Deep Learning-Driven Sonar Image Fish Species Recognition and Behavioral Feature Analysis

Yao Jin (Corresponding author) Feng Jing

School of mechanical and resource engineering, Wuzhou university, Guangxi Wuzhou 543000;

Abstract: In fields such as aquatic ecology research, fisheries management, and underwater monitoring, sonar imaging technology has become a key method for acquiring fish information due to its insensitivity to water quality and light interference. However, traditional analysis methods rely on manual expertise, suffering from low efficiency, poor accuracy, and strong subjectivity, making it difficult to meet large-scale real-time demands. The powerful feature extraction and pattern recognition capabilities of deep learning offer a new pathway to address these challenges.

This paper focuses on the application of deep learning in sonar image fish recognition and behavior analysis. It begins by explaining the imaging principles and characteristics of sonar images, analyzing the limitations of traditional methods. It then reviews models suitable for sonar image processing, such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Transformers, discussing their application logic in object detection, classification, and behavior extraction. Subsequently, targeting key issues like target blurring, noise interference, and target occlusion, it proposes model optimization and data augmentation strategies. Finally, considering practical scenarios,Outlook for the application prospects of this technology in fisheries assessment, endangered species protection, underwater ecological monitoring, and other areas, providing references for related practices and research.

Keywords: Behavioral Feature Analysis; Object Detection; Feature Extraction

DOI: 10.69979/3041-0843.25.04.076

1 Research Background and Significance

1.1 Research Background

With increasing human attention to aquatic ecosystems and the growing need for the sustainable utilization of fishery resources, accurately obtaining information on underwater fish populations, species composition, and behavioral patterns has become a core prerequisite for ecological research and resource management. Sonar technology, as an active underwater detection technique, generates images by transmitting and receiving acoustic wave reflection signals. It can effectively detect underwater targets in scenarios where optical detection fails, such as turbid water and dark environments, and is thus widely used for fish monitoring in oceans, lakes, rivers, and other water bodies.

Traditional sonar image fish analysis primarily relies on manual interpretation by professionals, covering tasks such as target identification, species differentiation, and trajectory recording. However, it has significant limitations. On one hand, sonar images have low resolution, blurred target edges, and complex background noise; the accuracy of manual interpretation is easily affected by subjective factors, leading to considerable result variability. On the other hand, as monitoring scopes expand and data volumes surge, the efficiency of manual methods is extremely low, unable to process large-scale data in real-time, thus failing to provide timely support for ecological research and resource management.

1.2 Research Significance

Deep learning technology has achieved remarkable breakthroughs in the field of computer vision. Through multi-layer neural networks, it can automatically learn deep features of targets from massive datasets, greatly enhancing the performance of tasks like image classification and object detection. Applying it to sonar image fish recognition and behavior analysis holds significant theoretical and practical importance.

At the theoretical level, fundamental differences exist between sonar images and optical images. Their unique imaging mechanisms make traditional computer vision algorithms difficult to apply. The application of deep learning models can promote innovation in theories and methods for analyzing images in special underwater environments, enrich research outcomes of deep learning in non-optical image domains, and provide new ideas for the development of cross-modal image analysis technologies.

At the practical level, this technology can achieve automatic fish target recognition, species classification, and behavior extraction, significantly improving efficiency and accuracy. It can be widely applied in fishery resource assessment, endangered species protection, underwater ecological monitoring, and other fields, providing scientific basis for aquatic ecological protection and the sustainable management of fishery resources.

2 Sonar Image Imaging Principles and Limitations of Traditional Analysis Methods

2.1 Sonar Image Imaging Principles

A sonar system consists of a transmitter, transducer, receiver, signal processor, and display. Its imaging is based on the principles of sound wave propagation and reflection. The transmitter generates electrical signals, which are converted into sound waves by the transducer and emitted underwater. After the sound waves encounter targets such as fish or seabed topography and reflect back, the transducer converts the reflected sound waves back into electrical signals. These signals are then amplified and filtered by the receiver, analyzed by the signal processor, and finally used to generate sonar images.

Based on imaging methods, sonar images can be divided into three categories. Side-scan sonar images present the contours and distribution of targets using grayscale values, where fish targets appear as areas of grayscale difference but with blurred edges and few details. Synthetic Aperture Sonar (SAS) images improve resolution by synthesizing signals from multiple small apertures, capable of clearly displaying fish morphology; however, the large data volume increases processing difficulty. Multibeam sonar images can obtain the three-dimensional position of targets, supporting the analysis of fish spatial distribution and swimming trajectories, but they are susceptible to interference from water currents, water refraction, and other factors, leading to noise interference.

2.2 Limitations of Traditional Fish Recognition and Behavior Analysis Methods

Traditional methods, centered on manual interpretation supplemented by simple image processing techniques like threshold segmentation and edge detection, exhibit obvious limitations.

First, accuracy is low and subjectivity is strong. The grayscale of sonar images is influenced by various factors such as target size, distance, and background. The grayscale difference between fish and the background is often not distinct, and features of some species overlap. Manual interpretation relies on experience, is prone to misjudging targets with similar features, and fatigue from prolonged work further reduces accuracy, making it difficult to ensure consistency and reliability of results.

