

Article

Performance Comparison of Feature Detectors on Various Layers of Underwater Acoustic Imagery

Xiaoteng Zhou ¹, Shihao Yuan ², Changli Yu ^{1,*}, Hongyuan Li ³ and Xin Yuan ^{1,*}¹ School of Ocean Engineering, Harbin Institute of Technology, Weihai 264200, China² School of Computer Science and Technology, Harbin Institute of Technology, Weihai 264200, China³ College of Engineering, Peking University, Beijing 100871, China

* Correspondence: yuchangli@hitwh.edu.cn (C.Y.); xin.yuan@upm.es (X.Y.); Tel.: +86-138-6310-6670 (C.Y.); +86-178-6270-9967 (X.Y.)

Abstract: Image feature matching is essential in many computer vision applications, and the foundation of matching is feature detection, which is a crucial feature quantification process. This manuscript focused on detecting more features from underwater acoustic imageries for further ocean engineering applications of autonomous underwater vehicles (AUVs). Currently, the mainstream feature detection operators are developed for optical images, and there is not yet a feature detector oriented to underwater acoustic imagery. To better analyze the suitability of existing feature detectors for acoustic imagery and develop an operator that can robustly detect feature points in underwater imageries in the future, this manuscript compared the performance of well-established handcrafted feature detectors and that of the increasingly popular deep-learning-based detectors to fill the gap in the literature. The datasets tested are from the most commonly used side-scan sonars (SSSs) and forward-looking sonars (FLSs). Additionally, the detection idea of these detectors on the acoustic imagery phase congruency (PC) layer was innovatively proposed with the aim of finding a solution that balances detection accuracy and speed. The experimental results show that the ORB (Oriented FAST and Rotated BRIEF) and BRISK (Binary Robust Invariant Scalable Keypoints) detectors achieve the best overall performance, the FAST detector is the fastest, and the PC and Sobel layers are the most favorable for implementing feature detection.

Keywords: image matching; feature detection; underwater acoustic image; ocean engineering; AUV



Citation: Zhou, X.; Yuan, S.; Yu, C.; Li, H.; Yuan, X. Performance Comparison of Feature Detectors on Various Layers of Underwater Acoustic Imagery. *J. Mar. Sci. Eng.* **2022**, *10*, 1601. <https://doi.org/10.3390/jmse10111601>

Academic Editors: Tracianne B Neilsen and Haiqiang Niu

Received: 6 September 2022

Accepted: 18 October 2022

Published: 31 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

In recent years, human and industrial development has become more closely related to the ocean, with more infrastructure built off the coast and more renewable energy obtained from the deep sea. To more safely and efficiently explore the ocean, AUVs are becoming the predominant underwater exploration equipment, which can effectively replace divers for port and offshore infrastructure maintenance and can also perform exploration tasks in uncharted waters. When the AUVs perform exploration tasks in deep water, sonar sensors play an important role, with the forward-looking sonar (FLS) assisting AUVs to avoid obstacles for their safety and the side-scan sonar (SSS) providing high-resolution views of the seafloor to help researchers understand the deep-sea operating area [1]. With the continuous development of the electronics industry and remote-sensing technology, the detection performance of sonar equipment has been dramatically improved. The two main sonar types, FLS and SSS, can provide rich underwater acoustic imageries. The number and resolution of these images are constantly increasing. Therefore, it is crucial and far-reaching to design an efficient vision-processing algorithm to extract the information from underwater acoustic images, which, in turn, can serve deep-sea exploration activities. Image matching, as a principal technology in computer vision, applies the knowledge of graphics and corresponding mathematical theory to analyze the correspondence of

grayscale, content, structure, texture, and semantic information in one or more groups of images, finally, to find similar image targets. Typical matching methods can be divided into region-matching and feature-point-matching approaches [2]. Image matching is the essential work of image analysis and processing and also a fundamental prerequisite for tasks such as target tracking, visual navigation, map stitching, and 3D reconstruction [3].

The above tasks are bound to be faced in practical ocean engineering tasks, so there is an opportunity and far-reaching significance to developing an effective acoustic image-matching algorithm. Underwater acoustic imageries are different from natural light images in that they have low contrast, a narrow grayscale range, high spatial correlation of adjacent pixels, and poor quality. Traditional matching algorithms designed for optical images, such as SIFT (scale-invariant feature transform) [4] and SURF (speeded-up robust features) [5], cannot effectively perform the matching task of acoustic images. The contrast between the large and small size of underwater acoustic images and the lack of the number of targets of interest leads to the inappropriateness of area-based matching methods for underwater task scenarios and considering the real-time requirements of underwater detection tasks, and the current mainstream research is based on feature-point-matching methods [6,7], which usually include feature detection, feature description, and outlier rejection. Among them, feature detection is the basis of the whole process, representing the salient structural information in the image. It also determines the number of correct matching pairs and the accuracy of the point position in the two images to be matched. In recent years, more research results have been achieved in the feature description stage, but many bottlenecks have not been solved in the feature detection stage. Meanwhile, compared with the development of optical image matching technology, the technique in the field of underwater acoustic imagery matching is lagging, which is mainly limited by the imaging shortcomings of underwater acoustic imageries and external factors such as the lack of samples available for research [8].

In recent years, with the development of the electronics industry and the update of sensor technology, the quantity and quality of underwater acoustic imageries available to researchers have been guaranteed. Meanwhile, the research on sonar image feature detection has gradually started to be emphasized. When a search is performed on the Web of Science using the subject term “sonar image feature detection” and setting the search date from 2000 to the present, the following report could be derived, as shown in Figure 1. There has been a significant increase in the number of publications and citation frequency of related research since 2013, which may be related to the rise of deep-learning techniques within the field of computer vision, where new techniques have brought opportunities for underwater acoustic imagery feature detection tasks.

In this paper, a framework of comparative feature detection algorithms was proposed to find a high-quality feature detection scheme so as to fundamentally advance the underwater acoustic image matching technology to meet the needs of the downstream tasks such as the underwater visual navigation, seabed mapping, and environment reconstruction. Moreover, the relevant experimental reports are also summarized to help newcomers in this field as well as researchers who are proficient in image processing but do not understand sonar images to participate in acoustic perception tasks. Finally, some potential applications of deep-learning methods in this research field are summarized.

The rest of the manuscript is organized as follows: Section 2 briefly describes the research related to sonar image matching and the recent developments; Section 3 presents the feature detection approach taken in this paper; Section 4 describes the experimental setup, including dataset introduction, evaluation criteria, and parameter settings; Section 5 describes and compares the experimental results; Section 6 discusses the experimental shortcomings and outlook; finally, conclusions are drawn in Section 7.

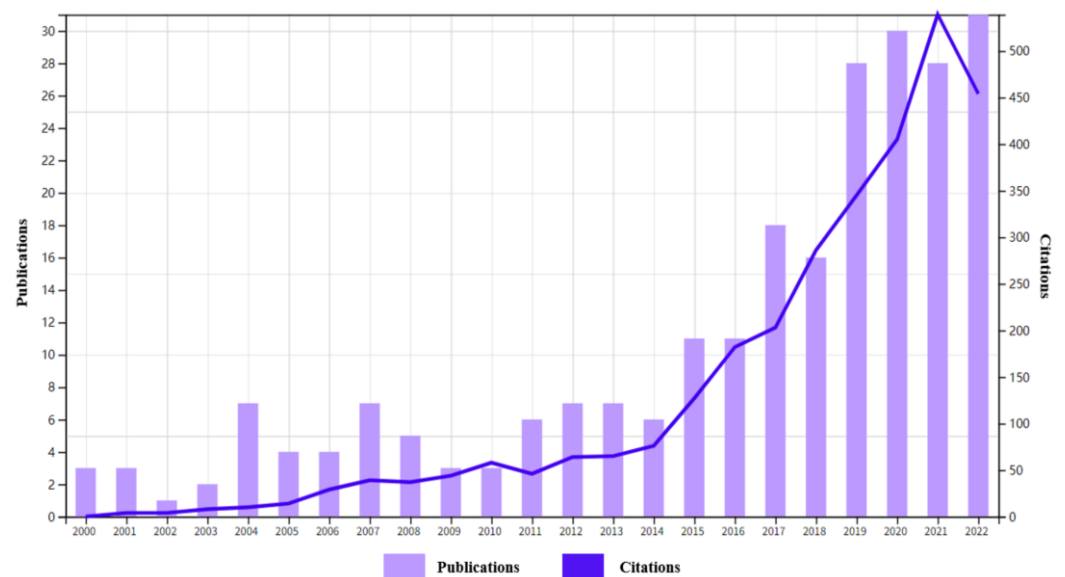


Figure 1. Status of publications related to sonar image feature detection since 2000.

2. Related Work

In natural light images, feature points are generally selected from pixel points that are easily distinguishable and visually salient in the image, such as straight-line segment intersections, corner points, or local center-of-gravity points [2]. Compared with natural light images, underwater acoustic images have poor imaging quality, severe noise interference, grayscale distortion, and weak edge features, all of which lead to the inefficiency of feature detectors initially designed for optical images. In addition, due to the significant differences between acoustic and optical imaging mechanisms, sonar images are usually accompanied by a narrow grayscale range, high spatial correlation of adjacent pixels, and insignificant grayscale variation, and there is also no publicly available and mature scheme for preprocessing sonar images. This situation also leads to almost failure of feature matching in the final underwater acoustic imageries, as evidenced by the low number of feature points detected and the absolute low matching accuracy.

