

Article

Research on Self-Noise Processing of Unmanned Surface Vehicles via DD-YOLO Recognition and Optimized Time-Frequency Denoising

Zhichao Lv ¹, Gang Wang ¹, Huming Li ¹, Xiangyu Wang ¹, Fei Yu ¹, Guoli Song ^{2,*} and Qing Lan ³

¹ College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China; lvzhichao@hrbeu.edu.cn (Z.L.); wanggang@sdust.edu.cn (G.W.); 2017051211@hrbeu.edu.cn (H.L.); wangxiangyu@sdust.edu.cn (X.W.); feiyu@sdust.edu.cn (F.Y.)

² The Key Laboratory of Underwater Acoustic Environment, Institute of Acoustics, Chinese Academy of Sciences (CAS), Beijing 100190, China

³ Wuhan Second Ship Design and Research Institute, Wuhan 430205, China; lanqing67@hrbeu.edu.cn

* Correspondence: songguoli@mail.ioa.ac.cn

Abstract

This research provides a new systematic solution to the essential issue of self-noise interference in underwater acoustic sensing signals induced by unmanned surface vehicles (USVs) operating at sea. The self-noise pertains to the near-field interference noise generated by the growing diversity and volume of acoustic equipment utilized by USVs. The generating mechanism of self-noise is clarified, and a self-noise propagation model is developed to examine its three-dimensional coupling properties within spatiotemporal fluctuation environments in the time-frequency-space domain. On this premise, the YOLOv11 object identification framework is innovatively applied to the delay-Doppler (DD) feature maps of self-noise, thereby overcoming the constraints of traditional time-frequency spectral approaches in recognizing noise with delay spread and overlapping characteristics. A comprehensive comparison with traditional models like YOLOv8 and SSD reveals that the suggested delay-Doppler YOLO (DD-YOLO) algorithm attains an average accuracy of 87.0% in noise source identification. An enhanced denoising method, termed optimized time-frequency regularized overlapping group shrinkage (OTFROGS), is introduced, using structural sparsity alongside non-convex regularization techniques. Comparative experiments with traditional denoising methods, such as the normalized least mean square (NLMS) algorithm, wavelet threshold denoising (WTD), and the original time-frequency regularized overlapping group shrinkage (TFROGS), reveal that OTFROGS outperforms them in mitigating USV self-noise. This study offers a dependable technological approach for optimizing the performance of USV acoustic systems and proposes a theoretical framework and methodology applicable to different underwater acoustic sensing contexts.

Keywords: unmanned surface vehicle; YOLO11 algorithm; optimized time-frequency regularized overlapping group shrinkage; self-noise



Academic Editor: Weicheng Cui

Received: 31 July 2025

Revised: 1 September 2025

Accepted: 2 September 2025

Published: 4 September 2025

Citation: Lv, Z.; Wang, G.; Li, H.; Wang, X.; Yu, F.; Song, G.; Lan, Q. Research on Self-Noise Processing of Unmanned Surface Vehicles via DD-YOLO Recognition and Optimized Time-Frequency Denoising. *J. Mar. Sci. Eng.* **2025**, *13*, 1710. <https://doi.org/10.3390/jmse13091710>

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

Maritime information technology possesses considerable utility in marine resource management and environmental protection, facilitating essential functions such as environmental monitoring, early warning, resource exploitation, and disaster prevention and response [1,2]. Modern maritime observation systems extensively utilize unmanned mobile vehicles outfitted with various sensors to attain thorough collecting of oceanographic

data [3]. These platforms are essential for marine data collecting, environmental monitoring, resource development, and disaster forecasting [4,5]. Nevertheless, the variety and multitude of devices incorporated in unmanned surface vehicles (USVs) can significantly impair the efficacy of onboard acoustic systems due to the multi-band self-noise produced during operation [6]. This problem is especially evident when the platform is equipped with underwater acoustic transceivers, since the resultant self-noise may disrupt or entirely conceal the intended target signals. An exhaustive examination of the spatiotemporal attributes of self-noise on USV platforms, coupled with the formulation of specialized noise identification and mitigation techniques, possesses considerable theoretical significance and practical relevance for augmenting marine target detection efficacy and refining the precision of underwater acoustic signal recognition.

The investigation of the spatiotemporal attributes of self-noise produced by unmanned surface vehicles identifies three principal mechanisms of noise generation: propeller cavitation noise [7], motor vibration noise [8], and structure resonance noise [9]. These noise sources have notable variability in the frequency domain and are characterized by considerable time delays and multipath propagation effects [10]. During USV operation, the propeller's rotation generates significant mechanical vibrations along the propulsion shaft, which are later transformed into two distinct types of noise characteristics: wideband noise and discrete tonal noise [11]. Broadband noise is chiefly defined by its aperiodic characteristics, and its generation mechanism is intricately linked to turbulent boundary layers and cavitation occurrences [12–14]. Conversely, discrete tonal noise is distinguished by significant peaks at particular frequency components inside the frequency domain, a characteristic notably evident in USV self-noise. This noise predominantly originates from periodic excitation processes, such as rotor–stator interactions caused by impeller rotation and periodic flow separation events. Common manifestations encompass distinct frequency components, including the rotation frequency (RF), blade passing frequency (BPF), and their harmonic frequencies. The self-noise of the USV displays intricate multi-source characteristics, along with time-delay effects and multipath propagation features in near-field transmission [15]. The spatiotemporal characteristics offer essential theoretical insights for the precise identification of noise sources and the formulation of efficient noise reduction strategies.

In target feature recognition, conventional machine learning techniques predominantly depend on manually crafted feature extraction methods that involve deriving explicit features like edges, textures, colors, and shapes from images, and utilizing classification algorithms such as support vector machines (SVM), decision trees, or random forests for target identification [16,17]. Nevertheless, handmade feature-based approaches demonstrate significant limits when addressing complex high-dimensional data or identifying small objects, frequently leading to diminished recognition accuracy and heightened false positive rates. Conversely, deep learning methodologies autonomously acquire intrinsic feature representations from data and utilize multi-layer neural networks to model intricate patterns, exhibiting considerable benefits in target recognition tasks. Single-stage detection approaches have garnered significant interest owing to their high efficiency, with the YOLO series of algorithms [18–20] being particularly notable. In recent advancements within the YOLO family, YOLOv5 has attained superior performance in feature processing by using sophisticated methods such as mosaic data augmentation, generalization algorithms, and feature pyramid networks (FPN). YOLOv11, the most recent evolutionary iteration of YOLOv5, integrates an advanced feature extraction architecture with a multi-scale detection technique, yielding substantial improvements in detection accuracy and performance. As a result, it has been extensively embraced across multiple application domains [21,22]. The YOLO series algorithms often utilize time-frequency processing techniques in visual

recognition tasks; however, underwater audio signals possess distinct attributes, including multi-source composition, multipath propagation effects, and Doppler shifts. These characteristics hinder the efficacy of traditional spectral analysis and convolutional neural networks in addressing noisy data with significant overlap and delay spread. Conversely, attributes in the DD domain can proficiently encapsulate both time-delay and frequency-shift data within signals, thereby markedly improving recognition reliability [23–29].

Recent years have witnessed substantial progress in signal separation and denoising methodologies within the domain of noise suppression research. Conventional techniques, such spectral subtraction and band-pass filtering, mainly address noise within designated frequency ranges; however, their efficacy diminishes when the noise coincides with the frequency range of the desired signal [30]. Despite the application of empirical mode decomposition (EMD) and its enhanced versions in source separation and denoising within marine soundscapes [31], they continue to encounter technical difficulties, including mode mixing and end effects [32].

Methods based on wavelet packet transform (WPT) facilitate band-selective denoising; nonetheless, their efficacy is significantly contingent upon the manually chosen wavelet basis functions [33–36]. Principal component analysis (PCA), a data-driven methodology, attains noise reduction by projecting the acoustic spectrum onto an orthogonal subspace, exhibiting specific benefits in the analysis of maritime acoustic environments [37–39]. Nonetheless, when faced with intricate non-stationary noise distributions—especially those characterized by significant spectral overlap or an absence of clear structural features—the denoising efficacy of PCA is markedly constrained.

This study innovatively applies the YOLOv11 object detection framework to learn and identify features in the DD domain, aiming to mitigate the substantial interference from complex self-noise generated during USV operations on underwater acoustic perception systems. Furthermore, taking into account the structurally sparse and delay-spread attributes of USV self-noise in the time-frequency domain, a novel time-frequency regularized denoising model is introduced. The primary contributions of this study are as follows:

1. This study elucidates the generation mechanisms of self-noise and constructs a propagation model under spatiotemporal fluctuation conditions, offering a comprehensive analysis of the characteristics of USV self-noise and revealing its three-dimensional coupling properties in the time-frequency-space domain within near-field environments.
2. A novel YOLOv11-based object detection system is introduced to analyze self-noise defined by delay spread and spectral overlap, employing DD feature maps. This framework utilizes transfer learning to modify a visual recognition network for auditory feature detection, facilitating swift identification of USV self-noise.
3. A refined time-frequency domain denoising model utilizing overlapping group shrinkage is established, incorporating structural sparsity restrictions and non-convex regularization techniques to create an optimal framework characterized by overlapping group shrinkage attributes. This model attains effective and resilient self-noise suppression and signal reconstruction in intricate noise environments.

2. Theoretical

2.1. Analysis of the Mechanism and Characteristics of Self-Noise

To initiate research on noise detection and management in unmanned platforms, it is essential to deliver a thorough examination of the characteristics and mechanisms of noise generation in these platforms. Table 1 categorizes self-noise into three types: flow noise, propeller noise, and mechanical noise. The mechanisms and characteristics of various types

of self-noise generation differ, with a specific analysis conducted on cavitation noise and blade passage frequency noise within designated classifications.

Table 1. Summary of radiation noise characteristics.

Noise	Noise Source
Mechanical noise	Diesel engine, motor reducer, etc.
Propeller noise	Cavitation on and near propellers Shell resonance caused by propellers
Hydrodynamic noise	Water flow radiation noise Resonance of cavities, plates, and attachments Cavitation on pillars and attachments

2.1.1. Cavitation Noise

A persistent cavitation noise was generated by a series of disintegrating air bubbles of differing diameters. The spectral level displays differing characteristics across distinct frequency bands as the frequency increases. The spectral strength amplifies in the low-frequency range and diminishes by 6 dB per octave in the high-frequency band. Figure 1 depicts this pattern [40].

