Journal of *Marine Science and Engineering*

MDPI

*Article*

# Robust Forward-Looking Sonar-Image Mosaicking Without External Sensors for Autonomous Deep-Sea Mining

**Xinran Liu [1,2,3], Jianmin Yang [1,2,3,*], Changyu Lu [1,2,3], Enhua Zhang [1,3] and Wenhao Xu [1,2,3]**

[1] State Key Laboratory of Ocean Engineering (SKLOE), Shanghai Jiao Tong University (SJTU), Shanghai 200240, China; lxr0719@sjtu.edu.cn (X.L.); luchangyu@sjtu.edu.cn (C.L.); zhangenhuagzy@gmail.com (E.Z.); xu_wenhao@sjtu.edu.cn (W.X.)
[2] Yazhou Bay Institute of Deepsea Technology, Shanghai Jiao Tong University, Sanya 572000, China
[3] School of Ocean and Civil Engineering, Shanghai Jiao Tong University (SJTU), Shanghai 200240, China
* Correspondence: jmyang@sjtu.edu.cn

**Abstract**

With the increasing significance of deep-sea resource development, Forward-Looking Sonar (FLS) has become an essential technology for real-time environmental mapping and navigation in deep-sea mining vehicles (DSMV). However, FLS images often suffer from a limited field of view, uneven imaging, and complex noise sources, making single-frame images insufficient for providing continuous and complete environmental awareness. Existing mosaicking methods typically rely on external sensors or controlled laboratory conditions, often failing to account for the high levels of uncertainty and error inherent in real deep-sea environments. Consequently, their performance during sea trials tends to be unsatisfactory. To address these challenges, this study introduces a robust FLS image mosaicking framework that functions without additional sensor input. The framework explicitly models the noise characteristics of sonar images captured in deep-sea environments and integrates bidirectional cyclic consistency filtering with a soft-weighted feature refinement strategy during the feature-matching stage. For image fusion, a radial adaptive fusion algorithm with a protective frame is proposed to improve edge transitions and preserve structural consistency in the resulting panoramic image. The experimental results demonstrate that the proposed framework achieves high robustness and accuracy under real deep-sea conditions, effectively supporting DSMV tasks such as path planning, obstacle avoidance, and simultaneous localization and mapping (SLAM), thus enabling reliable perceptual capabilities for intelligent underwater operations.

**Keywords:** Forward-Looking Sonar; image mosaicking; deep-sea mining

check for updates

Academic Editor: Dong-Sheng Jeng

Received: 1 June 2025
Revised: 22 June 2025
Accepted: 23 June 2025
Published: 30 June 2025

**Citation:** Liu, X.; Yang, J.; Lu, C.; Zhang, E.; Xu, W. Robust Forward-Looking Sonar-Image Mosaicking Without External Sensors for Autonomous Deep-Sea Mining. *J. Mar. Sci. Eng.* **2025**, *13*, 1291. https://doi.org/10.3390/jmse13071291

**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 depletion of terrestrial mineral resources and the growing strategic importance of oceanic assets, deep-sea mineral extraction has emerged as a critical priority for many nations [1]. To achieve efficient and safe deep-sea mining, the development of intelligent underwater equipment capable of environmental perception has become a key focus of current international research. The deep-sea environment is characterized by extreme darkness, high turbidity, and a complex terrain, which significantly limits the effectiveness of optical sensing methods. Forward-Looking Sonar (FLS), as an active acoustic sensing device, enables efficient real-time perception through turbid plumes generated by the movement and operation of deep-sea mining vehicles (DSMV) [2]. Compared to optical sensors and other acoustic systems such as mechanical scanning sonar and side-scan sonar,

FLS is unaffected by ambient lighting conditions and provides higher-resolution imagery. As a result, it plays a crucial role in tasks such as obstacle avoidance, path planning, and simultaneous localization and mapping (SLAM) for DSMVs, as illustrated in Figure 1a,b.

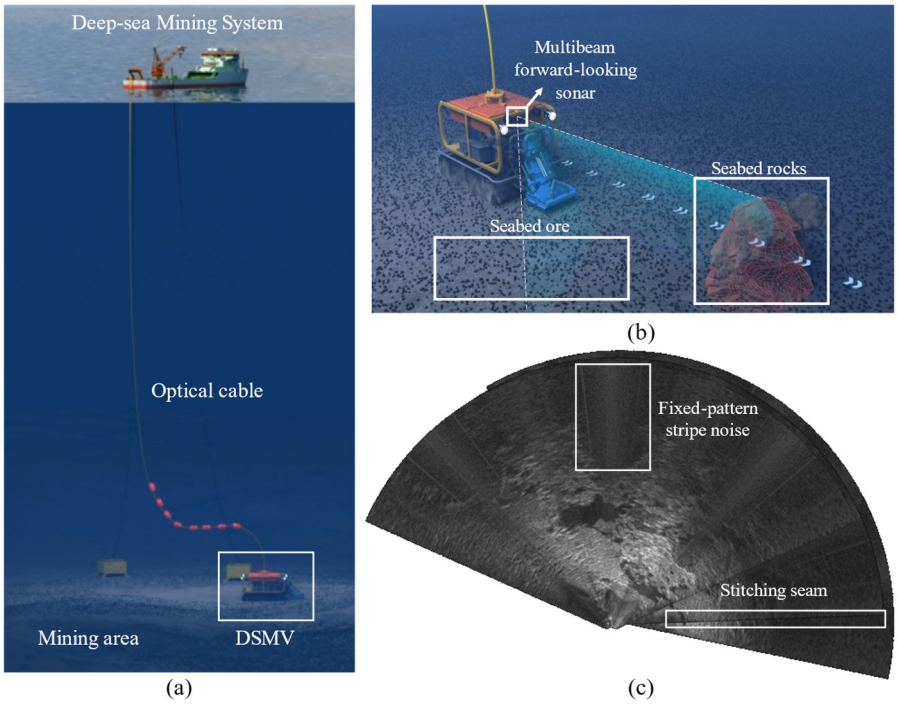


**Figure 1.** Deep-sea mining system and sonar-mosaic artefacts: (**a**) system overview; (**b**) DSMV with Forward-Looking Multibeam Sonar; (**c**) mosaic showing fixed-pattern stripe noise and stitching seam.

However, due to its characteristic fan-shaped imaging geometry, FLS suffers from uneven resolution across near and far fields, a limited field of view, and frequent issues such as echo noise and structural artifacts. These factors make single-frame images inadequate for capturing intuitive, high-quality, wide-area environmental information [3]. The field of view is inherently constrained by sonar-hardware design, making short-term improvement difficult. A common approach is to mosaic multiple overlapping FLS images captured from different viewpoints of the same scene, creating a seamless panoramic image with richer environmental information [4], as shown in Figure 1c. This technique has been successfully applied to engineering tasks such as the underwater inspection of marine structures and seafloor mapping.

In real-world deep-sea operations, however, the continuous motion of the FLS-equipped platform introduces significant changes in image brightness and contrast. Different regions of the fan-shaped image are affected to varying degrees by movement, making traditional alignment methods based on image intensity or handcrafted features unreliable. These methods often struggle to maintain alignment accuracy under such varying imaging conditions [4,5]. Moreover, mosaicked images frequently suffer from fixed-pattern noise and visible seams, which degrade overall image quality.

Many existing studies rely on auxiliary pose data provided by external sensors such as Inertial Measurement Units (IMU) and Doppler Velocity Logs (DVL) to improve mosaicking accuracy. The resulting outputs are often treated as approximate ground truth. However, in real deep-sea environments, measurements from such sensors are typically subject to high levels of noise and uncertainty, compromising the stability and reliability of sensor-assisted mosaicking methods in practical deployments. Other approaches utilize sonar data collected in simulators or controlled water-tank experiments. Although these datasets offer favorable imaging conditions suitable for algorithm validation, they

fail to reflect the degraded perception quality and robustness challenges encountered in real underwater scenarios. Additionally, some sensor-independent methods lack structural designs adapted to the non-ideal conditions of deep-sea environments. In particular, their components for feature extraction, mismatch suppression, and image fusion are not robustly designed, which limits mosaicking accuracy in practice.

To address these challenges, this paper proposes a high-precision mosaicking framework for wide-area FLS images that does not rely on external-sensor input. The main objective of this study is to design a fully image-driven and sensor-independent mosaicking framework that enhances the robustness of sonar-image registration and the quality of large-scale acoustic panorama construction under real deep-sea conditions. The framework explicitly models common imaging defects such as fixed-pattern structural noise and regional brightness inconsistency. During the feature-matching stage, a bidirectional cyclic consistency filtering mechanism and an expectation-guided feature refinement strategy are introduced. In the panoramic mosaicking stage, a radial adaptive fusion algorithm with protective frames is designed to smooth stitching seams and improve both global consistency and local detail fidelity, enabling the construction of high-quality, large-scale panoramic acoustic images. The proposed framework is readily integrable into DSMV perception systems. The resulting wide-area seafloor panoramas can provide reliable support for subsequent DSMV tasks such as path planning [6], obstacle avoidance control, and acoustic simultaneous localization and mapping (SLAM) [7], thereby significantly enhancing the safety and operational efficiency of deep-sea mineral extraction [8].

The remainder of this paper is organized as follows. Section 2 reviews related work and key technologies. Section 3 presents the FLS imaging model and details the proposed mosaicking framework and algorithms. Section 4 presents the experimental results, followed by a comprehensive analysis and evaluation. Section 5 concludes the paper.

## 2. Related Work

### 2.1. Review of Sonar-Image Denoising Methods

FLS images acquired during deep-sea operations are often affected by strong speckle noise, fixed-pattern stripe artifacts, and global brightness non-uniformity, all of which significantly impair the accuracy of subsequent feature extraction and matching. Existing research on noise suppression and image-quality enhancement for FLS images primarily falls into two categories: traditional image-filtering techniques and self-supervised denoising methods based on deep learning [9].

Traditional filters such as mean filtering, median filtering, and Gaussian filtering [10] have shown good performance on natural images, particularly in suppressing Gaussian and salt-and-pepper noise. However, these methods exhibit two major limitations when applied to FLS imagery under deep-sea conditions. First, they tend to blur image edges and suppress weak target signals during the filtering process. Second, they are generally ineffective at removing structured noise such as fixed-pattern stripes [11], making it difficult to balance noise suppression with the preservation of key echo features.

