Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Geomagnetic Navigation for AUV based on Deep Reinforcement Learning Algorithm

Cong Wang, Yun Niu*, Mingyong Liu, Tingchao Shi, Jiaqi Li, Lianggen You

Abstract -**For the problem of geomagnetic navigation in the case of autonomous underwater vehicle (AUV) without a prior geomagnetic information, this paper proposes a geomagnetic navigation algorithm based on deep reinforcement learning. Using the correlation between the geomagnetic parameters and the navigation trajectory, the Deep Q Network is used for choosing the navigation direction by multiple geomagnetic components, and the relative position is estimated by the difference between the geomagnetic parameters of the current position and the target position. Under the influence of the above navigation model ,the multiple geomagnetic component gradually approaches the target value with the movement of the aircraft, thereby achieving the navigation purpose. By comparing with the evolutionary algorithm (EA), it is proved that the DQN algorithm have better convergence ability.**

I. INTRODUCTION

As an intrinsic resource of the Earth, the Earth's magnetic field can provide stable and reliable navigation information for autonomous underwater vehicles (AUV), which is currently a popular navigation resource [1]. Geomagnetic navigation has the characteristics of strong anti-interference, strong concealment, and no error accumulated over time. Compared with inertial navigation and satellite navigation, the application of geomagnetic navigation on underwater carriers has great advantages. The use of geomagnetic parameters to provide navigation information for underwater carriers can compensate for the shortcoming of accumulation of errors in inertial navigation systems over time [2]. Moreover, it can overcome the shortcomings of the rapid attenuation of wireless signals in water in currently widely used satellite navigation systems [3]. Geomagnetic navigation has become an increasingly hot research direction.

The most popular geomagnetic navigation research is geomagnetic matching navigation. They are characterized by using real-time geomagnetic maps to match the prior geomagnetic maps to obtain the current position of the carrier. In reference [4], a new geomagnetic matching feature is proposed, and the geomagnetic entropy is extracted from the total geomagnetic intensity, which is used as an auxiliary feature to improve the navigation effect. In reference [5], a geomagnetic matching method using geomagnetic vector is proposed. However, when the geomagnetic information of the matching region is lacking, it would be difficult to accurately locate.

The geomagnetic matching navigation has strong dependence on the prior geomagnetic map. It is necessary to measure the geomagnetism of the target area in advance to obtain the geomagnetic map. The accuracy of geomagnetic matching navigation is highly dependent on the accuracy, completeness and density of the prior geomagnetic data. However, the underwater environment is extremely harsh, and the effects of complex reefs and ocean currents make it difficult to establish accurate underwater geomagnetic maps. In addition, the nature of the earth's magnetic field slowly changing and abnormal geomagnetism can also cause interference. Due to these factors, obtaining accurate underwater geomagnetic maps is extremely difficult and has become a key reason for restricting underwater geomagnetic matching navigation.

In nature, it is not uncommon to navigate without a priori geomagnetic information. Geomagnetic field intensity and the inclination of geomagnetic field lines can be used as animal "magnetic compasses" to guide the direction of animal movement, and such "magnetic compasses" are widely found in amphibians, birds, mammals, and fish [6]. At certain times, turtles on both sides of the coast will migrate to the other side. Capture the young turtles on both sides of the coast and bring them to the center of the coast. The turtles still use geomagnetic information to move to the target coast [7-8]. Pigeons taken to an unfamiliar environment, after being released, use geomagnetic information to return to the nest across thousands of kilometers, accurately and quickly [9]. In the dark environment, the lobsters are sent to multiple test sites, and they will move to the habitat even if the vision is disturbed. If the strong magnets are used for interference, the lobsters can be disturbed [10].

The geomagnetic navigation phenomenon of these animals proves that this geomagnetic navigation method independent of a prior geomagnetic map can be used for navigation.

II. PROBLEM FORMULATION

* Corresponding Author

This work was supported by Science and Technology on Underwater Information and Control Laboratory (6142218051806) and the Fundamental Research Funds for the Central Universities .

Cong Wang, Yun Niu, Mingyong Liu, Tingchao Shi, Jiaqi Li and Lianggen You are with the School of Marine Engineering, Northwestern Polytechnical University, Xi'an, 710072, China,

E-mail: 0930wc@mail.nwpu.edu.cn, niuyun010121@nwpu.edu.cn, liumingyong@nwpu.edu.cn, shi_tingchao@mail.nwpu.edu.cn, lijiaqi@mail.nwpu.edu.cn, youlianggen@mail.nwpu.edu.cn.

