Journal of *Marine Science and Engineering*

MDPI

Article

Intention Recognition for Multiple AUVs in a Collaborative Search Mission

Yinhuan Wang [1,2], Kaizhou Liu [1,*], Lingbo Geng [1] and Shaoze Zhang [1]

[1] State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China; wangyinhuan@sia.cn (Y.W.); genglingbo@sia.cn (L.G.); zhangshaoze@sia.cn (S.Z.)
[2] University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: liukzh@sia.cn

Abstract: This paper addresses the challenges of intent recognition in collaborative Autonomous Underwater Vehicle (AUV) search missions, where multiple AUVs must coordinate effectively despite environmental uncertainties and communication limitations. We propose a consensus-based intent recognition (CBIR) method grounded in the Belief–Desire–Intention (BDI) framework. The CBIR approach incorporates fuzzy inference and deep learning techniques to predict AUV intentions with minimal data exchange, improving the robustness and efficiency of collaborative decision making. The system uses a behavior modeling phase to map state features to actions and a deep learning-based intent inference phase, leveraging a residual convolutional neural network (ResCNN) for accurate intent prediction. The experimental results demonstrate that the proposed ResCNN network improves intent recognition accuracy, enhances the efficiency of collaborative search missions, and increases the success rate.

Keywords: autonomous underwater vehicle (AUV); collaborative search; intent recognition; residual convolutional neural network; fuzzy inference; underwater communication

check for updates

Academic Editor: Weicheng Cui

Received: 17 February 2025
Revised: 9 March 2025
Accepted: 12 March 2025
Published: 17 March 2025

Citation: Wang, Y.; Liu, K.; Geng, L.; Zhang, S. Intention Recognition for Multiple AUVs in a Collaborative Search Mission. *J. Mar. Sci. Eng.* **2025**, *13*, 591. https://doi.org/10.3390/jmse13030591

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the increasing complexity of underwater exploration missions, collaborative operations of Autonomous Underwater Vehicle (AUV) swarms have become a critical technology for addressing challenges in wide-area search, target localization, and environmental monitoring [1,2]. However, during complex and dynamic missions, AUV swarms must adapt their collaborative strategies in real time under conditions of limited communication, environmental uncertainty, and frequent mission changes [3]. Identifying individual intents and integrating them into swarm decision-making processes is a core challenge for enabling efficient collaborative search among AUVs.

Intent recognition, as a technique for inferring decision goals from observed behaviors, plays a crucial role in scenarios with incomplete or uncertain information [4]. By analyzing observed behaviors to infer underlying intents, this method provides essential support for swarm mission planning and role allocation. Combining intent recognition with collaborative control techniques can not only enhance the adaptability of AUV swarms in dynamic and complex environments but also offer innovative solutions for multi-mission collaboration and emergency response.

Intent recognition fundamentally involves analyzing behaviors or information to infer the goals or motivations of individuals or swarms within specific contexts. It transforms external observations (such as actions, states, or communication signals) into high-level cognitive constructs (such as intents or objectives), serving as a critical bridge between

observable behaviors and underlying decision-making processes. In real-world scenarios, observed information is often incomplete or noisy. Intent recognition leverages the integration of multi-source data, prior knowledge, and environmental context to reconstruct genuine intents from limited observations.

Intent recognition has found widespread applications in various domains [5,6], including human behavior prediction [7,8], vehicle lane change intent detection [9,10], aerial combat target intent prediction [11,12], and question–answering systems [13,14], supported by relatively mature methodologies. Currently, intent recognition approaches can be broadly categorized into two main types: model-based methods and data-driven methods.

Model-based intent recognition methods rely on predefined frameworks combined with adaptive parameter adjustments to construct deterministic models. These methods have made significant advancements and are increasingly reaching maturity. Prominent approaches in this category include template matching, expert systems, decision trees, and Bayesian networks. These methods excel in scenarios where domain knowledge is well-understood and can be explicitly modeled.

Data-driven intent recognition methods, on the other hand, leverage neural networks and deep learning techniques to learn latent patterns directly from data, without the need for prior assumptions about prototype models. The rapid advancements in data collection and computational capabilities have enabled the proliferation of data-driven algorithms in recent years. Key methods in this category include artificial neural networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs), as well as cutting-edge deep learning frameworks such as transformers. While model-based methods benefit from explainability and reliability, particularly in structured environments, data-driven approaches demonstrate superior adaptability and generalization in complex and dynamic scenarios. The integration of these two paradigms, combining the robustness of model-based inference with the flexibility of data-driven learning, represents a promising direction for advancing intent recognition technologies.

Although the theory of intent recognition has seen significant advancements and has reached a relatively mature stage, its application in the field of multi-agent collaboration remains underexplored and requires further development. In particular, research on intent recognition in the domain of underwater robotic collaboration is still in its infancy, with limited foundational work and significant gaps in practical implementation. The unique challenges posed by underwater environments, such as limited communication bandwidth, high latency, and the dynamic nature of oceanic conditions, further complicate the application of intent recognition techniques [15]. The existing intent recognition methods face several challenges when applied to underwater multi-AUV collaborative search missions, which are outlined as follows:

- In contrast to other domains where intention recognition often relies on inherent or well-established mappings, such as the intuitive association between human actions and intentions, no such predefined or standardized relationship exists in the context of AUV collaborative search missions.
- Unlike autonomous driving and collaborative surface or aerial robots, which can rely on high-precision sensors to obtain large volumes of high-resolution environmental data in real time, multi-AUV systems are constrained by low bandwidth, high latency, and high packet loss rates of underwater acoustic communication. These limitations restrict intent recognition to only a limited amount of data exchange. Additionally, accumulated errors in underwater navigation lead to inaccurate positioning information. These factors result in relatively low accuracy and completeness of the data used for intent recognition.

To address the challenges mentioned above, we propose an intent recognition method for AUVs, termed the consensus-based intent recognition (CBIR) approach. Specifically, this method is grounded in the Belief–Desire–Intention (BDI) framework [16]. The BDI framework is a cognitive architecture widely used in autonomous systems and multi-agent environments to model rational decision-making processes. The BDI framework consists of three core components: beliefs, desires, and intentions. Beliefs represent the agent's understanding of the environment, including information received from sensors, prior knowledge, and inferred data. Beliefs form the foundation of the agent's perception of the current state of the world. Desires reflect the goals or objectives the agent aims to achieve. Desires are not bound by practicality; instead, they represent the ideal outcomes or end states the agent wishes to reach. Intentions represent the specific plans or actions the agent commits to executing in order to achieve its desires, given its beliefs. Intentions bridge the gap between abstract goals and practical actions, guiding the agent's behavior in a structured manner.

In this study, beliefs represent the AUV's perception of the state information of both the swarm and the target. Desires refer to the goal of successfully locating the target, while intentions indicate the actions taken by the AUV, based on its understanding of global state information, to complete the collaborative search mission. The BDI framework is used to unify the decision-making process of the entire AUV swarm. Intent recognition enhances an AUV's understanding of the global situation by identifying the intentions of other AUVs, thereby optimizing its decision-making process.

As previously analyzed, intent recognition in the AUV domain lacks a clear mapping between behavioral states and intentions. Therefore, the first challenge is to determine which specific state information influences AUV intentions and how this information impacts decision making. However, in collaborative search missions, it is impossible for the AUV swarm to complete the target search with a single decision. Throughout the mission, maintaining communication and improving detection accuracy must also be considered, requiring multiple dynamic adjustments to accomplish the mission. This complexity makes the mapping between behavioral states and intentions highly intricate.

To address this, we introduce the concept of "landmarks" [17]. In this context, a "landmark" refers to a representative state that signifies potential short-term achievements during the intermediate stages of completing the mission. Specifically, to execute the collaborative search mission, the AUV needs to dynamically adjust its decisions at fixed intervals based on the current global situation, producing an immediate intention that serves as the "landmark". In this study, "landmarks" are categorized into three types: optimizing communication, enhancing detection, and rapidly approaching the target. This classification unifies the intention types throughout the entire swarm decision-making process.

In practical operations, the situational information and action sequences of the AUV swarm often constitute massive datasets. It is neither feasible nor practical to use all this information to establish a mapping between behavioral states and intentions. On one hand, underwater acoustic communication is limited and cannot support large-scale data exchange. On the other hand, only a subset of this information truly influences intentions, while most of the data are redundant.

To address these challenges, we propose using a fuzzy inference method to establish the mapping between state information and AUV intentions. Specifically, by combining the three intention types mentioned above, we select key information that impacts decision making. By adjusting the fuzzy inference model, we enable it to produce corresponding intention results based on varying input data. Additionally, this approach ensures that when different AUVs receive identical or similar state information, they can make consistent

decisions, achieving consensus within the swarm. We refer to this process as establishing the AUV's behavior model.

In an ideal scenario where no communication packet loss or delay exists between AUVs, the intentions of other AUVs could be directly inferred using identical state features based on the behavior model. However, each AUV can only infer the behavior intentions of others based on similar, rather than identical, state features. To overcome these challenges, we propose the use of a residual convolutional neural network (ResCNN) to achieve robust and accurate intention inference for AUVs, even under the constraints of unreliable communication.

The main innovations of the proposed method are summarized as follows:

- Consensus Mechanism Based on "Landmarks"

The method introduces the concept of "landmarks" to construct consensus among AUVs, focusing only on critical information that influences key behaviors. By leveraging the behavioral consistency of AUVs in similar or identical situations, this approach significantly reduces the data volume required for intent recognition, alleviates communication overhead in underwater acoustic channels, and enhances both the accuracy and efficiency of intent prediction.

- Dual-Stage Architecture for Behavior Modeling and Intent Inference

The proposed method adopts a dual-stage architecture comprising a behavior modeling phase and an intent inference phase. In the behavior modeling phase, fuzzy inference maps the state space to the action space, providing an initial characterization of the current intent. The intent inference phase employs a residual convolutional neural network (ResCNN) to further analyze and predict the target intent based on behavioral data. This modular design enhances the flexibility and predictive power of the system.

The remainder of this paper is organized as follows: Section 2 provides a brief overview of several intent recognition methods, highlighting their strengths and limitations. Section 3 introduces the problem of intent recognition for the AUV collaborative search mission. Section 4 presents a detailed explanation of the proposed behavior modeling and intent inference methods. Section 5 conducts experiments and analyzes the performance of the proposed model. Finally, Section 6 concludes the paper with a summary of the findings and discusses potential directions for future work.

2. Related Works

In this section, we provide a comprehensive overview of existing intent recognition methods, categorizing them into two primary approaches: model-based methods and data-driven methods. We highlight their respective strengths and limitations and discuss their applicability to AUV collaborative search missions.

2.1. Model-Based Intent Recognition Methods

Model-based intent recognition relies on predefined frameworks, combined with adaptive parameter adjustments, to create deterministic models. These methods are grounded in structured representations and domain knowledge, offering the advantage of interpretability and robustness. Common techniques are described below.

2.1.1. Template Matching

Template matching utilizes predefined behavioral templates to match observed actions against known patterns [18]. Template matching is a widely used technique for intent recognition, leveraging predefined templates to match observed behaviors and infer intentions. It has been applied in various domains, such as human activity recognition, facial expression

analysis, head motion detection, and air target intention recognition, demonstrating high efficiency and generalization, especially in scenarios with limited training samples.

