Research on Motion Attitude Control of Under-actuated Autonomous Underwater Vehicle Based on Deep Reinforcement Learning

Jie Jiang\(^{1,*}\), Runqiang Zhang\(^{2,b}\), Yuan Fang\(^{3,c}\), Xinlong Wang\(^{4,d}\)

\(^{1}\)92942 army, Beijing, China
\(^{2}\)Department of Energy and Power Engineering, Tsinghua University, Beijing, 100084, China
\(^{3}\)Department of Power Engineering, Naval University of Engineering, Wuhan, 430030, China
\(^{4}\)College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao, 266590, China
\(^{b}\)westlifejie@sina.cn, \(^{b}\)runqiangzhang@mail.tsinghua.edu.cn, \(^{c}\)fy_nue@foxmail.com, \(^{d}\)xinlongwang1995@163.com

Abstract. This paper adopts Deep Deterministic Policy Gradient (DDPG) algorithm in Deep Reinforcement Learning (DRL) to analyse the control of Autonomous Underwater Vehicle (AUV) motion attitude control based on Robot Operating System (ROS) and Gazebo simulation platform, within UUV Simulator underwater simulation environment. The heel angle \(\phi\), pitch angle \(\theta\), heading angle \(\psi\) are chosen as the agent state input and control variables, and the output angle of the four rudders is selected as the agent action output. The problem of strong coupling caused by X-type rudder and multi-degree-of-freedom control is solved with reinforcement learning and training. In addition, this paper proposes a multi-state space and multi-action space control scheme, which has achieved remarkable results for the AUV's fixed speed, constant heel, constant pitch, and constant heading motion control.

1. Introduction
The deep sea is a huge resource treasure, which needs to be further excavated and utilized. Scholars have conducted in-depth research on the flow and acoustic characteristics of ocean energy [1, 2], gas-liquid two-phase transportation [3], vibration and noise [4], etc., but the numerical calculation method still remains in traditional solution mode. Deep learning currently occupies an important position in many interdisciplinary fields, in addition to being widely used in life scenarios such as emotion recognition [5, 6], it also begins to play a significant role in engineering computation and application fields.

Nowadays, engineering practice and academic research are paying more and more attention to the development of underwater robot technology, and great progress has been made in AUV detection, communication and motion control, laying a solid foundation for further effective exploration of marine resources. In the meanwhile, the motion control of AUV under complex working conditions puts forward higher requirements. Due to the strong coupling when AUV interacts with the hydrodynamic environment, it is difficult to establish accurate dynamics and kinematics models for...
the calculation and simulation. Therefore, traditional control methods have achieved limited results in the field of AUV control. Meanwhile, the emerging reinforcement learning algorithms have great potential for such situations, especially model-free reinforcement learning, which utilizes the agent and the environment without making full use of the mathematical modelling. The interaction between data and environment, generating data to train the agent are conducted to achieve the set goals according to the defined reward function.

Straight line navigation at fixed depth is the most basic form of motion for an autonomous underwater vehicle, and the realization for the pitch angle control of an autonomous underwater vehicle is important. It is of great significance in the application of snorkeling and diving, seafloor tracking and special maneuvers. Relying on the ROS and Gazebo simulation platforms, Zhang et al. [7] selected the lauv-type AUV as the research object, and explored the AUV straight line navigation with fixed depth and fixed trajectory tracking, comparing the preset reward function, the artificial input reward function, and the combined DQN (Deep Q-learning) algorithm. Similarly, Huo et al. [8] combined LSTM (Long Short-Term Memory) and DDPG algorithms to study the simple trajectory tracking of AUV in the horizontal plane. However, the problem was that this type of research only focused on the three degrees of freedom motion as surge, sway and yaw, and the distance from simulation to actual application was relatively large. Furthermore, the task generalization ability was slightly insufficient, and it can only complete straight line navigation and single fixed trajectory tracking.

With the support of the MATLAB platform, Wu et al. [9] took the classic REMUS AUV as the research object, investigated three continuous depth control situations as the Constant Depth Control, Curved Depth Tracking, Seafloor Tracking of the AUV in the vertical plane based on the DPG algorithm. However, scholars also merely concentrate on the three degrees of freedom motion, which has limited reference value in the engineering field. And the accuracy of mathematical model has a greater impact on the generalization ability of the training model in practical applications. Carlucho et al. [10] chose Nessie VII AUV as the research object, utilized the DDPG algorithm to control six thrusters and made the AUV move toward the target point. However, the research of scholars remains on the motion control of the AUV in a three-dimensional space. The actual control variables are only three degrees of freedom as surge, sway and heave, so that the AUV reaches or approaches the target point in the spatial displacement is judged to complete the target. Without considering the three degrees of freedom as heal, pitch and heading, there is a certain distance from actual application.