In the study of underwater acoustic image matching, Daniel et al. [8] proposed to use acoustic shadow features and a segmentation algorithm to achieve SSS image matching, which is based on the geometric distribution characteristics of shadows in SSS images, echoes, and their corresponding critical position information. However, the authors used a manual combination of features rather than individual feature detectors, so the overall pipeline design lacked flexibility. Khater [9] proposed an SSS image feature detection method based on SUSAN (smallest univalue segment assimilating nucleus) and Harris corner point information, which can achieve better detection when the sonar image features are stable and uniform. However, in general, there is a large amount of noise and distortion on sonar images, and features are scarce and difficult to extract, so the generalization performance of this method is insufficient, and the application space in underwater scenes is limited. P. Vanish [6] investigated the effect of the classical SIFT algorithm on SSS images. The results showed that when the image features are rich and stable, SIFT could accomplish the SSS image-matching tasks. However, when there are apparent noise interference and image distortion in the sonar image, the matching effects of SIFT will also be significantly reduced. Zhang [10] proposed a SIFT-like feature detector and descriptor MBS-SIFT that could robustly cope with noise interference in multibeam imaging sonar images, which could more accurately capture the features of underwater objects compared with the classical SIFT method. Tao [11] proposed a combined SSS image-matching method in a priori position information, and SURF detectors are used to detect feature points from the SSS image that are stable to the affine transform for matching, and finally, the results are corrected with the help of RANSAC (random sample consensus). The results show that

the matching algorithm is time-consuming and has high accuracy. However, this method depends on a priori position data, which has limited application in practical underwater scenarios. Shang [12] proposed a way to automatically match and mosaic SSS images. First, the image overlap region between adjacent strips is automatically determined based on the sonar detection track lines and the width of the scanned strips. Then in the overlap region, the SURF detector is used to detect the feature points. Finally, the geographic coordinates of the feature points are used to constrain the method, which uses multiple constraint information to complete the matching between the SSS image pair, and the matching results are better but very dependent on external information. Peter King [7] compared the performance of several classical handcrafted feature-matching algorithms in matching sonar images using a pre-collected dataset. The results show that the SIFT and SURF methods match better when the sonar images are feature-rich and stable. In the complex landscape background, only the SURF method could barely complete the matching. However, the authors did not deeply dissect the feature detection link and only evaluated each algorithm regarding the overall matching effect. Tueller [13] compared the performance of seven commonly used feature detection algorithms on synthetic sonar image datasets with different background attributes by training an SVM (support vector machine) classifier to evaluate the performance of the detectors. However, the study only covered several traditional feature detectors and tested a single type of image. Ansari [14] investigated the detection effectiveness of two algorithms, SIFT and SURF, in underwater image matching, and the results showed that SIFT and SURF in feature extraction and matching of SSS images offer good performance; but this is limited to scenes where stable textures and independent targets exist, and their performance sharply degrades when hard sediments and rocks are encountered.

Although the above studies have achieved feature detection of sonar images in their respective assumed scenarios, it is evident that the detection is less than ideal, and there are strict requirements for the quality of the acoustic imageries themselves and the feature richness of the detected scenes, which reduce the generalization performances of these methods. When AUVs carry sonar equipment for underwater detection tasks, the idealized characteristics of acoustic images cannot be guaranteed in most cases due to the complexity and unknown nature of the underwater environment, so it is necessary to develop a robust acoustic imagery feature detection method, which is also the cornerstone for designing an overall algorithm for underwater acoustic imagery matching.

With the rapid development of deep-learning techniques within the field of image processing, CNN (convolutional neural network) based feature detectors have been generated, which can be used to extract shallow, in-depth, and combined multidimensional features in images to achieve feature point detection. Zhou [15] proposed to match underwater acoustic images using deep-learning-based detectors, feature descriptors, and style transfer algorithms, and the matching effect was quantified and visualized. The experimental results showed that even though the learned detectors and descriptors were obtained based on the optical image dataset training, they surpassed the traditional handcrafted operators, such as the SIFT, in terms of the overall effect. When the style transfer method was introduced, the matching results of acoustic images were further improved, and this study can provide a reference for the design of deep-learning-based matching algorithms for underwater acoustic images in the future.

The above review of related studies shows that the feature point-based matching method is more robust and flexible and is the leading way of feature matching for underwater acoustic images at present. This kind of method can also work well when there are complex geometric transformations and external interference between two images; therefore, it has been more widely used in underwater acoustic imagery matching. However, most of the literature studies used the existing optical matching methods due to the fact that it is very challenging to develop a high-precision acoustic matching algorithm with particular generalization performance, so it is necessary to individually design two links of feature detection and feature description, each of which has a significant impact on match-

ing results. Obviously, the first step of these processes is to find out a high-performance feature detection scheme according to the characteristics of underwater acoustic imagery.

According to the literature survey, existing literature does not independently develop an analysis of feature detection on underwater acoustic imagery, which is the motivation of this manuscript. We noticed that Oliveira [16] conducted research on the feature extraction of acoustic imagery involved in underwater navigation tasks, and the tested algorithms involved SURF, ORB, and BRISK. Their work achieved preliminary results, but the conclusion is limited to underwater navigation tasks, and the evaluated methods are not comprehensive enough to generalize to other underwater tasks, such as 3D reconstruction or underwater mapping.

In contrast, the research in this paper has expanded with multiple applications, devices, detection methods, and evaluation methods, aiming to obtain a more comprehensive solution. Meanwhile, the work in this manuscript is more focused, as reflected by the fact that the research content only explores the feature detection aspect in depth, as it is the cornerstone of feature description and matching. In addition to filling gaps in the existing literature, the study of this paper has made a more valuable contribution: on the one hand, due to the scarcity of underwater acoustic data, the research on deep-learning-based matching methods is still not mature enough, and many algorithms need further experimentation and validation—this manuscript carried out valuable work in this area; on the other hand, the research pipeline designed is flexible and can be updated according to the actual mission requirements to cope with the complex and changing underwater environment information.

The overall research pipeline proposed in this manuscript is shown in Figure 2. First, input underwater acoustic data, such as from FLS and SSS, and visualize them as a set of digital images. Meanwhile, the images can be pre-processed to generate regions of interest (ROIs) Second, extract layers ready for feature detection, such as grayscale, gradient, and PC layers. Third, set up detectors for feature extraction, such as handcrafted and deep-learning-based. Then, the best feature detection scheme tailored to the current scene is used for downstream tasks such as matching, mapping, and tracking.

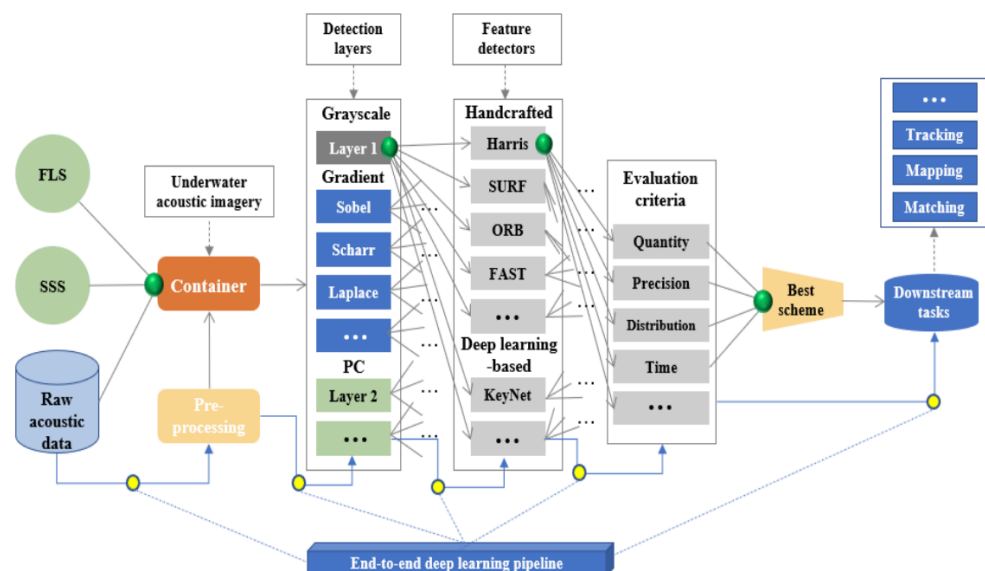


Figure 2. Research pipeline proposed in this manuscript.

3. Methodology

Current feature point-based matching methods usually use grayscale or gradient information to detect keypoints, which are a set of potential feature points. However, this kind of method is easily affected by contrast, grayscale distortion, and noise, which are negative factors commonly found in underwater detection scenarios, and they will make

the detection of keypoints unstable and inaccurate. On the one hand, due to the wide noise distribution, some points will be incorrectly detected as keypoints, whereas some correct keypoints may not be detected; on the other hand, low contrast and grayscale distortion phenomena on acoustic images will directly affect the number of detected feature points. Inaccurate feature point detection results will directly affect the design of subsequent descriptors, affecting the final matching results.

Therefore, to address these problems, this manuscript proposed to apply the PC (phase congruency) principle to underwater acoustic imagery feature detection. The PC layer is not affected by image illumination and contrast, and is robust to noise interference; it also provides rich texture, edge, and structure information consistent with human vision, so using the PC layer in underwater acoustic imagery feature detection tasks has the suitable space and apparent advantages.

To more clearly and intuitively understand the effect of PC processing of underwater acoustic imagery, Figure 3 shows the PC layer and original grayscale layer of the raw SSS image while comparing them with the classical gradient layers, which involve the Sobel, Scharr, and Laplacian layers.

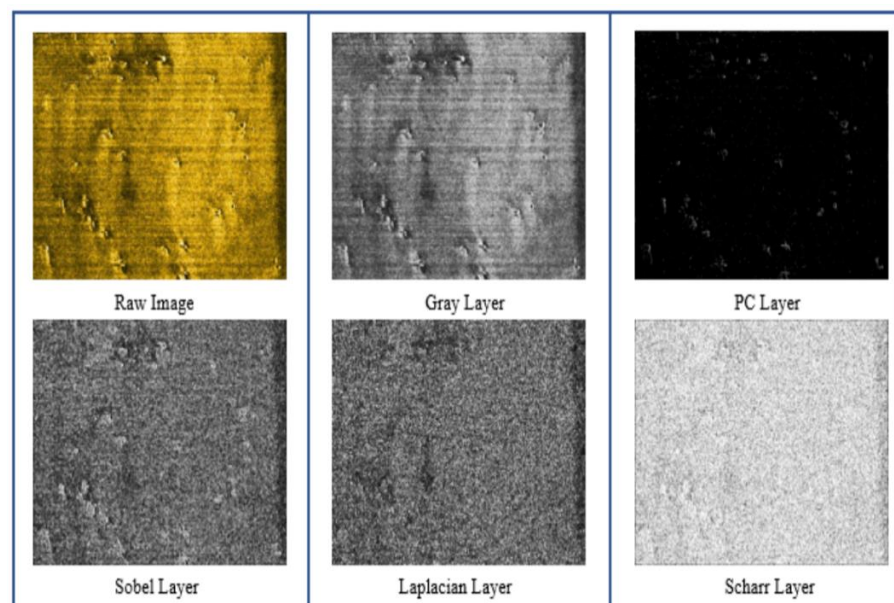


Figure 3. Schematic diagram of various layers of underwater acoustic imagery.

