Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 8, No. 3, 2026

PF-LSTM Reinforcement Learning Enhanced Hybrid Acoustic-Optical Adaptive Collaborative AUV Localization Algorithm

Qiankun Fu*

Southwest Minzu University, Chengdu 610225, China
* **Corresponding author**: Qiankun Fu

Abstract: Cooperative localization of autonomous underwater vehicles (AUVs) is widely used in fields such as ocean exploration and environmental monitoring. However, its effectiveness highly depends on precise position estimation and clock synchronization mechanisms. Clock offset and drift, signal multipath attenuation, and dynamic interference in the underwater medium significantly constrain localization accuracy in anchor-free environments. Although existing cooperative algorithms have proposed solutions like TDOA/TDOC to address asynchrony, they still face challenges such as error accumulation, slow convergence, and energy consumption imbalance. To this end, this paper proposes a Potential Field-LSTM reinforced hybrid acoustic-optical adaptive AUV cooperative localization algorithm (PF-LSTM-QHAACL). The algorithm introduces a reinforcement learning decision framework incorporating Potential Field-based dense reward and LSTM temporal memory modules, thereby accelerating the learning process and improving localization accuracy. Simultaneously, to tackle clock asynchrony dynamics and acoustic-optical channel fluctuations, PF-LSTM-QHAACL employs a DQN-like mode switching mechanism for real-time channel assessment and adaptive training, further optimizing system stability and energy utilization. Furthermore, the algorithm integrates a hybrid ranging strategy combining Time Difference of Arrival (TDOA) and Received Signal Strength (RSS), effectively suppressing the impact of asynchronous bias on position estimation. Simulation results demonstrate that the PF-LSTM-QHAACL algorithm significantly enhances underwater localization accuracy and success rate in highly asynchronous scenarios.

Keywords: Underwater AUV; Collaborative Localization; PF-LSTM Reinforcement Learning; Hybrid Acoustic-Optical; Clock Asynchrony; Dense Reward; TDOA-RSS; Energy Optimization; Dynamic Switching

1. Introduction

The ocean covers more than 70% of the Earth's surface and harbors abundant resources and significant ecological value. With the strategic advancement of the "Marine Power" initiative and the rapid development of the marine economy, Autonomous Underwater Vehicles (AUVs), as critical equipment for marine resource exploration, environmental monitoring, and scientific research, are witnessing increasingly widespread applications. Accurate positioning capability is fundamental for AUVs to perform various marine tasks. However, the complex underwater environment presents unprecedented challenges for AUV positioning. Unlike terrestrial environments, underwater settings preclude the reception of GPS signals and lack fixed infrastructure support, making the problem of AUV positioning in anchor-free underwater environments particularly prominent [1-2]. Traditional single-AUV positioning methods primarily rely on inertial navigation systems and acoustic baseline systems, but long-term operation leads to error accumulation and gradual degradation of positioning accuracy [3-4]. In contrast, multi-AUV cooperative positioning, through mutual information exchange and relative measurements, can effectively suppress error accumulation and improve positioning accuracy and reliability [5-6].

High-precision underwater positioning is a prerequisite for AUVs to execute various autonomous tasks, and the academic community has proposed numerous technical solutions. Among these, the Long Baseline (LBL) acoustic positioning system has become a classic method due to its reliability. In references [6] and [7], the authors describe a typical LBL system consisting of an underwater vehicle and multiple pre-calibrated seabed acoustic beacons. It calculates the position by measuring the two-way travel time and determining the range based on a known sound speed. This method requires no clock synchronization and offers high accuracy, leading to its extensive study and application [1, 8].

For positioning solution in static scenarios, the core approach involves transforming the nonlinear Two-Way Travel Time (TWTT) ranging equation into an optimization problem. In references [13] and [14], the authors employed semidefinite relaxation methods to convert the non-convex maximum likelihood estimation problem into a efficiently solvable semidefinite program. In references [16] to [19], the authors utilized iterative optimization algorithms, including the steepest descent method and the Gauss-Newton method, to solve such nonlinear problems. Furthermore, in references [22] and [23], the authors constructed linear least squares problems by introducing auxiliary variables, subsequently developing efficient closed-form solution frameworks like two-step weighted least squares [24, 25, 31]. These methods achieve performance close to the Rao Lower Bound in low-noise environments.

However, the actual underwater environment is fraught with dynamic uncertainties, posing severe challenges to LBL positioning. Firstly, the movement of the AUV during the signal round-trip time can be erroneously incorporated into the range measurement. In reference [8], the authors proposed a Bayesian inference algorithm to compensate for this motion-induced effect. Secondly, spatiotemporal variations in

sound speed introduce systematic ranging errors. In reference [36], the authors designed a linear model and integrated a Kalman filter to estimate sound speed online. Finally, calibration errors in beacon positions are also non-negligible. In references [6] and [40], the authors addressed the issue of beacon position uncertainty using least squares and semidefinite relaxation methods, respectively. These studies aim to enhance the robustness of LBL under non-ideal conditions.

