Fixed-time Stochastic Learning from Human-UAV Interaction with State-Input Constraints

Junkai Tan, Shuangsi Xue, *IEEE Member*, Qingshu Guan, Zihang Guo,
Hui Cao, *IEEE Member*, and Badong Chen, *IEEE Senior Member*

***Abstract*—Human-unmanned aerial vehicle (UAV) collaboration requires control frameworks that are both efficient and safe. This paper introduces a stochastic fixed-time inverse optimal control (FxT-IOC) approach designed for such systems. The proposed framework constructs inverse optimal control, enabling the extraction of human operator intent. It features a fixed-time adaptive learning mechanism that guarantees parameter convergence within a predetermined time, irrespective of initial conditions. Crucially, the design explicitly incorporates prescribed performance control (PPC) to enforce state constraints while handling input saturation, ensuring operational safety and reliability. Rigorous theoretical analysis establishes the fixed-time stability of the learning process and the closed-loop system under these constraints. The effectiveness of the FxT-IOC framework is validated through comprehensive numerical simulations and physical hardware experiments, demonstrating superior trajectory tracking precision, accelerated learning convergence, and robust constraint satisfaction compared to human demonstrations. This work offers a principled and practical solution for developing high-performance, reliable human-UAV collaborative systems.**

***Index Terms*—Human-UAV interaction, inverse optimal control, fixed-time learning, adaptive dynamic programming.**

I. Introduction

UNMANNED aerial vehicles (UAVs) are increasingly vital in applications ranging from infrastructure inspection to emergency response [1], [2]. This proliferation requires sophisticated control strategies capable of effectively integrating human operational expertise with autonomous system capabilities. A central challenge lies in achieving this synergy safely and efficiently [3]. While purely manual control can be demanding and susceptible to errors [4], [5], fully autonomous systems may lack the adaptability required for complex, dynamic environments [6]. Human-robot interaction (HRI) frameworks seek to overcome these limitations by leveraging the complementary strengths of human strategic guidance and autonomous precision execution [7], [8]. Developing robust collaborative control mechanisms is therefore crucial [9], drawing upon advancements in areas such as master-slave control [10], practical fixed-time methods [11], and cooperative games [12] to enhance overall system performance and reliability.

Integrating human operator expertise with autonomous UAV capabilities presents a significant challenge. This necessitates methods that can interpret human decision-making, such as learning directional preferences [13] or modeling behavior within the loop [6], and adapt control strategies to dynamic mission requirements using techniques like composite adaptive learning [14]. Inverse Optimal Control (IOC) [15] provides a principled framework for this by inferring underlying objectives from observed expert actions, enabling the synthesis of control policies that align with human intent [16]. Recent advancements have explored game-theoretic [17] and trust-region [18] formulations. This inference enables the synthesis of control policies aligned with human intent [19], while also considering optimality guarantees and safety aspects crucial for cooperative tasks [12], and potentially scaling to multi-agent coordination [20]. In [21], a IOC-based method for inferring constraints from demonstrations is proposed, enhancing the interpretability of human-UAV interactions. However, many traditional IOC approaches, including those applied to drone objective inference [22] or incorporating anomaly detection [23], often rely on asymptotic convergence properties. These may prove inadequate for the rapid adaptation and response times crucial in real-time human-UAV collaboration scenarios.

While traditional IOC methods, including recent variants [22], [24], offer valuable tools for inferring intent, their reliance on asymptotic convergence often falls short in dynamic human-UAV collaboration where rapid adaptation is paramount. Fixed-time stability theory [25], [26] provides a crucial advantage by guaranteeing convergence within a predetermined time, irrespective of initial conditions, even in stochastic [27] or discrete-time [28] settings. This predictable convergence is vital for ensuring the safety and reliability

This work was supported in part by the National Natural Science Foundation of China (No. U21A20485), the International (Regional) Cooperation and Exchange (ICE) Projects of the National Natural Science Foundation of China (NSFC) (No. 62311540022), and the China Postdoctoral Science Foundation (Project ID: 2024M762602).

Junkai Tan, Shuangsi Xue, Qingshu Guan and Hui Cao are with the Shaanxi Key Laboratory of Smart Grid, Xi'an Jiaotong University, Xi'an 710049, China, and also with the State Key Laboratory of Electrical Insulation and Power Equipment, School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: 15958024@stu.xjtu.edu.cn; xssxjtu@xjtu.edu.cn; huicao@mail.xjtu.edu.cn) (*Corresponding author: Shuangsi Xue*).

Badong Chen is with the National Key Laboratory of Human-Machine Hybrid Augmented Intelligence, National Engineering Research Center for Visual Information and Applications, and the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China.

demanded by applications such as practical tracking [29] and adaptive control [30]. Consequently, recent research increasingly studies fixed-time principles with learning-based control paradigms like Adaptive Dynamic Programming (ADP) and Inverse Reinforcement Learning (IRL) [31]. Efforts encompassing fixed-time system identification [32], event-triggered control [33], ADP [34], and adaptive optimal IRL [35], [36] demonstrate the potential for enhanced convergence speed compared to conventional methods [37], overcoming limitations in time-critical human-robot interaction scenarios. Despite these advancements, a significant challenge persists in simultaneously achieving guaranteed fixed-time convergence, effective human intent inference via IOC, and robust handling of state and input constraints within human-UAV systems. Existing IOC approaches often lack predictable convergence guarantees [15], [24], while many fixed-time control methods [34], [38] do not adequately integrate human factors or manage complex operational limits like Prescribed Performance Control (PPC) and input saturation. This paper introduces a novel Fixed-Time Inverse Optimal Control (FxT-IOC) framework specifically designed to address these concurrent requirements. The main contributions of this work include:

1) **FxT-IOC for Human Intent Learning:** A novel FxT-IOC framework is introduced to learn the human operator's reward function within a guaranteed fixed time, independent of initial conditions. This approach overcomes the asymptotic convergence limitations of traditional IOC methods [15], [22].
2) **Optimal Control with State-Input Constraints:** A unified optimal control synthesis is developed to systematically handle both state constraints via Prescribed Performance Control (PPC) and input saturation. This coordination achieves high-performance tracking while respecting physical actuator limits, a key challenge that is often overlooked in prior works [6], [38].
3) **Rigorous Fixed-Time Stability Guarantees:** A formal theoretical proof of fixed-time stability for both the parameter learning and the closed-loop system is provided in Theorem 1. This guarantee of convergence to a bounded residual set within a fixed time is a significant advancement over common asymptotic results [31], [39].

The paper is organized as follows: Section II presents preliminaries. Section III formulates the problem. Section IV details the FxT-IOC framework. Section V provides validation results. Section VI concludes the paper.

Notations: $\bar{\lambda}(\cdot), \underline{\lambda}(\cdot)$ denote maximum and minimum eigenvalues, respectively; $\mathrm{sat}(u)$ is the saturation function; $\mathbb{E}[\cdot], \mathbb{P}[\cdot]$ are expectation and probability, respectively; $\lceil x \rfloor^{\gamma} = |x|^{\gamma} \operatorname{sign}(x)$.

II. Preliminaries and System Description

A. Stochastic Fixed-Time Stability Framework

This subsection introduces stochastic fixed-time stability for analyzing convergence under uncertainty. Consider nonlinear stochastic systems described by the Itô differential equation:

$$\mathrm{d}X(t) = f(X(t), U(t))\mathrm{d}t + g(X(t), U(t))\mathrm{d}W(t) \quad (1)$$

TABLE I
Notations and their Meanings

Notation	Description
System Variables:	
t	Time variable
x_a, x_p	Attitude and position state vectors
$e(t)$	Tracking error vector
ϱ	Transformed tracking error (PPC)
X	Augmented state vector $[x^\top, \varrho^\top, x_d^\top]^\top$
Φ	Euler angles $[\phi, \theta, \psi]^\top$ (roll, pitch, yaw)
$\mathcal{V}(t)$	Lyapunov function
Control and Learning Parameters:	
u_a, u_h	Autonomous and human control inputs
U	Combined control input vector
γ_1, γ_2	Fixed-time exponents ($0 < \gamma_1 < 1$, $\gamma_2 > 1$)
$V(X, U)$	Value function
$\mathit{\Pi}(U)$	Input cost function (saturation-aware)
$\hat{W}_c, \hat{W}_a$	Critic and actor network weights
$\hat{\theta}$	Reward function parameters (IOC)
Transformation Functions:	
$\vartheta_i(t)$	Performance bound function
$\xi_{l,i}, \xi_{u,i}$	Lower and upper error bounds
$\phi(\cdot)$	Transformation function (e.g., tan, tanh)
$\mathcal{R}_\varrho$	Diagonal transformation matrix (PPC)
Υ	Compensation term in transformed dynamics
Learning Algorithm Components:	
$\mathcal{D}(t)$	Experience buffer
$\mathcal{E}(t)$	Current learning errors
$\mathcal{E}^k$	Historical learning samples
δ, δ_θ	Bellman error and reward parameter error
ζ^k	Temporal Bellman error (historical)
$\Gamma_c, \Gamma_a, \Gamma_\theta$	Positive definite gain matrices
α_1, α_2	Learning rate parameters

Here, $X(t) \in \mathbb{R}^n$ is the state, $U(t) \in \mathbb{R}^m$ is the control, f is the drift, g is the diffusion matrix, and $W(t)$ is a standard Wiener process. Assume $f(0,0) = 0$ and $g(0,0) = 0$.

Definition 1 (Stochastic Fixed-Time Stability [27])**.** *The origin of system (1) (with $U = 0$) is fixed-time stable in probability if it is stable in probability and the expected settling time $\mathbb{E}[\tau(X_0)]$ is finite for any $X_0 \neq 0$, where*

$$\tau(X_0) = \inf\{t \geq 0 : X(t, X_0) = 0\}$$

It is globally fixed-time stable in probability if $\mathbb{E}[\tau(X_0)] \leq T_{\max}$ for some $T_{\max} > 0$ and all $X_0 \in \mathbb{R}^n$.

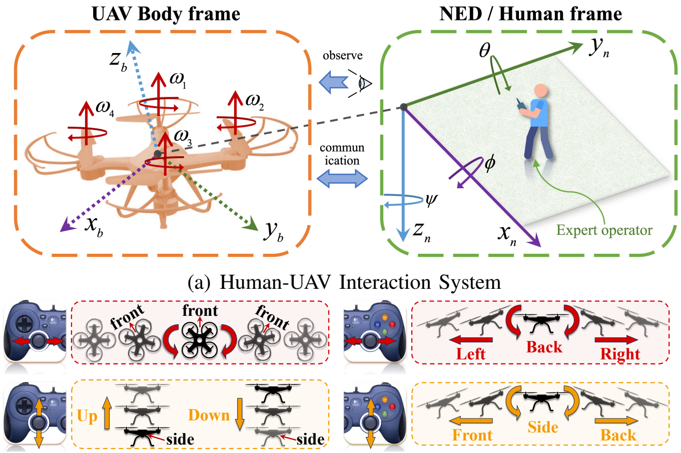
The following lemma provides sufficient conditions for establishing fixed-time stability based on Lyapunov theory.