3.1. Phase Congruency Layer

Phase Congruency Solving

The principle of PC started in 1981 when Oppenheim and Lim [17] first revealed the effect of phase information on image features, and subsequent researchers have extended PC in various directions [18–20]. One of the more classical PC-solving algorithms in image processing is the PC based on the multiscale multidirectional log-Gabor filter proposed by Kovess in 2003 [21]. The log-Gabor filter is expressed as Equation (1):

$$LG(\hat{w}) = \exp\left(\frac{-[\ln(\hat{w}/\hat{w}_0)]^2}{2[\ln(\sigma_{\hat{w}}/\hat{w}_0)]^2}\right) \quad (1)$$

assuming that $I(x, y)$ denotes the underwater acoustic imagery, and $M_{n\bar{o}}^e$ and $M_{n\bar{o}}^o$ represent the even- and odd-symmetric log-Gabor wavelets at the n_{th} scale and $\bar{o}_{th}$ direction, respectively. The two wavelet functions are convolved with the image signal to obtain the

corresponding components $e_{n\bar{o}}(x, y)$ and $o_{n\bar{o}}(x, y)$, respectively, and the overall computational process is described as Equation (2):

$$[e_{n\bar{o}}(x, y), o_{n\bar{o}}(x, y)] = [I(x, y) \times M_{n\bar{o}}^e, I(x, y) \times M_{n\bar{o}}^o] \quad (2)$$

The amplitude and phase of the underwater acoustic imagery after wavelet transform with scale n and direction $\bar{o}$ are as follows:

$$A_{n\bar{o}}(x, y) = \sqrt{e_{n\bar{o}}(x, y)^2 + o_{n\bar{o}}(x, y)^2} \quad (3)$$

$$\phi_{n\bar{o}}(x, y) = \arctan[o_{n\bar{o}}(x, y) / e_{n\bar{o}}(x, y)] \quad (4)$$

Combining the analysis results in each direction, the PC solving of the underwater acoustic imagery can be obtained as Equation (5):

$$PC_2(x, y) = \frac{\sum_n \sum_{\bar{o}} W_{\bar{o}}(x, y) [A_{n\bar{o}}(x, y) \Delta \Phi_{n\bar{o}}(x, y) - \hat{T}]}{\sum_n \sum_{\bar{o}} A_{n\bar{o}}(x, y) + \xi} \quad (5)$$

where $W_{\bar{o}}(x, y)$ indicates the two-dimensional (2D) frequency expansion weight factor, and $\Delta \Phi_{n\bar{o}}(x, y)$ is the 2D phase deviation function.

3.2. Classic Gradient Layers

3.2.1. Sobel Layer Solving

The Sobel operator is a discrete differentiation operator used to compute the approximate gradient of the image grayscale function. Furthermore, it is also a joint Gaussian smoothing plus differentiation operator with solid noise resistance.

Assume an underwater acoustic image I , derive it in both horizontal and vertical directions using the Sobel operator, and then perform a convolution operation on I applying a kernel of size K . To ensure the accuracy and operation speed, K is usually taken as 3, where G_x and G_y denote the convolution results in the horizontal and vertical directions, respectively, and the calculation formula is

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \times I \quad (6)$$

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \times I \quad (7)$$

Then, for any pixel on the underwater acoustic imagery, the gradient of its location can be calculated by using Equation (8).

$$G = \sqrt{G_x^2 + G_y^2} \quad (8)$$

3.2.2. Scharr Layer Solving

The Scharr operator has enhanced its difference extraction capability based on the Sobel operator, and the solution principle is similar to that of Sobel. In general, the size of the edge detection filter of the Scharr operator is set to (3×3) , and the detection is also divided into two directions, which are calculated as Equations (9) and (10):

$$G_x = \begin{bmatrix} -3 & 0 & 3 \\ -10 & 0 & 10 \\ -3 & 0 & 3 \end{bmatrix} \times I \quad (9)$$

$$G_y = \begin{bmatrix} -3 & -10 & -3 \\ 0 & 0 & 0 \\ 3 & 10 & 3 \end{bmatrix} \times I \quad (10)$$

where I represents the input underwater acoustic imagery, and G_x and G_y indicate the image gradient values obtained by solving.

3.2.3. Laplacian Layer Solving

The Laplacian operator is a kind of a spatial sharpening filter, which is a second-order differential operator in n -dimensional Euclidean space with rotational invariance. This operator usually considers the point whose second-order differential point of the pixel point value is 0 as an edge point. For the underwater acoustic imagery $I(x, y)$, the Laplace solving process of the image is

$$\nabla^2 I(x, y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad (11)$$

To find the second-order derivatives of the image $I(x, y)$ in the x and y directions, the second-order Laplacian operator expressing can be obtained as follows:

$$\nabla^2 I(x, y) = I(x, y + 1) + I(x, y - 1) + I(x + 1, y) + I(x - 1, y) - 4I(x, y) \quad (12)$$

Finally, the filter mask of the Laplacian operator can be solved using Equations (11) and (12).

$$G_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (13)$$

It can be seen from Figure 3 that the PC is well-adapted to the nonlinear features of underwater acoustic imagery. Specifically, it can well contain the structural feature information in the raw acoustic imagery, and the preservation of edges is more desirable, while it can also effectively resist noise interference. On the contrary, the operators' performances based on gradient information are poor, and it is not easy to adapt to the nonlinear gradient features on acoustic imagery; it is also more sensitive to the interference of noise. In the next section, this research will quantitatively and qualitatively evaluate the performances of the feature detectors on various layers from experiments.

4. Experiment

4.1. Experimental Method Setting

In terms of experimental method settings, several mainstream feature detection operators, including single-scale detection operators, multiscale detection operators, and deep-learning-based detectors, are selected in this manuscript, as depicted in Table 1. The purpose of the experiments is to fully explore the best performance of the detection operators of various modes on the underwater acoustic imagery layers and to verify whether the deep-learning-based feature detector will bring some opportunities for underwater acoustic imagery feature detection task.

The above-mentioned feature detectors have their own advantages; however, underwater scenes are often complex and variable, resulting in detected sonar images that may have noise interference, grayscale distortion, translation transformations, rotation transformations, scale transformations, mixed rotation and scale transformations, viewpoint transformations, and slight distortions, which make it tough for a single detector to cope with. What is worse, these transformations are often nonlinear, irregular, and difficult to describe in mathematical language. Therefore, it is necessary to further explore a scheme for underwater acoustic detection scenarios that should concentrate on the advantages of existing detection operators and be able to effectively deal with the limitations of images themselves while meeting the real-time needs of underwater applications.

The detectors and layers set up for the experiments in this research are separately shown in Tables 1 and 2. The detectors for sequences 1 to 9 are implemented using the benchmark OpenCV library [22], and a KeyNet [23] detector for sequence 10 is implemented using the benchmark Kornia library [24], both with default parameters to facilitate comparison and update for future researchers. In terms of the detection scene setting, three types of feature detection layers are selected, which are the grayscale, gradient, and PC layers, where the gradient layers are computed by the Sobel, Scharr, and Laplacian operators.

Table 1. Introduction to detectors used in the experiment.

Sequences	Name	Time	Creator	Characteristic
1	AKAZE [25]	2011	Pablo F Alcantarilla	Multiscale detector
2	BRISK ¹ [26]	2011	S Leutenegger	Multiscale detector
3	FAST [27]	2010	Edward Rosten	Single-scale detector
4	Harris [28]	1988	CG Harris	Single-scale detector
5	ORB ² [29]	2011	GARY R.Bradski	Multiscale detector
6	Shi-Tomasi [30]	1994	J.SHI & C. Tomasi	Single-scale detector
7	SIFT [4]	2004	Lowe	Multiscale detector
8	SURF [5]	2008	Bay	Multiscale detector
9	SAR-SIFT [31]	2014	Flora Dellinger	Multiscale detector
10	KeyNet [23]	2019	Axel Barroso	Learned detector

¹ Detector used by BRISK is FAST-9-16. ² Detector used by ORB is Oriented FAST.

Table 2. Detection layers used in the experiment.

Layer Name	Gradient Layers	PC Layer	Grayscale Layer
Operator	Sobel Scharr Laplacian	PC	Pixel

4.2. Dataset

To evaluate the detection accuracy and processing time of different detectors on various layers, a random dataset of underwater acoustic imageries was selected in this paper. These images were obtained from various types of SSS and FLS sensors, and they are depicted in Figure 4. For the internal target richness, the dataset can be classified into simple and complex scenes; in terms of external factors, it can be classified into general, scale-transformed, noise-interfered, grayscale-transformed, and other phenomena. The division of this dataset is specified in Table 3, and the test samples are described as follows:

1. This image was provided by the T-SEA Marine Technology company [32] and originated from an SSS device with an image size of 112×186 . The content of this image is the wreck site without any pre-processing of the image.
2. Image 2 is a scale transformation of image 1 with a scale-transformation factor of 2. The transformed image size is 224×372 , and the purpose of the transformation is to check the scale-invariant performances of the detectors.
3. Image 3 was provided by Peng Cheng Laboratory (PC LAB) [33] from an SSS device. The image size is 429×424 , and the image content is a “ridge-like zone” on the seafloor. This image is used to test the detection ability of the detector for protruding targets in a monotonous background, without any pre-processing of the image.
4. Image 4 was provided by PC LAB from an SSS device. The image size is 337×425 , and the image content is a seafloor protrusion with severe acoustic shadowing. This image is used to check the detection capability of the detector for targets with severe acoustic shadowing in a sandy background, without any processing of the image.