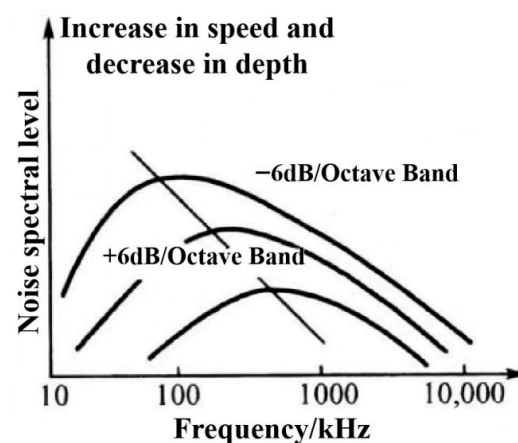


Figure 1. This is a figure. Schemes follow the same formatting.

2.1.2. Blade Passing Frequency Noise

Propellers operate in turbulent flow conditions, where the fluid disturbed by blade rotation produces low-frequency line spectrum noise known as blade rate spectral noise, with frequencies between 1 and 100 Hz:

$$I_w = k \cdot v^n \quad (1)$$

where k is a constant; v is the speed; and n is a quantity related to factors such as the underwater line shape of the ship.

2.2. Analysis of the Spatial Distribution of Self-Noise Based on Measured Data

In contrast to conventional permanent undersea platforms, unmanned surface vehicles (USVs) demonstrate notable dynamic attributes, including frequent activations and elevated cruising velocities. These qualities produce self-noise with intricate features, encompassing multi-source composition, considerable temporal fluctuation, and notable spatial distribution disparities. This study carefully examines self-noise data acquired under real operating situations to gain a greater knowledge of noise distribution patterns,

considering three dimensions: various spatial positions on the platform, differing water depths, and distinct operational states.

In this experiment, the parameters were as follows: the water depth was 15 m; the vessel's draft was 2 m; the receiving array comprised five sub-arrays, each equipped with 32 hydrophones, with 5 m spacing between adjacent elements; and the USV measured 9 m in length and 2.5 m in width [41].

As shown in Figure 2a, numerical simulations were performed for the near-field self-noise with the propeller mounted at the stern. The results reveal a pronounced axial asymmetry: peak sound pressure is concentrated in the vicinity of the propeller disk and diminishes progressively downstream, whereas levels toward the bow are markedly lower owing to hull shielding and geometric spreading. Vertically (along the z-axis), sound pressure decreases with increasing depth, indicating that the self-noise energy is largely confined to the propeller layer and attenuates rapidly in deeper water.

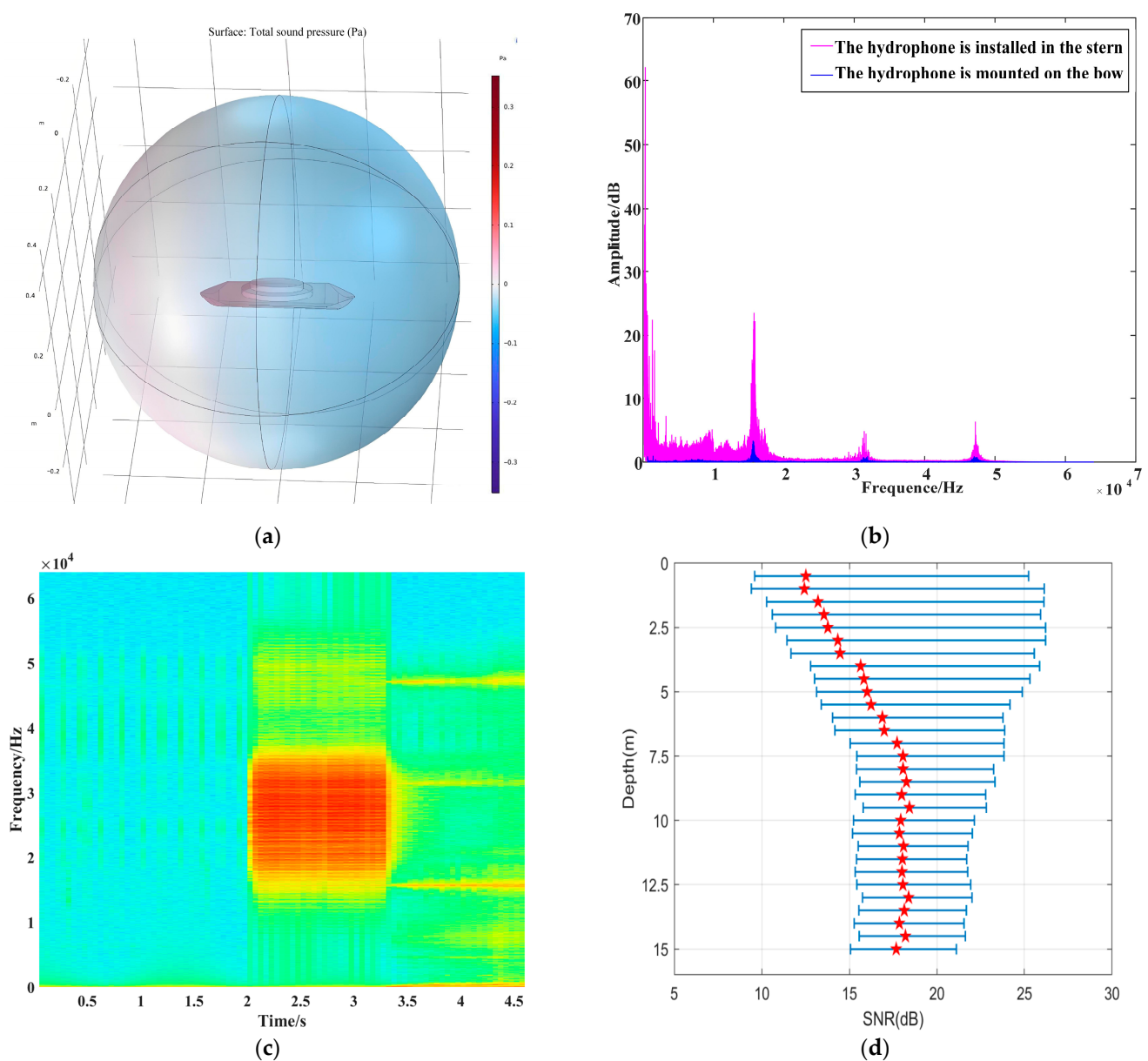


Figure 2. Spatiotemporal characteristic analysis: (a) simulation of the near-field distribution of USV self-noise; (b) comparison of different spatial deployment positions; (c) analysis of different operational states of the USV; (d) effect of spatial depth variation on system performance.

The geographic distribution of USV self-noise, seen in Figure 2b, is examined by positioning hydrophones at the bow and stern of the USV to record the self-noise spectrum features at various spatial positions. The findings reveal that the hydrophone positioned at the stern demonstrates elevated amplitude responses throughout all frequency bands, notably with pronounced spectral peak improvements at approximately 1.5×10^4 Hz, 3.2×10^4 Hz, and 4.7×10^4 Hz. This indicates the substantial impact of propeller function and localized backflow disruptions in the stern area. Conversely, the noise energy distribution near the front of the USV exhibits greater stability and a generally reduced amplitude, largely attributable to the characteristics of the USV's propulsion system. The research findings demonstrate that the spatial distribution of USV self-noise reveals certain regularities, with various places exhibiting distinct noise patterns. Regions adjacent to the vessel's propulsion system are generally linked to elevated self-noise intensity and more prominent structural characteristics. The intense and organized background noise markedly diminishes the detectability and distinguishability of external sound source signals, thereby impairing the target recognition and information extraction functions of underwater acoustic systems.

Conversely, Figure 2c further depicts the time-frequency spectrogram of USV self-noise in both stationary and dynamic conditions. During the initial 2 s, while the USV remains stationary, the noise frequency predominantly resides in the low-frequency spectrum with comparatively modest energy levels. Upon the initiation of USV, a distinct region of energy amplification is evident in the figure, especially within the frequency range of 1.5×10^4 – 3.7×10^4 Hz, characterized by a significant concentration of energy. This signifies that considerable self-noise is produced at startup, and a high-frequency tail lingers even post-shutdown. The results indicate that during the operation of the USV, its self-noise generates structural noise effects on high-frequency signals. The non-stationary and time-varying features of noise present significant obstacles for subsequent acoustic signal processing.

The examination of signal-to-noise ratio (SNR) variations in Figure 2d indicates considerable oscillations in SNR within the 0–3 m depth range, with an average value below 15 dB and a substantial error margin. This suggests that the self-noise strength is elevated in the near-surface region, accompanied by intricate background interference, resulting in inconsistent signal quality. As the depth increases, the SNR gradually stabilizes, particularly in the 8–13 m depth range, where the average SNR rises to 18–20 dB and shows a relatively convergent trend. This suggests that the mid-water layer provides better acoustic observation conditions in this measurement environment. However, at depths approaching 15 m, the SNR slightly decreases, which may be attributed to enhanced seabed reflection and increased background noise interference.

The systematic collection and analysis of self-noise across various deployment sites, depths, and operational states of the USV elucidate the complexity and pronounced non-stationary characteristics of self-noise in the spatiotemporal domain. The analytical results support the spatial variety and temporal evolution of USV self-noise, while also offering clear direction for developing further self-noise detection models and extracting critical characteristics. This establishes a basis for the formulation of targeted denoising techniques.

3. Model Formulation

To satisfy the real-time and computational requirements for on-board deployment of USVs, we utilize Ultralytics' lightweight object-detection framework YOLOv11 as the foundational implementation, which provides a minimal parameter footprint and reduced inference latency [42]. Nonetheless, YOLOv11 serves merely as a tool; the primary contributions of this section include: for the first time, explicitly mapping USV self-noise

into the DD domain to enhance separability amidst multipath, delay spread, and Doppler effects; developing a DD-YOLO framework that utilizes transfer learning to adapt a visual detection network for underwater acoustic self-noise recognition; and introducing an OT-FROGS denoising model that combines structural sparsity with non-convex regularization to improve noise suppression in intricate environments.

This study analyzes USV self-noise mechanisms and constructs simulation environments, revealing the distribution patterns and acoustic characteristics of self-noise in the spatiotemporal domain, thereby offering essential data for accurate feature recognition and effective noise reduction. Characteristic analysis alone is inadequate to fulfill the recognition and suppression demands in intricate noise settings, hence demanding the development of recognition and denoising models equipped with time-frequency modeling capabilities. This paper offers a YOLOv11 recognition model utilizing DD domain features to successfully extract critical auditory characteristics of self-noise, therefore greatly enhancing recognition accuracy. A denoising model utilizing overlapping group shrinkage is devised, incorporating structural sparsity and time-frequency regularization. This model executes advanced noise suppression according to recognition outcomes, effectively diminishing noise interference on system performance. The joint use of these two models offers substantial technical support for improving the stability and stealth of USVs in intricate acoustic environments, possessing considerable theoretical and practical significance for engineering applications.