To address the lack of clean ground truth in real FLS datasets, we constructed a static-view dataset using adjacent frame pairs, enabling the application of self-supervised denoising methods like Noise2Noise (N2N) [12]. However, the experimental results revealed that these methods exhibit limited effectiveness when applied to deep-sea FLS images. This is mainly due to the complex noise characteristics in such data, which violate the basic assumptions of zero-mean and pixel-wise independent noise required by these approaches.

In recent years, self-supervised learning methods such as N2N and Noise2Self [13] have made significant progress in domains such as natural and medical imaging. These

approaches learn denoising mappings by leveraging paired noisy images or internal image redundancy, without requiring ground-truth labels. In theory, they are suitable for label-scarce scenarios. However, these methods are typically built upon assumptions such as zero-mean noise, pixel-wise independence, or sufficient internal redundancy within the image. In FLS imagery, noise tends to exhibit strong spatial correlation and non-independent, non-identically distributed (non-IID) characteristics. Additionally, it is influenced by physical mechanisms such as multipath echoes and gain drift, making model convergence difficult and limiting both generalization and stability during training [14].

In summary, most existing denoising methods are designed for natural images or synthetic underwater datasets, and are not well suited to the extreme low signal-to-noise ratios and systematic noise interference commonly encountered in deep-sea mining scenarios. To address these limitations, this study proposes a two-stage preprocessing and denoising framework that integrates sonar-imaging physics with multi-frame image enhancement, aiming to improve the overall image quality and feature fidelity. The proposed approach is detailed in Section 3.2.

### 2.2. Review of Sonar-Image Matching and Mosaicking Algorithms

Research on sonar-image registration and mosaicking can be broadly categorized into two main approaches: region-based frequency-domain methods and feature-based spatial-domain methods.

Frequency-domain approaches typically utilize Fourier transforms to extract spectral features and perform image alignment based on frequency correlations. For instance, Hurtós et al. [15] proposed a phase-correlation-based Fourier registration method capable of robust mosaicking in low-visibility underwater environments, demonstrating the adaptability of frequency-domain strategies under weak texture conditions. Hansen et al. [16] introduced the FS2D method, which maps the Fourier spectrum onto a spherical surface and estimates rotations using SO(3) transformations. This method achieves robust registration under large viewpoint changes and high noise levels, highlighting the potential of frequency-domain techniques for highly disturbed sonar data. These methods are computationally efficient and less sensitive to translational shifts, making them suitable for low-texture or noisy conditions. However, they are generally less effective at handling rotation and spatial structural variations, which limits their accuracy under dynamic viewpoints and complex terrain.

To overcome the limitations of frequency-domain methods in handling complex geometric transformations, recent studies have increasingly focused on feature-based spatial-domain mosaicking techniques [17], which have become the dominant approach in sonar-image processing. Inspired by frameworks in optical-image mosaicking, these methods extract key-points and geometric descriptors, match features based on descriptor similarity, and estimate inter-image transformations using algorithms such as RANSAC. This category of methods demonstrates good robustness against scale changes, rotations, and local occlusions, particularly in well-structured, texture-rich imagery. However, extreme environments such as deep-sea mining still pose significant challenges. Repetitive textures and blurred edges in FLS images reduce the distinctiveness and stability of key-points. Furthermore, non-uniform sound propagation caused by complex seafloor topography, combined with low signal-to-noise ratios and occlusions, further undermines the reliability of feature extraction and matching, limiting the effectiveness of these methods in real-world conditions.

To enhance the performance of traditional feature-based registration methods for FLS mosaicking, various improvements have been proposed. These include advanced matching strategies, multi-stage registration pipelines, and fusion-aware mosaicking tech-

niques aimed at improving geometric consistency and accuracy. For example, Li et al. [18] proposed a hybrid approach combining feature-based region selection with region-level registration. Their method achieved favorable results on both commercial ship datasets and ROV-based pool experiments, validating the effectiveness of multi-stage strategies. Su et al. [19] addressed the issue of registration errors by introducing a local statistics-based weighting mechanism during the fusion stage to mitigate error propagation. Shang et al. [20] proposed a matching optimization framework based on density clustering and convolutional consistency analysis, designed to reduce the dependency on predefined geometric models. While their work primarily targets multi-source sonar-image matching, its motion modeling and outlier suppression concepts offer useful insights for improving geometric consistency. In another approach, Wei et al. [21] moved beyond traditional image-domain mosaicking and developed a beam-domain representation model based on FLS imaging principles, enabling improved alignment accuracy and efficiency through deformation modeling. Despite these advancements, many of the aforementioned methods still rely on external pose data from navigation sensors or assume well-structured image content. As a result, they struggle to remain deployable under real-world conditions such as image degradation, weak features, or unstable positioning.

Recently, deep-learning-based methods such as SuperPoint [22] and LoFTR [23] have achieved remarkable performance in feature extraction and matching for natural and remote-sensing imagery. However, these methods require large-scale, high-quality labeled datasets for training and are mainly optimized for texture-rich, edge-defined scenes. Due to the inherent low texture, high noise, and distortion in FLS images, the direct transferability of these models is severely restricted, leading to poor generalization and compromised robustness. SONIC [24], a more recent approach tailored to sonar images, employs a pose-aware feature-learning framework aided by IMU data and synchronized pose labels under a weakly supervised setting. Nevertheless, SONIC relies on data collected from simulators and external-sensor inputs. In actual deep-sea operations, accurate pose estimation is difficult, and high communication latency often prevents real-time deployment, limiting the applicability of such systems in real-world platforms.

In summary, current studies primarily focus on feature-matching pipelines that depend on high-quality training data and external-sensor assistance. There remains a lack of fully image-driven, sensor-independent mosaicking frameworks that can robustly operate under the extreme conditions of deep-sea environments. To address this gap, this study introduces a practical system based on uncertainty-aware feature matching and multi-scale incremental mosaicking. The proposed approach is entirely based on FLS imagery and achieves high robustness in both feature registration and image mosaicking. The detailed algorithms are presented in Sections 3.3 and 3.4.

## 3. Methodology

This study proposes a panoramic mosaicking framework based on sequences of FLS images, as illustrated in Figure 2. It is important to note that the focus of this work is on the engineering feasibility of applications in deep-sea mining operations. At depths ranging from 4000 to 6000 m, external positioning sensors such as IMU and DVL often fail to provide stable and reliable localization due to factors such as cumulative attitude drift, signal attenuation, and reflections from complex seafloor topography. To ensure the generality and practical applicability of the proposed method, the image-matching and mosaicking processes were performed entirely based on the information contained within FLS images, without reliance on any external-sensor data. Moreover, external measurements were not utilized for supervision or geometric alignment at any stage of the process.

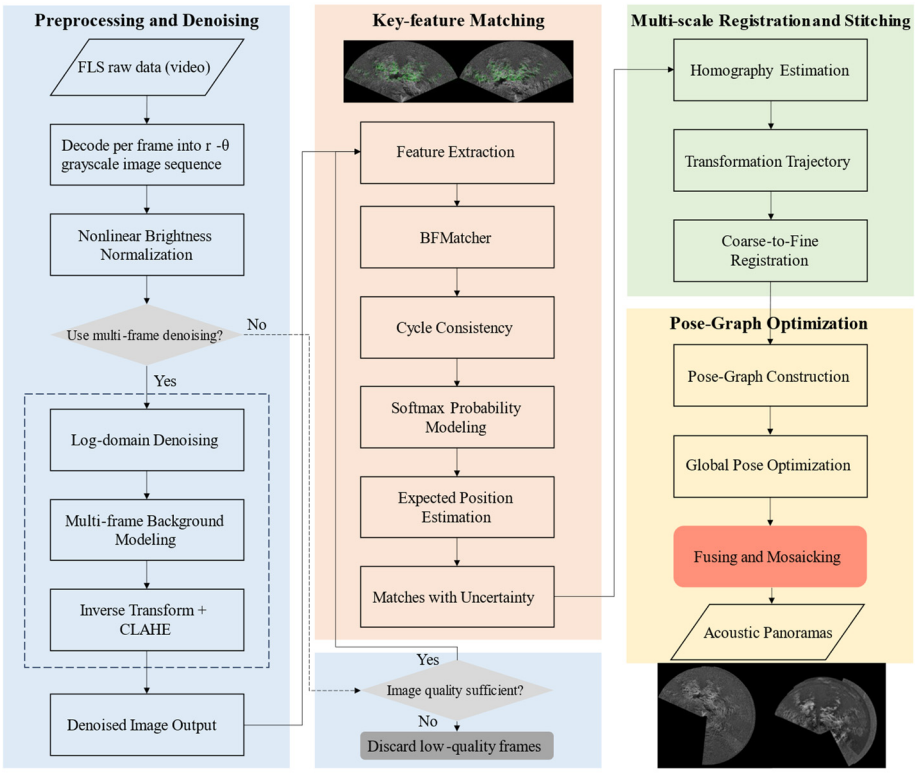


**Figure 2.** Flowchart of the proposed FLS image stitching and mosaicking framework.

### *3.1. Imaging Model and Platform Perturbation*

To achieve high-precision image matching, it is necessary to establish the FLS imaging model and its projection variations under motion perturbations, as illustrated in Figure 3. FLS captures the range $r$ and azimuth angle $\theta$ of a target point through multibeam scanning. However, due to the wide vertical-beam width and low resolution, the elevation angle $\phi$ of the target is lost during imaging, and the projection is approximated onto a plane with zero elevation angle. Let the 3D target point be defined as $\boldsymbol{P}_w = [X, Y, Z]^T$; its 2D projection onto the sonar-image coordinate system can be expressed as

$$\hat{\boldsymbol{P}}_s = \begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix}, \; r = \sqrt{X^2 + Y^2 + Z^2}, \; \theta = arctan2(Y, X) \tag{1}$$

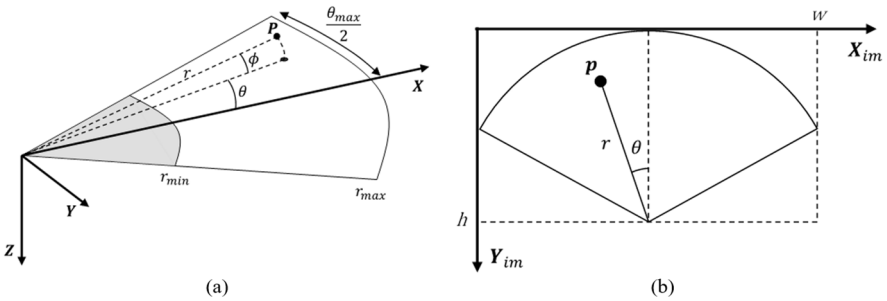


**Figure 3.** FLS imaging geometry. (**a**) 3D fan-shaped scanning model. (**b**) 2D image projection.

In practical applications, this study focuses on the movement of DSMVs equipped with FLS operating over complex seafloor terrains, where slight pose disturbances inevitably occur during task execution. Suppose the DSMV undergoes small translations $\Delta \boldsymbol{t}_{xy} = [\delta x, \delta y]^T$

along the x and y axes and a minor yaw rotation $\delta\varphi$ about the z-axis. The position of the target point in the DSMV coordinate system then transforms as

$$\boldsymbol{P}'_w = \boldsymbol{R}(\delta\varphi)^T \left(\boldsymbol{P}_w - \Delta\boldsymbol{t}_{xy}\right) \tag{2}$$

where the 2D rotation matrix is given by

$$\boldsymbol{R}(\delta\varphi) = \begin{bmatrix} cos\delta\varphi & -sin\delta\varphi \\ sin\delta\varphi & cos\delta\varphi \end{bmatrix} \tag{3}$$