Lemma 1 (Stochastic Fixed-Time Convergence Criteria [25])**.** *For system (1), let $V(X)$ be a C^2 positive definite function with $V(0) = 0$. If there exist constants $k_1, k_2 > 0$, $0 < \gamma_1 < 1$, $\gamma_2 > 1$, and $\sigma \geq 0$ such that the infinitesimal generator $\mathcal{L}V(X)$ satisfies:*

$$\mathcal{L}V(X) \leq -k_1 V(X)^{\gamma_1} - k_2 V(X)^{\gamma_2} + \sigma \quad (2)$$

for all $X \in \mathbb{R}^n \setminus \{0\}$, where

$$\mathcal{L}V(X) = \frac{\partial V}{\partial X} f(X, U) + \frac{1}{2}\mathrm{Tr}\left(g(X, U)^\top \frac{\partial^2 V}{\partial X^2} g(X, U) \right)$$



(a) Human-UAV Interaction System

(b) Maneuver 1: throttle/yaw control

(c) Maneuver 2: roll/pitch control

Fig. 1. Exemplar human maneuvers for UAV operation using a gamepad.

then the system trajectories converge in fixed time to a residual set Ω_V around the origin. The expected settling time $\mathbb{E}[\tau(X_0)]$ is bounded by $T_{\max}$, independent of the initial state X_0:

$$\mathbb{E}[\tau(X_0)] \le T_{\max} \approx \frac{1}{k_1\epsilon(1-\gamma_1)} + \frac{1}{k_2\epsilon(\gamma_2-1)} \qquad (3)$$

for any small $\epsilon \in (0,1)$. The size of Ω_V depends on σ, k_1, k_2. If $\sigma = 0$, the origin is globally fixed-time stable in probability.

Remark 1 (Stochastic Fixed-Time Convergence)**.** *Lemma 1 guarantees convergence to a residual set within a fixed time, independent of initial conditions. The exponents γ_1, γ_2 ensure rapid convergence, while σ bounds the residual error. This predictability is vital for real-time systems.*

B. Quadrotor UAV Dynamics with Input Constraints

Consider a quadrotor UAV system whose dynamics involve coupled attitude and position components, as illustrated in Fig. 1(a). The human operator interacts with the UAV via control inputs, depicted in Fig. 1(b) and Fig. 1(c). The attitude dynamics are described using Euler angles [40]:

$$M_a\ddot{\Phi} = -C_a(\Phi,\dot{\Phi})\dot{\Phi} + \mathcal{T} + \tau_d \qquad (4)$$

where $M_a = \mathrm{diag}([J_\phi, J_\theta, J_\psi])$ is the inertia matrix, $\Phi = [\phi, \theta, \psi]^\top$ are the Euler angles (roll, pitch, yaw) with constraints $\phi, \theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$, $\psi \in [-\pi, \pi]$. $C_a(\Phi,\dot{\Phi})$ is the Coriolis/centrifugal matrix, $\mathcal{T} = [\gamma_\phi, \gamma_\theta, \gamma_\psi]^\top$ is the control torque vector: $\gamma_\phi = \alpha_l\alpha_w u_\phi$, $\gamma_\theta = \alpha_l\alpha_w u_\theta$, $\gamma_\psi = \alpha_\gamma u_\psi$, with constants $\alpha_l, \alpha_w, \alpha_\gamma$. τ_d represents disturbances. The control inputs u_ϕ, u_θ, u_ψ derive from rotor speeds ω_j^2 $(j = 1,\ldots,4)$: $u_\phi = \omega_1^2 - \omega_3^2$, $u_\theta = \omega_2^2 - \omega_4^2$, $u_\psi = \omega_1^2 + \omega_3^2 - \omega_2^2 - \omega_4^2$. The translational dynamics are given by [12]:

$$M_p\ddot{\varrho} = F_T R(\Phi)e_3 - M_p g e_3 + d_p \qquad (5)$$

where $\varrho = [x, y, z]^\top$ is the position, M_p is the mass, $F_T = \sum_{j=1}^4 \omega_j^2$ is the total thrust, $R(\Phi)$ is the rotation matrix (body to inertial), $e_3 = [0, 0, 1]^\top$, g is gravity, and d_p represents disturbances. The attitude and position state vectors are:

$$\begin{cases} x_a = [\phi, \theta, \psi, \dot{\phi}, \dot{\theta}, \dot{\psi}]^\top \in \mathbb{R}^6 & \text{(attitude state)} \\ x_p = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^\top \in \mathbb{R}^6 & \text{(position state)} \end{cases}$$

The complete UAV dynamics, incorporating both autonomous control u_a and human operator input u_h, can be represented in state-space form:

$$\begin{bmatrix} \dot{x}_a \\ \dot{x}_p \end{bmatrix} = \begin{bmatrix} f_a(x_a) \\ f_p(x_a, x_p) \end{bmatrix} + \begin{bmatrix} g_a(x_a) \\ 0 \end{bmatrix} u_a + \begin{bmatrix} 0 \\ g_p(x_a) \end{bmatrix} u_h + d(t) \qquad (6)$$

where $u_a \in \mathbb{R}^{m_a}$ and $u_h \in \mathbb{R}^{m_h}$ are the autonomous and human control inputs, respectively, subject to saturation constraints: $|u_{a,i}| \le \bar{u}_{a,i}$ for $i = 1,\ldots,m_a$ and $|u_{h,i}| \le \bar{u}_{h,i}$ for $i = 1,\ldots,m_h$. The terms f_a, f_p represent the drift dynamics, g_a, g_p are the input matrices, and $d(t)$ lumps the disturbances τ_d, d_p. Specifically:

$$f_a(x_a) = \begin{bmatrix} \dot{\Phi} \\ -M_a^{-1}C_a(\Phi,\dot{\Phi})\dot{\Phi} \end{bmatrix}, \quad g_a(x_a) = \begin{bmatrix} 0_{3\times m_a} \\ M_a^{-1}B_a \end{bmatrix},$$

$$f_p(x_a, x_p) = \begin{bmatrix} \dot{\varrho} \\ -ge_3 \end{bmatrix}, \quad g_p(x_a) = \begin{bmatrix} 0_{3\times m_h} \\ \frac{R(\Phi)e_3}{M_p} \end{bmatrix}$$

where B_a maps u_a to the control torques $\mathcal{T}$. Note that the human input u_h typically controls the total thrust F_T, affecting the position dynamics.

For trajectory tracking tasks, let $x_d(t) \in \mathbb{R}^n$ be the desired reference trajectory, generated by a reference model $\dot{x}_d(t) = f_d(x_d(t))$. The tracking error is defined as $e(t) = x(t) - x_d(t)$, where $x(t) = [x_a(t)^\top, x_p(t)^\top]^\top$. The error dynamics are:

$$\dot{e}(t) = [f(x) - f_d(x_d)] + g_a(x)u_a + g_h(x)u_h + d(t) \qquad (7)$$

where $f(x) = [f_a(x_a)^\top, f_p(x_a, x_p)^\top]^\top$ and the input matrices are combined appropriately.

III. Problem Formulation: Optimal Control Design with Input and State Constraints

This section formulates the state-input constrained optimal control problem for the human-UAV system (6). The objective is precise trajectory tracking with guaranteed fixed-time convergence, while respecting state and input constraints and integrating human intent via IOC. A unified PPC transformation and saturation-aware costs framework given in Fig. 2 is used to achieve these goals.

A. PPC-based State Constraint Transformation

Prescribed Performance Control (PPC) enforces constraints on the tracking error $e(t)$ by ensuring it remains within predefined time-varying bounds: $\xi_{l,i}(t) < e_i(t) < \xi_{u,i}(t)$.

Definition 2 (Prescribed Performance Bound [29])**.** *A smooth, positive, and decreasing function $\vartheta_i(t) : \mathbb{R}_{\ge 0} \to \mathbb{R}_{>0}$ is a performance bound function if it satisfies:*

$$\lim_{t\to\infty} \vartheta_i(t) = \vartheta_{i\infty} > 0, \quad \vartheta_i(0) = \vartheta_{i0} > \vartheta_{i\infty} \qquad (8)$$

for $i = 1, 2, \ldots, n$. A common choice is an exponential decay:

$$\vartheta_i(t) = (\vartheta_{i0} - \vartheta_{i\infty})e^{-\lambda_i t} + \vartheta_{i\infty} \qquad (9)$$

where $\lambda_i > 0$ determines the convergence rate towards the steady-state bound $\vartheta_{i\infty}$.

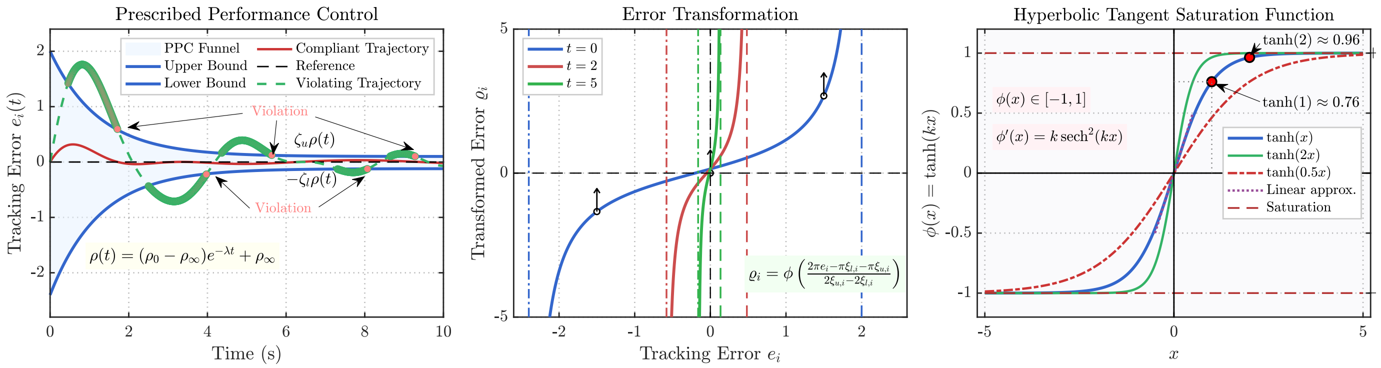


Fig. 2. Visualization: (Left) Unified PPC transformation (11) for state constraints ($e_i \rightarrow \varrho_i$). (Right) Tanh function for input saturation (29) ($|u_i| \leq \mu_i$).

