

## Letter

## Multi-AUV Inspection for Process Monitoring of Underwater Oil Transportation

Jingyi He, Jiabao Wen, Shuai Xiao, and  
Jiachen Yang, *Senior Member, IEEE*

Dear Editor,

This letter presents an inspection method for process monitoring of underwater oil transportation via multiple autonomous underwater vehicles (AUV). To improve the adaptability of our method in practice, we introduce the dynamic complex ocean current data to the previously mentioned case by using regional ocean modeling system (ROMS) for the first time, and elaborately design AUV's local information, as well as the deep reinforcement learning (DRL) tuple  $(S_t, A_t, R_t, S_{t+1})$ . Specifically, according to the local information of each AUV, including captured pipeline image, ocean current velocity, position coordinates and pose angles, an agent can be well trained to control the AUV for underwater oil pipeline tracking through vision-hydrology-motion-based soft actor critic (VHM-SAC) mechanism. Experimental results show that the proposed method can provide stable oil pipeline tracking performance in dynamic complex underwater environment.

Underwater oil transportation is an important part of offshore oil-field systems, which can continuously transport large amounts of oil in a fast, efficient and economical way [1]. However, the lifespan of oil pipelines is limited. With the increase of service time, the aging and corrosion of the pipelines occur from time to time. The resulting oil leakage will not only bring serious economic losses, but also cause long-term pollution to the environment. In this case, it is of great significance to be able to give an alarm in a timely manner through process monitoring [2] and take remedial measures when the pipeline leaks.

The main purpose of underwater oil pipeline inspection is to find out whether there is a leakage point on the pipeline, which can usually be divided into internal inspection methods and external inspection methods. The former mainly uses ultrasonic, magnetic flux, eddy current and other technical means to complete, the results are accurate but only suitable for large-diameter pipes. The latter mainly focuses on parameters such as flow difference, pressure difference, negative pressure wave, etc., the cost is lower than the internal inspection methods and there is no possibility of pipeline blockage due to detection equipment [3]–[5]. Thus, the external inspection methods are more common in the field of underwater oil pipeline inspection.

AUV is a typical representative of underwater unmanned system. With its small, flexible and low-energy-consumption features, it can play an important role in the external inspection of underwater oil pipelines [6] and [7]. However, the underwater environment is dynamic and complex, a single AUV may not be able to meet the needs of refined tasks. Faced with this situation, AUV cluster can realize multi-directional inspection of underwater oil pipelines through mutual cooperation [8]–[10].

Corresponding author: Jiabao Wen.

Citation: J. Y. He, J. B. Wen, S. Xiao, and J. C. Yang, "Multi-AUV inspection for process monitoring of underwater oil transportation," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 3, pp. 828–830, Mar. 2023.

The authors are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: seie\_hjy@tju.edu.cn; Wen\_jiabao@tju.edu.cn; xs611@tju.edu.cn; yangjiachen@tju.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JAS.2023.123117

With the outstanding performance of AlphaGo and other highly intelligent products, DRL method has led the development of various decision-making methods, such as traffic control, man-machine game, video coding, search referral, text processing etc. It can extract the important features from various observation states through deep neural networks, and hereafter realize the trial-and-error learning for agent through the reward-driven interaction between environment and agent. The combination of the features-extracting ability and the decision-making ability [11], making DRL method as if it is tailored for AUV underwater pipeline inspection.

**Related work:** With the development of underwater equipment technology, many researchers have carried out studies on underwater pipeline inspection and tracking by using underwater robots. Most of those studies are based on remotely operated vehicle (ROV). Matsumoto *et al.* [12] used Deep Ocean Phantom S2 ROV to design a Hough transform-based submarine cable detection system, which realized the tracking of 25 mm cables. Oceaneering company designed Magna ROV that can use electromagnetic ultrasonic technology to detect pipelines exposed on the seabed in the depth of 3 000 meters [13]. However, due to the influence of the umbilical cable, ROV must be supported by the mother ship and cannot carry out large-scale and long-distance missions. What is worse, ROV requires operators to keep their eyes on the screen for remote control. Then, researchers begin to turn their attention to applying AUV. Huang *et al.* [14] proposed an AUV-based underwater pipeline detection method by fusing multi-source information collected by AUV. Li *et al.* [15] proposed an adaptive control law for AUV pipeline tracking, which includes the estimation of uncertain parameters related to the hydrodynamic damping coefficients.

