

ADHDP-based robust self-learning 3D trajectory tracking control for underactuated UUVs

Chunbo Zhao, Huaran Yan and Deyi Gao

Merchant Marine College, Shanghai Maritime University, Shanghai, China

ABSTRACT

In this work, we propose a robust self-learning control scheme based on action-dependent heuristic dynamic programming (ADHDP) to tackle the 3D trajectory tracking control problem of underactuated uncrewed underwater vehicles (UUVs) with uncertain dynamics and time-varying ocean disturbances. Initially, the radial basis function neural network is introduced to convert the compound uncertain element, comprising uncertain dynamics and time-varying ocean disturbances, into a linear parametric form with just one unknown parameter. Then, to improve the tracking performance of the UUVs trajectory tracking closed-loop control system, an actor-critic neural network structure based on ADHDP technology is introduced to adaptively adjust the weights of the action-critic network, optimizing the performance index function. Finally, an ADHDP-based robust self-learning control scheme is constructed, which makes the UUVs closed-loop system have good robustness and control performance. The theoretical analysis demonstrates that all signals in the UUVs trajectory tracking closed-loop control system are bounded. The simulation results for the UUVs validate the effectiveness of the proposed control scheme.

Subjects Adaptive and Self-Organizing Systems, Autonomous Systems, Robotics

Keywords Unmanned underactuated vehicles (UUVs), Robust adaptive control, Trajectory tracking, Action-dependent heuristic dynamic programming (ADHDP)

Submitted 31 May 2024
Accepted 21 November 2024
Published 10 December 2024

Corresponding author
Huaran Yan,
huaranyan202202@163.com

Academic editor
Valentina Emilia Balas

Additional Information and
Declarations can be found on
page 17

DOI 10.7717/peerj-cs.2605

© Copyright
2024 Zhao et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

INTRODUCTION

Recently, uncrewed underwater vehicles (UUVs) have garnered substantial interest from research teams due to potential engineering applications such as underwater archaeology, torpedo deployment, and cable installation (Peng, Wang & Han, 2019; Qiao & Zhang, 2020; Wang et al., 2021a; Wang, Gao & Zhang, 2021; Yuan et al., 2023). Most oceanic missions require the tracking of a predefined desired trajectory. In practice, UUVs are often underactuated, displaying significant nonlinearity, model parameter uncertainties, and vulnerability to unknown time-varying disturbances in the ocean environment (Li et al., 2023; Zhou et al., 2023). These features greatly complicate and challenge the development of trajectory tracking controllers.

Since the majority of UUVs are outfitted with three propellers to control surge, pitch, and yaw motions, they are classified as underactuated systems, where the number of independent actuators is less than the degrees of freedom. Yet, the existence of underactuated characteristic makes the design of UUV trajectory tracking control law more difficult. Further, the incomplete development of the hydraulic system complicates

the acquisition of UUV model parameters. The neural networks (NNs) have gained substantial attention in research because of its universal approximation property. In [Shi et al. \(2024\)](#), [Van \(2019\)](#), [Liu & Du \(2021\)](#), [Wang et al. \(2019\)](#), NNs were applied to approximate the unknown nonlinear dynamics. Nevertheless, achieving precise estimation with NNs necessitates weight identification, which adds considerable computational complexity. To overcome this issue, [Shen et al. \(2020\)](#) proposes combining minimum learning parameter technology with RBFNNs and a fixed-time sliding mode control scheme with RBFNN disturbance observer and the single-parameter learning was proposed in [Zhu et al. \(2023\)](#). A robust trajectory tracking control scheme is proposed in [Heshmati-Alamdari, Nikou & Dimarogonas \(2021\)](#) for underactuated underwater vehicle with obstacles in the constrained workspace. In [Chen et al. \(2017\)](#), a robust control method is developed though based on sliding mode control and fuzzy control for underactuated underwater vehicle with uncertain model parameters and unknown disturbances. Accurate trajectory tracking with good robustness is achieved based on a finite-time extended state observer in [Wang et al. \(2021b\)](#).

Currently, most proposed control schemes focus on achieving stable tracking without considering the optimal performance of UUVs under uncertain dynamics and time-varying disturbances. Nonetheless, the intricate marine environment inevitably impacts the control performance of systems with fixed design parameters. In recent times, there has been a wealth of research on the application of adaptive dynamic programming (ADP) in numerous systems to achieve optimal control ([Tong, Sun & Sui, 2018](#); [Li, Sun & Tong, 2019](#)). The fault-tolerant tracking control problem for UUVs subjected to time-varying ocean current disturbance is converted into an optimal control problem through the ADP approach in [Che & Yu \(2020\)](#). Through the integration of the backstepping design method, a fault-tolerant control scheme is formulated using a single-critic network-based ADP approach in [Che \(2022\)](#). In [Wen et al. \(2019\)](#), virtual and actual control are designed as the optimization solution of corresponding subsystems, and an optimized backstepping technique based on an actor network is utilized to enhance the control performance of the surface vessel system. A fault-tolerant fuzzy-optimized control scheme for tracking control of underactuated autonomous underwater vehicles is proposed in [Gong, Er & Liu \(2024\)](#) for complicated oceanic environments, particularly with the existence of unknown actuator failures and uncertain dynamics. Based on the above analyses, it is imperative to explore an intelligent control scheme with self-learning and self-adaptive capability to improve the control performance of UUVs under uncertain dynamics and time-varying ocean disturbances.

Motivated by the above discussions, this work develops a robust self-learning control scheme based on action-dependent heuristic dynamic programming (ADHDP) for underactuated UUVs subject to uncertain dynamics and unknown time-varying external disturbances. The main contributions of this work are as follows.

- The fact that a large number of control design parameters (e.g., adaptive gain, leakage term gain, etc.) must be artificially set to fixed values makes it difficult to ensure the optimization of control performance metrics for existing robust adaptive control

methods in [Li et al. \(2021\)](#) and [Yang et al. \(2021\)](#). The ADHDP-based control scheme proposed in this article can automatically update the control parameters in response to dynamic uncertainties and unknown time-varying perturbations, and our work not only reduces the need for manually designing the control design parameters, but also greatly improves the control performance and adaptability of the system.

- Unlike traditional optimal tracking control schemes, such as those in [Zhang, Li & Liu \(2018\)](#) and [Zheng et al. \(2020\)](#), which usually require partial model information (e.g., the control gain is known or partially known) for parameter tuning and strategy optimization, the ADHDP-based control scheme proposed in this article no longer relies on the model parameter information, but only the input and output data of the UUV.

The remaining segments of this work are organized as follows. The problem formulation and preliminaries are introduced in “Problem Formulation and Preliminaries”. The details of design process and stability proof of the robust self-learning are given in “Control Law Design”. “Simulation Results” presents the simulation results. The conclusions of this work are summed up in “Conclusion”.

Notations: In this work, $\text{diag}(\cdot)$ denotes the diagonal matrix; $\lambda(\cdot)$ denotes the minimum eigenvalue of the matrix; $\|\cdot\|$ denotes the 2-norm value of the matrix or vector; $s(\cdot)$ and $c(\cdot)$ denote the sine and cosine functions, respectively.

PROBLEM FORMULATION AND PRELIMINARIES

Motion mathematical model of UUVs

The mathematical model of underactuated UUVs ([Yan et al., 2024](#)) is given as:

$$\dot{\boldsymbol{\eta}} = \mathbf{J}(\boldsymbol{\eta})\mathbf{v} \quad (1)$$

$$\mathbf{M}\dot{\mathbf{v}} + \mathbf{C}(\mathbf{v})\mathbf{v} + \mathbf{D}\mathbf{v} + \mathbf{g}(\boldsymbol{\eta}) = \boldsymbol{\tau} + \mathbf{d} \quad (2)$$

where $\boldsymbol{\eta} = [x, y, z, \theta, \varphi]^T$ is position vector consist of the surge position x , the sway position y , the heave position z , the pitch angle θ and the yaw angle φ of UUVs in the earth-fixed frame. $\mathbf{v} = [u, v, w, q, r]^T$ is the velocity vector consist of the surge velocity u , the sway velocity v , the heave velocity w , the pitch velocity q and the yaw velocity r of UUVs in the body-fixed frame. In this article, we define the earth-fixed and the body-fixed reference frames of the UUV as indicated in [Fig. 1](#). $\mathbf{D} = \text{diag}(d_{11}, d_{22}, d_{33}, d_{44}, d_{55})$ represents the hydrodynamic damping matrix. $\mathbf{M} = \text{diag}(m_{11}, m_{22}, m_{33}, m_{44}, m_{55})$ is the positive-definite inertia matrix. $\boldsymbol{\tau} = [\tau_u, 0, 0, \tau_q, \tau_r]^T$ is the control input vector consist of surge force τ_u , pitch moment τ_q and yaw moment τ_r . $\mathbf{d} = [d_1, d_2, d_3, d_4, d_5]^T$ represents the environmental disturbance vector. $\mathbf{g}(\boldsymbol{\eta}) = [0, 0, 0, \rho g \nabla GM_L s(\theta), 0]^T$ with ρ, g, ∇ and GM_L are water density, gravity acceleration, displaced volume and longitudinal metacentric

height, respectively. $\mathbf{J}(\boldsymbol{\eta}) = \begin{bmatrix} (\theta)c(\varphi) & -s(\varphi) & s(\theta)c(\varphi) & 0 & 0 \\ c(\theta)s(\varphi) & c(\varphi) & s(\theta)s(\varphi) & 0 & 0 \\ -s(\theta) & 0 & c(\varphi) & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{c(\varphi)} \end{bmatrix}$ represents the

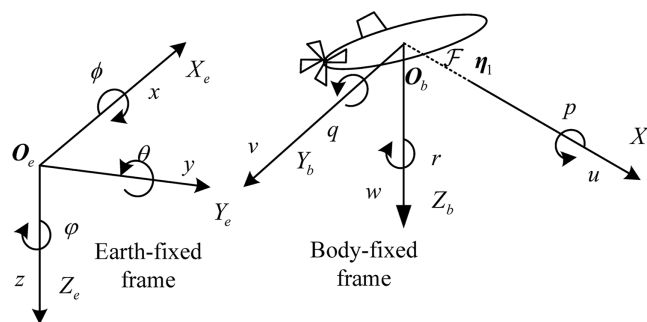


Figure 1 The earth-fixed and the body-fixed reference frames of the UAV.