3.1. YOLOv11-Based Recognition Model Incorporating DD-Domain Features

Due to the high autonomy and system integration of USVs, there are still certain limitations in terms of system interoperability, real-time data processing capabilities, and data storage. On one hand, current processors face performance constraints when handling complex or time-sensitive tasks, leading to computational delays. On the other hand, the high computational load brought by complex algorithms further exacerbates this issue. To address these problems, this study selects the YOLOv11 neural network, which is lightweight, highly accurate, and easy to integrate, as the target recognition model to enhance the platform's real-time processing capabilities.

Furthermore, in high-velocity motion circumstances, the USV self-noise signal generally displays attributes including multipath propagation, multi-source interference, and Doppler effects. This paper presents a DD transformation approach to efficiently extract pertinent information by mapping the signal to a two-dimensional DD domain. This method concurrently acquires the propagation delay and relative velocity data of the noise sources. The resultant transformation is subsequently utilized as input for the YOLOv11 network, augmenting the model's capacity to assess the propagation pathways and motion properties of noise sources, thus facilitating the effective identification of various noise source components.

3.1.1. Feature Extraction in the DD Domain

The DD domain seamlessly integrates the time-domain and frequency-domain attributes of the signal by creating a two-dimensional grid along the delay and Doppler axes, facilitating a planar representation of the signal with physical significance. This two-dimensional mapping improves the stability and discernibility of signals in dynamic situations [24]. The noise produced during USV operation is affected by the relative velocity between the platform and the receiver, usually resulting in considerable Doppler frequency changes. Examining signals in the DD domain effectively captures the frequency shift phenomena and elucidates the distinct reactions of different noise sources concerning delay and frequency shift. Propeller noise and voice-type interference display unique

distribution properties in the DD map, attributable to their differing physical causes, hence offering independent feature inputs and classification criteria for subsequent deep learning recognition models.

Self-noise signals can be mapped to the DD domain through the transformation relationship between the time-frequency domain and the DD domain. Specifically, self-noise can be transformed between time-frequency domain signal $X[n, m]$ and DD domain signal $x[k, l]$ using the symplectic finite Fourier transform (SFFT) and the inverse symplectic finite Fourier transform (ISFFT). The specific formulas are as follows:

$$x(p, j) = \frac{1}{\sqrt{PQ}} \sum_{n=0}^{P-1} \sum_{m=0}^{Q-1} X(l, k) e^{-j2\pi(\frac{qk}{Q} - \frac{jl}{P})} \quad (2)$$

$$X(l, k) = \frac{1}{\sqrt{PQ}} \sum_{n=0}^{Q-1} \sum_{m=0}^{P-1} x(p, j) e^{j2\pi(\frac{pk}{Q} - \frac{jl}{P})} \quad (3)$$

where $X(l, k)$ denotes the time-domain signal subsequent to Fourier transformation and sampling; m and n signify the quantities of time delay and Doppler dimensions, respectively; and $0 \leq p \leq P, 0 \leq j \leq Q$.

The implementation process is shown in Figure 3. According to Formulas (2) and (3), the time-frequency domain matrix (with time interval $1/T$ and frequency interval Δf) can be transformed into the DD domain (with delay interval $1/N\Delta f$ and Doppler interval $1/NT$). This transformation allows the signal to be mapped into the DD domain. Furthermore, as shown in Figure 4, the amplitude distribution characteristics of the signal in the DD domain are presented. Notable energy peaks are observed in the figure, with the signal in the DD domain exhibiting typical multipath delays and frequency shift characteristics.

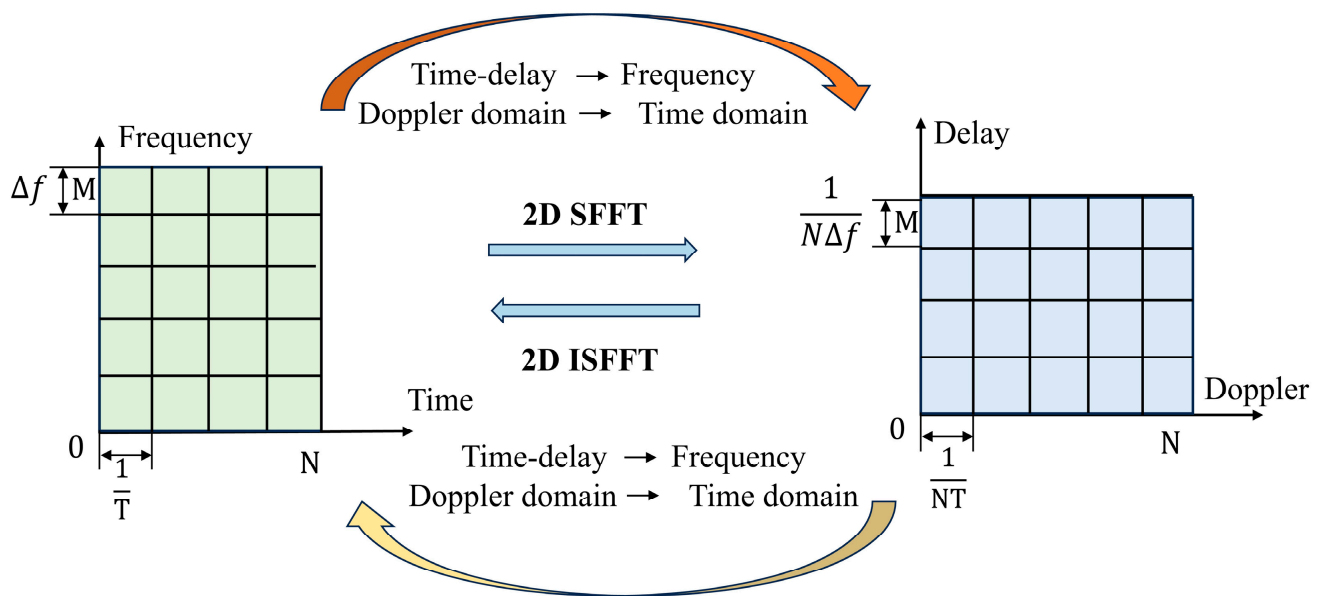


Figure 3. DD domain implementation.

3.1.2. Introduction to the YOLOv11 Model

YOLOv11, an effective real-time object identification algorithm, possesses a streamlined model structure and an optimized backbone framework. In contrast to Darknet53 utilized in YOLOv7 and YOLOv8, YOLOv11 incorporates the CSPDarknet network design, providing enhanced efficiency and more thorough feature extraction. This study employs

the lightweight YOLOv11 model due to the minimal complexity needs of small platforms such as the USV. The YOLOv11n architecture comprises four essential components: input, backbone, neck, and decoupled head.

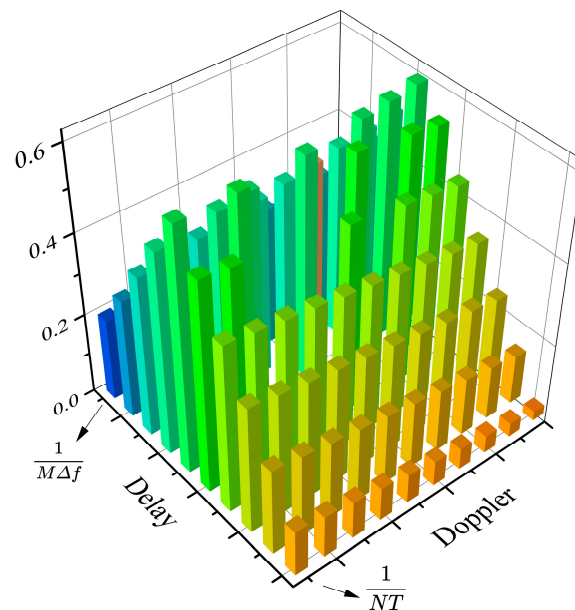


Figure 4. Visualization of signals in the DD Domain.

Meanwhile, YOLOv11, created by Ultralytics, represents the most recent iteration in the YOLO series. Figure 5 illustrates that the backbone network is a crucial element of YOLOv11, tasked with extracting multi-scale information from the input images. This module employs C3k2 as its foundational framework, enhancing feature extraction through the stacking of convolutional layers and modules to produce feature maps at varying resolutions. Moreover, YOLOv11 incorporates the SPPF module from YOLOv8, with the addition of a C2PSA module subsequent to the SPPF to augment the model's spatial attention skills.

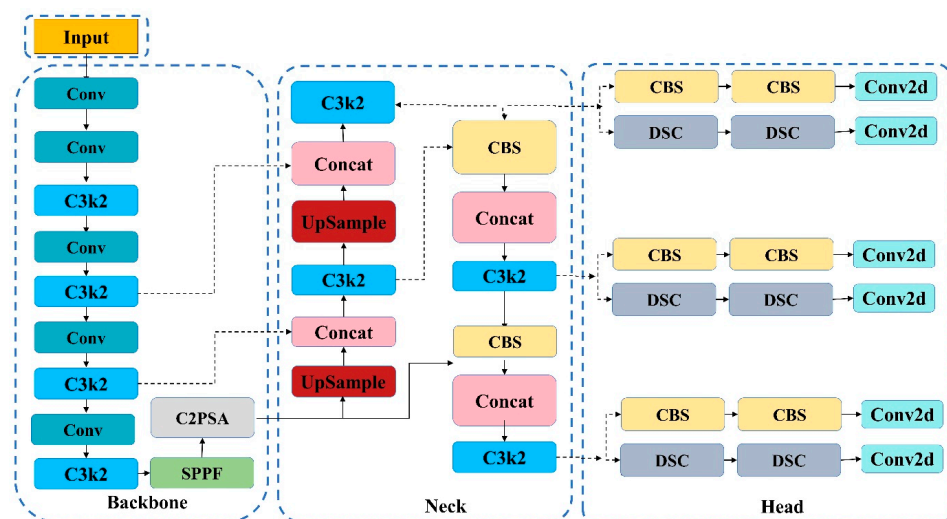


Figure 5. Architecture of the YOLOv11 model.

The neck, situated between the backbone and the output layers, is crucial for feature fusion and enhancement. This section employs the SPPF module to optimize performance by more effectively capturing items of varying sizes, therefore enhancing the detection of

small objects. These modifications and enhancements to the framework enable YOLOv11 to improve performance without compromising speed.

In the training process, we utilized the conventional YOLO-family losses for bounding-box regression, objectness (confidence), and classification, and selected mAP as the primary assessment metric [43,44].

