}$$

Further, substituting $\ddot{\eta}_j$ into $\ddot{\Lambda}_i$ and doing some manipulations yield

$$\begin{aligned}
\ddot{\Lambda}_i &= b_i\ddot{\eta}_d + (b_i + d_i)\ddot{\nu}_i + \sum_{j=1}^N a_{ij}\left(\ddot{\eta}_j - \ddot{\nu}_j\right)\\
&= b_i\ddot{\eta}_d + (b_i + d_i)\ddot{\nu}_i + \sum_{j=1}^N a_{ij}\left(-\frac{k_{t,j}}{m_j}\dot{\eta}_j + u_{\eta,j} + \Delta_{\eta,j} - \ddot{\nu}_j\right)\\
&= \Lambda_{i0} + \sum_{j=1}^N a_{ij}\Delta_{\eta,j},
\end{aligned} \tag{11}$$

where

$$\Lambda_{i0} = b_i\ddot{\eta}_d + (b_i + d_i)\ddot{\nu}_i + \sum_{j=1}^N a_{ij}(-\frac{k_{t,j}}{m_j}\dot{\eta}_j + u_{\eta,j} - \ddot{\nu}_j)$$

is a known variable. The $\ddot{\eta}_d$ and $\ddot{\nu}_i$ in 'λ_{i0}' are manually defined and the $u_{\eta,j}$ and $\ddot{\nu}_j$ have to be accessed from other quadrotors. However, these two items are controlled by the a_{ij}. Specifically, as defined and described in the Laplacian matrix, $a_{ij} = 0$ means these messages cannot be transmitted from the jth quadrotor to the ith quadrotor; otherwise, $a_{ij} = 1$. Then, the error dynamics of the translational loop can be given by

$$\ddot{e}_{\eta,i} = -\frac{(d_i + b_i)k_{t,i}}{m_i}\dot{\eta}_i + (b_i + d_i)u_{\eta,i} - \Lambda_{i0} + (b_i + d_i)\Delta_{\eta,i} - \sum_{j=1}^N a_{ij}\Delta_{\eta,j}. \tag{12}$$

Table 1
Some variables used in controller design.

Variable	Definition	Variable	Definition
$m_{\rho 1\sim\rho 3}$ [a]	constants in rotational observer design	$z_{\rho 1}$	estimation of e_ρ
$z_{\rho 2}$	estimation of $\dot{e}_\rho$	$\beta_{1\sim 3}$ [a]	positive constants
$z_{\rho 3}$	estimation of Δ_ρ	$\alpha_{1\sim 3}$ [a]	positive constants
$p_{1\sim 6}$ [a]	constants in rotatinoal controllers	$k_{\rho 1\sim\rho 4}$ [a]	rotational loop control gains
$z_{\eta 1}$	estimation of e_η	$m_{\eta 1\sim\eta 3}$	constants in translational observer design
$z_{\eta 2}$	estimation of $\dot{e}_\eta$	$n_{\eta 1\sim\eta 3}$	constants in translational observer design
$z_{\eta 3}$	estimation of Δ_η	$k_{\eta 1\sim\eta 4}$	translational loop control gains
$q_{1\sim 6}$	constants in translational controllers		

[a] Subscript 'i' in this section represents the index of the quadrotor.

2.3. Problem formulation

Based on the Assumption. 1–3, the control objective of this paper is formulated as follows:

Control Objective: Given a quadrotor group of N quadrotors subject to airflow disturbances generated by fans and additional suspended payloads with uncertain mass, we assume the topological graph $\mathcal{G}$ satisfies the conditions in Assumption. 1. For a set of reference trajectories generated by the virtual leader node with $\eta_d = [x_d, y_d, z_d]^\top$ being the reference position, ψ_d being the reference yaw angle, and $\nu_i = [\nu_{x,i}, \nu_{y,i}, \nu_{z,i}]^\top$ denoting the offset of the ith quadrotor from η_d, we design an adaptive controller such that for $i = 1, 2, \dots, N$

$$\lim_{t\to T_\eta}\left[\eta_i(t) - \left(\eta_d + \nu_i\right)\right] = 0,\ \lim_{t\to T_\rho}\left[\psi_i(t) - \psi_d\right] = 0, \tag{13}$$

with T_η and T_ρ being the settling time of the translational and rotational subsystems, respectively.

Remark 2. Note that η_d and ν_i can be either constants or time-varying variables, which implies that the UAV formation may either maintain a fixed configuration or undergo a change in formation. The uncertainties considered in the experiments of this work are classified into two types: (i) the mass uncertainty of the UAVs, and (ii) the wind disturbances generated by the fan. The control framework designed in this work theoretically guarantees that the consensus tracking error converges to zero within a fixed time.

3. Controller design

This section provides a consensus controller design for the quadrotor formation. Before starting this section, we first provide Table 1 of some important mathematical symbols and their definitions in this section for easy reference.

3.1. Rotational subsystem stability

For simplicity, the subscript 'i' in the rotational loop controller design is omitted since the quadrotors are all homogeneous and the design of the FTDO and FNTSMC in the rotational loop does not require the information from other quadrotors [38].

To begin with, an FTDO can be designed as

$$\begin{aligned}
\dot{z}_{\rho 1} &= \aleph_\rho m_{\rho 1}\left\lfloor\tilde{e}_\rho\right\rceil^{\alpha_1} + (1-\aleph_\rho) m_{\rho 1}\left\lfloor\tilde{e}_\rho\right\rceil^{\beta_1} + z_{\rho 2},\\
\dot{z}_{\rho 2} &= \aleph_\rho m_{\rho 2}\left\lfloor\tilde{e}_\rho\right\rceil^{\alpha_2} + (1-\aleph_\rho) m_{\rho 2}\left\lfloor\tilde{e}_\rho\right\rceil^{\beta_2} + z_{\rho 3} + A_\rho + B_\rho\tau,\\
\dot{z}_{\rho 3} &= \aleph_\rho m_{\rho 3}\left\lfloor\tilde{e}_\rho\right\rceil^{\alpha_3} + (1-\aleph_\rho) m_{\rho 3}\left\lfloor\tilde{e}_\rho\right\rceil^{\beta_3},
\end{aligned} \tag{14}$$

where $z_{\rho 1}$, $z_{\rho 2}$, and $z_{\rho 3}$ are the estimations of e_ρ, $\dot{e}_\rho$, and Δ_ρ, respectively, $\tilde{e}_{\rho,i} = e_\rho - z_{\rho 1}$ is the estimation error of e_ρ. $\aleph_\rho$ is a switching parameter; α_1, α_2, α_3, β_1, β_1, and β_3 are positive constants. Specifically, $\alpha_1 = \frac{3}{4}$, $\alpha_2 = \frac{2}{4}$, $\alpha_3 = \frac{1}{4}$, $\beta_1 = \frac{5}{4}$, $\beta_2 = \frac{6}{4}$, $\beta_3 = \frac{7}{4}$, and $\aleph_\rho = 0$ if $||\tilde{e}_\rho||_2 > e_\rho^*$ else 1 with e_ρ^* being the threshold of the estimation error. Apart from that, parameters $m_{\rho 1}$, $m_{\rho 2}$, and $m_{\rho 3}$ are designed such that matrix $\Gamma_{m,\rho} = \begin{bmatrix} -m_{\rho 1} & 1 & 0 \\ -m_{\rho 2} & 0 & 1 \\ -m_{\rho 3} & 0 & 0 \end{bmatrix}$ is Hurwitz.

With observer (14) and Lemma. 2, Δ_ρ can be estimated in a fixed-time $\mathcal{T}_{\rho 1}$, and the estimation error converges to a neighborhood of the origin Ω_ρ. Due to space limitations, the detailed stability analysis is omitted here. Interested readers are referred to Theorem 1 in [39] and Theorem 1 in [40] for a comprehensive proof.

Remark 3. Observer (14) is a second-order FTDO. $z_{\rho 1}$, $z_{\rho 2}$, and $z_{\rho 3}$ are utilized to estimate the e_ρ, $\dot{e}_\rho$, and Δ_ρ, respectively. The key principle of this observer lies in estimating the Δ_ρ by using the $\tilde{e}_\rho$, the observation error of $e_{\rho 1}$. Theoretically, as $||\tilde{e}_\rho||_2$ decreases, $z_{\rho 1}$, $z_{\rho 2}$, and $z_{\rho 3}$ would gradually converge to the real values of e_ρ, $\dot{e}_\rho$, and Δ_ρ, respectively. The higher the order of the observer, the longer it takes for the observation error to propagate to the highest derivative. To satisfy the requirements on $\alpha_{1\sim 3}$ and $\beta_{1\sim 3}$ in observer [39], for a P-order observer, the values of α_1 to α_P need to be designed as $\frac{P}{P+1}, \frac{P-1}{P+1}, \dots, \frac{1}{P+1}$ accordingly. Similarly, β_1 to β_P should be set as $\frac{P+2}{P+1}, \frac{P+3}{P+1}, \dots, \frac{2P+1}{P+1}$. In this study, $P = 3$. Therefore, $\alpha_1 = \frac{3}{4}$, $\alpha_2 = \frac{2}{4}$, $\alpha_3 = \frac{1}{4}$, $\beta_1 = \frac{5}{4}$, $\beta_2 = \frac{6}{4}$, $\beta_3 = \frac{7}{4}$. As for the selection of $\aleph_\rho$, it is a threshold for switching the observer. Empirically, the value of $\aleph_\rho$ can be designed to be slightly less than 0.5 in practical applications. This choice ensures system stability while enhancing the observer's sensitivity to large errors, thereby enabling faster convergence.

Remark 4. Although the stability of the observer can be ensured by guaranteeing the stability of the matrix $\Gamma_{m,\rho}$, in practical applications it is often challenging to determine whether a third-order matrix is Hurwitz simply by inspecting its parameters. Therefore, we utilize a method that integrates linear system theory to explicitly compute $m_{\rho 1}$, $m_{\rho 2}$, and $m_{\rho 3}$. Solving $|\lambda I_3 - \Gamma_{m,\rho}| = 0$ yields

$$\lambda^3 + m_{\rho 1}\lambda^2 - m_{\rho 2}\lambda + m_{\rho 3} = 0. \tag{15}$$

Simultaneously, assuming three negative real roots of a third-order linear equation are $-\omega_1$, $-\omega_2$, and $-\omega_3$ with $\omega_1, \omega_2, \omega_3 > 0$. Then, we have

$$\begin{aligned}
&\lambda^3 + m_{\rho 1}\lambda^2 - m_{\rho 2}\lambda + m_{\rho 3}\\
=&(\lambda+\omega_1)(\lambda+\omega_2)(\lambda+\omega_3)\\
=&\lambda^3 + (\omega_1+\omega_2+\omega_3)\lambda^2 + (\omega_1\omega_2 + \omega_2\omega_3 + \omega_1\omega_3)\lambda + \omega_1\omega_2\omega_3 = 0
\end{aligned} \tag{16}$$

which yields

$$m_{\rho 1} = \omega_1 + \omega_2 + \omega_3,\ m_{\rho 2} = \omega_1\omega_2 + \omega_2\omega_3 + \omega_1\omega_3,\ m_{\rho 3} = \omega_1\omega_2\omega_3. \tag{17}$$

Therefore, we can compute $m_{\rho 1}$, $m_{\rho 2}$, and $m_{\rho 3}$ by selecting appropriate ω_1, ω_2, and ω_3.

Larger values of ω_i ($i = 1, 2, 3$) indicate that the observer has a higher bandwidth, which implies faster convergence. However, prior to convergence, the observer output may exhibit large magnitudes and heightened sensitivity to noise. Conversely, smaller values of ω_i correspond to lower bandwidth, suggesting that the observer produces less overshoot and is less sensitive to noise, but at the cost of slower convergence. In practical applications, it is essential to strike a balance between convergence speed and overshoot based on specific requirements. In numerical simulations, where conditions are relatively idealized, the values of ω_i, ($i = 1, 2, 3$) can be set around 4 to increase the observer bandwidth and shorten convergence time. However, in practical implementations, the system control frequency typically ranges from 50 Hz to 200 Hz. Under

such conditions, an excessively large observer bandwidth may lead to divergence.

Thereafter, a sliding mode surface is defined as

$$s_\rho = e_\rho + k e_{\rho 1}^{\frac{p_1}{p_2}} + k \dot{e}_{\rho 2}^{\frac{p_3}{p_4}}, \tag{18}$$

where $\rho l > 0$, $\rho 2 > 0$. p_1, p_2, p_3, and p_4 are all positive odd numbers satisfying $\frac{p_1}{p_2} > \frac{p_3}{p_4} > 1$ and $2 > \frac{p_3}{p_4} > 1$. Assuming there are no disturbances or uncertain terms in the system, an equivalent control law can be given by

$$\tau_{eq} = -B_\rho^{-1}\left[A_\rho + \frac{p_4}{k_{\rho 1} p_3} \dot{e}_\rho^{2-\frac{p_3}{p_4}} \circ \left(\mathbf{I}_3 - \frac{k_{\rho 1} p_1}{p_3} e_\rho^{\frac{p_1}{p_2}-1}\right)\right]. \tag{19}$$

By Wang et al. [41], it can be easily verified that the control matrix B_ρ is of full rank and invertible.

In addition, a switching control law is further required to maintain s_ρ at the origin when there exist disturbances or uncertainty in the system, which is given by

$$\tau_{sw} = -B_\rho^{-1}\left[z_{\rho 3} + k_{\rho 3}\,\mathrm{sgn}\,(s_\rho) + k r_{\rho 4} s_\rho^{\frac{p_5}{p_6}}\right], \tag{20}$$

where $k_{\rho 3} > 0$, $k_{\rho 4} > 0$ are positive constants. $p_5 > p_6 > 1$ are all positive odd parameters. Then, the complete control law for the rotational loop can be designed as

$$\tau = \tau_{eq} + \tau_{sw}. \tag{21}$$

Based on the analysis and derivation aforementioned, the following theorem can be concluded.

Theorem 1. *For the rotational subsystem of the quadrotor (3) disturbed by Δ_ρ, the system is fixed-time stable with the FNTSMC (21) and the FTDO (14).*

Proof. Firstly, we need to prove that the sliding mode surface converges to the origin in a fixed time.