Like the methods mentioned above, most of the existing methods need to establish an accurate kinematic model for AUV, which is usually difficult when not having enough prior information. DRL method can provide control strategy without knowing the kinematic model, by interacting between agent and environment. Yu *et al.* [16] proposed an underwater motion control system based on DRL method to solve the problem of underwater robot trajectory tracking, getting more accurate results than traditional proportion-integration-differentiation (PID) control systems when the trajectory is a complex curve. Wu *et al.* [17] proposed a discrete-time model-free reinforcement learning framework to solve three depth control problems for AUV, namely constant depth control, curved depth tracking, and seafloor tracking. Unfortunately, there is few works to consider the impact of the real underwater environment when training the agent.

**VHM-SAC mechanism for underwater oil pipeline inspection:** DRL method is based on Markov decision process (MDP) theory and has trial and error exploration mechanism, which totally differs from supervised learning or semi-supervised learning [18]. As shown in Fig. 1, AUV chooses an action  $a_t$  at the state  $s_t$ , and then get the next state  $s_{t+1}$  with a reward  $r_t$  for the next interaction.

During continuous trial-and-error learning, the agent gradually has the ability to choose the best action with the highest long-term discounted cumulative rewards which can be expressed as

$$G_t = r_{t+1} + \gamma \times r_{t+2} + \gamma^2 \times r_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k \times r_{t+k+1} \quad (1)$$

where  $\gamma^k$  ( $\gamma \in [0, 1]$ ) is the discount factor and  $r_{t+k+1}$  represents the future rewards. Then, the value functions can be expressed as

$$V_{\pi}(s) = E_{\pi} \left( \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s \right) \quad (2)$$

where  $V_{\pi}(s)$  is the expectation of the long-term discounted cumulative rewards for the agent at the state  $s$ . The goal of DRL is to obtain the optimal policy  $\pi^*$  by maximizing the value functions

$$\pi^*(s) = \arg \max_{a \in A} V_{\pi}(s). \quad (3)$$

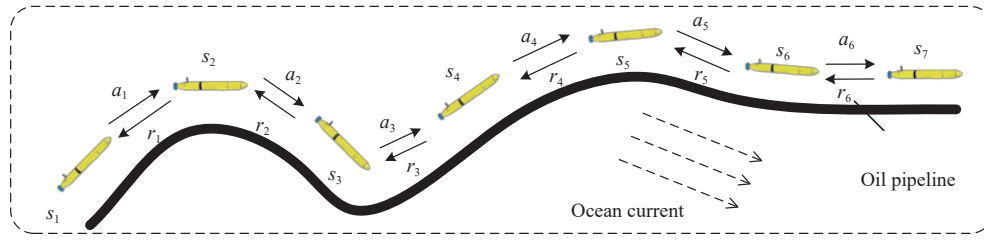


Fig. 1. The dynamic process of DRL-based AUV underwater oil pipeline inspection.

Combining the advantages of policy-based DRL algorithm and value-based DRL algorithm, actor critic (AC) algorithm becomes popular in recent years. Based on the environmental information, the actor network is used to estimate the action through the policy gradient which can be expressed as

$$\nabla_{\theta} J(\theta) = E_{\pi_{\theta}}[\nabla_{\theta} \log \pi_{\theta}(s, a) Q_{\pi}(s, a)] \quad (4)$$

where  $\theta$  is actor network's parameter,  $Q_{\pi}(s, a)$  is the expected long-term discounted cumulative rewards for an agent applying the action  $a$  at the state  $s$ . The critic network is used to estimate the value on the basis of the policy by temporal difference (TD) error which can be expressed as

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t). \quad (5)$$

In order to improve the randomness of action selection, entropy regularization is introduced into soft actor critic (SAC) algorithm. The greater the entropy, the higher the randomness of action selection. Then, the goal of the algorithm should be

$$\pi^* = \arg \max_{\pi} E_{\tau \sim \pi} \left[ \sum_{t=0}^{\infty} \gamma^t (R(s_t, a_t, s_{t+1}) + \alpha H(\pi(\cdot|s_t))) \right] \quad (6)$$

where  $H(\pi(\cdot|s_t))$  is the policy entropy representing the entropy of the output action under the policy  $\pi(a_t|s_t)$ , and  $\alpha$  is the temperature parameter.

For the VHM-SAC mechanism, the tuple  $(S_p, A_p, R_p, S_{t+1})$  can be defined as follows.