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

In the two-dimensional space, the AUV is treated as a particle, and its equation of movement is:

$$\begin{cases} x(j+1) = x(j) + v(j)\cos\theta(j)\Delta T \\ y(j+1) = y(j) + v(j)\sin\theta(j)\Delta T \end{cases} \tag{1}$$

Where j represents the time step size, v represents the total speed of the AUV, ΔT represents the sampling period, and θ represents the heading angle of the AUV.

A. Navigation search based on geomagnetic component

The Earth's magnetic field contains multiple components, which is a mixed field, and the geomagnetic parameters [11] can be described as:

$$\mathbf{B} = \{B_1, \cdots, B_n\} \tag{2}$$

where $B_1, \cdots, B_n$ are geomagnetic components, which include the north geomagnetic field component, the east geomagnetic field component, the vertical geomagnetic field component, the total intensity, the horizontal magnetic field, the declination angle, and the inclination angle (B_X, B_Y, B_Z, B_F, B_H, B_D and B_I).

From the perspective of bionics, the changing trend of geomagnetic field has a significant influence on biological movement behavior [12]. Therefore, the process of bio-inspired geomagnetic navigation can be described as the process of convergence of multiple geomagnetic components with the AUV moving toward the target position [13], and there is a strong connection between the navigation trajectory and the geomagnetic parameters.

In the process of bio-inspired geomagnetic navigation, the AUV's navigation trajectory is guided by the convergence of multiple geomagnetic components. Therefore, the constraint relationship between the navigation trajectory and the geomagnetic parameters can be established as follows:

$$\begin{cases} \min F\left(\mathbf{B}, \boldsymbol{\theta}^j\right) = \sum_{i=1}^{n} f_i(\mathbf{B}) \\ t_i = f\left(\mathbf{B}_i^j, \mathbf{B}_i^t, \boldsymbol{\theta}^j\right) \end{cases} \tag{3}$$

Where $\mathbf{B} = \{B_1, \cdots, B_n\}$ is the geomagnetic parameter; j is the iterative step number(k=0,1,2,3...); $F\left(\mathbf{B}, \boldsymbol{\theta}^j\right)$ is the objective function; $f_i(\mathbf{B})$ is the objective function of the i-th geomagnetic component; t_i is the relationship between the movement parameter and the geomagnetic parameter; $\mathbf{B}_i^j$ is the geomagnetic parameter of the current position; $\mathbf{B}_i^t$ represents the geomagnetic parameter of the target position.

The relative position can be reflected by the difference between the geomagnetic parameters of the target position and the geomagnetic parameters of the current position of the AUV. Therefore, the squared difference between the two is used to constitute the objective function of the navigation, where the objective function of the i-th geomagnetic component is:

$$f_i^j(\mathbf{B}) = \left(\mathbf{B}_i^t - \mathbf{B}_i^j\right)^2, i \in n \tag{4}$$

The objective function is normalized to make the magnitude of the multiple geomagnetic components equal, which can be expressed as:

$$F\left(\mathbf{B}, \boldsymbol{\theta}^j\right) = \sum_{i=1}^{n} \frac{f_i^j(\mathbf{B})}{f_i^{'}(\mathbf{B})} = \sum_{i=1}^{n} \frac{\left(\mathbf{B}_i^t - \mathbf{B}_i^j\right)^2}{\left(\mathbf{B}_i^t - \mathbf{B}_i^0\right)^2} \tag{5}$$

The signification of the objective function is the relative distance between the target position and the current position, so the purpose of the geomagnetic navigation is to bring the objective function closer to zero, namely:

$$F\left(\mathbf{B}, \boldsymbol{\theta}^j\right) \to 0 \tag{6}$$

Through the above expression, the geomagnetic navigation for AUV problem without prior geomagnetic information is transformed into a multi-target search problem.

B. Reinforcement learning and Deep Q-Network algorithm

Reinforcement learning is the interaction between the agent and the surrounding environment and simultaneously produces a feedback signal that evaluates the current observation of the agent and the actions taken. If the feedback from the environment is positive, then the trend of taking the current action will increase, and vice versa. In essence, reinforcement learning is a process of improving strategies through continuous trial, with the goal of learning the strategies that bring the greatest rewards.