Full-size DOI: 10.7717/peerj-cs.2605/fig-1

rotation matrix. $C(\mathbf{v}) = \begin{bmatrix} 0 & 0 & 0 & m_{33}w & -m_{22}v \\ 0 & 0 & 0 & 0 & m_{11}u \\ 0 & 0 & 0 & -m_{11}u & 0 \\ -m_{33}w & 0 & m_{11}u & 0 & 0 \\ m_{22}v & -m_{11}u & 0 & 0 & 0 \end{bmatrix}$ represents the

Coriolis and centripetal force matrix.

Assumption 1 The model parameters matrices $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{D}$ are unavailable.

Assumption 2 The disturbance vector $\mathbf{d}$ is bound yet unknown time-varying and $\|\mathbf{d}\| \leq d_m$, d_m is a positive constant.

Assumption 3 The desired trajectory of UAVs in this article is notated as $\boldsymbol{\eta}_d = [x_d, y_d, z_d]^T$. The component of $\boldsymbol{\eta}_d$ and their first two time derivatives are bound.

Assumption 4 The sway velocity v and the heave velocity w are bounded.

Transformation of UAV's position

To initiate the discussion, the subsequent coordinate transformation is employed to tackle the underactuation issue of UAVs:

$$\boldsymbol{\eta}_1 = [x + \mathcal{F}c(\theta)c(\varphi), y + \mathcal{F}c(\theta)s(\varphi), z - \mathcal{F}s(\varphi)]^T \quad (3)$$

where $\mathcal{F}$ is a small positive constant.

According to Eqs. (1)–(3), we can obtain the following equation as:

$$\dot{\boldsymbol{\eta}}_1 = \mathbf{J}_1(\boldsymbol{\eta})\mathbf{v}_1 + \mathbf{R}_1(\boldsymbol{\eta}, \mathbf{v}) \quad (4)$$

$$\mathbf{M}_1\dot{\mathbf{v}}_1 + \mathbf{C}_1(\mathbf{v})\mathbf{v}_1 + \mathbf{D}_1\mathbf{v}_1 + \mathbf{g}_1(\boldsymbol{\eta}) = \boldsymbol{\tau}_1 + \mathbf{d}_1 \quad (5)$$

$$\text{where } \mathbf{J}_1(\boldsymbol{\eta}) = \begin{bmatrix} (\theta)c(\varphi) & -\mathcal{F}s(\theta)c(\varphi) & -\mathcal{F}s(\varphi) \\ c(\theta)s(\varphi) & -\mathcal{F}s(\theta)s(\varphi) & \mathcal{F}c(\varphi) \\ -s(\theta) & -\mathcal{F}c(\theta) & 0 \end{bmatrix}, \mathbf{M}_1 = \text{diag}(m_{11}, m_{44}, m_{55}),$$

$$\mathbf{v}_1 = [u, q, r]^T, \mathbf{C}_1(\mathbf{v}) = \begin{bmatrix} 0 & m_{33}w & -m_{22}v \\ (m_{11} - m_{33})w & 0 & 0 \\ (m_{22} - m_{11})v & 0 & 0 \end{bmatrix},$$

$$\mathbf{R}_1(\boldsymbol{\eta}, \mathbf{v}) = \begin{bmatrix} -vs(\varphi) + ws(\theta)c(\varphi) \\ vc(\varphi) + ws(\theta)s(\varphi) \\ wc(\theta) \end{bmatrix}, \boldsymbol{\tau}_1 = [\tau_u, \tau_q, \tau_r]^T, \mathbf{d}_1 = [d_1, d_4, d_5]^T,$$

$$\mathbf{D}_1 = \text{diag}(d_{11}, d_{44}, d_{55}) \text{ and } \mathbf{g}_1(\boldsymbol{\eta}) = [0, \rho g \nabla GM_L s(\theta), 0]^T.$$

The goal of this work is to propose a robust self-learning control law based on ADHDP to ensure that the UUV trajectory tracking error converges to a small compact set in the presence of uncertain dynamics, and unknown time-varying disturbances, while all signals in the UUV trajectory tracking closed-loop control system are bounded.

Lemma 1: (*Luo et al., 2020; Na et al., 2020*) Notate $f(\mathbf{s})$ is a continuous function which is defined on a compact set $\mathbf{s} \subset R^n$. There has a radial basis function neural network (RBFNN) for an arbitrarily small constant Θ satisfy

$$f(\mathbf{s}) = \omega^{*T} \mathbf{G}(\mathbf{s}) + \Theta \quad (6)$$

where ω^* denotes optimal weight vector. Θ denotes the reconstruction error of NNs, which satisfies $\|\Theta\| \leq \bar{\Theta}$. $\mathbf{G}(\mathbf{s}) = [\mathbf{G}_1(\mathbf{s}), \mathbf{G}_2(\mathbf{s}), \dots, \mathbf{G}_n(\mathbf{s})]^T$ denotes the basic function vector. Herein, $\mathbf{G}_n(\mathbf{s})$ is chosen as the Gaussian function.

CONTROL LAW DESIGN

Robust adaptive NN control law design

Notate the tracking error $\mathbf{s}_1 = [s_{11}, s_{12}, s_{13}]^T$ as:

$$\mathbf{s}_1 = \boldsymbol{\eta}_1 - \boldsymbol{\eta}_d. \quad (7)$$

According to Eqs. (5) and (7), we have

$$\dot{\mathbf{s}}_1 = \mathbf{J}_1(\boldsymbol{\eta}) \mathbf{v}_1 + \mathbf{R}_1(\boldsymbol{\eta}, \mathbf{v}) + \dot{\boldsymbol{\eta}}_d. \quad (8)$$

Design the virtual control law $\boldsymbol{\varsigma} \in R^3$ for $\mathbf{v}_1$ as follows

$$\boldsymbol{\varsigma} = \mathbf{J}_1^{-1}(\boldsymbol{\eta})(-\mathbf{k}_1 \mathbf{s}_1 - \mathbf{R}_1(\boldsymbol{\eta}, \mathbf{v}) + \dot{\boldsymbol{\eta}}_d) \quad (9)$$

where $\mathbf{k}_1 \in R^{3 \times 3}$ is positive-definite design matrix.

To proceed, we can define the velocity error vector $\mathbf{s}_2 = [s_{21}, s_{22}, s_{23}]^T$ as follows:

$$\mathbf{s}_2 = \mathbf{v}_1 - \boldsymbol{\varsigma} \quad (10)$$

In the light of Eqs. (6) and (10), we can obtain the time derivative of $\mathbf{s}_2$

$$\mathbf{M}_1 \dot{\mathbf{s}}_2 = -\mathbf{C}_1(\mathbf{v}) \mathbf{v}_1 - \mathbf{D}_1 \mathbf{v}_1 - \mathbf{g}_1(\boldsymbol{\eta}) + \boldsymbol{\tau}_1 + \dot{\mathbf{d}}_1 - \mathbf{M}_1 \dot{\boldsymbol{\varsigma}} \quad (11)$$

Using the RBF NN to approximate the uncertain term $-\mathbf{C}_1(\mathbf{v}) \mathbf{v}_1 - \mathbf{D}_1 \mathbf{v}_1 - \mathbf{g}_1(\boldsymbol{\eta}) - \mathbf{M}_1 \dot{\boldsymbol{\varsigma}}$. We have

$$-\mathbf{C}_1(\mathbf{v}) \mathbf{v}_1 - \mathbf{D}_1 \mathbf{v}_1 - \mathbf{g}_1(\boldsymbol{\eta}) - \mathbf{M}_1 \dot{\boldsymbol{\varsigma}} = \omega^{*T} \mathbf{G}(\boldsymbol{\mathfrak{S}}) + \Theta \quad (12)$$

where ω^* is the ideal weight matrix, $\boldsymbol{\mathfrak{S}} = [\mathbf{v}^T, \boldsymbol{\varsigma}^T, \theta]^T$ denotes the input vector, $\mathbf{G}(\boldsymbol{\mathfrak{S}})$ represents the basis function vector. Θ is the NN reconstruction error vector with $\|\Theta\| \leq \bar{\Theta}$.

With the aid of virtual parameter learning technology, the following inequality can be obtained

$$\begin{aligned} \|\omega^{*T}G(\mathfrak{S}) + \Theta + d_1\| &\leq \omega_m \|G(\mathfrak{S})\| + \bar{\Theta} + d_m \\ &\leq H\phi \end{aligned} \quad (13)$$

where $H = \max\{\omega_m, \bar{\Theta} + d_m\}$ is an virtual parameter without physical meaning. $\phi = \|G\mathfrak{S}\| + 1$ is a scalar function.

Design the control law for control input as:

$$\tau_1 = -k_2 s_2 - J_1(\eta)s_1 - \hat{H}\phi^2 s_2 \quad (14)$$

where $s_2 \in R^{3 \times 3}$ denotes positive-definite design matrix, $\hat{H}$ denotes the estimate of H with estimation error $\tilde{H} = H - \hat{H}$.