Choose a Lyapunov function candidate as $V_{\rho 1} = \frac{1}{2} s_\rho^\top s_\rho$. Differentiating $V_{\rho 1}$ yields

$$\dot{V}_{\rho 1} = s_\rho^\top \left\{ \dot{e}_\rho + \frac{k_{\rho 1} p_1}{p_2} e_\rho^{\frac{p_1}{p_2}-1} \circ \dot{e}^\rho + \frac{k_{\rho 2} p_3}{p_4} \dot{e}_\rho^{\frac{p_3}{p_4}-1} \circ \left[A_\rho + B_\rho(\tau_{eq} + \tau_{sw}) + \Delta_\rho\right] \right\} \tag{22}$$

Substituting controller (21) into $\dot{V}_{\rho 1}$ and doing some manipulations yield

$$\dot{V}_{\rho 1} = s_\rho^\top \left[\frac{k_{\rho 2} p_3}{p_4} \dot{e}_\rho^{\frac{p_3}{p_4}-1} \circ \left(\Delta_\rho - z_{\rho 3} - k_{\rho 3}\,\mathrm{sgn}\,(s_\rho) - k_{\rho 4} s_\rho^{\frac{p_5}{p_6}}\right)\right]. \tag{23}$$

Denote $\tilde{\Delta}_\rho = \Delta_\rho - z_{\rho 3}$ as the estimation error of Δ_ρ and $k_{\rho 0} = \frac{k_{\rho 2} p_3}{p_4} \dot{e}_\rho^{\frac{p_3}{p_4}-1}$, which yields

$$\begin{aligned}\dot{V}_{\rho 1} = -k_{\rho 0}^\top \circ s_\rho^\top \left(k_{\rho 3}\,\mathrm{sgn}\,(s_\rho) + k_{\rho 4} s_\rho^{\frac{p_5}{p_6}} - \tilde{\Delta}_\rho\right) = &- k_{\rho 4} k_{\rho 0}^\top s_\rho^{\frac{p_5+p_6}{p_6}} \\ &- k_{\rho 5} k_{\rho 0}^\top |s_\rho|\end{aligned} \tag{24}$$

where $k_{\rho 5} = k_{\rho 3} - ||\tilde{\Delta}_\rho||_2 > 0$.

Note the fact that all elements in $k_{\rho 0}$ are non-negative. Demoting the minimum element in $k_{\rho 0}$ as k_ρ and using Lemma. 3 yield

$$\begin{aligned}\dot{V}_{\rho 1} \le -k_{\rho 4} k_\rho ||s_\rho||_2^{\frac{p_5+p_6}{2p_6}} - k_{\rho 5} k_\rho ||s_\rho||_2^{\frac{1}{2}} = &- k_{\rho 4} k_\rho 2^{\frac{p_5+p_6}{2p_6}} V_{\rho 1}^{\frac{p_5+p_6}{2p_6}} \\ &- k_{\rho 5} k_\rho \sqrt{2} V_{\rho 1}^{\frac{1}{2}}.\end{aligned} \tag{25}$$

Using Lemma. 2 and the fact $p_5 > p_6$ yield that s_ρ is fixed-time stable, and the settling time $\mathcal{T}_{\rho 2}$ is bounded by

$$\mathcal{T}_{\rho 2} \le \frac{\sqrt{2}}{k_{\rho 5} k_\rho} + \frac{p_6}{k_{\rho 4} k_\rho (p_5 - p_6)} 2^{\frac{2p_6}{p_5 - p_6}}. \tag{26}$$

Secondly, we need to prove that e_ρ converges to the origin in a fixed time when the states are maintained on the sliding mode surface. On the sliding mode surface, we have $e_\rho + k_{\rho 1} e_\rho^{\frac{p_1}{p_2}} + k_{\rho 2} \dot{e}_\rho^{\frac{p_3}{p_4}} = 0$, yielding $\dot{e}_\rho^{\frac{p_3}{p_4}} = -\frac{1}{k_{\rho 2}}\left(e_\rho + k_{\rho 1} e_\rho^{\frac{p_1}{p_2}}\right)$. Choose a Lyapunov function candidate as $V_{\rho 2} = \frac{1}{2} e_\rho^\top e_\rho$. Differentiating $V_{\rho 2}$ along the system trajectory and using Eq. (3) yield

$$\begin{aligned}\dot{V}_{\rho 2} &= -\left[\frac{1}{k_{\rho 2}}\left(e_\rho^\top\right)^{\frac{p_3}{p_4}}\left(e_\rho + k_{\rho 1} e_\rho^{\frac{p_1}{p_2}}\right)\right]^{\frac{p_4}{p_3}} \\ &\le -\left[\frac{1}{k_{\rho 2}}||e_\rho||_2^{\frac{p_3+p_4}{p_4}} + \frac{k_{\rho 1}}{k_{\rho 2}}||e_\rho||_2^{\frac{p_1}{p_2}+\frac{p_3}{p_4}}\right]^{\frac{p_4}{p_3}} \\ &= -\left[\kappa_{\rho 1} V_{\rho 2}^{\frac{p_3+p_4}{2p_4}} + \kappa_{\rho 2} V_{\rho 2}^{\frac{1}{2}\left(\frac{p_1}{p_2}+\frac{p_3}{p_4}\right)}\right]^{\frac{p_4}{p_3}},\end{aligned} \tag{27}$$

where $\kappa_{\rho 1} = \frac{\sqrt{2^{(p_3+p_4)/p_4}}}{k_{\rho 2}}$ and $\kappa_{\rho 2} = \frac{k_{\rho 1}\sqrt{2^{p_1/p_2+p_3/p_4}}}{k_{\rho 2}}$.

By Lemma 2, one concludes e_ρ converges to the origin in a fixed time, and the settling time $\mathcal{T}_{\rho 3}$ can be bounded by

$$\mathcal{T}_{\rho 3} = \frac{2p_3}{(p_3 - p_4)\kappa_{\rho 1}^{p_4/p_3}} + \frac{2p_2 p_3}{(p_1 p_4 - p_2 p_3)\kappa_{\rho 2}^{p_4/p_3}}. \tag{28}$$

Therefore, the convergence time of the system with external disturbance converges within $\mathcal{T}_\rho \le \mathcal{T}_{\rho 1} + \mathcal{T}_{\rho 2} + \mathcal{T}_{\rho 3}$. The proof is completed. □

Remark 5. In this study, the term *non-singular* refers to the issue where the denominator appearing in the first derivative of the sliding surface variable becomes zero, leading to an undefined system. For example, a sliding mode surface $s = \dot{x} + c x^p$, $c > 0$, $0 < p < 1$. Differentiating s gives $\dot{s} = \ddot{x} + cpx^{p-1}\dot{x}$. By observing $\dot{s}$ we find the singularity probelm occurs when $x = 0$ because $p - 1 < 0$. However, in matrix theory, the term singular refers to a square matrix that is non-invertible. When certain matrices are invertible but have large condition numbers, they are considered nearly singular and exhibit poor numerical stability. Such matrices are referred to as ill-conditioned. Therefore, in the context of this study and its research domain, the term *non-singular* refers to the first definition.

Remark 6. In Eq. ((18)), it is obvious to conclude that $\frac{p_3+p_4}{2p_4} \cdot \frac{p_4}{p_3} = \frac{p_3+p_4}{2p_3} \in (0,1)$ and $\frac{1}{2}\left(\frac{p_1}{p_2} + \frac{p_3}{p_4}\right) \cdot \frac{p_4}{p_3} = \frac{1}{2}\left(\frac{p_1 p_4}{p_2 p_3} + 1\right) > 1$ hold for $\frac{p_1}{p_2} > \frac{p_3}{p_4} > 1$, satisfying the conditions required in Lemma 2.

3.2. Translational subsystem stability

Similarly, the fixed-time disturbance observer can be designed as

$$\begin{aligned}\dot{z}_{\eta 1,i} &= \aleph_{\eta,i} m_{\eta 1,i}\left[\tilde{e}_{\eta 1,i}\right]^{\alpha_1} + (1 - \aleph_{\eta,i}) m_{\eta 1,i}\left[\tilde{e}_{\eta 1,i}\right]^{\beta_1} + z_{\eta 2,i}, \\ \dot{z}_{\eta 2,i} &= \aleph_{\eta,i} m_{\eta 2,i}\left[\tilde{e}_{\eta 1,i}\right]^{\alpha_2} + (1 - \aleph_{\eta,i}) m_{\eta 2,i}\left[\tilde{e}_{\eta 1,i}\right]^{\beta_2} + z_{\eta 3,i} - \frac{k_t}{m_i}\dot{\eta}_i + u_{\eta,i}, \\ \dot{z}_{\eta 3,i} &= \aleph_{\eta,i} m_{\eta 3,i}\left[\tilde{e}_{\eta 1,i}\right]^{\alpha_3} + (1 - \aleph_{\eta,i}) m_{\eta 3,i}\left[\tilde{e}_{\eta 1,i}\right]^{\beta_3},\end{aligned} \tag{29}$$

where $z_{\eta 1,i}$, $z_{\eta 2,i}$, and $z_{\eta 3,i}$ are the estimates of $e_{\eta,i}$,$\dot{e}_{\eta,i}$,and $\Delta_{\eta,i}$,respectively, $\tilde{e}_{\eta 1,i} = e_{\eta,i} - z_{\eta 1,i}$ is the estimate error of $e_{\eta,i}$. $\aleph_{\eta,i}$ is a switching parameter and $\aleph_{\eta,i} = 0$ if $||\tilde{e}_{\eta,i,1}||_2 > e_{\eta,i}^*$ else 1 with $e_{\eta,i}^*$ being the threshold of the estimation error. Moreover, hyper-parameters $m_{\eta 1,i}$, $m_{\eta 2,i}$, and $m_{\eta 3,i}$ are designed such that matrix $\Gamma_{m,\eta,i} = \begin{bmatrix} -m_{\eta 1,i} & 1 & 0 \\ -m_{\eta 2,i} & 0 & 1 \\ -m_{\eta 3,i} & 0 & 0 \end{bmatrix}$ is Hurwitz. Similarly, $\Delta_{\eta,i}$ can be estimated in fixed-time $\mathcal{T}_{\eta 1,i}$, and the estimation error converges to a neighborhood of the origin $\Omega_{\eta,i}$ [39].

Thereafter, a sliding mode surface for the ith quadrotor can be defined as

$$s_{\eta,i} = e_{\eta,i} + k_{\eta 1,i} e_{\eta,i}^{\frac{q_1}{q_2}} + k_{\eta 2,i} \dot{e}_{\eta,i}^{\frac{q_3}{q_4}}, \tag{30}$$

where $k_{\eta 1.i} > 0$, $k_{\eta 2,i} > 0$. q_1, q_2, q_3, and q_4 are all positive odd numbers satisfying $\frac{q_1}{q_2} > \frac{q_3}{q_4} > 1$ and $2 > \frac{q_3}{q_4} > 1$. An equivalent control law is then proposed to maintain $s_{\eta,i}$ on the sliding mode surface, which is given by

$$u_{\eta,i,eq} = -\frac{1}{b_i + d_i}\left[-\frac{(b_i + d_i)k_{t,i}}{m_i}\dot{\eta}_i - \Lambda_{i0} + \frac{q_4}{q_3 k_{\eta 2,i}}\dot{\eta}_i^{2-\frac{q_3}{q_4}} \circ \left(\mathbf{I}_3 - \frac{q_1 k_{\eta 1.i}}{q_2} e_{\eta,i}^{\frac{q_1}{q_2}-1}\right)\right]. \tag{31}$$

To handle uncertainties and external disturbances in the translational subsystem, a switching control law is required, which is designed as

$$u_{\eta,i,sw} = -\frac{1}{b_i + d_i}\left[(b_i + d_i) z_{\eta 3,i} + \sum_{i=1}^{N} a_{ij} z_{\eta 3,j} + k_{\eta 3,i}\,\mathrm{sgn}(s_{\eta,i}) - k_{\eta 4i} s_{\eta,i}^{\frac{q_5}{q_6}}\right], \tag{32}$$

where $k_{\eta 3,i}$ and $k_{\eta 4i}$ are positive gains. $q_5 > q_6 > 1$ are all positive odd parameters. Finally, the complete control law is given by

$$u_{\eta,i} = u_{\eta,i,eq} + u_{\eta,i,sw}. \tag{33}$$

Similar to that of the rotational subsystem, the following theorem guarantees the stability of the translational subsystem of the entire quadrotor group.

Theorem 2. *For the consensus tracking error of the translational subsystems of the quadrotor formation (12) disturbed by Δ_i, the system is fixed-time stable with FNTSMC (33) and FTDO (29).*

Proof. Firstly, we need to prove $s_{\eta,i}$, $i = 1, 2, \ldots, N$ converge to the origin in fixed time. Choose a Lyapunov function candidate as $V_{\eta 1} = \frac{1}{2}\sum_{i=1}^{N} s_{\eta,i}^{\top} s_{\eta,i}$. Differentiating $V_{\eta 1}$ yields

$$\dot{V}_{\eta 1} = \sum_{i=1}^{N} s_{\eta,i}^{\top}\left(\dot{e}_{\eta,i} + \frac{q_1 k_{\eta 1.i}}{q_2} e_{\eta,i}^{\frac{q_1}{q_2}-1} \circ \dot{e}_{\eta,i} + \frac{q_3 k_{\eta 3,i}}{q_4}\dot{e}_{\eta,i}^{\frac{q_3}{q_4}-1} \circ \ddot{e}_{\eta,i}\right). \tag{34}$$

Substituting Eq. (12) into $\dot{V}_{\eta 1}$, using controller (33), and doing some manipulations yield

$$\begin{aligned}
\dot{V}_{\eta 1} &= \sum_{i=1}^{N} s_{\eta,i}^{\top}\left\{\dot{e}_{\eta,i} + \frac{q_1 k_{\eta 1.i}}{q_2} e_{\eta,i}^{\frac{q_1}{q_2}-1} \circ \dot{e}_{\eta,i}\right. \\
&\quad \left. + \frac{q_3 k_{\eta 3,i}}{q_4}\dot{e}_{\eta,i}^{\frac{q_3}{q_4}-1}\left[-\frac{(b_i + d_i)k_{t,i}}{m_i}\dot{\eta}_i + (b_i + d_i)u_{\eta,i} - \Lambda_{i0} + (b_i + d_i)\Delta_{\eta,i} - \sum_{j=i}^{N} a_{ij}\Delta_{\eta,j}\right]\right\} \\
&= \sum_{i=1}^{N} s_{\eta,i}^{\top}\left\{\frac{q_3 k_{\eta 2,i}}{q_4}\dot{e}_{\eta,i}^{\frac{q_3}{q_4}-1} \circ \left[(b_i + d_i)(\Delta_{\eta,i} - z_{\eta 3,i}) + \sum_{j=i}^{N} a_{ij}\left(z_{\eta 3,j} - \Delta_{\eta,j}\right)\right]\right\}.
\end{aligned} \tag{35}$$

Define the estimatation error of $\Delta_{\eta,i}$ as $\tilde{\Delta}_{\eta,i} = \Delta_{\eta,i} - z_{\eta 3,i}$ and $k_{\eta 0,i} = \frac{q_3 k_{\eta 2,i}}{q_4}\dot{e}_{\eta,i}^{\frac{q_3}{q_4}-1}$. $\dot{V}_{\eta 1}$ can be simplified as

$$\dot{V}_{\eta 1} = -\sum_{i=1}^{N} k_{\eta 0,i}^{\top} \circ s_{\eta,i}^{\top}\left\{\left[k_{\eta 3,i}\,\mathrm{sgn}(s_{\eta,i}) + k_{\eta 4i} s_{\eta,i}^{\frac{q_5}{q_6}} - (b_i + d_i)\tilde{\Delta}_{\eta,i} - \sum_{j=i}^{N} a_{ij}\tilde{\Delta}_{\eta,j}\right]\right\}. \tag{36}$$