3.2. Optimized Time-Frequency Regularized Overlapping Group Shrinkage Denoising Model

In intricate marine environments, utilizing a USV as a platform for underwater acoustic communication signal transmission and reception, along with essential marine information collection, necessitates the effective reduction of self-noise when significant self-noise components are identified in the received signal. This is crucial for obtaining the target signal. This paper presents an enhanced time-frequency regularized OTFROGS based on the investigation and identification of USV self-noise characteristics. This approach incorporates sparse representation optimization into the TFROGS framework to address the spatiotemporal non-stationary attributes of self-noise. In contrast to traditional normalized least-mean-squares (NLMS), wavelet thresholding (WTD), and the original TFROGS, OTFROGS integrates structural sparsity and non-convex regularization into its modeling, rendering it more adept at addressing the time-frequency-space non-stationarity of USV self-noise.

The algorithm is delineated as follows:

The noisy signal $y(t)$ is composed of the original signal $x(t)$ and self-noise $n(t)$:

$$y(t) = x(t) + n(t) \quad (4)$$

The essence of this model is to recover a denoised signal from $y(t)$ that is as close as possible to the original signal $x(t)$. To obtain this signal, the short-time Fourier transform (STFT) is applied to convert the noise signal $y(t)$ into a time-frequency domain signal $Y(\tau, f)$, as shown in Formula (9).

$$Y(\tau, f) = \int_{-\infty}^{+\infty} y(t)w(t - \tau)e^{-j2\pi ft} dt \quad (5)$$

where $w(t)$ represents the window function, τ denotes the time shift, and f represents the frequency.

Furthermore, the sparse matrix $\mathbf{A}$ of the signal $Y(\tau, f)$ can be represented as:

$$\mathbf{A} = \Phi \mathbf{S} \quad (6)$$

where Φ represents the dictionary matrix, and $\mathbf{S}$ is the sparse matrix, representing the coefficients of the signal in the time-frequency basis.

The majority of items will approximate 0, with only a limited number of elements being substantial in the time-frequency domain. A non-convex overlapping group regularization approach is employed to foster organized sparsity and improve the suppression of correlated self-noise while maintaining prominent components. The sparse matrix is approximated by minimizing:

$$J(\mathbf{S}) = \frac{1}{2} \|\mathbf{Y} - \Phi \mathbf{S}\|_F^2 + \lambda \sum_{g \in G} (\|\mathbf{S}_g\|_2^2 + \varepsilon)^{\frac{p}{2}}, 0 < p < 1 \quad (7)$$

where G denotes the set of overlapping time-frequency groups, $\mathbf{S}_g$ is the sub-vector of coefficients belonging to group g , ε is a small positive constant ensuring stability, and

$p = 0.5$ is employed in the experiments. The non-convex ℓ_p -type group penalty enforces stronger sparsity while retaining local structure.

The Frobenius norm component is given as:

$$\|\mathbf{Y} - \Phi \mathbf{S}\|_F^2 = \sum_{\tau, f} |\mathbf{Y}(\tau, f) - (\Phi \mathbf{S})(\tau, f)|^2 \quad (8)$$

while the group penalty is explicitly expressed as:

$$\Omega(\mathbf{S}) = \sum_{g \in G} (\|\mathbf{S}_g\|_2^2 + \varepsilon)^{\frac{p}{2}} \quad (9)$$

To solve Equation (6), a reweighted iterative scheme is applied. At iteration k , each group is assigned a weight

$$\alpha_g^k = \frac{p}{2} (\|\mathbf{S}_g^{(k)}\|_2^2 + \varepsilon)^{\frac{p}{2}-1} \quad (10)$$

The update of $\mathbf{S}$ is then formulated as:

$$\mathbf{S}^{k+1} = \frac{\mathbf{Y}}{1 + \lambda \cdot \text{Cov}_f(\text{Conv}_t(R^{(k)}))} \quad (11)$$

where

$$R^{(k)}(\tau, f) = \sum_{g \in (\tau, f)} \alpha_g^{(k)} \quad (12)$$

denotes the total weights of all overlapping groups encompassing coefficient (τ, f) . This represents the overlapping group shrinkage (OGS) operator, which adaptively mitigates redundant noise while maintaining structured signal components.

A normalized weight map is then defined as:

$$\mathbf{W}(\tau, f) = \frac{R^{(k)}(\tau, f)}{\max_{\tau, f} R^{(k)}(\tau, f)} \quad (13)$$

It guarantees increased penalties for coefficients situated within robust noise clusters, while allocating comparatively diminished penalties to coefficients associated with major signal structures. Upon completing adequate iterations, the denoised sparse matrix $\mathbf{S}$ is acquired. The resulting denoised signal is reconstructed using the inverse short-time Fourier transform (ISTFT):

$$x_{\text{denoised}}(t) = \text{ISTFT}(\mathbf{S}) \quad (14)$$

In summary, the detailed procedure of OTFROGS is presented in Table 2:

Table 2. Steps of OTFROGS algorithm.

Input: observation signal $y(t)$, measurement matrix Φ , and convergence criterion σ
Output: reconstructed signal $x(t)$
1. Initialize the parameters.
2. Compute the time-frequency representation $\mathbf{Y}(\tau, f)$ of the signal using Equation (4).
3. Update the reconstruction cost function $J(\mathbf{S})$ according to Equation (7).
4. Update $\mathbf{S}^{k+1}$ and $\mathbf{W}(\tau, f)$ using Equation (11) and Equation (13), respectively, corresponding to the non-convex overlapping group shrinkage operator.
5. Termination condition: if the change in the cost function falls below the predefined threshold σ , stop the iteration and output the reconstructed signal; otherwise, return to Step 2 and continue iterating.

4. Experiment Results

This study utilized two datasets regarding data sources and theoretical justification. Dataset-A consisted of a historical 5×32 hydrophone array with 5 m spacing between components, utilized to examine the spatial distribution of USV self-noise across varying depths, locations, and operational states. This dataset established a robust basis for the theoretical examination of three-dimensional coupling effects in the time-frequency-space domain; nevertheless, its restricted scale rendered it inadequate for extensive deep model training and validation. Consequently, Dataset-B was acquired in near-shore waters off Qingdao with a self-contained hydrophone. To provide practical deployment and operational flexibility, a single receiver was utilized, and recordings were made at various depths to somewhat mitigate the lack of array data. It is important to acknowledge that the two datasets were not gathered during the same experimental effort, and the vessels were not identical. Nonetheless, both platforms exhibited analogous hull dimensions and propeller diameters, with measurements taken under identical tranquil near-shore water conditions. Consequently, Dataset-A guarantees the dependability of spatial pattern analysis in the theoretical segment, whilst Dataset-B supplies adequate data for the systematic training and assessment of the DD-YOLO recognition model and the OTFROGS denoising algorithm. The comparability of vessel scale, propeller configuration, and sea conditions guarantees the applicability of Dataset-B to the spatial characteristics specified in Dataset-A. Furthermore, as the self-noise signals were represented in the DD domain, which intrinsically demonstrates sparsity and separability of multipath and Doppler components, the utilization of single-channel hydrophone data remains applicable. It is important to note that the existing recognition and denoising experiments do not fully utilize spatial-domain diversity; subsequent research will augment the dataset by integrating multi-hydrophone recordings to improve the generality of the suggested models.

Furthermore, each continuous recording (defined as a session) was taken as the smallest partitioning unit, with each session collected under fixed ship speed and engine RPM conditions to ensure consistent operating states. During preprocessing, the raw audio was segmented into 3-second samples for subsequent training and evaluation. To avoid data leakage, all samples were first grouped by session and then split at the session level into training, validation, and test sets in a 10:3:1 ratio. This ensures that all 3-second samples from the same session are contained within a single subset, thereby enabling fair and independent performance evaluation.

This work gathered propeller noise, a significant element of USV self-noise, using the experimental setup depicted in Figure 6. Figure 6a illustrates the deployment of a USV in near-shore waters off Qingdao, where its near-field self-noise was captured using an integrated hydrophone. The comprehensive experimental parameters are delineated in Figure 6b, where L and H denote the hydrophone depth and the distance from the vessel's keel to the bottom, respectively. For ease of deployment and operational flexibility, the receiving system employed a single self-contained hydrophone.

4.1. Evaluation of Recognition Model Performance

This section provides a comprehensive assessment of the operational efficacy of the YOLOv11 identification model—trained on DD domain characteristics—in detecting USV self-noise. Figure 7 outlines the preprocessing workflow used for the self-noise recordings: routine electrical interference is first removed from the raw acoustic data to suppress background clutter, after which Equations (2) and (3) are applied to transform the self-noise from the time-frequency (TF) domain into the delay-Doppler domain. This conversion enables extraction of the embedded propagation-delay and Doppler-shift signatures, as illustrated in Figure 8.

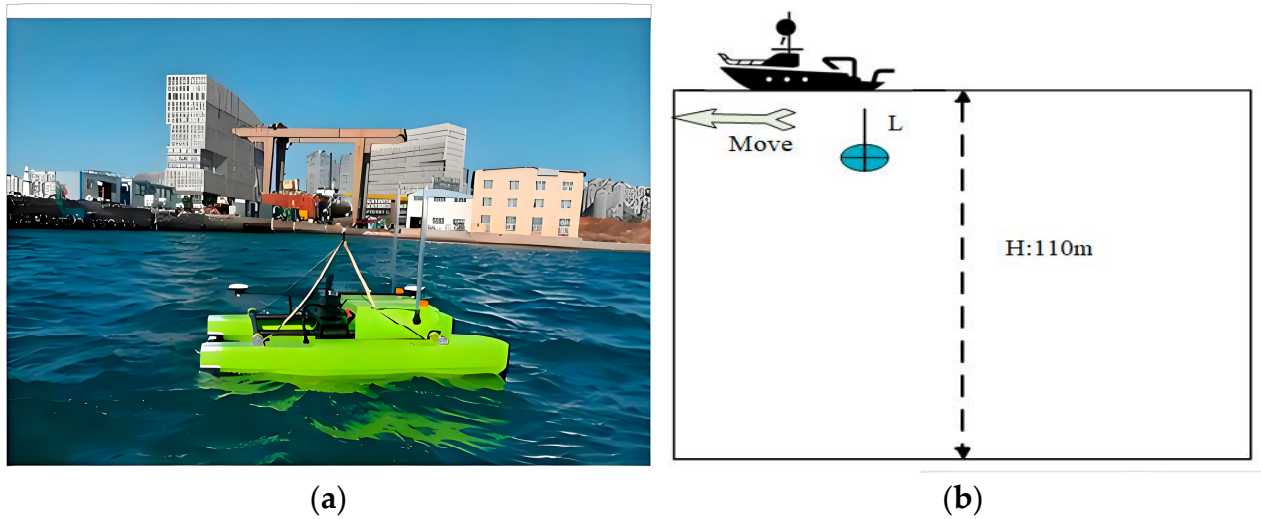


Figure 6. Data collection for experiments: (a) real-world experimental scenarios; (b) experimental environment and settings.