In the study of Jenny Margarito et al., the signal recorded by the three-axis accelerometer was automatically extracted and the user and axial-independent template was matched with the actual observed active signal [19]. The study further combines template matching with feature extraction techniques, such as feature extraction based on dynamic time warping, for human action recognition and complex activity classification. By simplifying the characteristics of the observed signal, template matching technology can effectively process high-dimensional data and improve computational efficiency, which makes it a significant advantage in wearable device data analysis [20,21]. In complex scenes such as facial expression and head posture recognition, template matching is successfully applied to human–computer interaction system and auxiliary equipment design by detecting the user's face in real time and matching the head motion feature template [22]. This study presents a case-based inference system for classifying opponent configurations in adversarial air combat environments, demonstrating improved accuracy and resilience compared to standalone learned behavior models, even under noisy conditions and high model errors [23].

The strengths of template matching lie in its simplicity, robustness to noise, and effectiveness in small-sample learning problems. However, its limitations include reliance on predefined templates, high computational costs for large-scale or high-dimensional data, and limited scalability for unseen behaviors or dynamic environments.

2.1.2. Expert Systems

Expert systems encode domain knowledge into rule-based systems to infer intent. Chang et al. proposed an expert system based on belief rules (BRB) to deal with classification problems under uncertainty [24]. Carling analyzed the naval posture and threat assessment process and designed an expert system using real-time knowledge to implement naval battlefield situation assessment [25]. Zhou et al. improved the combat intention recognition expert system to solve the problem of insufficient expert knowledge [26].

Expert systems offer significant advantages in intent recognition by leveraging domain knowledge and rule-based inference to provide highly interpretable and reliable results, particularly in scenarios requiring transparency and trust. They excel in handling uncertainty through probabilistic methods like Bayesian networks and belief rule-based systems, ensuring robust performance even with incomplete or noisy data. However, expert systems also face limitations, such as a strong dependence on manually crafted rules and domain expertise, which can hinder scalability and adaptability to dynamic or complex environments. Additionally, their computational complexity may increase significantly with the expansion of rule sets, limiting their applicability in large-scale or highly dynamic scenarios.

2.1.3. Bayesian Networks

Bayesian networks employ probabilistic graphical models to represent dependencies between variables, allowing intent inference under uncertainty. Dynamic Bayesian networks (DBN) are widely used for intent recognition of complex missions, which can infer intent from time series data under uncertain conditions.

Schrempf et al. proposed a DBN-based human–computer interaction system that enables robots to infer human intentions in real time and adjust their behavior through an intent–action–state modeling mechanism, thus achieving a more natural interaction [27]. To solve the problem of target intention recognition in air combat, a robust method based on dynamic Bayesian networks is proposed [28]. The method combines self-organizing feature mapping to pre-process the target track information, and uses the tactical features of the

target to construct a dynamic Bayesian network, which realizes the effective recognition of the enemy target's tactical intent. Bayesian networks are also used in dynamic behavior analysis, combining conditional dependence and time series modeling to achieve real-time inference of multi-objective intentions [29]. To solve the dynamic intention recognition problem in complex battlefield environment [30], the extended multi-entity Bayesian network (EMEBN) model is proposed.

Bayesian networks excel in intent recognition due to their strong capability to handle uncertainty by integrating prior knowledge with dynamic observations, making them highly suitable for complex environments [31]. They offer transparent inference through directed acyclic graphs (DAGs), which provide an intuitive representation of causal relationships, and their ability to model temporal dependencies using extensions like dynamic Bayesian networks further enhances their applicability in dynamic scenarios. However, they face challenges such as high computational complexity as the number of nodes and state space increase, and a strong reliance on initial network structure and conditional probability tables, which require extensive domain knowledge and data support.

2.1.4. Decision Trees

Because of its simple structure, high computational efficiency, and strong interpretability, the decision tree-based intention recognition method is widely used in many fields of intention inference missions. Belief decision tree introduces a belief function to model uncertainty, which improves the adaptability of the decision tree in a complex environment.

The researchers proposed a novel random forest method based on the belief decision tree, which combines belief structures by a weighted average of the weights of each tree [32]. Compared with the traditional random forest simple voting or average method, the accuracy of intention recognition is significantly improved. The decision tree algorithm is used to analyze the entrepreneurial intention of young people. By analyzing factors including demographic characteristics, social environment, attitude, and so on, a classification model based on QUEST (fast, unbiased, efficient, and statistical tree) algorithm was built [33]. Aiming at the problem of spatial object intention inference, a method of sequence intention inference based on the meta-fuzzy decision tree was proposed [34]. By defining three situation functions, namely "circling", "collision", and "observation", the internal characteristics of target trajectory and relative attitude are modeled, and the time-space modeling is carried out in combination with the Long Short-Term Memory network (LSTM), and the task transfer is realized by using meta-learning technology.

Decision tree-based intent recognition methods offer several advantages, including high computational efficiency, making them suitable for real-time scenarios, and strong interpretability, as their branching structure provides clear decision paths [35]. They are highly adaptable across various domains, from space target prediction to smart home monitoring, and demonstrate enhanced performance when integrated with techniques like fuzzy logic, meta-learning, and deep learning. However, decision trees are sensitive to data quality, prone to overfitting in noisy or skewed datasets, and may struggle with scalability in high-dimensional or dynamic environments. Additionally, extending decision trees with fuzzy or meta-learning techniques often requires complex rule design and parameter optimization.

2.2. Data-Driven Intent Recognition Methods

Data-driven methods leverage machine learning and deep learning techniques to learn intent patterns directly from data, eliminating the need for extensive prior knowledge. These methods have gained popularity due to advancements in computational power and data availability. Key approaches are included below.

2.2.1. Artificial Neural Networks

Artificial Neural Networks (ANNs) have been widely used in intention recognition because of their strong ability in feature learning and pattern recognition.

In the field of human motion intention recognition, Leap Motion Controller (LMC) combined with ANN are used to infer human operation intention [36]. By capturing hand movement data, the ANN model is used to predict which target boxes the operator will select during the stacking boxes task. In order to improve the synchronization of man-machine haptic collaboration, Liu and Hao propose an intention recognition method based on a radial basis function neural network (RBFNN) [37]. The method predicts the motion intention of a human partner by analyzing the interaction force and contact point motion characteristics (such as position and speed) in real time. The experimental results show that this model can improve the synchronization of cooperation and decrease the cooperation force significantly while eliminating the difficulty of estimating the impedance parameters of the human body. In the field of network security, an attack intent recognition method (SAIRF) based on the Fuzzy Min-Max Neural Network (FMMNN) is proposed [38]. By analyzing the attack characteristics, the network attack intent is divided into general intent (breach of security indicators such as confidentiality and integrity) and specific intent (specific attack mode).

ANNs excel in intent recognition due to their strong feature extraction and nonlinear modeling capabilities, making them highly effective in complex data scenarios. They are versatile, demonstrating cross-domain applicability from physical human–robot interaction to cybersecurity, and support efficient processing of temporal data, enabling real-time applications in dynamic environments. However, ANNs require large labeled datasets for training, limiting their use in data-scarce domains. Their black box nature reduces interpretability, posing challenges in scenarios requiring high transparency and trust. Additionally, the computational resources needed for complex neural architectures can be a bottleneck, especially in resource-constrained settings.

2.2.2. Deep Learning Approaches

In recent years, deep learning has made remarkable progress in the field of intention recognition, showing a wide range of cross-domain application potential. From gesture recognition to airborne target intent inference to motion intent analysis, deep learning models offer new solutions for intent recognition with their efficient feature learning capabilities and robustness.

In the field of gesture intention recognition, researchers use LMC to obtain gesture data and propose a variety of deep learning strategies based on the Visual Geometry Group (VGG) convolutional neural network [39]. These strategies include classical VGG-16 networks, VGG-16 combined with dynamic time warping (DTW), and classification methods based on principal component analysis (PCA) data dimensionality reduction. Three deep learning methods based on a fully connected network, CNN, and RNN are proposed to solve the problem of target intent recognition in complex air combat environments [40]. These methods are combined with expert experience to simulate air combat scenarios and extract target attitude information and script labels for model training. Compared with the traditional Dempster–Shafer theory (D-S) inference method, this method shows stronger robustness and universality. In multi-target formation intent recognition, a new information fusion method is proposed by combining deep learning with D-S [41]. This method generates uncertain information through a deep learning network and introduces fuzzy discount weighting rules to improve the reliability of evidence. In the field of rehabilitation and assisted motion exoskeleton, a ResNet based motion intention recognition method is proposed [42]. By introducing a multi-source data fusion module, convolutional block

attention module (CBAM), and adaptive pooling module, the stability and recognition accuracy of the model are improved in the case of insufficient data. Through the analysis of lower limb motion data, the feasibility and potential application value of the algorithm in the field of exoskeleton are verified.

Deep learning-based intent recognition methods excel in automatically extracting critical features from high-dimensional data, eliminating the need for manual feature design. They demonstrate remarkable cross-domain applicability, from gesture recognition to motion assistance, and offer robust performance in handling uncertainty and complex environmental data, enhancing model generalization. However, these methods require large-scale labeled datasets for training, limiting their application in data-scarce domains. They also demand significant computational resources, which can hinder real-time deployment and scalability. Additionally, the black-box nature of deep learning models results in a lack of interpretability, posing challenges in high-trust applications.

2.3. Summary of Related Works

Analysis of the above research reveals that both model-based and data-driven intent recognition methods have been widely applied, each offering distinct advantages and facing specific limitations, as summarized in Table 1. However, none of these methods can be directly applied to the multi-AUV domain. As previously analyzed, the above studies rely on fixed mapping relationships within their respective domains, and all of these methods require a large amount of data exchange to achieve intent recognition.

Table 1. Intent recognition methods summary.

Method	Strengths	Limitations
Template Matching	Simplicity, robustness to noise, effective in small datasets.	Relies on predefined templates, high computational costs for large-scale data.
Expert Systems	High interpretability, reliable under uncertain conditions.	Depends on domain expertise, not highly adaptable to dynamic environments.
Bayesian Networks	Handles uncertainty well, transparent inference process.	High computational complexity, requires prior knowledge and data support.
Decision Trees	Fast computation, easy to interpret, adaptable across domains.	Sensitive to data quality, prone to overfitting in noisy datasets.
ANNs	High adaptability, good for real-time applications.	Requires large labeled datasets, low interpretability.
DL	Strong feature extraction, robust performance in dynamic scenarios.	Needs large-scale data, high computational resource demands, limited interpretability.

3. Description of the Intent Recognition Problem in AUV Collaborative Search Mission

In multi-AUV collaborative search missions, AUV swarms are tasked with performing target search, tracking, and encirclement in unknown or partially observable complex marine environments. Given the dynamic and uncertain nature of the marine environment, along with the concealment and maneuverability of targets, AUVs must make rapid and accurate collaborative decisions under limited environmental information and communication constraints. During this process, each AUV predicts the intentions of other AUVs based on local sensor observations and inter-agent interaction data (e.g., position, velocity, direction), infers the behavioral objectives of each AUV, and designs multi-AUV collaborative strategies in each decision cycle to ensure the efficiency and robustness of mission execution.

The description of the intent space and intent feature input is shown below.

3.1. AUV Intention Space

The collaborative search mission typically involves a long duration and a wide spatial span. In such missions, a swarm of AUVs must collaborate intensively to complete missions such as target detection, tracking, and localization. We assume the entire process as follows:

At the initial stage of the mission, the AUV swarm performs an area patrol mission aimed at covering a broad marine area and identifying potential targets. Due to the advantages of low power consumption and high concealment, passive sonar sensors are commonly used for detection in the AUV swarm. At the beginning of the mission, one AUV uses its passive sonar sensor to preliminarily detect a target. The AUV then shares the target information with the entire swarm to ensure that other AUVs receive the preliminary location data of the target.