Second, processing efficiency is low. With the advancement of sonar technology, detection ranges continue to expand, and data acquisition speeds accelerate, with daily data volumes potentially reaching tens to hundreds of GB. Manual methods require analyzing images frame by frame, cannot process large-scale data in real-time, and vast amounts of image data struggle to be converted into effective information promptly, delaying decision-making in ecological research and resource management.

Third, the capability for behavioral feature analysis is weak. Traditional methods can only manually record simple swimming trajectories of fish. They are unable to quantitatively analyze complex behaviors of groups (such as aggregation, migration) and individuals (such as acceleration, predation), struggle to capture dynamic behavioral changes, and cannot reveal the correlations between fish behavior and environmental factors like water temperature and currents, limiting in-depth research into behavioral mechanisms.

3 Deep Learning Models Suitable for Sonar Image Processing

Deep learning models can automatically learn the deep features of fish targets in sonar images through multiple layers of nonlinear transformations, thereby overcoming the limitations of traditional methods. According to task requirements, commonly used models include Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Transformers. Their working principles and application scenarios are as follows.

3.1 Convolutional Neural Networks (CNN)

The core advantage of CNNs lies in extracting local features of images through convolution operations and reducing dimensionality while enhancing translation invariance through pooling operations, making them suitable for fish target detection and species classification.

In object detection, two-stage models represented by Faster R-CNN first generate candidate regions through a Region Proposal Network (RPN), then perform feature extraction and classification, offering high accuracy and strong resistance to background noise. However, such models suffer from high computational complexity and slow running speeds. Single-stage models, typified by YOLO and SSD, predict target categories and locations through a single convolution operation, possess faster running speeds, and can adapt to real-time monitoring tasks. Among them, the YOLO model also enhances the detection capability for small targets through multi-scale fusion technology, but has the drawback of relatively lower accuracy and susceptibility to missed and false detections.

In species classification tasks, CNNs extract shallow and deep features through multiple layers of convolution and pooling operations, and achieve classification via fully connected layers. Researchers often introduce attention mechanisms to focus on key regions, use feature fusion methods to enrich feature information, and employ transfer learning to address issues of small dataset size and difficult annotation, thereby improving model performance.

3.2 Recurrent Neural Networks (RNN)

Recurrent Neural Networks (RNNs) possess the ability to process sequential data. They capture temporal dependencies within the data through the transmission of hidden states, making them suitable for the dynamic analysis of fish behavior. A single sonar image can only reflect instantaneous information, whereas a sequence of sonar images contains information about fish movement trajectories and behavioral changes. Long Short-Term Memory networks (LSTM), utilizing their gating structure, solve the gradient problem in long sequences and can capture long-term behavioral features such as fish migration. Gated Recurrent Units (GRU) simplify the gating structure, balancing computational efficiency while maintaining performance. Feeding the spatial features of frame images extracted by a CNN into an RNN to build a hybrid model enables spatiotemporal joint analysis, thereby more accurately identifying complex behaviors such as fish school aggregation and predation.

3.3 Transformer Models

Transformer models, based on the self-attention mechanism, can capture dependencies between any positions in a sequence, possessing powerful global modeling capabilities and high parallel computational efficiency. In fish recognition tasks, this model can simultaneously focus on the global and local features of fish, addressing the issue of insufficient global feature extraction in CNNs when processing large-size images. For example, the DETR model divides an image into multiple patches and calculates their relationships, achieving end-to-end object detection, simplifying the model structure, and possessing strong capabilities against occlusion and complex background interference. In fish behavior analysis, Transformer models can process sequential sonar images in parallel, offering faster processing speeds than RNNs, and can also capture long-range dependencies in lengthy sequences, analyzing dynamic patterns such as seasonal fish migration. The Vision Transformer (ViT) model combines the advantages of CNNs and Transformers, integrating local feature extraction and global modeling capabilities, further enhancing model performance.

4 Key Issues in Sonar Image Processing and Solution Strategies

Although deep learning technology shows promising application prospects in the field of sonar image fish analysis, it still faces three key challenges due to the inherent characteristics of sonar images and the underwater environment: target blurring and noise interference, target occlusion, and data scarcity and imbalanced distribution. Corresponding solution strategies are as follows:

4.1 Target Blurring and Noise Interference & Solution Strategies

Sonar images have low resolution, resulting in blurred edges and details of fish targets. Furthermore, influenced by factors like water scattering and equipment noise, the images contain considerable background noise, adversely affecting the feature extraction precision of models.

To address this issue, during the data preprocessing stage, methods such as adaptive median filtering and wavelet transform can be used for denoising. Operations like grayscale stretching and histogram equalization can enhance the contrast between the target and the background. Super-resolution reconstruction techniques can be employed to improve image resolution. In terms of model optimization, residual connections can be introduced into CNNs to reduce the impact of noise on deep features, and dilated convolutions can be used to expand the receptive field, thereby capturing the overall features of blurred targets. In Transformer models, the self-attention mechanism can be improved to reduce the weight of noisy regions.