The Q-learning algorithm is a traditional model-free reinforcement learning algorithm, which was proposed by Watkins [14] in 1989. The Q-value is an important part of the algorithm, which refers to the value brought by the agent's taking a certain action. The key of the Q-learning algorithm is that the Q-value corresponding to each group of observations and actions of the agent is stored in the Q-table, and the agent selects the action that can obtain the maximum Q value according to the obtained Q-table. The Q-table will be updated according to the feedback reward obtained after the action is performed. The update is based on the following formula:

$$Q_{t+1}(s,a) = Q_t(s,a) + \alpha\left[r + \gamma \max_{a'}\left(Q_t\left(s', a'\right)\right) - Q_t(s,a)\right] \tag{7}$$

Where $\gamma (0 \le \gamma \le 1)$ is the discount factor, whose meaning is the degree of influence of the subsequent observation on the reward of the current observation; $\alpha (0 \leqslant \alpha \leqslant 1)$ is the learning rate; s and a are the observation of the current moment and the current action performed; s' and a' are the next observation and the next action performed; t represents the time step size.

As the number of iterations increases, Q-table will gradually converge, and the Q-value will approach the ideal value gradually. Finally, the agent gets the optimal strategy by looking for the maximum Q-value of each observation in the Q-table.

Traditional reinforcement learning algorithms can only be applied to situations where the observation space and action space dimensions are low, such as Q-Learning algorithm. However, in many practical problems, such as AUV autonomous navigation, there is often a huge observation space. In this case, Q-Learning needs to build a huge Q-table and update it according to the reward, which not only requires huge computing resources but also is impractical.

With the rapid development of deep learning, deep reinforcement learning is gradually getting hot. In 2015, Deepmind [15] proposed the Deep Q-Network (DQN) algorithm, which uses a neural network as a value function to approximate Q-table. The formula is as follows:

$$f(s,a,w) \approx Q^*(s,a) \tag{8}$$

The Q-table can be replaced by a neural network regardless of the dimension of the observation space. The trained deep Q-network takes the observations of the agent as input, obtains each action-value (Q) and selects the highest valued action. The typical architecture of DQN is shown in Fig.1.

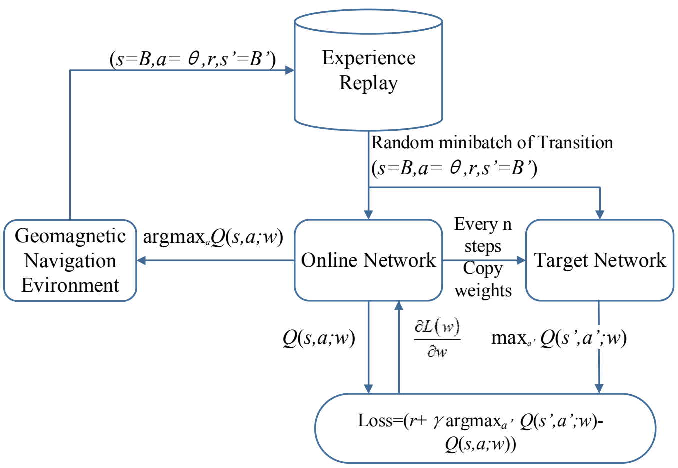

Fig. 1 DQN Network Architecture within geomagnetic navigation environment

III. Navigetion Algorithm Steps

In order to obtain the best strategy for navigation, the geomagnetic navigation algorithm based on deep reinforcement learning needs to train N rounds. The steps of each round of navigation algorithm are as follows:

Step 1: Initialize the parameters related to the DQN algorithm. The experience replay pool D, the online network parameter $\boldsymbol{w}$, and the target network parameter $\boldsymbol{w'}$ are initialized.

Step 2: Initialize the navigation for AUV related parameters. Set the geomagnetic parameters of the start position, the geomagnetic parameters of the target position, the movement step size of the AUV (L), and the threshold to reach the target position (μ). The multiple geomagnetic components of the position are used as the observation feature (s) of the AUV; the heading angle is taken as an action (a); the action space is set as:

$$\theta = \{\theta_1, \theta_2, \cdots, \theta_m\}, m = 2\pi / \Delta\theta \tag{9}$$

Step 3: Select and execute the action. Input the current observation (s) of the AUV (B_x, B_y, B_z) into the online network, and according to the epsilon-greedy strategy, select and execute the action with the largest Q value from the action space with the probability of 1-ε or randomly select and execute the action with a small probability ε, and get the next observation (s') and reward.