To enhance the accuracy of target localization and gather more information about the target, the AUV swarm needs to perform a collaborative localization and tracking mission. The entire mission is divided into two stages:

1. Positioning Stage: To quickly approach the target and prepare for subsequent collaborative localization, the AUV swarm first needs to design a series of positioning points based on the target's location. The selection of these points takes into account providing better spatial positioning for collaborative detection and localization. Each AUV navigates from its initial position to the corresponding positioning point based on the collaborative localization requirements, preparing for continuous tracking and localization. During this stage, the AUV swarm needs to ensure that no AUV is left behind while reaching the positioning point as quickly as possible.
2. Tracking Stage: Once the AUV reaches the positioning points, the AUV adopts a bearing-only passive target tracking (BOT) method to continuously track and localize the target [43]. The BOT method enables the AUV swarm to adjust its headings and speeds based on the relative direction of the target, thereby improving localization accuracy and stability.

In the entire collaborative search mission, in addition to the target position and bearing information, the AUV swarm also needs to share other relevant information in real time, such as detection errors. This allows for flexible strategy adjustment based on changes in the environment and the progress of the mission, ensuring the efficiency and robustness of the mission. To achieve this, we collected the aforementioned information during the decision-making cycle to determine the AUV's behavior in the next cycle, which is essentially the AUV's intent. Specifically, based on the two main phases of the mission, we have designed corresponding intent strategies.

In the positioning phase, the AUV's intent can be classified into two types: communication priority and positioning priority. The communication priority intent emphasizes maintaining stable communication connections within the swarm, ensuring smooth information flow and preventing swarm disintegration due to communication interruptions. On the other hand, the positioning priority intent focuses on rapidly reaching the predetermined positioning points to ensure a quick response for the subsequent target tracking mission.

In the tracking phase, the AUV swarm must minimize target detection errors to enhance localization accuracy. However, while the BOT method helps reduce detection errors, it overlooks the detection range, which may cause the AUV to exceed the detection radius. To address this limitation, the AUV's intent in this phase can be categorized into two types: detection priority and positioning priority. Under the detection priority intent, the AUV continues to track the target using the BOT method, further refining the target's localization accuracy. Once the distance to the target reaches a predefined threshold, the

intent shifts to positioning priority. The positioning priority intent prioritizes rapidly approaching the target within a specified heading range to minimize the distance. Upon reaching the threshold, the intent transitions back to detection priority.

The positioning priority in both phases is directed toward the target, thus the entire search phase can be divided into three types of intent: communication priority, positioning priority, and detection priority.

3.2. AUV Intention Characteristics

The intent of AUVs is manifested through their actions and states, with the mapping from behavior and state to intent being determined by the collaborative search decision-making process. As such, it is imperative to first identify the factors that influence this decision-making process. Building on our prior work [44], we identify several key factors in the positioning phase, including the distance to the positioning point, the success rate of the received data, the distance to the parent node in the communication topology (PNCT), and the number of companions. In the tracking phase, the principal factors include the distance to the target, current detection errors, and the variance of detection errors over time. If these factors are obtained in a timely manner, the intentions of other AUVs can be predicted using the same decision-making method, facilitating adjustments to the collaborative strategy. However, due to the inherent characteristics of underwater acoustic communication, such as significant delays and packet loss, the transmission of this information may become unreliable. As a result, intention recognition is crucial for predicting the intentions of other AUVs.

Intent recognition necessitates the fusion and analysis of information obtained from sensors, making the selection of appropriate feature inputs for the model crucial. The aforementioned factors are key contributors to intent recognition, but certain information, such as packet loss rate and detection error, exhibits significant temporal variation and has a more direct impact on intent. These factors can adversely affect the intent recognition of AUVs under conditions of high latency and packet loss. Therefore, we ultimately select the following information as inputs for the model: distance to the positioning point, distance to the PNCT, relative direction to the positioning point, relative direction to the PNCT, number of companions, and the average detection error. By integrating and analyzing these factors, we can more accurately capture the key elements influencing the variation in AUV intentions, thereby improving the accuracy and robustness of intent recognition. Table 2 further lists the specific meanings and related descriptions of each AUV intent feature indicator.

Table 2. The specific meanings and related descriptions of AUV intention characteristics.

Characteristic	Description
Distance to the positioning point	This distance reflects the overall progress of the search mission: As the AUV approaches the positioning point, it indicates that the search mission is nearing completion; once the positioning point is reached, the system will transition to the next phase of the mission or execute new instructions.
Distance to the PNCT	This distance reflects the current communication status: The farther the distance from the PNCT, the poorer the communication quality. In this case, the AUV is more likely to prioritize “communication” to ensure stable data transmission and network security.
Relative direction to the positioning point	It is determined by the deviation between the current heading of the AUV and the positioning point, providing an intuitive reflection of whether the AUV is moving toward the positioning point and the strength of its movement intention.

Table 2. *Cont.*

Characteristic	Description
Relative direction to the PNCT	Similarly to the relative direction to the positioning point, this is used to measure the bearing relationship between the AUV and the PNCT, to determine whether the AUV needs to adjust its heading to improve communication quality.
Number of companions	Represents the number of potential communication nodes available for the AUV during the mission. The greater the number of companions, the more complete and optimized the subsequent communication topology can be. As a result, the AUV may be more inclined to prioritize the completion of the positioning mission, as there are fewer concerns regarding communication.
Average detection error	This can be used to determine the current stage of the mission: During the positioning phase, since the AUV is still at a relatively large distance from the target, the detection error generally remains at a high level. Once the tracking phase begins, the AUV's trajectory gradually optimizes, and the detection error decreases and stabilizes, indicating that the tracking process is improving.

3.3. Dataset Construction for Intent Recognition in Multi-AUV Collaborative Search Mission

In the context of the multi-AUV collaborative search mission, the dataset construction process is critical for training effective intent recognition models. The quality of the data, as well as the selection of relevant features, greatly impacts the model's ability to predict and understand the intentions of AUVs in complex marine environments. The simulation dataset and real-time feature set used in this study were obtained through a collaborative search simulation platform developed in our previous work. The platform comprises multiple modules, including AUV kinematics and dynamics simulation, underwater acoustic channel simulation, underwater acoustic networking communication simulation, and the simulation of relevant control algorithms. During the simulation of the collaborative search mission, the platform can record the AUV state features in real time and construct the corresponding intent recognition dataset from these features.

During the simulation, various state information of the AUV swarm is recorded. These include, but are not limited to, AUV motion characteristics, such as speed, posture, and position information, reflecting the dynamic state of the AUV in space, communication quality indicators, such as packet loss rate and latency, representing the stability of the underwater acoustic communication network and the efficiency of information transmission, and networking communication status, including the communication state between AUVs, signal strength, and effective connections. These state feature data are used to construct the AUV behavior model, from which the AUV's intent can be inferred.

To achieve cooperative decision-making and intent recognition for the AUV swarm, the collected state features are matched during the simulation process, and the corresponding intent labels are assigned based on the AUV's behavior and mission requirements. Specifically, within each decision cycle, we determine the AUV's behavior intent based on the feature information from the following two main phases: In the positioning phase, the AUV's mission is to reach the designated position quickly. State features, such as the distance to the positioning point, relative direction to the PNCT, and the number of companions, can indicate whether the AUV has deviated from the mission requirements. In the tracking phase, the AUV's goal is to precisely locate the target. At this stage, information such as detection error and distance to the target is critical for determining the AUV's behavioral goal. Through the mapping relationships in the behavior model, we match the collected state features with the corresponding intent labels (e.g., "communication priority", "positioning priority") to form the intent recognition dataset.

To ensure the training effectiveness of the intent recognition model, we have added corresponding intent labels to each data entry in the simulation dataset. These labels are

annotated based on the actual performance of the AUV's state features in the mission. For example, during the positioning phase, the AUV's intent label can be either "positioning priority" or "communication priority", depending on the AUV's mission requirements and communication status. During the tracking phase, the AUV's intent label may be "detection priority" or "positioning priority", depending on the current detection accuracy and mission requirements.

To improve the model's generalization ability and avoid overfitting, we applied data augmentation techniques to enhance the dataset. By injecting noise and performing data normalization, the dataset is better suited to the diversity and uncertainty of real-world environments.

Once the dataset is ready, we divide it into training, validation, and test sets for model training and evaluation. The training set is used to train the model, the validation set is used for parameter tuning and preventing overfitting, and the test set is used to evaluate the model's performance and generalization ability.

4. CBIR Method in the AUV Collaborative Search Mission

To address the challenges faced by underwater robots in collaborative search missions, we propose a method called consensus-based intent recognition (CBIR), shown in Figure 1. We first unified the AUV swarm decision-making process based on the BDI framework. To establish the mapping between behavioral states and intentions, we introduced the concept of "landmarks", standardizing the intention types throughout the swarm decision-making process. These intention types include optimizing communication, enhancing detection, and rapidly approaching the target. Building on our previous work [44], we identified the key information influencing AUV decision-making: distance to the positioning point (DPP), success rate of received data (PRR), distance to the PNCT (DPN), number of companions (NoCs), detection error (DE), and variance of detection error (VDE).

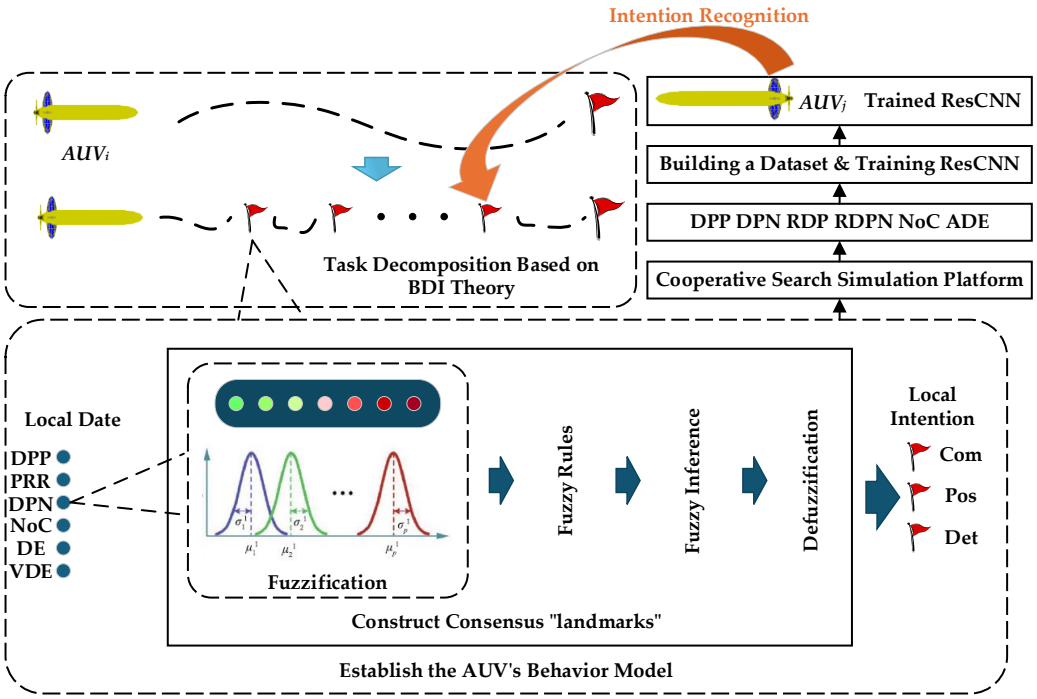


Figure 1. The CBIR framework and intention recognition process.