According to the existing literature [25], FLS imaging is typically modeled as a 2D polar projection, considering mainly in-plane translations and yaw rotations of the sensor platform. Under this model, the influence of such perturbations on the geometric structure of targets can be neglected, and the variations in target positions can be represented by a 2D rigid-body transformation:

$$\hat{\boldsymbol{P}}'_s = \boldsymbol{R}(\delta\varphi)\cdot\hat{\boldsymbol{P}}_s + \Delta\boldsymbol{t}_{xy} \tag{4}$$

It is worth noting that in deep-sea mining scenarios of interest in this study, DSMVs may frequently experience minor pitch perturbations during movement and operation on sloped terrains. Although pitch disturbances do not affect the spatial relative positions or beam indices of the targets and thus keep the observed $r$ and $\theta$ stable, they alter the incidence angles of the sonar beams. This affects local illumination angles and occlusion patterns, leading to variations in brightness and shadowing in the images, which can degrade the stability of inter-frame image matching.

Based on this analysis, the subsequent registration modeling continues to employ a 2D rigid-body transformation to describe geometric variations in the images, under the assumption that pitch disturbances do not affect the geometric contour positions of targets but do induce changes in image-brightness features. This assumption reveals the underlying conditions for geometric invariance relative to platform attitude stability while also exposing the potential interference risks posed by brightness variations during the registration process. Accordingly, to address brightness variations caused by pitch disturbances, this study further designs an image preprocessing and denoising strategy that integrates dynamic normalization and physical modeling to enhance the image quality and improve the robustness of feature matching.

### *3.2. Brightness Normalization and Noise Suppression*

During the process of panoramic mosaicking based on FLS images, non-uniform brightness variations and noise interference often become critical factors limiting registration accuracy. Under specific FLS systems and the extreme operational conditions of deep-sea mining, images are frequently affected by fixed-pattern stripe noise and non-uniform brightness, which can introduce abrupt boundaries and local artifacts at stitching seams, thereby degrading the overall reconstruction quality. To improve grayscale consistency and local contrast in sonar images, this study proposes a two-stage image preprocessing and noise suppression approach that combines physical modeling and image enhancement techniques. The two stages, respectively, perform nonlinear brightness normalization and logarithmic-domain multi-frame noise suppression.

First, to better understand the noise components, the formation process of sonar images was modeled based on a multiplicative-additive noise model. The observed image can be expressed as

$$I_{observed}(x,y) = I_{true}(x,y) \times N_m(x,y) + N_a(x,y) \tag{5}$$

where $I_{true}(x, y)$ represents the ideal echo image, $N_m(x, y)$ denotes the multiplicative noise components such as systematic gain drift and stripe noise, and $N_a(x, y)$ represents additive noise components such as electronic and thermal noise. In deep-sea applications, the echo signals are extremely weak, making the effects of multiplicative noise particularly significant.

To validate the reasonableness of this noise modeling assumption, multiple FLS images collected under various deep-sea conditions were subjected to frequency-domain analysis, as shown in Figures 4 and 5. Figure 4 reveals prominent stripe-like interference patterns, corresponding to high-intensity frequency components along specific directions in the Fourier spectrum, forming typical directional spectral peaks. Furthermore, Figure 5 presents the normalized noise characteristic curves extracted from the analyzed images. Despite differences in acquisition environments and scenes, the images exhibit similar distribution trends in metrics such as average spectral energy, peak spectral energy, and local contrast. This observation further demonstrates the common structural properties of sonar-image noise, particularly the dominance of multiplicative structures and directional stripe interference. These findings provide strong theoretical and practical support for the adoption of a denoising strategy that integrates frequency-domain suppression with structural modeling in the proposed algorithm.

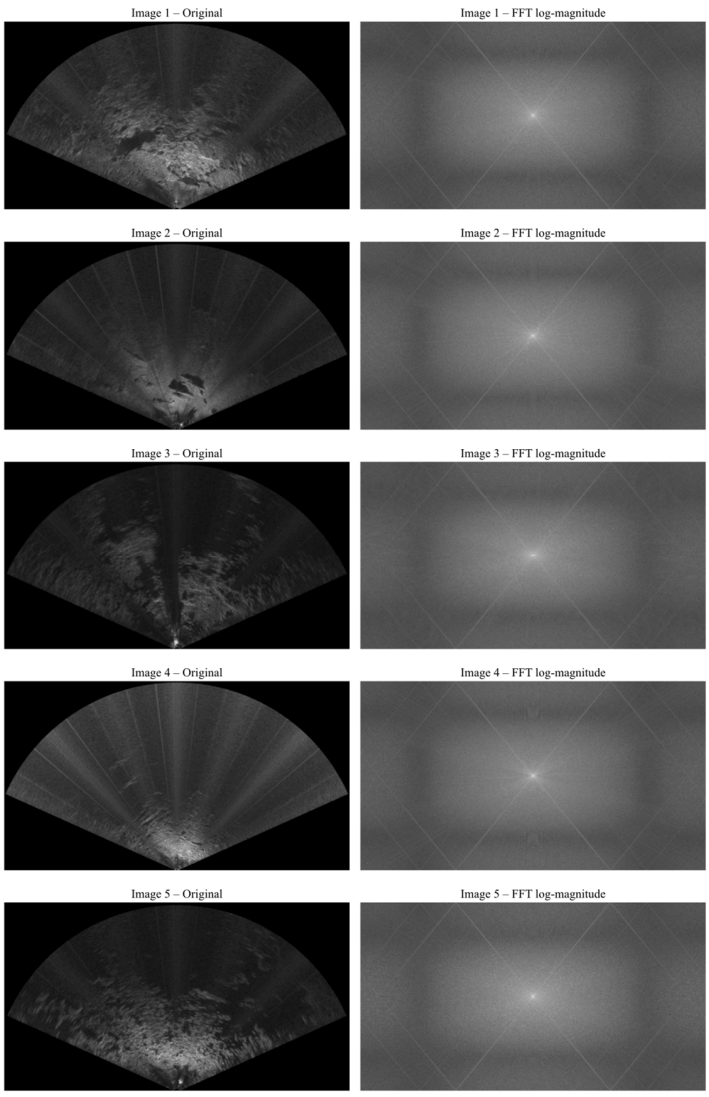


**Figure 4.** FLS images and their spectra from various deep-sea scenes. Directional peaks indicate striping-like multiplicative noise.

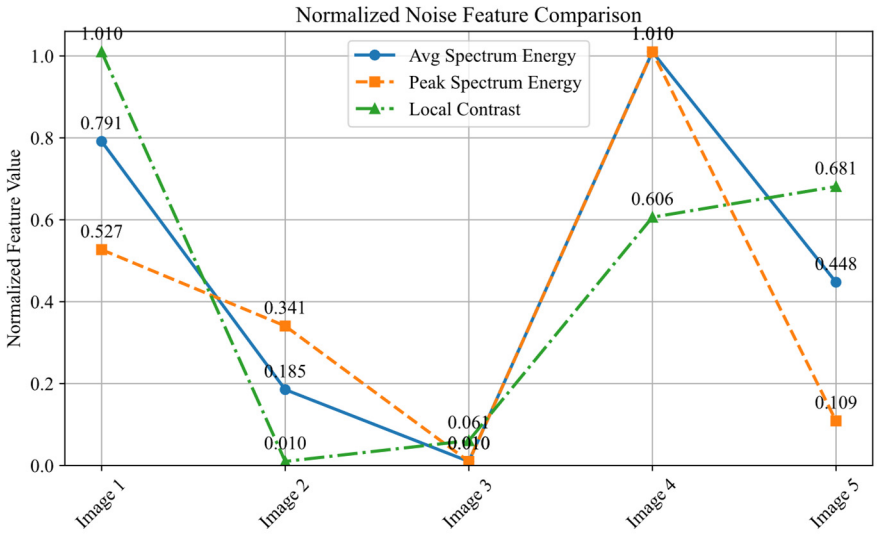


**Figure 5.** Normalized feature comparison across images. Consistent trends support the proposed noise modeling.

The proposed method consists of a two-stage image processing pipeline:

(1) Raw Data Preprocessing and Brightness Normalization

First, raw sonar data are decoded frame-by-frame to generate range–azimuth $r$-$\theta$ grayscale images, preserving spatial geometric information. Due to pitch disturbances, the dynamic range of image brightness may become unstable. To address this, a nonlinear normalization strategy is employed to adjust brightness dynamically. Specifically, a combination of min–max normalization and exponential enhancement is used to stretch low-intensity regions and suppress high-intensity saturation, thereby improving the overall brightness balance. The transformation model is expressed as

$$I_{adjusted}(x,y) = \left(\frac{I(x,y) - I_{min}}{I_{max} - I_{min}}\right)^{p} \times (v_{max} - v_{min}) + v_{min} \tag{6}$$

where $p$ is a parameter controlling brightness enhancement. This stage aims to generate standardized images with a balanced dynamic range and preserved details, providing a stable input for subsequent noise suppression;

(2) Log-Domain Modeling and Multi-Frame Fixed-Pattern Noise Suppression

To further mitigate fixed-pattern stripe noise, this stage adopts a log-domain background modeling and multi-frame fusion strategy. A logarithmic transformation is first applied to the images, converting multiplicative noise into additive interference:

$$\log(1 + I_{observed}(x,y)) = \log(1 + I_{true}(x,y)) + \log(1 + N_m(x,y)) \tag{7}$$

Subsequently, for image sequences of the same dimensions, a truncated mean is computed at each pixel by retaining only the central 90% of pixel values, thereby enhancing robustness against outliers and modeling the background distribution. Each frame is then background-subtracted in the log domain and restored to the linear domain using the inverse exponential transformation, achieving the simultaneous suppression of fixed-pattern noise and global brightness non-uniformity. Finally, contrast-limited adaptive histogram equalization (CLAHE) is applied to enhance local details, further improving texture clarity and overall visual quality.