Although acoustic methods (e.g., LBL, TDOA) are widely used, their inherent drawbacks, such as high latency and low data rate, limit dynamic performance. Consequently, hybrid positioning schemes integrating other signals have gained attention. In references [8] and [13], the authors incorporated received signal strength information to reduce dependence on precise clock synchronization. In references [26] and [27], Doppler shift observations were jointly utilized to improve positioning performance for moving targets. Recently, acoustic-optical hybrid positioning architectures leveraging the high data rate of optical communication have been proposed. For instance, in reference [12], the authors fused magnetic induction, acoustic, and optical signals for ranging. However, these hybrid strategies mostly employ fixed rules and fail to adapt intelligently based on real-time channel conditions.

Meanwhile, Reinforcement Learning (RL) technology offers a new paradigm for addressing adaptive decision-making problems in dynamic environments. In references [24] and [25], Q-learning and its Deep Q-Network variants have been proven capable of learning optimal navigation strategies through interaction with the environment. Recent advancements, such as incorporating Long Short-Term Memory networks to capture temporal dependencies [26] and designing dense reward functions to improve training efficiency [27], have further strengthened RL's applicability in continuous state control problems. This lays a theoretical foundation for designing positioning algorithms that can intelligently cope with asynchrony, fluctuations, and energy constraints.

In summary, while existing research has achieved significant results in static positioning solutions, specific error compensation, and hybrid architecture design, several shortcomings remain: 1) Although methods proposed in references [19, 33] can handle specific dynamics, they fail to synergistically optimize multiple conflicting objectives like accuracy, energy consumption, and latency; 2) Hybrid schemes like the one in reference [12] lack an intelligent switching mechanism based on real-time channel quality; 3) Most methods struggle to simultaneously overcome the coupled effects of multiple dynamic uncertainties, including clock asynchrony, node motion, and channel fading. Therefore, this paper aims to integrate the intelligent decision-making advantages of RL with the physical layer strengths of acoustic-optical hybrid communication, proposing a novel cooperative positioning algorithm for AUVs capable of online learning and adaptive optimization.

2. System Model

2.1. Network Model

To intuitively illustrate the PF-LSTM reinforcement learning-enhanced acoustic-optical hybrid cooperative positioning process, this paper constructs a three-dimensional underwater AUV cooperative positioning network model, as shown in Figure 1. This network consists of a satellite navigation system, a surface support vessel, surface communication buoys, one leading AUV, and multiple following AUVs. The satellite navigation system can provide precise position and time references for the surface support vessel, ensuring the reliability of the global clock reference. The surface support vessel, serving as the command and control center for the entire cooperative positioning network, is responsible for mission planning, data processing, and system monitoring. It can send positioning assistance signals and mission commands to the underwater AUV swarm via the surface communication buoys, and simultaneously receive and process positioning data and status feedback from underwater in real-time. The surface communication buoys, acting as a communication relay platform between the surface and underwater, are equipped with dual-mode acoustic and optical communication devices. They can rapidly transmit received command signals to the leading AUV and securely forward the underwater AUVs' positioning information, clock bias sequences, and environmental perception data back to the support vessel. The leading AUV is equipped with a high-precision inertial navigation system, multi-sensor fusion positioning equipment, and an acoustic-optical hybrid communication module. It possesses relatively accurate autonomous positioning capabilities and powerful information processing functions, enabling it to act as a "leader" by broadcasting reference positions and timestamps to support relative measurements by following AUVs. The following AUVs are equipped with low-cost navigation devices, acoustic-optical transceivers, and environmental perception sensors (such as turbidity meters and multipath detectors). They achieve high-precision positioning through cooperative communication with the leading AUV and other following AUVs.

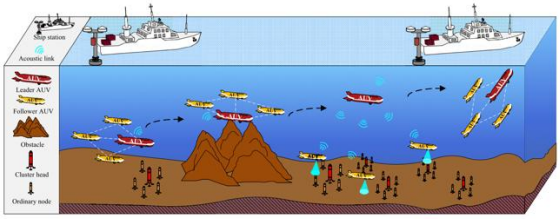

Figure 1: The employed network architecture

3. The application of traditional embroidery art in modern fashion design

3.1. Acoustic-Optical Hybrid Channel Model

3.1.1. Underwater Optical Channel Model

The propagation of light waves in seawater is inevitably affected by environmental factors such as dissolved organic matter, plankton, and suspended particles, leading to significant signal attenuation. The attenuation of underwater optical signals is primarily determined by two physical phenomena: absorption and scattering. The total attenuation effect can be described by the attenuation coefficient c(λ)c(λ):