These features were used as inputs for the fuzzy inference model, with the three intention types as outputs. Using a fuzzy inference method, we constructed a mapping between this information and the intended actions, achieving consensus among the AUV

swarm. This approach ensures that when different AUVs receive the same or similar state information, they make consistent decisions.

Once the mapping relationship is established, intent recognition for other AUVs can be achieved by classifying this information. However, due to the large delays and high packet loss characteristics of underwater acoustic communication, certain key information with significant temporal fluctuations, such as PRR, DE, and VDE, could negatively impact classification accuracy. Therefore, we replaced these variables with alternative features that exhibit less temporal variability but still reflect the necessary characteristics. The final input features for the classification model include the DPP, DPN, relative direction to the positioning point (RDP), relative direction to the PNCT (RDPN), number of companions (NoCs), and the average detection error (ADE). The outputs of the classification model remain consistent with the three intention types: communication priority, detection priority, and positioning priority.

In the collaborative search process of AUVs, information flow begins with collecting key state data from other AUVs, including parameters such as DPP, DPN, RDP, RDPN, NoCs, and ADE. These data are processed by a pre-trained ResCNN to predict the intent of other AUVs. Using both predicted intent and data received through underwater acoustic communication, the system computes the future state of other AUVs. The fuzzy inference model uses the following information: DPP, PRR, DPN, NoC, DE, and VDE to determine the local AUV's intent, ensuring consensus within the swarm. Finally, the identified intent guides the path planning process, enabling the AUV to make adaptive decisions and contribute effectively to the collaborative search mission. The overall process is illustrated in Figure 2.

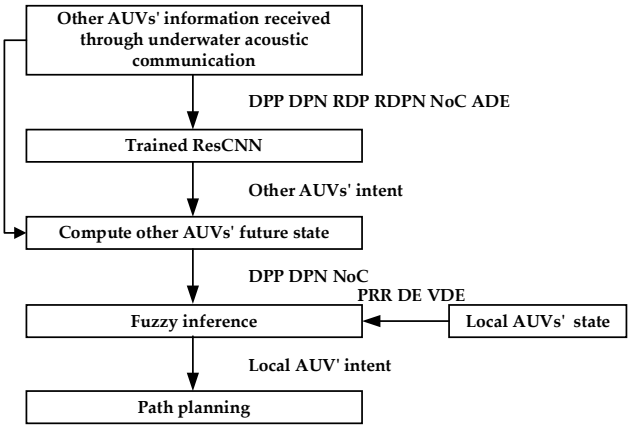


Figure 2. The data flow of the system.

In summary, by combining fuzzy inference and deep learning, and utilizing landmark consensus, the CBIR method effectively reduces communication overhead, enhancing the robustness and scalability of intent recognition in collaborative search missions. This provides an efficient and practical solution for AUVs to perform collaborative missions and achieve common objectives in complex underwater environments.

4.1. Construction of Behavior Models Based on Fuzzy Inference

Fuzzy inference is used in this study to establish the AUV behavior model, essentially serving to unify the decision-making method of the AUV swarm. The core idea is to ensure that the same or similar information consistently leads to the same decision outcome, thereby achieving decision-making consensus among the AUVs. We applied Gaussian membership functions and triangular membership functions to fuzzify the inputs and outputs, respectively. Through multiple tests, we established a comprehensive fuzzy rule

table. The Mamdani-type fuzzy implication operator and the "max-min" fuzzy composition operation were used to compute the inference result for each rule. The weighted average defuzzification method was then applied. The complete process is illustrated in Figure 3.

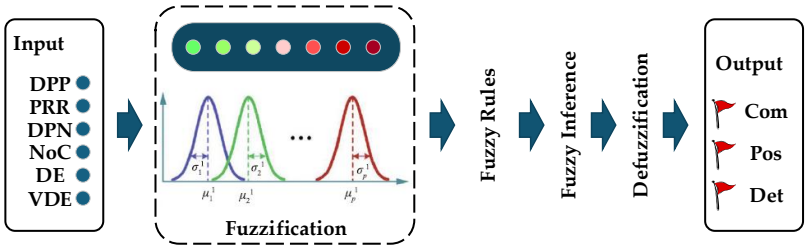


Figure 3. Flow of behavior models based on fuzzy inference.

Fuzzy inference is a method based on fuzzy logic, used to handle situations involving uncertainty and vagueness [45]. Fuzzy inference allows for the processing of fuzzy information in a system, simulating the way humans handle ambiguous and complex decisions to draw reasonable conclusions. Fuzzy inference relies on fuzzy set theory and fuzzy logic. Traditional binary logic can only represent "true" or "false", but fuzzy logic allows for continuous values between "true" and "false", commonly using numbers between 0 and 1 to represent the degree of membership of an element. Fuzzy inference consists mainly of four components: fuzzification, fuzzy rules, fuzzy inference, and defuzzification.

4.1.1. Fuzzification

The fuzzification process involves defining several fuzzy subsets for each input variable within the fuzzy domain and assigning membership functions to these subsets. The membership function serves to describe the degree of mapping of each fuzzy subset within the domain, and it is typically presented in a graphical form.

We define the fuzzy domain for NoCs as [0, n − 1], where n represents the total number of AUVs. The fuzzy domain for PRR is [0, 1], the fuzzy domain for DPP is [−3000, 0], the fuzzy domain for DPN is [−1000, 0], the fuzzy domain for detection error is [0, 10], and the fuzzy domain for detection error variance is [0, 3]. The linguistic values for all these domains are represented as "Large (L)", "Medium (M)", and "Small (S)".

Because the smooth curve characteristics of the Gaussian membership function are well-suited for handling continuous and fuzzy data, we choose the Gaussian membership function here. The mathematical expression of the Gaussian membership function is as follows:

$$\mu(x) = \exp(-\frac{(x-c)^2}{2\sigma^2}) \tag{1}$$

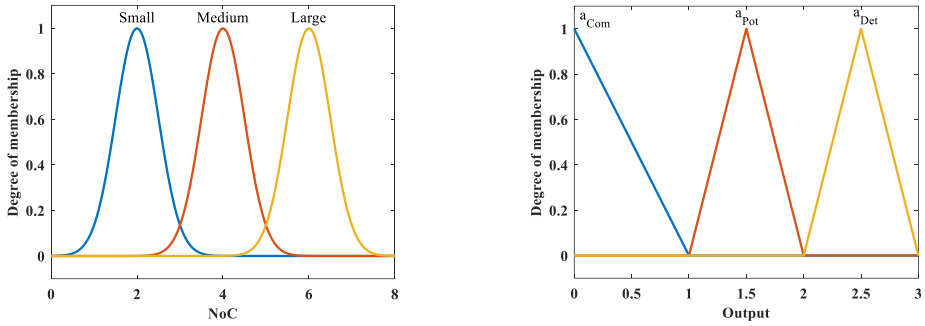
where x is the input value, c is the center of the Gaussian function, and σ is the standard deviation (which controls the width of the curve).

For the output, we designed three behavioral outcomes $A = \{a_{Com}, a_{Pot}, a_{Det}\}$, where a_{Com} is communication priority, a_{Pot} is positioning priority, and a_{Det} is detection priority. The membership function chosen for these outcomes is the triangular membership function, and the mathematical expression is as follows:

$$\mu(x) = \begin{cases} 0, & if\ x \le a\ or\ x \ge c \\ \frac{x-a}{b-a}, & if\ a < x \le b \\ \frac{c-x}{c-b}, & if\ b < x < c \end{cases} \tag{2}$$

where a is the left endpoint (minimum value) with a membership degree of 0, b is the center point (maximum value) with a membership degree of 1, and c is the right endpoint (maximum value) with a membership degree of 0.

Due to space limitations, we have presented the membership function curves for one input (NoCs) and the output separately, as shown in Figure 4.



(**a**) (**b**)

Figure 4. The membership function curves for one input (NoCs) and the output. (**a**) The membership function curves for NoCs. (**b**) The membership function curves for the output.

4.1.2. Fuzzy Rules

Fuzzy rules are used to define the relationship between input conditions and output variables. In fuzzy inference theory, fuzzy rules are expressed in the form of "IF-THEN", which determines the search strategy of the AUV. In this work, the overall goal of the AUV is to reach the positioning point as quickly as possible. However, during the process, adjustments may be made in individual cycles to maintain communication and optimize detection. For example, when communication quality deteriorates, the AUV is more likely to prioritize maintaining communication and to optimize detection, it may even move away from the positioning point. By listing all possible combinations of antecedents and corresponding consequents, we can derive a rule table (see Table 3).

Table 3. List of fuzzy rules.

No.	Rule
i	IF NoCs is L, THEN a_{Pot}
ii	IF NoCs is M, PRR is M, DPP is M and DPN is M, THEN a_{Pot}
iii	IF NoCs is S, PRR is S and DPN is M, THEN a_{Com}
iv	IF NoCs is S, PRR is L and DPN is S, THEN a_{Pot}
v	IF DPP is L, THEN a_{Pot}
vi	IF DPP is M, VDE is S and DE is L, THEN a_{Pot}
vii	IF DPP is M, VDE is L and DE is L, THEN a_{Pot}
viii	IF DPP is M, VDE is L and DE is S, THEN a_{Det}
ix	IF DPP is M, VDE is S and DE is S, THEN a_{Det}
x	IF DPP is S, THEN a_{Det}

4.1.3. Fuzzy Inference

We use the Mamdani-type fuzzy implication operator and the "max-min" fuzzy composition operation to compute the inference result for each rule. Afterward, the maximum value method is employed to aggregate the outputs of all fuzzy rules.

4.1.4. Defuzzification

The weighted average defuzzification method is used. The formula and specific process are as follows:

$$y = \frac{\sum_{i=1}^{n} \mu(x_i) \cdot x_i}{\sum_{i=1}^{n} \mu(x_i)} \tag{3}$$

where y is the defuzzified output, $\mu(x_i)$ is the membership degree of the i-th output value, x_i is the i-th output value, and n is the total number of output values.

At this point, we have completed the mapping from the state space set to the action space set based on the observed data. If communication is good and the information exchange delay between AUVs is small, we can directly use this method to infer the intentions of other AUVs. However, as discussed earlier, the acoustic communication delay is quite large and variable. The same data arriving at the receiving AUV may have already lost its timeliness. Therefore, it is necessary to use data that reflects both communication quality and mission completion status but is less sensitive to time, to recognize the intentions of the AUV.

4.2. Intent Recognition Model Based on Residual Convolutional Neural Networks

Intent recognition in underwater collaborative search missions is essentially a classification problem. We referred to the residual block design concept proposed in [46] and based on this idea, constructed the intent recognition network in this work. This architecture introduces skip connections, which allow deep neural networks (DNNs) to maintain effective gradient flow in deeper layers, thus avoiding the vanishing gradient problem during the training process [47]. This structure significantly improves the training efficiency and enhances the model's performance. The network structure is shown in Figure 5, consisting of multiple layers of residual blocks, which effectively learn the features of the input data and infer the intent of the AUV.