Algorithm 1 summarizes the proposed preprocessing and denoising procedure. By integrating the physical modeling of sonar imaging, image enhancement, and multi-frame noise suppression, the method achieves brightness consistency and detail preservation

under extremely low signal-to-noise ratio conditions, providing a high-quality image foundation for subsequent image registration and mosaicking tasks.

It is worth noting that the design of the subsequent registration and mosaicking strategy also takes into account the imaging characteristics of the sonar used in this study. The high angular resolution (0.18° beam spacing) provides denser and more consistent spatial sampling across adjacent frames, which facilitates stable feature extraction and matching. Meanwhile, the narrow horizontal beamwidth (1°) results in sharper but more limited fan-shaped views per frame, making multi-scale registration and geometric consistency filtering essential for constructing seamless mosaics over extended seafloor areas.

**Algorithm 1**: Two-stage preprocessing and denoising framework for FLS images.

**Input :** Raw sonar frames $\mathcal{F}$, brightness adjustment exponent $p$, trimming fraction $\alpha$
**Output:** Denoised sonar images $\mathcal{S}$

**for** *each frame* $f_i$ *in* $\mathcal{F}$ **do**
  Generate polar-coordinate grayscale image $I_{r\theta}(x, y)$;
  Normalize $I_{r\theta}(x, y)$ by min-max scaling;
  Apply brightness adjustment with exponent $p$:
  $I_{adj}(x, y) = \left(\frac{I_{r\theta}(x,y) - I_{\min}}{I_{\max} - I_{\min}}\right)^p \times (v_{\max} - v_{\min}) + v_{\min}$;
  Save the preprocessed image $I_{adj}(x, y)$;
**end**
**for** *each group* $\mathcal{G}$ **do**
  **for** *each image* $g \in \mathcal{G}$ **do**
    Compute log-transformed image: $I_{\log}(x, y) = \log(1 + g(x, y))$;
  **end**
  Stack all $I_{\log}(x, y)$ into tensor $\mathcal{T}$;
  Compute trimmed mean background $B_{\log}(x, y)$ by removing top and bottom $\alpha\%$;
  **for** *each* $I_{\log}(x, y)$ **do**
    Subtract background: $I_{corr}(x, y) = I_{\log}(x, y) - B_{\log}(x, y)$;
    Inverse log-transform: $I_{final}(x, y) = \exp(I_{corr}(x, y)) - 1$;
    Normalize $I_{final}(x, y)$ to $[0, 255]$;
    Apply CLAHE enhancement to $I_{final}(x, y)$;
    Save the denoised image;
  **end**
**end**
**return** $\mathcal{S}$

### *3.3. Key-Point Matching Optimization for FLS Images*

To achieve high-quality mosaicking of FLS images, it is essential to ensure accurate key-point matching across multiple frames. Although the A-KAZE algorithm can stably extract feature points and generate descriptors in a multi-scale nonlinear diffusion space, and the BFMatcher provides an efficient brute-force matching method, significant challenges arise when applying these techniques to FLS imagery. Specifically, due to the superposition of multi-directional echoes, FLS images of terrains such as densely rocky seabeds and soft sediment areas contain numerous structurally similar patterns. These characteristics degrade the discriminative power of the descriptors, leading to a flattened distance distribution. Consequently, the distances among multiple candidate points become very close, making it difficult to select matches decisively based on minimum distance alone. This issue results in frequent mismatches using conventional methods, adversely affecting the subsequent geometric transformation estimation.

Moreover, in regions near feature edges or with significant scale variations, pure pixel-level matching lacks precision control, making it difficult to achieve the sub-pixel accuracy required for high-quality mosaicking. To overcome these limitations, two key improvements are proposed: bidirectional cyclic consistency filtering and expectation-guided match refinement. These strategies address the problems of matching accuracy and geometric consistency, respectively. Algorithm 2 summarizes the proposed matching algorithm.

In conventional BFMatcher, for each key-point $d_i^A$ in image A, the corresponding point is found in the descriptor set $\mathcal{D}_B$ of target image B by minimizing the Hamming distance, as expressed by

$$f_{A\to B}(i) = \underset{j}{argmin} d_H\left(d_i^A, d_j^B\right) \tag{8}$$

Although this method is simple and efficient and performs well on natural images, several challenges have been observed in FLS images collected from real deep-sea environments. Due to the local repetitiveness of structures such as rocks and gullies, as well as speckle noise caused by multipath echoes, descriptors often exhibit ambiguity and non-uniqueness in spatial distribution. As a result, descriptors for multiple key-points tend to be closely clustered, causing the matching results to favor a set of similar points rather than a unique and stable physical correspondence. This ambiguity severely undermines the stability of subsequent geometric transformation estimations. To enhance geometric consistency, a reverse matching validation function is introduced, defined as

$$f_{B\to A}(j) = \underset{i}{argmin} d_H\left(d_j^B, d_i^A\right) \tag{9}$$

Only those matching pairs that satisfy the cyclic consistency condition are retained:

$$f_{B\to A}(f_{A\to B}(i)) = i \tag{10}$$

This strategy introduces a redundancy constraint into the matching process, requiring that each key-point not only projects to its most reliable match in the target image but that it must also be reciprocally matched back from the target image using the same nearest-neighbor criterion. By enforcing mutual nearest-neighbor relationships, this approach effectively suppresses mismatches caused by descriptor ambiguity or structural repetition. In practice, polar-coordinate sonar images frequently exhibit clusters of neighboring points with highly similar descriptor features; bidirectional consistency aids in selecting the most representative pair within such ambiguous matches. Additionally, this symmetric-matching logic reinforces geometric consistency between images, providing a more stable and reliable set of initial point correspondences for subsequent affine transformation estimation.

Even after applying the above filtering strategy to obtain a relatively reliable set of matches, challenges remain in weak-texture regions and near the fan-shaped edges, where minimal Hamming-distance differences among candidate points reduce descriptor discriminability. As a result, matching outcomes are still prone to fluctuation among similar candidates, preventing stable identification of optimal correspondences. Moreover, pixel-level matching lacks sub-pixel precision control, often leading to misalignment or visible seams in the mosaicking process, particularly near image edges.

To address these issues, an expectation-guided match refinement strategy is proposed to replace traditional hard-decision matching. Specifically, for each query descriptor $d_q^A$, the top $K$ nearest candidate matches are selected, denoted as $\{(x_i, d_i)\}_{i=1}^K$, and their distances

are converted into a probability distribution. A softmax-based probability distribution is constructed to model the matching confidence for each candidate point:

$$p_i = \frac{exp(-\alpha d_i)}{\sum_{j=1}^{K} exp(-\alpha d_j)} \tag{11}$$

where $d_i$ is the Hamming distance of the $i$-th candidate, and $\alpha > 0$ is a smoothing factor controlling the concentration of the weight distribution. Formally, this distribution is equivalent to a Boltzmann distribution with $d_i$ as the energy function, which is widely used in probabilistic modeling of expected cost minimization. Thus, it not only provides an interpretable confidence distribution for matching but also serves as a cost representation in descriptor space. Based on the probability distribution, the matching-position estimation problem is further formulated as a weighted least-squares minimization:

$$x = \underset{x}{argmin} \sum_{i=1}^{K} p_i \cdot \|x - x_i\|^2 \tag{12}$$

This objective seeks an optimal position that is, in a weighted sense, closest to all candidate matches. The optimization problem admits a closed-form solution obtained by differentiating the objective function with respect to $x$ and setting the derivative to zero, yielding

$$\hat{x} = \sum_{i=1}^{K} p_i \cdot x_i \tag{13}$$

The estimated position is continuous in coordinate space, achieving sub-pixel accuracy without the need for interpolation or regression models. Furthermore, the approach exhibits strong robustness by integrating the spatial-distribution information of multiple candidates, enabling stable outputs even in ambiguous regions where descriptors lack clear minima. This significantly reduces matching fluctuations. Additionally, the probability distribution allows for the construction of an uncertainty measure for the matching point, defined as the weighted variance of the candidate set:

$$\sigma^2 = \sum_{i=1}^{K} p_i \cdot \|x_i - \hat{x}\|^2 \tag{14}$$

This uncertainty measure can be employed in weighted RANSAC frameworks as a reliability indicator during subsequent geometric estimation, further improving the accuracy of affine matrix estimation.

In summary, by introducing the above strategies into the traditional A-KAZE and BFMatcher framework, the stability and accuracy of key-point matching in sonar images are significantly improved. Bidirectional cyclic consistency filtering effectively eliminates mismatches through symmetric validation, enhancing the geometric reliability of matched pairs. Expectation-guided match refinement, through optimization-based soft estimation, demonstrates notable advantages in weak-texture regions and under sub-pixel precision requirements. These two enhancements provide a more robust foundation for high-precision image registration and mosaicking in complex deep-sea sonar-imaging environments, facilitating more reliable affine estimation and image fusion in subsequent stages.

### *3.4. Multi-Scale Registration and Stitching*

After obtaining high-quality matching points, achieving accurate geometric registration across multiple FLS images is critical for high-quality mosaicking. Directly fitting a geometric model to original-resolution images often leads to unstable RANSAC-based

affine estimation due to the effects of speckle noise, viewpoint variations, and low-texture regions. To address this, a multi-scale registration framework was adopted in this study, incorporating the uncertainty-aware matching results proposed in Section 3.3 to enhance the adaptability and robustness of the strategy for sonar-image mosaicking.

Specifically, a Gaussian pyramid $\mathcal{I}^0, \mathcal{I}^1 \ldots \mathcal{I}^n$, In was constructed for each input image pair through Gaussian filtering and downsampling. Coarse registration was performed at the lowest resolution layer $\mathcal{I}^n$, where key-point matching followed the bidirectional consistency filtering and expectation-guided refinement algorithm described in Section 3.3. Unlike traditional methods, the proposed approach introduces uncertainty weights into the affine estimation model during coarse registration. The coarse affine matrix $\boldsymbol{A}_L$ is obtained through a weighted least-squares model:

$$\boldsymbol{A}_L = \underset{A}{argmin} \sum_{i=1}^{N} w_i \cdot \left\| x_i' - A \cdot \widetilde{x_i} \right\|_2^2 \tag{15}$$

where $\widetilde{x_i} = [x_i, y_i, 1]^T$ represents the homogeneous coordinates, and $w_i$ is the reciprocal of the matching uncertainty derived from the softmax-based confidence scores introduced in Section 3.3. This design effectively mitigates the influence of low-texture regions or outlier matches on the geometric estimation, enhancing the robustness of the fitting process.