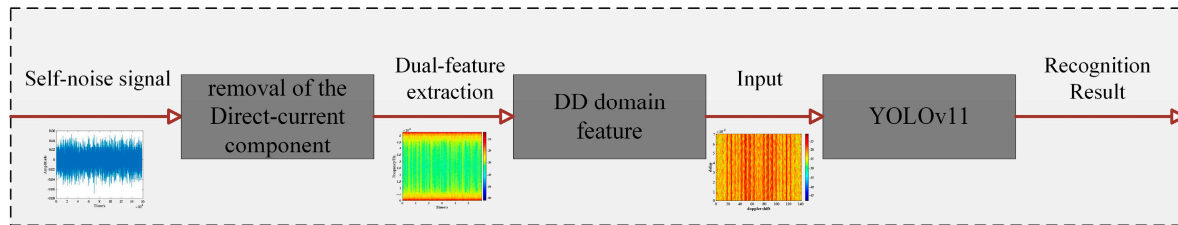


Figure 7. DD domain processing of self-noise signals.

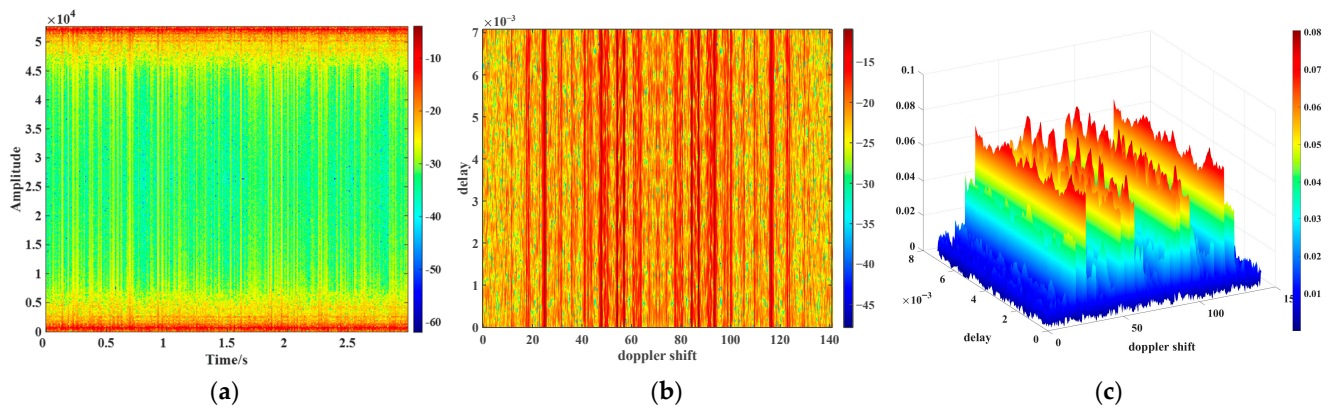


Figure 8. Time-frequency domain and DD domain: (a) time-frequency domain; (b) DD domain signal response; (c) 3D-DD domain signal response.

To enhance alignment with practical experimental circumstances, the detection task utilizes a singular category—propeller self-noise. This category is not further split into cavitation or mechanical noise components, as these elements frequently co-occur and are generally interrelated under small-USV operating conditions; additionally, in the DD domain, both typically manifest as concentrated energy peaks or stripe-like formations. Joint modeling is hence more coherent and operationally feasible. This study split raw acoustic recordings with self-noise into 3-second samples, resulting in a total of 1735 samples; the dataset was divided into training, validation, and test sets in a ratio of 10:3:1. Signal processing adhered to an SFFT/ISFFT architecture, incorporating parameters such as window

length, hop size, and FFT points (NFFT); each segment generated a DD energy map with a constant grid resolution of $M \times N = 374 \times 273$. The annotation protocol was executed by trained experimenters following standardized procedures: raw acoustic data were initially transformed into DD energy maps, after which annotators visually identified regions of concentrated energy and delineated axis-aligned bounding boxes, uniformly labeled “propeller-noise”. To guarantee annotation quality and consistency, each sample was independently labeled by a minimum of two experimenters, followed by cross-verification and consensus resolution of any discrepancies.

Figure 8a illustrates that the time-frequency (TF) spectrogram displays the temporal-spectral evolution of signal energy; however, due to the compounded effects of various noise sources and multipath phenomena, the features are densely clustered with indistinct contours, resulting in considerable overlap among signal components, which complicates effective feature separation and identification. Conversely, the DD-domain energy maps in Figure 8b,c display several discrete energy peaks linked to different propagation paths and Doppler components. The distinct structural and geographical qualities provide YOLOv11 with a considerable advantage in performing target detection and multi-source recognition tasks. The DD map is annotated with bounding boxes to enclose the peak or stripe-like regions associated with propeller self-noise in the succeeding training photos, following the aforementioned procedures.

To objectively assess the efficacy of the DD representation, we performed an ablation experiment comparing the detection performance of identical YOLO models utilizing standard TF spectrograms and DD energy maps as inputs. Table 3 demonstrates that DD inputs regularly yield superior mAP values, with YOLOv11 increasing from 0.67 (TF) to 0.87 (DD) and YOLOv8 from 0.64 (TF) to 0.82 (DD). Comparable benefits are noted in AP@[.5:.95], recall, and F1-score. For example, YOLOv11 achieves an F1 score of 0.86 with DD input, in contrast to 0.63 with TF input, and YOLOv8 improves from 0.60 (TF) to 0.81 (DD). The results demonstrate that the more distinct structural cues offered by DD features improve visual separability and lead to significant enhancements in quantitative performance, hence corroborating the findings in Figure 8.

Table 3. YOLO performance on TF vs. DD representations.

Model	mAP@0.5	AP@[.5:.95]	Precision	Recall	F1 (at IoU = 0.5, Best-thr)
DD-YOLOv11	0.87	0.59	0.89	0.84	0.86
TF-YOLOv11	0.67	0.50	0.65	0.61	0.63
DD-YOLOv8	0.82	0.50	0.83	0.80	0.81
TF-YOLOv8	0.64	0.47	0.63	0.58	0.60

A systematic performance comparison is thereafter performed comparing YOLOv11, its predecessor YOLOv8, and the traditional SSD neural network. The findings are depicted in Figures 9 and 10:

Figures 9 and 10 compare the training and validation dynamics of YOLOv11 with YOLOv8 regarding loss components and detection performance metrics (mAP50 and mAP50-95). Overall, YOLOv11 demonstrates enhanced efficacy in loss convergence. Both box_loss and cls_loss diminish more swiftly during training, whilst the associated validation losses not only converge more rapidly but also stabilize at lower values, exhibiting significantly reduced fluctuations. The results indicate that YOLOv11 exhibits superior training stability and generalization capacity. Conversely, YOLOv8 exhibits more significant oscillations and elevated residuals in the loss curves, signifying inadequate convergence behavior and diminished robustness.

Figure 11 illustrates that the PR curve of YOLOv11 continuously surpasses that of YOLOv8, exhibiting good precision even when recall exceeds 0.7, hence indicating superior robustness. Conversely, YOLOv8 demonstrates a more rapid decrease in precision as recall escalates. The AUC-PR results further validate that YOLOv11 attains a superior precision/recall trade-off, rendering it more appropriate for situations necessitating both high confidence and high recall.

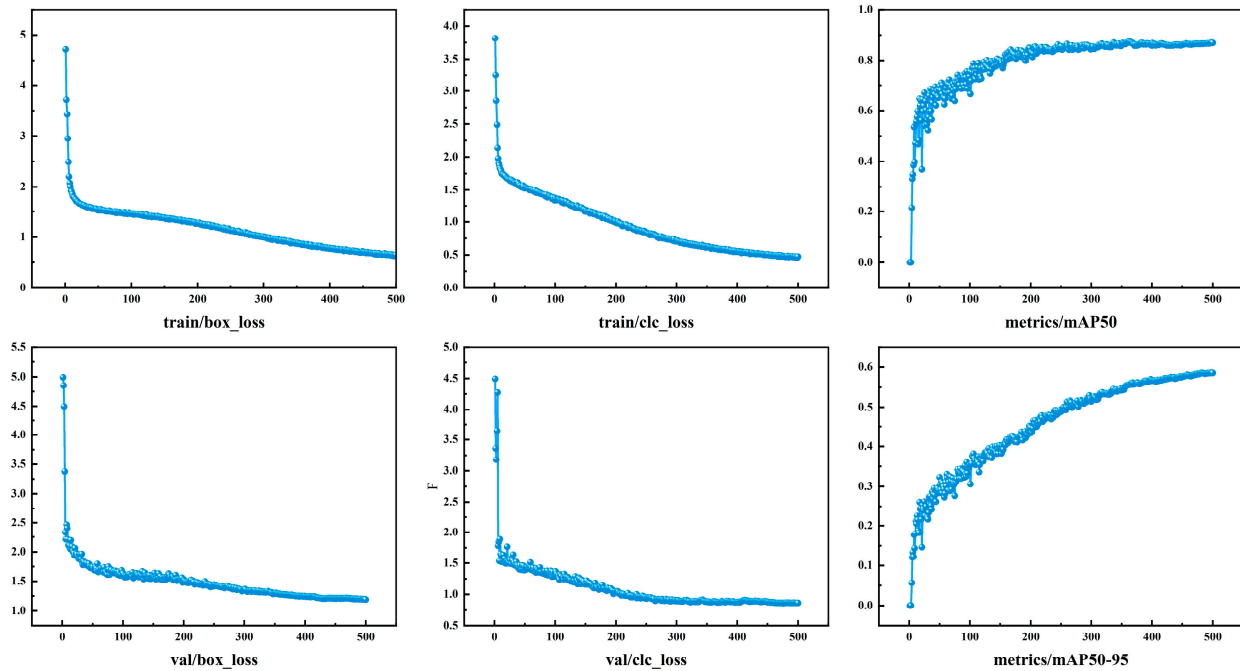


Figure 9. Evaluation of YOLOv11 performance.