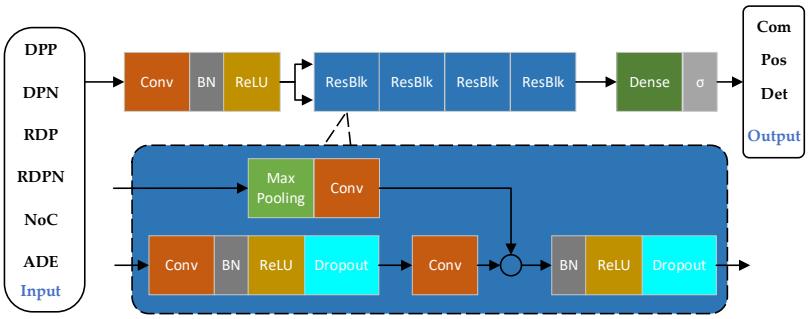


Figure 5. The network structure used for AUV intent recognition in a collaborative search mission.

The network architecture consists of an initial convolutional layer (Conv) followed by several residual blocks, each containing multiple convolutional layers. Specifically, our network features a sequence of residual blocks, with each block containing two convolutional layers. The output of the last residual block is then passed through a fully connected layer (Dense), followed by a softmax activation function. This activation function is used because the classes are not mutually exclusive; meaning, multiple classes may be predicted simultaneously. Each convolutional layer's output is rescaled using batch normalization (BN) and is then passed through a rectified linear activation unit (ReLU). Additionally, dropout is applied after the nonlinearity to prevent overfitting.

The network consists of a series of convolutional and residual blocks, designed for 1D data processing. It begins with an initial convolutional layer, which applies a 1D convolution with 16 filters, followed by batch normalization, a ReLU activation function,

and max pooling with a stride of 2. This first block reduces the spatial dimensions of the input data while extracting low-level features.

Following the initial block, there are four residual blocks, each constructed using the optimized residual network module. These blocks progressively increase the number of filters from 16 to 256, allowing the network to capture increasingly complex patterns as the depth increases. Each residual block contains skip connections to help the network learn deeper features without suffering from the vanishing gradient problem.

After passing through the residual blocks, the network uses an adaptive average pooling layer to reduce the spatial dimension to a fixed size of 1, regardless of the input length. This is followed by a flattening operation to prepare the output for the final fully connected layer. The output from the flattening operation is passed through a fully connected layer with 256 input units, which outputs a prediction for categories.

Overall, this architecture effectively balances feature extraction and complexity, utilizing residual connections to improve learning capacity while keeping the model relatively simple and efficient for classification missions.

Compared to traditional convolutional neural networks, the introduction of ResCNN enables the model to learn complex patterns more flexibly and achieve better generalization with smaller training samples. In AUV intent recognition, the combination of ResCNN and CNN not only captures the deep relationships between AUV behaviors and states but also improves the recognition accuracy and robustness in complex environments.

The training process of the ResCNN for intent recognition in AUV collaborative search missions involves several key steps. First, the training dataset is preprocessed using a data loader to convert state features and intent labels into PyTorch tensors. The ResCNN model is constructed with convolutional layers and residual blocks to maintain gradient flow and prevent vanishing gradients. During training, the model uses a cross-entropy loss function with the Adam optimizer and a learning rate scheduler to ensure efficient convergence. The training loop involves forward propagation, loss computation, backpropagation, and model weight updates over 300 epochs. Accuracy is monitored during training by comparing predicted and true labels, and the training loss is recorded to assess convergence. After training, the model is saved and tested using a custom testing function, validating its performance on unseen data. This approach ensures the model's robustness and accuracy in dynamic underwater environments.

5. Numerical Experimental Analysis

5.1. Numerical Experimental Data and Environment

The simulation experiment is based on the LP-AUV [48], as shown in Figure 6, which features a wide speed range and high load capacity, allowing it to carry various sensors for rapid response operations. Due to space limitations, detailed information on the LP-AUV mathematical model and the relevant parameters of the LP-AUV can be found in reference [48]. For a detailed modeling process of the underwater acoustic channel, please refer to our previous work [44].

The simulation scenario is illustrated in Figure 7. During the dataset generation, the initial positions of the AUVs are randomly distributed. In the initialization phase, the communication network is ensured to be connected. We assume that initially, only one AUV (referred to as AUV_1) detects the target's position during the simulation. To enhance detection accuracy, collaborative detection and localization are required. Therefore, based on the target's initial position and the needs of collaborative localization, several positioning points are designed to guide the remaining AUVs to reach these points quickly. During the initial phase, the Hungarian method is used to assign each AUV a positioning point, with the primary criterion being the minimization of the total travel distance of the AUV swarm.



Figure 6. LP-AUV at the lake test site.

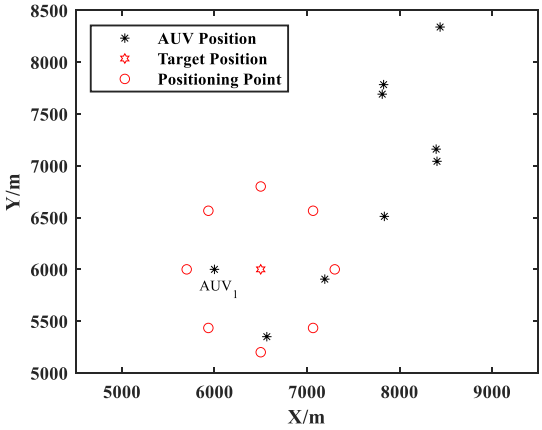


Figure 7. A schematic diagram of the initial state of the collaborative search mission.

Communication among the AUV swarm uses a multi-hop communication approach. In multi-hop communication, data are transmitted from one AUV to another through intermediate nodes rather than directly between the source and the destination. This method is particularly beneficial in underwater environments where acoustic communication has limited range, high latency, and high packet loss rates. The communication follows a hierarchical flooding routing protocol, ensuring efficient data dissemination while reducing redundant transmissions. During the simulation, the information exchanged between AUVs is affected by end-to-end delay and packet loss rate. For detailed calculation methods, please refer to our previous work [44]. During the collaborative search process, if any AUV loses connection, the search is considered unsuccessful. If all AUVs enter the tracking phase, the collaborative search is deemed successful.

The number of AUVs significantly impacts the communication topology and the complexity of collaborative decision making. If the number of AUVs is too small, the problem becomes overly simplistic, failing to demonstrate the effectiveness of the decision-making process. The speed difference between the AUVs and the target also influences the success rate of the collaborative search. AUVs moving only slightly faster than the target may struggle to achieve effective positioning and tracking, while a significant speed advantage could make the mission less challenging and reduce the need for advanced intent recognition and collaborative strategies. Although the actual communication radius of underwater acoustic networks can exceed 1 km, we set a 1 km communication and detection radius in this study to simplify the problem and focus on the proposed method's core aspects. To focus on specific research questions, we set the number of AUVs to nine, with each AUV having a cruising speed of 4 knots. The target consists of a single object, which moves at a speed of 2 knots. The parameters for underwater acoustic communication are configured to give the AUVs a communication radius of 1 km, as detailed in Table 4. The detection radius was also set to 1 km.

Table 4. The parameters for underwater acoustic channel.

Parameter	Value
Source Level (SL)	113 dB
Frequency (f)	34 kHz
Bits	144

The experimental computer system is Windows 11, with Python version is 3.8.0. The laptop was sourced from Lenovo, Beijing, China, and is equipped with NVIDIA GeForce RTX 3060 GPU. CUDA 11.0 is used for acceleration, and PyTorch 1.8.0 deep learning framework is used.

5.2. Impact of Underwater Acoustic Channels on Communication

In underwater collaborative search missions, numerous state variables are considered, including depth, position, velocity, acceleration, attitude, and detection error. The proposed method simplifies this complexity by constructing a behavior model that requires only key information rather than all available state data. This approach effectively reduces the data volume in communication packets, thereby lowering the packet loss rate. This is particularly meaningful, as the packet error rate P_{PER} is directly influenced by data size, as shown in the following formula:

$$P_{PER} = 1 - (1 - P_{BER})^n \tag{4}$$

where P_{BER} is bit error rate, and can be calculated based on Signal to Noise Ratio (SNR) and the modulation method, and n represents the number of bits in a packet. The probability of error-free reception is equivalent to the probability that all bits are received correctly. Figure 8 illustrates the impact of different data lengths on the packet loss rate under consistent underwater acoustic channel parameters. Table 5 lists the packet error rates at distances of 600 m, 700 m, and 800 m.

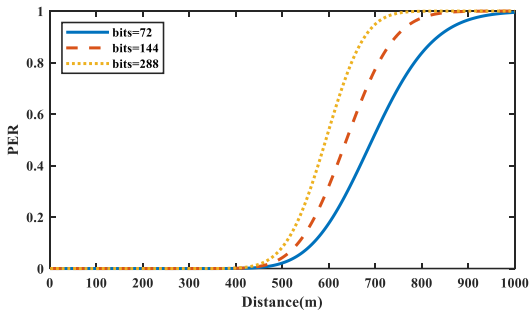


Figure 8. Variation in packet loss rate with distance under different data size.

Table 5. Packet error rate at different distances and data size.

Bits	600 m	700 m	800 m
72	17.51%	52.45%	83.68%
144	31.96%	77.39%	97.34%
288	53.70%	94.89%	99.93%

Figure 8 illustrates that as the communication distance increases, the PER rises sharply, particularly beyond 600 m. The data size significantly affects this trend, with larger data packets (288 bits) exhibiting a much steeper increase in PER compared to smaller packets (72 bits). This is because larger data packets are more susceptible to errors over longer

distances, highlighting the importance of minimizing data size to maintain reliable communication in underwater environments. The data in Table 5 further quantifies this effect.

Since there is currently no existing research on intent recognition in the AUV domain, it is not possible to directly compare the data volume of the proposed method with other approaches. But, since intent recognition can combine multi-dimensional information to predict the development trend of the swarm from a decision-making perspective, it not only reduces the need for real-time data compared to trajectory prediction but also better handles unexpected situations. This capability enables swarm collaboration under more challenging conditions, further reducing dependence on communication. Although the data packet length set in this study may not fully match real-world scenarios, the strategy of reducing data volume by constructing a behavior model to lower the packet loss rate is feasible. This approach provides valuable insights and a practical foundation for future research in this area.

5.3. Parameter Tuning

When training deep learning models, properly setting and tuning hyperparameters is a key factor in improving model performance. The main hyperparameters include, but are not limited to epoch n_e, batch size n_b, and learning rate lr. These hyperparameters significantly impact the model's convergence speed, training effectiveness, and final classification performance.

We use the accuracy metric in classification problems to evaluate the impact of each parameter on the classification mission, and the results are shown in Table 6. The accuracy of the classification task is calculated as the ratio of correctly predicted labels to the total number of samples and can be expressed as follows:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Samples} \times 100\% \tag{5}$$

Table 6. The accuracy of AUV intent recognition under different hyperparameters.

Batch Size	**Accuracy (%)**					
	n_e **= 350**			n_e **= 400**		
	lr **= 0.01**	lr **= 0.001**	lr **= 0.0001**	lr **= 0.01**	lr **= 0.001**	lr **= 0.0001**
$n_b = 128$	94.42	94.98	90.89	94.81	95.53	93.83
$n_b = 256$	94.68	95.40	94.68	95.10	95.83	94.93
$n_b = 512$	95.02	95.57	94.38	95.10	95.57	95.10

Based on the results in Table 6, we maximized the intent recognition accuracy as the criterion and ultimately determined that $n_e = 400$, $n_b = 256$, and $lr = 0.001$.