The coarse affine result $\boldsymbol{A}_L$ is then used to initialize a fine registration process on the original-resolution images $\mathcal{I}^0$. For each predicted matching point $x^0$, refined feature matching is conducted within a local neighborhood window to achieve fine-grained matching. The final affine matrix $\boldsymbol{A}_0$ is used to align the source image $\mathcal{I}_{src}$ with the reference image $\mathcal{I}_{ref}$, following the transformation:

$$x_{ref} = \boldsymbol{A}_0 \cdot \widetilde{x_{src}} \tag{16}$$

After geometric registration, the newly transformed frame is mapped onto a global canvas. The newly added regions relative to the current mosaic are cropped and merged, implementing an incremental image fusion based on geometric relationships. The proposed approach integrates the uncertainty-weighted feature matching strategy from Section 3.3 and further explores its practical effectiveness within a multi-scale registration system. Considering the high sampling rate of sonar data, frame-by-frame mosaicking can lead to image blurring and redundancy. To address this, keyframes are selected at intervals of 5–15 frames and are used as nodes to construct a pose-graph, incorporating distance constraints. A graph optimization method is applied to jointly refine the pose relationships among all nodes. This strategy significantly reduces computational load while effectively suppressing cumulative registration drift caused by dramatic inter-frame viewpoint changes, sparse textures, and noise interference. It enhances robustness against outlier matches and achieves superior registration stability and mosaicking accuracy, particularly in the low-texture and high-noise conditions typical of FLS imagery. The key steps of the proposed process are summarized in Algorithm 2.

To enhance the quality of image fusion, a radial adaptive fusion algorithm with protected frames is designed. Specifically, several early keyframes in the mosaicking sequence are designated as global reference frames. The pixels from these frames are given priority in writing to the panoramic canvas and are subsequently protected from being overwritten by later frames. For overlapping regions of subsequent frames, a set of piecewise-smooth functions based on the radial distance from the image origin is introduced as fusion weights. Lower weights are assigned to new frames near the center to strengthen global consistency, while the weights gradually increase toward the periphery, thereby preserving the edge texture details carried by the later frames.

**Algorithm 2**: Uncertainty-aware multi-scale registration for FLS images.

**Input :** Sequential FLS images $\{I_1, I_2, \ldots, I_T\}$, sampling interval $k$, softmax factor $\alpha$, pyramid levels $n$
**Output:** Stitched panorama canvas $C$

Initialize node set $\mathcal{N} \leftarrow \{I_i \mid i \bmod k = 0\}$;
Initialize pose-graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$;
**for** *each consecutive pair* $(I_A, I_B) \in \mathcal{N}$ **do**
  $F_A, D_A \leftarrow \mathrm{AKAZE}(I_A)$;
  $F_B, D_B \leftarrow \mathrm{AKAZE}(I_B)$;
  **for** *each descriptor* $d_i^A \in D_A$ **do**
    $f_{A\to B}(i) \leftarrow \arg\min_j d_H(d_i^A, d_j^B)$;
  **end**
  **for** *each descriptor* $d_j^B \in D_B$ **do**
    $f_{B\to A}(j) \leftarrow \arg\min_i d_H(d_i^A, d_j^B)$;
  **end**
  $\mathcal{M} \leftarrow \{(i,j) \mid f_{B\to A}(f_{A\to B}(i)) = i\}$;
  ; `// Cycle-consistency filter`
  **for** *each* $d_q^A$ *in* $\mathcal{M}$ **do**
    Select top-$K$ candidates $\{(x_i, d_i)\}_{i=1}^K$;
    $p_i = \frac{\exp(-\alpha d_i)}{\sum_{j=1}^K \exp(-\alpha d_j)}$;
    $\hat{x} = \sum_{i=1}^K p_i x_i$;
  **end**
  Build Gaussian pyramid $\{I^0, I^1, \ldots, I^n\}$;
  **for** $\ell = n, n-1, \ldots, 0$ **do**
    **if** $\ell = n$ **then**
      Estimate coarse affine $A_n$ by weighted least squares:
      $A_n \leftarrow \arg\min_A \sum_i w_i \|x_i' - A\tilde{x}_i\|_2^2$;
    **end**
    **else**
      Refine affine $A_\ell$ using matches at level $\ell$;
      $A_\ell \leftarrow \arg\min_A \sum_i w_i \|x_i' - A\tilde{x}_i\|_2^2$;
    **end**
  **end**
  Add edge $e_{AB}$ with constraint $A_0$ to pose-graph $\mathcal{G}$;
**end**
Optimize pose-graph $\mathcal{G}$ to obtain globally consistent transforms $\{\hat{A}_i\}$;
**for** *each image* $I_i \in \mathcal{N}$ **do**
  Warp $I_i$ into canvas $C$ using $\hat{A}_i$;
**end**
**return** $C$

This strategy balances the stability of early frames in maintaining the overall panoramic structure with the richness of texture details provided by later frames. It ensures that the central region retains clear structural information while significantly improving texture quality in the outer areas. The fusion approach effectively enhances boundary smoothness and detail preservation in the mosaicking results. The key steps of the proposed fusion strategy are summarized in Algorithm 3.

**Algorithm 3**: Protected-frame radial-adaptive blending algorithm.

**Input :** Canvas $C$, warped image $W$, canvasMask $M$, frameIndex $i$, seamWidth $W_s$, protectedFrames $N_0$, origin $(c_x, c_y)$, innerRadius $R_{in}$, outerRadius $R_{out}$
**Output:** Updated canvas $C$ and mask $M$
Init $maskI \leftarrow (W > 0)$;
Init $overlap \leftarrow maskI \wedge (M > 0)$;
Init $newArea \leftarrow maskI \wedge (M = 0)$;
$C[newArea] \leftarrow W[newArea]$;
**if** $i \leq N_0$ **then**
  // Protect first $N_0$ frames: skip blending
**else**
  // Compute overlap blending weights $invMask \leftarrow \neg newArea$;
  $distMap \leftarrow \text{DistanceTransform}(invMask)$;
  $wNew \leftarrow \text{clip}(distMap/W_s, 0, 1)$;
  // Old weight is zero in overlap $wOld \leftarrow 0$ for all $(x, y)$ in overlap;
  // Radial factor $r(x, y) \leftarrow \sqrt{(x - c_x)^2 + (y - c_y)^2}$;
  $\alpha_r(x, y) \leftarrow \text{clip}\big((r(x, y) - R_{in})/(R_{out} - R_{in}), 0, 1\big)$;
  // Final blending weight
  $wFinal(x, y) \leftarrow (1 - \alpha_r(x, y))\, wOld + \alpha_r(x, y)\, wNew(x, y)$;
  // Blend overlapping pixels **foreach** *pixel* $(x, y)$ *in overlap* **do**
    $C(x, y) \leftarrow (1 - wFinal(x, y))\, C(x, y)$
      $+wFinal(x, y)\, W(x, y)$
// Update mask $M \leftarrow M \vee maskI$;
**return** $C, M$

## 4. Experiments and Discussion

### *4.1. Experimental Platform and Dataset*

The experimental data used in this study were collected during a sea trial conducted in the western Pacific Ocean using the DSMV *Pioneer II*, developed by Shanghai Jiao Tong University. The platform is equipped with an inertial navigation system, electronic compass, an ultra-short baseline system, a Forward-Looking Sonar, an underwater camera, and lighting systems, enabling it to perform deep-sea mineral extraction tasks under complex ocean conditions. To obtain high-resolution images of the seafloor mining area, a BlueView M900 D6-Mk2 FLS (Teledyne BlueView, Bellevue, WA, USA) was mounted on the front of the DSMV. The performance specifications of the FLS are summarized in Table 1. This imaging system is particularly suited for low-visibility underwater environments and can effectively detect seafloor contours and obstacles ahead. The appearance of the DSMV *Pioneer II* and sonar system, along with the imaging principles and example images, are shown in Figure 6.

During the data-preparation phase, several representative FLS image sequences were selected from the raw sea trial data for algorithm validation. The dataset covers various typical seafloor terrains, including rocky areas, gullies, and fine sediment regions, exhibiting sonar-imaging characteristics such as sparse textures, repetitive structures, and strong noise. These features provide a challenging testbed for algorithm evaluation. Unlike commonly used public synthetic datasets, the data used in this study are entirely derived from real sea trial measurements, offering a more realistic reflection of practical performance. It is noteworthy that although the *Pioneer II* is equipped with multiple onboard sensors, factors such as water turbulence, unstable environmental conditions, and the extreme pressures and low temperatures at depths of several thousand meters introduce significant drift and cumulative errors in sensor measurements. Consequently, some sensor data cannot serve as reliable ground truth during practical operations. To enhance the robustness and gener-

alization capability of the proposed registration and mosaicking methods for engineering applications, all processing is conducted purely based on image content without relying on external pose-sensor information.

**Table 1.** Key performance parameters of the BlueView M900 D6-Mk2 sonar.

| Parameters | Value |
|---|---|
| Operating Frequency | 900 kHz |
| Field of View | 130° |
| Maximum Detection Range | 100 m |
| Optimal Detection Range | 2–60 m |
| Horizontal Beamwidth | 1° |
| Vertical Beamwidth | 20° |
| Maximum Number of Beams | 768 |
| Beam Spacing | 0.18° |
| Range Resolution | 1.3 cm |
| Update Rate | Up to 25 Hz |

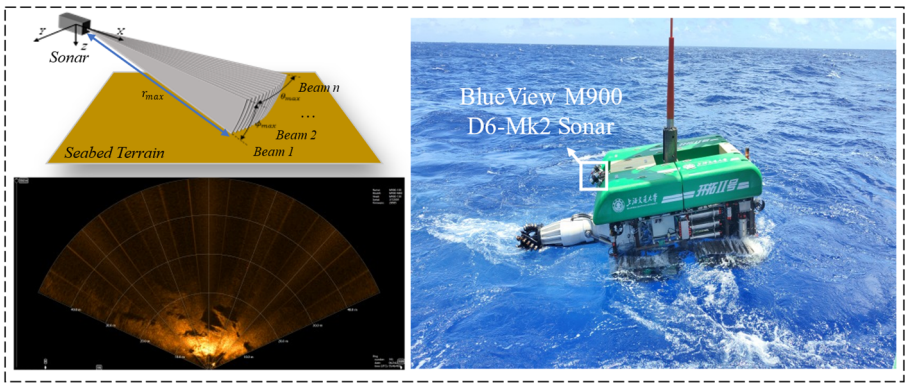


**Figure 6.** Overview of the DSMV *Pioneer II* platform and Forward-Looking Sonar system used in this study.

### *4.2. No-Reference Quantitative Evaluation of Denoising Results*

Given these constraints, two commonly used metrics from the current No-Reference Image Quality Assessment (NR-IQA) framework, namely the Naturalness Image Quality Evaluator (NIQE) [26] and the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [27], were adopted for the comparative analysis of images before and after denoising. It should be noted that both NIQE and BRISQUE are designed based on the statistical properties of natural optical images, relying heavily on assumptions regarding texture, brightness distribution, and noise characteristics that differ significantly from those of Forward-Looking Sonar images. FLS imagery is characterized by strong speckle noise, low contrast, and polar coordinate distortions, limiting the applicability of these metrics for sonar-image denoising evaluation [28].