Note the fact that all elements in $k_{\eta 0,i}$ are non-negative. Using Eq. (3) and denoting the minimum element in $k_{\eta 0,i}$ as $k_{\eta,i}$ yield

$$\begin{aligned}
\dot{V}_{\eta 1} &\le -\sum_{i=1}^{N} k_{\eta,i} s_{\eta,i}^{\top}\left\{\left[k_{\eta 3,i}\,\mathrm{sgn}(s_{\eta,i}) + k_{\eta 4i} s_{\eta,i}^{\frac{q_5}{q_6}} - (b_i + d_i)\tilde{\Delta}_{\eta,i} - \sum_{j=i}^{N} a_{ij}\tilde{\Delta}_{\eta,j}\right]\right\} \\
&= -\sum_{i=1}^{N} k_{\eta,i} s_{\eta,i}^{\top}\left[k_{\eta 3,i}\,\mathrm{sgn}(s_{\eta,i}) + k_{\eta 4i} s_{\eta,i}^{\frac{q_5}{q_6}}\right] = \\
&\quad -\sum_{i=1}^{N} k_{\eta,i}\left[k_{\eta 3,i}\sum_{j=1}^{3}\left|s_{\eta,i,j}\right| + k_{\eta 4i}\sum_{j=1}^{3}\left|s_{\eta,i,j}\right|^{1+\frac{q_5}{q_6}}\right] \\
&\le -\sum_{i=1}^{N} k_{\eta,i}\left[k_{\eta 3,i}||s_{\eta,i}||_2^{\frac{1}{2}} + k_{\eta 4i}||s_{\eta,i}||_2^{\frac{1}{2}\left(1+\frac{q_5}{q_6}\right)}\right] \\
&\le -\underline{k}_{\eta 3}\sum_{i=1}^{N}\sum_{j=1}^{3}\left|s_{\eta,i,j}\right| - \underline{k}_{\eta 4}\sum_{i=1}^{N}\sum_{j=1}^{3}\left|s_{\eta,i,j}\right|^{1+\frac{q_5}{q_6}},
\end{aligned} \tag{37}$$

where $\underline{k}_{\eta 3} = \min(k_{\eta,i}, k_{\eta 3,i})$, $\underline{k}_{\eta 4} = \min(k_{\eta,i}, k_{\eta 4i})$, and $k_{\eta,i} = k_{\eta 3,i} - (b_i + d_i)||\tilde{\Delta}_{\eta,i}||_2 - \sum_{j=1}^{N} a_{ij}||\tilde{\Delta}_{\eta,j}||_2 > 0$ for $i = 1, 2, \ldots, N$. Using Lemma 3 again in Eq. 37 yields

$$\begin{aligned}
\dot{V}_{\eta 1} &\le -\underline{k}_{\eta 3}\sum_{i=1}^{N}\left(||s_{\eta,i}||_2^2\right)^{\frac{1}{2}} - \underline{k}_{\eta 4}\sum_{i=1}^{N}\left(||s_{\eta,i}||_2^2\right)^{\frac{1}{2}\left(1+\frac{q_5}{q_6}\right)} \\
&= -\underline{k}_{\eta 3}\left(\sum_{i=1}^{N}||s_{\eta,i}||_2^2\right)^{\frac{1}{2}} - \underline{k}_{\eta 4}\left(\sum_{i=1}^{N}||s_{\eta,i}||_2^2\right)^{\frac{1}{2}\left(1+\frac{q_5}{q_6}\right)} \\
&= -\kappa_{\eta 1} V_{\eta 1}^{\frac{1}{2}} - \kappa_{\eta 2} V_{\eta 1}^{\frac{1}{2}\left(1+\frac{q_5}{q_6}\right)},
\end{aligned} \tag{38}$$

where $\kappa_{\eta 1} = \underline{k}_{\eta 3}\sqrt{2}$ and $\kappa_{\eta 2} = \underline{k}_{\eta 4}\sqrt{2^{(q_6+q_6)/q_6}}$. Using Lemma 2 yields that $s_{\eta,i}, i = 1, 2, \ldots, N$ are fixed-time stable, and the settling time $\mathcal{T}_{\eta 2}$ can be bounded by $\mathcal{T}_{\eta 2} \le \frac{2}{\kappa_{\eta 1}} + \frac{2q_6}{\kappa_{\eta 2}(q_5 - q_6)}$.

Secondly, similar to that of the rotational subsystem, we need to prove that the tracking errors of the quadrotor group converge to the origin in a fixed time. On the sliding mode surface, there is $s_{\eta,i} = e_{\eta,i} + k_{\eta 1.i} e_{\eta,i}^{\frac{q_1}{q_2}} + k_{\eta 2,i}\dot{e}_{\eta,i}^{\frac{q_3}{q_4}} = 0$, yielding $\dot{e}_{\eta,i}^{\frac{q_3}{q_4}} = -\frac{1}{k_{\eta 2,i}}\left(e_{\eta,i} + k_{\eta 1.i} e_{\eta,i}^{\frac{q_1}{q_2}}\right)$. A Lyapunov function candidate can be defined as $V_{\eta 2} = \frac{1}{2}\sum_{i=1}^{N} e_{\eta,i}^{\top} e_{\eta,i}$. Differentiating $V_{\eta 2}$ yields

$$\begin{aligned}
\dot{V}_{\eta 2} &= -\sum_{i=1}^{N}\left[\frac{1}{k_{\eta 2,i}}\left(e_{\eta,i}^{\top}\right)^{\frac{q_3}{q_4}}\left(e_{\eta,i} + k_{\eta 1.i} e_{\eta}^{\frac{p_1}{p_2}}\right)\right]^{\frac{q_4}{q_3}} \\
&= -\sum_{i=1}^{N}\left[\frac{1}{k_{\eta 2,i}}\left(||e_{\eta,i}||_2^2\right)^{\frac{1}{2}\left(1+\frac{q_3}{q_4}\right)} + \frac{k_{\eta 1.i}}{k_{\eta 2,i}}\left(||e_{\eta,i}||_2^2\right)^{\frac{1}{2}\left(\frac{q_3}{q_4}+\frac{q_1}{q_2}\right)}\right]^{\frac{q_4}{q_3}}.
\end{aligned} \tag{39}$$

Using Lemma 3 yields

$$\begin{aligned}
\dot{V}_{\eta 2} &\le -\left\{\sum_{i=1}^{N}\left[\frac{1}{k_{\eta 2,i}}\left(||e_{\eta,i}||_2^2\right)^{\frac{1}{2}\left(1+\frac{q_3}{q_4}\right)} + \frac{k_{\eta 1.i}}{k_{\eta 2,i}}\left(||e_{\eta,i}||_2^2\right)^{\frac{1}{2}\left(\frac{q_3}{q_4}+\frac{q_1}{q_2}\right)}\right]\right\}^{\frac{q_4}{q_3}} \\
&\le -\left[\frac{1}{\underline{k}_{\eta 2}}\left(\sum_{i=1}^{N}||e_{\eta,i}||_2^2\right)^{\frac{1}{2}\left(1+\frac{q_3}{q_4}\right)} + \frac{\underline{k}_{\eta 1}}{\underline{k}_{\eta 2}}\left(\sum_{i=1}^{N}||e_{\eta,i}||_2^2\right)^{\frac{1}{2}\left(\frac{q_3}{q_4}+\frac{q_1}{q_2}\right)}\right]^{\frac{q_4}{q_3}} \\
&= -\left[\kappa_{\eta 1} V_{\eta 2}^{\frac{1}{2}\left(1+\frac{q_3}{q_4}\right)} + \kappa_{\eta 2} V_{\eta 2}^{\frac{1}{2}\left(\frac{q_3}{q_4}+\frac{q_1}{q_2}\right)}\right]^{\frac{q_4}{q_3}},
\end{aligned} \tag{40}$$

where $\bar{k}_{\eta 2} = \max(k_{\eta 1.i}, k_{\eta 2,i}, \ldots, k_{\eta N,i})$, $\kappa_{\eta 1} = 2^{\left(\frac{1}{2}+\frac{q_3}{2q_4}\right)}/\bar{k}_{\eta 2}$, $\underline{k}_{\eta 1} = \min(k_{\eta 1.i}, k_{\eta 2,i}, \ldots, k_{\eta N,i})$, and $\kappa_{\eta 2} = \underline{k}_{\eta 1} 2^{\left(\frac{q_3}{2q_4}+\frac{q_1}{2q_2}\right)}/\bar{k}_{\eta 2}$. Using Lemma 2 indicates that $e_{\eta,i}, i = 1, 2, \ldots, N$ converge to the origin in a fixed-time $\mathcal{T}_{\eta 3}$, which can be bounded by

$$\mathcal{T}_{\eta 3} \le \frac{2q_3}{(q_3 - q_4)\kappa_{\eta 1}^{q_4/q_3}} + \frac{2q_2 q_3}{(q_1 q_4 - q_2 q_3)\kappa_{\eta 2}^{q_4/q_3}}. \tag{41}$$

Therefore, the translational subsystem of the quadrotor group is fixed-time stable, and the settling time can be bounded by $\mathcal{T}_{\eta} \le \mathcal{T}_{\eta 1} + \mathcal{T}_{\eta 2} + \mathcal{T}_{\eta 3}$. The proof is completed. □

Remark 7. The topological graph used in this study is time-invariant and undirected. In practical scenarios, however, directed and time-varying graphs are required, for example, when the information transmission rate is limited. The works proposed in [42,43] provide strong theoretical support for extending from undirected, time-invariant graphs to directed, switching graphs. The results in [42,43] ensure that the formation system remains stable under directed and time-varying topologies. Nevertheless, changes in the topological graph do not affect the DRL component in this paper, because DRL-based parameter optimization is independent of the traditional control framework. Regardless of the specific control method, the control gains can be adaptively tuned in real time by DRL, provided that the underlying control scheme guarantees system stability.

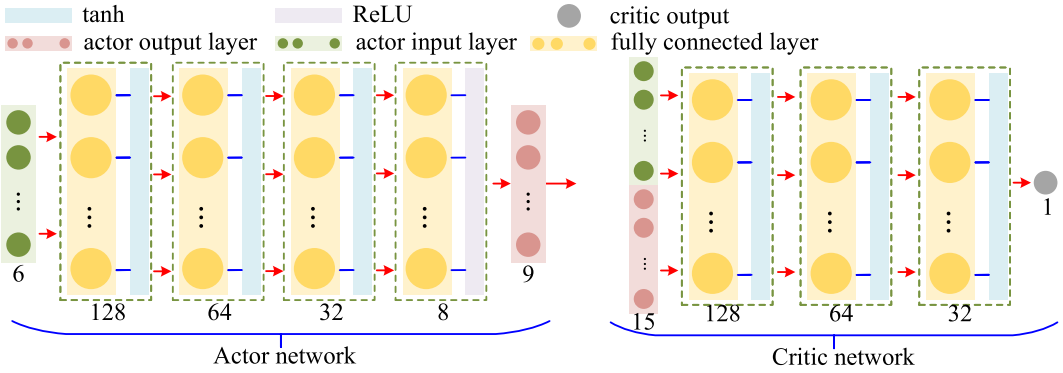


Fig. 1. Architecture of the actor and the critic networks.

Table 2
Some related parameters of the PPO optimizer.

Symbol	Value	Symbol	Value	Symbol	Value	Symbol	Value
T_m	10	dt	0.01	std_0	0.45	$std_{\min}$	0.2
std_d	0.05	std_{dN}	250	γ	0.99	K_{ep}	10
b_s	$T_m/dt * 2$	a_{lr}	10^{-4}	c_{lr}	10^{-3}	c_{en}	0.01
λ	0.95	$c_{\min}$	0.8	$c_{\max}$	1.2	N_m	1000

4. DRL for parameter optimization

In Section 3, a consensus control protocol is designed for the quadrotor formation. However, tuning the hyperparameters remains a critical issue that needs to be addressed. In this section, DRL is utilized as a hyperparameter optimizer for FNTSMCs to achieve improved control performance. The basic DRL algorithm employed in this study is Proximal Policy Optimization (PPO) with Generalized Advantage Estimation (GAE) [44]. PPO is designed to ensure monotonic improvement in the value function throughout the policy learning process, thereby enhancing learning robustness. This property is a key reason why PPO has become widely adopted in control-oriented learning tasks. Table 2 summarizes the relevant parameters of PPO with GAE used in this work.

Algorithm 1 PPO with GAE.

r is the immediate reward

Require: $e_p = 0$, N_e, critic net $C()$, actor net $\mathcal{A}()$.

while $e_p < N_e$ **do**
  Collect data buffer $\mathcal{B}$.
  $V_{gae} = 0$, $V_{adv} = [\ ]$, $\mathcal{A}_{lg} = [\ ]$
  for each $b \in \mathcal{B}$ **do**
    $b = (s, a, a_{lp}, r, s\_done, success)$, $v = C(s), v' = C(s')$, $\delta = r + \gamma(1 - success)v' - v$.
    $V_{gae} = \delta + \gamma\lambda V_{gae}(1 - done)$, $V_{adv}.add(V_{gae})$, $\mathcal{A}_{lg}.add(a_{lp})$
  end for
  Calculate the target value function: $V_{tar} = V_{adv} + v$.
  Calculate the distribution of actor $\mathcal{N}_{ac}$, the entropy of the distribution $\mathcal{N}_{en}$, and the
  Calculate the log-probability of the distribution $\mathcal{N}_{lg}$.
  $\mathcal{N}_a = e^{\mathcal{N}_{lg} - \mathcal{A}_{lg}}$.
  Calculate the surrogate objective: $s_1 = \mathcal{N}_a * V_{adv}$, $s_2 = V_{adv} * [\mathcal{N}_a.clip(c_{\min}, c_{\max})]$.
  Calculate loss function for $\mathcal{A}()$: $\mathcal{L}_a = -min(s_1, s_2) - c_{en} * \mathcal{N}_{en}$.
  Update actor net weights $\mathcal{A}().learn()$.
  Calculate loss function for $C()$: $\mathcal{L}_c = \frac{1}{2}(V_{tar} - V)^2$.
  Update critic net weights $C().learn()$.
  $e_p += 1$.
end while
return $\mathcal{A}()$

The parameter tuning of the DRL-based optimizer is also crucial. Compared to other DRL methods such as Soft Actor-Critic, Deep Deterministic Policy Gradient, and Deep Q-Network, PPO is relatively less sensitive to hyperparameter selection. This robustness to hyperparameters is the main reason for choosing PPO as the training framework. Considering the limited computational capacity of the onboard computer for online deployment, the neural network should be kept lightweight. After extensive tuning and validation, it is recommended that the network consist of 3 to 5 layers, with no more than 128 neurons per layer. In addition to network size, the exploration standard deviation of the policy is a critical parameter. To ensure sufficient exploration while maintaining rapid convergence, the standard deviation is initialized at $std_0 = 0.45$ and gradually reduced by $std_d = 0.05$ every $std_{dN} = 250$ training episodes, with a minimum limit of $std_{\min} = 0.2$. For each training iteration, the data volume should include at least two complete simulation trajectories, each with a duration of $T_m = 10s$ and a sampling period of $dt = 0.01s$. Accordingly, the data buffer should store a minimum of $b_s = T_m/dt * 2 = 2000$ samples. Other parameters largely follow the guidelines proposed in [44].