5.4. Comparative Analysis with Other Intention Recognition Methods

In this experiment, we compared the proposed ResCNN model with CNN, LSTM, and GAN models, evaluating their classification performance on the same dataset. To ensure fairness, all models used identical training and testing sets, with consistent optimizers and early stopping mechanisms. Hyperparameters for all models were tuned using grid search to ensure optimal configurations for the experiments. We use the accuracy metric in classification problems to evaluate the performance of each method. The intent recognition results are shown in Figure 9 and Table 7.

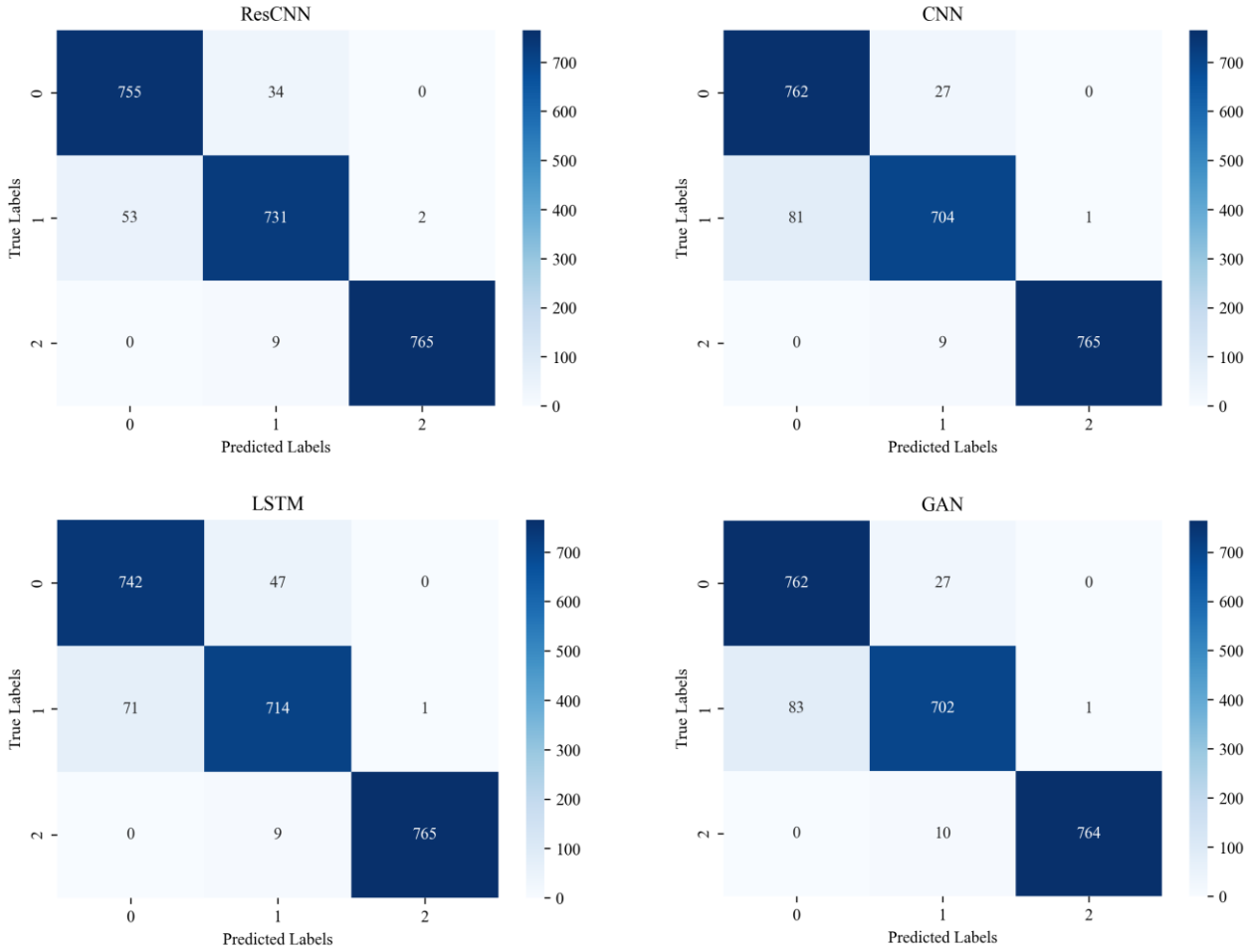


Figure 9. The confusion matrix results for different methods.

Table 7. The intent recognition accuracy for different methods.

Model	Accuracy (%)
ResCNN	95.83
CNN	94.98
LSTM	94.55
GAN	94.85

The results demonstrate that the proposed ResCNN model outperforms the other models, achieving a classification accuracy of 95.83%. This improvement is primarily attributed to the introduction of residual connections, which effectively mitigate the vanishing gradient problem, allowing the network to maintain strong training performance even at deeper layers. The CNN model achieved an accuracy of 94.98%, ranking second among the compared models. While CNN effectively captures local features in the data, its performance is somewhat limited by the absence of a deeper network structure. The LSTM model achieved an accuracy of 94.55%, slightly lower than both CNN and ResCNN, suggesting that the advantages of LSTM in time-series modeling did not fully translate into better performance for the current mission, potentially due to its weaker ability to extract local spatial features. The GAN model achieved an accuracy of 94.85%, similar to CNN, but slightly lower than ResCNN, indicating that while GAN generates richer feature representations, its classifier and feature extraction capabilities did not fully match those of ResCNN.

From the training process shown in Figure 10, it can be observed that the training loss curve of ResCNN shows an overall rapid decrease, and it converges to a relatively low loss value quickly within the first 50 epochs. As training progresses, the loss fluctuations gradually decrease and eventually stabilize, demonstrating good convergence. The training loss curve of CNN also declines quickly, but there are larger fluctuations in the later stages, especially after 250 epochs, where the loss does not show significant improvement and the fluctuation amplitude is higher than that of ResCNN. The training loss curve of LSTM exhibits a clear downward trend, with rapid loss reduction within the first 50 epochs, followed by a gradual stabilization, maintaining low fluctuations in the later stages. The training loss curve of GAN decreases slowly in the early stages, and its fluctuations are significantly higher than those of the other models throughout the entire training process. Even after a substantial number of epochs, the loss still shows noticeable oscillations.

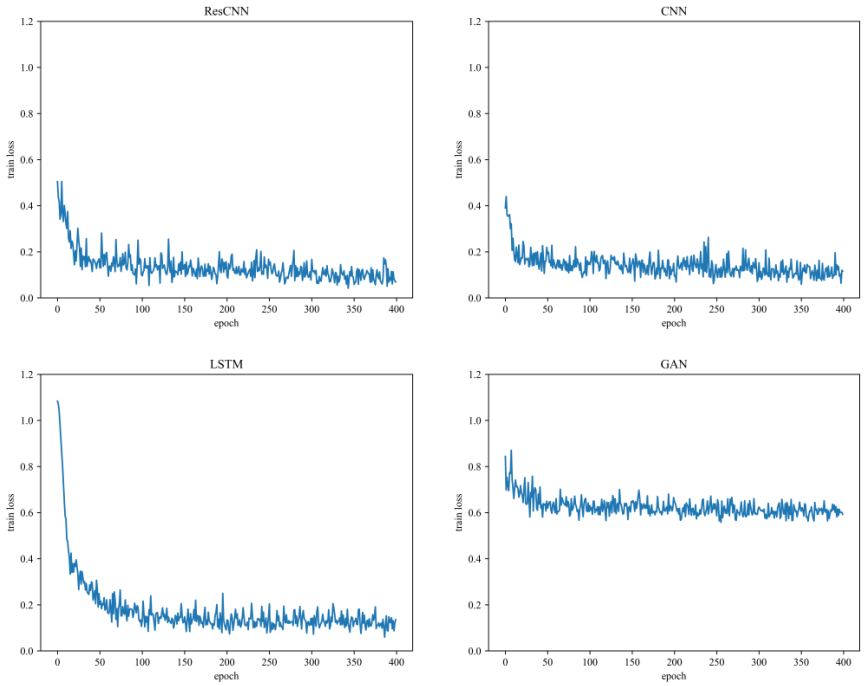


Figure 10. The loss variation during the training process for different methods.

From the analysis of the loss curves, it is clear that ResCNN outperforms the other methods in terms of training, with faster convergence and higher stability. CNN and LSTM also demonstrate certain advantages during training, but they still have limitations in deep feature extraction and handling complex missions. GAN requires further optimization of the training strategy to improve its stability and efficiency.

Overall, the experimental results strongly validate the effectiveness and superiority of the proposed ResCNN model. Compared to traditional CNN, LSTM, and GAN models, ResCNN exhibits enhanced generalization ability and stability in classification missions, providing a more efficient solution for practical applications. Future research could further enhance model performance by integrating attention mechanisms or other advanced techniques.

5.5. Comparative Analysis with Traditional Collaborative Search Methods

To validate the performance of the proposed decision consensus-based intent recognition method in collaborative search missions, we designed a comparative experiment to compare this method with the collaborative search method discussed in [44]. We designed two typical experimental scenarios to test the search efficiency and mission completion

capabilities of the method under different communication topology conditions (as shown in Figure 11):

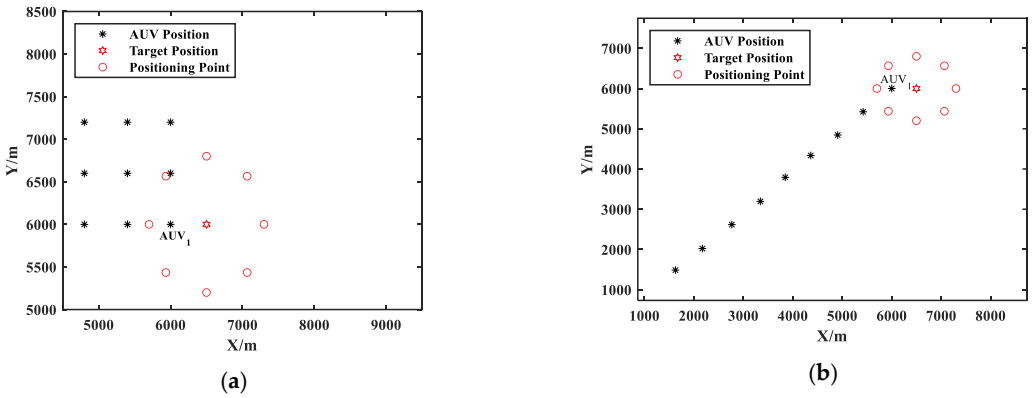


Figure 11. The initial states of the two typical simulation scenarios. (**a**) Scenario A. (**b**) Scenario B.

Scenario A: Good Communication Topology Conditions

Scenario A simulates a collaborative search environment with favorable communication conditions. Under the initial conditions, the communication topology between the AUVs is fully connected, allowing each AUV to exchange state information in real time with all other AUVs. The objective of this scenario is to compare the efficiency of the two methods in completing the search mission, with the mission completion time being the primary evaluation metric. By testing the mission duration of both methods, we can assess the advantage of the proposed method in optimizing search paths and improving mission efficiency.