Step 4: Determine if the target location is reached. The normalized objective function is calculated using the next observation (multiple geomagnetic component of the current position) obtained in step 3, see equation (5). If the objective function is less than the threshold μ, that is,

$$F\left(\mathbf{B}, \boldsymbol{\theta}^j\right) < \mu \tag{10}$$

then the AUV is considered to reach the target position, and the current round of training ends, otherwise step 5 is entered.

Step 5: Save the sample. Save the sample to the experience replay pool in the form of (s, a, r, s').

Step 6: Extract the sample. m samples are randomly taken from the experience replay pool.

Step 7: Calculate the online Q-value and the target Q-value. The current observation (s) of the extracted m samples and the execution action are input to the online network, and action-value ($Q(s, a; w)$) is calculated; the next observation (s') is input to the target network, and the target Q-value can be calculated by the following formula:

$$\text{Target } Q = r + \gamma \max_{a'} Q\left(s', a'; w'\right) \tag{11}$$

Step 8: Update the network parameters. The loss function of the DQN algorithm is:

$$Loss = (\text{Target } Q - Q(s,a;w))^2 \tag{12}$$

The parameters of the online network are updated by the mini-batch gradient descent method, and the online network copies its parameters to the target network every n iteration steps. Go back to step 3.

Finally, after N rounds of training, the number of steps in which the AUV navigates to the target position gradually converges, and the online network at this time is used as the optimal strategy for heading decision.

IV. Simulation Results

A. Simulation setup

So as to verify the validity of the algorithm, we simulated on the basis of the International Geomagnetic Reference Field 2010 (IGRF2010), which provides the required geomagnetic components. In the simulation, three independent parameters are selected from the multiple geomagnetic components as the navigation feature and observation feature, which are the geomagnetic north component (B_x), the geomagnetic east component (B_y), and the geomagnetic vertical component (B_z).

The simulation area is shown in Fig.2. A square area from (34N, 108E) to (35N, 109E) is selected. The starting position is located at (34.16N, 108.54E), as shown by the circle, and its three geomagnetic components are: B_x=25198nT,

B_y=-1241nT, B_z=31272nT. The target position represented by a triangle is located at (34.55N, 108.35E), and its three geomagnetic parameters are: B_x=25035nT, B_y=-1238nT, B_z=31614nT.

The reward function plays an important role in reinforcement learning. It evaluates the actions performed by the AUV and has a guiding effect on the selection of actions. The reward function of the DQN algorithm in the simulation is set as follows:

$$r=\begin{cases} 10, & F(k)\le\mu \\ -10, & k>max\_step \\ -\left(F(k)-F(k-1)\right), & others \end{cases} \tag{13}$$

Where k represents the current iteration step of the AUV navigation; μ is the threshold for determining the arrival of the target position; $F(k)$ represents the objective function value of the current position; max_step is the maximum number of steps allowed for the iteration.

Under the above dynamic environment and reward function, the N-round training is performed using the DQN algorithm, and finally the navigation iteration step and the objective function converge.

The parameter values set in the simulation experiment are given in Table I.

Table I
Simulation Parameters

Parameters	Size
L	100m
$\Delta\theta$	30°
μ	0.005
max_step	20000
ε	0.1
N	500

B. Simulation result and analysis

In order to prove the effectiveness of this algorithm, the DQN algorithm is compared with the evolutionary algorithm proposed by reference [16]. The following is the simulation result.