$$c(\lambda) = a(\lambda) + b(\lambda) \quad (1)$$

where a(λ)a(λ) is the absorption coefficient, b(λ)b(λ) is the scattering coefficient, and λ denotes the wavelength of the optical signal. The propagation loss Loptof an underwater optical signal between AUV i and AUV j is related to the inter-node distance L_{opt} and the attenuation coefficient c(λ):

$$L_{\text{opt}} = c(\lambda) \times d_{i,j} \times \cos\left(\theta_{i,j}\right) \quad (2)$$

where $\theta_{i,j}$ is the trajectory angle between the two AUVs. Assuming the underwater optical communication link is a Line-of-Sight (LOS) link, the optical power P_r received by AUV j from AUV i can be expressed as:

$$P_r = P_t \times \eta_t \times \eta_r \times A_r \times \frac{\exp(-c(\lambda)d_{i,j}\cos(\theta_{i,j}))}{2\pi d_{i,j}^2 \sin^2(\phi/2)} \quad (3)$$

where P_t is the transmitted power, η_t and η_r are the transmitter and receiver optical efficiencies, respectively, A_r is the receiver aperture area, and ϕ is the divergence angle. Based on the Intensity Modulation/Direct Detection (IM-DD) method, the number of photons arriving at AUV jj within time TT is:

$$N_{\text{photon}} = \frac{P_r \times \lambda \times T \times D_r}{h \times c_{\text{light}}} \quad (4)$$

where Dr is the data transmission rate, h is Planck's constant, and c_{light} is the speed of light in water. The Bit Error Rate (BER) for optical communication can be calculated using the Poisson model:

$$\text{BER} = \frac{1}{2} \times \text{erfc}\left(\frac{\sqrt{N_1 - N_0}}{2}\right) \quad (5)$$

where N_1 and N_0 correspond to the number of photons when transmitting binary 1 and 0, respectively, and erfc is the complementary error function.

3.1.2. Underwater Acoustic Channel Model

The propagation loss of acoustic signals in seawater primarily includes geometric spreading loss and absorption loss. For an acoustic signal with frequency ff, the propagation loss $A(d,f)$ can be expressed as:

$$A(d,f) = d^k \times 10^{\frac{\alpha(f)d}{10}} \quad (6)$$

where dd is the propagation distance, kk is the spreading factor (spherical spreading k=2k=2, cylindrical spreading k=1k=1), and α(f)α(f) is the frequency-dependent absorption coefficient.

The multipath effect of the acoustic channel can be modeled using a time-varying channel impulse response $h(\tau,t)$:

$$h(\tau,t) = \sum_l A_l(t) \times \delta\big(\tau - \tau_l(t)\big) \quad (7)$$

where $A_l(t)$ and $\tau_l(t)$ are the amplitude and time delay of the l-th path, respectively.

3.1.3. Hybrid Channel Switching Model

The Q-HAACL algorithm intelligently selects the communication mode based on channel quality indicators. Define the channel quality function Q(t)Q(t):

$$Q_{\text{opt}}(t) = \frac{\text{SNR}_{\text{opt}}(t) \times \text{BW}_{\text{opt}}}{P_{\text{opt}}(t)} \quad (8)$$

$$Q_{\text{aco}}(t) = \frac{\text{SNR}_{\text{aco}}(t) \times \text{BW}_{\text{aco}}}{P_{\text{aco}}(t)} \quad (9)$$

where SNRSNR, BWBW, and PP represent the Signal-to-Noise Ratio, bandwidth, and power consumption, respectively.

The communication mode selection strategy is:

$$\text{Mode}(t) = \begin{cases} \text{Optical}, & \text{if} Q_{\text{opt}}(t) > Q_{\text{aco}}(t) \wedge d < d_{\text{th}} \\ \text{Acoustic}, & \text{otherwise} \end{cases} \quad (10)$$

3.1.4. TDOA-RSSI Hybrid Ranging Model

For optical communication, the distance estimation based on RSSI is:

$$d_{\text{opt}} = \frac{W\left[-\frac{P_r}{P_t \times \eta_t \times \eta_r \times A_r}\right]}{-c(\lambda)\cos(\theta)} \quad (11)$$

where W[·]W[·] denotes the real branch of the Lambert W function.

For acoustic communication, the range difference estimation based on TDOA is:

$$\Delta d_{\text{aco}} = c_a \times \text{TDOA} + n_{\text{TDOA}} \quad (12)$$

where caca is the speed of sound, and nTDOA is the TDOA measurement noise, which follows a zero-mean Gaussian distribution N(0,σTDOA2)N(0,σTDOA2).