The adaptive law is given as

$$\dot{\hat{H}} = \Gamma(\phi^2 \|s_2\|^2 - \vartheta \hat{H}) \quad (15)$$

where Γ and ϑ are positive design constants.

Theorem 1 Consider the UUV trajectory tracking closed-loop control system Eqs. (1), (2) under Assumptions 1–4, the virtual control law Eq. (9), the control law Eq. (14), with the adaptive law Eq. (15). The actual trajectory η_1 can track the desired trajectory η_d , while all signals in the UUV trajectory tracking closed-loop control system are bounded.

Proof: Select the following Lyapunov function candidate

$$V = \frac{1}{2} \left(s_1^T s_1 + s_2^T M_1 s_2 + \Gamma^{-1} \tilde{H}^2 \right) \quad (16)$$

The time derivative of Eq. (16) is

$$\dot{V} = s_1^T \dot{s}_1 + s_2^T M_1 \dot{s}_2 + \Gamma^{-1} \tilde{H} \left(-\dot{\tilde{H}} \right) \quad (17)$$

In view of Eqs. (8)–(10), we have

$$s_1^T \dot{s}_1 = s_1^T J_1(\eta)s_2 - s_1^T k_1 s_1 \quad (18)$$

Substituting Eqs. (13) and (14) into Eq. (11) and rearranging Eq. (11) yields

$$\begin{aligned} s_2^T M_1 \dot{s}_2 &\leq \|s_2\| H\phi + s_2^T (-k_2 s_2 - J_1(\eta)s_1 - \hat{H}\phi^2 s_2) \\ &\leq \frac{H}{4} + s_2^T (-k_2 s_2 - J_1(\eta)s_1 - \tilde{H}\phi^2 s_2) \end{aligned} \quad (19)$$

In view of Eq. (15), one has:

$$\begin{aligned} \Gamma^{-1} \tilde{H} \left(-\dot{\tilde{H}} \right) &= -\tilde{H}(\phi^2 \|s_2\|^2 - \vartheta \hat{H}) \\ &\leq -\tilde{H}\phi^2 \|s_2\|^2 - \frac{1}{2} \vartheta \tilde{H}^2 + \frac{1}{2} \vartheta H^2. \end{aligned} \quad (20)$$

Substituting Eqs. (18)–(20) into (17) and rearranging Eq. (17) yields:

$$\begin{aligned}\dot{V} &= \mathbf{s}_1^T \dot{\mathbf{s}}_1 + \mathbf{s}_2^T \mathbf{M}_1 \dot{\mathbf{s}}_2 + \Gamma^{-1} \tilde{H} \left(-\dot{\tilde{H}} \right) \\ &\leq -\mathbf{s}_1^T \mathbf{k}_1 \mathbf{s}_1 - \mathbf{s}_2^T \mathbf{k}_2 \mathbf{s}_2 - \frac{1}{2} \vartheta \tilde{H}^2 + \frac{1}{2} \vartheta H^2 + \frac{H}{4} \\ &\leq -2\sigma V + \psi\end{aligned}\quad (21)$$

where $\sigma = \min \left\{ \lambda_{\min}(\mathbf{k}_1), \lambda_{\min}(\mathbf{k}_2 \bar{\mathbf{M}}_1^{-1}), \frac{1}{2} \vartheta \Gamma \right\}$ and $\psi = \frac{1}{2} \vartheta H^2 + \frac{H}{4}$.
Solving Eq. (21) yields

$$0 \leq V \leq \frac{\psi}{2\sigma} + \left[V(0) - \frac{\psi}{2\sigma} \right] \exp^{-(2\sigma t)}.\quad (22)$$

It shows that V is uniformly ultimately bounded. From Eq. (17), $\mathbf{s}_1$, $\mathbf{s}_2$ and $\tilde{H}$ are uniformly ultimately bounded. Thus, Theorem 1 is proved.

Optimal control law design

To optimize the tracking performance online, an optimal control law based on ADHDP is developed in this section. Schematic diagram of the proposed closed-loop control system of the UUV is presented in Fig. 2. An action-critic network structure is developed, the actor network is constructed to learn the optimal control scheme $\boldsymbol{\tau}_a = [\tau_{a1}, \tau_{a2}, \tau_{a3}]^T$, and the critic network is constructed to approximate the cost function. Notate $\bar{\mathbf{s}} = [\mathbf{s}(t - \Delta t), \mathbf{s}(t)]^T$ with $\mathbf{s} = [s_{11}, s_{12}, s_{13}, s_{21}, s_{22}, s_{23}]^T$. Δt denotes the uniform time interval.

Herein, the cost function is formulated as follows:

$$P(\bar{\mathbf{s}}(t), \boldsymbol{\tau}(t)) = \sum_{c=1}^{\infty} \beta^{c-1} Y(\bar{\mathbf{s}}(t + b\Delta t), \boldsymbol{\tau}_a(t + b\Delta t))\quad (23)$$

where $0 < \beta \leq 1$ is a discount factor. $Y(\bar{\mathbf{s}}(t), \boldsymbol{\tau}_a(t)) = \bar{\mathbf{s}}^T \mathbf{F} \bar{\mathbf{s}} + \boldsymbol{\tau}_a^T \mathbf{L} \boldsymbol{\tau}_a$ represents the utility function. Herein, $\mathbf{F} \in R^{12 \times 12}$ and $\mathbf{L} \in R^{3 \times 3}$ are positive-definite design matrices.

The Bellman equation is as follows (Liang, Xu & Zhang, 2023):

$$P^*(t - \Delta t) = \min_{\boldsymbol{\tau}_a(t - \Delta t)} [Y(t) + \beta P^*(t)]\quad (24)$$

The goal of this section is to make the system output $\boldsymbol{\eta}_1$ can track the desired trajectory $\boldsymbol{\eta}_d$ in an optimal manner, while minimizing the cost function.

1) Critic network:

A multilayer perceptron with a three-layer network containing input, hidden, and output layers is introduced to approximate the cost function $P(t)$. The input vector is $\mathbf{I}_c = [\mathbf{s}(t - \Delta t), \mathbf{s}(t), \boldsymbol{\tau}_a]$. I_{ch} is the number of hidden nodes. $\omega_{cij}^{(1)}(t)$ and $\omega_{cj}^{(2)}(t)$ denote the weights of critic network, where $i = 1, \dots, 15$ and $j = 1, \dots, I_{ch}$. The hyperbolic tangent function $\Xi(t) = [(1 - e^{-t}) / (1 + e^{-t})]$ is introduced in this work. Here, $p_{cj}(t)$ and $q_{cj}(t)$ are the intermediate variables. Further, one can derive the approximation $\hat{P}(t)$ as

$$p_{cj}(t) = \sum_{i=1}^{12} \omega_{cij}^{(1)}(t) \bar{s}_{1i}(t) + \sum_{i=1}^3 \omega_{c(i+12)j}^{(1)}(t) \tau_{ai}(t)\quad (25)$$

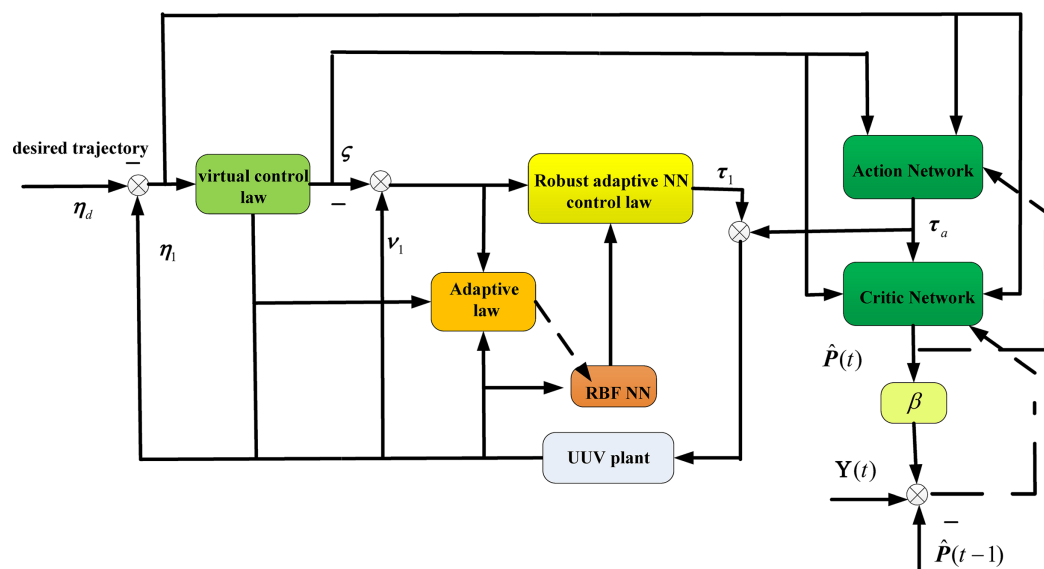


Figure 2 Schematic diagram of the proposed closed-loop control system of the UUV.