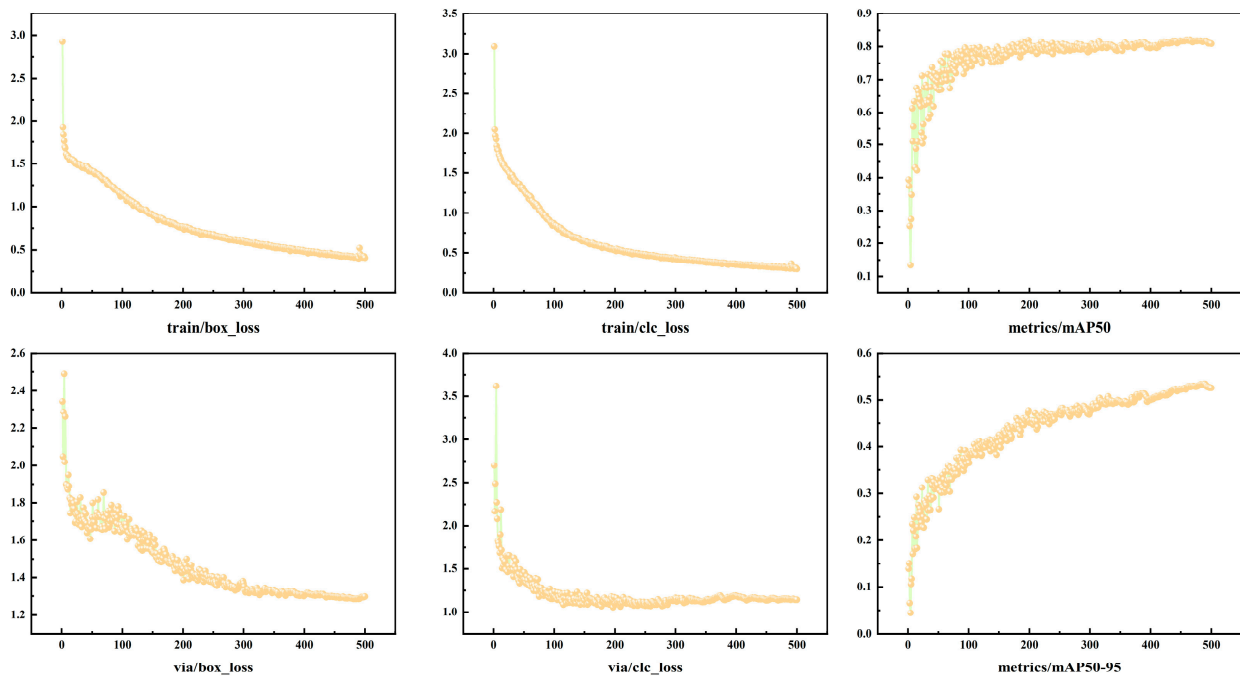


Figure 10. Evaluation of YOLOv8 performance.

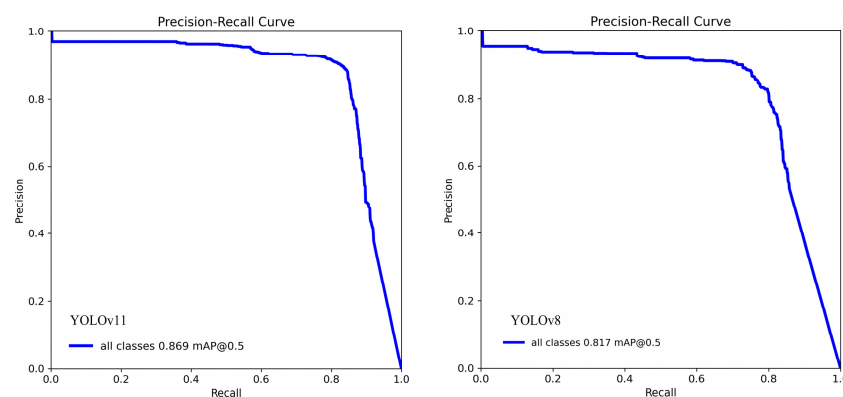


Figure 11. PR curves of YOLOv11 and YOLOv8.

As indicated by the detection performance comparison in Table 4, YOLOv11 attains superior results across all metrics, with an AP@0.5 of 0.87, an AP@[.5:.95] of 0.59, a precision of 0.89, a recall of 0.84, and an F1-score of 0.86, illustrating a well-balanced trade-off between precision and recall, alongside enhanced robustness across varying IoU thresholds. YOLOv8 achieves an AP@0.5 of 0.82, an AP@[.5:.95] of 0.50, a precision of 0.83, a recall of 0.80, and an F1-score of 0.81, securing second place overall while preserving high detection accuracy. Conversely, SSD achieves an AP@0.5 of 0.78, exhibiting a high precision of 0.90, yet a recall of merely 0.37, culminating in an F1-score of only 0.52. This indicates an excessively cautious detection behavior, resulting in the model generating fewer predictions that are predominantly accurate, but at the cost of recall. YOLOv11 markedly surpasses YOLOv8 and SSD in detection accuracy and resilience, establishing it as a more dependable option for practical applications necessitating high-confidence target identification.

Table 4. Model Performance Comparison.

Model	mAP@0.5	AP@[.5:.95]	Precision	Recall	F1 (at IoU = 0.5, Best-thr)
YOLOv11	0.87	0.59	0.89	0.84	0.86
YOLOv8	0.82	0.50	0.83	0.80	0.81
SSD	0.78	\	0.90	0.37	0.52

4.2. Denoising Model

This research presents an OTFROGS denoising model. The model presents a coefficient-matrix representation designed for self-noise signal characteristics; by utilizing sparse representation, it retains the principal frequency components while effectively eliminating the USV self-noise, particularly the low-energy components, leading to efficient and precise denoising performance. The suggested method's validity is confirmed using experimentally obtained signals that encompass significant levels of USV self-noise.

4.2.1. Performance Comparison Under Simulated Conditions

This section presents the application of the proposed OTFROGS approach alongside several traditional denoising techniques to signals containing USV self-noise. As seen in Figure 12, the OTFROGS approach displays enhanced denoising efficacy, proficiently removing the majority of noise while accurately reconstructing the target signal. Conversely, conventional denoising methods—such as NLMS, WTD, and TFROGS—attain merely partial noise reduction and preserve discernible residual noise. To elucidate the performance disparities among the algorithms more intuitively, the SNR is shown in absolute terms; hence, a lower number signifies superior denoising efficacy. As illustrated in Figure 13, all methods yield negative SNR values under -15 dB and -25 dB noise conditions. For

clarity and ease of comparison, the absolute values of SNR are reported. The proposed OTFROGS technique achieves the lowest $|\text{SNR}|$ values of 0.99 and 7.85 in the respective noise environments, significantly outperforming traditional denoising algorithms such as NLMS, WTD, and TFROGS. These results underscore the superior noise suppression capability of OTFROGS in challenging acoustic environments.

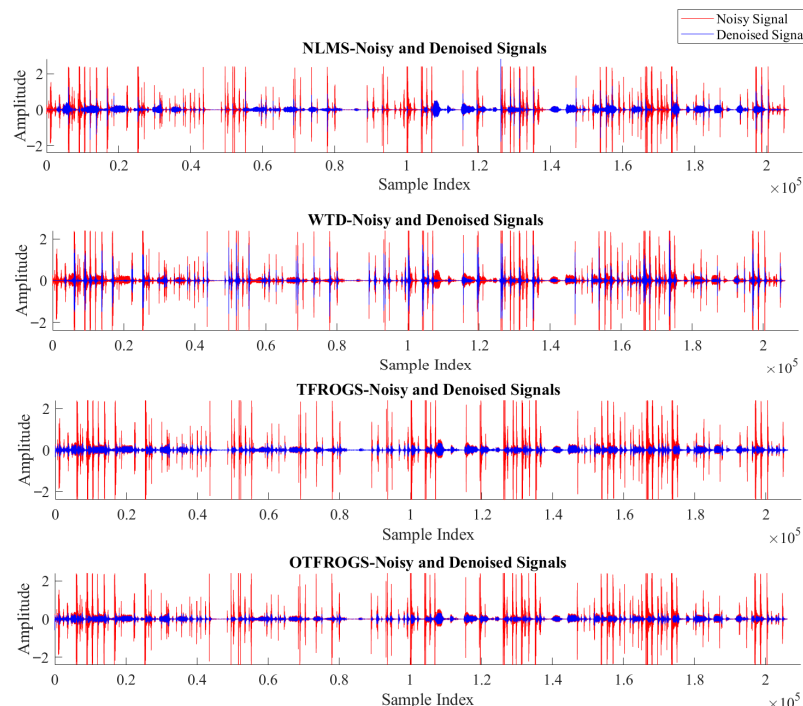


Figure 12. Comparison of noisy and denoised signal waveforms in the time domain.

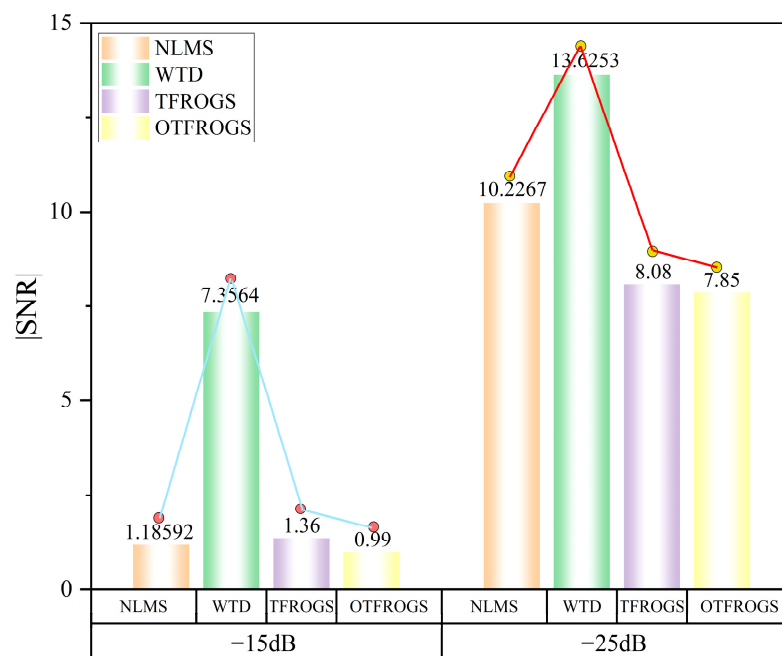


Figure 13. Assessment of denoising performance under different SNR scenarios.

4.2.2. Denoising Performance on Real Self-Noise Data

This section illustrates the denoising efficacy of OTFROGS on acoustic signals with USV self-noise; the relevant results are displayed in Figures 14 and 15.