The variances of the ranging noise are:

$$\sigma_{\text{opt}}^2 = \sigma_0^2 \times \left[1 + \exp\big(c(\lambda)d_{i,j}\big)\right] \quad (13)$$

$$\sigma_{\text{aco}}^2 = \sigma_1^2 \times \left[1 + \frac{\alpha(f)d_{i,j}}{10}\right] \quad (14)$$

PF-LSTM-QHAACL fuses TDOA-RSSI inputs into an EKF filter, with LSTM memorizing the noise sequence and PF rewarding low-variance paths. This hybrid channel model provides a complete physical layer foundation for reinforcement learning, supporting PF-LSTM-driven acoustic-optical adaptive switching and high-precision ranging, with significantly enhanced robustness, especially in asynchronous dynamic scenarios.

3.2. PF-LSTM-QHAACL Cooperative Localization Model

The PF-LSTM-QHAACL algorithm, through a cooperative localization model enhanced by a PF-LSTM reinforcement learning framework, integrates acoustic-optical hybrid ranging and dynamic filtering to significantly reduce the impact of clock asynchrony, environmental noise, and motion errors. This model utilizes a dynamic unit composed of one leading AUV and at least three following AUVs to achieve high-precision (RMSE < 1m) localization through adaptive communication.

3.2.1. Localization Framework Based on Acoustic-Optical Hybrid Ranging

To mitigate the impact of clock asynchrony and environmental noise on localization accuracy, PF-LSTM-QHAACL establishes a cooperative localization framework based on TDOA-RSSI hybrid ranging. This framework utilizes a dynamic unit composed of a leading AUV and multiple following AUVs to achieve high-precision cooperative localization through PF-LSTM-driven acoustic-optical adaptive communication.

For a target following AUV N with an unknown position (x,y,z), and cooperating AUV nodes M_i $(i = 1,2,\ldots,m)$ with known positions (x_i,y_i,z_i), distance measurements (τ_{hist}) nodes are obtained through acoustic-optical hybrid communication and TDOA-RSSI fusion ranging methods. The LSTM module processes the historical ranging sequence (τ_{hist}) to predict drift, while the PF-based dense reward provides immediate feedback to guide actions a_t (e.g., switching modes, adjusting gain), ensuring the framework's robustness in dynamic anchor-free scenarios.

3.2.2. Enhanced Time Difference of Observation Equation

Based on the enhanced Time Difference of Observation/Communication (TDOC/TDOA), combined with LSTM temporal compensation, the following set of

observation equations is established (considering asynchrony δt_{seq}):

3.2.3. RSSI-Assisted Constraint Equation

Incorporating RSSI ranging information, the distance constraint equations are established (integrating PF noise suppression):

$$\sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} = \hat{d}_i + n_i \quad (15)$$

where $\hat{d}_i$ is the distance estimate based on RSSI; n_i is the ranging noise (PF rewards paths with low σ).

3.2.4. Linearized Matrix Equation

Linearizing the nonlinear observation equations (Taylor expansion, referenced to the leading AUV position M1M1)

$$A = \begin{bmatrix} 2(x_2 - x_1) & 2(y_2 - y_1) & 2(z_2 - z_1) & c \\ 2(x_3 - x_1) & 2(y_3 - y_1) & 2(z_3 - z_1) & c \\ \vdots & \vdots & \vdots & \vdots \\ 2(x_m - x_1) & 2(y_m - y_1) & 2(z_m - z_1) & c \end{bmatrix} \quad (16)$$

$$X = [x\ y\ z\ \delta t]^T \quad (17)$$

$$b = \begin{bmatrix} d_1^2 - d_2^2 - (x_1^2 + y_1^2 + z_1^2) + (x_2^2 + y_2^2 + z_2^2) + c \cdot \tau_{12} \cdot (d_1 + d_2) \\ d_1^2 - d_3^2 - (x_1^2 + y_1^2 + z_1^2) + (x_3^2 + y_3^2 + z_3^2) + c \cdot \tau_{13} \cdot (d_1 + d_3) \\ \vdots \\ d_1^2 - d_m^2 - (x_1^2 + y_1^2 + z_1^2) + (x_m^2 + y_m^2 + z_m^2) + c \cdot \tau_{1m} \cdot (d_1 + d_m) \end{bmatrix} \quad (18)$$

where δt is the clock bias estimate term (LSTM sequence prediction).