In view of the insufficiency in previous work, this paper conducts research on the motion control of under-actuated autonomous underwater vehicle, the eca_a9 autonomous underwater vehicle is chosen as the object and adopts the X-type rudder as the actuator relying on ROS and Gazebo simulation platforms. Based on the deep reinforcement learning framework, the DDPG algorithm [11] is utilized to explore the motion at fixed speed, constant heel, pitch, and heading under six degrees of freedom, and achieves remarkable control results.

2. Materials and Methods
As the action space of the algorithm [7] is one, and the output action of the agent controls two rudders at the same time. While the action space of the algorithm in this paper is four, and the output actions of the agent independently control the four rudders, so that the AUV keeps the heel at 0 degree and the pitch at 0 degree when it sails straight.

Based on the ROS and Gazebo simulation platform, this paper selects the UUV Simulator underwater simulation environment [12] to study the motion attitude control of the under-actuated autonomous underwater vehicle based on deep reinforcement learning.

The DDPG algorithm employed in this paper belongs to the framework of reinforcement learning, and the basic concept of reinforcement learning is present as follows. When the agent is in a certain state during the process of interacting with the environment, it will choose the corresponding action according to its own strategy, and then interacts with the environment to obtain the reward function, next the state is updated. In the process of cycling the above steps, the algorithm is used to
continuously update the agent's own strategy, so as to achieve the purpose of training the agent and learning to complete the task. Usually the reinforcement learning process is based on Markov Decision Process (MDP) [13].

MDP is mainly composed of four parts: S, A, P, and R. S represents the state space of the agent in the environment, A represents the action space of the agent in the environment, and P represents the state transition probability when the agent performs actions in a certain state, and R represents the reward function. At the time of t, the agent selects the action at from the action space according to its own strategy according to its state st, so that the state of the agent in the environment is transferred from st to st+1 according to the state transition probability, and than obtain the reward function r(st, at). In the above process of the loop, the state transition probability and r(st, at) are only related to the current state st and the current action at, which have nothing to do with the previous state and action. Reinforcement learning is based on MDP, which interacts with the environment through the agent, and than obtain the reward value. The parameters are updated according to the algorithm to optimize the further results, and obtain the strategy that can obtain the maximum reward.

The research goal of this paper is to control the AUV to move at a fixed speed, constant heel, pitch and heading. The AUV thruster output is set to be constant, and the target heel angle, target pitch angle and target heading angle are all 0 degrees, so the state s=(φ, θ, ϕ) is selected. Where φ represents the heel angle, θ represents the pitch angle, and ϕ represents the heading angle. In this paper, the AUV's thruster is set as a fixed speed, so that we can just control the four rudder angles of the AUV. Therefore, the action a=(fin_0, fin_1, fin_2, fin_3) is selected, where the four quantities in a represent the output angle of four rudders, as shown in Figure 1. At the same time, in order to prevent the AUV from interacting with the underwater environment immoderately, the angle range of each rudder is limited to (-35°, 35°), and the reward function is set as R = -|φ| - |θ| - |ϕ|.

Fig.1 The output angle of four rudders

3. Results and Discussions
In order to effectively overcome the problem that DQN algorithm [14] cannot handle continuous action space, this paper employs the DDPG algorithm in reinforcement learning to perform fixed speed, constant heel, pitch and heading motions on the AUV. When the agent is trained for 110 episodes, it can be seen from Figure 2 that the cumulative rewards from 0 to 40th episodes gradually increase, and converge from 40th episode. The following is a phased study of the AUV's motion attitude.
Fig. 2 Cumulative rewards obtained by DDPG agent in the fixed speed, constant heal angle, pitch angle and heading motion

**Phase A The primary episode**

In this phase, the training has not yet started. It can be seen from Figure 3 that the output angle of the four rudders in the AUV present a certain degree of randomness. As can be seen from Figure 4 and Figure 5 that the motion behavior of the AUV cannot reach the predetermined goal.