In Table 2, aside from the parameters mentioned above, N_m denotes the maximum number of learning episodes, γ is the discount factor, K_{ep} represents the number of times the neural network (NN) gradients are updated in a single learning iteration, and b_s is the buffer size. The learning rates of the actor and critic networks are denoted by a_{lr} and c_{lr}, respectively. The parameters c_{en}, λ, $c_{\min}$, and $c_{\max}$ are associated with the GAE technique, as referenced in [44]. In addition, the maximum training episodes are calculated as $N_m = 1000 \cdot \frac{(std_0 - std_{\min})}{std_d} + 1000$. The pseudocode for PPO with GAE is illustrated in Algorithm 1. For more details on the PPO algorithm, readers may refer to [45] and several widely cited GitHub repositories[1].

The architectures of the actor and critic networks are illustrated in Fig. 1. The actor network is a five-layer fully connected neural network, with input and output dimensions of 6 and 9, respectively. The input represents the tracking error of the quadrotor, while the output corresponds to the learned hyperparameters of the FNTSMC. Note that the activation function of the actor network's output layer is ReLU rather than tanh, since the FNTSMC hyperparameters are required to be positive. The critic network has an input dimension of 15, which equals the sum of the input and output dimensions of the actor network. Its output is a scalar representing the state-action value function for the current input, which serves as an indicator to evaluate the quality of the selected hyperparameters. The learning-based control framework for a single quadrotor is shown in Fig. 2. In this framework, the rotational and translational loop controls are coupled through the desired roll and pitch angles, denoted as φ_d and θ_d. The FNTSMC hyperparameters for both loops are optimized simultaneously yet independently by DRL.

[1] https://github.com/HKPolyU-UAV/ReinforcementLearningPlatform/tree/main/algorithm

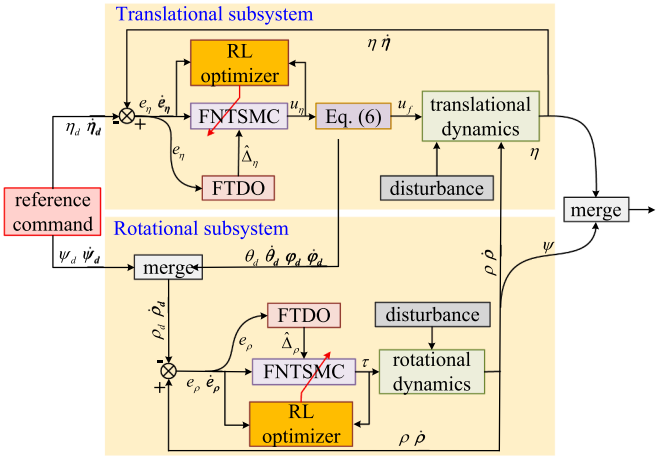


Fig. 2. Diagram of the learning-based control framework.

4.1. Rotational subsystem parameter optimizer training

The controller for the rotational subsystem is utilized in Eq. (21) and is tuned by $k_{\rho 1,i}$, $k_{\rho 2,i}$, $k_{\rho 3,i}$, $k_{\rho 4,i}$, and p_1 to p_6. First, to ensure the rapid convergence of the training process, we select $p_1 = 9$, $p_2 = 7$, $p_3 = 5$, $p_4 = 3$, $p_5 = 7$, and $p_6 = 5$. Second, to reduce the gap between numerical simulations and real-world experiments, we opted not to rely solely on the results learned by DRL. Specifically, the regulation of $k_{\rho 1,i}$, $k_{\rho 2,i}$, and $k_{\rho 4,i}$ is assigned to DRL, while $k_{\rho 3,i}$ is retained to further enhance the robustness of the controller during real-world experiments.

The input and output of the rotational-subsystem optimizer are defined as follows, respectively:

$$\chi_{\rho,i} = \left[e_{\rho,i}^{\top}, \dot{e}_{\rho,i}^{\top}\right]^{\top} \in \mathbb{R}^6, \tag{42}$$

$$\Theta_{\rho,i} = \left[k_{\rho 1,i,x}, k_{\rho 1,i,y}, k_{\rho 1,i,z}, k_{\rho 2,i,x}, k_{\rho 2,i,y}, k_{\rho 2,i,z}, k_{\rho 3,i,x}, k_{\rho 3,i,y}, k_{\rho 3,i,z}\right]^{\top} \in \mathbb{R}^9.$$

The reward function is defined as

$$J_{\rho,i}(t) = -\int_{t=0}^{\infty} e^{-\gamma_\rho (s-t)} \left(\chi_{\rho,i}^{\top} Q_{\rho,i} \chi_{\rho,i} + \tau^{\top} R_{\rho,i} \tau\right) \mathrm{d}s, \tag{43}$$

where $Q_{\rho,i} = \mathrm{diag}(Q_{e_{\rho,i}}, Q_{\dot{e}_{\rho,i}})$ with $Q_{e_{\rho,i}} = \mathbf{I}_3$, $Q_{\dot{e}_{\rho,i}} = 0.01\mathbf{I}_3$, $R_{\rho,i} = 0.01\mathbf{I}_3$, and $\gamma_\rho = 0.99$ is the discount factor.

Remark 8. Note that the DRL training process can be regarded as a mathematical approximation to solving an optimal control problem. While DRL focuses on maximizing rewards, optimal control aims to minimize costs. To reconcile this discrepancy between reward maximization and cost minimization, a negative sign is applied to the integral-form cost function.

The reward curves during the training process are shown in Fig. 3, where a multi-stage training technique [46] is employed to accelerate the training process and enhance the robustness of the trained neural network (NN). In the first training stage, the control performance exhibits significant fluctuations, as the NNs have not yet fully converged. In the subsequent three stages, the initial policies are set based on the results from the corresponding previous stage. Furthermore, lower learning rates are adopted for the NNs to reduce fluctuations and improve the robustness of the learned optimizer.

Moreover, we collected the control costs of the rotational subsystem under various initial conditions and control frameworks to preliminarily demonstrate the superiority of the proposed control framework. During the evaluation process, the reference angular commands are randomly generated as $\rho_d = A\sin(2\pi t/T)$, where $A \in [0, \pi/3]$ is the amplitude of the reference attitude and $T \in [3\text{ s}, 6\text{ s}]$ is the period of the reference signal. This is illustrated in Fig. 3, which shows that the proposed FNTSMC-DRL-FTDO control framework outperforms controllers that do not incorporate DRL. Notably, pure FNTSMC (without FTDO or DRL) exhibits significant fluctuations under strong external disturbances, further highlighting the robustness of the proposed control framework.

Remark 9. To expedite the training of the DRL-based optimizer, the neural networks are initially trained in a single-agent interactive environment, leveraging the homogeneity of the quadrotors. The well-trained NN-based optimizer is then integrated into the FNTSMC of each quadrotor for extensive simulation validation and physical experiments, enabling real-time tuning of the FNTSMC hyperparameters.

4.2. Translational subsystem parameter optimizer training

Similarly, we set $q_1 = 9$, $q_2 = 7$, $q_3 = 5$, $q_4 = 3$, $q_5 = 7$, and $q_6 = 5$. and $k_{\eta 3,i}$ is retained out of the DRL-based optimization framework. The input and output of the optimizer of the translational loop are respectively defined as

$$\chi_{\eta,i} = \left[e_{\eta,i}^{\top}, \dot{e}_{\eta,i}^{\top}\right]^{\top} \in \mathbb{R}^6, \tag{44}$$

$$\Theta_{\eta,i} = \left[k_{\eta 1,i,x}, k_{\eta 1,i,y}, k_{\eta 1,i,z}, k_{\eta 2,i,x}, k_{\eta 2,i,y}, k_{\eta 2,i,z}, k_{\eta 3,i,x}, k_{\eta 3,i,y}, k_{\eta 3,i,z}\right]^{\top} \in \mathbb{R}^9.$$

The reward function is defined as

$$J_{\eta,i}(t) = -\int_{t=0}^{\infty} e^{-\gamma_\eta (s-t)} \left(\chi_{\eta,i}^{\top} Q_{\eta,i} \chi_{\eta,i} + u_{\eta,i}^{\top} R_{\eta,i} u_{\eta,i}\right) \mathrm{d}t, \tag{45}$$

where $Q_{\eta,i} = \mathrm{diag}(Q_{e_{\eta,i}}, Q_{\dot{e}_{\eta,i}})$ with $Q_{e_{\eta,i}} = \mathbf{I}_3$, $Q_{\dot{e}_{\eta,i}} = 0.1\mathbf{I}_3$ and $R_{\eta,i} = 0.01\mathbf{I}_3$, and $\gamma_\eta = 0.99$ is the discount factor.

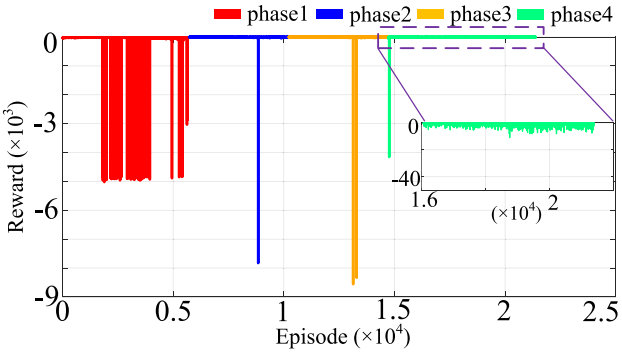
The curves for the reward during the training process are recorded in Fig. 4. The trend of the reward curve is very similar to that of the rotational subsystem. In Fig. 4, the red, blue, orange, and green curves correspond to the four training stages of the attitude-loop controller, with each stage building upon the outcome of the preceding one. Following the approach in [46], partitioning the training into multiple stages proves more effective for achieving robustness than applying a single policy until convergence. This is mainly because, during DRL training, the agent may otherwise become trapped in local optima due to insufficient exploration, or the network may diverge or overfit under an excessively high learning rate. By interrupting training after a specified number of episodes and resuming with the current outcome as the initial point-while simultaneously reinitializing the value-function network, learning rate, and exploration probability-the procedure facilitates the derivation of policies with stronger robustness and improved generalization. As shown in Fig. 4, the network exhibits oscillations during the first three training stages but gradually stabilizes in the final stage. Consequently, the results obtained at the end of the third stage were ultimately deployed in the real-world UAV experiments.

The cost surfaces of the translational subsystem control under different initial conditions and various control frameworks are presented in Fig. 4, which reveals that the patterns of translational control performance are fundamentally similar to those of the rotational loop. The reference trajectory we utilized to test the performance is $\eta_d = A\sin(2\pi t/T)$, where $A \in [0m, 2.5m]$ and $T \in [5s, 8s]$ respectively denote the amplitude and period of the position reference signal. The performance of the proposed FNTSMC-DRL-FTDO control framework outperforms the other three methods. Additionally, the cost of the FNTSMC (represented by the green surface) is lower than that of FNTSMC-DRL (represented by the cyan surface), further indicating the superiority of the DRL technique.

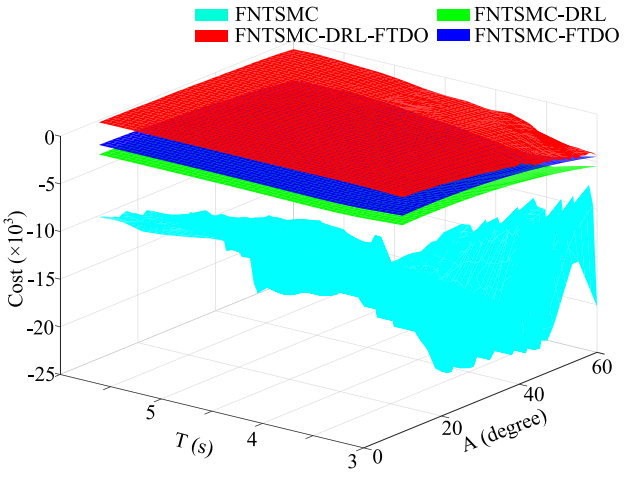
4.3. Stabilization analysis

Adaptively tuning the gains of a controller using DRL provides significant advantages over directly learning the controller itself [30]. Optimizing the parameters through reinforcement learning preserves the structure of the controller and does not compromise the stability of the closed-loop system.

The primary reason for maintaining the stability of the closed-loop system lies in the design of the activation function in the output layer

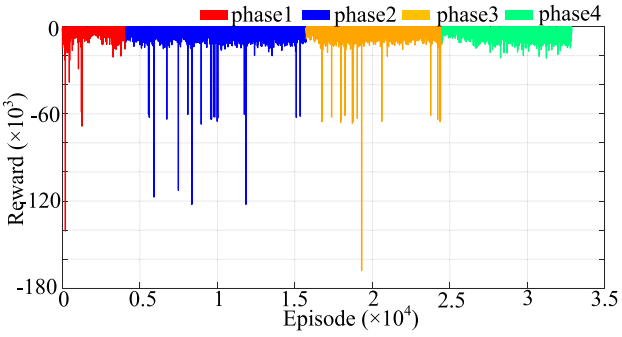


(a) Reward of the training process of the rotational subsystem.

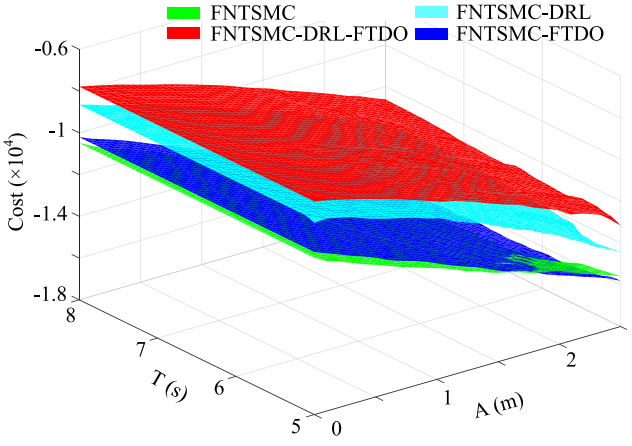


(b) Comparative cost under different control frameworks and different initial conditions for the rotational subsystem.

Fig. 3. Training and evaluation processes of the rotational subsystem. Subfigure (a) illustrates the variation of the reward function with the number of training episodes during the four-stage training process of the rotational-loop controller. The four colors correspond to the four training stages. Subfigure (b) presents a comparison of the reward functions under different initial conditions and methods, tested using the well-trained NN parameter optimizer.



(a) Reward of the training process of the translational subsystem.



(b) Comparative cost under different control frameworks and different initial conditions for the translational subsystem.

Fig. 4. Training and evaluation processes of the translational subsystem. Subfigure (a) illustrates the variation of the reward function with the number of training episodes during the four-stage training process of the translational-loop controller. The four colors correspond to the four training stages. Subfigure (b) presents a comparison of the reward functions under different initial conditions and methods, tested using the well-trained NN parameter optimizer.

of the actor network. As shown in Fig. 1, the output layer employs the ReLU activation function rather than tanh or other alternatives. The use of ReLU restricts the control gains of the FNTSMC to the range $[0, \infty)$. By adding a small positive value ε, the output of the actor network is effectively bounded within $(0, \infty)$, naturally satisfying the requirements of the FNTSMC design. Even if the DRL parameters are not optimally selected, potentially resulting in suboptimal performance of the neural network optimizer, the system stability can still be theoretically guaranteed. This is because the DRL design in this study operates independently of the FNTSMC framework, which is the core principle of the hybrid learning-based control method proposed herein. Therefore, after training, the deployment of the neural networks in the online control framework does not compromise system stability. In the following, guidelines are provided for several important parameters that are closely related to the training stability of the neural networks.