5. Image 5 was provided by PC LAB from an SSS device. The image size is 210×305 , and the image content is a depressed area on the seafloor. This image is used to check the detection capability of the detector for depressed targets in a monotonous background, without any pre-processing of the image.
6. Image 6 was provided by PC LAB from an SSS device. The image size is 245×264 , and the image content is a tubular object on the seafloor. This image is used to check the detection capability of the detector for tubular targets in a monotonous background, without any pre-processing of the image.
7. Image 7 was provided by PC LAB from an SSS device. The image size is 286×258 , and the image content is a submarine tubular target with significant grayscale distortion. This image is used to check the detector's detection capability for tubular targets under the influence of grayscale distortion, without any pre-processing of the image.
8. Image 8 was provided by PC LAB from an SSS device. The image size is 495×374 , and the image content is a rocky area of the seafloor with strong noise interference. The image was denoised using a median filter in order to check the detector's feature extraction capability under strong noise interference.
9. Image 9 is the original image of image 8 without any pre-processing of the image, the purpose of which is to compare with the test results of image 8.
10. Image 10 was provided by PC LAB from an SSS device. The image size is 446×371 , and the image content is a seabed area with a lot of noise interference. The image is intended to complement the test scenes of test images 8 and 9. The targets in image 10 are mostly raised, and no pre-processing of the image has been performed.
11. Image 11, from [34], is an FLS image with an image size of 116×127 . The image contains a tire laid underwater, and the image has significant noise interference. This image is used to check the detection ability of the detector under substantial noise interference, and the image is denoised using a median filter.
12. Image 12 is the original image of image 11, without any pre-processing of the image, which is intended to be compared with the test results of image 11.
13. Image 13, from [34], is an FLS image with an image size of 44×50 . The image content is a steel frame laid underwater, without any pre-processing of the image.
14. Image 14 is obtained based on image 13 by scale transformation with a scale factor of 2. The transformed image size is 88×100 . The purpose of the transformation is to check the scale-invariant performances of the detectors.
15. Image 15 is from the Sound Metrics database [35] and is a DIDSON (dual-frequency identification sonar) image with an image size of 351×359 . The image content is a floor tile underwater, and the image is used to check the feature extraction capability of the detectors for DIDSON images without any pre-processing of the image.
16. Image 16 was provided by DeepVision AB company [36] from an SSS image device with an image size of 800×400 . The image content is a low-texture, low-contrast depressed region of the seafloor. The image was designed to check the detectors' feature extraction capability for low contrast and texture areas without any pre-processing of the image.

Table 3. Characteristics division of dataset scenarios.

Characteristics	Simple Scenes	Complex Scenes
General	2, 11	9, 14
Scale-transformed	1	13
Noise-interfered	5, 12	8, 10
Grayscale-transformed	6, 7	3, 4
Others	15	16

Next, some representative underwater acoustic imageries are taken to demonstrate the display of each layer to help the reader visualize the information represented by various layers. The demonstration results are described in Figures 5 and 6.

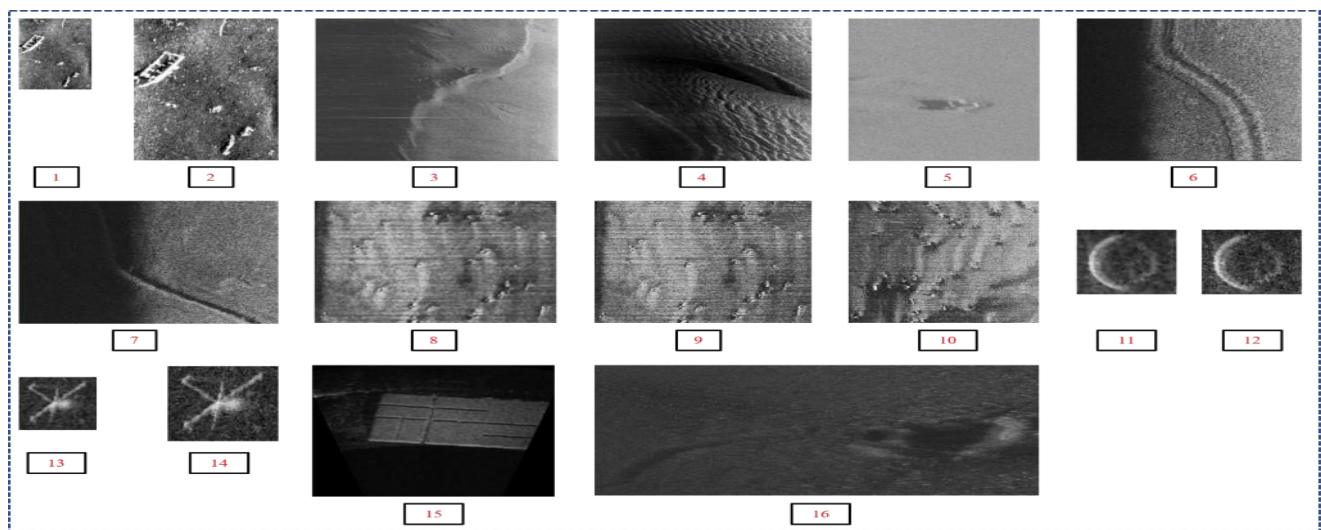


Figure 4. Schematic diagram of experimental dataset, the serial numbers 1–16 in the figure caption correspond to the 16 samples above. For the analysis of their characteristics, please refer to the above.

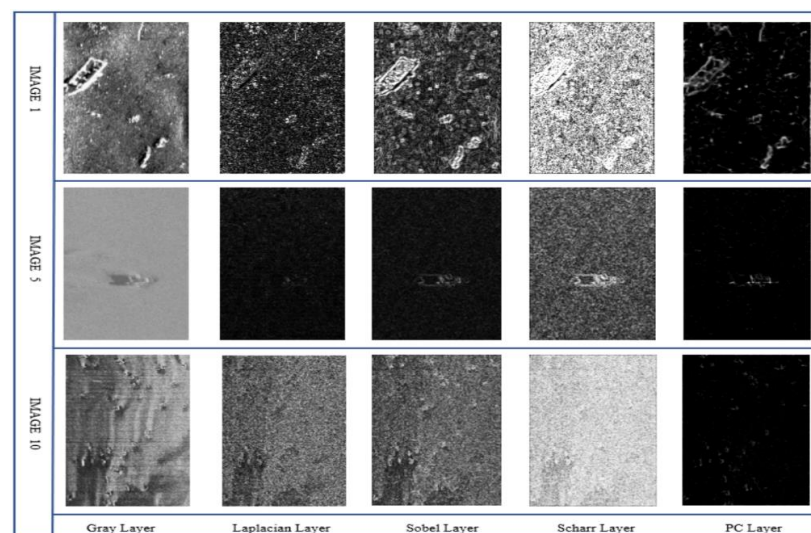


Figure 5. Display of various layers of some side-scan sonar images.

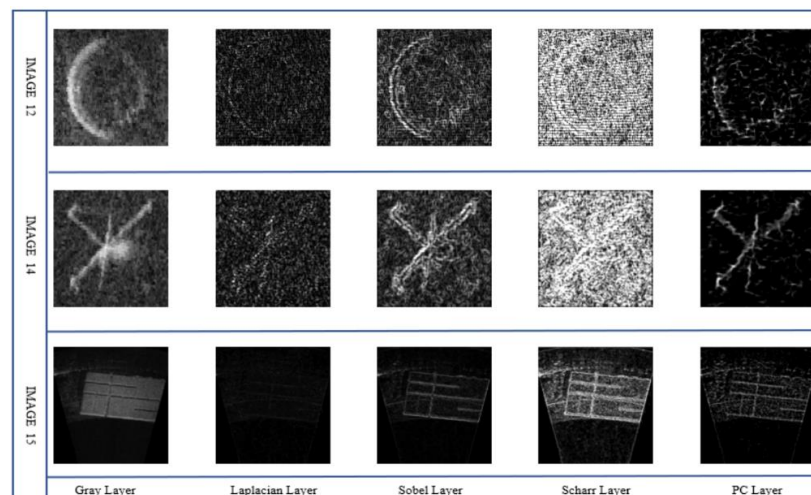


Figure 6. Display of various layers of some forward-looking sonar images.

Finally, the feature detectors in Table 1 were used for three gradient layers, one PC layer, and one grayscale layer. Quantitative and qualitative evaluation criteria were used to find the feature detector with the best detection performance, the layer best suited for feature detection, and the best combination of the two; these criteria will be given in the next section.

4.3. Evaluation Criterion

In terms of the evaluation criterion, this paper introduced the Number of keypoints (N), Precision (P), Distribution (D), and Time consumption (T). It is worth mentioning that these evaluation criteria are developed for the region of interest (ROI), not the whole image, which is mapped by experts in the field of underwater acoustics. The detection performance of the detector is reflected by comparing the number of features correctly identified within the ROI, the precision, and the distribution of feature points; the real-time performance of the detector is reflected by comparing the time consumed for feature point detection. In traditional optical image-matching studies, the metric used to evaluate the performance of detector is the feature repeatability, which requires prior knowledge of the image pair corresponding to the transformation relationship. However, due to the scarcity of underwater acoustic datasets, the corresponding image pairs cannot be effectively obtained, so the test dataset used in this manuscript is single, random, and irregular, including both SSS and FLS image data. Moreover, the image size is not adjusted and is kept in its raw state to maximize the fairness of the test.

Firstly, experts in the field of sonar image interpretation are requested to manually map the ROIs, as shown in Figure 7. These ROIs all have a clear detection value, and the closer the feature detection effect is to these regions, the better the detector's performance. The keypoints that fall into these regions are considered valid feature points, and the detector's performance is accordingly evaluated. In addition, this manuscript fine-tuned the thresholds of some detectors during the experiment to ensure that the number of their detection points will not explode, making the final evaluation results in a valid range.