The time-varying lower and upper bounds for the tracking error $e_i(t)$ are defined using the performance bound function:

$$\xi_{l,i}(t) = -\zeta_{l,i}\vartheta_i(t), \quad \xi_{u,i}(t) = \zeta_{u,i}\vartheta_i(t) \tag{10}$$

where $\zeta_{l,i}, \zeta_{u,i} \in (0, \infty)$ are constants defining the shape and asymmetry of the performance funnel. To transform the constrained error $e_i(t)$ into an unconstrained variable $\varrho_i(t)$, the following mapping (visualized in Fig. 2, Left) is employed:

$$\begin{cases} \varrho_i = \phi\left(\dfrac{2\pi e_i - \pi(\xi_{l,i} + \xi_{u,i})}{2(\xi_{u,i} - \xi_{l,i})}\right) \\ e_i = \dfrac{\xi_{l,i} + \xi_{u,i}}{2} + \dfrac{\xi_{u,i} - \xi_{l,i}}{2\pi}\phi^{-1}(\varrho_i) \end{cases} \tag{11}$$

where $\phi(\cdot) = \tan(\cdot)$ maps the normalized error to $(-\infty, \infty)$. Bounded $\varrho_i(t)$ ensures that $e_i(t)$ remains within the bounds $(\xi_{l,i}(t), \xi_{u,i}(t))$. The dynamics of the transformed error ϱ_i are obtained by differentiation:

$$\dot{\varrho}_i = \frac{\partial \varrho_i}{\partial e_i}\dot{e}_i + \frac{\partial \varrho_i}{\partial \xi_{l,i}}\dot{\xi}_{l,i} + \frac{\partial \varrho_i}{\partial \xi_{u,i}}\dot{\xi}_{u,i} = \mathcal{R}_{\varrho_i}\dot{e}_i + \Upsilon_i \tag{12}$$

where $\mathcal{R}_{\varrho_i} = \frac{\partial \varrho_i}{\partial e_i} = \pi \sec^2\left(\frac{2\pi e_i - \pi(\xi_{l,i} + \xi_{u,i})}{2(\xi_{u,i} - \xi_{l,i})}\right)/(\xi_{u,i} - \xi_{l,i}) > 0$ and $\Upsilon_i = \frac{\partial \varrho_i}{\partial \xi_{l,i}}\dot{\xi}_{l,i} + \frac{\partial \varrho_i}{\partial \xi_{u,i}}\dot{\xi}_{u,i}$ is a term involving derivatives of the bounds. The vectorized transformed error dynamics are:

$$\dot{\varrho} = \mathcal{R}_\varrho \dot{e} + \Upsilon \tag{13}$$

where $\varrho = [\varrho_1, \ldots, \varrho_n]^\top$, $\mathcal{R}_\varrho = \mathrm{diag}(\mathcal{R}_{\varrho_1}, \ldots, \mathcal{R}_{\varrho_n}) \succ 0$, and $\Upsilon = [\Upsilon_1, \ldots, \Upsilon_n]^\top$. Combining (6), (7), (13), and $\dot{x}_d = f_d(x_d)$, the augmented system is:

$$\dot{X} = F(X) + G(X)U + D(t) \tag{14}$$

where the augmented state is $X = [x^\top, \varrho^\top, x_d^\top]^\top \in \mathbb{R}^{N_X}$ (with $N_X = \dim(x) + \dim(\varrho) + \dim(x_d)$), the combined control input is $U \in \mathbb{R}^m$, and the augmented drift, input matrix, and disturbance terms are:

$$F(X) = \begin{bmatrix} f(x) \\ \mathcal{R}_\varrho(f(x) - f_d(x_d)) + \Upsilon \\ f_d(x_d) \end{bmatrix}, \tag{15}$$
$$G(X) = \begin{bmatrix} g_a(x) & g_h(x) \\ \mathcal{R}_\varrho g_a(x) & \mathcal{R}_\varrho g_h(x) \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad D(t) = \begin{bmatrix} d(t) \\ \mathcal{R}_\varrho d(t) \\ \mathbf{0} \end{bmatrix}$$

Here, $U = [u_a^\top, u_h^\top]^\top$ combines autonomous and human inputs, and $\mathbf{0}$ denotes zero matrices/vectors of appropriate dimensions. The augmented system dynamics (14) are assumed to satisfy the following property:

Assumption 1 (System Properties [41])**.** *The dynamics $F(X)$ and $G(X)$ are locally Lipschitz continuous. The disturbance $D(t)$ is bounded, i.e., $\|D(t)\| \leq \bar{D}$ for some constant $\bar{D} > 0$.*

$$\|F(X_1) - F(X_2)\| \leq L_F\|X_1 - X_2\| \tag{16}$$
$$\|G(X_1) - G(X_2)\| \leq L_G\|X_1 - X_2\| \tag{17}$$

Furthermore, the disturbance $D(t)$ is bounded, i.e., $\|D(t)\| \leq \bar{D}$ for some constant $\bar{D} > 0$.

B. Optimal Control with Input Constraints

The objective is to design a control policy $U(t)$ for the augmented system (14) that minimizes an infinite-horizon cost functional, ensuring fixed-time convergence and respecting input saturation constraints. The cost functional is defined as:

$$J(X_0, U(\cdot)) = \mathbb{E}\left\{\int_0^\infty e^{-\gamma t} r(X(\tau), U(\tau))\, d\tau \;\middle|\; X(0) = X_0\right\}$$

where $\gamma > 0$ is the discount factor, and the instantaneous cost $r(X, U)$ incorporates fixed-time convergence terms and a saturation-aware input penalty:

$$r(X, U) = \underbrace{\|X\|_Q^{\gamma_1} + \|X\|_Q^{\gamma_2}}_{\text{Fixed-Time State Cost}} + \underbrace{\mathit{\Pi}(U)}_{\text{Input Cost}} \tag{18}$$

where $Q \succ 0$ weights the state, $0 < \gamma_1 < 1$ and $\gamma_2 > 1$ are fixed-time exponents for fast convergence, and $\|X\|_Q^\alpha = (X^\top Q X)^{\alpha/2}$. The input cost $\mathit{\Pi}(U)$ addresses the input saturation constraints $|u_i| \leq \mu_i$ through a non-quadratic penalty function. This function, visualized in Fig. 2 and based on the formulation in [42], is defined by the integral:

$$\mathit{\Pi}(U) = \int_0^U 2\mu R \tanh^{-1}(\nu/\mu)\, d\nu \tag{19}$$

where μ contains the saturation limits and $R \succ 0$ weights the control effort. The gradient is $\nabla_U \mathit{\Pi}(U) = 2\mu R \tanh^{-1}(U/\mu)$, which penalizes inputs near saturation.

Remark 2 (Role of the $\tanh^{-1}$ Function)**.** *The $\tanh^{-1}(\cdot)$ function within the cost functional $\mathit{\Pi}(U)$ serves as a mathematical tool to derive the optimal control structure, a technique employed in similar ADP contexts [41]. Crucially, the final implemented control policy (22) utilizes the standard hyperbolic tangent function. This design ensures that the control output is inherently bounded within $(-\mu, \mu)$, thus sidestepping any numerical singularities at the saturation boundary during execution.*

The optimal value function $V^*(X)$ is the minimum cost starting from state X:

$$V^*(X) = \min_{U(\cdot)} J(X, U(\cdot)) \tag{20}$$

According to Bellman's principle of optimality, $V^*(X)$ satisfies the Hamilton-Jacobi-Bellman (HJB) equation:

$$\begin{aligned} 0 = \min_{U \in \Omega_U} \{ & r(X,U) + \nabla V^*(X)^\top (F(X) + G(X)U) \\ & - \gamma V^*(X) \} \end{aligned} \tag{21}$$

where $\nabla V^*(X) = \partial V^*(X)/\partial X$ is the gradient of the value function, and $\Omega_U = \{U \in \mathbb{R}^m : |u_i| \leq \mu_i, \forall i\}$ is the set of admissible controls. The optimal control policy $U^*(X)$ is the control that minimizes the expression within the braces in (21). By setting the partial derivative with respect to U to zero:

$$\begin{aligned} \frac{\partial}{\partial U} \{ r(X,U) + \nabla V^*(X)^\top G(X) U \} &= 0 \\ \nabla_U \mathit{\Pi}(U) + G(X)^\top \nabla V^*(X) &= 0 \\ 2\mu R \tanh^{-1}\left(\frac{U^*}{\mu}\right) + G(X)^\top \nabla V^*(X) &= 0 \end{aligned}$$

Solving for U^* yields the optimal control law:

$$U^*(X) = -\mu \tanh\left(\frac{1}{2\mu} R^{-1} G(X)^\top \nabla V^*(X)\right) \tag{22}$$

This control law inherently respects the saturation limits μ due to the properties of the hyperbolic tangent function. The following standard assumption regarding the cost matrices is made:

Assumption 2 (Cost Function Properties [42])**.** *The cost weighting matrices Q and R from reward (18) are positive definite, satisfying:*

$$\underline{\lambda}_Q \mathbf{I} \preceq Q \preceq \bar{\lambda}_Q \mathbf{I}, \quad \underline{\lambda}_R \mathbf{I} \preceq R \preceq \bar{\lambda}_R \mathbf{I} \tag{23}$$

where $\underline{\lambda}_Q, \bar{\lambda}_Q, \underline{\lambda}_R, \bar{\lambda}_R > 0$ are positive constants and $\mathbf{I}$ is the identity matrix.

Solving the HJB equation (21) directly is computationally challenging, which may cause the 'curses of dimensionality' [43]. The next section presents the FxT-IOC framework, which approximates the solution and ensures fixed-time convergence.

IV. FxT-IOC Framework Design

This section details the FxT-IOC framework, which is shown in Fig. 3, This Framework combines Inverse Optimal Control (IOC) for human intent learning with a fixed-time Actor-Critic structure for autonomous control synthesis. The design ensures guaranteed convergence times and state-input constraints satisfaction.

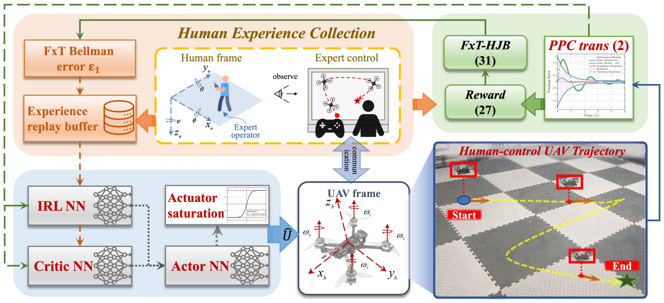


Fig. 3. Architecture of the FxT-IOC framework, integrating IOC for human intent learning and fixed-time Actor-Critic control synthesis.

A. Inverse Optimal Control for Human Intent Learning

To understand and leverage the human operator's expertise, the framework first employs IOC to estimate the underlying reward function guiding the human's control actions U_h.

1) Problem Formulation: To find reward parameters θ that best explain the observed human behavior U_h. This is formulated as minimizing the discrepancy between the human's actions and the optimal actions $U_\theta^*(X)$ derived from the estimated reward function r_θ:

$$\min_{\theta \in \Theta} \mathcal{L}(\theta) = \mathbb{E}\left[\|U_h - U_\theta^*(X)\|^2\right] \tag{24}$$

where $U_\theta^*(X)$ is the policy that minimizes the cost associated with the reward r_θ. The human's reward is parameterized linearly:

$$r_\theta(X,U) = \theta^\top \phi_r(X,U) + \frac{1}{2} U^\top R_h U \tag{25}$$

Here, $\theta \in \mathbb{R}^p$ are the unknown parameters, $\phi_r(X,U)$ is a vector of known basis functions representing task-relevant features, and $R_h \succ 0$ is a known matrix weighting the human's control effort.

2) Fixed-Time Parameter Estimation: Estimate θ using a fixed-time gradient-based update law. Let $V_h(X;\theta)$ be the value function corresponding to the human reward r_θ. The associated Bellman error is:

$$\delta_\theta = r_\theta(X, U_h) + \nabla V_h^\top (F + GU_h) - \gamma V_h \tag{26}$$

Assuming V_h can be approximated, the reward parameters $\hat{\theta}$ are updated using a composite fixed-time rule based on both the current reward error $\delta_\theta(t)$ and historical errors δ_θ^k:

$$\begin{aligned} \dot{\hat{\theta}} = & -\alpha_{\theta 1} \Gamma_\theta \nabla_{\hat{\theta}} \delta_\theta \left(\lceil \delta_\theta(t) \rfloor^{\gamma_1} + \lceil \delta_\theta(t) \rfloor^{\gamma_2}\right) \\ & - \frac{\alpha_{\theta 2}}{N} \Gamma_\theta \sum_{k=1}^{N} \nabla_{\hat{\theta}} \delta_\theta^k \left(\lceil \delta_\theta^k \rfloor^{\gamma_1} + \lceil \delta_\theta^k \rfloor^{\gamma_2}\right) \end{aligned} \tag{27}$$

where $\alpha_{\theta 1}, \alpha_{\theta 2} > 0$ are learning rates, $\Gamma_\theta \succ 0$ is a gain matrix, δ_θ^k is the historical reward error, and the exponents $0 < \gamma_1 < 1, \gamma_2 > 1$ ensure fixed-time convergence of $\hat{\theta}$.