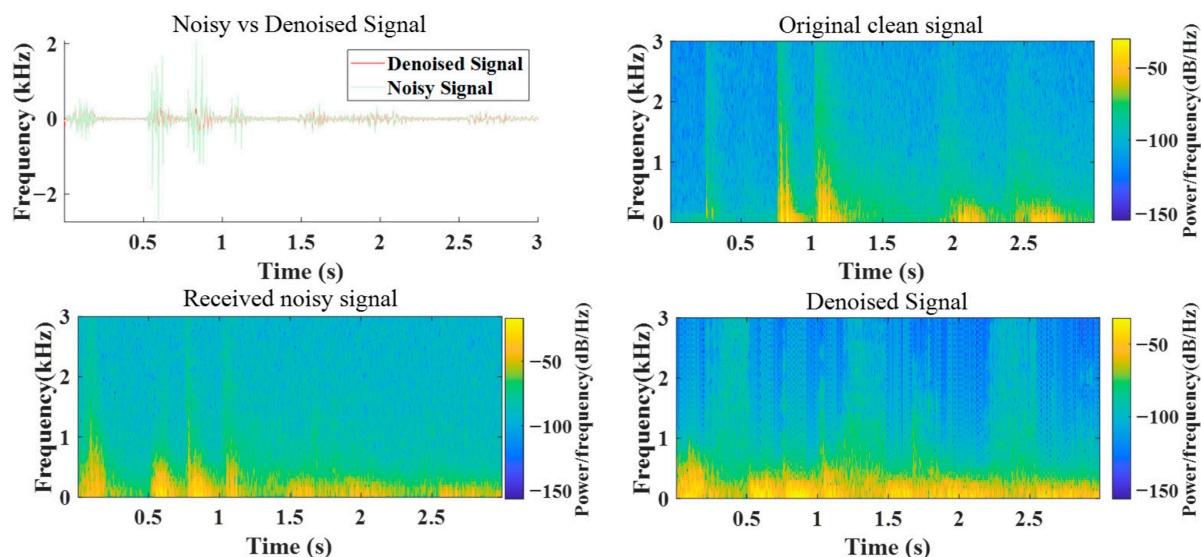


Figure 14. Denoising effectiveness for real signals in high-SNR environments.

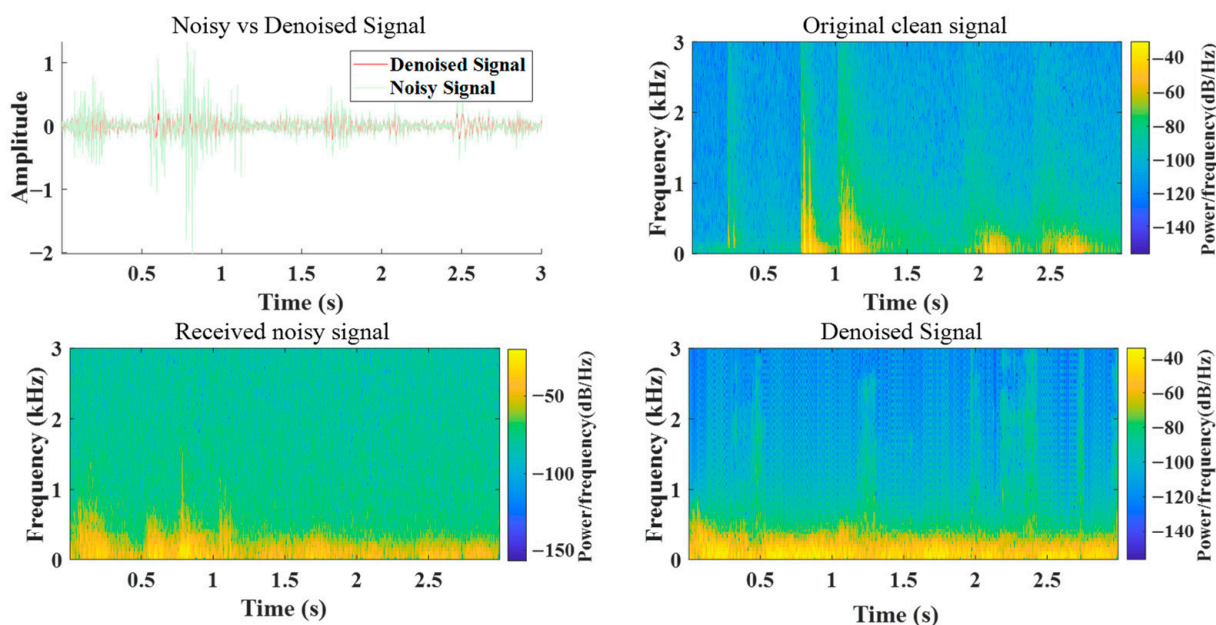


Figure 15. Evaluation of denoising results on real-world signals with low SNR.

Figure 14 represents the high-SNR scenario, wherein a distinct differentiation is seen between the denoised signal and the noise-affected input. The self-noise's high- and low-frequency components are successfully mitigated. Conversely, in the low-SNR conditions depicted in Figure 15, despite the presence of some residual low-frequency noise in the denoised output, the overall noise reduction is significant.

In addition to the denoising efficacy demonstrated in Figures 14 and 15, the computational expense of the two modules was assessed. On a standard desktop CPU, DD-YOLOv11 necessitates 43.67 ms per DD feature map with a memory use of 1.64 GB, whereas the OTFROGS module processes 1 s of acoustic data in approximately 0.03 s. These measures suggest capacity for near-real-time functionality. To assess the practicality of compact USVs, previous studies indicate that lightweight YOLO models generally get 20–30 FPS with less than 1 GB of memory on the Jetson Xavier NX and 8–15 FPS on the Jetson Nano [18,45]. Furthermore, GPU-based FFT/IFFT implementations typically provide 1.5–3 $\times$ acceleration, suggesting additional speed enhancements for OTFROGS on embedded GPUs [46,47]. The

results demonstrate compatibility with edge-class CPU/GPU devices often incorporated in small USVs.

5. Conclusions and Discussion

5.1. Conclusions

This work examines self-noise interference in USVs and presents a comprehensive solution—feature modeling, noise identification, and denoising—whose efficacy is experimentally proven. A self-noise propagation model is initially developed under spatiotemporal fluctuation conditions, clarifying the coupling properties of self-noise within the time-frequency-space domain and offering a theoretical foundation for subsequent signal processing. Secondly, in the identification phase, YOLOv11 is tailored for the DD domain, facilitating accurate detection of non-stationary, multi-source interference noise and attaining an average accuracy of 87.0%, markedly surpassing YOLOv8 and SSD. The proposed OTFROGS denoising model incorporates structural sparsity and non-convex regularization to efficiently diminish low-energy redundant noise while maintaining essential acoustic characteristics; it outperforms NLMS, wavelet thresholding, and the original TFROGS in both interference suppression and signal quality enhancement. It is important to note that YOLOv11 in this study functions solely as an implementation tool; the advancements are found in the delay-Doppler domain feature modeling, the DD-YOLO framework, and the OTFROGS denoising technique. The technology is generalizable and can be adapted to various detection networks and a wider array of underwater acoustic applications.

The proposed integrated identification–denoising framework markedly enhances the signal quality and stability of USV acoustic systems, while also offering a practical approach for self-noise suppression in intricate marine environments.

5.2. Discussion

5.2.1. Main Results Analysis

This study aims to achieve two objectives: firstly, to characterize USV self-noise in the DD domain, which more effectively distinguishes delay spread and Doppler shifts compared to conventional TF spectrograms; and secondly, to formulate a tailored denoising strategy that leverages structural sparsity in the TF plane to mitigate self-noise while maintaining target signal integrity. The experimental findings corroborate both objectives. Specifically, modifying the YOLOv11 detector for DD feature maps results in enhanced detection performance compared to standard baselines, suggesting that the DD domain presents self-noise as more distinct, learnable, and compact structures, while TF inputs frequently display overlapping textures with indistinct boundaries. The OTFROGS method surpasses NLMS, wavelet thresholding, and the original TFROGS in both simulations and actual recordings, accomplishing non-convex, structure-aware shrinking in the TF plane that more accurately aligns with the morphology of self-noise.

5.2.2. Actual Value and Practical Significance

The results of this study offer practical recommendations for the design of future USV platforms and the placement of sensors. Experimental findings demonstrate that acoustic pressure peaks near the propeller; energy across various frequency bands is more pronounced at the stern than at the bow; and the bow displays generally lower energy levels due to geometric shielding. From an engineering perspective, hydrophone installation should be favored in front of the propeller or at increased depths. Simultaneously, we are developing a noise-recognition model utilizing DD-domain characteristics and an OT-FROGS approach for denoising. From a systemic viewpoint, the close integration of “recognition + denoising” is essential: the recognized self-noise category and its DD-domain

locus can be utilized to control or parameterize the denoiser, facilitating more selective attenuation and consequently enhancing the stability and acoustic stealth of USV acoustic systems in intricate sea conditions.

5.2.3. Limitations and Future Work

This research possesses multiple limitations. The spatial distribution analysis of USV self-noise was performed using a historical array dataset (Dataset-A), whereas the recognition and denoising tests were conducted using a newly acquired single-channel dataset (Dataset-B). The delay-Doppler representation provides intrinsic sparsity that facilitates model training with single-hydrophone data; nevertheless, the absence of comprehensive array information in experimental validation limits the demonstration of spatial-domain generalizability. Secondly, embedded real-time verification is constrained: existing outcomes predominantly stem from offline/desktop assessments, and there is a deficiency in closed-loop integration and sea-trial validation across various low-power hardware platforms.

In subsequent research, we will broaden the trials to include multi-hydrophone and array-based recordings to provide a more thorough validation of the proposed DD-YOLO and OT-FROGS frameworks across temporal, frequency, and spatial dimensions. Simultaneously, across diverse sea states and kinematic conditions, we will augment mAP/recall with downstream task-level metrics—such as communication reliability and detection range—to determine if the characterization of motion-induced frequency shifts by DD features results in measurable operational advantages.

Author Contributions: Conceptualization, H.L. and G.W.; validation, X.W. and F.Y.; investigation, Z.L. and G.W.; resources, X.W. and H.L.; writing—original draft preparation, H.L. and G.W.; writing—review and editing, Q.L., F.Y. and G.S.; supervision, Z.L. and G.S.; funding acquisition, Z.L. and G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work has received funding from the National Key Research and Development Program of China (2023YFE0201900); the Underwater Unmanned Platform Acoustic Signal Perception and Early Warning Technology Research Project (SKLA202504); and the Basic and Frontier Exploration Project Independently Deployed by Institute of Acoustics, Chinese Academy of Sciences (JCQY202408).

Data Availability Statement: Restrictions apply to the datasets. The datasets presented in this article are not readily available because the data are part of an ongoing study. Requests for access to the datasets should be directed to the first author, Zhichao Lv (lvzhichao@hrbeu.edu.cn).