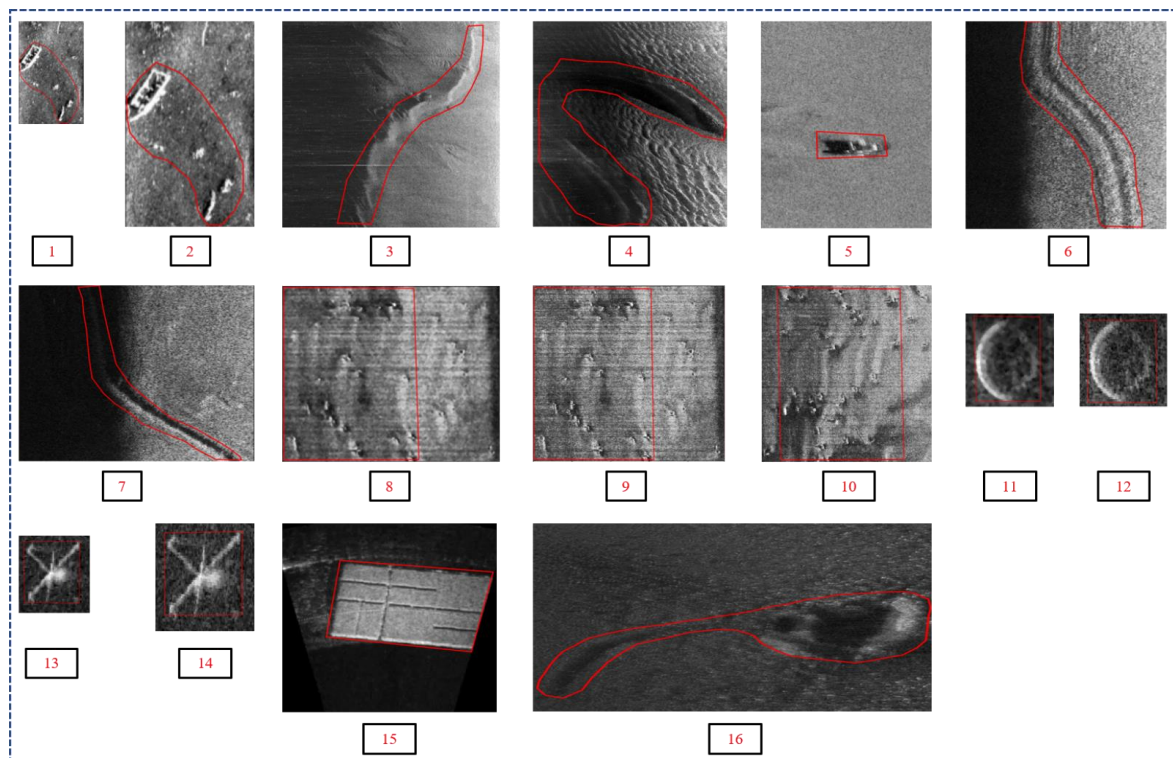


Figure 7. Schematic diagram of ROIs (regions delineated by red line) on tested images drawn by experts. The serial numbers 1–16 in the figure correspond to the 16 samples in Section 4.2.

4.3.1. Number of keypoints (N)

N is computed to evaluate the richness of the detection effect. It directly affects the starting quantity of subsequent matches and the accuracy of the solution of the transformation matrix.

$$N = \sum_{i=1} [(x_i, y_i) \in ROI_s] \quad (14)$$

where (x_i, y_i) represents the coordinate point that falls within the ROIs.

4.3.2. Precision (P)

P is used as the criterion to evaluate the accuracy of detection effect, which is calculated as Equation (15):

$$P = \frac{N}{N_{all}} \quad (15)$$

where N_{all} indicates the number of detected keypoints on the whole image.

4.3.3. Distribution (D)

The distribution criterion is used to measure the uniformity of the detection results, since an ideal feature detector should detect feature points that are uniformly distributed. For the distribution measure, the traditional histogram idea is used to divide the whole image into a set of uniform regions. In the experimental part, the width and height of each region are 0.1 of the entire imagery, and there are 100 regions in total. Since this work only considered the homogeneity of feature points inside the ROI, where the area is covered in red in the figure, only the feature points inside the ROI are counted, and only the regions (*bins*) overlapping with the ROI with an area larger than 0 are considered, while the rest of the regions are ignored, as shown in Figure 8. After counting the number of keypoints n_i in each small area, the chi-square test is used to check the distribution of keypoints. Specifically, this part of the research proposed the hypothesis that the distribution of feature points in the H_0 : ROI follows a two-dimensional uniform distribution. Therefore, the probability p_i that a feature point falls in each cell region is equal to the ratio of the area S_i of the region to the total area S_{ROI} of the ROI. Then the expectation of the number of feature points in each small area i is calculated as Equation (16):

$$E_i = p_i \times N = \frac{S_i}{S_{ROI}} \times N \quad (16)$$

where N is the total number of keypoints in the ROI. From the above, the chi-square statistic can be calculated based on the actual distribution of keypoints and the expected distribution, which is computed as follows:

$$X^2 = \sum_i \frac{(n_i - E_i)^2}{E_i} \quad (17)$$

The probability is then calculated based on the value of the chi-square distribution function cdf at X^2 for a sample with degrees of freedom $N_{bin} - 1$. Then the final calculation formula is derived, where the significance level parameter was set to 0.05.

$$D = 1 - cdf(X^2, N_{bin} - 1) \quad (18)$$

4.3.4. Time consumption (T)

T denotes the time consumed to detect per keypoint and is computed as

$$T = \frac{T_{all}}{N_{all}} \quad (19)$$

where T_{all} denotes the total time taken to detect all keypoints (N_{all}) on the whole image.

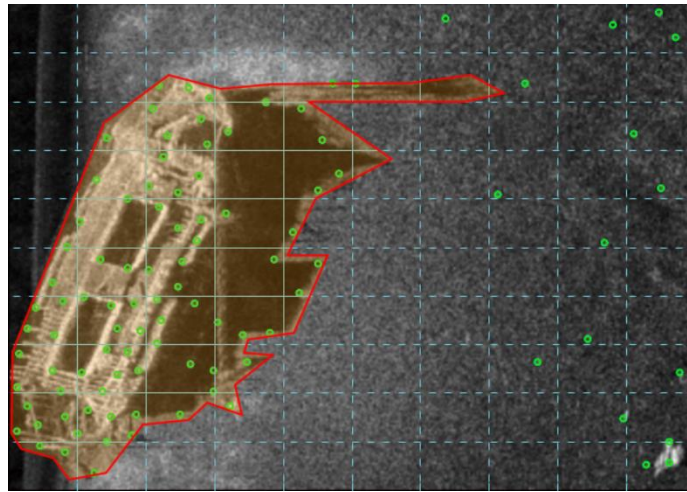


Figure 8. Schematic diagram of Distribution (D) calculation. The area enclosed by the red line is the ROI to be calculated, and the detected feature points are represented by green circles. Image is offered by DeepVision AB company [36].

4.4. Testing Environment

All tests were conducted under a 64-bit Windows 10 operating system with an Intel Core i7-10750 2.60 GHz processor, 16 GB of physical memory, and an NVIDIA GeForce RTX2060 graphics card. All algorithms were compiled using Python 3.7. To ensure the reliability of the test results and avoid the influence of coarse differences, the evaluation results were averaged over ten test results.

5. Experiment Results

The evaluation criterion can evaluate the detectors in four aspects: richness, accuracy, distribution uniformity, and timeliness, and the evaluation scope is more comprehensive. In order to more intuitively find the most suitable feature detector for underwater acoustic images, this paper proposed a normalized evaluation strategy that normalizes the first three evaluation criteria, the number of keypoints (N), Precision (P), and Distribution (D), into each index of the evaluation algorithm, and then multiplies them by their respective weights to calculate the final evaluation score. Each index is assigned a score (0–10) according to ranking order (0–10), and then the final score S is calculated as follows:

$$S = \sum w_i s_i = w_N s_N + w_P s_P + w_D s_D \quad (20)$$

where s_i is the score of each individual metric of the feature detector, and w_i is the weight value corresponding to the individual metric. s_N is the metric score obtained from the evaluation of the number of keypoints, w_N is the weight value corresponding to s_N , s_P is the metric score obtained from the evaluation of precision, w_P is the weight value corresponding to s_P , s_D is the metric score obtained from the evaluation of the distribution, and w_D is the weight value corresponding to s_D . S is the final score of the performance evaluation of each detector. Considering the practical applications in ocean engineering, the experts suggest that the w_N , w_P , and w_D distributions take the values 0.35, 0.60, and 0.05, respectively. The results on each test image are presented in the form of a table, where, for the evaluation of gradient layers, it selected the one with the greatest Precision (P) among the three layers for presentation. The specific experimental results are shown in Tables 4–7 and A1–A13. In order to accommodate the layout of the manuscript, three typical table results were selected for analysis in this section; please refer to Appendix A for the completed test results. The highest scoring items are marked with *.

Table 4. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 6.

Detectors \ Layers	Grayscale Layer			Gradient Layer (Sobel) *			PC Layer *		
	N	P	D	N	P	D	N	P	D
AKAZE	41	0.8913	0.0378	100	0.5882	0.0217	46	0.5349	0.2422
BRISK	559	0.4815	0.0000	1088	0.4131	0.0000	441	0.4608	0.0000
FAST	332	0.4618	0.0000	836	0.3696	0.0038	331	0.4206	0.0009
Harris	9	0.4286	0.3444	11	0.2973	0.0539	66	0.3158	0.5465
ORB *	332	0.8448	0.0000	322	0.7368	0.0000	315	0.7360	0.0000
Shi-Tomasi	43	0.4300	0.9462	43	0.4300	0.9879	50	0.5000	0.8790
SIFT	151	0.4535	0.3539	184	0.4044	0.3060	185	0.4780	0.0000
SURF	255	0.4374	0.0135	299	0.3705	0.0390	236	0.4574	0.0140
SAR-SIFT	12	0.7059	0.0000	30	0.3409	0.0000	295	0.3203	0.0000
KeyNet	159	0.3850	0.7262	170	0.3020	0.5512	160	0.3922	0.3252

Table 5. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 8.

Detectors \ Layers	Grayscale Layer			Gradient Layer (Laplacian)			PC Layer *		
	N	P	D	N	P	D	N	P	D
AKAZE	291	0.6063	0.0000	16	0.8000	0.0241	334	0.6095	0.0000
BRISK	614	0.6221	0.0000	1254	0.6006	0.0000	2234	0.5935	0.0000
FAST	194	0.6978	0.0000	924	0.6067	0.0000	1410	0.5975	0.0001
Harris	123	0.6089	0.0000	105	0.6069	0.0000	381	0.5443	0.0240
ORB *	344	0.6880	0.0000	288	0.6973	0.0000	324	0.6480	0.0000
Shi-Tomasi	70	0.7000	0.0091	69	0.6900	0.0008	67	0.6700	0.6352
SIFT	950	0.5675	0.0000	77	0.5461	0.0000	1363	0.5623	0.0000
SURF	1125	0.6134	0.0109	522	0.5781	0.0000	1487	0.6165	0.0000
SAR-SIFT *	146	0.8795	0.0000	2278	0.6142	0.0000	2322	0.6180	0.0000
KeyNet	1059	0.5970	0.9472	1021	0.6241	0.9671	1119	0.6159	0.0957

Table 6. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 10.