Remark 3 (On the Human Optimality Assumption)**.** *The assumption of perfect human optimality is a simplification. The IOC approach interprets demonstrations as near-optimal, aligning with the concept of 'bounded rationality' [44]. While more sophisticated models like Level-k theory [6], [45] or Gaussian noise distributions [46] can be used to model suboptimal decision-making, The current method effectively extracts the operator's primary intent. This provides a strong foundation for synthesizing a robust control policy, and future work will integrate more advanced human models based on the related research [6].*

Remark 4 (Learning from Human Interaction)**.** *The FxT-IOC framework learns operator intent from expert control data.*

As depicted in Fig. 3, its IOC module infers the human's reward function (r_θ) by estimating parameters ($\hat{\theta}$) that explain the observed actions. The core innovation is the fixed-time update law (27), which guarantees learning completes within a predictable, bounded time. This learned intent then guides the autonomous control policy. Unlike traditional IOC methods with only asymptotic guarantees [15], [22], proposed approach ensures finite-time convergence. While other works explore fixed-time learning [31], proposed framework offers a more comprehensive solution by integrating state constraints (PPC) and input saturation, features often unaddressed in prior learning-based control schemes [6].

B. Fixed-Time Synthesis Optimal Control

The autonomous control component uses an Actor-Critic structure to achieve optimal control according to (18), ensuring fixed-time stability and constraint satisfaction, guided by the learned human reward $\hat{\theta}$.

1) Critic Network: The critic approximates the optimal value function $V^*(X)$ of Eq. (20), which is the solution to the HJB equation (21), using a neural network:

$$V^*(X) \approx \hat{V}(X) = \hat{W}_c^\top \varphi_c(X) \tag{28}$$

where $\hat{W}_c \in \mathbb{R}^{N_c}$ are the critic weights updated via a fixed-time learning law (Detailed in Sec. IV-C), and $\varphi_c(X)$ is a basis function.

2) Actor Network: The actor approximates the optimal policy $U^*(X)$ of Eq. (22). Using $\nabla V^*(X) \approx \nabla\varphi_c(X)^\top \hat{W}_c$, the estimated control is given by:

$$\hat{U}(X;\hat{W}_a) = -\mu \tanh\left(\frac{1}{2\mu} R^{-1} G(X)^\top \nabla\varphi_a(X)^\top \hat{W}_a\right) \tag{29}$$

where $\hat{W}_a$ are the actor weights. This policy inherently respects the saturation limits μ.

3) Bellman Error: The critic weights $\hat{W}_c$ are updated to minimize the following Bellman error derived from the HJB equation (21):

$$\begin{aligned}\delta(X) = {} & \underbrace{\|X\|_Q^{\gamma_1} + \|X\|_Q^{\gamma_2} + \mathit{\Pi}(\hat{U}(X))}_{\text{Instantaneous Cost } r(X,\hat{U})} \\ & + \nabla\hat{V}(X)^\top (F(X) + G(X)\hat{U}(X)) - \gamma\hat{V}(X)\end{aligned} \tag{30}$$

This Bellman error $\delta(X)$ drives the fixed-time update for the critic weights $\hat{W}_c$, enabling the learning of an optimal constrained policy with a guaranteed convergence time.

C. Fixed-Time Composite Learning Update

The FxT-IOC framework uses fixed-time composite learning with experience replay, leveraging current and historical data for enhanced efficiency and robustness.

1) Experience Replay Buffer: A structured experience buffer $\mathcal{D}(t)$ stores past states and associated learning signals:

$$\mathcal{D}(t) = \{\mathcal{E}(t)\} \cup \{\mathcal{E}^k\}_{k=1}^N \tag{31}$$

where $\mathcal{E}(t) = \{X(t), \hat{U}(t), r(t), X(t+1)\}$ represents the current transition data, and $\{\mathcal{E}^k\}_{k=1}^N = \{X^k, \hat{U}^k, r^k, X^{k+1}\}_{k=1}^N$ contains N historical samples.

2) Temporal Bellman Error: For each stored experience k, the temporal Bellman error ζ^k evaluates the consistency of the current value function estimate $\hat{V}(X) = \hat{W}_c^\top \varphi_c(X)$ with the observed transition and reward:

$$\zeta^k = r^k + \gamma\hat{W}_c^\top \varphi_c(X^{k+1}) - \hat{W}_c^\top \varphi_c(X^k) \tag{32}$$

This error assesses the value function's prediction accuracy using past data. The current Bellman error $\delta(t)$ uses the current data $\mathcal{E}(t)$ and policy $\hat{U}(t)$ per (30). The reward parameter error δ_θ is given by (26).

3) Composite Fixed-Time Update Laws: The updates minimize a composite objective with fixed-time terms for current and historical errors. Inspired by [31], the critic weights $\hat{W}_c$ are updated using a fixed-time rule based on the current Bellman error $\delta(t)$ and historical temporal errors ζ^k:

$$\begin{aligned}\dot{\hat{W}}_c = & -k_{c1}\Gamma_c \frac{\Psi(t)}{\|\Psi(t)\|^2 + \epsilon_1}\left(\lceil\delta(t)\rfloor^{\gamma_1} + \lceil\delta(t)\rfloor^{\gamma_2}\right) \\ & - \frac{k_{c2}}{N}\Gamma_c \sum_{k=1}^N \frac{\Xi^k}{\|\Xi^k\|^2 + \epsilon_2}\left(\lceil\zeta^k\rfloor^{\gamma_1} + \lceil\zeta^k\rfloor^{\gamma_2}\right)\end{aligned} \tag{33}$$

where $k_{c1}, k_{c2} > 0$ are adaptation gains, $\Psi(t) = \nabla\psi_c[F(X(t)) + G(X(t))\hat{U}(X(t))]$ is the current feature difference vector, $\Xi^k = \nabla\psi_c[F(X^k) + G(X^k)\hat{U}(X^k)]$ is the historical feature difference vector, $\Gamma_c \succ 0$ is a positive definite gain matrix, $\epsilon_1, \epsilon_2 > 0$ ensure numerical stability, and $0 < \gamma_1 < 1, \gamma_2 > 1$ provide fixed-time convergence. For the actor network, the weights $\hat{W}_a$ are updated using a fixed-time composite law based on the actor error $\delta_a = \hat{U}(X;\hat{W}_a) - \hat{U}(X;\hat{W}_c)$. The update is:

$$\begin{aligned}\dot{\hat{W}}_a = & -\alpha_{a1}\Gamma_a \frac{\Psi(t)}{\|\Psi(t)\|^2 + \epsilon_1}\left(\lceil\delta_a(t)\rfloor^{\gamma_1} + \lceil\delta_a(t)\rfloor^{\gamma_2}\right) \\ & - \frac{\alpha_{a2}}{N}\Gamma_a \sum_{k=1}^N \frac{\Xi^k}{\|\Xi^k\|^2 + \epsilon_2}\left(\lceil\delta_a^k\rfloor^{\gamma_1} + \lceil\delta_a^k\rfloor^{\gamma_2}\right)\end{aligned} \tag{34}$$

where $\alpha_{a1}, \alpha_{a2} > 0$ are learning rates, $\Gamma_a \succ 0$ is a gain matrix, and δ_a^k is the historical actor error. This fixed-time update uses current and past data. See Algorithm 1.

Remark 5 (Coordination of PPC and Input Saturation)**.** *A critical challenge in human-UAV systems is reconciling aggressive state constraint enforcement with inherent input saturation from actuators and human interfaces. Demanding rapid error convergence via PPC can lead to control commands that exceed physical limits. Proposed framework resolves this by coordinating the PPC convergence rate λ_i with the input saturation level μ in the control law (22). This co-tuning ensures that state constraints are met without demanding unrealizable control actions, thereby guaranteeing both safety and performance, a challenge also addressed in [29], [41].*

Remark 6 (Comparison with Existing Methods)**.** *As summarized in Table II, prior methods often lack guaranteed time-critical convergences [15], comprehensive constraint handling [38], or full integration of human intent [6]. Proposed FxT-IOC framework uniquely combines these critical features: fixed-time convergence, state (PPC) and input constraints, human integration via IOC, and stability guarantees, offering a holistic solution for human-UAV collaboration.*

Algorithm 1 FxT-IOC Framework Algorithm

```
Input: $\hat{W}_c(0)$, $\hat{W}_a(0)$, $\hat{\theta}(0)$, $\Gamma_c$, $\Gamma_a$, $\Gamma_\theta$, $k_{c1}, k_{c2}$, $k_{a1}, k_{a2}$, $\alpha_{\theta 1}$, $\alpha_{\theta 2}$, $\gamma_1$, $\gamma_2$, $N$, $\epsilon_{term}$, $T_{end}$.
Output: $\hat{U}(X), \hat{W}_c, \hat{\theta}$.
Initialize buffer $\mathcal{D} = \emptyset$.
while $t < T_{end}$ and not converged do
    Get state $X(t)$.
    Compute $\hat{U}(t)$ using (29).
    Apply $U(t)$; Observe $X(t+1), r(t)$.
    Store $\{X(t), \hat{U}(t), r(t), X(t+1)\}$ in $\mathcal{D}$.
    Compute $\delta$, $\zeta$, $\delta_\theta$ using (30), (32), (26).
    Sample minibatch $\{\mathcal{E}^k\}_{k=1}^{N_{batch}}$ from $\mathcal{D}$.
    Compute $\zeta^k, \delta_\theta^k$ for the minibatch.
    Update $\hat{\theta}$ using (27).
    Update $\hat{W}_c$ using (33).
    Update $\hat{W}_a$ using (34).
    $t \leftarrow t + dt$.
end while
return $\hat{U}(X), \hat{W}_c, \hat{\theta}$.
```

TABLE II
COMPARISON OF FXT-IOC WITH EXISTING METHODS

Method	Fixed-Time Convergence	State Constraints	Input Constraints	Human Integration	Stability Guarantees
Conventional IOC [15]	✗	✗	!	!	✓
ADP-based Control [6]	✗	✗	✓	!	✓
Game-based IOC [17]	✗	✗	✗	✓	✓
Trust-region IOC [18]	✗	✓	✗	✓	✓
Fixed-time Control [38]	✓	!	!	✗	✓
Proposed FxT-IOC	✓	✓	✓	✓	✓

Legend: ✓Full support; ✗Not supported; **!** Partial support. Colors indicate the level of support.

Remark 7 (Actor Update Stability)**.** *Actor updates can become unstable when driven by inaccurate critic gradients [29], [47]. This framework employs two mechanisms to ensure stability. First, the critic's composite learning update (33) uses historical data to produce a smoother, more stable value estimate $\hat{V}(X)$. Second, the actor's policy (29) is bounded by a hyperbolic tangent function, $\hat{U}(X) = -\mu \tanh(\cdot)$. This inherently prevents excessive control signals, even with transient critic errors, promoting a robust learning process for actor.*

Remark 8 (Motivation for Human Intent Learning)**.** *Human intent learning combines the strategic adaptability of human operators with the precision of autonomous control. While manual control is error-prone and full autonomy may lack situational judgment, the proposed FxT-IOC framework offers a synergistic solution. It uses Inverse Optimal Control (IOC) to learn the operator's reward function from demonstrations. Unlike traditional IOC methods with only asymptotic convergence, proposed approach guarantees this learning occurs within a fixed time—a critical advantage for real-time collaboration. The learned intent then guides an autonomous policy that ensures high-precision tracking and robust constraint satisfaction, effectively blending human guidance with machine reliability.*

D. Stability Analysis

This section establishes theoretical guarantees for the FxT-IOC framework.