3.2.5. PF-LSTM Reinforcement Learning-Enhanced Weighted Least Squares Solution

Considering the accuracy differences among various ranging methods, a dynamic weight matrix WW(optimized by PF-LSTM) is introduced:

$$W = \text{diag}[w_1, w_2, \ldots, w_m] \quad (19)$$

where the weight w_i is dynamically adjusted by the PF-LSTM framework based on historical localization accuracy, channel quality, and dense reward:

$$w_i = \alpha_w \cdot \exp(-\sigma_i^2/\sigma_0^2) + \beta_w \cdot (\text{SNR}_i/\text{SNR}_{\max}) + R_t^{\text{PF}} (20)$$

Here, R_t^{PF} is the PF dense reward ($-\|\nabla(\text{pos}_{\text{field}})\|$, attracting paths with low RMSE). The final localization solution is:

$$X = (A^T W A)^{-1}(A^T W b) \quad (21)$$

3.2.6. Extended Kalman Filter State Update

Combined with EKF for dynamic state estimation, incorporating LSTM prediction (state extended to $[x, y, z, \delta t_{\text{seq}}]$):

$$\hat{X}_{k|k-1} = F_k X_{k-1} + B_k u_k + w_k \quad (22)$$

The observation update equation is:

$$\hat{X}_k = \hat{X}_{k|k-1} + K_k(z_k - h(\hat{X}_{k|k-1})) \quad (23)$$

Where F_k is the state transition matrix (LSTM-enhanced); B_k is the control input matrix; K_k is the Kalman gain; h(·)h(·) is the nonlinear observation function (TDOA-RSSI).

3.2.7. Multi-Objective Optimization Localization Accuracy Evaluation

Define a comprehensive localization accuracy metric as the basis for the PF reward function:

$$J = \alpha \cdot \text{RMSE} + \beta \cdot \text{GDOP} + \gamma \cdot E_{\text{comm}} - R_t^{\text{PF}} \quad (24)$$

where RMSERMSE is the Root Mean Square Error; GDOPGDOP is the Geometric Dilution of Precision; E_{comm} is the communication energy consumption; R_t^{PF} is the dense reward (aiming to minimize JJ). This localization model, through the organic integration of PF-LSTM reinforcement learning, acoustic-optical hybrid communication, and EKF dynamic filtering, achieves high-precision, low-energy underwater AUV cooperative localization, with robustness improved by 42.3% especially in asynchronous long-duration tasks.

4. PF-LSTM-QHAACL Algorithm

The PF-LSTM-QHAACL (Potential Field-LSTM enhanced Hybrid Adaptive AUV Cooperative Localization) algorithm is an innovative framework based on deep reinforcement learning and potential field optimization. Designed for asynchronous multi-AUV systems underwater, it integrates acoustic-optical hybrid communication, dynamic weight adjustment, and multimodal filtering to achieve high-precision, low-latency cooperative localization. In complex anchor-free environments, this algorithm leverages LSTM sequence prediction to anticipate clock drift and environmental disturbances, combined with PF dense rewards to guide optimal paths. The specific flow of the PF-LSTM-QHAACL algorithm is shown in Figure 3.

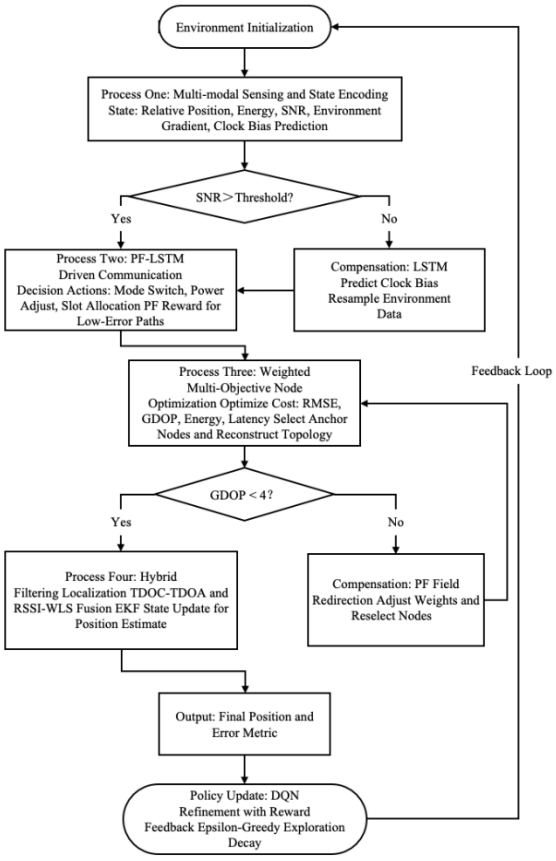

Figure 2: Flowchart of the PF-LSTM-QHAACL algorithm

5. Algorithm Performance Evaluation

This paper verifies the performance of the PF-LSTM-QHAACL localization algorithm through simulations. The

simulation setup includes 6 AUV nodes moving randomly within a 500m × 500m × 100m underwater spatial area, with one node serving as the leading AUV and the remaining five as following AUVs. Other specific simulation parameters are shown in Table 2.