1) State space  $S$ : The goal of the agent is to control AUV cluster to move from starting point to end point along the underwater oil pipeline without collision in dynamic complex underwater environment. Therefore, the state space mainly includes captured pipeline image, ocean current velocity, position coordinates and pose angles, where the raw captured pipeline image is two dimensional and the rest raw data is one dimensional. As shown in Fig. 2, all state data is flattened and then concatenated to form a one-dimensional vector through deep neural networks, where  $(x, y, z)$  denotes the position coordinates in the earth frame,  $(\varphi, \theta, \psi)$  denotes the pose angles in the body frame,  $(u, v, w)$  denotes the ocean current velocities in the directions of the three coordinate axes. The captured image is decomposed into three matrices according to the red-green-blue (RGB) components.

2) Action space  $A$ : The discrete action space is composed of the movements in six directions: up, down, left, right, front and back. Each movement represents a fixed timestep in the corresponding direction, and only one action will be chosen at a certain time.

3) Reward function  $R$ : The design of the reward function is of great significance to the convergence of the DRL method [19]. Since both the state space and the action space are large and complex, traditional sparse reward function cannot be expected to work. Therefore, a composite reward function is designed

i) The vertical distance between the AUV and the oil pipeline,  $r_d = -0.1|d_j - d_{TH}|$ , where  $d_j$  is the vertical distance between the AUV and the oil pipeline,  $d_{TH}$  (set to 2 m) is the threshold of the vertical distance between the AUV and the oil pipeline.

ii) The distance to the target,  $r_t = -0.01 d_2$ , where  $d_2$  is the distance to the target.

iii) The collision,  $r_c = 0$  (if no collision) or  $-1$  (if collision).

Thus, the total reward for each step is  $r = r_d + r_t + r_c$ , which is clipped in  $[0, 1]$ .

The pseudo code of the SAC algorithm used is as follows:

---

#### Algorithm 1 VHM-SAC

---

**Input:** The captured image of the oil pipeline  $I_{op}$ , The position coordinates of the AUV  $P_A$ , The pose angles of the AUV  $A_A$ , The velocity of the ocean current  $V_c$

**Output:** Optimal action decision policy  $\pi^*$

Initialize parameters: Target point  $P_t$ , Number of steps  $N$ , Number of episodes  $M$ , Batch size  $K$ , Temperature parameter  $\alpha$ , Learning rate  $\lambda$ , Policy parameters  $\theta_p$ , Q-function parameters  $\theta_1, \theta_2$ , target parameters  $\theta_1' \leftarrow \theta_1, \theta_2' \leftarrow \theta_2$ , Replay memory  $D$ , Vertical distance threshold  $d_{TH}$ , Discounting factor  $\gamma$

**repeat**

**if** it is the beginning of the episode **then**

    Initialize the position of the AUV  $P_A$

**end if**

  Observe the state  $s$  and select an action  $a \sim \pi_{\theta_p}(a|s)$

  Execute the action  $a$ , get the reward  $r$ , the next state  $s'$ , and the done signal  $d$  to indicate whether  $s'$  is the terminal

  Store the transition  $(s, a, r, s')$  in  $D$

  Sample random minibatch  $K$  of transitions  $(s, a, r, s')$  from  $D$

  Compute targets for the Q functions

  Update Q-functions by one step of gradient descent

  Update policy by one step of gradient ascent

  Update the temperature parameter

  Update target networks

**until** convergence

---

**Experiments:** Our simulation and the experiments are carried out on Ubuntu 20.04 system, using an Intel Core i7-12700K CPU @ 3.60 GHz, two NVIDIA GeForce RTX 3080 GPU. The necessary software includes Gazebo 11.4, ROS Noetic Ninjemys, Python 3.8, Pytorch 1.10, NetCDF 4.1, Hdf5 1.8, MATLAB 2016b and other commonly used dependencies for data processing [20].

To evaluate the performance of the proposed method, we laid the pipeline model in *ocean\_wave.world* of Gazebo simulation software, and imported a dynamic complex ocean environment generated by ROMS to simulate the actual underwater oil pipeline inspection. The reason for using ROMS generated data is that the generated data can be more uniform and spatiotemporally denser through data processing when compared with the measured data, providing more reliable data support. The environment is a subarea of the South China Sea, covering the spatial extent of 20.75°N–22.25°N, 118.5°E–120.0°E. The average water depth is 2495.88 meters and the maximum water depth is 3650.54 meters. Linear interpolation was adopted to fill the data into continuous.