Detectors \ Layers	Grayscale Layer			Gradient Layer (Scharr)			PC Layer *		
	N	P	D	N	P	D	N	P	D
AKAZE	436	0.9316	0.0000	707	0.8559	0.0000	45	0.9783	0.0059
BRISK	7804	0.7963	0.0000	9992	0.7930	0.0000	614	0.8899	0.0000
FAST	5541	0.7354	0.0000	7810	0.7223	0.9939	390	0.8387	0.0000
Harris	719	0.6548	0.0423	4	0.6667	0.7437	56	0.8000	0.0405
ORB *	491	0.9820	0.0000	470	0.9514	0.0000	432	0.9841	0.0000
Shi-Tomasi	84	0.8400	0.0927	73	0.7300	0.2023	84	0.8400	0.0234
SIFT	994	0.7821	0.0000	1231	0.7245	0.2000	46	0.8519	0.0000
SURF	2400	0.7846	0.0000	3292	0.7735	0.0000	123	0.9111	0.0000
SAR-SIFT	203	0.9144	0.0000	3	1.0000	0.0000	3015	0.7320	0.0000
KeyNet	1197	0.7389	0.9982	1365	0.7708	0.9445	668	0.7687	0.0024

Table 7. Statistical table of results of experimental tests. (n) indicates that the number of times with optimal performance is n. Total number of tests is 16.

Test items	Detector	Layer
Best	ORB ¹ (7)	PC (7)
Next best	BRISK ² (6)	Sobel (5)

¹ Specific detection operator is Oriented FAST. ² Specific detection operator is FAST-9-16.

Tables 4–6 are utilized as typical schematic tables to illustrate the statistical contents. The horizontal axis of each table is used to illustrate the performance comparison of detection layers, and the vertical axis of each table is used to illustrate the performance comparison of detectors. The performance evaluation metrics include the Number of

keypoints (N), Precision (P), and Distribution (D). Finally, all the metrics are weighted to determine the optimal detector and detection layer.

From the statistical results in Table 7, it can be seen that the most suitable detectors for the underwater acoustic imagery are ORB and BRISK, which appear seven and six times, respectively. It is worth mentioning that both detectors use the detection algorithm of the FAST series.

In terms of detection layers, the PC layer has the best performance, appearing seven times, followed by the classic Sobel layer appearing five times. This result shows the PC layer's great potential in underwater acoustic imagery feature detection.

Next, the paper evaluated the detection speed of various detectors and detection layers, and the results show that the FAST detector and PC layer have an apparent advantage. In comparison, grayscale layers are the slowest to detect. The specific detection time difference is demonstrated in Figure 9.

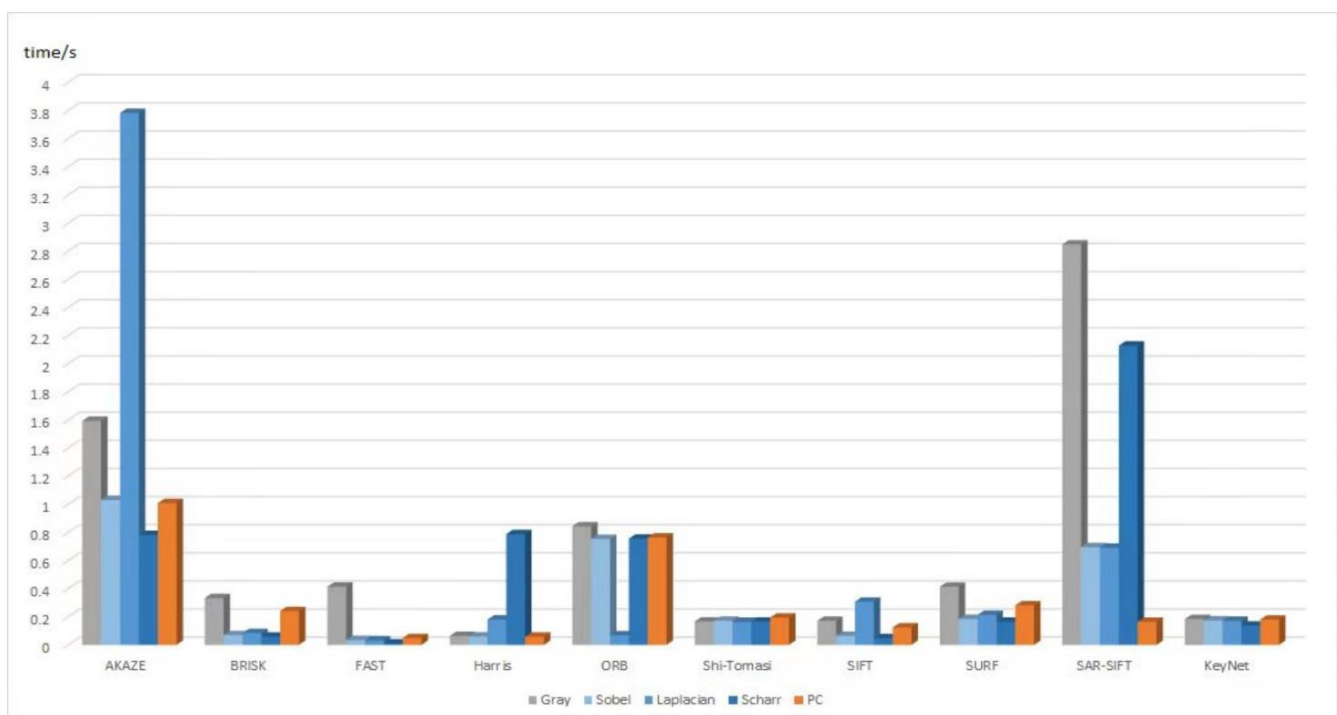


Figure 9. Comparison of average time consumed to detect single feature point.

As shown in Figure 10, images with good visual perception of the detection results are also labeled as optimal items in quantitative statistical Tables 4–6, which are the quantitative evaluation results for images 6, 8, and 10, respectively, fully validating the validity and rationality of our proposed evaluation framework.

In the following, this manuscript will use the detection scheme that is the ORB detector with the PC layer, which achieves the best performance for the feature detection of the acoustic imagery dataset. The detailed detection effects are shown in Figure 10. In order to observe the details of feature detection, the feature points detected by the PC layer are mapped to the original image.

As shown in Figure 10, it can be seen that the ORB detector has an overall good detection effect on the PC layer, which can resist the noise interference on the underwater acoustic imageries, and the detected feature points can be accurately around the target area and more uniformly distributed. This detection method allows for effective feature detection on underwater acoustic imageries while meeting the need for real time.

In addition, the detection effect of images 1 and 2 shows that the scheme can resist scale transformation; the detection effect of image pairs 8 and 9, and image pairs 11 and 12 indicates that the scheme based on the PC layer does not need to denoise the acoustic

images in advance and can achieve good results by directly detecting the raw images. It is worth mentioning that this almost noise-independent feature detection scheme can bring significant help to the underwater sonar target tracking tasks [37–39], which, in turn, can effectively enhance the autonomous capability of AUVs.

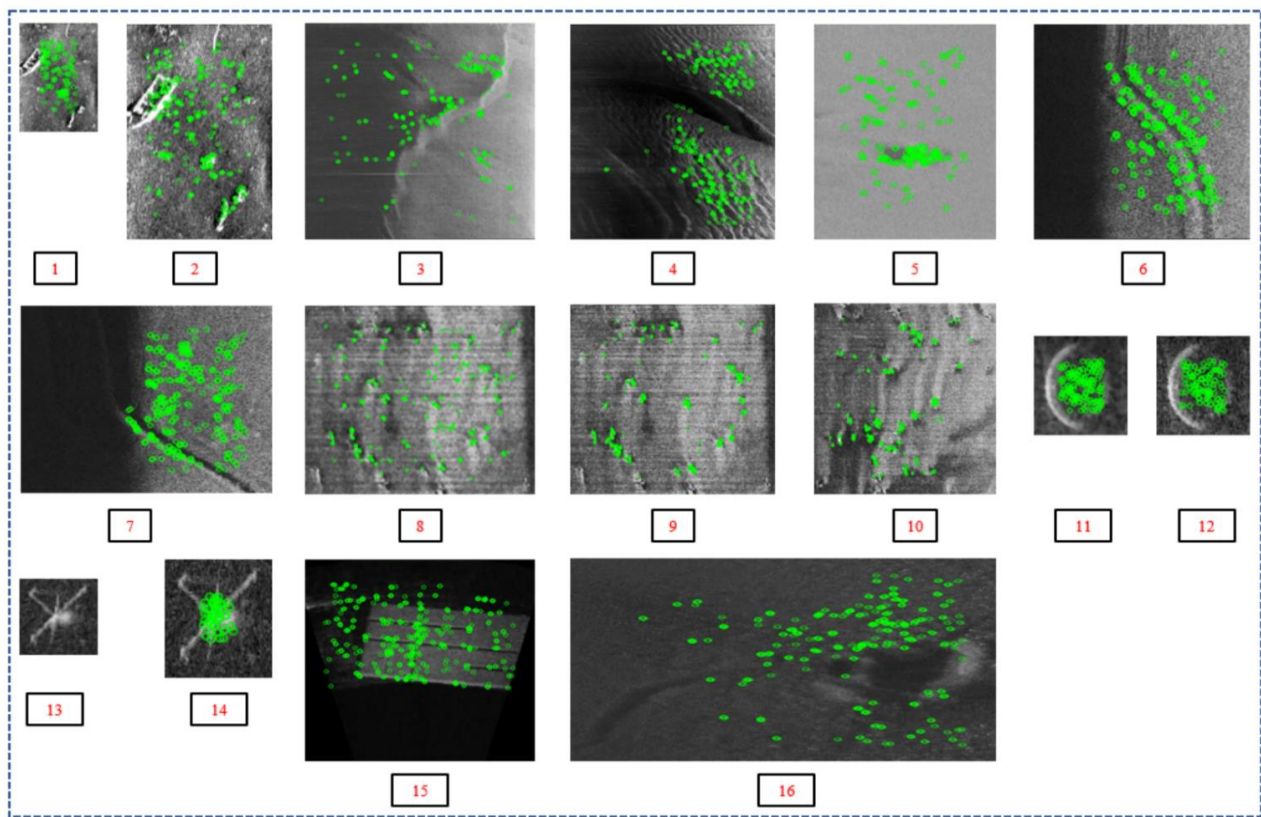


Figure 10. Effects of feature detection on PC layer using ORB detector; feature points are highlighted with green. The serial numbers 1–16 in the figure correspond to the 16 samples in Section 4.2.