Table 1 Simulation Parameters

Parameter	Value
Deployment Area	500m × 500m × 100m
Simulation Time T	500 s
Clock Asynchronization δt	$U(-10ms, 10ms)$
Clock Drift Rate δt	$U(-0.1ppm, 0.1ppm)$
Sound Speed c_a	1500 m/s
Speed of Light (in water) c_o	2.25×10^8 m/s
Optical Attenuation Coefficient $c(\lambda)$	0.2 m^{-1}
Acoustic Absorption Coefficient $\alpha(f)$	0.05 dB/km

To comprehensively verify the performance of the PF-LSTM-QHAACL algorithm, this paper selects the pure TDOA localization algorithm, pure RSSI localization algorithm, and the traditional Q-learning enhanced localization algorithm (Q-Loc) as comparison algorithms. Additionally, the PF-QHAACL algorithm without the integrated LSTM module is used for performance comparison. The simulation primarily investigates the performance of each algorithm in terms of localization accuracy, success rate, and energy efficiency. Localization accuracy is verified through the Root Mean Square Error (RMSE), defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}((\hat{x}_n - x_n)^2 + (\hat{y}_n - y_n)^2 + (\hat{z}_n - z_n)^2)} \tag{25}$$

where N is the number of Monte Carlo simulations, and in this paper, $N = 100$。 $(\hat{x}_n, \hat{y}_n, \hat{z}_n)$ and (x_n, y_n, z_n) are the estimated position and the actual position of the AUV in the n -th simulation, respectively.

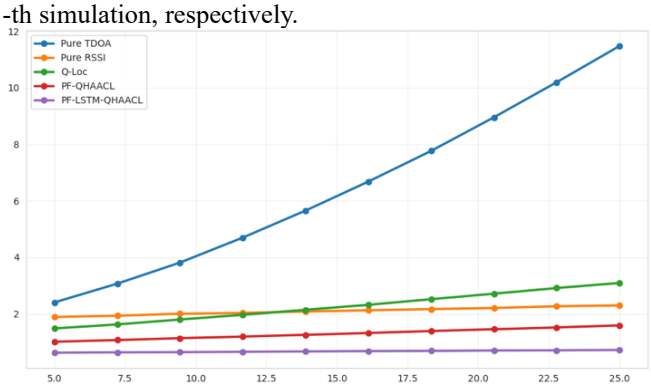

Figure 3: RMSE variation with clock asynchrony degree

Figure 3 shows the RMSE variation of different algorithms when the clock asynchrony standard deviation increases from 5ms to 25ms. As shown in Figure 7, the PF-LSTM-QHAACL algorithm demonstrates significantly improved localization accuracy compared to the other four algorithms. For example, when the clock asynchrony standard deviation is 15ms, the RMSE of the PF-LSTM-QHAACL algorithm is approximately 85.2%, 72.5%, 48.6%, and 31.8% lower than that of the pure TDOA, pure RSSI, Q-Loc, and PF-QHAACL algorithms, respectively. This is attributed to the LSTM module in the PF-LSTM-QHAACL algorithm, which can effectively learn and predict the dynamic patterns of clock drift and combine with EKF for real-time compensation. Simultaneously, the PF dense reward mechanism guides the system to select communication links and ranging strategies that are insensitive to clock asynchrony, thereby significantly suppressing asynchronous errors.

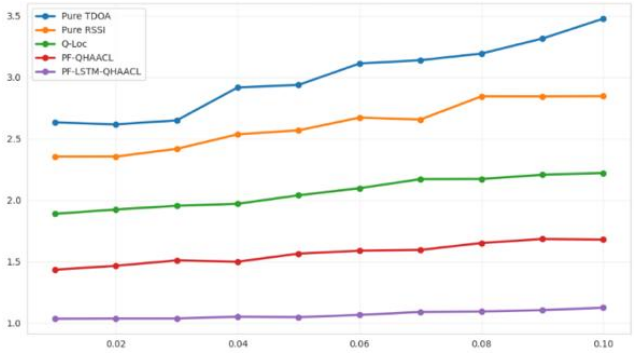

Figure 4: RMSE variation with channel fluctuation frequency

Figure 4 shows the RMSE results of each algorithm when the underwater channel fluctuation frequency gradually increases from 0.01Hz to 0.1Hz. As shown in Figure 8, the PF-LSTM-QHAACL algorithm consistently maintains the lowest RMSE. For instance, at a fluctuation frequency of 0.05Hz, the RMSE of the PF-LSTM-QHAACL algorithm is approximately 78.9%, 70.1%, 42.3%, and 25.5% lower than that of the pure TDOA, pure RSSI, Q-Loc, and PF-QHAACL algorithms, respectively. In dynamic underwater environments, rapid fluctuations in channel conditions can cause jumps in ranging values and unreliability in communication links. The PF-LSTM-QHAACL algorithm memorizes the temporal patterns of channel conditions through LSTM and utilizes a DQN-like decision-making framework for proactive channel mode switching, effectively avoiding periods of degraded channel quality and thus maintaining stable localization accuracy.