1) Learning rate. The learning rate of the neural network in DRL is a critical parameter that directly influences training robustness. Empirically, the learning rate of the critic network should be maintained within $(10^{-3}, 5\times10^{-3})$, while that of the actor network should remain in $(10^{-4}, 5\times10^{-4})$. If the learning rate is excessively large, the training process may diverge. Conversely, if the learning rate is too small, the control gains of the FNTSMCs tend to remain near zero, exhibiting negligible variation. Therefore, selecting appropriate learning rates is essential for achieving both rapid learning and stable network performance.

2) Exploration policy. In the PPO framework, the policy is defined as a Gaussian policy, where the standard deviation serves as the level of exploration. A common practice is to initialize the standard deviation at a relatively large value and gradually decrease it. If the initial exploration is too conservative, the policy may converge to a suboptimal local minimum. Conversely, if the standard deviation remains excessively large, the policy may fail to converge. Empirically, a range of 0.2 to 0.5 is suitable for the standard deviation of the Gaussian policy.

3) Data volume per learning iteration. PPO is an on-policy DRL algorithm, which requires all buffer data to be utilized during training. If the maximum duration of each trajectory is T_m and the sampling period is $\mathrm{d}t$, the data size for each trajectory is $N_0 = \frac{T_m}{\mathrm{d}t}$. The data buffer capacity should be maintained within the range of $2N_0$ to $4N_0$. If the buffer capacity is too small, the network may fail to capture sufficient information during training; conversely, an excessively large buffer may introduce redundant or irrelevant data, especially during the initial training phase. Therefore, buffer capacity is a crucial factor affecting the training stability of the network.

Algorithm 2 Pseudocode of the control framework.

Require: N, $\mathcal{G}_B$, η_d, $v_i, i = 1, 2, \ldots, N$, T_m, $\mathrm{d}t$
Require: rotational parameter optimizer Act_{att}, translational parameter optimizer Act_{pos}
Load system model, load initial controllers, load observers
Load disturbances $\Delta_{\rho,i}$, $\Delta_{\eta,i}$, $i = 1, 2, \ldots, N$
$t = 0$
while $t < T_m$ **do**
  for $i = 1 \rightarrow N$ **do**
    Use observer (29) to estimate $\Delta_{\eta,i}$
    Calculate $e_{\eta,i}$ by (7) and $\dot{e}_{\eta,i}$ by (8)
    Get control gains for translational FNTSMC (44) with Act_{pos}
    Generate expected virtural input $u_{\eta,i}$ with (33)
    Generate expected attitude and throttle with (6)
    Use observer (14) to estimate $\Delta_{\rho,i}$
    Get control gains for rotational FNTSMC (42) with Act_{att}
    Generate torque inputs with (21)
    Quadrotor state update
  end for
  $t+ = \mathrm{d}t$
end while
Data save and figure plot

5. Simulation

This section conducts some numerical simulations to verify the superiority and effectiveness of the proposed FNTSMC-FTDO-DRL control framework. Building upon the results in Sections 3 and 4, the pseudocode of the entire control framework can be illustrated in Algorithm 2.

5.1. Simulation group 1

The topological graph is shown in Fig. 5. Correspondingly, the adjacent matrix $\mathcal{A}_{s1}$, in-degree matrix $\mathcal{D}_{s1}$, communication matrix $\mathcal{B}_{s1}$, and Laplacian matrix $\mathcal{L}_{s1}$ are respectively defined as

$$\mathcal{A}_{s1} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathcal{D}_{s1} = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathcal{B}_{s1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{46}$$

and $\mathcal{L}_{s1} = \mathcal{D}_{s1} - \mathcal{A}_{s1}$. The equation of the geometric center $\mathcal{O}_d$ is defined as

$$x_d = r_d \sin(0.2\pi t) + 2, \ y_d = r_d \cos(0.2\pi t) + 3, \ z_d = \sin(0.4\pi t) + 2 \tag{47}$$

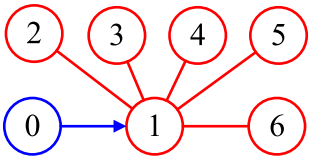


Fig. 5. Topological graph of simulation 1. The red line segment means the information can transmit between two nodes. The blue arrow indicates that information can only be transmitted from the tail node to the head node. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

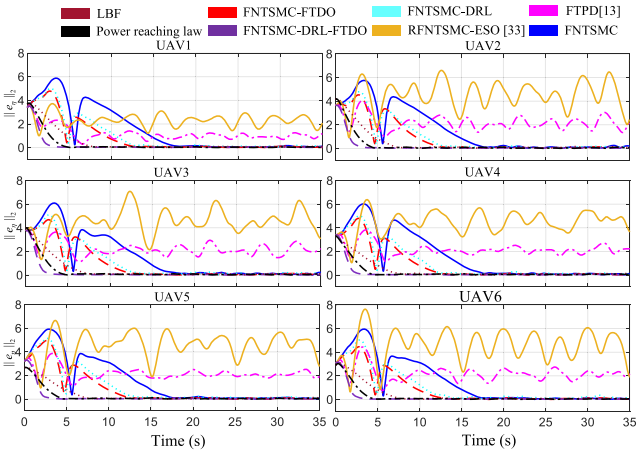


Fig. 6. $||e_{\eta,i}||_2$ of each quadrotor under different control frameworks in simulation 1.

with $r_d = 5m$. The offsets of each quadrotor to $\mathcal{O}_d$, denoted by $v_i, i = 1, 2, \ldots, 6$, are defined as

$$v_{i,x} = r_v \sin(0.2\pi t + \phi_{x,i}), \ v_{i,y} = r_v \sin(0.2\pi t + \phi_{y,i}), \ v_{i,z} = 0, \tag{48}$$

where $\phi_{x,i} = \frac{\pi}{2} + (i-1)\frac{\pi}{3}$, $\phi_{y,i} = (i-1)\frac{\pi}{3}$ and $r_v = 2m$.

Fig. 6 illustrates the 2-norm of the tracking errors under six different control frameworks. It is evident that the purple curve demonstrates the best performance. The red curve also eventually converges to the origin; however, in the absence of an adaptive parameter adjustment mechanism, significant fluctuations are observed during the initial phase. Other controllers maintain system stability but consistently exhibit comparatively larger tracking errors.

We further conduct a simulation in which the reference trajectories on the XOY plane remain fixed, while the altitude is set to different constant values. The corresponding position response of the quadrotor group under the "FNTSMC + DRL + FTDO" control framework is shown in Fig. 7, where the reference trajectories consist of the superposition of two circles with different radii. The disturbances in the simulation are modeled as combinations of sine functions with varying amplitudes, phases, and periods. Mass uncertainties are applied only to the first four quadrotors. Specifically, for $i = 1, 2, 3, 4$, $\Delta m_i = -0.2kg$ when $0 < t \leq 10$, $\Delta m_i = 0.2kg$ when $10 < t \leq 20$, and $\Delta m_i = 0kg$ when $30 < t \leq 40$.

The output of the FTDOs is presented in Fig. 8. As shown, the proposed observer converges within approximately 1 second and accurately estimates the external disturbances. In particular, the red rectangles highlight sudden changes in disturbances caused by abrupt mass variations in quadrotors 1 to 4. The plotted curves demonstrate that the observer can rapidly track the desired signals even under such abrupt changes, confirming its effectiveness and robustness. These results validate the fast convergence and strong robustness of the proposed observer against abrupt disturbances.

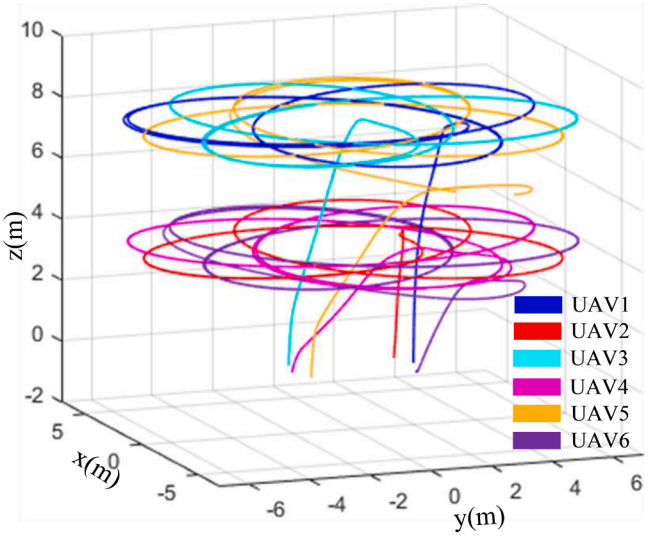


Fig. 7. Graphic demonstration of the quadrotor formation in a 3D view of simulation 1.

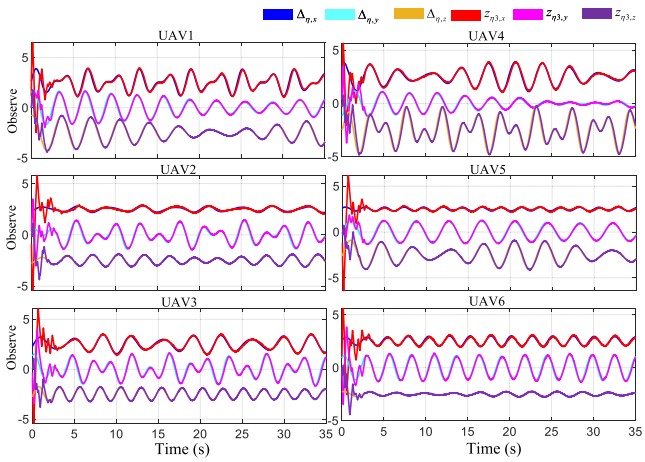


Fig. 8. Output of the observers in simulation 1.

5.2. Simulation group 2

We further test our algorithm with a more complicated topological graph and a more aggressive reference trajectory. The topological graph of simulation group 2 is shown in Fig. 9. Correspondingly, the adjacency matrix $\mathcal{A}_{s2}$, in-degree matrix $\mathcal{D}_{s2}$, communication matrix $\mathcal{B}_{s2}$, and Laplacian matrix $\mathcal{L}_{s2}$ are defined as

$$\mathcal{A}_{s2}=\begin{bmatrix}0&1&0&1&0&0\\1&0&1&0&0&0\\0&1&0&1&0&0\\1&0&1&0&1&0\\0&0&0&1&0&1\\0&0&0&0&1&0\end{bmatrix},\ \mathcal{D}_{s2}=\begin{bmatrix}2&0&0&0&0&0\\0&2&0&0&0&0\\0&0&2&0&0&0\\0&0&0&3&0&0\\0&0&0&0&2&0\\0&0&0&0&0&1\end{bmatrix},$$

$$\mathcal{B}_{s2}=\begin{bmatrix}1&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&0&0\end{bmatrix}, \tag{49}$$

and $\mathcal{L}_{s2}=\mathcal{D}_{s2}-\mathcal{A}_{s2}$. The equation of the reference trajectory of the geometric center $\mathcal{O}_d$ is defined as

$$x_d=r_d\cos(0.2\pi t)+2,\ y_d=r_d\sin(0.4\pi t)+3,\ z_d=\sin(0.4\pi t)+2 \tag{50}$$

with $r_d=5m$. The offsets of each quadrotor to $\mathcal{O}_d$, denoted by $v_i, i=1,2,\dots,6$, are defined as

$$\begin{aligned}v_1&=[r_v,0,0]^\top, v_2=[r_v\sin(\theta_0),r_v\cos(\theta_0),0]^\top,\\ v_3&=[-r_v\sin(\theta_0),r_v\cos(\theta_0),0]^\top,\\ v_4&=[-r_v,0,0]^\top, v_5=[-r_v\sin(\theta_0),-r_v\cos(\theta_0),0]^\top,\\ v_6&=[r_v\sin(\theta_0),-r_v\cos(\theta_0),0]^\top,\end{aligned} \tag{51}$$

where $\theta_0=60°$ and $r_v=2m$.

Fig. 10 shows the 2-norm of the tracking errors under different control frameworks. The corresponding three-dimensional position response of the quadrotor group is presented in Fig. 11, while the outputs of the FTDOs are shown in Fig. 12. Similar to the results in Figs. 6 and 10, the control performances of FNTSMC + DRL (cyan curves) and FNTSMC + FTDO (red curves) are slightly better than that of the traditional FNTSMC (blue curves), although the tracking errors remain relatively large. In contrast, under the FNTSMC + DRL + FTDO framework, the tracking errors converge rapidly to zero for both the double-circle and ∞-shaped reference trajectories.

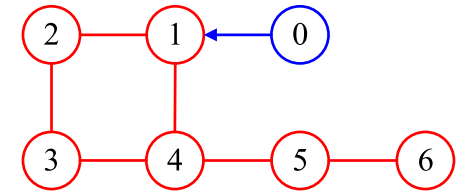


Fig. 9. Topological graph of simulation group 2. The red line segment means the information can transmit between two nodes. The blue arrow indicates that information can only be transmitted from the tail node to the head node. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

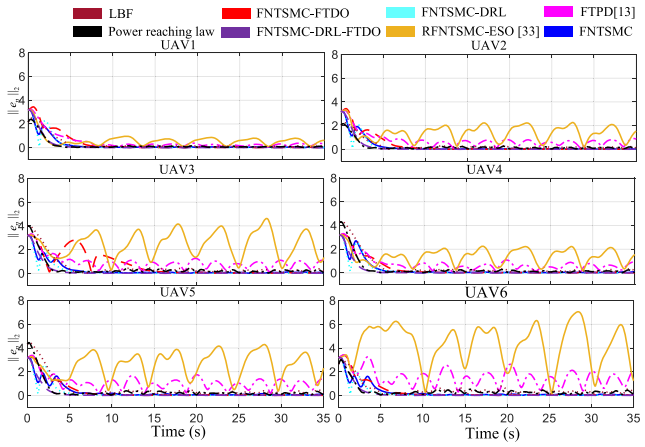


Fig. 10. $||e_\eta||_2$ of each quadrotor under different control frameworks in simulation 2.

To clearly highlight the performance differences among the various control frameworks, we summarize the L_1 and L_2 norms of the consensus tracking errors in the two simulation scenarios. Table 3 provides a detailed comparison of the performance of each control method. The convergence times of the different methods in the two sets of simulations are listed in Table 4. As shown in Tables 3 and 4, the proposed method achieves the fastest convergence and exhibits the smallest control errors under strong disturbances.

Table 3
Tracking errors of the quadrotor formation under two groups of simulations.