However, this detection scheme is currently unable to overcome the challenges posed by acoustic shadowing, as shown in test images 4, 7, and 16. Moreover, the detection results on small-scale forward-looking sonar images are also unsatisfactory, such as image 13. Both of these are the directions for continued research in the future.

In order to compare the detection effects of the detector on various layers, Figure 11 displays the detection effects of the ORB detector and grayscale layer, Sobel layer, Scharr layer, Laplacian layer, and PC layer combinations. In order to visually compare the detailed effects of feature detection, the detection performances on each layer are mapped to the original image. As can be seen in Figure 11, the PC layer has stronger noise resistance than the gradient layer, whereas richer feature points can be detected in the PC layer than in the grayscale layer. This detection scheme can achieve a good balance between the precision and velocity, making it more suitable for underwater sensing.

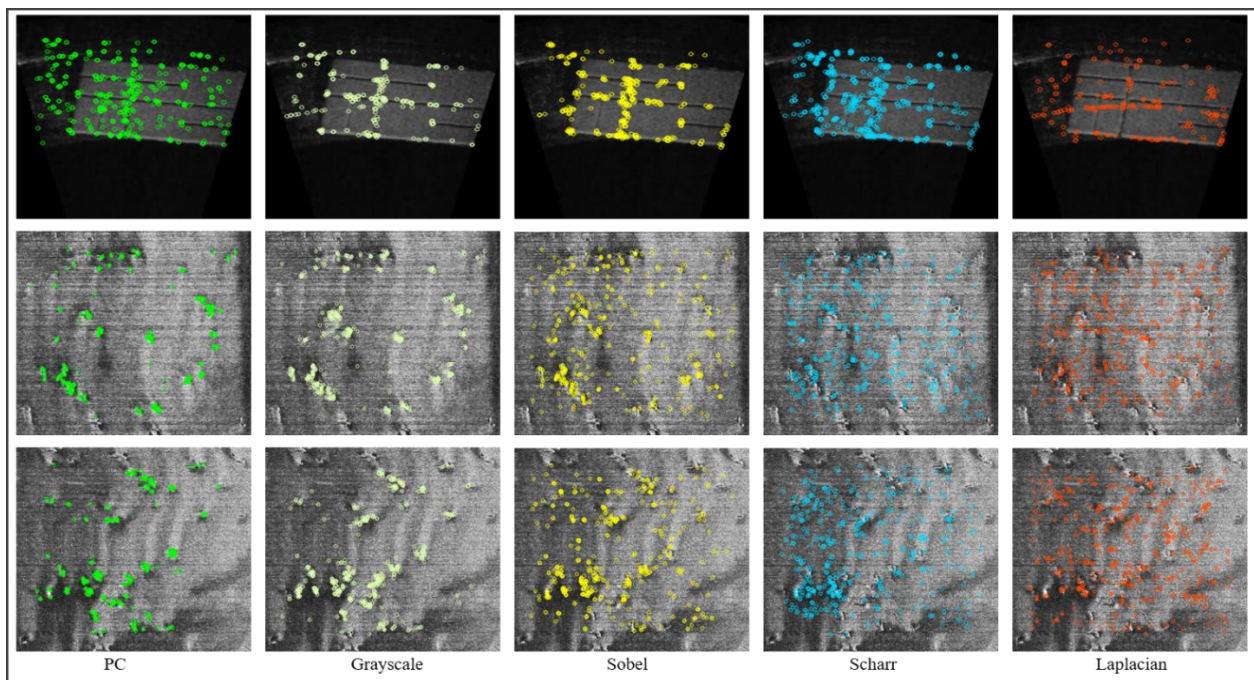


Figure 11. Comparison of detection effects of ORB detector on PC layer, grayscale layer, Sobel layer, Scharr layer, and Laplacian layer.

6. Discussion

In this manuscript, we designed feature detection experiments from several scientific perspectives, including the sonar-type differences, scale transformation, noise interference, detector design methods, and the properties of detection layers. In addition to the intuitive optimal results above, there are many valuable research insights that can be found from the experimental results: (1) KeyNet, a deep-learning-based feature detector, is not the optimal choice, which may be related to its lack of pretraining on acoustic datasets. In the future, if more underwater acoustic datasets can be collected and fine-tuned on the original model, the final detection results may be better. (2) Both the ORB and BRISK detectors perform well, and the underlying detection operators they used are based on the FAST derivation, which is very clever. In addition, the FAST detector achieved the best results in velocity evaluation, so developing better FAST-like detectors may present opportunities for underwater acoustic imagery feature detection. (3) Layers are the basis of feature detection. In the experiments of this paper, the best results were achieved by PC-layer-based detection, which achieves a balance between accuracy and speed. The PC layer has excellent resistance to noise interference and brightness variations, which is well-suited for applications in underwater acoustic matching tasks. In the future, if a feature detector based on PC layers can be developed, it will become more convenient to engage in underwater perception tasks. (4) The test results show that the combination of the ORB detector and PC layer is not perfect. They are not sensitive to the line features in the image, and how to enhance their line feature extraction capability will be essential work in the future because it will affect underwater applications such as underwater pipeline maintenance and underwater cable laying. (5) From the comparison of all the experimental results, the detection results of various detectors on side-scan sonar images are better than those on forward-looking sonar, which may be due to the higher resolution and better quality of side-scan sonar images, which are closer to the optical images in terms of image display effects. In the future, it is worth considering how to use deep-learning methods to achieve domain transfer between acoustic and optical images and reduce the imaging differences between the two, which will also provide an idea for underwater acoustic imagery detection and matching research.

7. Conclusions

This manuscript systematically investigated how to more effectively detect feature points on underwater acoustic imagery by enumerating the most commonly used detectors for experiments and comparing their effects on the grayscale, gradient, and PC layers. Tests were conducted on a dataset containing side-scan sonar and forward-looking sonar images, and a new evaluation framework was proposed. The experimental results show that the feature detection using the ORB detector and PC layer is the best, and this combined detection mode can ensure the detection accuracy while also reducing the interference of image noise, and has obvious advantages in terms of speed. It is worth mentioning that this manuscript is the first comprehensive proposal to use the PC layer for underwater acoustic imagery feature detection. Throughout the testing process, no assumptions were made on the sonar type and imagery attributes, and the test results proved that the evaluation framework proposed in this manuscript has good generalization performance.

The research work in this manuscript selected typical acoustic images in practical ocean applications, analyzed the characteristics of each image and classified them, and, finally, displayed the experimental report of each of them in detail to provide reference for practitioners. The conclusions can also provide new ideas for novices in the field of acoustic image processing and researchers who are proficient in image processing but do not understand acoustic images. However, there can be more detector attempts beyond the test methods set in this research, as well as the use of methods such as the hyperparameters to explore the optimal parameter settings for the detector instead of the default parameters. For these cases, the evaluation framework in this manuscript preserves the pipeline and deep-learning interfaces to facilitate testing.

In the future, the research will introduce deep-learning-based assessment metrics while expanding the underwater acoustic dataset to form a more comprehensive and objective assessment framework and open-source it in the community for researchers to use; moreover, it will consider the performance of feature detectors and feature descriptors when combined to conclude a complete solution for underwater acoustic imagery matching.

Author Contributions: X.Z. provided the idea. X.Z. and S.Y. designed the experiments. H.L. and C.Y. analyzed the experiments. X.Z. wrote the paper. X.Y. edited and proofread the article. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Chinese Shandong Provincial Key Research and Development Plan (Grant No. 2019GHZ011; 2021CXGC010702) and Wenhai Program of the S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (NO. 2021WHZZB2000).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to thank the editors and anonymous reviewers for their valuable suggestions.

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

Abbreviations

The following abbreviations are used in this manuscript:

AUV	Autonomous Underwater Vehicle
BRISK	Binary Robust Invariant Scalable Keypoints
DIDSON	Dual-Frequency Identification Sonar
FLS	Forward-Looking Sonar
FAST	Features from Accelerated Segment Test
ORB	Oriented FAST and Rotated BRIEF

PC	Phase Congruency
ROI	Region of Interest
SIFT	Scale-Invariant Feature Transform
SURF	Speeded-Up Robust Features
SSS	Side-Scan Sonar

Appendix A

The appendix is a supplement to the experimental results in Section 5 of this manuscript. The table format and test content are consistent with the schematic table in Section 5, and the test images are derived from the dataset presented in Section 4 of this paper.

Table A1. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 1.

Detectors	Layers	Grayscale Layer			Gradient Layer (Scharr) *			PC Layer		
		N	P	D	N	P	D	N	P	D
AKAZE		15	1.0000	0.3298	74	0.7708	0.1713	16	0.8889	0.2303
BRISK		180	0.5556	0.0005	674	0.5570	0.0000	96	0.6713	0.0000
FAST		188	0.4574	0.0130	650	0.4159	0.9488	116	0.5604	0.0000
Harris		30	0.3093	0.4894	1	1.0000	0.9846	20	0.3704	0.8499
ORB *		167	0.7076	0.0000	232	0.8315	0.0000	110	0.7746	0.0000
Shi-Tomasi		36	0.3600	0.9899	36	0.3600	0.9994	36	0.3956	0.9919
SIFT		90	0.4615	0.0953	111	0.5139	0.2532	63	0.7326	0.0000
SURF		108	0.5482	0.9383	166	0.5061	0.0176	62	0.6667	0.4383
SAR-SIFT		51	0.9107	0.0000	2	0.2857	0.9624	201	0.4752	0.0000
KeyNet		72	0.4865	0.3438	91	0.4866	0.9121	76	0.5135	0.2934

Table A2. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 2.