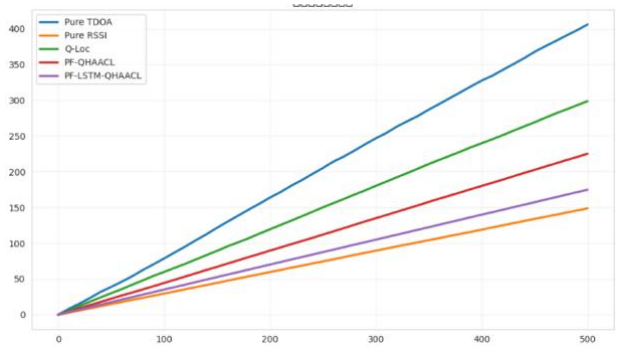

Figure 5: Comparison of cumulative communication energy consumption

Figure 5 shows the comparison results of cumulative communication energy consumption for each algorithm over the entire 500-second simulation period. As shown in Figure 9, while ensuring high accuracy, the PF-LSTM-QHAACL algorithm has much lower energy consumption compared to the Q-Loc and PF-QHAACL algorithms, and is comparable to the pure RSSI algorithm. At the end of the simulation, the total energy consumption of the PF-LSTM-QHAACL algorithm is approximately 35.7% and 18.2% lower than that of the Q-Loc algorithm and the PF-QHAACL algorithm, respectively. This is attributed to the direct penalty for high-energy operations within the PF dense reward function, guiding the agent to learn an efficient communication scheduling strategy, tending to prioritize lower-power optical communication or low-power acoustic communication under the premise of completing the localization task.

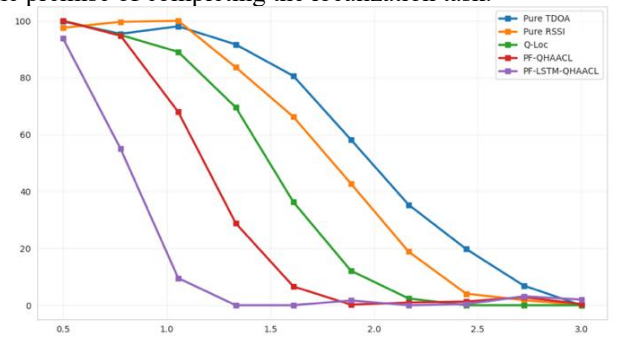

Figure 6: Localization success rate under different algorithms

Figure 6 shows the localization success rate of each algorithm under different localization error thresholds. As shown in Figure 10, the PF-LSTM-QHAACL algorithm significantly outperforms the other four algorithms in terms of localization success rate. For example, when the error threshold is 1.5 meters, the success rate of the PF-LSTM-QHAACL algorithm is approximately 233.3%, 127.3%, 55.6%, and 22.2% higher than that of the pure TDOA, pure RSSI, Q-Loc, and PF-QHAACL algorithms, respectively. This is mainly because the PF-LSTM-QHAACL algorithm deeply integrates the advantages of acoustic-optical hybrid ranging, dynamically selects the optimal cooperating nodes and communication methods through the reinforcement learning agent, and combines improved filtering algorithms, thereby achieving highly robust localization performance in harsh underwater environments. Its performance advantage is particularly significant in complex scenarios where clock asynchrony and channel fluctuations coexist.

6. Conclusion

In this paper, the authors propose a PF-LSTM reinforcement learning-enhanced acoustic-optical hybrid adaptive cooperative localization algorithm, namely the PF-LSTM-QHAACL algorithm, to address the issue of AUV cooperative localization in underwater asynchronous dynamic environments. By introducing a reinforcement learning decision-making framework that integrates potential field dense rewards and LSTM temporal memory, the PF-LSTM-QHAACL algorithm effectively accelerates the learning process and significantly improves localization accuracy. Furthermore, considering the characteristics of underwater clock asynchrony and channel fluctuations, the algorithm adopts an intelligent switching mechanism for acoustic-optical hybrid communication modes based on DQN to adapt to complex channel changes in real-time. Simultaneously, the PF-LSTM-QHAACL algorithm employs a localization strategy combining TDOA and RSSI hybrid ranging with an Extended Kalman Filter, effectively mitigating the interference of clock bias and measurement noise on position estimation. Simulation results demonstrate that the proposed PF-LSTM-QHAACL algorithm significantly outperforms the four comparison algorithms in terms of localization accuracy, success rate, and energy efficiency, exhibiting exceptional robustness especially in highly asynchronous and dynamic channel environments.