Full-size DOI: 10.7717/peerj-cs.2605/fig-2

$$q_{cj}(t) = \Xi_{cj}(p_{cj}(t)) = \frac{1 - e^{-p_{cj}(t)}}{1 + e^{-p_{cj}(t)}} \quad (26)$$

$$\hat{P}(t) = \omega_c^{(2)}(t) \Xi_c(t) = \sum_{j=1}^{I_{ch}} \omega_{cj}^{(2)}(t) q_{cj}(t). \quad (27)$$

2) Action network:

A multilayer perceptron with a three-layer network containing input, hidden, and output layers is introduced to approximate the optimal control law τ_a . The input vector of action network is $\bar{s}$, and the output vector of action network is τ_a . I_{ah} is the number of hidden nodes. $\omega_{akn}^{(1)}(t)$ and $\omega_{ank}^{(2)}(t)$ denote the weights of action network, where $\kappa = 1, \dots, 12$, $n = 1, \dots, I_{ah}$ and $k = 1, \dots, 3$. Here, $p_{an}(t)$ and $q_{an}(t)$ are the intermediate variables. Further, we have

$$p_{an}(t) = \sum_{\kappa=1}^{12} \omega_{akn}^{(1)}(t) \bar{s}_{1\kappa}(t) \quad (28)$$

$$q_{an}(t) = \Xi_{an}(p_{an}(t)) = \frac{1 - e^{-p_{an}(t)}}{1 + e^{-p_{an}(t)}} \quad (29)$$

$$\tau_{ak}(t) = \sum_{n=1}^{I_{ah}} \omega_{ank}^{(2)}(t) q_{an}(t). \quad (30)$$

3) Adaptation of the critic network:

The critic network's error function can be described as:

$$O_c(t) = \beta \hat{P}(t) - [\hat{P}(t - \Delta t) - Y(t)] \quad (31)$$

Hence, the critic network's objective function for weights update can be defined as:

$$E_c(t) = \frac{1}{2} O_c^2(t). \quad (32)$$

A gradient descent algorithm is applied to update the weight law. Here, $\omega_{cij}^{(1)}(t + \Delta t) = \omega_{cij}^{(1)}(t) + \Delta\omega_{cij}^{(1)}(t)$, $\omega_{cj}^{(2)}(t + \Delta t) = \omega_{cj}^{(2)}(t) + \Delta\omega_{cj}^{(2)}(t)$. According to Eqs. (28)–(30), the weight updating laws $\Delta\omega_{cij}^{(1)}$ and $\Delta\omega_{cj}^{(2)}$ of critic network can be calculated using the chain derivation rule

$$\begin{aligned} \Delta\omega_{cij}^{(1)} &= -\mu_c \frac{\partial E_c(t)}{\partial \hat{P}(t)} \frac{\partial \hat{P}(t)}{\partial q_{cj}(t)} \frac{\partial q_{cj}(t)}{\partial p_{cj}(t)} \frac{\partial p_{cj}(t)}{\partial \omega_{cij}^{(1)}(t)} \\ &= -\mu_c \beta O_c(t) \omega_{cj}^{(2)}(t) \left[\frac{1}{2} (1 - \Xi_{cj}^2(t)) \right] I_{ci}(t) \end{aligned} \quad (33)$$

$$\Delta\omega_{cj}^{(2)} = -\mu_c \frac{\partial E_c(t)}{\partial \hat{P}(t)} \frac{\partial \hat{P}(t)}{\partial \omega_{cj}^{(2)}(t)} = -\mu_c \beta O_c(t) \Xi_{cj}(t) \quad (34)$$

where μ_c is the learning rate.

4) Adaptation of the action network:

The action network can be adjusted by backpropagating the error between the desired value and the approximate value of the critic network.

Herein, we have

$$O_a(t) = \hat{P}(t) - P(t) \quad (35)$$

$$E_a(t) = \frac{1}{2} O_a^2(t) \quad (36)$$

Similarly, a gradient descent algorithm is applied to update the weight law of the action network. Here, $\omega_{akn}^{(1)}(t + \Delta t) = \omega_{akn}^{(1)}(t) + \Delta\omega_{akn}^{(1)}(t)$, $\omega_{ank}^{(2)}(t + \Delta t) = \omega_{ank}^{(2)}(t) + \Delta\omega_{ank}^{(2)}(t)$. According to Eqs. (31)–(33), the weight updating laws $\omega_{akn}^{(1)}(t)$ and $\Delta\omega_{ank}^{(2)}$ of action network can be calculated using the chain derivation rule:

$$\begin{aligned} \Delta\omega_{akn}^{(1)} &= -\mu_a \frac{\partial E_a(t)}{\partial \hat{P}(t)} \left[\frac{\partial \hat{P}(t)}{\partial \tau_{ak}(t)} \right]^T \frac{\partial \tau_{ak}(t)}{\partial q_{an}(t)} \frac{\partial q_{an}(t)}{\partial p_{an}(t)} \frac{\partial p_{an}(t)}{\partial \omega_{akn}^{(1)}(t)} \\ &= -\mu_a O_a(t) \left(\sum_{j=1}^{I_{ch}} \omega_{cj}^{(2)}(t) \left[\frac{1}{2} (1 - \Xi_{cj}^2(t)) \right] \right) \end{aligned} \quad (37)$$

$$\begin{aligned} &\times \left(\sum_{i=1}^3 \omega_{c(i+12)j}^{(1)}(t) \right) \omega_{ank}^{(2)}(t) \left[\frac{1}{2} (1 - \Xi_{an}^2(t)) \right] \left(\sum_{\kappa=1}^{12} \bar{s}_{1\kappa}(t) \right) \\ \Delta\omega_{ank}^{(2)} &= -\mu_a \frac{\partial E_a(t)}{\partial \hat{P}(t)} \frac{\partial \hat{P}(t)}{\partial \tau_{ak}(t)} \frac{\partial \tau_{ak}(t)}{\partial \omega_{ank}^{(2)}(t)} \\ &= -\mu_a O_a(t) \Xi_{an}(t) \left(\sum_{j=1}^{I_{ch}} \omega_{cj}^{(2)}(t) \left[\frac{1}{2} (1 - \Xi_{cj}^2(t)) \right] \right) \times \sum_{i=1}^3 \omega_{c(i+12)j}^{(1)}(t) \end{aligned} \quad (38)$$

where μ_a is the learning rate.

Remark 1 As a matter of convenience, only the output layer weights of the critic-action network are tuned during the learning process, whereas the weights in Eqs. (25) and (28) are randomly initiated. As described by Liang, Xu & Zhang (2023), as the number of NN's hidden nodes increases, the NN's approximation error is able to converge to a value that is adequately small.

Stability analysis

The optimal control law based on ADHDP is developed to optimize the tracking performance online. It can be obtained that actual control input signal is compound of the robust adaptive NN control law and optimal control law. The stability of robust adaptive control has been presented in Theorem 1. The ADHDP technique is looking for an optimal control action while approximating the Bellman equation with the critic network.

The ideal weights ω_c^* and ω_a^* for critic and action networks are bounded.

$$\|\omega\|_c^* \leq \omega_{cm}, \|\omega_a^*\| \leq \omega_{am} \quad (39)$$

where ω_{cm} and ω_{am} are positive constants.

In the light of Eqs. (27) and (30), we have:

$$\hat{P}(t) = \omega_c^{(2)}(t)^T \Xi_c(t) \quad (40)$$

$$\tau_a = \omega_a^{(2)}(t)^T \Xi_a(t) \quad (41)$$

where $\tilde{\omega}_c^{(2)} = \omega_c^{(2)} - \omega_c^*$ and $\tilde{\omega}_a^{(2)} = \omega_a^{(2)} - \omega_a^*$ are the estimation error of $\omega_c^{(2)}$ and $\omega_a^{(2)}$, respectively.

Theorem 2 Consider the updating law Eq. (38) for actor network weights, and the updating law Eq. (34) for critic network weights, and the utility function $Y(t)$ is assumed as a bounded positive semidefinite function. Then, the estimation error $\tilde{\omega}_c^{(2)}$ and $\tilde{\omega}_a^{(2)}$ are bounded.

Proof: Select the following Lyapunov function candidate as:

$$L_V(t) = L_{V1}(t) + L_{V2}(t) \quad (42)$$

where $L_{V1}(t) = \frac{1}{\mu_c} \text{tr}(\tilde{\omega}_c^{(2)}(t)^T \tilde{\omega}_c^{(2)}(t))$, $L_{V2}(t) = \frac{1}{\gamma \mu_a} \text{tr}(\tilde{\omega}_a^{(2)}(t)^T \tilde{\omega}_a^{(2)}(t))$ with γ being a design constant.

The first difference of $L_V(t)$ can be written as:

$$\Delta L_V(t) = \Delta L_{V1}(t) + \Delta L_{V2}(t). \quad (43)$$

Here, $\Delta L_{V1}(t)$ and $\Delta L_{V2}(t)$ can be calculated as

$$\Delta L_{V1}(t) = \frac{1}{\mu_c} \text{tr}(\tilde{\omega}_c^{(2)}(t + \Delta t)^T \tilde{\omega}_c^{(2)}(t + \Delta t) - \tilde{\omega}_c^{(2)}(t)^T \tilde{\omega}_c^{(2)}(t)) \quad (44)$$

$$\Delta L_{V2}(t) = \frac{1}{\gamma \mu_a} \text{tr}(\tilde{\omega}_a^{(2)}(t + \Delta t)^T \tilde{\omega}_a^{(2)}(t + \Delta t) - \tilde{\omega}_a^{(2)}(t)^T \tilde{\omega}_a^{(2)}(t)). \quad (45)$$

From Eq. (34), $\tilde{\omega}_c^{(2)}(t + \Delta t)$ can be get:

$$\begin{aligned}\tilde{\omega}_c^{(2)}(t + \Delta t) &= \tilde{\omega}_c^{(2)}(t) + \Delta \tilde{\omega}_c^{(2)}(t) \\ &= -\mu_c \beta \Xi_c(t) \left[\beta \omega_c^{(2)}(t)^T \Xi_c(t) + Y(t) - \omega_c^{(2)}(t - \Delta t)^T \Xi_c(t - \Delta t) \right]^T \\ &\quad + \tilde{\omega}_c^{(2)}(t).\end{aligned}\quad (46)$$

According to Eqs. (44) and (46), we have:

$$\begin{aligned}\Delta L_{V1} &= -2\beta \delta_c(t) \left[Y(t) - \omega_c^{(2)}(t - \Delta t)^T \Xi_c(t - \Delta t) + \beta \omega_c^{(2)}(t)^T \Xi_c(t) \right]^T \\ &\quad + \mu_c \beta^2 \left\| \Xi_c(t) \right\|^2 \left\| \beta \omega_c^{(2)}(t)^T \Xi_c(t) + Y(t) - \omega_c^{(2)}(t - \Delta t)^T \Xi_c(t - \Delta t) \right\|^2\end{aligned}\quad (47)$$

where $\delta_c(t) = \tilde{\omega}_c^{(2)}(t)^T \Xi_c(t)$.