4.2 Target Occlusion & Solution Strategies

In sonar images, fish may occlude each other or be occluded by underwater obstacles, preventing the model from obtaining complete target features and leading to missed or false detections.

Solution strategies mainly include: Strategy 1, adopting multi-scale feature fusion and context modeling methods, fusing feature maps of different resolutions, using shallow features to capture details of unoccluded parts, using deep features to infer the contours of occluded parts, and combining environmental features to assist in target judgment. Strategy 2, designing occlusion-aware attention mechanisms to identify occluded regions in the image and adjust their weights, or using Generative Adversarial Networks (GANs) to complete occluded features. Strategy 3, simulating occlusion scenarios through data augmentation methods, artificially adding occlusion blocks to construct training datasets, thereby enhancing the model's generalization ability.

4.3 Data Scarcity and Imbalanced Distribution & Solution Strategies

The acquisition of sonar images requires specialized equipment and sites, and image annotation is labor-intensive, leading to small dataset sizes and uneven distribution of species samples. This results in insufficient model training and a tendency to overfit, with low recognition accuracy for minority categories.

In terms of data augmentation, semi-supervised/unsupervised learning methods can be used to leverage unlabeled data, for example, using unsupervised clustering to assist annotation, label propagation, and other methods to expand the dataset size, or using GANs to generate synthetic images that conform to sonar imaging characteristics. For sample balancing, resampling methods (oversampling to increase the number of minority class samples, undersampling to reduce majority class samples) or optimizing the loss function (such as weighted cross-entropy, giving higher penalties for misclassifying minority classes) can be employed. However, practical considerations must be taken to avoid overfitting or loss of information.

5 Technical Application Prospects and Future Development Directions

5.1 Main Application Scenarios

With high efficiency and accuracy as its core advantages, this technology holds significant value in three major areas: In fishery resource assessment, utilizing sonar and deep learning enables automated fish monitoring, helping to understand population dynamics, formulate fishing policies, and ensure resource sustainability. In endangered species protection, it can enable non-intrusive monitoring of species like the Chinese sturgeon, tracking their trajectories and habitats, providing alerts for anomalies, and enhancing protection precision. In underwater ecological monitoring, correlating fish analysis with environmental data can reveal the impact of the environment on fish, and changes in fish populations can help detect ecological issues, providing support for restoration efforts.

5.2 Future Development Directions

Breakthroughs are needed in four aspects: First, developing lightweight models through techniques like pruning to reduce complexity and adapt to portable devices. Second, advancing multi-modal fusion, combining sonar with optical, acoustic, and other data to enhance analytical capabilities. Third, improving cross-scene generalization by building diverse datasets and using techniques like domain adaptation to enhance model adaptability. Fourth, delving deeper into behavioral mechanisms, integrating behavioral ecology, and using deep learning and reinforcement learning to reveal the laws governing fish behavior.

6 Acknowledgements

Funding Support: Wuzhou Science and Technology Plan Project for 2024, "Research and Development of Wideband Parametric Array Fishery Monitoring Sonar" (Contract No.: 202402020)， and "Research on Key Technologies of Medical Bionic Soft Robotics" (Contract No.: 202402024).

References

[1] Ship and Ocean Structure Design and Manufacturing. Deep Learning-Based Method for Small Underwater Target Recognition in Sonar Images [D]. 2023.
[2] Hu Gang. Research on Underwater Target Recognition and Motion Behavior Analysis Technology Based on Deep Learning [D]. Harbin Engineering University, 2021.
[3] Pattern Recognition and Intelligent Systems. Research on Underwater Target Recognition and Motion Behavior Analysis Technology Based on Deep Learning [D]. 2021.
[4] Zhou Qiuru. Research on Sonar Image Target Detection and Data Augmentation Methods Based on Deep Learning [D]. Harbin Engineering University, 2023.
[5] Xu Huanyu. Research on Classification of Forward-Looking Sonar Images Based on Deep Learning [D]. Harbin Engineering University, 2020.

Author Profile: Yao Jin (1982—), male, Han ethnicity, born in Pingyi County, Linyi City, Shandong Province, holds a master's degree in engineering, is an associate professor and senior engineer at Wuzhou University, with research interests in:Industrial robots, marine engineering information systems, and underwater target detection and recognition technologies.
Funding Support: Wuzhou Science and Technology Plan Project for 2024, "Research and Development of Wideband Parametric Array Fishery Monitoring Sonar" (Contract No.: 202402020).
Author's Profile: Feng Jing (1985-), female, Han ethnicity, born in Zhongxiang City, Hubei Province, holds a master's degree in Engineering with a specialization in Precision Machinery Design. Currently an associate professor at Wuzhou University, she is a senior engineer specializing in precision machinery design.
Funding Support: Wuzhou Science and Technology Plan Project for 2024, "Research on Key Technologies of Medical Bionic Soft Robotics" (Contract No.: 202402024).