Assumption 3 (Approximation Properties [39])**.** *The optimal value function $V^*(X)$ and human value function $V_h(X)$ are approximated using basis functions $\varphi_c(X), \varphi_h(X)$ with ideal weights W_c^*, W_h^* and bounded errors $\varepsilon_c(X), \varepsilon_h(X)$, respectively, i.e., $\|\varepsilon_c\| \leq \bar{\varepsilon}_c, \|\varepsilon_h\| \leq \bar{\varepsilon}_h$. The basis functions $\varphi_c, \varphi_h, \phi_r$ and their gradients are bounded. All approximation errors, Bellman errors, and estimated weights are assumed to be bounded.*

Assumption 4 (Persistent Excitation Condition [37])**.** *The regressors $\Psi(t)$ (current) and Ξ^k (historical), including those for the $\hat{\theta}$ update, satisfy a composite Persistent Excitation (PE) condition: there exist $T, \rho > 0$ such that*

$$\lambda_{\min}\left\{\int_t^{t+T} \frac{\Psi(\tau)\Psi^\top(\tau)}{\|\Psi(\tau)\|^2 + \epsilon_1} d\tau + \frac{1}{N}\sum_{k=1}^{N} \frac{\Xi^k[\Xi^k]^\top}{\|\Xi^k\|^2 + \epsilon_2}\right\} \geq \rho$$

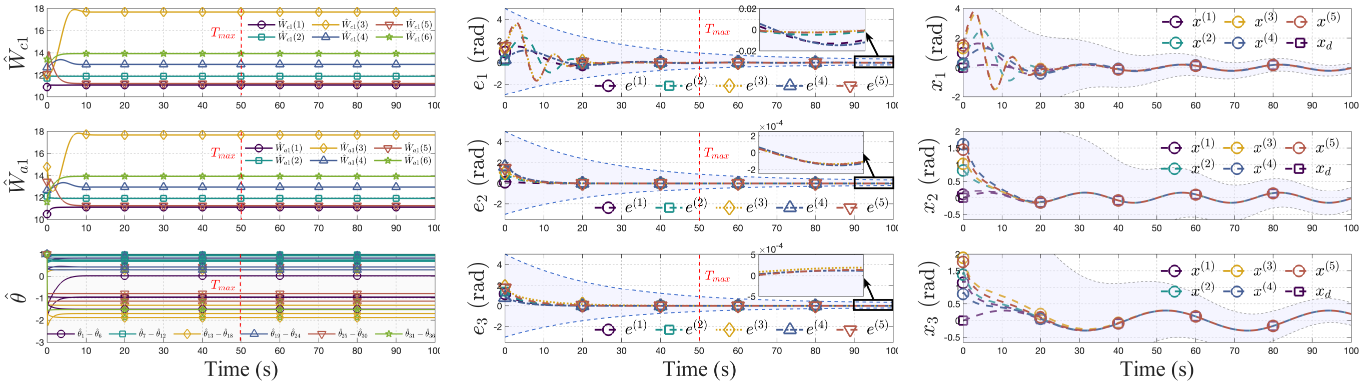
Remark 9 (Practical Considerations for the PE Condition)**.** *The Persistent Excitation (PE) condition is crucial for parameter convergence in adaptive control, though its analytical verification for nonlinear systems is often intractable. In practice, the PE condition is often satisfied by designing sufficiently rich reference trajectories [28]. Furthermore, proposed composite learning approach, which leverages a replay buffer of historical data (Eq. (31)), relaxes the requirement to an integral PE condition, enhancing robustness against temporary data sparsity [37]. The controller finds parameters sufficient for stabilization, demonstrating robustness to weakened excitation. Further research could explore sparse PE conditions [48] to ensure persistent excitation in practical scenarios.*

Let the parameter estimation errors be $\tilde{W}_c = W_c^* - \hat{W}_c$, $\tilde{\theta} = \theta^* - \hat{\theta}$, and $\tilde{W}_a = W_a^* - \hat{W}_a$. The combined error vector is $\tilde{Z}(t) = [\tilde{W}_c^\top, \tilde{\theta}^\top, \tilde{W}_a^\top]^\top$. The combined gain matrix is $\Gamma_Z = \text{blkdiag}(\Gamma_c, \Gamma_\theta, \Gamma_a)$.

Theorem 1 (Fixed-Time Convergence of Parameter Estimation)**.** *Under Assumptions 1-4, consider the augmented system (14) controlled by the policy derived from (29) with parameter updates from Algorithm 1 ($0 < \gamma_1 < 1, \gamma_2 > 1$). Let $\tilde{Z}(t) = [\tilde{W}_c(t)^\top, \tilde{\theta}(t)^\top, \tilde{W}_a(t)^\top]^\top$ be the combined parameter estimation error vector. Then, the estimation error converges in fixed time to a residual set around the origin. Specifically, consider the Lyapunov function $L(\tilde{Z}) = \frac{1}{2}\tilde{Z}^\top \Gamma_Z^{-1} \tilde{Z}$. There exist positive constants c_1, c_2, exponents $\eta_1 = (1+\gamma_1)/2 \in (1/2, 1)$, $\eta_2 = (1+\gamma_2)/2 > 1$, and a bound $\Pi_L \geq 0$ (dependent on approximation errors and disturbances) such that the time derivative of L satisfies:*

$$\dot{L}(\tilde{Z}) \leq -c_1 L^{\eta_1} - c_2 L^{\eta_2} + \Pi_L \tag{35}$$

1) *The error $\tilde{Z}$ converges to the residual set $\Omega_L = \{\tilde{Z} \mid L(\tilde{Z}) \leq \max\{(\Pi_L/(c_1 - c_1\epsilon))^{1/\eta_1}, (\Pi_L/(c_2 - c_2\epsilon))^{1/\eta_2}\}\}$ for any small constant $\epsilon \in (0, 1)$.*



(a) Actor-critic and IOC weights (b) Position tracking errors e_1, e_2, e_3 (c) State trajectory tracking x_1, x_2

Fig. 4. Simulation results: (a) Convergence of critic and IOC weights, demonstrating learning stability. (b) Position tracking errors over time, showing adherence to performance bounds. (c) Tracking of selected state components (e.g., roll ϕ and pitch θ) against their desired trajectories.

2) *The convergence time T required to reach Ω_L is bounded by $T \leq T_{\max}$, where*

$$T_{\max} \approx \frac{1}{c_1\epsilon(1-\eta_1)} + \frac{1}{c_2\epsilon(\eta_2-1)} \tag{36}$$

This upper bound $T_{\max}$ is independent of the initial estimation error $\tilde{Z}(0)$.

The constants c_1, c_2, Π_L depend on the system dynamics (Assumption 1), cost function (Assumption 2), approximation capabilities (Assumption 3), PE levels (Assumption 4), and the chosen learning gains and fixed-time exponents.

Proof. The proof of fixed-time convergence for the estimation error $\tilde{Z}$, based on Assumptions 1-4 and Lemma 1, is detailed in Appendix A. □

Remark 10 (Convergence to a Residual Set)**.** *Theorem 1 establishes that parameter errors converge to a residual set Ω_L, a practical outcome for adaptive systems with NN approximators and disturbances. The size of this set depends on the approximation errors ($\bar{\varepsilon}_c, \bar{\varepsilon}_h$) and system disturbances ($\bar{D}$). While the error bound can be reduced by improving the NN or increasing learning gains, the key contribution is guaranteeing this convergence occurs within a fixed time $T_{\max}$, which is critical for real-time performance guarantees.*

Remark 11 (Novelty of the FxT-IOC Approach)**.** *The core novelty of this work is a **fixed-time learning mechanism** for Inverse Optimal Control (IOC), designed for real-time human-UAV collaboration. Traditional IOC methods offer only asymptotic convergence guarantees [24], [49], which are often insufficient for dynamic, safety-critical applications. Proposed FxT-IOC framework overcomes this by using the update law (27), which guarantees that human reward parameters ($\hat{\theta}$) converge within a predetermined time, as proven in Theorem 1. This predictable, fast convergence is a key distinction from conventional IOC and is critical for reliable human-in-the-loop systems.*

V. Experimental Validation

This section evaluates the FxT-IOC framework using quadrotor UAV simulations and hardware experiments. The fixed-time learning convergence and trajectory tracking under prescribed constraints are validated.

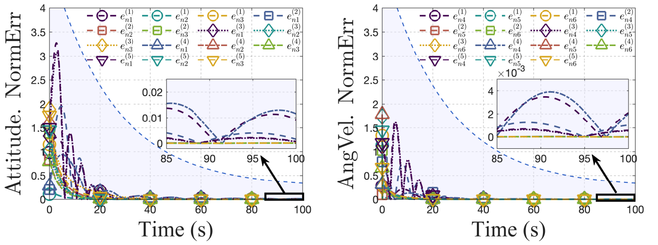


Fig. 5. Normalized tracking error demonstrating FxT convergence.

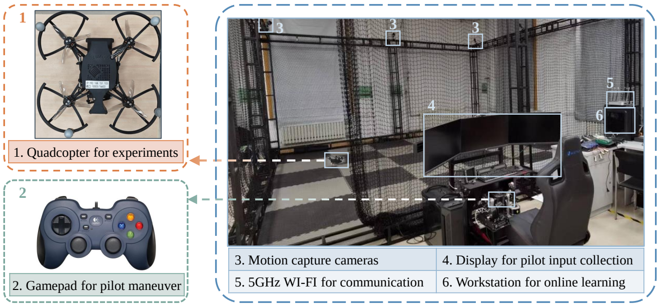

1
1. Quadcopter for experiments
2
2. Gamepad for pilot maneuver
3
4
5
6
3. Motion capture cameras
4. Display for pilot input collection
5. 5GHz WI-FI for communication
6. Workstation for online learning

Fig. 6. Hardware experimental platform comprising the Droneyee-X150 UAV, OptiTrack motion capture system, and control station.