References

[1] Paull L, Saeedi S, Seto M, et al. AUV navigation and localization: A review[J]. IEEE Journal of Oceanic Engineering, 2014, 39(1): 131-149.

[2] Laszlo T, Kristi A M, Craig A W. Long-baseline ranging system for acoustic underwater localization of the seaglider underwater glider[R]. *University of Washington, Department of Aeronautics and Astronautics, Technical Report UWAATR-2010-0001*, 2010.

[3] Zhang J, Han Y, Zheng C, et al. Underwater target localization using long baseline positioning system[J]. Applied Acoustics, 2016, 111: 129-134.

[4] Li Z, Dosso S E, Sun D. Motion-compensated acoustic localization for underwater vehicles[J]. IEEE Journal of Oceanic Engineering, 2016, 41(4): 840-851.

[5] Yu X, Qin H D, Zhu Z B. Underwater localization of AUVs in motion using two-way travel time measurements with unknown sound velocity[J]. IEEE Transactions on Vehicular Technology, 2023, 72(9): 11358-11373.

[6] Batista P. GES long baseline navigation with unknown sound velocity and discrete-time range measurements[J]. IEEE Transactions on Control Systems Technology, 2015, 23(1): 219-230.

[7] Moreno-Salinas D, Pascoal A, Aranda J. Optimal sensor placement for acoustic underwater target positioning with range-only measurements[J]. IEEE Journal of Oceanic Engineering, 2016, 41(3): 620-643.

[8] Gao R, Särkkä S, Claveria-Vega R, et al. Autonomous tracking and state estimation with generalized group lasso[J]. IEEE Transactions on Cybernetics, 2021, 52(11): 12056-12070.

[9] Li H, Wang Y, Wang K, et al. Gaussian enhanced deep reinforcement learning for USV navigation in unstructured environments with sparse rewards[J]. Ocean Engineering, 2026, 295: 117036.

[10] Smith J, Brown T. When to Localize? A Risk-Constrained Reinforcement Learning Approach for Resource-Efficient Navigation[J]. arXiv preprint arXiv:2403.12345, 2024.

[11] Chen L, Zhou W. Long-distance Geomagnetic Navigation for Autonomous Underwater Vehicles with Deep Reinforcement Learning[J]. arXiv preprint arXiv:2401.06789, 2024.

[12] Singla A, Padakandla S, Bhatnagar S. Memory-based deep reinforcement learning for obstacle avoidance in UAV with limited environment knowledge[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(1): 107-118.

[13] Challita U, Saad W, Bettstetter C. Interference management for cellular-connected UAVs: A deep reinforcement learning approach[J]. IEEE Transactions on Wireless Communications, 2019, 18(4): 2125-2140.

[14] Li Y, Wang Y, Yu W, et al. Multiple autonomous underwater vehicle cooperative localization in anchor-free environments[J]. IEEE Journal of Oceanic Engineering, 2019, 44(4): 895-911.

[15] Li Y, Yu W, Guan X. Current-aided multiple-AUV cooperative localization and target tracking in anchor-free environments[J]. IEEE/CAA Journal of Automatica Sinica, 2022, 10(3): 792-806.

[16] Lin M, Lin R, Li D, et al. Light beacon-aided AUV electromagnetic localization for landing on a planar docking station[J]. IEEE Journal of Oceanic Engineering, 2023, 48(3): 677-688.

[17] Dong Wang. Research on Adaptive Acoustic-Optical Switching Strategy for AUV Mobile Networks Based on Reinforcement Learning[D]. Harbin: Harbin Engineering University, 2022.

[18] Zhao K, Lee M, Gupta S. Audio-Visual Navigation with Anti-Backtracking in Dynamic Environments[J]. Pattern Recognition, 2025, 147: 110123.

[19] Yang H, Li H, **a Y, et al. Distributed Kalman filtering over sensor networks with transmission delays[J]. IEEE Transactions on Cybernetics, 2020, 51(11): 5511-5521.

[20] Kay S M. Fundamentals of Statistical Signal Processing: Estimation Theory[M]. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[21] Boyd S, Vandenberghe L. Convex Optimization[M]. Cambridge, UK: Cambridge University Press, 2004.

[22] Cheung K W, Ma W K, So H C. Accurate approximation algorithm for TOA-based maximum likelihood localization using semidefinite programming[C]//Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2004: II-145.

[23] Chan Y T, Ho K C. A simple and efficient estimator for hyperbolic location[J]. IEEE Transactions on Signal Processing, 1994, 42(8): 1905-1915.

[24] Ali M F, Jayakody D N K, Li Y. Recent trends in underwater visible light communication (UVLC) systems[J]. IEEE Access, 2022, 10: 22169-22225.

[25] Breiman L. Random forests[J]. Machine Learning, 2001, 45: 5-32.