Currently, there is no widely accepted NR-IQA standard specifically designed for FLS images [29]. Nonetheless, the NIQE and BRISQUE scores for multiple deep-sea scenes, including rugged rocky terrains and fine sediment areas, are reported before and after denoising to provide a reference. The experimental results are summarized in Figure 7 and Table 2. Although both the NIQE and BRISQUE scores increased after denoising, visual inspection revealed that the stripe noise along the fan-shaped radial direction was significantly suppressed. Moreover, key regions such as highlights and shadows, which are crucial for obstacle-height estimation and three-dimensional mapping [7], showed enhanced edge and detail features. The effective delineation of the FLS detection range was

also improved, substantially increasing the reliability of subsequent image matching and mosaicking tasks.

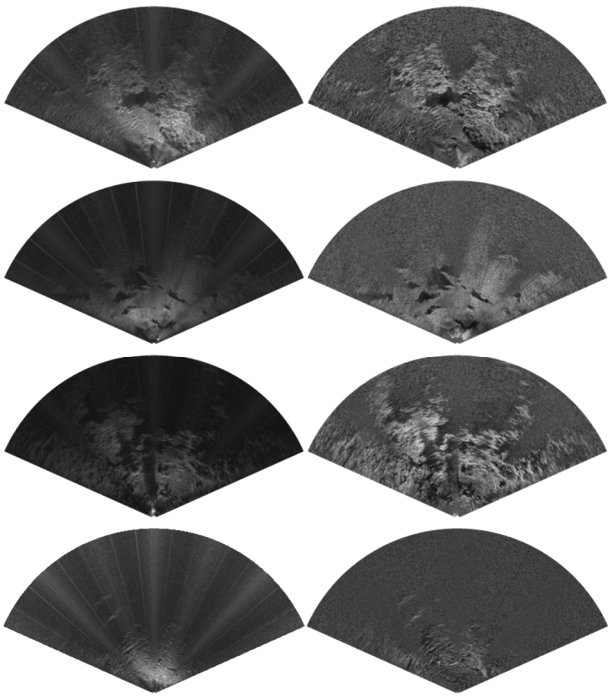


**Figure 7.** Visual comparison of denoising results across different deep-sea scenes. From top to bottom: S1 to S4. Each row shows (**a**) the original image and (**b**) the denoised result. The approximate detection ranges of S1–S4 are 40 m, 50 m, 80 m, and 80 m, respectively, with a horizontal field of view of ~130°.

**Table 2.** Quantitative comparison of image quality before and after denoising across four deep-sea scenes using NIQE and BRISQUE metrics. Lower scores indicate better perceptual quality.

| Scene ID | NIQE (Before) | BRISQUE (Before) | NIQE (After) | BRISQUE (After) |
|---|---|---|---|---|
| S1 | 5.138 | 43.451 | 6.016 | 54.147 |
| S2 | 4.665 | 38.006 | 5.609 | 53.102 |
| S3 | 4.987 | 40.441 | 5.477 | 46.230 |
| S4 | 4.914 | 37.371 | 5.555 | 50.392 |

Previous studies have pointed out the limitations of existing NR-IQA methods for sonar-image evaluation and have highlighted the need for further research in sonar-imaging quality assessment [30]. Therefore, in subsequent sections, practical performance metrics from the mosaicking tasks are employed as supplementary validation for denoising effectiveness. In the future, there are research plans to develop FLS-specific NR-IQA methods tailored to deep-sea mining tasks, aiming to establish more scientifically rigorous and application-oriented evaluation standards for sonar-image denoising algorithms.

### *4.3. Performance Evaluation of Two-Frame Feature Matching and Registration*

To systematically evaluate the performance of the proposed method, multiple pairs of adjacent FLS images with overlapping regions were selected. The effectiveness of the proposed method was compared with that of SuperPoint [22], SIFT [31], A-KAZE [32], and the Midline Template Matching (MTM) method [33] on the image mosaicking task. First, each algorithm was applied to extract and match features on the same pair of adjacent frames, and the number of matching points was recorded to assess the feature-extraction

capability on denoised images. Subsequently, RANSAC was used to fit an affine model to the matching points, and the inlier ratio was calculated to evaluate the spatial consistency of the matches. Furthermore, based on the estimated affine transformation, the matching points from the source image were projected onto the target image, and the reprojection error was computed as the mean Euclidean distance over all inliers to quantify registration accuracy. The reprojection error is calculated as

$$E = \frac{1}{N}\sum_{i=1}^{N}\left\|\hat{x}_i - x_i'\right\|_2 \tag{17}$$

where $\hat{x}_i$ denotes the transformed source feature point using the estimated affine matrix, and $x_i'$ represents the corresponding ground-truth matching point in the target image.

To further analyze the robustness of the algorithms under different RANSAC thresholds, the number of inliers was recorded as a function of the threshold, as shown in Figure 8. This metric reflects the stability of the matching performance under varying tolerance levels. The results indicate that traditional algorithms such as SIFT and A-KAZE are sensitive to noise under low threshold conditions, resulting in sparse matches. Although their inlier counts increase with higher thresholds, they tend to saturate. SuperPoint maintains a certain number of matches under high thresholds but suffers from insufficient accuracy. In contrast, the proposed method consistently maintains the highest inlier growth curve across the entire threshold range, demonstrating superior error tolerance and stability. This performance is particularly advantageous under challenging deep-sea imaging conditions characterized by strong speckle noise and weak feature responses.

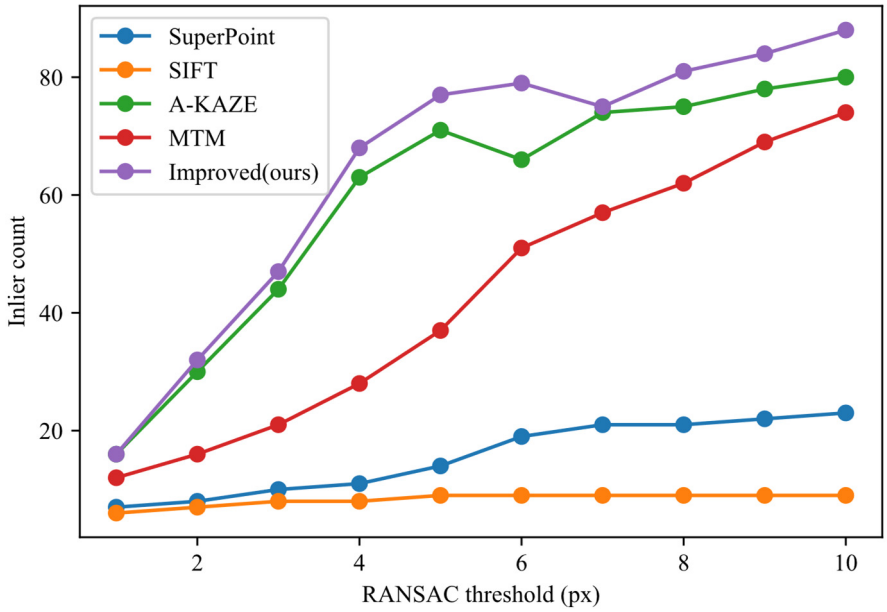


**Figure 8.** Effect of RANSAC threshold on inlier count across matching algorithms.

Furthermore, Figure 9 presents a comparison of the average reprojection error and matching accuracy across four typical scene pairs for each algorithm. The results show that the proposed method consistently achieves lower reprojection errors across multiple scenarios, while maintaining a matching accuracy exceeding 70%, outperforming traditional feature-based and template-matching methods. In particular, in challenging scenes such as S2 and S3, which exhibit significant local repetitive structures or shadow interference, the proposed method maintains good precision and consistency. These outcomes highlight the effectiveness of the uncertainty modeling incorporated into the algorithm. This is because

the proposed matching strategy effectively handles repetitive textures and shadow-induced ambiguities through uncertainty-aware refinement and geometric consistency filtering. It is also worth noting that the performance on S3 is slightly lower than in other scenarios, which aligns with its nature as a feature-sparse flat sediment terrain. This reflects the inherent difficulty of such environments and highlights the potential for further enhancing adaptability in future work.

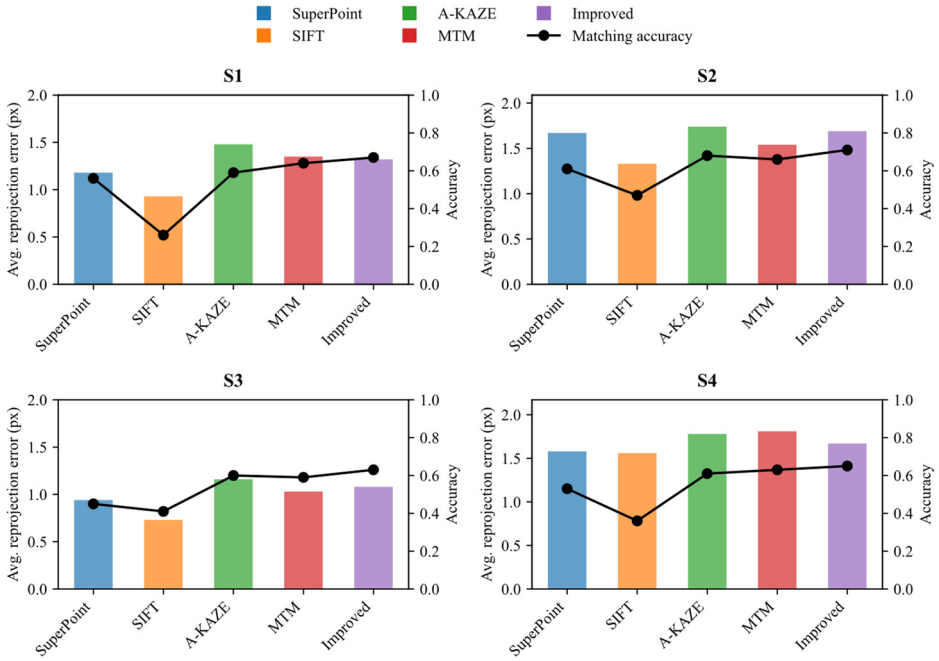


**Figure 9.** Comparison of average reprojection error and matching accuracy of different feature-matching algorithms on four neighboring FLS frame pairs.