Three homogeneous AUVs with same attributes were used to form a cluster for multi-angle coordinated inspections of underwater oil pipelines. Each AUV has its corresponding agent to control itself. In our training progress, learning rate is 0.0005, batch size is 256, target entropy is  $-6$ , and temperature parameter is 1. Fig. 3 shows the result of an oil pipeline inspection via multi-AUV. It can be seen that all three AUVs can keep a certain distance and move along the oil pipeline stably. Table 1 shows the success rate of the AUV cluster to complete underwater oil pipeline inspection under the influence of different ocean current intensity. Even in the case of ocean current velocity of 1 m/s, our method can still maintain good control effect with an average success rate 86.67%.

**Conclusions:** This letter presents a VHM-SAC mechanism for process monitoring of underwater oil transportation by multi-AUV in dynamic complex underwater environment. Three agents are trained

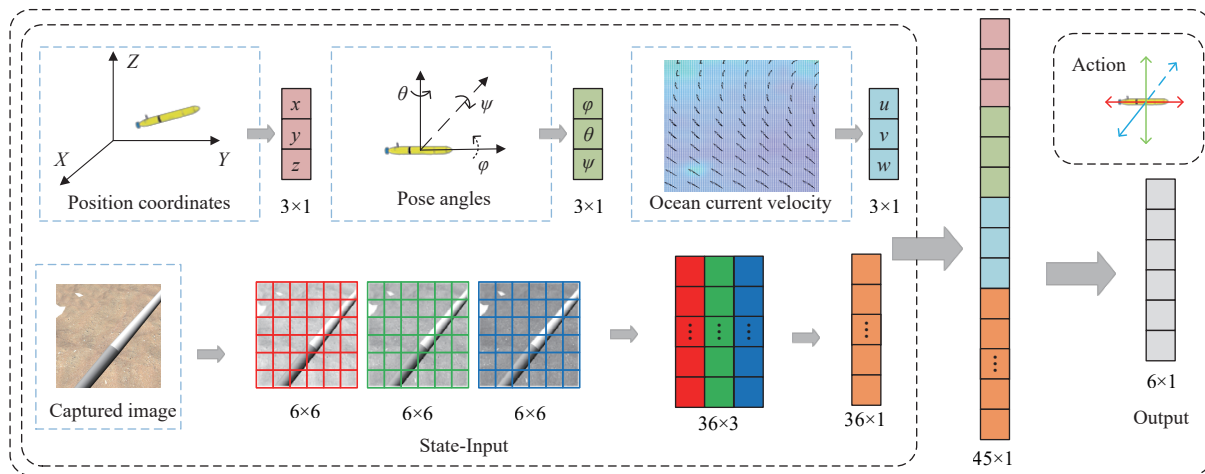


Fig. 2. The illustration of VHM-SAC mechanism.



Fig. 3. The result of an oil pipeline inspection via multi-AUV.

Table 1. Comparison of Different Ocean Current Intensity

Success rate	Ocean current intensity (m/s)			
	0.25	0.50	0.75	1.00
AUV1	100%	97%	91%	89%
AUV2	99%	94%	88%	85%
AUV3	99%	95%	89%	86%
<b>Average</b>	<b>99.33%</b>	<b>95.33%</b>	<b>89.33%</b>	<b>86.67%</b>

to control their corresponding AUV based on the local information, and perform well even with severe ocean current, providing some meaningful references for the application of AUV clusters and deep reinforcement learning in the field of process monitoring.