Scenario B: Extreme Communication Topology Conditions

Scenario B simulates an extreme environment under sparse communication conditions. In this scenario, the communication topology between the AUVs is limited, and information exchange can only occur between nodes within the local neighborhood range. This setting simulates real-world situations where communication may be interrupted or constrained. The goal is to evaluate the differences between the two methods in prediction and decision-making capabilities, particularly in maintaining mission completion rates and search efficiency when communication is incomplete. By testing the mission success rate and search efficiency of both methods, we analyzed the robustness and adaptability of the proposed method under extreme conditions. In the experiment, all models were tested under the same initial conditions, with evaluation metrics including search time and mission success rate.

As shown in Table 8, the time taken to reach the target location using the two methods is 4219 s and 3140.5 s, respectively. The proposed method reduces the mission completion time by approximately 25.6%, demonstrating a significant improvement in efficiency.

Table 8. Time for all AUVs to reach the positioning point in Scenario A.

Method	Without IR	Proposed
Arrival Time (s)	4219	3140.5

At each sampling moment, we calculated the total distance of all AUVs to their respective positioning points to quantify the progress of the collaborative search mission. The specific calculation method is as follows:

$$\text{total DPP} = \sum_{i=1}^{n} \| Coord_AUV_i - Coord_PP_i \| \quad (6)$$

where n represents the number of AUVs, $Coord_AUV_i$ represents the coordinates of AUV_i, and $Coord_PP_i$ represents the coordinates of their corresponding occupying sites.

We visualized the results of the two methods in a graph, as shown in Figure 12. The plot illustrates the decline in total DPP over time, where the method proposed in this paper (with IR) achieves a faster reduction compared to the method without IR. This demonstrates that IR enhances the efficiency of collaborative search by enabling AUVs to reach their target positions more quickly.

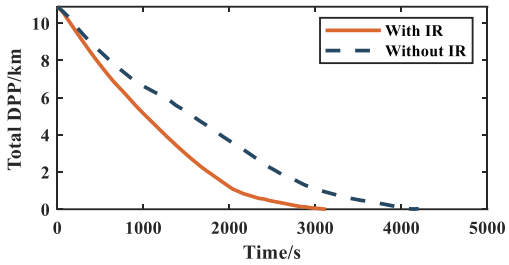
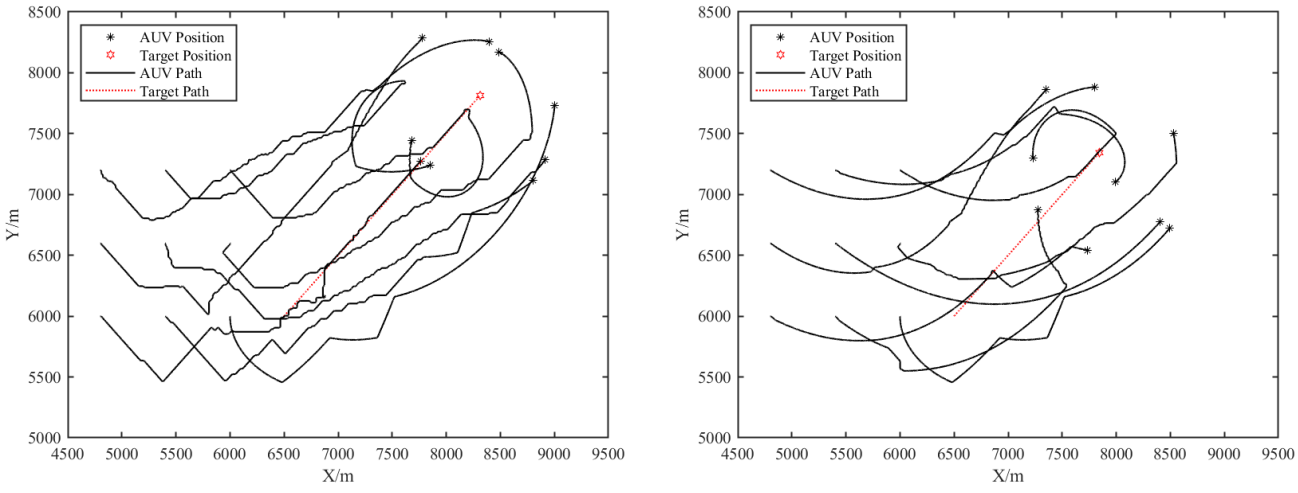


Figure 12. The variation in the total DPP in the collaborative search process in Scene A.

As shown in Figure 13, the simulation comparison results for Scenario A clearly demonstrate the advantages of the proposed CBIR method in collaborative search missions. From Figure 13a, it can be observed that the AUV's search path exhibits considerable redundant movement. This is primarily because traditional methods rely on global communication and strictly follow communication constraints for path planning, making it difficult to optimize mission execution through smarter decision making. In contrast, Figure 13b shows a more efficient path distribution, where the intent of other AUVs is predicted and used to make the collaboration between AUVs more rational, thus avoiding unnecessary path overlap. This indicates that the CBIR method, through its intent prediction and inference mechanism, effectively reduces redundant paths and enhances the efficiency of path planning.



(**a**) (**b**)

Figure 13. The trajectories of all AUVs in completing the collaborative search mission for both methods in Scenario A. (**a**) Scenario A. (**b**) CBIR method.

As shown in Table 9, the proposed method takes 15,340.5 s to complete the mission in Scenario B. Due to the more demanding topological structure, the mission completion time increases significantly. However, the proposed method is still capable of accomplishing the collaborative search mission, demonstrating better robustness compared to traditional methods.

Table 9. Time for all AUVs to reach the positioning point in Scenario B.

Method	Without IR	Proposed
Arrival Time (s)	inf	15,340.5

As shown in Figure 14, despite the increasingly stringent communication topology, the proposed method still ensures superior search efficiency.

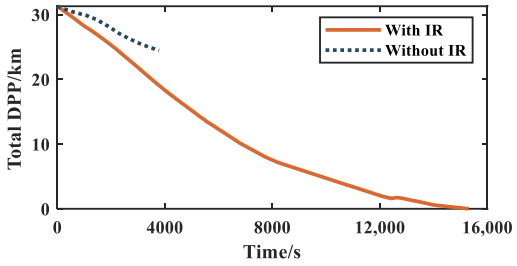


Figure 14. The variation in the total DPP in the collaborative search process in Scene B.

As shown in Figure 15, under extreme communication topology conditions for collaborative search missions, significant differences in performance were observed between the traditional method and the intent prediction-based CBIR method in the simulation. Figure 15a shows the search path of the traditional method under sparse communication conditions. Due to limited communication, AUVs are unable to exchange information in real-time, leading to a failure in collaboration. Some AUVs fall into a passive state after losing communication, unable to continue effectively participating in the mission, which results in team separation and ultimately failure of the search mission. The distance of 1.58 km indicated by the arrow exceeds the communication radius (1 km). Figure 15b illustrates the search path using the intent prediction-based CBIR method. Under the same extreme communication conditions, despite long periods without communication between AUVs, the CBIR method still enables global decision making based on local information through intent prediction and inference mechanisms. Each AUV is able to infer the intent of the entire team and dynamically adjust its own path, effectively avoiding the mission failure caused by communication interruptions in traditional methods. As a result, the collaborative search mission was successfully completed.

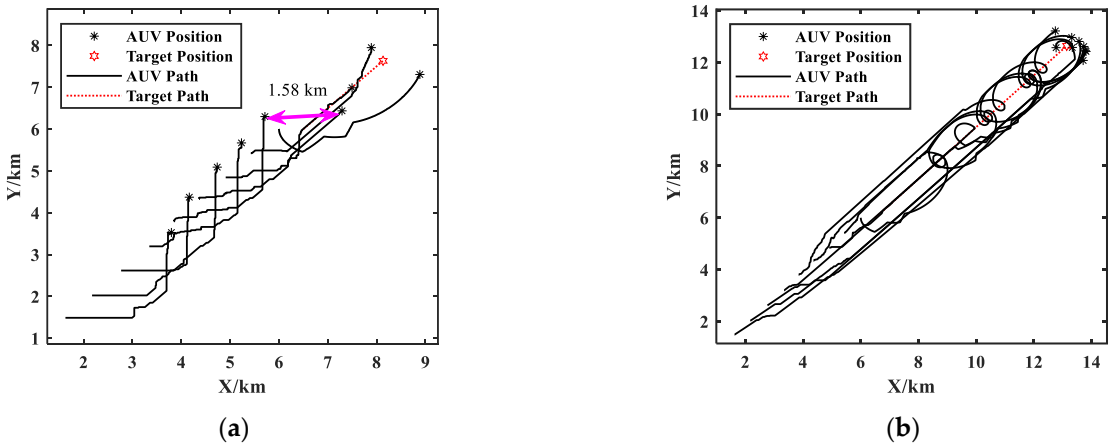


Figure 15. The trajectories of all AUVs in completing the collaborative search mission for both methods in Scenario B. (**a**) Scenario A. The distance of 1.58 km indicated by the arrow exceeds the communication radius. (**b**) CBIR method.

From the analysis, it is evident that the proposed method excels in adaptability and robustness under extreme communication conditions. By introducing a decision consensus-based intent recognition mechanism, this method effectively reduces dependence on global communication and significantly enhances the dynamic adjustment capabilities of multi-AUV collaborative search missions. In conclusion, the proposed method demonstrates superior search performance in complex environments and provides an efficient solution for multi-AUV collaborative search missions.

6. Conclusions

In this study, we introduced a novel CBIR method for collaborative AUV search missions, effectively addressing the challenges posed by underwater environments. According to the authors' research, this is the first study to explore intent recognition in the AUV domain. By combining fuzzy inference for behavior modeling with a deep learning-based intent inference phase, the CBIR approach ensures accurate intent recognition with minimal data exchange, even under communication constraints.

Although there is no relevant reference for intent recognition in the field of underwater collaborative search missions, we conducted a comparative study from two perspectives: the accuracy of the intent classification model and its performance compared to traditional collaborative search methods. The experimental results demonstrate that the proposed method outperforms in both intent classification accuracy and collaborative search mission efficiency.

While there are no comparable studies on intent recognition in AUVs to evaluate the types of data used, our simulation analysis of the impact of data packet size on underwater acoustic communication indicates that selecting key information through behavior modeling is far more effective than using more extensive data for decision making. Additionally, under more extreme communication topology conditions, the proposed method maintains high efficiency and success rates in collaborative search missions by predicting the intent of AUVs within the swarm—something traditional methods cannot achieve.

This study introduces a novel approach to underwater collaborative operations, providing valuable insights and a practical foundation for future research in multi-AUV systems.

Future efforts will focus on refining the model to handle more diverse mission scenarios and optimizing its performance in practical underwater operations.