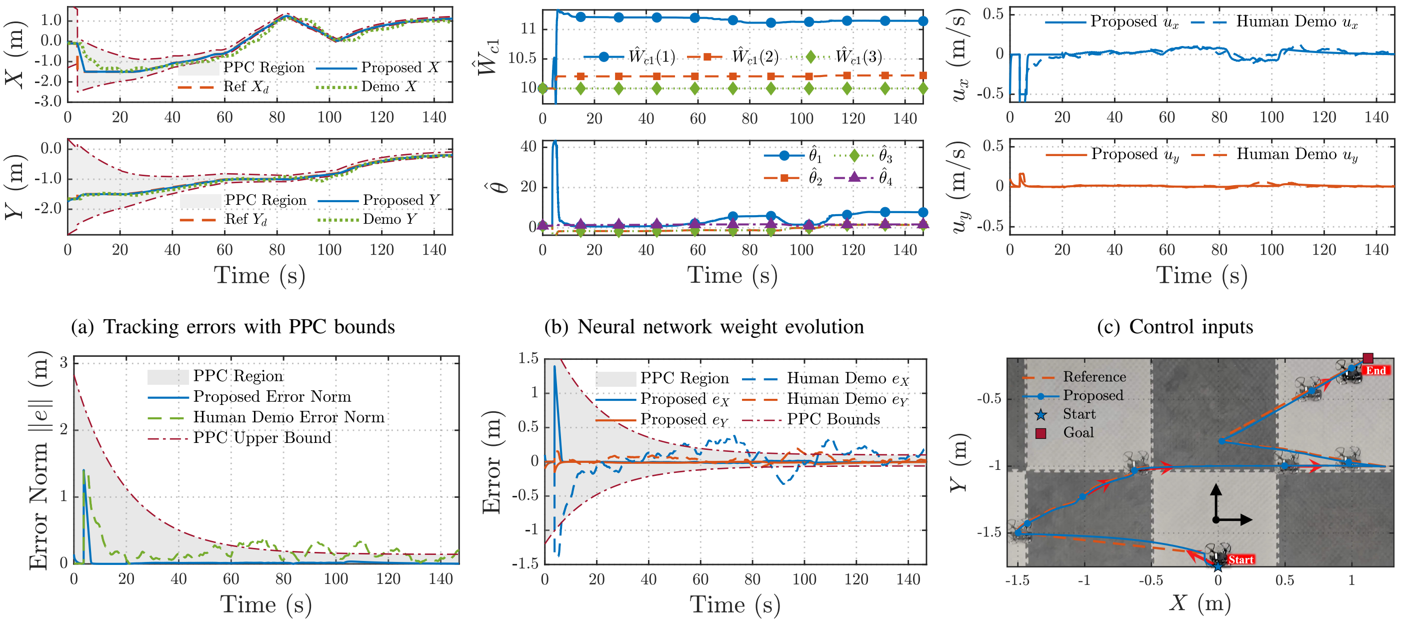
TABLE III
Key Simulation Parameters

Parameter	Value	Parameter	Value
State Penalty Q	$2\mathbf{I}_6$	FxT Exponent γ_1	0.8
Control Penalty R	$\mathbf{I}_4$	FxT Exponent γ_2	1.2
Gain Γ_* $(*=a,c)$	diag(10)	Buffer Size N	30
IOC Gain Γ_θ	diag(5)	Time Step Δt	0.001 s
First Rate k_{*1}	0.05	Sim. Duration T_{sim}	100 s
Second Rate k_{*2}	0.15	Discount γ	0.90

A. Simulation Results

First, the simulation results are presented, followed by hardware experiments to demonstrate the practical applicability of the proposed framework.

1) Simulation Setup: The simulations utilize a standard quadrotor attitude dynamics model as described in Section II, focusing on tracking a desired attitude trajectory



(a) Tracking errors with PPC bounds

(b) Neural network weight evolution

(c) Control inputs

(d) Fixed-time convergence validation

(e) Position tracking performance

(f) Trajectory tracking performance

Fig. 7. Hardware results: (a) Weight convergence. (b) Control inputs. (c) Error norm convergence. (d) Bounded Errors. (e) Positions. (f) Trajectory.

$x_d(t) = [\phi_d(t), \theta_d(t), \psi_d(t), \dot{\phi}_d(t), \dot{\theta}_d(t), \dot{\psi}_d(t)]^\top$. The reference trajectory involves sinusoidal components designed to excite the system dynamics: $\phi_d(t) = 0.2\sin(0.25t)$, $\theta_d(t) = 0.15\sin(0.25t)$, $\psi_d(t) = 0.3\sin(0.15t)$. The control goal, with tracking error $e = x - x_d$, is to minimize state deviations within a fixed time boundary. For approximation, neural network basis functions are employed as:

$$\varphi_i = \left[e_1^2, e_2^2, e_1e_2, e_3^2, e_1e_3, e_2e_3, e_4^2, e_5^2, e_6^2, e_1e_4, e_2e_5, e_3e_6\right]^\top$$

The simulation parameters are summarized in Table III. The fixed-time exponents are set to $\gamma_1 = 0.8$ and $\gamma_2 = 1.2$ to leverage both fast initial adaptation and precise final convergence. The simulations were performed in MATLAB/Simulink with a step of $\Delta t = 0.001$s over a duration of $T_{sim} = 100$ s.

2) Fixed-Time Convergence Analysis: Theorem 1 establishes that the parameter estimation errors converge within a fixed time $T_{\max}$, bounded by:

$$T_{\max} \approx 1/(c_1\epsilon(1-\eta_1)) + 1/(c_2\epsilon(\eta_2 - 1))$$

where $\eta_1 = (1+\gamma_1)/2$ and $\eta_2 = (1+\gamma_2)/2$. The coefficients c_1 and c_2 are positive constants influenced by learning rates, the PE condition (Assumption 4), and neural network approximation accuracy. While a precise analytical derivation is complex, an illustrative calculation is provided. Using the simulation parameters from Table III ($\gamma_1 = 0.8, \gamma_2 = 1.2$), it comes $\eta_1 = 0.9$ and $\eta_2 = 1.1$. For illustrative purposes, assuming the learning gains and PE conditions yield effective coefficients $c_1 \approx 1.0$ and $c_2 \approx 0.25$, and setting $\epsilon = 1$, the terms evaluate as:

$$\text{First term: } 1/(c_1\epsilon(1-\eta_1)) \approx 10 \text{ s}$$
$$\text{Second term: } 1/(c_2\epsilon(\eta_2-1)) \approx 40 \text{ s}$$

Consequently, the theoretical upper bound is $T_{\max} \approx 50$ seconds in this example.

Remark 12 (Observed vs. Theoretical Convergence)**.** *The theoretical $T_{\max}$ provides a conservative upper bound for parameter convergence (approximately 50 s in this example). In practice, convergence is much faster, as simulations show that the learning weights (Fig. 4(a)) and estimation error (Fig. 5) stabilize in about 20 s. This gap highlights the efficiency of the fixed-time scheme, ensuring rapid, predictable convergence well within the theoretical bound, regardless of initial conditions.*

3) Performance Analysis: The simulation results validate the core features of the FxT-IOC framework.

1) **Learning Convergence:** As analyzed above, the critic ($\hat{W}_c$) and IOC ($\hat{\theta}$) weights converge smoothly to stable values within a fixed time (Fig. 4(a)). The normalized estimation error also converges rapidly (Fig. 5), confirming the fixed-time stability guaranteed by Theorem 1.

2) **Tracking Performance:** Position tracking errors are quickly minimized and maintained within small bounds (Fig. 4(b)), demonstrating effective control and adherence to prescribed performance constraints. The UAV accurately follows the desired trajectories for position states (Fig. 4(c)).

The simulations confirm that the FxT-IOC framework provides stable, rapid fixed-time learning convergence and precise trajectory tracking, outperforming traditional asymptotic methods in predictability and efficiency.

TABLE IV
HARDWARE EXPERIMENTAL PARAMETERS

Parameter	Value	Parameter	Value
Control Freq. f_c	30 Hz	FxT Exponent γ_1	0.5
Time Step Δt	33.3 ms	FxT Exponent γ_2	2.0
UAV Mass m	310 g	Initial $\hat{W}_c(0)$	$[10, 10, 10]^\top$
State Penalty Q	$\mathrm{diag}(100, 100)$	Critic Rate k_{c1}	1.0
Input Penalty R	$\mathrm{diag}(500, 500)$	Critic Rate k_{c2}	0.01
Input Saturation μ	±1.0 m/s	Discount γ	0.99

B. Hardware Implementation

To validate the practical applicability and real-world performance of the FxT-IOC framework, experiments were conducted on a physical quadrotor platform.

1) Experimental Setup: The hardware setup (Fig. 6) included a Droneyee-X150 UAV (RK3566 onboard), an OptiTrack system (120 Hz), a control station (30 Hz WiFi link), and a Logitech F310 gamepad (Fig. 1). Key parameters are listed in Table IV, including fixed-time exponents $\gamma_1 = 0.5, \gamma_2 = 2.0$. Simplified basis functions $\varphi_i = [e_X^2, e_X e_Y, e_Y^2]^\top$ were used for reduced computation. Other parameters matched the simulations.

2) Experimental Protocol and Results: The experiments proceeded in three stages: 1. **Demonstration:** Expert manual flight data was collected along predefined trajectories using the gamepad shown in Fig. 6. The UAV was flown in a controlled environment, capturing position and control inputs. 2. **Learning:** The FxT-IOC processed the data to estimate human reward parameters ($\hat{\theta}$) via IOC and train the critic network ($\hat{W}_c$). 3. **Validation:** The learned controller autonomously flew the UAV with the simulation platform being the same as in V-A using position dynamics from (6), tracking reference trajectories while adhering to constraints. Tracking accuracy and convergence were assessed.

The hardware experiments on the Droneyee-X150 platform validated the FxT-IOC framework's real-world applicability. Key findings include:

1) **Fixed-Time Learning Validation:** The experiments empirically confirmed the fixed-time convergence guarantee. Estimation error norms converged rapidly within a predictable time, as shown in Fig. 7(d), and network weights evolved stably, as shown in Fig. 7(b), supporting Theorem 1. Similar to the simulation analysis, the theoretical fixed-time convergence bound for the experiments is estimated. According to Theorem 1, the convergence time is bounded by $T_{\max} \approx 1/(c_1\epsilon(1-\eta_1)) + 1/(c_2\epsilon(\eta_2 - 1))$. Using the hardware parameters from Table IV ($\gamma_1 = 0.5, \gamma_2 = 2.0$), it comes $\eta_1 = 0.75$ and $\eta_2 = 1.5$. Assuming illustrative effective coefficients $c_1 \approx 0.5$ and $c_2 \approx 0.1$ based on the experimental setup, the theoretical upper bound is estimated as $T_{\max} \approx 8 + 20 = 28$ seconds. The experimental results in Fig. 7(d) show that the error norm stabilizes in approximately 10-15 seconds, well within this conservative theoretical bound, confirming the rapid and predictable convergence of the FxT-IOC framework in a real-world setting.

2) **Tracking and Constraint Handling:** High-fidelity trajectory tracking was achieved and shown in Figs. 7(a) and 7(f), which is an illustrative snapshot of the UAV's trajectory similar to previous work [50]. Position errors were kept small and strictly within the prescribed performance bounds, given in Fig. 7(e), validating the PPC mechanism. Fig. 7(c) shows that control inputs consistently respected saturation limits and were smoother than the manual inputs, demonstrating the controller's robustness and stability.

A notable feature in the results is the initial sharp decrease observed in the tracking error norm of Fig. 7(a). This transient behavior is caused by the initial discrepancy between the UAV's physical starting position and the beginning of the human operator's demonstration trajectory. The controller is designed to rapidly close this initial gap to commence tracking. It has been verified that this initial transient does not negatively impact the overall performance or stability of the learning system, provided the initial error is within the prescribed performance bounds. This demonstrates the controller's robustness in handling practical initialization mismatches.

Additionally, the experiments highlighted the importance of tuning the fixed-time exponents to achieve optimal performance across varying conditions. Future work will focus on refining these parameters to enhance adaptability in dynamic environments.