**Acknowledgments:** This work was supported by the National Natural Science Foundation of China (61871283).

## References

- [1] C. Xia, W. Liu, and Q. Deng, "Cost minimization of wireless sensor networks with unlimited-lifetime energy for monitoring oil pipelines," *IEEE/CAA J. Autom. Sinica*, vol. 2, no. 3, pp. 290–295, Jul. 2015.
- [2] X. Yuan, Y. Gu, Y. Wang, C. Yang, and W. Gui, "A deep supervised learning framework for data-driven soft sensor modeling of industrial processes," *IEEE Trans. Neural Networks Learning Systems*, vol. 31, no. 11, pp. 4737–4746, Nov. 2020.
- [3] T. Zhou, M. Chen, and J. Zou, "Reinforcement learning based data fusion method for multi-sensors," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 6, pp. 1489–1497, Nov. 2020.
- [4] K. Zhu and Y. Wang, "Event-triggered sensor fault estimation of unreliable networked unmanned surface vehicle system with correlated noises," *IEEE Trans. Vehicular Technology*, vol. 71, no. 3, pp. 2527–2537, Mar. 2022.
- [5] W. Chen, K. Gu, T. Zhao, G. Jiang, and P. L. Callet, "Semi-reference sonar image quality assessment based on task and visual perception," *IEEE Trans. Multimedia*, vol. 23, pp. 1008–1020, Apr. 2020.
- [6] H. Xia, M. A. Khan, Z. Li, and M. Zhou, "Wearable robots for human underwater movement ability enhancement: A survey," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 6, pp. 967–977, Jun. 2022.
- [7] J. Wen, J. Yang, and T. Wang, "Path planning for autonomous underwater vehicles under the influence of ocean currents based on a fusion heuristic algorithm," *IEEE Trans. Vehicular Technology*, vol. 70, no. 9, pp. 8529–8544, Sept. 2021.
- [8] S. Wang, L. Chen, D. Gu, and H. Hu, "Cooperative localization of AUVs using moving horizon estimation," *IEEE/CAA J. Autom. Sinica*, vol. 1, no. 1, pp. 68–76, Jan. 2014.
- [9] Z. Gao and G. Guo, "Fixed-time sliding mode formation control of AUVs based on a disturbance observer," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 2, pp. 539–545, Mar. 2020.
- [10] H. Li, D. Wang, M. Zhou, Y. Fan, and Y. Xia, "Multi-swarm co-evolution based hybrid intelligent optimization for BI-objective multi-workflow scheduling in the cloud," *IEEE Trans. Parallel Distributed Systems*, vol. 33, no. 9, pp. 2183–2197, Sept. 2022.
- [11] J. Wang, Q. Zhang, and D. Zhao, "Highway lane change decision-making via attention-based deep reinforcement learning," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 3, pp. 567–569, Mar. 2022.
- [12] Y. Liu, L. Wan, M. Sheng, T. Liu, and Y. Li, "An improved elbows detection algorithm for underwater blurred images," in *Proc. Conf. Electrical, Automation and Mechanical Engineering*, Apr. 2017, pp. 264–267.
- [13] G. P. Drummond, I. P. Pasqualino, B. C. Pinheiro, and S. F. Estefen, "Pipelines, risers and umbilicals failures: A literature review," *Ocean Engineering*, vol. 148, pp. 412–425, Jan. 2018.
- [14] X. Huang, Y. Li, F. Du, and S. Jin, "Horizontal path following for underactuated AUV based on dynamic circle guidance," *Robotica*, vol. 35, no. 4, pp. 876–891, Apr. 2017.
- [15] Z. Li, J. Deng, R. Lu, Y. Xu, J. Bai, and C. Su, "Trajectory-tracking control of mobile robot systems incorporating neural-dynamic optimized model predictive approach," *IEEE Trans. Systems, Man, Cybernetics: Systems*, vol. 46, no. 6, pp. 740–749, Jun. 2016.
- [16] R. Yu, Z. Shi, C. Huang, T. Li, and Q. Ma, "Deep reinforcement learning based optimal trajectory tracking control of autonomous underwater vehicle," in *Proc. Chinese Control Conf.*, Jul. 2017, pp. 495–4965.
- [17] H. Wu, S. Song, K. You, and C. Wu, "Depth control of model-free AUVs via reinforcement learning," *IEEE Trans. Systems, Man, Cybernetics: Systems*, vol. 49, no. 12, pp. 2499–2510, Dec. 2019.
- [18] Y. Li and X. Chao, "Semi-supervised few-shot learning approach for plant diseases recognition," *Plant Methods*, vol. 17, no. 68, pp. 1–10, Jun. 2021.
- [19] M. Xi, J. Yang, J. Wen, H. Liu, Y. Li, and H. H. Song, "Comprehensive ocean information-enabled AUV path planning via reinforcement learning," *IEEE Internet of Things J.*, vol. 9, no. 18, pp. 17440–17451, Sept. 2022.
- [20] J. Yang, S. Xiao, A. Li, W. Lu, X. Gao, and Y. Li, "MSTA-Net: Forgery detection by generating manipulation trace based on multi-scale self-texture attention," *IEEE Trans. Circuits Systems Video Technology*, vol. 32, no. 7, pp. 4854–4866, Jul. 2022.