Simulation Group 1								
	①	②	③	④	⑤	⑥	⑦	Proposed
L_2-norm: $\sum_i \int \|\|e_{\eta,i}\|\|_2 (\times 10^5)$	0.79	0.44	0.51	0.91	1.86	0.28	0.17	0.09
L_1-norm: $\sum_i \int \|\|e_{\eta,i}\|\|_1 (\times 10^5)$	0.91	0.54	0.62	1.23	2.45	0.20	0.14	0.12
MSE:[a] $\frac{1}{T}\sum_i \int \|\|e_{\eta,i}\|\|_2^2$	249.35	113.17	132.40	171.01	703.91	35.62	21.25	15.45
EV:[a] $\frac{1}{T}\sum_i \int \|\|e_{\eta,i} - \mu_{e,i}\|\|_2^2$	134.73	77.13	84.76	12.82	39.53	27.89	18.37	14.05
Simulation Group 2								
	[b]①	[b]②	③[b]	④[b]	⑤[b]	⑥[b]	⑦[b]	Proposed
L_2-norm: $\sum_i \int \|\|e_{\eta,i}\|\|_2 (\times 10^4)$	1.30	1.95	1.05	4.24	9.77	2.16	1.47	0.81
L_1-norm: $\sum_i \int \|\|e_{\eta,i}\|\|_1 (\times 10^5)$	0.18	0.27	0.15	0.59	1.41	0.33	0.22	0.12
MSE: $\frac{1}{T}\sum_i \int \|\|e_{\eta,i}\|\|_2^2$	16.74	31.54	15.36	52.28	268.27	31.25	18.65	13.12
EV: $\frac{1}{T}\sum_i \int \|\|e_{\eta,i} - \mu_{e,i}\|\|_2^2$	13.60	23.80	13.32	14.76	35.54	22.19	14.43	11.89

[a] MSE: Mean Square Error, EV: Error Variance.
[b] ①: FNTSMC ②: FNTSMC-FTDO ③: FNTSMC-DRL ④: [33] ⑤: [13] ⑥: LBF-based ⑦: Power-rate reaching law

Table 4
The convergence time of different control methods.

Simulation Group 1								Simulation Group 2							
[a]①	[a]②	[a]③	[a]④	[a]⑤	[a]⑥	[a]⑦	Proposed	[a]①	[a]②	[a]③	[a]④	[a]⑤	[a]⑥	[a]⑦	Proposed
18.2s	12.5s	12.5s	6.5s	−−	2.5s	7.6s	5.0s	7.5s	10.0s	7.5s	8.1s	−−	5.2s	5.6s	5.0s

[a] ①: FNTSMC ②: FNTSMC-FTDO ③: FNTSMC-DRL ④: [33] ⑤: [13] ⑥: LBF-based ⑦: Power-rate reaching law.

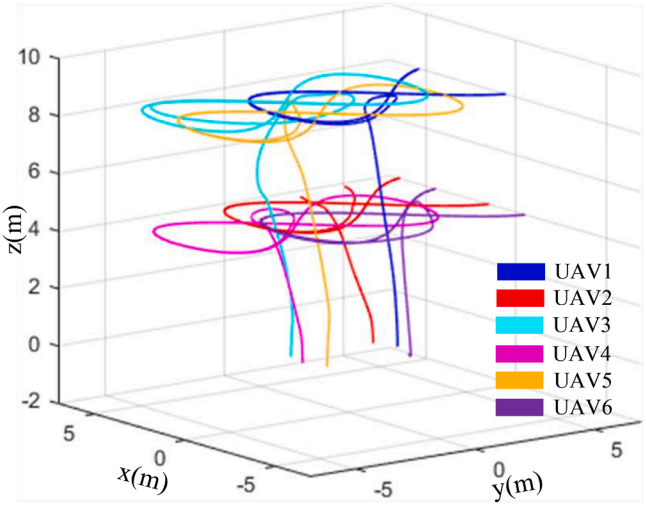


Fig. 11. Graphic demonstration of the quadrotor formation in a 3D view of simulation 2.

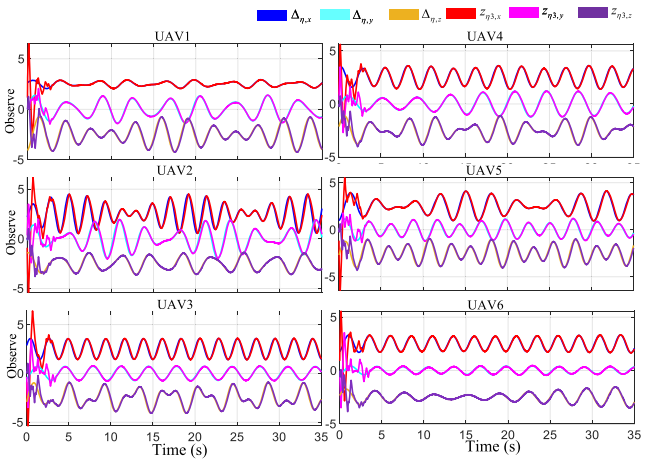


Fig. 12. Output of the observers in simulation 2.

6. Physical experiments

This section presents real-world experiments to further validate the proposed control framework. The overall experimental setup is illustrated in Fig. 13, with the quadrotors used in the experiments shown in the bottom part of the figure. Each quadrotor has a mass of 0.722 kg and a wheelbase of 250 mm. To ensure safety during the experiments, the maximum velocity and acceleration of the quadrotors are limited to 3 m/s and $2g \approx 19.6$ m/s^2, respectively. Four quadrotors are employed to demonstrate consensus formation control. To evaluate the robustness of the proposed method under strong disturbances, two types of external perturbations are introduced in the experimental environment: high-speed rotating fans and weights suspended by elastic ropes. Moreover, aerodynamic interactions among quadrotors in formation, as well as discrepancies between each quadrotor' s center of mass and geometric centroid, further complicate flight control. The complete hardware configuration of the experimental environment is depicted in the top part of Fig. 13. Four self-designed quadrotors, each equipped with distinct night-light colors, are used to form the formation. The flight control unit (FCU) of each quadrotor is the Holybro Kakute H7 v1.3, which runs the open-source PX4-Autopilot firmware. The on-board computer is a LattePanda Alpha 864s running Ubuntu 20.04 with ROS Noetic. The FCU communicates with the on-board computer via a USB-to-TTL module and utilizes the MAVROS protocol for real-time data transmission. Quadrotor localization is provided by a VICON indoor positioning system equipped with 14 high-resolution optical cameras, delivering precise position and attitude feedback at a frequency exceeding 200 Hz. Velocity feedback is obtained by fusing VICON measurements with data from the Inertial Measurement Unit (IMU) using a Kalman filter. The detailed hardware configuration of the quadrotors is summarized in Table 5. The ground station computer is used solely for monitoring the quadrotor states and collecting experimental data; it does not send any control-related commands, as the proposed control protocol is fully distributed and decentralized.

6.1. Experiment group 1

The topological graph of the quadrotor group is shown in Fig. 14. The adjacent matrix $\mathcal{A}_{p1}$, in-degree matrix D_{p1}, communication matrix

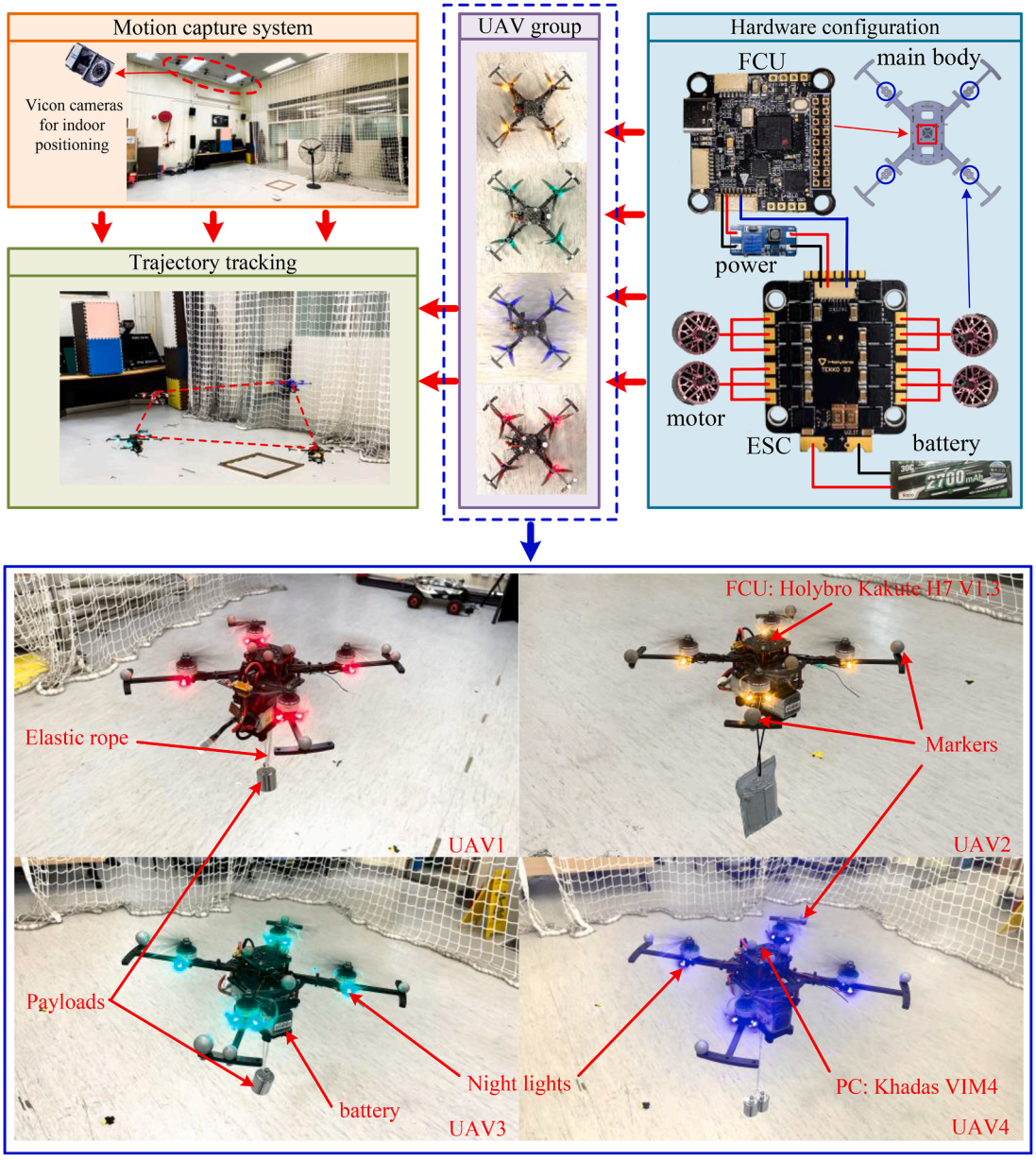


Fig. 13. The entire experiment configuration. The flight formation consists of four quadrotors. Each quadrotor is composed of a frame, flight controller, onboard computer, propulsion system, and a safety remote controller.

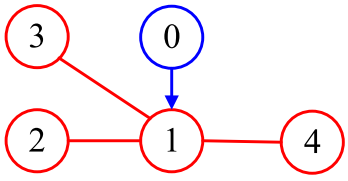


Fig. 14. Topological graph in Experiment Group 1. The red line segment means the information can transmit between two nodes. The blue arrow indicates that information can only be transmitted from the tail node to the head node. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Hardware configuration of the quadrotors.

Hardware	Configuration	Hardware	Configuration
Mass of the quadrotor	$0.722kg$	On-board computer	LattePanda Alpha 864s
FCU	Holybro Kakute H7 v1.3	Firmware of the FCU	PX4-Autopilot
Operating system	Ubuntu 20.04 - ROS Noetic	IMU	BMI270 (integrated in the FCU)
Positioning system	VICON	Motor	VELOX V2207 KV1750
Battery	4S 2700mAh	Propeller	5.5-inch three-blade propeller
Wi-Fi transmission delay	$4-8ms$	ESC	Tekko32 F4 4in1 $60A$ ESC (AM32)

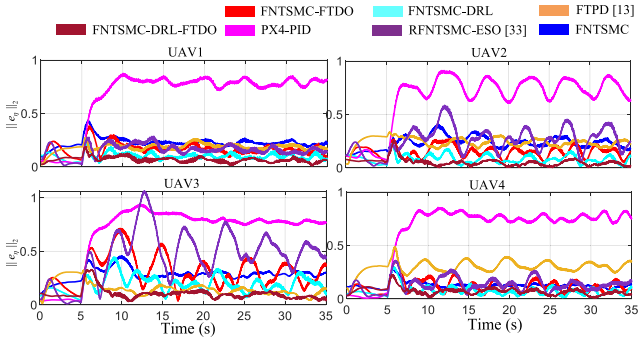


Fig. 15. $||e_\eta||_2$ of each quadrotor under different control frameworks.

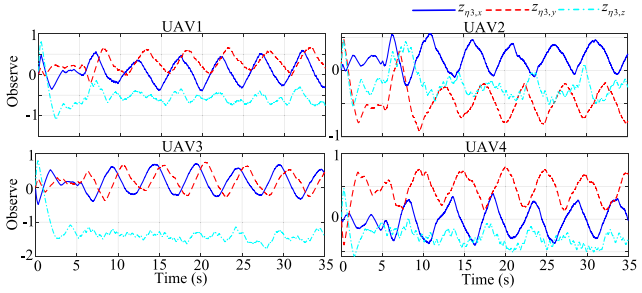


Fig. 17. Output of the observers in experiment group 1.

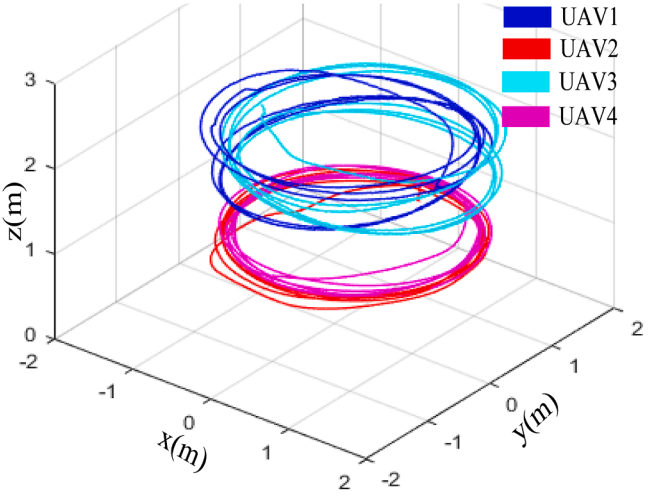


Fig. 16. Graphic demonstration of the quadrotor formation in a 3D view.