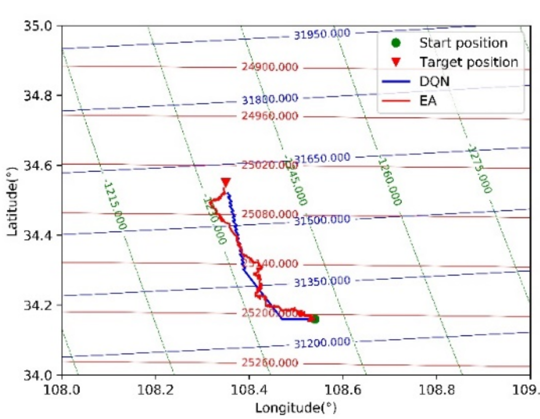

Fig. 2 AUV navigation trajectory under two algorithms

Fig. 2 shows the navigation trajectory of the AUV under the two algorithms. Both algorithms can navigate to the target position without a priori geomagnetic information, but the trajectories of the two navigation algorithms are quite different. The trajectory under the evolutionary algorithm is rather tortuous, while the trajectory is smoother under the DQN algorithm.

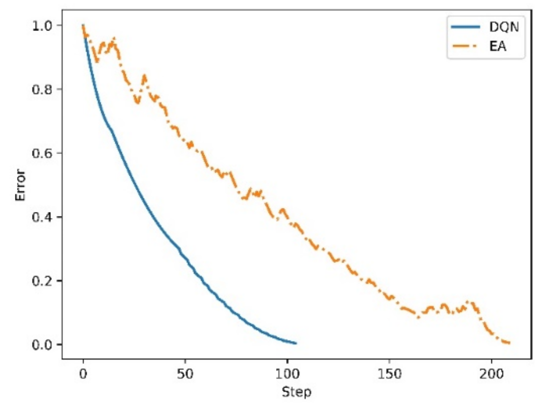

Fig. 3 Convergence curves of the objective functions of the two algorithms

It can be seen from Fig. 3 that both navigation algorithms can converge the objective function to the threshold μ, but the number of iteration steps required is different. The number of iteration steps required by the DQN algorithm is 102 steps, and the number of iteration steps required by the evolutionary algorithm is 210 steps. Therefore, the DQN algorithm has a shorter search trajectory than the evolutionary algorithm and has higher navigation efficiency.

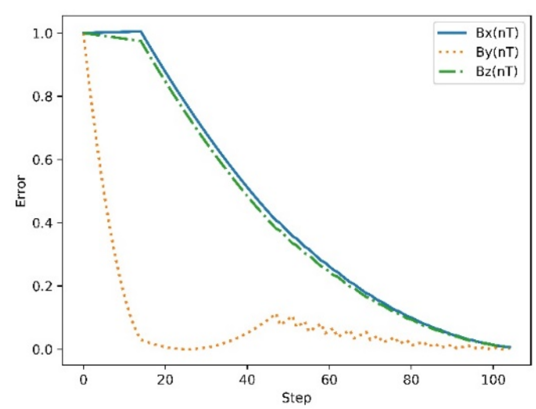

Fig. 4 Convergence curve of three geomagnetic components under the DQN algorithm

The convergence curves of the three geomagnetic components under the DQN algorithm and the evolutionary algorithm are shown by Fig. 4 and Fig. 5 respectively. From the figures, it can be seen that the convergence curves of the geomagnetic components under the evolutionary algorithm are more oscillating, while the convergence curves under the DQN algorithm are smoother and less oscillating. Therefore, after comparison, the DQN algorithm has better convergence ability than the evolutionary algorithm.

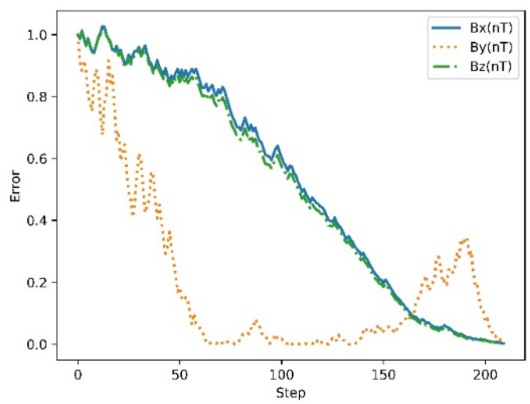

Fig. 5 Convergence curve of three geomagnetic components under evolutionary algorithm

The two algorithms take 50 simulation results each and count the average number of iteration steps. As shown in Fig.6, the horizontal axis represents two algorithms, the vertical axis represents the number of navigation iteration steps under the two algorithms, and the triangle in the box diagram represents the average number of iteration steps of 50 times. As can be seen from the figure, the DQN algorithm has a shorter average number of search steps than the evolutionary algorithm without a priori geomagnetic information.