In addition to matching accuracy, we also evaluated the runtime performance of the algorithm. The results show that the proposed method offers a slight advantage in per-frame processing time compared to other approaches, with the matching stage achieving an average reduction of approximately 4–10 milliseconds. The overall average processing time per frame is 42 milliseconds across different scenarios, which meets the real-time requirements of DSMV operations.

To further compare the matching quality from a visual perspective, Figure 10 presents the visualization of matching lines generated by various methods under different seafloor terrains. It can be observed that traditional methods suffer from dense mismatches and missing matches in texture-sparse regions. In contrast, the proposed method produces more uniform and concentrated matching lines, with key-points accurately aligned along structural edges or salient regions, demonstrating good adaptability across complex backgrounds such as rocks, gullies, and sediments.

In summary, the experiments in this section validate the effectiveness of the proposed strategy for adjacent-frame mosaicking tasks in sonar imagery. Compared with traditional methods, the proposed approach achieves higher matching robustness and geometric registration accuracy without relying on any external-sensor information. This is particularly important in practical deep-sea operations, where sensor drift and signal degradation often render external pose estimates unreliable. The image-driven design of our method ensures greater robustness and stability, providing a more reliable foundation for subsequent multi-frame mosaicking.

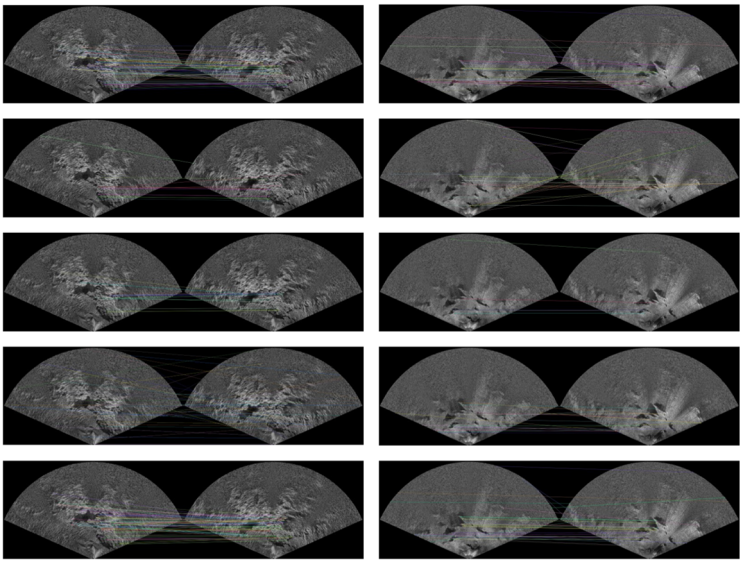

**Figure 10.** Visualization of matching results under different deep-sea scenes. From top to bottom: A-KAZE, SIFT, SuperPoint, template matching method, and the proposed improved method.

### *4.4. Demonstration of Deep-Sea Image Stitching Using Multi-Frame Matching*

After validating the local registration accuracy in Section 4.3, large-scale multi-frame mosaicking experiments were conducted to further evaluate the generality and robustness of the proposed framework on two different datasets. The first dataset consists of the Barrel Roll sonar-image dataset, captured in a water tank using the ARIS Explorer 3000 system (Sound Metrics [34]); the second dataset comprises the real-world FLS image sequences from the deep-sea mining area described in Section 4.1.

The water-tank dataset poses typical challenges such as prominent acoustic shadows, blurring, and echo interference. Figure 11a,b show examples of raw frames and the corresponding mosaicking results, respectively. As observed, even in the noisiest regions, the stitching boundaries are naturally aligned, and local details are well preserved. No noticeable misalignments or discontinuities are found in texture-sparse areas, indicating that the proposed fusion and mosaicking strategy effectively suppresses brightness non-uniformity and striping noise, maintaining a smooth and coherent panoramic image. Subsequently, the same mosaicking process was applied to multiple real-world FLS sequences collected from the deep-sea mining area. It is noteworthy that due to constraints such as tether drag and water currents, the DSMV typically scans slowly with minimal amplitude, resulting in a concentric fan-like overlay structure in most mosaicking results. Figure 12 illustrates several typical mosaicking examples. It can be observed that the outermost pixels of the initial frame, representing the first detection, are strictly preserved, and the seafloor terrain features are largely matched correctly with continuous and complete details. These results demonstrate the framework's excellent noise suppression and seamless fusion capabilities, consistent with the outcomes observed in the water-tank experiments.

In terms of overall coverage performance, the proposed method maintains natural alignment and detailed preservation even in areas with significant structural variations or strong echo interference, for both water-tank and deep-sea datasets. It also demonstrates good geometric consistency and texture fidelity in weak-texture or high-noise regions, highlighting the robustness of the global optimization framework under multi-interference conditions. It should be noted that due to inherent sonar speckle noise and inter-frame brightness variations, the boundary smoothing method currently employed can still leave faint striping artifacts along the seams. Future work will consider incorporating more

sophisticated brightness compensation methods or multi-band fusion strategies to further suppress noise discrepancies.

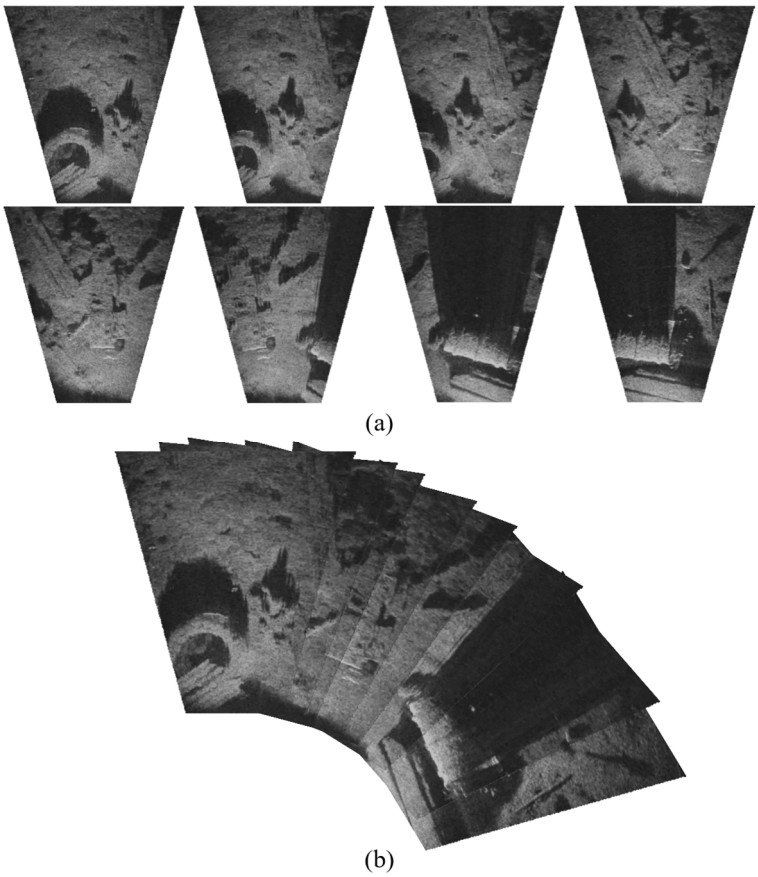


**Figure 11.** Stitching results of the ARIS Explorer 3000 Barrel Roll water-tank experiment: (**a**) raw fan-shaped frames; (**b**) stitched sonar panorama. Each frame covers a horizontal field of view of ~130°, with an effective range of approximately 3–5 m.

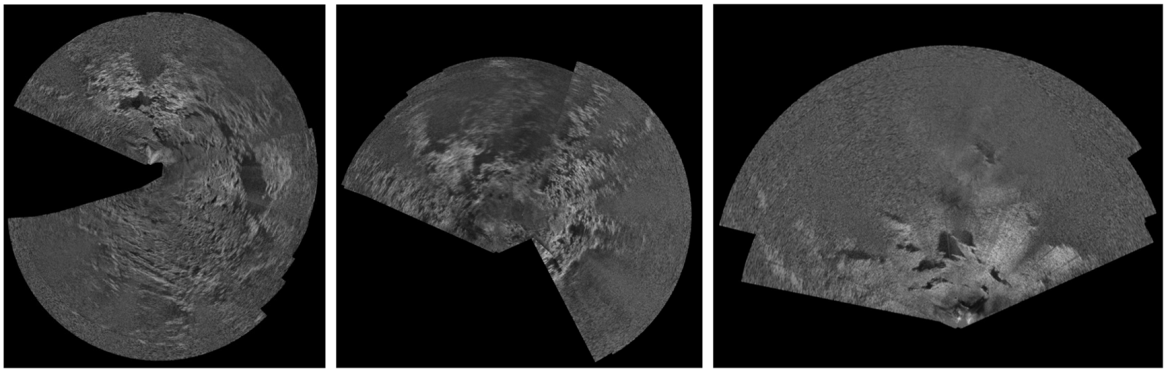

**Figure 12.** Stitched panoramas of DSMV deep-sea mining site sonar data.

To quantitatively evaluate the system's stability and perception range extension capability, the Stitching Success Rate (SSR) and the Perception Range Extension Factor (PRE) were introduced as evaluation metrics, defined in the following way:

$$SSR = \frac{N_{success}}{N_{total}} \times 100\% \tag{18}$$

where $N_{success}$ is the number of successfully fused node pairs with registration errors below a predefined threshold, and $N_{total}$ is the total number of node pairs involved in the mosaicking process.

$$PRE = \frac{A_{stitched}}{A_0} \tag{19}$$

where $A_{stitched}$ represents the effective coverage area of the resulting panoramic image. The theoretical coverage area of a single sonar frame, $A_0$, is calculated based on the horizontal field of view $\theta_h$, and the maximum and minimum detection ranges $R_{max}$ and $R_{min}$, respectively, as

$$A_0 = \frac{1}{2}\theta_h\left(R_{max}^2 - R_{min}^2\right) \tag{20}$$

The stitching success rate and perception range extension factors across different deep-sea scenarios are summarized in Table 3. The proposed method achieves an average SSR of 93% across various complex environments, with only a small number of mismatches or missed matches occurring in extremely noisy or textureless scenes, which remains within acceptable limits. Meanwhile, an average PRE of 297% is achieved, indicating significant perception range enhancement.