Defined $-2\beta \delta_c(t) [Y(t) - \omega_c^{(2)}(t - \Delta t)^T \Xi_c(t - \Delta t) + \beta \omega_c^{(2)}(t)^T \Xi_c(t)]^T = A$. The first term in Eq. (47) can be defined as:

$$\begin{aligned}A &= \left\| \beta \omega_c^{(2)}(t)^T \Xi_c(t) + Y(t) - \omega_c^{(2)}(t - \Delta t)^T \Xi_c(t - \Delta t) - \beta \delta_c(t) \right\|^2 \\ &\quad - \left\| \beta \omega_c^{(2)}(t)^T \Xi_c(t) + Y(t) - \omega_c^{(2)}(t - \Delta t)^T \Xi_c(t - \Delta t) \right\|^2 - \left\| \beta \delta_c(t) \right\|^2.\end{aligned}\quad (48)$$

Substituting Eqs. (48) into (47), we have:

$$\begin{aligned}\Delta L_{V1}(t) &= -\beta^2 (1 - \mu_c \beta^2 \left\| \Xi_c(t) \right\|) \times \left\| \beta^{-1} Y(t) - \beta^{-1} \omega_c^{(2)}(t - \Delta t)^T \Xi_c(t - \Delta t) \right. \\ &\quad \left. + \omega_c^{(2)}(t)^T \Xi_c(t) \right\|^2 + \left\| \beta \omega_c^{(2)}(t)^T \Xi_c(t) - \omega_c^{(2)}(t - \Delta t)^T \Xi_c(t - \Delta t) + Y(t) \right\|^2 \\ &\quad - \beta^2 \left\| \delta_c(t) \right\|^2.\end{aligned}\quad (49)$$

From Eq. (38), we have:

$$\begin{aligned}\tilde{\omega}_a^{(2)}(t + \Delta t) &= \tilde{\omega}_a^{(2)}(t) + \Delta \tilde{\omega}_a^{(2)}(t) \\ &= \tilde{\omega}_a^{(2)}(t) - \mu_a \Xi_a(t) \omega_c^{(2)}(t)^T N(t) \left[\omega_c^{(2)}(t)^T \Xi_c(t) \right]^T\end{aligned}\quad (50)$$

According to Eqs. (45) and (50), we have:

$$\begin{aligned}\Delta L_{V2}(t) &= \frac{1}{\gamma_1} \text{tr} \left(-2\delta_a(t) \omega_c^{(2)}(t)^T N(t) \left[\omega_c^{(2)}(t)^T \Xi_c(t) \right]^T \right. \\ &\quad \left. + \mu_a \left\| \Xi_a(t) \right\|^2 \left\| \omega_c^{(2)}(t)^T N(t) \right\|^2 \left\| \omega_c^{(2)}(t)^T \Xi_c(t) \right\|^2 \right)\end{aligned}\quad (51)$$

where $\delta_a(t) = \tilde{\omega}_a^{(2)}(t)^T \Xi_a(t)$.

The last term in Eq. (51) can be further calculated as:

$$\begin{aligned}&-2\delta_a(t) \omega_c^{(2)}(t)^T N(t) \left[\omega_c^{(2)}(t)^T \Xi_c(t) \right]^T \\ &= \left\| \omega_c^{(2)}(t)^T \Xi_c(t) - \omega_c^{(2)}(t)^T N(t) \delta_a(t) \right\|^2 - \left\| \omega_c^{(2)}(t)^T \Xi_c(t) \right\|^2 \\ &- \left\| \omega_c^{(2)}(t)^T N(t) \delta_a(t) \right\|^2.\end{aligned}\quad (52)$$

Substituting Eqs. (52) into (51) yields

$$\begin{aligned}\Delta L_{V2}(t) = & \frac{1}{\gamma_1} \text{tr}(\| \boldsymbol{\omega}_c^{(2)}(t)^T \Xi_c(t) - \boldsymbol{\omega}_c^{(2)}(t)^T N(t) \delta_a(t) \|^2 \\ & + \mu_a \| \Xi_a(t) \|^2 \| \boldsymbol{\omega}_c^{(2)}(t)^T N(t) \|^2 \| \boldsymbol{\omega}_c^{(2)}(t)^T \Xi_c(t) \|^2 \\ & - \| \boldsymbol{\omega}_c^{(2)}(t)^T \Xi_c(t) \|^2 - \| \boldsymbol{\omega}_c^{(2)}(t)^T N(t) \delta_a(t) \|^2)\end{aligned}\quad (53)$$

By using Cauchy-Schwarz inequality, we have:

$$\begin{aligned}& \| \boldsymbol{\omega}_c^{(2)}(t)^T \Xi_c(t) - \boldsymbol{\omega}_c^{(2)}(t)^T N(t) \delta_a(t) \|^2 - \| \boldsymbol{\omega}_c^{(2)}(t)^T \Xi_c(t) \|^2 \\ & \leq 4 \| \delta_c(t) \|^2 + 4 \| \boldsymbol{\omega}_c^{*T} \Xi_c(t) \|^2 + \| \boldsymbol{\omega}_c^{(2)}(t)^T N(t) \delta_a(t) \|^2.\end{aligned}\quad (54)$$

Substituting Eqs. (54) into (53), we have:

$$\begin{aligned}\Delta L_{V2}(t) \leq & \frac{1}{\gamma} [-(1 - \mu_a \| \Xi_a(t) \|^2 \| \boldsymbol{\omega}_c^{(2)}(t)^T N(t) \|^2) \times \| \boldsymbol{\omega}_c^{(2)}(t)^T \Xi_c(t) \|^2 \\ & + 4 \| \delta_c(t) \|^2 + 4 \| \boldsymbol{\omega}_c^{*T} \Xi_c(t) \|^2 + \| \boldsymbol{\omega}_c^{(2)}(t)^T N(t) \delta_a(t) \|^2].\end{aligned}\quad (55)$$

According to Eqs. (49) and (55), we have:

$$\begin{aligned}\Delta L_V(t) = & \Delta L_{V1}(t) + \Delta L_{V2}(t) \\ = & -\beta^2(1 - \mu_c \beta^2 \| \Xi_c \|^2) \times \| \boldsymbol{\omega}_c^{(2)}(t)^T \Xi_c(t) \\ & + \beta^{-1} Y(t) - \beta^{-1} \boldsymbol{\omega}_c^{(2)}(t - \Delta t)^T \Xi_c(t - \Delta t) \|^2 \\ & - \left(\beta^2 - \frac{4}{\gamma} \right) \| \delta_c(t) \|^2 - \frac{1}{\gamma_1} \| \boldsymbol{\omega}_c^{(2)}(t)^T \Xi_c(t) \|^2 \\ & \times (1 - \mu_a \| \Xi_c(t) \|^2 \| \boldsymbol{\omega}_c^{(2)}(t)^T N(t) \|^2) + \Omega\end{aligned}\quad (56)$$

where $\Omega = \frac{4}{\gamma} \| \boldsymbol{\omega}_c^{*T} \Xi_c(t) \|^2 + \frac{1}{\gamma} \| \boldsymbol{\omega}_c^{(2)}(t)^T N(t) \delta_a(t) \|^2 + \| \beta \boldsymbol{\omega}_c^{*T} \Xi_c(t) + Y(t) - \boldsymbol{\omega}_c^{(2)}(t - \Delta t)^T \Xi_c(t - \Delta t) \|^2$.

By using Cauchy-Schwarz inequality, we have:

$$\begin{aligned}\Omega(t) \leq & (4\beta^2 + \frac{4}{\gamma}) \| \boldsymbol{\omega}_c^{*T} \Xi_c(t) \|^2 + \frac{1}{\gamma} \| \boldsymbol{\omega}_c^{(2)}(t)^T N(t) \delta_a(t) \|^2 \\ & + 4 \| Y(t) \|^2 + 2 \| \boldsymbol{\omega}_c^{(2)}(t - \Delta t)^T \Xi_c(t - \Delta t) \|^2 \\ \leq & (4\beta^2 + \frac{4}{\gamma} + 2) \bar{\omega}_c^2 \Xi_{cm}^2 + 4 Y_m^2 + \frac{1}{\gamma} \omega_{cM}^2 N_m \bar{\omega}_{am}^2 \Xi_{am}^2 \\ = & D_m\end{aligned}\quad (57)$$

where $\bar{\omega}_{am}$, ω_{cM} , Ξ_{cm} , Ξ_{am} , Y_m and N_m are the upper bounds of $\tilde{\omega}_a(t)$, $\tilde{\omega}_c(t)$, $\Xi_c(t)$, $\Xi_a(t)$, $Y(t)$ and $N(t)$, respectively, and $\bar{\omega}_c = \max\{\omega_{cm}, \omega_{cM}\}$.