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

References

1. Fang, Y.; Huang, Z.; Pu, J.; Zhang, J. AUV position tracking and trajectory control based on fast-deployed deep reinforcement learning method. *Ocean Eng.* **2022**, *245*, 110452. [\[CrossRef\]](#)
2. Ding, J.; Pang, S.; Chen, Z. Optimization of the chamber of OWC to improve hydrodynamic performance. *Ocean Eng.* **2023**, *287*, 115782. [\[CrossRef\]](#)
3. Zhuo, X.; Liu, M.; Wei, Y.; Yu, G.; Qu, F.; Sun, R. AUV-aided energy-efficient data collection in underwater acoustic sensor networks. *IEEE Internet Things J.* **2020**, *7*, 10010–10022. [\[CrossRef\]](#)
4. Qiu, F.; Zhang, W. AUV-aided joint time synchronization and localization of underwater target with propagation speed uncertainties. *Ocean Eng.* **2023**, *283*, 115060. [\[CrossRef\]](#)
5. Chen, H.; Cai, W.; Zhang, M. AUV-aided computing offloading for multi-tier underwater computing: A stackelberg game learning approach. *Ocean Eng.* **2024**, *297*, 117109. [\[CrossRef\]](#)
6. Chettri, A.P.; Narayanamoorthi, R. Comprehensive review of wireless power transfer for autonomous underwater vehicles: Technological innovations, challenges, and future prospects. *e-Prime—Adv. Electr. Eng. Electron. Energy* **2025**, *13*, 101079. [\[CrossRef\]](#)

7. Jalkanen, J.P.; Johansson, L.; Liefvendahl, M.; Bensow, R.; Sigray, P.; Östberg, M.; Karasalo, I.; Andersson, M.; Peltonen, H.; Pajala, J. Pajala, Marine science. *Ocean Sci.* **2018**, *14*, 1373–1383. [\[CrossRef\]](#)
8. Legaz, M.; Sergio, A.; Busquier, S. Marine propulsion shafting: A study of whirling vibrations. *J. Ship Res.* **2021**, *65*, 55–61. [\[CrossRef\]](#)
9. Lei, J.; Zhou, R.; Chen, H.; Gao, Y.; Lai, G. Experimental investigation of effects of ship propulsion shafting alignment on shafting whirling and bearing vibrations. *J. Mar. Sci. Technol.* **2022**, *27*, 151–162. [\[CrossRef\]](#)
10. Bocanegra, J.A.; Borelli, D.; Gaggero, T.; Picó, R.; Tani, G. Acoustic characterization of a cavitation tunnel for ship propeller noise studies. *J. Ocean Eng. Sci.* **2025**, *10*, 330–341. [\[CrossRef\]](#)
11. Zou, D.; Zhang, J.; Ta, N.; Rao, Z. Study on the axial exciting force characteristics of marine propellers considered the effect of the shaft and blade elasticity. *Appl. Ocean Res.* **2019**, *89*, 141–153. [\[CrossRef\]](#)
12. Dong, L.; Zhao, Y.; Dai, C. Detection of inception cavitation in centrifugal pump by fluid-borne noise diagnostic. *Shock Vib.* **2019**, *3*, 1–15. [\[CrossRef\]](#)
13. Donmez, A.H.; Yumurtaci, Z.; Kavurmacioglu, L. The effect of inlet blade angle variation on cavitation performance of a centrifugal pump: A parametric study. *J. Fluids Eng.-Trans. ASME* **2019**, *141*, 021101. [\[CrossRef\]](#)
14. Wei, Z.; Li, X.; Tao, R.; Sun, D.; Xiao, R.; Hu, H. Clocking effect of relative position matching on the hydrodynamic and flow-induced noise characteristics in a centrifugal pump. *Appl. Acoust.* **2025**, *240*, 110899. [\[CrossRef\]](#)
15. Lin, Y.; Li, X.; Li, B.; Jia, X.; Zhu, Z. Influence of impeller sinusoidal tubercle trailing-edge on pressure pulsation in a centrifugal pump at nominal flow rate. *J. Fluids Eng.-Trans. ASME* **2021**, *143*, 493–514. [\[CrossRef\]](#)
16. Lee, S.; Kim, B. Machine learning model for leak detection using water pipeline vibration sensor. *Sensors* **2023**, *23*, 8935. [\[CrossRef\]](#)
17. Sayed, A.N.; Ramahi, O.M.; Shaker, G. Machine learning for UAV classification employing mechanical control information, IEEE Trans. *Aerosp. Electron. Syst.* **2024**, *60*, 68–81. [\[CrossRef\]](#)
18. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
19. YOLOv9: Learning what you want to learn using programmable gradient information. *arXiv* **2025**, arXiv:2402.13616.
20. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
21. Zhou, S.; Yang, L.; Liu, H.; Zhou, C.; Liu, J.; Zhao, S.; Wang, K. A lightweight drone detection method integrated into a linear attention mechanism based on improved YOLOv11. *Remote Sens.* **2025**, *17*, 705. [\[CrossRef\]](#)
22. Nandal, P.; Bohra, N.; Mann, P.; Das, N.N. YOLOv11 with transformer attention for real-time monitoring of ships: A federated learning approach for maritime surveillance. *Results Eng.* **2025**, *27*, 106297. [\[CrossRef\]](#)
23. Zhang, Y.; Wang, Y.; Liu, Y.; Shi, L.; Zang, Y. A deep learning receiver for underwater acoustic OTFS communications with Doppler squint effect. *IEEE Wirel. Commun. Lett.* **2025**, *14*, 1179–1183. [\[CrossRef\]](#)
24. Li, L.; Wang, B.; Huang, Y. A new scheme of low-frequency long-range underwater acoustic communication with high spectral efficiency. *Appl. Acoust.* **2025**, *235*, 110668. [\[CrossRef\]](#)
25. Liu, L.; Ma, C.; Duan, Y.; Liu, X.; Zhang, W. Kernel adaptive filter-based channel prediction for adaptive underwater acoustic OFDMA system. *Appl. Ocean Res.* **2025**, *158*, 104586. [\[CrossRef\]](#)
26. Wang, Q.; Liang, Y.; Zhang, Z.; Fan, P. 2D off-grid decomposition and SBL combination for OTFS channel estimation. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 3084–3098. [\[CrossRef\]](#)
27. Wei, Z.; Yuan, W.; Li, S.; Yuan, J.; Bharatula, G.; Hadani, R. Orthogonal time-frequency space modulation: A promising next-generation waveform. *IEEE Wirel. Commun.* **2021**, *28*, 136–144. [\[CrossRef\]](#)
28. Kumar, A.; Gaur, N.; Nanthamornphong, A. Signal detection of M-MIMO-orthogonal time frequency space modulation using hybrid algorithms: ZFE + MMSE and ZFE + MF. *Results Eng.* **2024**, *24*, 103311. [\[CrossRef\]](#)
29. Hlawatsch, F.; Matz, G. *Wireless Communications over Rapidly Time-Varying Channels*; Academic Press: New York, NY, USA, 2011.
30. Xie, J.; Colonna, J.G.; Zhang, J. Bioacoustic signal denoising: A review. *Artif. Intell. Rev.* **2021**, *54*, 3575–3597. [\[CrossRef\]](#)
31. Li, Y.X.; Wang, L. A novel noise reduction technique for underwater acoustic signals based on complete ensemble empirical mode decomposition with adaptive noise, minimum mean square variance criterion and least mean square adaptive filter. *Def. Technol.* **2020**, *16*, 543–554. [\[CrossRef\]](#)
32. Zhang, J.; Jin, Y.; Sun, B.; Han, Y.; Hong, Y. Study on the improvement of the application of complete ensemble empirical mode decomposition with adaptive noise in hydrology based on RBFNN data extension technology. *CMES-Comp. Model. Eng. Sci.* **2020**, *126*, 755–770. [\[CrossRef\]](#)
33. Babalola, O.P.; Versfeld, J. Wavelet-based feature extraction with hidden Markov model classification of Antarctic blue whale sounds. *Eco. Inform.* **2024**, *80*, 102468. [\[CrossRef\]](#)
34. Evrendilek, F.; Karakaya, N. Regression model-based predictions of diel, diurnal and nocturnal dissolved oxygen dynamics after wavelet denoising of noisy time series. *Phys. A* **2014**, *404*, 8–15. [\[CrossRef\]](#)

35. Hu, H.; Ao, Y.; Yan, H.; Bai, Y.; Shi, N. Signal denoising based on wavelet threshold denoising and optimized variational mode decomposition. *J. Sens.* **2021**, *2021*, 5599096. [[CrossRef](#)]
36. Frusque, G.; Fink, O. Robust time series denoising with learnable wavelet packet transform. *Adv. Eng. Inform.* **2024**, *62*, 102669. [[CrossRef](#)]
37. Huang, B.; Wu, Y.; Lyu, Y.; Yan, X.; Tong, M.; Wang, X. PCA-based denoising and automatic recognition of marine biological sounds to estimate Bio-voice Count Index for marine monitoring. *Ecol. Inform.* **2025**, *90*, 103280. [[CrossRef](#)]
38. Chang, C.I.; Du, Q. Interference and noise-adjusted principal components analysis. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 2387–2396. [[CrossRef](#)]
39. Peter, K.J.; Kannan, K.S.; Arumugan, S.; Nagarajan, G. Two-stage image denoising by Principal Component Analysis with Self Similarity pixel Strategy. *Int. J. Comput. Sci. Netw. Secur.* **2011**, *11*, 296–301.
40. Liu, B.; Lei, J.B. *Principles of Hydroacoustics*, 3rd ed.; Science Press: Beijing, China, 2009.
41. Zhichao, L.; Qi, L.; Ren, C. Acoustic Analysis and Design of Unmanned Ship Underwater Acoustic Communication. In Proceedings of the 2019 Academic Conference of the Chinese Society of Acoustics Underwater Acoustics Branch, Hersonissos, Greece, 5 July 2019; Chinese Society of Acoustics Underwater Acoustics Branch: Beijing, China, 2019; pp. 292–294.
42. Jocher, G.; Qiu, J. Ultralytics YOLO11. 2024. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 1 October 2024).
43. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal visual object classes (VOC) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [[CrossRef](#)]
44. Lin, T.-Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common objects in context. *arXiv* **2015**, arXiv:1405.0312. [[CrossRef](#)]
45. Basulto-Lantsova, A.; Gutiérrez-Ruiz, J.; Martínez, L. Benchmarking deep learning models for object detection on embedded GPU platforms: A case study on Jetson Nano and Xavier NX. *J. Imaging* **2020**, *6*, 16–29.
46. Adámek, M.; Armour, W.; Mort, B. AstroAccelerate: A many-core accelerated time-domain radio astronomy signal processing library. *Astron. Comput.* **2017**, *20*, 16.
47. Dimoudi, A.; Nikolic, B.; Hessels, J.W.T. Efficient GPU-accelerated dedispersion using sub-band partitioning. *Astron. Comput.* **2018**, *23*, 136.

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.