**Table 3.** Evaluation of stitching stability and perception expansion across typical DSMV trajectories.

| Scene ID | Trajectory Type | Frame Count | SSR(%) | PRE(%) |
|---|---|---|---|---|
| S1 | Non-loop | 16 | 93.8 | 383 |
| S2 | Loop | 27 | 92.6 | 247 |
| S3 | Loop | 31 | 90.3 | 256 |
| S4 | Non-loop | 25 | 92 | 303 |

Both the qualitative and quantitative analyses demonstrate that the proposed mosaicking framework not only produces high-quality panoramic images under controlled conditions but also maintains excellent performance in real-world, complex marine environments. These capabilities provide strong support for large-scale deep-sea terrain mapping and DSMV path planning.

## 5. Conclusions

This study addresses the challenges of strong noise interference, brightness inconsistency, and cumulative positioning errors in FLS imagery for deep-sea mining operations. A robust FLS image registration and mosaicking framework is proposed, operating independently of external-sensor assistance. The system integrates a two-stage preprocessing strategy—combining physical modeling with multi-frame fusion—for effective noise suppression and brightness normalization. With an uncertainty-aware feature matching mechanism and a multi-scale weighted registration process, the framework significantly improves matching robustness and geometric consistency. Experimental results demonstrate that the proposed method consistently achieves over 70% matching accuracy with reduced reprojection error across multiple terrain types, and it reaches a 93% stitching success rate and a 297% perception range extension on real-world deep-sea datasets.

Future work will focus on developing FLS-specific NR-IQA metrics to improve the scientific rigor of image-quality evaluation. Such metrics are expected to leverage the unique imaging principles of FLS and provide more accurate and application-relevant image-quality evaluation by capturing its statistical and structural characteristics, which are not well-represented by existing metrics designed for natural images. Moreover, efforts will be made to further optimize computational efficiency to meet the increasing demands for real-time performance in practical engineering applications. The results

confirm that the proposed framework successfully achieves the intended objective under realistic deep-sea conditions.

**Author Contributions:** Conceptualization, X.L.; Methodology, X.L.; Software, X.L., C.L., E.Z. and W.X.; Validation, X.L. and E.Z.; Formal analysis, X.L. and W.X.; Investigation, X.L.; Resources, J.Y.; Data curation, C.L. and W.X.; Writing—original draft, X.L.; Writing—review & editing, X.L., J.Y., C.L., E.Z. and W.X.; Visualization, X.L. and C.L.; Supervision, J.Y.; Project administration, J.Y.; Funding acquisition, J.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Shanghai Innovation Action Plan of Science and Technology grant number (19DZ1207300) and Major Projects of Strategic Emerging Industries in Shanghai grant number (BH3230001).

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

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Leng, D.; Shao, S.; Xie, Y.; Wang, H.; Liu, G. A brief review of recent progress on deep sea mining vehicle. *Ocean Eng.* **2021**, *228*, 108565. [CrossRef]
2. Liu, X.; Yang, J.; Xu, W.; Chen, Q.; Lu, H.; Chai, Y.; Lu, C.; Xue, Y. DSM-Net: A multi-scale detection network of sonar images for deep-sea mining vehicle. *Appl. Ocean Res.* **2025**, *158*, 104551. [CrossRef]
3. Cao, Y.; Xu, C.; Li, J.; Zhou, T.; Lin, L.; Chen, B. Underwater gas leakage flow detection and classification based on multibeam forward-looking sonar. *J. Mar. Sci. Appl.* **2024**, *23*, 674–687. [CrossRef]
4. Ferreira, F.; Djapic, V.; Micheli, M.; Caccia, M. Forward looking sonar mosaicing for mine countermeasures. *Annu. Rev. Control* **2015**, *40*, 212–226. [CrossRef]
5. Zhou, X.; Mizuno, K.; Zhang, Y.; Tsutsumi, K.; Sugimoto, H. Acoustic Camera-Based Adaptive Mosaicking Framework for Underwater Structures Inspection in Complex Marine Environments. *IEEE J. Ocean Eng.* **2024**, *49*, 1549–1573. [CrossRef]
6. Lu, C.; Yang, J.; Leira, B.J.; Skjetne, R.; Mao, J.; Chen, Q.; Xu, W. High-traversability and efficient path optimization for deep-sea mining vehicles considering complex seabed environmental factors. *Ocean Eng.* **2024**, *313*, 119500. [CrossRef]
7. Xu, W.; Yang, J.; Wei, H.; Lu, H.; Tian, X.; Li, X. Seabed mapping for deep-sea mining vehicles based on forward-looking sonar. *Ocean Eng.* **2024**, *299*, 117276. [CrossRef]
8. Lu, C.; Yang, J.; Lu, H.; Lin, Z.; Wang, Z.; Ning, J. Adaptive bi-level path optimization for deep-sea mining vehicle in non-uniform grids considering ocean currents and dynamic obstacles. *Ocean Eng.* **2025**, *315*, 119835. [CrossRef]
9. Wang, N.; Chen, Y.; Wei, Y.; Chen, T.; Karimi, H.R. UP-GAN: Channel-spatial attention-based progressive generative adversarial network for underwater image enhancement. *J. Field Robot.* **2024**, *41*, 2597–2614. [CrossRef]
10. Petillot, Y.; Ruiz, I.T.; Lane, D.M. Underwater vehicle obstacle avoidance and path planning using a multi-beam forward looking sonar. *IEEE J. Ocean Eng.* **2001**, *26*, 240–251. [CrossRef]
11. Liu, X.; Yang, J.; Xu, W.; Zhang, E.; Lu, C. FLS-GAN: An end-to-end super-resolution enhancement framework for FLS terrain in deep-sea mining vehicles. *Ocean Eng.* **2025**, *332*, 121369. [CrossRef]
12. Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; Aila, T. Noise2Noise: Learning image restoration without clean data. *arXiv* **2018**, arXiv:1803.04189.
13. Batson, J.; Royer, L. Noise2self: Blind Denoising by Self-Supervision. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 524–533.
14. Ye, T.; Deng, X.; Cong, X.; Zhou, H.; Yan, X. Parallelisation Strategy of Non-local Means Filtering Algorithm for Real-time Denoising of Forward-looking Multi-beam Sonar Images. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 13226–13243. [CrossRef]
15. Hurtos, N.; Ribas, D.; Cufi, X.; Petillot, Y.; Salvi, J. Fourier-based registration for robust forward-looking sonar mosaicing in low-visibility underwater environments. *J. Field Robot.* **2015**, *32*, 123–151. [CrossRef]
16. Hansen, T.; Birk, A. Using Registration with Fourier-Soft in 2d (fs2d) for Robust Scan Matching of Sonar Range Data. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023; pp. 3080–3087.
17. Tsutsumi, K.; Mizuno, K.; Sugimoto, H. Accuracy Enhancement of High-Resolution Acoustic Mosaic Images Using Positioning and Navigating Data. In Proceedings of the Global Oceans 2020: Singapore–US Gulf Coast, Biloxi, MS, USA, 5–30 October 2020; pp. 1–4.

18. Li, B.; Yan, W.; Li, H. A combinatorial registration method for forward-looking sonar image. *IEEE Trans. Ind. Inform.* **2023**, *20*, 2682–2691. [CrossRef]
19. Su, J.; Li, H.; Qian, J.; An, X.; Qu, F.; Wei, Y. A blending method for forward-looking sonar mosaicing handling intra-and inter-frame artifacts. *Ocean Eng.* **2024**, *298*, 117249. [CrossRef]
20. Shang, X.; Dong, L.; Fang, S. Sonar Image Matching Optimization Using Convolution Approach Based on Clustering Strategy. *IEEE Geosci. Remote Sens. Lett.* **2024**, *22*, 1500705. [CrossRef]
21. Wei, M.; Bian, H.; Zhang, F.; Jia, T. Beam-domain image mosaic of forward-looking sonar using expression domain mapping model. *IEEE Sens. J.* **2022**, *23*, 4974–4982. [CrossRef]
22. DeTone, D.; Malisiewicz, T.; Rabinovich, A. Superpoint: Self-Supervised Interest Point Detection and Description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 224–236.
23. Sun, J.; Shen, Z.; Wang, Y.; Bao, H.; Zhou, X. LoFTR: Detector-Free Local Feature Matching with Transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 8922–8931.
24. Gode, S.; Hinduja, A.; Kaess, M. Sonic: Sonar Image Correspondence Using Pose Supervised Learning for Imaging Sonars. In Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, 13–17 May 2024; pp. 3766–3772.
25. Johannsson, H.; Kaess, M.; Englot, B.; Hover, F.; Leonard, J. Imaging Sonar-Aided Navigation for Autonomous Underwater Harbor Surveillance. In Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan, 18–22 October 2010; pp. 4396–4403.
26. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "completely blind" image quality analyzer. *IEEE Signal Process. Lett.* **2012**, *20*, 209–212. [CrossRef]
27. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [CrossRef]
28. Chen, W.; Cai, B.; Zheng, S.; Zhao, T.; Gu, K. Perception-and-Cognition-Inspired Quality Assessment for Sonar Image Super-Resolution. *IEEE Trans. Multimed.* **2024**, *26*, 6398–6410. [CrossRef]
29. Zheng, S.; Chen, W.; Zhao, T.; Wei, H.; Lin, L. Utility-Oriented Quality Assessment of Sonar Image Super-Resolution. In Proceedings of the OCEANS 2022, Hampton Roads, VA, USA, 17–22 October 2022; pp. 1–5.
30. Cai, B.; Chen, W.; Zhang, J.; Junejo, N.U.R.; Zhao, T. Unified No-Reference Quality Assessment for Sonar Imaging and Processing. *IEEE Trans. Geosci. Remote Sens.* **2024**, *63*, 5902711. [CrossRef]
31. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
32. Alcantarilla, P.F.; Solutions, T. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell.* **2011**, *34*, 1281–1298.
33. Liu, H.; Ye, X. Forward-looking sonar image stitching based on midline template matching in polar image. *IEEE Trans. Geosci. Remote Sens.* **2023**, *62*, 4201210. [CrossRef]
34. Sound Metrics. *Image Gallery*; Sound Metrics: Washington, DC, USA, 2023. Available online: http://www.soundmetrics.com (accessed on 31 May 2025).


**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.