Choosing $\mu_a < [1/(\| \Xi_a(t) \|^2 \| \boldsymbol{\omega}_c^{(2)}(t)^T M(t) \|^2)]$, $\| \delta_a(t) \|^2 > [D_m/(\beta^2 - \frac{4}{\gamma})]$, $\mu_c < [1/(\beta^2 \| \Xi_c(t) \|^2)]$ and $\beta^2 - \frac{4}{\gamma} > 0$, then $\Delta L_V(t) \leq 0$.

According to the Lyapunov stability theorem, the estimation error $\tilde{\omega}_c^{(2)}$ and $\tilde{\omega}_a^{(2)}$ are bounded.

Remark 2 The actual input signal of the UUV consists of the robust adaptive control law and the optimal control law. The velocity error and trajectory tracking error under the robust

adaptive control law are the driving signals of the optimal control law based on ADHDP. In other words, the optimal control law based on ADHDP is inactive and the equivalent output is 0 if the velocity and trajectory are able to track the desired trajectory under the robust adaptive control law.

Remark 3 With respect to the trajectory tracking control law developed in [Li et al. \(2021\)](#) and [Yang et al. \(2021\)](#), manual adjustment of the control law design parameters does not easily guarantee control performance. Once the UUV suffers from the bigger uncertain dynamics and time-varying disturbances, the fixed control law design parameters can't attain the desirable tracking accuracy. With the incorporation of ADHDP into the control law design, the control law design parameters are updated automatically, thus effectively improving the control performance of the system.

SIMULATION RESULTS

In this section, we perform some simulations on an underactuated UUV to show the effectiveness and superiority of the proposed control scheme. As a comparison, the robust adaptive NN control scheme without the optimal control law based on ADHDP is labeled as "RC". As a comparison, an advanced kalman filter-Based control scheme based on [Vafamand, Arefi & Anvari-Moghaddam \(2023\)](#) is labeled as "AKF" and the design parameters about the augmented Kalman filter is selected the same as [Vafamand, Arefi & Anvari-Moghaddam \(2023\)](#). The ADHDP-based self-learning control scheme proposed is labeled as the "ADHDP". The model parameters of the UUV ([Liang et al., 2020](#)) are given as follows: $m_{11} = 47.5\text{kg}$, $m_{22} = 94\text{kg}$, $m_{33} = 94\text{kg}$, $m_{44} = 13.5\text{kg} \cdot \text{m}^2$, $m_{55} = 13.4\text{kg} \cdot \text{m}^2$, $d_{11} = 13.5\text{kg} \cdot \text{s}^{-1}$, $d_{22} = 45\text{kg} \cdot \text{s}^{-1}$, $d_{33} = 45\text{kg} \cdot \text{s}^{-1}$, $d_{44} = 23.8\text{kg} \cdot \text{m}^2 \cdot \text{s}^{-1}$, $d_{55} = 27.2\text{kg} \cdot \text{m}^2 \cdot \text{s}^{-1}$ and $\rho g \nabla GM_L = 8.8$.

The unknown disturbance vector is given as $\mathbf{d}(t) = [5 \sin(0.15t) + 2 \cos(0.15t), 0.5 \sin(0.15t) + 0.5 \cos(0.15t), 0.5 \sin(0.15t) + 0.5 \cos(0.15t), \sin(0.15t) + \cos(0.15t), \sin(0.15t) + \cos(0.15t)]^T$. The desired trajectory is given as $x_d = 10 \cos(0.1t)$, $y_d = 10 \sin(0.1t)$ and $z_d = -10 - 0.1t$.

The initial conditions are chosen as $\boldsymbol{\eta}(0) = [8.5\text{m}, 1\text{m}, -9\text{m}, 0.1\pi\text{rad}, 0\text{rad}]^T$, $\mathbf{v}(0) = [0\text{m/s}, 0\text{m/s}, 0\text{m/s}, 0\text{rad/s}, 0\text{rad/s}]^T$. The design parameters of the robust adaptive NN control law are selected as $\mathcal{F} = 0.3$, $\Gamma = 0.8$, $\vartheta = 0.002$, $\mathbf{k}_1 = \text{diag}(0.2, 0.1, 0.2)$, $\mathbf{k}_2 = \text{diag}(10, 10, 10)$.

The design parameters of optimal control law are chosen as $\beta = 0.95$, $\mathbf{F} = 0.001\mathbf{I}^{12 \times 12}$, $L = 0.001\mathbf{I}^{3 \times 3}$, respectively. The learning rates are selected as $\mu_a(0) = 0.1$ and $\mu_c(0) = 0.1$. The learning rate gradually decreases to the final values $\mu_a(\infty) = 0.005$ and $\mu_c(\infty) = 0.005$ over time. The number of hidden nodes for the action-critic network are given as $I_{ch} = 6$, $I_{ah} = 5$ and the initial weights of action-critic network are randomly generated within the range $[-0.2, 0.2]$. The maximal iteration numbers are set as $n_a = 200$ for the action network and $n_c = 200$ for the critic network. The completion of internal training is determined by reaching either the maximal iteration numbers. Herein, define the performance metrics as $J_{s1} = \int_0^T (|s_{11}| + |s_{12}| + |s_{13}|)dt$, $J_{s2} = \int_0^T (|s_{21}| + |s_{22}| + |s_{23}|)dt$.

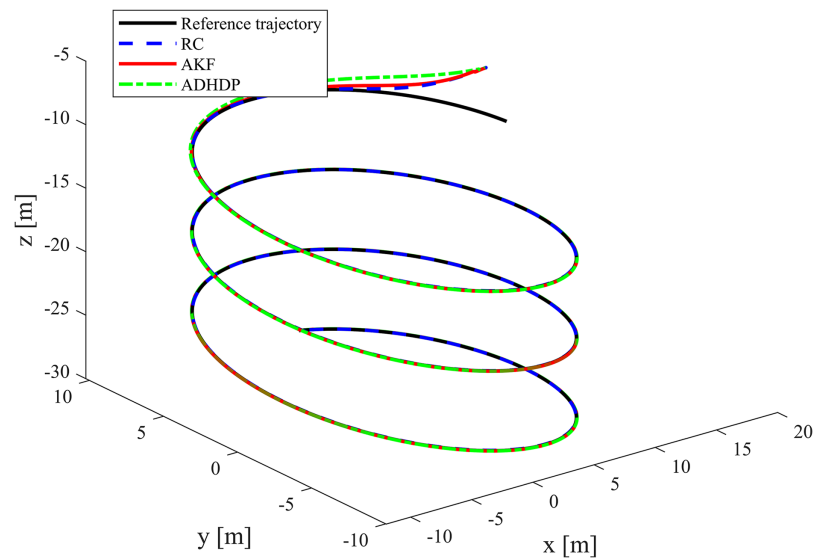


Figure 3 Desired and actual trajectories of the UUV. Full-size [DOI: 10.7717/peerj-cs.2605/fig-3](https://doi.org/10.7717/peerj-cs.2605/fig-3)

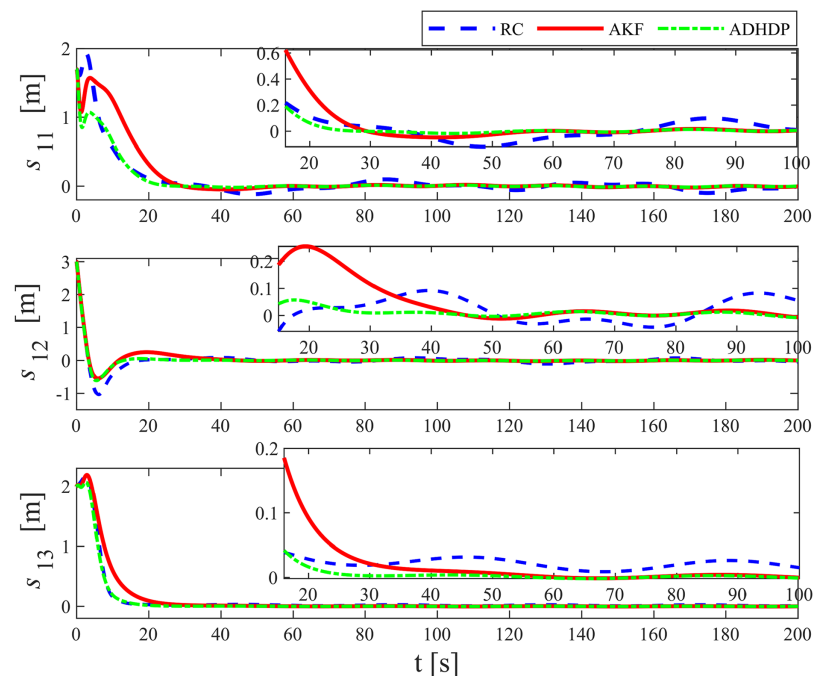


Figure 4 The tracking error of the UUV. Full-size [DOI: 10.7717/peerj-cs.2605/fig-4](https://doi.org/10.7717/peerj-cs.2605/fig-4)

Figures 3–8 show the simulation results of the proposed control scheme and the comparison control scheme. From Fig. 3, it can be seen that the underactuated UUV can accurately track the 3D spiral drive trajectory by the proposed control scheme. Specifically, it can be seen from Fig. 4 that the proposed control scheme has a smaller tracking error compared to the these two control scheme. It is observed in Fig. 4 that the tracking error under “AKF” is smaller than the tracking error under “RC”. The reason for this is that the main role of Kalman filtering in control is to provide accurate estimates of the state of the