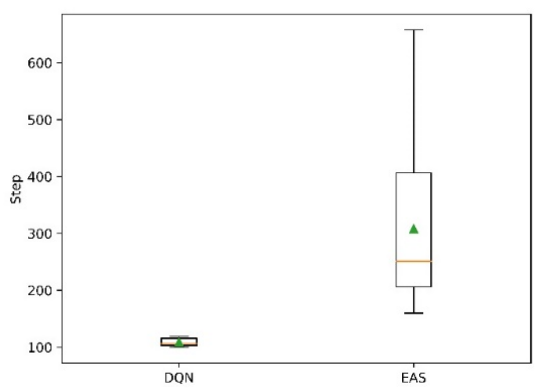

Fig. 6 Statistics of the number of iteration steps under the two algorithms

VI. Conclusion

For the problem of geomagnetic navigation for AUV without prior geomagnetic information, a geomagnetic navigation algorithm based on deep reinforcement learning is proposed. Select the discrete heading angle as the action space, take the multiple geomagnetic components as the observation features and input them into the DQN, use the online network of DQN to make the heading decision, and judge whether the AUV reaches the target position by the normalized objective function. The simulation results show that the DQN algorithm can make the AUV reach the target position without a priori information. In addition, compared with the evolutionary algorithm, the DQN algorithm has a shorter average iteration step and has better search performance.

Acknowledgment

This work was supported by Science and Technology on Underwater Information and Control Laboratory（6142218051806）and the Fundamental Research Funds for the Central Universities .

References

[1] Goldenberg, Felix. "Geomagnetic navigation beyond the magnetic compass." *2006 IEEE/ION Position, Location, And Navigation Symposium*. IEEE, 2006.
[2] Jun-wei, W. U., K. M. Tsang, and N. I. E. Li-juan. "Estimation of the INS's Errors [J]." *Journal of Chinese Inertial Technology* 6.1 (2002).
[3] Steiner, Ladina, Michael Meindl, and Alain Geiger. "Characteristics and limitations of GPS L1 observations from submerged antennas." *Journal of Geodesy* 93.2 (2019): 267-280.
[4] Guo, Caifa, Hong Cai, and G. H. M. van der Heijden. "Feature extraction and geomagnetic matching." *The Journal of Navigation* 66.6 (2013): 799-811.
[5] Song, Zhongguo, et al. "The vector matching method in geomagnetic aiding navigation." *Sensors* 16.7 (2016): 1120.
[6] Lohmann, Kenneth J. "Q&A: Animal behaviour: Magnetic-field perception." *Nature* 464.7292 (2010): 1140.
[7] Lohmann, Kenneth J., et al. "Animal behaviour: geomagnetic map used in sea-turtle navigation." *Nature* 428.6986 (2004): 909.
[8] Luschi, Paolo, Graeme C. Hays, and Floriano Papi. "A review of long□distance movements by marine turtles, and the possible role of ocean currents." *Oikos* 103.2 (2003): 293-302.
[9] Wiltschko, Roswitha, Ingo Schiffner, and Wolfgang Wiltschko. "A strong magnetic anomaly affects pigeon navigation." *Journal of Experimental Biology* 212.18 (2009): 2983-2990.
[10] Boles, Larry C., and Kenneth J. Lohmann. "True navigation and magnetic maps in spiny lobsters." *Nature* 421.6918 (2003): 60.
[11] Tyren, Carl. "Magnetic anomalies as a reference for ground-speed and map-matching navigation." *The Journal of navigation* 35.2 (1982): 242-254.
[12] Liu, Mingyong, Hong Li, and Kun Liu. "Geomagnetic navigation of AUV without a priori magnetic map." *OCEANS 2014-TAIPEI*. IEEE, 2014.
[13] Liu, Mingyong, et al. "Bio-inspired navigation based on geomagnetic." *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2013.
[14] Watkins, Christopher JCH, and Peter Dayan. "Q-learning." *Machine learning* 8.3-4 (1992): 279-292.
[15] Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." *Nature* 518.7540 (2015): 529.
[16] Li, Hong, Mingyong Liu, and Kun Liu. "Bio-inspired geomagnetic navigation method for autonomous underwater vehicle." *Journal of Systems Engineering and Electronics* 28.6 (2017): 1203-1209.