Fig. 3 The output angle of the four rudders in the primary episode

Fig. 4 Angle attitude change of AUV at the primary episode
**Phase B The 40th episode**

As for this phase, the control strategy of the AUV has just begun to converge, and the output angles of the four rudders start to show regular output trend. The angle attitude of the AUV has been initially controlled, but the set goals have not been achieved.

---

**Fig. 5** Space displacement of AUV in the primary episode

**Fig. 6** The output angle of the four rudders in the 40th episode

**Fig. 7** Angle attitude change of AUV in the 40th episode
Phase C The 100th episode
At this time, the control strategy of the AUV has converged, and the four rudders can output the control angle stably. The angle attitude of the AUV has been well controlled, and the set goal can be achieved.
4. Conclusion
This paper employs the DDPG algorithm to control the eca_a9 AUV motion attitude based on the ROS, Gazebo simulation platform and the UUV Simulator underwater simulation environment. The heel angle $\phi$, the pitch angle $\theta$, and the heading angle $\varphi$ are selected as the agent state input and control variables. The output angles of the four rudders are used as the agent action output. After reinforcement learning and training, the strong coupling problem caused by the X-type rudder and the multi-degree-of-freedom control has been overcome. And the control problems of multi-state space and multi-action space is solved. The AUV's motion control with fixed speed, constant heel, pitch, and heading has achieved remarkable results. The cumulative reward for the 40th episode is -45, which is 92.13% higher than the primary episode of -571.90, and the cumulative reward of the 100th episode is -23.46, which present an increase of 95.90% compared to the primary episode. The exploration and study on more forms of motion control problems in AUV can be conducted based on the simulation platform and real AUV test platform, according to DDPG and other reinforcement learning algorithms in the future.

References
[1] Huang Z, Huang ZY, Fan H. Influence of C groove on energy performance and noise source of a NACA0009 hydrofoil with tip clearance [J]. Renewable Energy, 2020.
[2] Huang ZW, Zheng H, Guo L, et al. Influence of the Position of Artificial Boundary on Computation Accuracy of Conjugated Infinite Element for a Finite Length Cylindrical Shell [J]. Acoustics Australia, 2020:1-8.
[3] Zhang J, Fan H, Zhang W, et al. Energy performance and flow characteristics of a multiphase pump with different tip clearance sizes[J]. Advances in Mechanical Engineering, 2019, 11(1): 1687814018823356.
[4] Lin Zhang, Jia-xi Duan, Liang-long Da, Guo-jun Xu, Xue-hai Sun. Vibroacoustic radiation and propagation properties of slender cylindrical shell in uniform shallow sea [J]. Ocean Engineering 195 (2020) 106659.
[5] Du X, Ma C, Zhang G, et al. An Efficient LSTM Network for Emotion Recognition from Multichannel EEG Signals[J]. IEEE Transactions on Affective Computing, 2020.
[6] ZHANG G, YU M, CHEN G, et al. A review of EEG features for emotion recognition[J]. SCIENTIA SINICA Informationis, 2019, 49(9): 1097-1118.
[7] Zhang Q, Lin J, Sha Q, et al. Deep Interactive Reinforcement Learning for Path Following of Autonomous Underwater Vehicle[J]. IEEE Access, 2020, 8: 24258-24268.
[8] Huo Y, Li Y, Feng X. Model-Free Recurrent Reinforcement Learning for AUV Horizontal Control[J]. MS&E, 2018, 428(1): 012063.
[9] Wu H, Song S, You K, et al. Depth control of model-free auvs via reinforcement learning[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2018, 49(12): 2499-2510.
[10] Carlucho I, De Paula M, Wang S, et al. AUV Position Tracking Control Using End-to-End Deep Reinforcement Learning. In the proceeding of OCEANS 2018 MTS/IEEE Charleston. IEEE, 2018: 1-8.
[11] Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous control with deep reinforcement learning[J]. arXiv preprint arXiv:1509.02971, 2015.
[12] Manhães M M M, Scherer S A, Voss M, et al. UUV simulator: A gazebo-based package for underwater intervention and multi-robot simulation. In proceeding of OCEANS 2016 MTS/IEEE Monterey. IEEE, 2016: 1-8.
[13] Sutton R S, Barto A G. Reinforcement learning: An introduction[M]. MIT press, 2018.
[14] Mnih V, Kavukcuoglu K, Silver D, et al. Playing atari with deep reinforcement learning[J]. arXiv preprint arXiv:1312.5602, 2013.