VI. Conclusion

This paper presented a Fixed-Time Inverse Optimal Control (FxT-IOC) framework for human-UAV collaboration under constraints. By integrating IOC for intent inference, fixed-time learning for predictable convergence, PPC for state constraints, and input saturation handling, the framework ensures efficient and safe operation. Theoretical analysis guarantees fixed-time stability, while simulations and experiments validated superior tracking, faster convergence, and robust constraint satisfaction. This work offers a practical and robust solution for developing reliable human-UAV systems. Several limitations, such as the simplified human intent model discussed in Remark 3 and the reliance on the PE condition discussed in Remark 9, are acknowledged. Future work will explore advanced human modeling such as Level-k theory and applications using Gaussian preferences regression.

Appendix

Proof of Theorem 1. Define the augmented parameter estimation error $\tilde{Z}(t) = [\tilde{W}_c(t)^\top, \tilde{\theta}(t)^\top, \tilde{W}_a(t)^\top]^\top$, where $\tilde{W}_c = W_c^* - \hat{W}_c$, $\tilde{\theta} = \theta^* - \hat{\theta}$, and $\tilde{W}_a = W_a^* - \hat{W}_a$ (assuming an explicit actor network $\hat{U}(X; \hat{W}_a)$ with ideal weights W_a^* is used). Let $\Gamma_Z = \text{blkdiag}(\Gamma_c, \Gamma_\theta, \Gamma_a)$ be the combined positive definite gain matrix, including the actor gain Γ_a. Consider the Lyapunov function candidate $L(\tilde{Z}) = \frac{1}{2}\tilde{Z}^\top \Gamma_Z^{-1} \tilde{Z} = \frac{1}{2}\tilde{W}_c^\top \Gamma_c^{-1} \tilde{W}_c + \frac{1}{2}\tilde{\theta}^\top \Gamma_\theta^{-1} \tilde{\theta} + \frac{1}{2}\tilde{W}_a^\top \Gamma_a^{-1} \tilde{W}_a$. Its time derivative along the error dynamics is:

$$\dot{L}(\tilde{Z}) = -\tilde{W}_c^\top \Gamma_c^{-1} \dot{\hat{W}}_c - \tilde{\theta}^\top \Gamma_\theta^{-1} \dot{\hat{\theta}} - \tilde{W}_a^\top \Gamma_a^{-1} \dot{\hat{W}}_a \tag{37}$$

Substituting the update laws (27), (33), and (34) into the expression for $\dot{L}(\tilde{Z})$, and relating the errors $(\delta, \zeta^k, \delta_\theta, \delta_\theta^k, \delta_a, \delta_a^k)$ to the parameter errors $(\tilde{W}_c, \tilde{\theta}, \tilde{W}_a)$ and bounded approximation errors $(\varepsilon_c, \varepsilon_\theta, \varepsilon_a)$ under the extended Assumption 3, it has:

$$\dot{L}(\tilde{Z}) \leq -\sum_{i \in \{c,\theta,a\}} \sum_{j \in \{1,2\}} k_{ij} \left(\text{PE}_i\right) \|\tilde{Z}_i\|^{1+\gamma_j} + \Pi_\varepsilon(\tilde{Z}, \varepsilon)$$

where k_{ij} are positive constants, PE_i represents terms involving the Persistent Excitation condition (Assumption 4, extended to actor regressors), $\tilde{Z}_c = \tilde{W}_c$, $\tilde{Z}_\theta = \tilde{\theta}$, $\tilde{Z}_a = \tilde{W}_a$, and Π_ε contains bounded terms from approximation errors, satisfying $\|\Pi_\varepsilon\| \leq \Pi_1 \|\tilde{Z}\| + \Pi_0$. Using standard inequalities and the PE condition, this simplifies to:

$$\dot{L}(\tilde{Z}) \leq -c_1' \|\tilde{Z}\|^{1+\gamma_1} - c_2' \|\tilde{Z}\|^{1+\gamma_2} + \Pi_1 \|\tilde{Z}\| + \Pi_0 \tag{38}$$

Expressing this in terms of L using $\underline{\lambda}(\Gamma_Z^{-1})\|\tilde{Z}\|^2 \leq 2L \leq \bar{\lambda}(\Gamma_Z^{-1})\|\tilde{Z}\|^2$ and applying Young's inequality to bound the linear term in $\|\tilde{Z}\| \propto L^{1/2}$, it obtains:

$$\dot{L}(\tilde{Z}) \leq -c_3 L^{\eta_1} - c_4 L^{\eta_2} + \Pi_L \tag{39}$$

where $\eta_1 = (1+\gamma_1)/2 \in (1/2, 1)$, $\eta_2 = (1+\gamma_2)/2 > 1$, and c_3, c_4, Π_L are positive constants depending on system parameters, gains, PE levels, and approximation bounds. By Lemma 1, inequality (39) ensures that the augmented parameter error $\tilde{Z}(t)$ converges in fixed time to a residual set Ω_Z. The convergence time $T_{\max}$ is bounded by $T_{\max} \approx \frac{1}{c_3\epsilon(1-\eta_1)} + \frac{1}{c_4\epsilon(\eta_2-1)}$, independent of the initial error $\tilde{Z}(0)$. The proof is completed. □

References

[1] S. Islam, P. X. Liu, A. E. Saddik, R. Ashour, J. Dias, and L. D. Seneviratne, "Artificial and Virtual Impedance Interaction Force Reflection-Based Bilateral Shared Control for Miniature Unmanned Aerial Vehicle," *IEEE Transactions on Industrial Electronics*, vol. 66, DOI 10.1109/TIE.2018.2793178, no. 1, pp. 329–337, Jan. 2019.

[2] J. Tan, S. Xue, Q. Guan, T. Niu, H. Cao, and B. Chen, "Unmanned aerial-ground vehicle finite-time docking control via pursuit-evasion games," *Nonlinear Dynamics*, DOI 10.1007/s11071-025-11021-6, Mar. 2025.

[3] M. Marcano, S. Díaz, J. Pérez, and E. Irigoyen, "A Review of Shared Control for Automated Vehicles: Theory and Applications," *IEEE Transactions on Human-Machine Systems*, vol. 50, DOI 10.1109/THMS.2020.3017748, no. 6, pp. 475–491, Dec. 2020.

[4] E. Eraslan, Y. Yildiz, and A. M. Annaswamy, "Shared Control Between Pilots and Autopilots: An Illustration of a Cyberphysical Human System," *IEEE Control Systems*, vol. 40, DOI 10.1109/MCS.2020.3019721, no. 6, pp. 77–97, Dec. 2020.

[5] A. B. Farjadian, A. M. Annaswamy, and D. Woods, "Bumpless Reengagement Using Shared Control between Human Pilot and Adaptive Autopilot," *IFAC-PapersOnLine*, vol. 50, DOI 10.1016/j.ifacol.2017.08.925, no. 1, pp. 5343–5348, Jul. 2017.

[6] J. Tan, J. Wang, S. Xue, H. Cao, H. Li, and Z. Guo, "Human–Machine Shared Stabilization Control Based on Safe Adaptive Dynamic Programming With Bounded Rationality," *International Journal of Robust and Nonlinear Control*, DOI 10.1002/rnc.7931, p. rnc.7931, Mar. 2025.

[7] K. Ghonasgi, T. Higgins, M. E. Huber, and M. K. O'Malley, "Crucial hurdles to achieving human-robot harmony," *Science Robotics*, vol. 9, DOI 10.1126/scirobotics.adp2507, no. 96, p. eadp2507, Nov. 2024.

[8] A. Broad, I. Abraham, T. Murphey, and B. Argall, "Data-driven Koopman operators for model-based shared control of human–machine systems," *The International Journal of Robotics Research*, vol. 39, DOI 10.1177/0278364920921935, no. 9, pp. 1178–1195, Aug. 2020.

[9] M. Ma and L. Cheng, "A Human-Robot Collaboration Controller Utilizing Confidence for Disagreement Adjustment," *IEEE Transactions on Robotics*, DOI 10.1109/TRO.2024.3370025, pp. 1–17, 2024.

[10] M. Li, J. Qin, Q. Ma, Y. Shi, and W. X. Zheng, "Master-Slave Safe Cooperative Tracking via Game and Learning Based Shared Control," *IEEE Transactions on Automatic Control*, DOI 10.1109/TAC.2024.3462254, pp. 1–8, 2024.

[11] Y. Yang, H. Jiang, C. Hua, and J. Li, "Practical Preassigned Fixed-Time Fuzzy Control for Teleoperation System Under Scheduled Shared-Control Framework," *IEEE Transactions on Fuzzy Systems*, vol. 32, DOI 10.1109/TFUZZ.2023.3300847, no. 2, pp. 470–482, Feb. 2024.

[12] S. Xue, J. Tan, Z. Guo, Q. Guan, K. Qu, and H. Cao, "Cooperative game-based optimal shared control of unmanned aerial vehicle," *Unmanned Systems*, DOI doi.org/10.1142/S2301385026500342, Apr. 2025.

[13] G. Ning, H. Liang, X. Zhang, and H. Liao, "Inverse-Reinforcement-Learning-Based Robotic Ultrasound Active Compliance Control in Uncertain Environments," *IEEE Transactions on Industrial Electronics*, vol. 71, DOI 10.1109/TIE.2023.3250767, no. 2, pp. 1686–1696, Feb. 2024.

[14] J. Lin, M. Wang, and H.-N. Wu, "Composite adaptive online inverse optimal control approach to human behavior learning," *Information Sciences*, vol. 638, DOI 10.1016/j.ins.2023.118977, p. 118977, Aug. 2023.

[15] W. Jin, D. Kulić, J. F.-S. Lin, S. Mou, and S. Hirche, "Inverse Optimal Control for Multiphase Cost Functions," *IEEE Transactions on Robotics*, vol. 35, DOI 10.1109/TRO.2019.2926388, no. 6, pp. 1387–1398, Dec. 2019.

[16] H. Wu, Q. Hu, J. Zheng, F. Dong, Z. Ouyang, and D. Li, "Discounted Inverse Reinforcement Learning for Linear Quadratic Control," *IEEE Transactions on Cybernetics*, vol. 55, DOI 10.1109/TCYB.2025.3540967, no. 4, pp. 1995–2007, Apr. 2025.

[17] K. Cao and L. Xie, "Game-Theoretic Inverse Reinforcement Learning: A Differential Pontryagin's Maximum Principle Approach," *IEEE Transactions on Neural Networks and Learning Systems*, DOI 10.1109/TNNLS.2022.3148376, pp. 1–8, 2022.

[18] K. Cao and L. Xie, "Trust-Region Inverse Reinforcement Learning," *IEEE Transactions on Automatic Control*, DOI 10.1109/TAC.2023.3274629, pp. 1–8, 2023.

[19] W. Jin, D. Kulić, S. Mou, and S. Hirche, "Inverse optimal control from incomplete trajectory observations," *The International Journal of Robotics Research*, vol. 40, DOI 10.1177/0278364921996384, no. 6-7, pp. 848–865, Jun. 2021.

[20] V. S. Donge, B. Lian, F. L. Lewis, and A. Davoudi, "Multiagent Graphical Games With Inverse Reinforcement Learning," *IEEE Transactions on Control of Network Systems*, vol. 10, DOI 10.1109/TCNS.2022.3210856, no. 2, pp. 841–852, Jun. 2023.

[21] Q. Wei, T. Li, J. Zhang, H. Li, X. Wang, and J. Xiao, "Isoperimetric Constraint Inference for Discrete-Time Nonlinear Systems Based on Inverse Optimal Control," *IEEE Transactions on Cybernetics*, vol. 54, DOI 10.1109/TCYB.2024.3367884, no. 9, pp. 5493–5505, Sep. 2024.