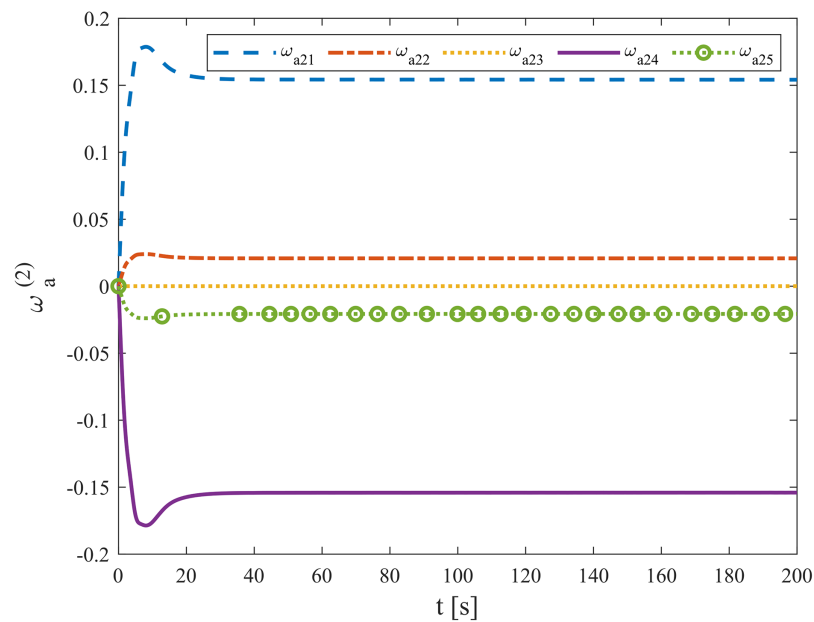


Figure 5 The weights of action network.

Full-size DOI: 10.7717/peerj-cs.2605/fig-5

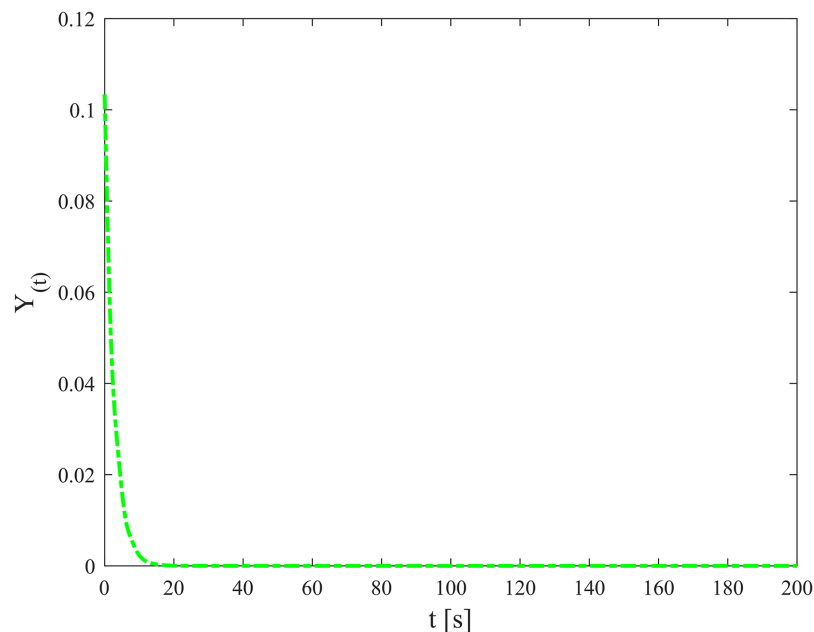


Figure 6 Utility function.

Full-size DOI: 10.7717/peerj-cs.2605/fig-6

system, especially in the presence of noise and uncertainty, and these estimates help to improve the control strategy and hence the stability and performance of the system. As can be seen in Figs. 5, 6, the optimal control part is enacted to minimize the tracking error resulted from the initial state deviation. As the weights are adaptively adjusted, the tracking error and utility function gradually decrease. As the tracking error approaches 0, the utility function decreases to 0 and the weights remain constant. Moreover, the control

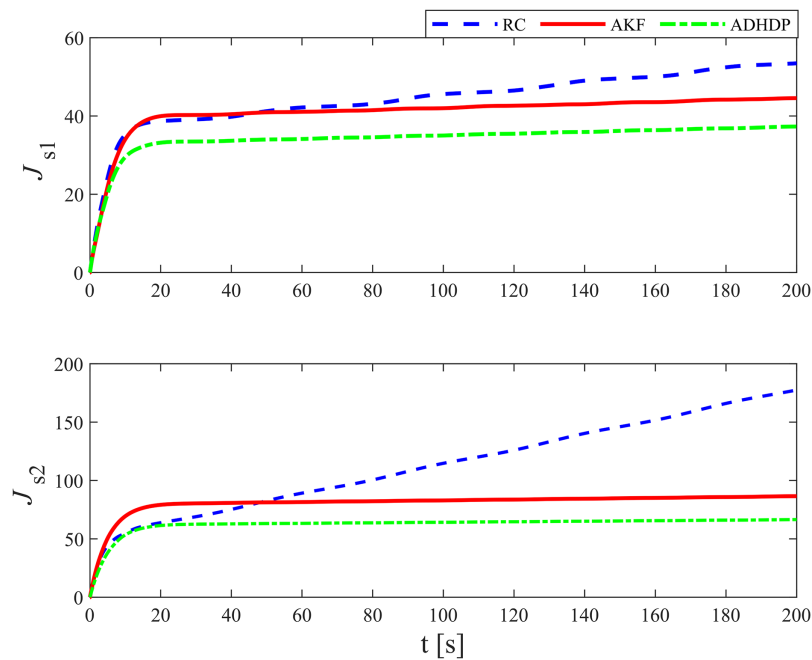


Figure 7 The performance metrics.

Full-size DOI: 10.7717/peerj-cs.2605/fig-7

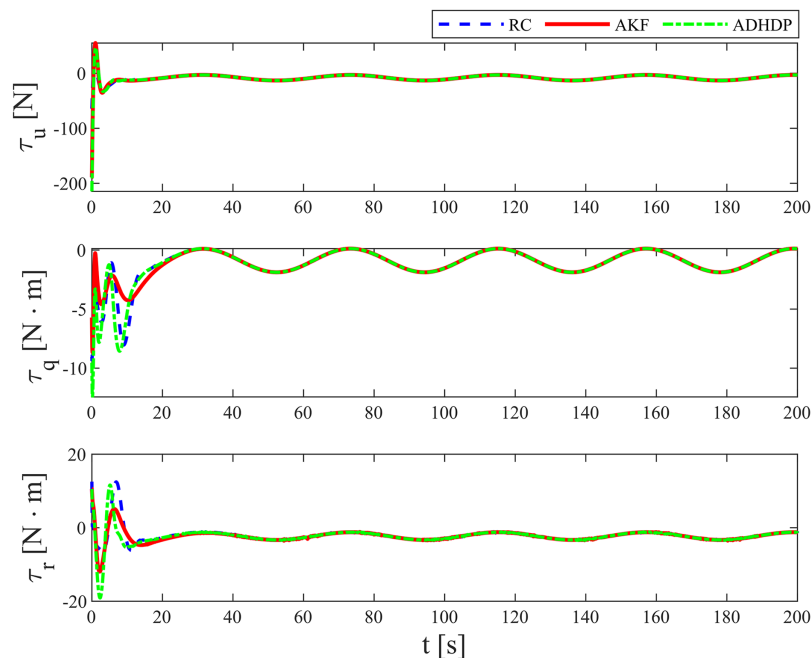


Figure 8 The actual control input.

Full-size DOI: 10.7717/peerj-cs.2605/fig-8

performance can be assessed through the performance metrics as shown in Fig. 7. Clearly, the performance metric values of the proposed control scheme are smaller than the metric values of the compared control scheme, corroborating the feature that the proposed control scheme is able to optimize the tracking performance. The performance metric values under “AKF” is smaller than under “RC”. The effect of Kalman filtering is further

demonstrated in Fig. 7. As shown in Fig. 8, the control input of the proposed control scheme is within a reasonably bounded range.

CONCLUSION

In this work, we have proposed a robust self-learning control scheme based on ADHDP to deal with the 3D trajectory tracking control problem of an UUV with uncertain dynamics and time-varying ocean disturbances. By combining the ADHDP optimization control scheme with the robust adaptive control scheme, an adaptive self-learning optimal scheme with online learning function is proposed. The proposed control method requires less model information and is more suitable for the actual situation of the system. In addition, the control parameters can be automatically updated according to the changes of external environment and unknown dynamics. Theoretical analyses as well as the comparison of simulations show that the proposed control scheme has significant effectiveness and superiority.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Chunbo Zhao performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Huaran Yan conceived and designed the experiments, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Deyi Gao performed the computation work, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The model is available in the [Supplemental File](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.2605#supplemental-information>.

REFERENCES

- Che G. 2022.** Single critic network based fault-tolerant tracking control for underactuated auv with actuator fault. *Ocean Engineering* **254**(7):111380 DOI [10.1016/j.oceaneng.2022.111380](https://doi.org/10.1016/j.oceaneng.2022.111380).
- Che G, Yu Z. 2020.** Neural-network estimators based fault-tolerant tracking control for auv via adp with rudders faults and ocean current disturbance. *Neurocomputing* **411**(4):442–454 DOI [10.1016/j.neucom.2020.06.026](https://doi.org/10.1016/j.neucom.2020.06.026).