$\mathcal{B}_{p1}$, and Laplacian matrix $\mathcal{L}_{p1}$ are respectively defined as

$$\mathcal{A}_{p1}=\begin{bmatrix}0&1&1&1\\1&0&0&0\\1&0&0&0\\1&0&0&0\end{bmatrix},\ \mathcal{D}_{p1}=\begin{bmatrix}3&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix},\ \mathcal{B}_{p1}=\begin{bmatrix}1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}, \tag{52}$$

and $\mathcal{L}_{p1}=\mathcal{D}_{p1}-\mathcal{B}_{p1}$. The geometric center $\mathcal{O}_d=[0,0.2,1.5]^\top$ remains unchanged. The offsets of each quadrotor to $\mathcal{O}_d$, denoted by $v_i, i=1,2,3,4$, are defined as

$$\begin{aligned}v_1&=[1.3\cos(0.4\pi t),1.3\sin(0.4\pi t),0.3\sin(0.2\pi t)+0.5]^\top,\\ v_2&=[-1.3\sin(0.4\pi t),1.3\cos(0.4\pi t),-0.5]^\top,\\ v_3&=[-1.3\cos(0.4\pi t),-1.3\sin(0.4\pi t),0.3\sin(0.2\pi t)+0.5]^\top,\\ v_4&=[1.3\sin(0.4\pi t),-1.3\cos(0.4\pi t),-0.5]^\top.\end{aligned}$$

Fig. 15 illustrates the 2-norm of the consensus tracking errors under different control frameworks, while the corresponding 3D position response is shown in Fig. 16. The response of the PID controller is relatively smooth; however, it exhibits noticeable steady-state errors and phase delays. In contrast, the proposed FNTSMC-DRL-FTDO method outperforms all other controllers. Benefiting from the integration of the FTDO and the DRL-based parameter optimizer, the proposed framework demonstrates superior control performance under strong disturbances and uncertainties. As highlighted in Fig. 15, the proposed method stabilizes the system within a short duration while avoiding overshoot or oscillations, even in the presence of external disturbances. Fig. 18 further illustrates that the control gains are time-varying rather than fixed pre-tuned constants. The gains are initialized at 0 and converge to constant values in less than 2 seconds. According to Eq. (44), this convergence is primarily attributed to the convergence of the consensus tracking errors e_η and $\dot{e}_\eta$.

Fig. 17 presents the corresponding outputs of the FTDO. As observed, the outputs in the x and y directions exhibit noticeable oscillations, with periods closely matching those of the reference trajectories for the quadrotor formation. This periodic behavior results from the influence of the fans, which introduce external disturbances to the quadrotors. In the z direction, the FTDO output reflects an equivalent weight that closely aligns with the masses suspended beneath the quadrotors, further demonstrating the effectiveness of the observer. Fig. 18 records the real-time control gains $k_{\eta 1}$, $k_{\eta 2}$, and $k_{\eta 4}$, which are adaptively tuned by the DRL-based optimizer.

Remark 10. It should be noted that the DRL-based parameter optimizer employed in this study follows an *offline training-online deployment* paradigm. Specifically, the neural network (NN) optimizer is first trained offline in a simulation environment. Subsequently, the trained NN is deployed across all quadrotors for online execution. During each control cycle, the program simultaneously adjusts the controller parameters and generates the corresponding control commands. Although the offline training process may be time-consuming, the online execution is highly efficient. In particular, on the LettaPanda Alpha 864s processor used in our system, the execution time for a single run of the NN optimizer is less than $5ms$, which ensures that the real-time performance of the system is not compromised.

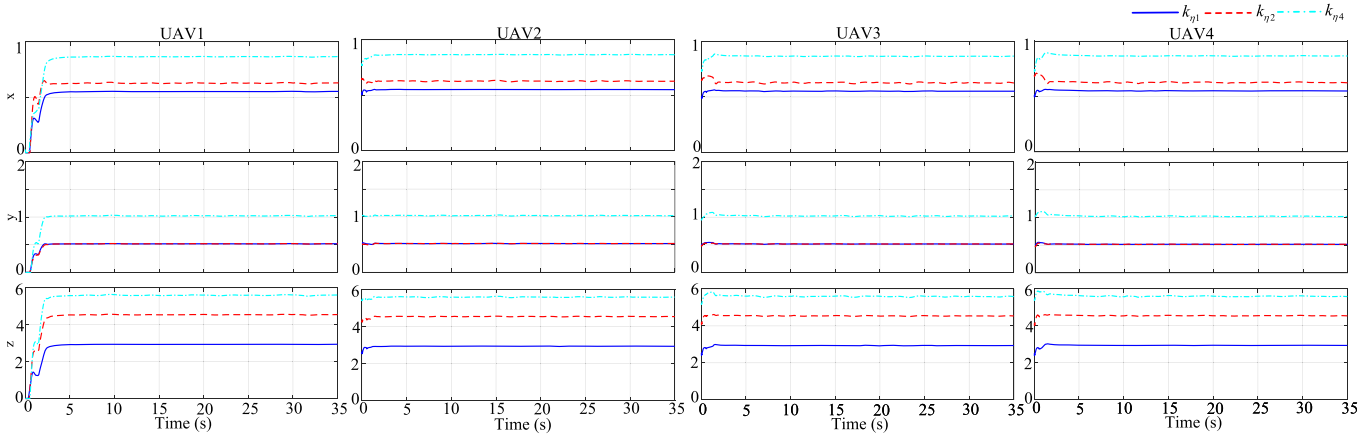


Fig. 18. Hyperparameters $k_{\eta 1}$, $k_{\eta 2}$, and $k_{\eta 4}$ tuned by DRL in experiment group 1. During the actual experiments, these parameters were optimized separately along the x, y, and z directions.

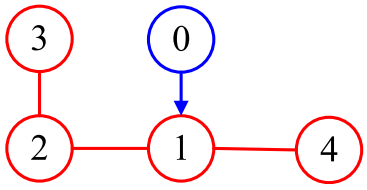


Fig. 19. Topological graph in Experiment Group 2. The red line segment means the information can transmit between two nodes. The blue arrow indicates that information can only be transmitted from the tail node to the head node. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

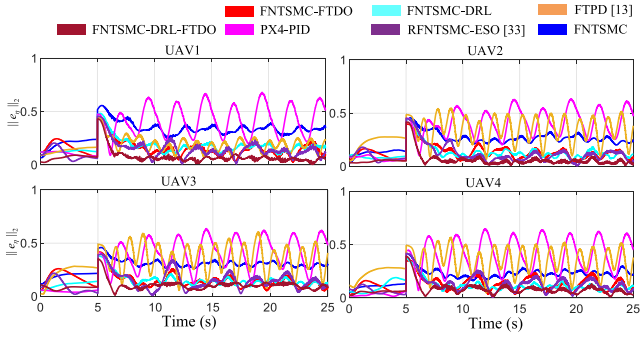


Fig. 20. $||e_\eta, i||_2$ of each quadrotor under different control frameworks.

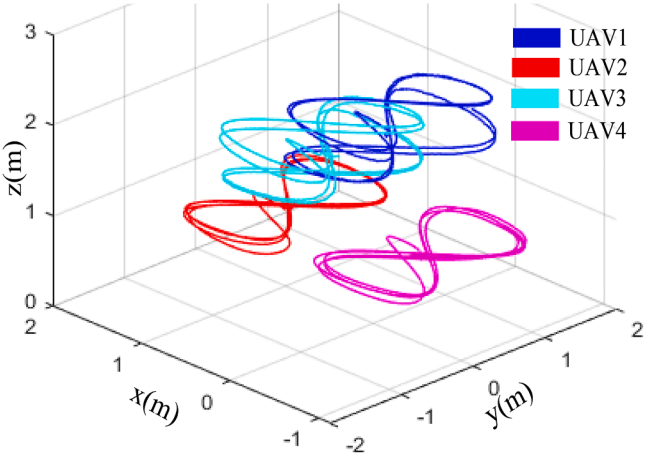


Fig. 21. Graphic demonstration of the quadrotor formation in a 3D view.

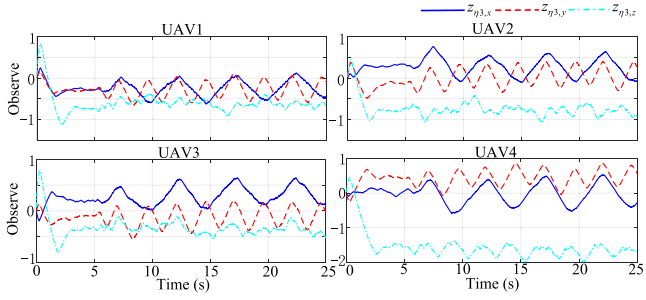


Fig. 22. Output of the observers in experiment group 2.

6.2. Experiment group 2

Furthermore, we employed a more aggressive reference trajectory to test the performance of the quadrotor formation under more extreme environmental conditions. The topological graph of the quadrotor group is shown in Fig. 19. The adjacent matrix $\mathcal{A}_{p2}$, in-degree matrix $\mathcal{D}_{p2}$, communication matrix $\mathcal{B}_{p2}$, and Laplacian matrix $\mathcal{L}_{p2}$ are respectively defined as

$$\mathcal{A}_{p2} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \ \mathcal{D}_{p2} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \ \mathcal{D}_{p2} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{53}$$

and $\mathcal{L}_{p2} = \mathcal{D}_{p2} - \mathcal{B}_{p2}$. The equation for the geometric center $\mathcal{O}_d$ is defined as

$$x_d = \cos(0.4\pi t), \ y_d = \sin(0.8\pi t) + 0.2, \ z_d = 1.5. \tag{54}$$

The offsets of each quadrotor to $\mathcal{O}_d$, denoted by $v_i, i = 1, 2, 3, 4$, are defined as

$$\begin{aligned} v_1 &= [0.5, 0, 0.3\sin(0.2\pi t) + 0.5]^\top, v_2 = [0, 0.8, -0.5]^\top, \\ v_3 &= [-0.5, 0, 0.3\sin(0.2\pi t) + 0.5]^\top, v_4 = [0, -0.8, -0.5]^\top. \end{aligned} \tag{55}$$

Similarly, the gains tuned by DRL, the 2-norm of the tracking errors, the position response in a 3D view, and the corresponding output of the FTDO are recorded in Figs. 20–23, respectively. The error statistical curves from Experiment 2 exhibit the same pattern as those from Experiment 1, indicating that the proposed control framework consistently outperforms other methods. Specifically, we calculated the sum of the L_2 and L_1 norms of the consensus errors for the quadrotor formation, as presented in Table 6, which demonstrates the superiority of the FNTSMC-DRL-FTDO method. Finally, as shown in Fig. 22, the oscillation frequency of the observer output curve in the y direction is

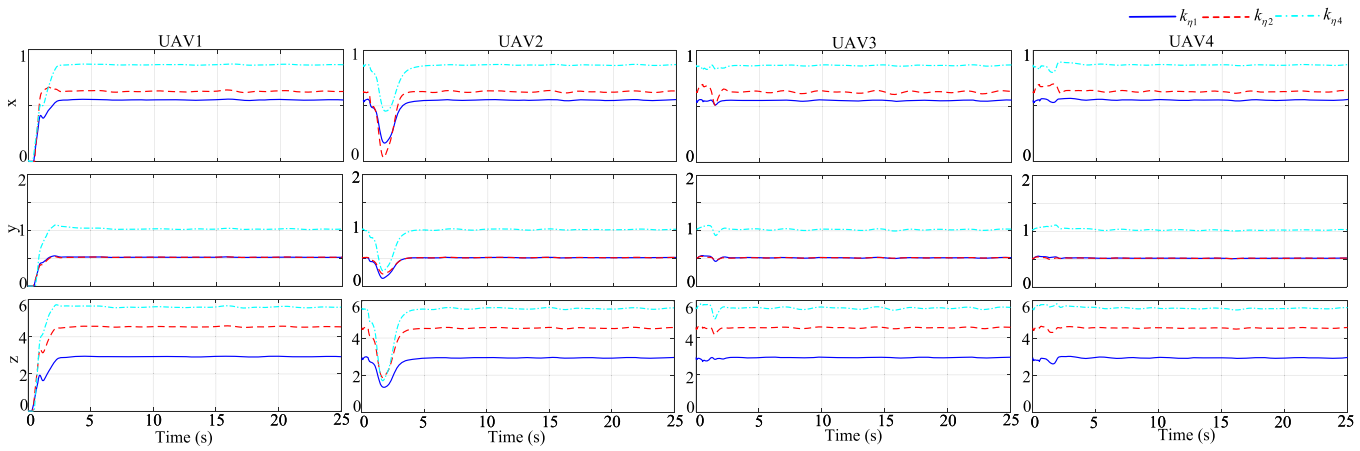


Fig. 23. Hyperparameters $k_{\eta 1}$, $k_{\eta 2}$, and $k_{\eta 4}$ tuned by DRL in experiment group 2. During the actual experiments, these parameters were optimized separately along the x, y, and z directions.

Table 6
Tracking errors of the quadrotor formation under two groups of experiments.

Experiment Group1							
	①	②	③	④	⑤	⑥	Proposed
L_2-norm: $\sum_i \int \|\|e_{\eta,i}\|\|_2 (\times 10^4)$	0.7699	0.6668	0.3946	2.3050	0.7705	0.8876	0.2606
L_1-norm: $\sum_i \int \|\|e_{\eta,i}\|\|_1 (\times 10^4)$	1.1409	0.9841	0.5905	3.0267	1.1450	1.2361	0.3897
MSE:[a] $\frac{1}{T}\sum_i \int \|\|e_{\eta,i}\|\|_2^2$	1.3516	1.2391	0.4388	12.6218	1.3572	2.5000	0.1808
EV:[a] $\frac{1}{T}\sum_i \int \|\|e_{\eta,i} - \mu_{e,i}\|\|_2^2$	0.0615	0.1987	0.0640	1.7743	0.0653	0.4766	0.0340
Experiment Group2							
	①[b]	②[b]	③[b]	④[b]	⑤[b]	⑥[b]	Proposed
L_2-norm: $\sum_i \int \|\|e_{\eta,i}\|\|_2 (\times 10^3)$	6.6189	3.5926	3.5141	8.5403	7.3302	3.2625	1.9650
L_1-norm: $\sum_i \int \|\|e_{\eta,i}\|\|_1 (\times 10^4)$	0.9820	0.5438	0.5319	1.1355	1.0839	0.4707	0.2926
MSE: $\frac{1}{T}\sum_i \int \|\|e_{\eta,i}\|\|_2^2$	1.9232	0.6389	0.6055	3.7708	2.5496	0.5862	0.2336
EV: $\frac{1}{T}\sum_i \int \|\|e_{\eta,i} - \mu_{e,i}\|\|_2^2$	0.1279	0.1193	0.1006	0.8509	0.3004	0.1579	0.0741

[a] MSE: Mean square error, EV: Error variance.
[b] ①: FNTSMC ②: FNTSMC-FTDO ③: FNTSMC-DRL ④: PX4-PID ⑤: [33] ⑥: [13].

approximately twice that of the x direction, which aligns perfectly with the characteristics of the pre-defined reference trajectory.

7. Conclusions

This paper presents a novel robust control framework for quadrotor formations that integrates FNTSMCs, DRL techniques, and FTDOs. First, FNTSMCs are designed for the closed-loop system to ensure rapid and robust tracking. Next, FTDOs are employed to accurately estimate the uncertainties and external disturbances affecting the quadrotors within a fixed time. The observer outputs are then incorporated into the switching control laws, enhancing controller robustness and improving overall control performance. The fixed-time stability of the multi-agent system is rigorously guaranteed in the Lyapunov sense. To further improve performance, DRL is used to train a parameter optimizer that adaptively tunes selected hyperparameters of the FNTSMCs based on the quadrotors' tracking errors. Finally, extensive simulations and physical experiments are conducted to validate the effectiveness and robustness of the proposed framework. Our future work will focus on extending this framework to multi-quadrotor formation systems with communication delays and switching topologies.