Author Contributions: Conceptualization, Y.W.; methodology, Y.W.; software, Y.W.; validation, Y.W.; formal analysis, Y.W.; investigation, Y.W.; resources, K.L.; data curation, Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, K.L., L.G. and S.Z.; visualization, Y.W.; supervision, K.L., L.G. and S.Z.; project administration, K.L.; funding acquisition, K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 61821005 and No. 61973297) and the Liaoning Revitalization Talents Program (No. XLYC1902032).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, Z.; Du, J.; Jiang, C.; Xia, Z.; Ren, Y.; Han, Z. Task Scheduling for Distributed AUV Network Target Hunting and Searching: An Energy-Efficient AoI-Aware DMAPPO Approach. *IEEE Internet Things J.* **2023**, *10*, 8271–8285. [CrossRef]

2. Guo, J.; Li, D.; He, B. Intelligent Collaborative Navigation and Control for AUV Tracking. *IEEE Trans. Ind. Inform.* **2021**, *17*, 1732–1741. [CrossRef]
3. Jiang, B.; Du, J.; Jiang, C.; Han, Z.; Debbah, M. Underwater Searching and Multiround Data Collection via AUV Swarms: An Energy-Efficient AoI-Aware MAPPO Approach. *IEEE Internet Things J.* **2024**, *11*, 12768–12782. [CrossRef]
4. Zhang, Z.; Zeng, Y.; Jiang, W.; Pan, Y.; Tang, J. Intention recognition for multiple agents. *Inf. Sci.* **2023**, *628*, 360–376. [CrossRef]
5. Zhang, J.; Chen, Y.; Zhu, S.; Li, Y. An RGB-D Fusion Based Semantic Segmentation Algorithm Based on Neighborhood Metric Relations. *ROBOT* **2023**, *45*, 156–165.
6. Yuqing, P.; Xiaosong, Z.; Huifang, T.A.O.; Xianzi, L.I.U.; Tiejun, L.I. Hand Gesture Recognition against Complex Background Based on Deep Learning. *ROBOT* **2019**, *41*, 534–542.
7. Jain, S.; Argall, B. Probabilistic human intent recognition for shared autonomy in assistive robotics. *ACM Trans. Hum.-Robot Interact. (THRI)* **2019**, *9*, 1–23. [CrossRef]
8. Huang, C.; Xiao, Y.; Xu, G. Predicting human intention-behavior through EEG signal analysis using multi-scale CNN. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2020**, *18*, 1722–1729. [CrossRef]
9. Xing, Y.; Lv, C.; Wang, H.; Wang, H.; Ai, Y.; Cao, D.; Velenis, E.; Wang, F.-Y. Driver lane change intention inference for intelligent vehicles: Framework, survey, and challenges. *IEEE Trans. Veh. Technol.* **2019**, *68*, 4377–4390. [CrossRef]
10. Zhang, H.; Fu, R. Target vehicle lane-change intention detection: An approach based on online transfer learning. *Comput. Commun.* **2021**, *172*, 54–63. [CrossRef]
11. Guanglei, M.; Runnan, Z.; Biao, W.; Mingzhe, Z.; Yu, W.; Xiao, L. Target tactical intention recognition in multiaircraft cooperative air combat. *Int. J. Aerosp. Eng.* **2021**, *2021*, 9558838. [CrossRef]
12. Zhang, Y.; Huang, F.; Deng, X.; Mingda, L.I.; Jiang, W. Air target intention recognition and causal effect analysis combining uncertainty information reasoning and potential outcome framework. *Chin. J. Aeronaut.* **2024**, *37*, 287–299. [CrossRef]
13. Pearce, K.; Alghowinem, S.; Breazeal, C. Build-a-bot: Teaching conversational ai using a transformer-based intent recognition and question answering architecture. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; pp. 16025–16032.
14. Trewhela, A.; Figueroa, A. Text-based neural networks for question intent recognition. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105933. [CrossRef]
15. Rice, J.; Green, D. Underwater Acoustic Communications and Networks for the US Navy's Seaweb Program. In Proceedings of the 2008 Second International Conference on Sensor Technologies and Applications, Cap Esterel, France, 25–31 August 2008.
16. Kaelbling, L.P.; Littman, M.L.; Cassandra, A.R. Planning and acting in partially observable stochastic domains. *Artif. Intell.* **1998**, *101*, 99–134. [CrossRef]
17. López Diez, P.; Sundgaard, J.V.; Patou, F.; Margeta, J.; Paulsen, R.R. Facial and Cochlear Nerves Characterization Using Deep Reinforcement Learning for Landmark Detection. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2021, Strasbourg, France, 27 September–1 October 2021; pp. 519–528.
18. Yan, X.; Ai, T.; Zhang, X. Template matching and simplification method for building features based on shape cognition. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 250. [CrossRef]
19. Margarito, J.; Helaoui, R.; Bianchi, A.M.; Sartor, F.; Bonomi, A.G. User-independent recognition of sports activities from a single wrist-worn accelerometer: A template-matching-based approach. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 788–796. [CrossRef]
20. Raju, M.; Ananna, S.; Meraz, S.; Azam, M.; Serikawa, S.; Ahad, M.A.R. Human Action Recognition: A Template Matching-based Approach. *J. Inst. Ind. Appl. Eng.* **2017**, *5*, 15–23. [CrossRef]
21. Siddiqi, M.H.; Alshammari, H.; Ali, A.; Alruwaili, M.; Alhwaiti, Y.; Alanazi, S.; Kamruzzaman, M. A template matching based feature extraction for activity recognition. *CMC-Comput. Mater. Contin.* **2022**, *72*, 611–634.
22. Bankar, R.T.; Salankar, S.S. Design of Eye Template Matching Method for Head Gesture Recognition System. In *Smart Innovations in Communication and Computational Sciences*; Advances in Intelligent Systems and Computing; Springer: Singapore, 2019; pp. 3–10.
23. Floyd, M.W.; Karneeb, J.; Aha, D.W. Case-Based Team Recognition Using Learned Opponent Models. In Proceedings of the Case-Based Reasoning Research and Development, Trondheim, Norway, 26–28 June 2017; pp. 123–138.
24. Chang, L.; Zhou, Z.; You, Y.; Yang, L.; Zhou, Z. Belief rule based expert system for classification problems with new rule activation and weight calculation procedures. *Inf. Sci.* **2016**, *336*, 75–91. [CrossRef]
25. Carling, R.L. Naval situation assessment using a real-time knowledge-based system. *Nav. Eng. J.* **1999**, *111*, 173–187. [CrossRef]
26. Zhou, W.-w.; Zhang, J.-y.; Gu, N.-n.; Yan, G.-q. Recognition of combat intention with insufficient expert knowledge. In Proceedings of the 3rd International Conference on Computational Modeling, Simulation and Applied Mathematics, Wuhan, China, 27–28 September 2018; pp. 27–28.
27. Schrempf, O.C.; Albrecht, D.; Hanebeck, U.D. Tractable probabilistic models for intention recognition based on expert knowledge. In Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA, USA, 29 October–2 November 2007; pp. 1429–1434.

28. Xiao, Q.; Liu, Y.; Deng, X.; Jiang, W. A robust target intention recognition method based on dynamic bayesian network. In Proceedings of the 2021 33rd Chinese Control and Decision Conference (CCDC), Kunming, China, 22–24 May 2021; pp. 6846–6851.
29. Jiang, J.; Liu, J.; Kadziński, M.; Liao, X. A Bayesian network approach for dynamic behavior analysis: Real-time intention recognition. *Inf. Fusion* **2025**, *118*, 102873. [CrossRef]
30. Li, J.; Zhang, P.; Hao, R. Unmanned Aerial Vehicle Tactical Intention Recognition Method Based on Dynamic Series Bayesian Network. In Proceedings of the 2023 IEEE International Conference on Unmanned Systems (ICUS), Hefei, China, 13–15 October 2023; pp. 427–432.
31. Qing, J.; Xiantai, G.; Weidong, J.; Nanfang, W. Intention recognition of aerial targets based on Bayesian optimization algorithm. In Proceedings of the 2017 2nd IEEE International Conference on Intelligent Transportation Engineering (ICITE), Singapore, 1–3 September 2017; pp. 356–359.
32. Li, X.; Li, M.; Zhang, Y.; Deng, X. A new random forest method based on belief decision trees and its application in intention estimation. In Proceedings of the 2021 33rd Chinese Control and Decision Conference (CCDC), Kunming, China, 22–24 May 2021; pp. 6008–6012.
33. Djordjevic, D.; Cockalo, D.; Bogetic, S.; Bakator, M. Predicting entrepreneurial intentions among the youth in Serbia with a classification decision tree model with the QUEST algorithm. *Mathematics* **2021**, *9*, 1487. [CrossRef]
34. Wang, X.; Yang, Z.; Han, Y.; Li, H.; Shi, P. Method of sequential intention inference for a space target based on meta-fuzzy decision tree. *Adv. Space Res.* **2024**, *74*, 4050–4067. [CrossRef]
35. Sheng, Y.; Phoha, V.V.; Rovnyak, S.M. A parallel decision tree-based method for user authentication based on keystroke patterns. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **2005**, *35*, 826–833. [CrossRef] [PubMed]
36. Erdoğan, K.; Durdu, A.; Yilmaz, N. Intention recognition using leap motion controller and Artificial Neural Networks. In Proceedings of the 2016 International Conference on Control, Decision and Information Technologies (CoDIT), Saint Julian's, Malta, 6–8 April 2016; pp. 689–693.
37. Liu, Z.; Hao, J. Intention Recognition in Physical Human-Robot Interaction Based on Radial Basis Function Neural Network. *J. Robot.* **2019**, *2019*, 4141269. [CrossRef]
38. Ahmed, A.A.; Mohammed, M.F. SAIRF: A similarity approach for attack intention recognition using fuzzy min-max neural network. *J. Comput. Sci.* **2018**, *25*, 467–473. [CrossRef]
39. Ding, I., Jr.; Zheng, N.-W.; Hsieh, M.-C. Hand gesture intention-based identity recognition using various recognition strategies incorporated with VGG convolution neural network-extracted deep learning features. *J. Intell. Fuzzy Syst.* **2021**, *40*, 7775–7788. [CrossRef]
40. Qu, C.; Guo, Z.; Xia, S.; Zhu, L. Intention recognition of aerial target based on deep learning. *Evol. Intell.* **2024**, *17*, 303–311. [CrossRef]
41. Zhang, Z.; Wang, H.; Geng, J.; Jiang, W.; Deng, X.; Miao, W. An information fusion method based on deep learning and fuzzy discount-weighting for target intention recognition. *Eng. Appl. Artif. Intell.* **2022**, *109*, 104610. [CrossRef]
42. Sun, H.; Gu, X.; Zhang, Y.; Sun, F.; Zhang, S.; Wang, D.; Yu, H. An enhanced ResNet deep learning method for multimodal signal-based locomotion intention recognition. *Biomed. Signal Process. Control* **2025**, *101*, 107254. [CrossRef]
43. He, S.; Shin, H.S.; Tsourdos, A. Trajectory Optimization for Target Localization With Bearing-Only Measurement. *IEEE Trans. Robot.* **2019**, *35*, 653–668. [CrossRef]
44. Wang, Y.; Liu, K.; Geng, L.; Zhang, S. Knowledge hierarchy-based dynamic multi-objective optimization method for AUV path planning in cooperative search missions. *Ocean Eng.* **2024**, *312*, 119267. [CrossRef]
45. Mizumoto, M.; Zimmermann, H.-J. Comparison of fuzzy reasoning methods. *Fuzzy Sets Syst.* **1982**, *8*, 253–283. [CrossRef]
46. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity Mappings in Deep Residual Networks. In Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; pp. 630–645.
47. Ribeiro, A.H.; Ribeiro, M.H.; Paixão, G.M.M.; Oliveira, D.M.; Gomes, P.R.; Canazart, J.A.; Ferreira, M.P.S.; Andersson, C.R.; Macfarlane, P.W.; Meira, W., Jr.; et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat. Commun.* **2020**, *11*, 1760. [CrossRef] [PubMed]
48. Liu, K.; Wang, Y.; Cui, J.; Zhao, B.; Hu, F. Single-neuron adaptive pitch-depth control for a lift principle AUV with experimental verification. *Ocean Eng.* **2023**, *280*, 114621. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.