- Chen Y, Li J, Wang K, Ning S. 2017.** Robust trajectory tracking control of underactuated underwater vehicle subject to uncertainties. *Journal of Marine Science and Technology* **25(3)**:5 DOI [10.6119/JMST-016-1219-1](https://doi.org/10.6119/JMST-016-1219-1).
- Gong H, Er MJ, Liu Y. 2024.** Fuzzy optimal fault-tolerant trajectory tracking control of underactuated auvs with prescribed performance in 3-D space. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* Epub ahead of print 7 October 2024 DOI [10.1109/TSMC.2024.3452122](https://doi.org/10.1109/TSMC.2024.3452122).
- Heshmati-Alamdari S, Nikou A, Dimarogonas DV. 2021.** Robust trajectory tracking control for underactuated autonomous underwater vehicles in uncertain environments. *IEEE Transactions on Automation Science and Engineering* **18(3)**:1288–1301 DOI [10.1109/TASE.2020.3001183](https://doi.org/10.1109/TASE.2020.3001183).
- Li Y, Liang S, Xu B, Hou M. 2021.** Predefined-time asymptotic tracking control for hypersonic flight vehicles with input quantization and faults. *IEEE Transactions on Aerospace and Electronic Systems* **57(5)**:2826–2837 DOI [10.1109/TAES.2021.3068442](https://doi.org/10.1109/TAES.2021.3068442).
- Li Y, Sun K, Tong S. 2019.** Observer-based adaptive fuzzy fault-tolerant optimal control for SISO nonlinear systems. *IEEE Transactions on Cybernetics* **49(2)**:649–661 DOI [10.1109/TCYB.2017.2785801](https://doi.org/10.1109/TCYB.2017.2785801).
- Li Z, Wang M, Ma G, Zou T. 2023.** Adaptive reinforcement learning fault-tolerant control for AUVs with thruster faults based on the integral extended state observer. *Ocean Engineering* **271(11)**:113722 DOI [10.1016/j.oceaneng.2023.113722](https://doi.org/10.1016/j.oceaneng.2023.113722).
- Liang X, Qu X, Wang N, Zhang R, Li Y. 2020.** Three-dimensional trajectory tracking of an underactuated AUV based on fuzzy dynamic surface control. *IET Intelligent Transport Systems* **14(5)**:364–370 DOI [10.1049/iet-its.2019.0347](https://doi.org/10.1049/iet-its.2019.0347).
- Liang S, Xu B, Zhang Y. 2023.** Robust self-learning fault-tolerant control for hypersonic flight vehicle based on ADHDP. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **53(9)**:5295–5306 DOI [10.1109/TSMC.2023.3264552](https://doi.org/10.1109/TSMC.2023.3264552).
- Liu J, Du J. 2021.** Composite learning tracking control for underactuated autonomous underwater vehicle with unknown dynamics and disturbances in three-dimension space. *Applied Ocean Research* **112(8)**:102686 DOI [10.1016/j.apor.2021.102686](https://doi.org/10.1016/j.apor.2021.102686).
- Luo C, Lei H, Li J, Zhou C. 2020.** A new adaptive neural control scheme for hypersonic vehicle with actuators multiple constraints. *Nonlinear Dynamics* **100(4)**:3529–3553 DOI [10.1007/s11071-020-05707-2](https://doi.org/10.1007/s11071-020-05707-2).
- Na J, Wang S, Liu YJ, Huang Y, Ren X. 2020.** Finite-time convergence adaptive neural network control for nonlinear servo systems. *IEEE Transactions on Cybernetics* **50(6)**:2568–2579 DOI [10.1109/TCYB.2019.2893317](https://doi.org/10.1109/TCYB.2019.2893317).
- Peng Z, Wang J, Han QL. 2019.** Path-following control of autonomous underwater vehicles subject to velocity and input constraints via neurodynamic optimization. *IEEE Transactions on Industrial Electronics* **66(11)**:8724–8732 DOI [10.1109/TIE.2018.2885726](https://doi.org/10.1109/TIE.2018.2885726).
- Qiao L, Zhang W. 2020.** Trajectory tracking control of AUVs via adaptive fast nonsingular integral terminal sliding mode control. *IEEE Transactions on Industrial Informatics* **16(2)**:1248–1258 DOI [10.1109/TII.2019.2949007](https://doi.org/10.1109/TII.2019.2949007).
- Shen Z, Bi Y, Wang Y, Guo C. 2020.** MLP neural network-based recursive sliding mode dynamic surface control for trajectory tracking of fully actuated surface vessel subject to unknown dynamics and input saturation. *Neurocomputing* **377(10)**:103–112 DOI [10.1016/j.neucom.2019.08.090](https://doi.org/10.1016/j.neucom.2019.08.090).
- Shi Y, Xie W, Zhang G, Zhang W, Silvestre C. 2024.** Event-triggered quantitative prescribed performance neural adaptive control for autonomous underwater vehicles. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **54(6)**:3381–3392 DOI [10.1109/TSMC.2024.3357252](https://doi.org/10.1109/TSMC.2024.3357252).

- Tong S, Sun K, Sui S. 2018.** Observer-based adaptive fuzzy decentralized optimal control design for strict-feedback nonlinear large-scale systems. *IEEE Transactions on Fuzzy Systems* **26**(2):569–584 DOI [10.1109/TFUZZ.2017.2686373](https://doi.org/10.1109/TFUZZ.2017.2686373).
- Vafamand N, Arefi MM, Anvari-Moghaddam A. 2023.** Advanced kalman filter-based backstepping control of ac microgrids: a command filter approach. *IEEE Systems Journal* **17**(1):1060–1070 DOI [10.1109/JSYST.2022.3221522](https://doi.org/10.1109/JSYST.2022.3221522).
- Van M. 2019.** Adaptive neural integral sliding-mode control for tracking control of fully actuated uncertain surface vessels. *International Journal of Robust and Nonlinear Control* **29**(5):1537–1557 DOI [10.1002/rnc.4455](https://doi.org/10.1002/rnc.4455).
- Wang N, Gao Y, Zhang X. 2021.** Data-driven performance-prescribed reinforcement learning control of an unmanned surface vehicle. *IEEE Transactions on Neural Networks and Learning Systems* **32**(12):5456–5467 DOI [10.1109/TNNLS.2021.3056444](https://doi.org/10.1109/TNNLS.2021.3056444).
- Wang N, Gao Y, Zhao H, Ahn CK. 2021a.** Reinforcement learning-based optimal tracking control of an unknown unmanned surface vehicle. *IEEE Transactions on Neural Networks and Learning Systems* **32**(7):3034–3045 DOI [10.1109/TNNLS.2020.3009214](https://doi.org/10.1109/TNNLS.2020.3009214).
- Wang J, Wang C, Wei Y, Zhang C. 2019.** Command filter based adaptive neural trajectory tracking control of an underactuated underwater vehicle in three-dimensional space. *Ocean Engineering* **180**(8):175–186 DOI [10.1016/j.oceaneng.2019.03.061](https://doi.org/10.1016/j.oceaneng.2019.03.061).
- Wang N, Zhu Z, Qin H, Deng Z, Sun Y. 2021b.** Finite-time extended state observer-based exact tracking control of an unmanned surface vehicle. *International Journal of Robust and Nonlinear Control* **31**(5):1704–1719 DOI [10.1002/rnc.5369](https://doi.org/10.1002/rnc.5369).
- Wen G, Ge SS, Chen CLP, Tu F, Wang S. 2019.** Adaptive tracking control of surface vessel using optimized backstepping technique. *IEEE Transactions on Cybernetics* **49**(9):3420–3431 DOI [10.1109/TCYB.2018.2844177](https://doi.org/10.1109/TCYB.2018.2844177).
- Yan T, Xu Z, Yang SX, Gadsden SA. 2024.** Distributed neurodynamics-based backstepping optimal control for robust constrained consensus of underactuated underwater vehicles fleet. *IEEE Transactions on Cybernetics* **54**(8):4666–4677 DOI [10.1109/TCYB.2023.3301737](https://doi.org/10.1109/TCYB.2023.3301737).
- Yang X, Yan J, Hua C, Guan X. 2021.** Trajectory tracking control of autonomous underwater vehicle with unknown parameters and external disturbances. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **51**(2):1054–1063 DOI [10.1109/TSMC.2019.2894171](https://doi.org/10.1109/TSMC.2019.2894171).
- Yuan C, Shuai C, Ma J, Fang Y. 2023.** An efficient control allocation algorithm for over-actuated auvs trajectory tracking with fault-tolerant control. *Ocean Engineering* **273**(1):113976 DOI [10.1016/j.oceaneng.2023.113976](https://doi.org/10.1016/j.oceaneng.2023.113976).
- Zhang Y, Li S, Liu X. 2018.** Adaptive near-optimal control of uncertain systems with application to underactuated surface vessels. *IEEE Transactions on Control Systems Technology* **26**(4):1204–1218 DOI [10.1109/TCST.2017.2705057](https://doi.org/10.1109/TCST.2017.2705057).
- Zheng Z, Ruan L, Zhu M, Guo X. 2020.** Reinforcement learning control for underactuated surface vessel with output error constraints and uncertainties. *Neurocomputing* **399**(2):479–490 DOI [10.1016/j.neucom.2020.03.021](https://doi.org/10.1016/j.neucom.2020.03.021).
- Zhou W, Fu J, Yan H, Du X, Wang Y, Zhou H. 2023.** Event-triggered approximate optimal path-following control for unmanned surface vehicles with state constraints. *IEEE Transactions on Neural Networks and Learning Systems* **34**(1):104–118 DOI [10.1109/TNNLS.2021.3090054](https://doi.org/10.1109/TNNLS.2021.3090054).
- Zhu Z, Duan Z, Qin H, Xue Y. 2023.** Adaptive neural network fixed-time sliding mode control for trajectory tracking of underwater vehicle. *Ocean Engineering* **287**(1):115864 DOI [10.1016/j.oceaneng.2023.115864](https://doi.org/10.1016/j.oceaneng.2023.115864).