CRediT authorship contribution statement

Yefeng Yang: Writing – review & editing, Writing – original draft, Software, Methodology; **Kang Liu:** Validation, Supervision; **Li-Yu Lo:** Visualization, Validation; **Tao Huang:** Writing – review & editing, Methodology; **Yanming Fu:** Supervision; **Chih-Yung Wen:** Supervision, Project administration, Funding acquisition.

Data availability

No data was used for the research described in the article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study was supported by the Research Center of Unmanned Autonomous Systems, the Hong Kong Polytechnic University (P0046487), Research Centre for Low Altitude Economy, the Hong Kong Polytechnic University (P0054124), and the Natural Science Foundation of China under grant number (61104059).

Appendix A. Abbreviations and mathematical notations

Abbreviations:

SMC	Sliding Mode Control
FNTSMC	Fast Nonsingular Terminal Sliding Mode Control
FTDO	Fixed-Time Disturbance Observer
DRL	Deep Reinforcement Learning
ADP	Adaptive Dynamic Programming
NN	Neural Network
PPO	Proximal Policy Optimization
TRPO	Trust Region Policy Optimization
GAE	Generalized Advantage Estimation
FCU	Flight Control Unit
IMU	Inertial Measurement Unit
ESC	Electronic Speed Controller

Notations related to quadrotor mathematical model:

$\eta = [x, y, z]^\top$	Position of the quadrotor in world frame
$\rho = [\phi, \theta, \psi]^\top$	Attitude of the quadrotor in world frame
$\omega = [p, q, r]^\top$	Angular rate of the quadrotor in body frame
k_t, k_r	Translational and rotational drag coefficients
$J = \mathbf{diag}(J_{xx}, J_{yy}, J_{zz})$	Inertia tensor matrix of the quadrotor
m	Mass of the quadrotor
g	Gravitational acceleration
$\tau = [\tau_x, \tau_y, \tau_z]^\top$	Torque
u_f	Throttle
δ_ρ	Disturbance acted on the rotational subsystem of the quadrotor
δ_η	Disturbance acted on the translational subsystem of the quadrotor
$\rho_d = [\phi_d, \theta_d, \psi_d]^\top$	The expected attitude angle
η_d	Expected geometric center of the quadrotor formation
ν_i	The offset between the i-th quadrotor and η_d

Notations related to topological graph:

N	Number of the quadrotors
$\mathcal{G}$	Graph that represents the quadrotor formation
$\mathcal{V}$	Node set of graph $\mathcal{G}$
v_i	The i-th node in node set $\mathcal{V}$
$\mathcal{E}$	Edge set of graph $\mathcal{G}$
$\mathcal{N}_i$	The neighbor set of v_i
$\mathcal{A}$	The adjacency matrix
a_{ij}	The i-th row, j-th column element of $\mathcal{A}$
d_i	$d_i = \sum_{j=1}^{N} a_{ji}$
$\mathcal{D} = \mathbf{diag}(d_1, d_2, \cdots, d_N)$	In-degree matrix
b_i	The information accessment between the i-th node and the leader node
$\mathcal{B}$	Leader adjacency matrix (or communication matrix)
v_b	Leader node
$\mathcal{G}_B = (\mathcal{V}, \mathcal{E}, v_b, b_i), i = 1, 2, \cdots, N$	The augmented graph

Notations related to controller design and DRL:

$p_1, p_2, p_3, p_4, p_5, p_6$	Constants in rotational subsystem control design
$k_{\rho 1}, k_{\rho 2}, k_{\rho 3}, k_{\rho 4}$	Control gains in rotational subsystem control design
$q_1, q_2, q_3, q_4, q_5, q_6$	Constants in translational subsystem control design
$k_{\eta 1}, k_{\eta 2}, k_{\eta 3}, k_{\eta 4}$	Control gains in translational subsystem control design
T_m	Maximum simulation time of an episode
$\mathrm{d}t$	Sampling period
std_0	Initial standard deviation of the exploration policy
std_{min}	Minimum standard deviation of the exploration policy
std_d	Step of the decrease of standard deviation of the exploration policy
std_{dN}	The number of learning episodes after each reduction of standard deviation
γ	Discount factor
K_{ep}	Number of NN gradient descends in one learning iteration
b_s	Buffer size
a_{lr}, c_{lr}	Learning rate of the Actor network and Critic network
$c_{en}, \lambda, c_{min}, c_{max}$	Parameters used in GAE technique

References

[1] D. Zhang, H. Duan, Variational resolution probability maps for multi-UAV search tasks with changeable searching behaviors, Aerosp. Sci. Technol. 155 (2024) 109669.

[2] K.D. Julian, M.J. Kochenderfer, Distributed wildfire surveillance with autonomous aircraft using deep reinforcement learning, J. Guid. Control Dyn. 42 (8) (2019) 1768–1778.

[3] K. Hovell, S. Ulrich, M. Bronz, Learned multiagent real-time guidance with applications to quadrotor runway inspection, J. Field Robot. 2 (1) (2022) 1105–1133.

[4] S. Barawkar, M. Kumar, M. Bolender, Decentralized adaptive controller for multi-drone cooperative transport with offset and moving center of gravity, Aerosp. Sci. Technol. 145 (2024) 108960.

[5] N. Hao, F. He, Y. Hou, W. Song, D. Xu, Y. Yao, Decentralized cooperative localization: a communication-efficient dual-fusion consistent approach, IEEE Robot. Autom. Lett. 10 (1) (2024) 636–643.

[6] N. Hao, F. He, C. Tian, Y. Hou, On the consistency of multi-robot cooperative localization: a transformation-based approach, IEEE Robot. Autom. Lett. 10 (1) (2024) 280–287.

[7] Q. Shen, Z. Wu, X. Yi, P. Zhou, H. Zhang, S. Yan, X. Wang, Gamba: marry Gaussian splatting with mamba for single-view 3D reconstruction, IEEE Trans. Patt. Anal. Mach. Intell. (2025) 1–14. Early access.

[8] G. Zhang, C. Xue, R. Zhang, SuperNeRF: high-precision 3D reconstruction for large-scale scenes, IEEE Trans. Geosci. Remote Sensing 62 (2024) 5635313.

[9] X. Zhang, H. Zong, W. Wu, Cooperative obstacle avoidance of unmanned system swarm via reinforcement learning under unknown environments, IEEE Trans. Instrum. Meas. 74 (2024) 7500615.

[10] Y.-K. Fu, Y. Liu, C. Deng, Deep reinforcement learning-based cooperative control for multimobile robots with obstacle avoidance, IEEE Trans. Ind. Electron. (2025) 1–9. Early access.

[11] Y. Chang, K. Ebadi, C.E. Denniston, M.F. Ginting, A. Rosinol, A. Reinke, M. Palieri, J. Shi, A. Chatterjee, B. Morrell, et al., LAMP 2.0: A robust multi-robot SLAM system for operation in challenging large-scale underground environments, IEEE Robot. Autom. Lett. 7 (4) (2022) 9175–9182.

[12] J. Wu, C. Luo, G. Min, S. McClean, Formation control algorithms for multi-UAV systems with unstable topologies and hybrid delays, IEEE Trans. Veh. Technol. 73 (9) (2024) 12358–12369.

[13] H. Wang, J. Shan, H. Alkomy, Fully distributed edge-based dynamic event-Triggered control for multiple quadrotors, IEEE/ASME Trans. Mechatron. 22 (4) (2024) 3203–3214.

[14] M. Khodaverdian, S. Hajshirmohamadi, A. Hakobyan, S. Ijaz, Predictor-based constrained fixed-time sliding mode control of multi-UAV formation flight, Aerosp. Sci. Technol. 148 (2024) 109113.

[15] Y.-J. Liu, X. Shang, L. Tang, S. Zhang, Finite-time consensus adaptive neural network control for nonlinear multiagent systems under PDE models, IEEE Trans. Neural Netw. Learn. Syst. 36 (4) (2024) 6218–6228.

[16] H. Zhang, X. Yue, J. Sun, X. Guo, Prescribed finite-time fuzzy consensus control for multiagent systems with aperiodic updates, IEEE Trans. Syst. Man Cybern. -Syst. 55 (2) (2024) 1362–1373.

[17] R. Nie, W. Du, Z. Li, S. He, Finite-time consensus control for MASs under hidden Markov model mechanism, IEEE Trans. Autom. Control 69 (7) (2024) 4726–4733.

[18] J. Liang, Y. Wang, H. Zhong, Y. Chen, H. Li, H. Hua, W. Wang, Robust adaptive tracking control for aerial transporting a cable-suspended payload using backstepping sliding mode techniques, IEEE Trans. Autom. Sci. Eng. 22 (2024) 4490–4500.

[19] Y. Hou, D. Chen, S. Yang, Adaptive robust trajectory tracking controller for a quadrotor UAV with uncertain environment parameters based on backstepping sliding mode method, IEEE Trans. Autom. Sci. Eng. 22 (2023) 4446–4456.

[20] J. Liang, Y. Chen, Y. Wu, Z. Miao, H. Zhang, Y. Wang, Adaptive prescribed performance control of unmanned aerial manipulator with disturbances, IEEE Trans. Autom. Sci. Eng. 20 (3) (2023) 1804–1814.

[21] S. Ijaz, M. Galea, M.T. Hamayun, H. Ijaz, U. Javaid, A new output integral sliding mode fault-tolerant control and fault estimation scheme for uncertain systems, IEEE Trans. Autom. Sci. Eng. 21 (3) (2024) 4214–4225.

[22] H. Zhao, L. Peng, L. Xie, H. Yu, Dynamic event-triggered sliding-mode bipartite consensus for multi-agent systems with unknown dynamics, IEEE Trans. Autom. Sci. Eng. 22 (2024) 7170–7180.

[23] Q. Miao, K. Zhang, B. Jiang, Fixed-time collision-free fault-tolerant formation control of multi-UAVs under actuator faults, IEEE Trans. Cybern. 54 (6) (2024) 3679–3691.

[24] Y.-H. Su, P. Bhowmick, A. Lanzon, A fixed-time formation-containment control scheme for multi-agent systems with motion planning: applications to quadcopter UAVs, IEEE Trans. Veh. Technol. 73 (7) (2024) 9495–9507.

[25] J. Rodriguez, V. Lechappe, S. Chesne, New methodology for adaptive sliding mode control with self-tuning threshold based on chattering detection, Mech. Syst. Signal Proc. 235 (2025) 112854.

[26] Y. Yan, H. Zhang, J. Sun, Y. Wang, Sliding mode control based on reinforcement learning for TS fuzzy fractional-order multiagent system with time-varying delays, IEEE Trans. Neural Netw. Learn. Syst. 35 (8) (2023) 10368–10379.

[27] T. Wang, G. Zong, X. Zhao, N. Xu, Data-driven-based sliding-mode dynamic event-triggered control of unknown nonlinear systems via reinforcement learning, Neurocomputing 601 (2024) 128176.

[28] T. Wang, H. Wang, N. Xu, L. Zhang, K.H. Alharbi, Sliding-mode surface-based decentralized event-triggered control of partially unknown interconnected nonlinear systems via reinforcement learning, Inf. Sci. 641 (2023) 119070.

[29] T. Wang, B. Niu, N. Xu, L. Zhang, ADP-based online compensation hierarchical sliding-mode control for partially unknown switched nonlinear systems with actuator failures, ISA Trans. 155 (2024) 69–81.

[30] G. Wen, B. Li, Optimized leader-follower consensus control using reinforcement learning for a class of second-order nonlinear multiagent systems, IEEE Trans. Syst. Man Cybern. Syst. 52 (9) (2021) 5546–5555.

[31] Z. Hou, I. Fantoni, Interactive leader–follower consensus of multiple quadrotors based on composite nonlinear feedback control, IEEE Trans. Control Syst. Technol. 26 (5) (2017) 1732–1743.

[32] K. Liu, R. Wang, X. Wang, X. Wang, Anti-saturation adaptive finite-time neural network based fault-tolerant tracking control for a quadrotor UAV with external disturbances, Aerosp. Sci. Technol. 115 (2021) 106790.

[33] L. Chen, Z. Liu, Q. Dang, W. Zhao, G. Wang, Robust trajectory tracking control for a quadrotor using recursive sliding mode control and nonlinear extended state observer, Aerosp. Sci. Technol. 128 (2022) 107749.

[34] Y. Hong, J. Hu, L. Gao, Tracking control for multi-agent consensus with an active leader and variable topology, Automatica 42 (7) (2006) 1177–1182.

[35] A. Polyakov, Nonlinear feedback design for fixed-time stabilization of linear control systems, IEEE Trans. Autom. Control 57 (8) (2011) 2106–2110.

[36] C. Qian, W. Lin, A continuous feedback approach to global strong stabilization of nonlinear systems, IEEE Trans. Autom. Control 46 (7) (2001) 1061–1079.

[37] S.P. Bhat, D.S. Bernstein, Finite-time stability of continuous autonomous systems, SIAM J. Control Optim. 38 (3) (2000) 751–766.

[38] B. Li, W. Gong, Y. Yang, B. Xiao, Distributed fixed-time leader-following formation control for multiquadrotors with prescribed performance and collision avoidance, IEEE Trans. Aerosp. Electron. Syst. 59 (5) (2023) 7281–7294.

[39] L. Zhang, Y. Xia, G. Shen, B. Cui, Fixed-time attitude tracking control for spacecraft based on a fixed-time extended state observer, Sci. China Inf. Sci. 64 (11) (2021) 212201.

[40] M.T. Angulo, J.A. Moreno, L. Fridman, Robust exact uniformly convergent arbitrary order differentiator, Automatica 49 (8) (2013) 2489–2495.

[41] B. Wang, X. Yu, L. Mu, Y. Zhang, Disturbance observer-based adaptive fault-tolerant control for a quadrotor helicopter subject to parametric uncertainties and external disturbances, Mech. Syst. Signal Process. 120 (2019) 727–743.
[42] J. Cai, J. Feng, J. Wang, Y. Zhao, Tracking consensus of multi-agent systems under switching topologies via novel SMC: an event-triggered approach, IEEE Trans. Netw. Sci. Eng. 9 (4) (2022) 2150–2163.
[43] B. An, H. Fan, B. Wang, L. Liu, Y. Wang, Event-triggered finite-time formation tracking of multi-agent systems with mismatched disturbances under switching topologies, ISA Trans. 151 (2024) 19–32.
[44] J. Schulman, P. Moritz, S. Levine, M. Jordan, P. Abbeel, High-dimensional continuous control using generalized advantage estimation, arXiv preprint arXiv:1506.02438 (2015).
[45] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms, arXiv preprint arXiv:1707.06347 (2017).
[46] T. Huang, Y. Liang, X. Ban, J. Zhang, X. Huang, The control of magnetic levitation system based on improved Q-network, in: 2019 IEEE Symposium Series on Computational Intelligence (SSCI), 2019, pp. 191–197.