[22] A. Perrusquía and W. Guo, "Drone's Objective Inference Using Policy Error Inverse Reinforcement Learning," *IEEE Transactions on Neural Networks and Learning Systems*, DOI 10.1109/TNNLS.2023.3333551, pp. 1–12, 2024.

[23] B. Lian, Y. Kartal, F. L. Lewis, D. G. Mikulski, G. R. Hudas, Y. Wan, and A. Davoudi, "Anomaly Detection and Correction of Optimizing Autonomous Systems With Inverse Reinforcement Learning," *IEEE Transactions on Cybernetics*, vol. 53, DOI 10.1109/TCYB.2022.3213526, no. 7, pp. 4555–4566, Jul. 2023.

[24] J. Town, Z. Morrison, and R. Kamalapurkar, "Pilot Performance Modeling via Observer-Based Inverse Reinforcement Learning," *IEEE Transactions on Control Systems Technology*, DOI 10.1109/TCST.2024.3410128, pp. 1–8, 2024.

[25] D. Li, S. Ge, and T. Lee, "Fixed-Time-Synchronized Consensus Control of Multi-Agent Systems," *IEEE Transactions on Control of Network Systems*, vol. PP, DOI 10.1109/TCNS.2020.3034523, pp. 1–1, Oct. 2020.

[26] Y. Liu, H. Li, R. Lu, Z. Zuo, and X. Li, "An Overview of Finite/Fixed-Time Control and Its Application in Engineering Systems," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, DOI 10.1109/JAS.2022.105413, no. 12, pp. 2106–2120, Dec. 2022.

[27] F. Tatari and H. Modares, "Deterministic and Stochastic Fixed-Time Stability of Discrete-time Autonomous Systems," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, DOI 10.1109/JAS.2023.123405, no. 4, pp. 945–956, Apr. 2023.

[28] F. Tatari, N. Niknejad, and H. Modares, "Discrete-Time Nonlinear System Identification: A Fixed-Time Concurrent Learning Approach," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, DOI 10.1109/TSMC.2024.3508267, pp. 1–9, 2024.

[29] J. Tan, S. Xue, H. Li, Z. Guo, H. Cao, and D. Li, "Prescribed Performance Robust Approximate Optimal Tracking Control Via Stackelberg Game," *IEEE Transactions on Automation Science and Engineering*, DOI 10.1109/TASE.2025.3549114, pp. 1–1, 2025.

[30] M. He, C. Li, H. Huang, F. Zhou, Y. He, and W. Shang, "Adaptive State Feedback Shared Control for Unmanned Surface Vehicle With Fixed-Time Prescribed Performance Control," *IEEE Access*, vol. 12, DOI 10.1109/ACCESS.2024.3417484, pp. 93 781–93 790, 2024.

[31] J. Lin and H.-N. Wu, "Online Human Behavior Learning via Dynamic Regressor Extension and Mixing With Fixed-Time Convergence," *IEEE Transactions on Industrial Informatics*, vol. 21, DOI 10.1109/TII.2024.3485814, no. 2, pp. 1764–1772, Feb. 2025.

[32] F. Tatari, M. Mazouchi, and H. Modares, "Fixed-Time System Identification Using Concurrent Learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, DOI 10.1109/TNNLS.2021.3125145, no. 8, pp. 4892–4902, Aug. 2023.

[33] K. Wang and C. Mu, "Learning-Based Control With Decentralized Dynamic Event-Triggering for Vehicle Systems," *IEEE Transactions on Industrial Informatics*, vol. 19, DOI 10.1109/TII.2022.3168034, no. 3, pp. 2629–2639, Mar. 2023.

[34] Z. Zhang, K. Zhang, X. Xie, and V. Stojanovic, "ADP-Based Prescribed-Time Control for Nonlinear Time-Varying Delay Systems With Un-

certain Parameters," *IEEE Transactions on Automation Science and Engineering*, DOI 10.1109/TASE.2024.3389020, pp. 1–11, 2024.

[35] B. Lian, W. Xue, F. L. Lewis, and T. Chai, "Inverse reinforcement learning for multi-player noncooperative apprentice games," *Automatica*, vol. 145, DOI 10.1016/j.automatica.2022.110524, p. 110524, Nov. 2022.

[36] W. Xue, P. Kolaric, J. Fan, B. Lian, T. Chai, and F. L. Lewis, "Inverse Reinforcement Learning in Tracking Control Based on Inverse Optimal Control," *IEEE Transactions on Cybernetics*, vol. 52, DOI 10.1109/TCYB.2021.3062856, no. 10, pp. 10 570–10 581, Oct. 2022.

[37] J. Tan, S. Xue, T. Niu, K. Qu, H. Cao, and B. Chen, "Fixed-time concurrent learning-based robust approximate optimal control," *Nonlinear Dynamics*, Apr. 2025.

[38] H. Yue, J. Xia, J. Zhang, J. H. Park, and X. Xie, "Event-based adaptive fixed-time optimal control for saturated fault-tolerant nonlinear multiagent systems via reinforcement learning algorithm," *Neural Networks*, vol. 183, DOI 10.1016/j.neunet.2024.106952, p. 106952, Mar. 2025.

[39] H.-N. Wu, "Online Learning Human Behavior for a Class of Human-in-the-Loop Systems via Adaptive Inverse Optimal Control," *IEEE Transactions on Human-Machine Systems*, vol. 52, DOI 10.1109/THMS.2022.3155369, no. 5, pp. 1004–1014, Oct. 2022.

[40] Q. Ma, P. Jin, and F. L. Lewis, "Guaranteed Cost Attitude Tracking Control for Uncertain Quadrotor Unmanned Aerial Vehicle Under Safety Constraints," *IEEE/CAA Journal of Automatica Sinica*, vol. 11, DOI 10.1109/JAS.2024.124317, no. 6, pp. 1447–1457, Jun. 2024.

[41] H. Yang, H. Dong, and X. Zhao, "ADP-Based Spacecraft Attitude Control Under Actuator Misalignment and Pointing Constraints," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 9, 2022.

[42] L. N. Tan and T. C. Pham, "Optimal Tracking Control for PMSM With Partially Unknown Dynamics, Saturation Voltages, Torque, and Voltage Disturbances," *IEEE Transactions on Industrial Electronics*, vol. 69, DOI 10.1109/TIE.2021.3075892, no. 4, pp. 3481–3491, Apr. 2022.

[43] W. B. Powell, *Approximate dynamic programming: solving the curses of dimensionality*, 2nd ed., ser. Wiley series in probability and statistics. Hoboken, N.J: Wiley, 2011.

[44] R. Tian, L. Sun, M. Tomizuka, and D. Isele, "Anytime Game-Theoretic Planning with Active Reasoning About Humans' Latent States for Human-Centered Robots," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, DOI 10.1109/ICRA48506.2021.9561463, pp. 4509–4515, May. 2021.

[45] C. O. Yaldiz and Y. Yildiz, "Driver Modeling Using Continuous Reasoning Levels: A Game Theoretical Approach," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, DOI 10.1109/CDC51059.2022.9992839, pp. 5068–5073, Dec. 2022.

[46] Z. Jin, A. Liu, W.-A. Zhang, L. Yu, and C.-Y. Su, "A Learning Based Hierarchical Control Framework for Human–Robot Collaboration," *IEEE Transactions on Automation Science and Engineering*, vol. 20, DOI 10.1109/TASE.2022.3161993, no. 1, pp. 506–517, Jan. 2023.

[47] C. Mu, K. Wang, X. Xu, and C. Sun, "Safe Adaptive Dynamic Programming for Multiplayer Systems With Static and Moving No-entry Regions," *IEEE Transactions on Artificial Intelligence*, DOI 10.1109/TAI.2023.3325780, pp. 1–13, 2023.

[48] M. L. Greene, P. Deptula, S. Nivison, and W. E. Dixon, "Approximate Optimal Trajectory Tracking with Sparse Bellman Error Extrapolation," *IEEE Transactions on Automatic Control*, DOI 10.1109/TAC.2022.3194040, pp. 1–8, 2022.

[49] R. Self, M. Abudia, S. N. Mahmud, and R. Kamalapurkar, "Model-based inverse reinforcement learning for deterministic systems," *Automatica*, vol. 140, DOI 10.1016/j.automatica.2022.110242, p. 110242, Jun. 2022.

[50] K. Kim, P. Spieler, E.-S. Lupu, A. Ramezani, and S.-J. Chung, "A bipedal walking robot that can fly, slackline, and skateboard," *Science Robotics*, vol. 6, DOI 10.1126/scirobotics.abf8136, no. 59, p. eabf8136, Oct. 2021.

Junkai Tan received the B.E. degree in electrical engineering at the School of Electrical Engineering in Xi'an Jiaotong University, Xi'an, China. He is currently working toward the M.E. degree in electrical engineering at the School of Electrical Engineering, Xi'an Jiaotong University.

His current research interest includes adaptive dynamic programming and inverse reinforcement learning.

Shuangsi Xue (M'24) received the B.E. degree in electrical engineering and automation from Hunan University, Changsha, China, in 2014, and the M.E. and Ph.D. degrees in electrical engineering from Xian Jiaotong University, Xian, China, in 2018 and 2023, respectively. He is currently an Assistant Professor at the School of Electrical Engineering, Xian Jiaotong University.

His current research interest includes adaptive control and data-driven control of networked systems.

Qingshu Guan received the B.S. degree in automation from North China Electric Power University, Beijing, China, in 2020, and the M.S. degree from the School of Electric Engineering, Xi'an Jiaotong University, Xi'an, China, in 2023. He is currently pursuing the Ph.D. degree with the School of Electric Engineering, Xi'an Jiaotong University, Xi'an, China.

His current interests include deep reinforcement learning and strategy optimization.

Zihang Guo received the B.E. degree in electrical engineering at the School of Electrical Engineering in Xi'an Jiaotong University, Xi'an, China. He is currently working toward the M.E. degree in electrical engineering at the School of Electrical Engineering, Xi'an Jiaotong University.

His current research interest includes neural network and sliding mode-based path planning and tracking methods.

Hui Cao (M'11) received the B.E., M.E., and Ph.D. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2000, 2004, and 2009, respectively. He is a Professor at the School of Electrical Engineering, Xi'an Jiaotong University. He was a Postdoctoral Research Fellow at the Department of Electrical and Computer Engineering, National University of Singapore, Singapore, from 2014 to 2015. He has authored or coauthored over 30 scientific and technical papers in recent years.

His current research interest includes knowledge representation and discovery. Dr. Cao was a recipient of the Second Prize of the National Technical Invention Award.

Badong Chen received the Ph.D. degree in Computer Science and Technology from Tsinghua University, Beijing, China, in 2008. He is currently a professor with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China.

His research interests are in signal processing, machine learning, artificial intelligence and robotics. He has authored or coauthored over 200 articles in various journals and conference proceedings (with 17000+ citations in Google Scholar), and has won the 2022 Outstanding Paper Award of IEEE Transactions on Cognitive and Developmental Systems. Dr. Chen serves as a Member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society, and serves (or has served) as an Associate Editor for several international journals including IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Cognitive and Developmental Systems, IEEE Transactions on Circuits and Systems for Video Technology, Neural Networks and Journal of The Franklin Institute. He has served as a PC or SPC Member for prestigious conferences including UAI, IJCAI and AAAI, and served as a General Co-Chair of the 2022 IEEE International Workshop on Machine Learning for Signal Processing.