Detectors	Layers	Grayscale Layer			Gradient Layer (Scharr)			PC Layer *		
		N	P	D	N	P	D	N	P	D
AKAZE		183	0.6489	0.0000	447	0.5992	0.0863	107	0.5912	0.0000
BRISK *		1096	0.4740	0.0000	3150	0.4827	0.0000	645	0.5051	0.0000
FAST		625	0.3714	0.0000	2329	0.3971	0.3134	403	0.4022	0.0000
Harris		130	0.2882	0.7285	1	1.0000	0.9153	111	0.3801	0.1492
ORB		364	0.8125	0.0000	357	0.7661	0.0000	309	0.7087	0.0000
Shi-Tomasi		43	0.4300	0.3483	39	0.3900	0.9779	42	0.4200	0.8083
SIFT		545	0.3929	0.2014	422	0.4302	0.0000	332	0.4094	0.0000
SURF		429	0.4400	0.5324	810	0.4574	0.0870	390	0.4407	0.3716
SAR-SIFT		120	0.8163	0.0000	30	0.8333	0.0000	802	0.4284	0.0000
KeyNet		315	0.4280	0.6049	380	0.4279	0.9222	321	0.4315	0.2497

Table A3. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 3.

Detectors	Layers	Grayscale Layer			Gradient Layer (Sobel) *			PC Layer		
		N	P	D	N	P	D	N	P	D
AKAZE		24	0.9231	0.0007	48	0.4364	0.0000	30	0.3226	0.0000
BRISK *		68	0.2500	0.0002	546	0.2423	0.0000	337	0.2334	0.0000
FAST		25	0.1773	0.3382	334	0.2001	0.1089	235	0.2083	0.0001
Harris		103	0.1791	0.0276	109	0.1542	0.9619	43	0.1937	0.0046
ORB		164	0.4505	0.0001	188	0.3917	0.0000	150	0.3119	0.0000
Shi-Tomasi		19	0.1900	0.9040	18	0.1800	0.8117	16	0.1600	0.8694
SIFT		66	0.4853	0.0001	214	0.1994	0.0384	65	0.2471	0.0000
SURF		243	0.4812	0.0485	418	0.2401	0.0918	233	0.3324	0.0000
SAR-SIFT		13	0.7222	0.0077	273	0.1716	0.0000	597	0.1743	0.0000
KeyNet		246	0.2064	0.9848	300	0.1795	0.6008	255	0.1941	0.6637

Table A4. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 4.

Layers Detectors	Grayscale Layer			Gradient Layer (Scharr)			PC Layer *		
	N	P	D	N	P	D	N	P	D
AKAZE	6	0.0274	0.9868	473	0.3546	0.0000	36	0.0859	0.0000
BRISK	172	0.0899	0.0000	3929	0.3714	0.0000	295	0.1427	0.0000
FAST	123	0.0838	0.0000	3034	0.3308	0.0000	225	0.1314	0.0000
Harris	52	0.2158	0.0000	111	0.5722	0.0000	52	0.1297	0.0002
ORB	10	0.0207	0.0000	141	0.2820	0.0000	23	0.0472	0.0000
Shi-Tomasi	4	0.0400	0.9423	26	0.2600	0.8645	11	0.1100	0.7849
SIFT	67	0.0847	0.0000	672	0.3630	0.0024	119	0.0869	0.0000
SURF	183	0.1801	0.0000	1113	0.3862	0.3628	151	0.1465	0.0000
SAR-SIFT *	132	0.6839	0.0000	327	0.8015	0.0000	720	0.3124	0.0000
KeyNet	303	0.2803	0.0003	572	0.3801	0.6949	256	0.2681	0.0000

Table A5. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 5.

Layers Detectors	Grayscale Layer			Gradient layer (Sobel) *			PC Layer		
	N	P	D	N	P	D	N	P	D
AKAZE	8	1.0000	0.6830	14	1.0000	0.5247	14	1.0000	0.0001
BRISK	21	0.9545	0.0007	74	0.4744	0.0000	54	0.4186	0.0000
FAST	5	1.0000	0.0811	31	0.4247	0.0193	31	0.3974	0.0331
Harris	11	0.1170	0.6387	6	0.2727	0.7380	7	0.8750	0.1628
ORB *	74	0.8409	0.0000	165	0.5935	0.0000	140	0.5072	0.0000
Shi-Tomasi	11	0.1100	0.9012	12	0.1200	0.9505	11	0.1122	0.9780
SIFT	7	0.8750	0.3071	26	0.4906	0.5564	21	0.7241	0.0204
SURF	19	0.7600	0.9658	35	0.2258	0.6841	29	0.8286	0.3505
SAR-SIFT	2	1.0000	0.0857	68	0.7556	0.0000	60	0.0338	0.1002
KeyNet	24	0.1043	0.0080	28	0.0532	0.0130	27	0.0961	0.0294

Table A6. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 7.

Layers Detectors	Grayscale Layer			Gradient Layer (Sobel) *			PC Layer		
	N	P	D	N	P	D	N	P	D
AKAZE	13	0.7222	0.0884	34	0.3148	0.0002	22	0.2095	0.0107
BRISK	162	0.1826	0.0000	313	0.1389	0.0000	145	0.1284	0.0000
FAST	106	0.1715	0.0000	260	0.1284	0.0000	115	0.1192	0.0000
Harris	3	0.1034	0.3431	3	0.0667	0.9466	14	0.0680	0.1860
ORB *	105	0.3079	0.0000	153	0.3533	0.0000	80	0.1882	0.0000
Shi-Tomasi	16	0.1600	0.5171	19	0.1900	0.4193	17	0.1700	0.6678
SIFT	43	0.1946	0.0000	71	0.1659	0.0000	73	0.0874	0.0000
SURF	69	0.1364	0.0000	104	0.1453	0.0000	65	0.1275	0.0000
SAR-SIFT	12	1.0000	0.0004	26	0.3333	0.0000	102	0.1067	0.0000
KeyNet	45	0.1184	0.0292	75	0.1276	0.8836	50	0.1160	0.0029

Table A7. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 9.

Layers Detectors	Grayscale Layer			Gradient Layer (Sobel)			PC Layer *		
	N	P	D	N	P	D	N	P	D
AKAZE	529	0.6861	0.0000	1205	0.6646	0.0000	44	0.6875	0.0000
BRISK	7747	0.6357	0.0000	10,499	0.6281	0.0000	676	0.6877	0.0000
FAST	5964	0.6395	0.4485	8640	0.6219	0.9960	529	0.7178	0.0000
Harris	696	0.6005	0.1376	24	0.3810	0.0005	67	0.6700	0.0000
ORB	360	0.7200	0.0000	383	0.7660	0.0000	286	0.6560	0.0000
Shi-Tomasi	71	0.7100	0.2824	67	0.6700	0.0769	75	0.7500	0.0808
SIFT	1060	0.6397	0.0000	1308	0.6384	0.0002	52	0.6500	0.0000
SURF *	2211	0.6159	0.0000	2593	0.6149	0.0000	141	0.7790	0.0000
SAR-SIFT	157	0.8870	0.0000	157	0.8351	0.0000	2747	0.5889	0.0000
KeyNet	1226	0.6284	0.9184	1285	0.6403	0.9594	728	0.6547	0.0036

Table A8. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 11.

Layers Detectors	Grayscale Layer			Gradient Layer (Sobel)			PC Layer *		
	N	P	D	N	P	D	N	P	D
AKAZE	2	1.0000	0.9330	5	1.0000	0.9475	21	1.0000	0.6758
BRISK *	7	1.0000	0.0012	125	1.0000	0.0000	195	0.9949	0.0014
FAST	6	1.0000	0.0001	129	0.8543	0.0000	200	0.7692	0.4320
Harris	46	0.7667	0.3578	31	0.6327	0.8912	52	0.5417	0.9936
ORB	14	1.0000	0.0660	130	1.0000	0.0000	177	1.0000	0.0000
Shi-Tomasi	56	0.6154	1.0000	64	0.6882	1.0000	58	0.6824	1.0000
SIFT	68	0.8831	0.2643	187	0.7924	0.0001	218	0.7676	0.0011
SURF	88	0.9362	0.6021	125	0.9615	0.4207	105	0.9211	0.8922
SAR-SIFT	18	0.8571	0.0000	177	0.8009	0.0000	179	0.7366	0.0000
KeyNet	76	0.9383	0.0249	95	0.9048	0.0258	115	0.9274	0.7302

Table A9. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 12.

Layers Detectors	Grayscale Layer *			Gradient Layer (Scharr)			PC Layer		
	N	P	D	N	P	D	N	P	D
AKAZE	6	1.0000	0.4100	60	1.0000	0.0573	16	1.0000	0.9797
BRISK *	123	1.0000	0.0000	783	0.9962	0.0000	196	1.0000	0.0040
FAST	47	0.7966	0.0001	811	0.7502	0.9989	198	0.8082	0.6602
Harris	59	0.6705	0.9429	1	1.0000	0.9112	43	0.6615	0.9870
ORB	118	1.0000	0.0000	247	1.0000	0.0000	139	1.0000	0.0000
Shi-Tomasi	57	0.6404	1.0000	66	0.6804	1.0000	56	0.7000	1.0000
SIFT	220	0.7885	0.3889	98	0.8448	0.0001	104	0.8125	0.0051
SURF	105	0.9906	0.1416	271	0.9679	0.0134	155	0.9810	0.1540
SAR-SIFT	24	0.8276	0.0000	4	1.0000	0.0000	207	0.7782	0.0000
KeyNet	97	0.9604	0.6042	126	0.9618	0.9298	112	0.9492	0.4372

Table A10. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 13.

Layers Detectors	Grayscale Layer *			Gradient Layer (Laplacian)			PC Layer		
	N	P	D	N	P	D	N	P	D
AKAZE	0	1.0000	1.0000	0	1.0000	1.0000	0	1.0000	1.0000
BRISK	6	1.0000	0.1759	15	1.0000	0.0385	4	1.0000	0.9495
FAST *	31	1.0000	0.8846	64	0.7805	0.8736	39	0.9512	0.3238
Harris	1	0.2500	0.9724	1	1.0000	0.9724	8	1.0000	0.3102
ORB	0	1.0000	1.0000	0	1.0000	1.0000	0	1.0000	1.0000
Shi-Tomasi	11	0.8462	0.8202	8	0.5714	0.9901	9	0.9000	0.2691
SIFT	26	0.9630	0.2636	20	0.8333	0.0044	22	1.0000	0.3143
SURF	3	1.0000	0.9135	7	1.0000	0.9678	5	1.0000	0.9197
SAR-SIFT	16	1.0000	0.0000	17	1.0000	0.0000	32	0.6275	0.0000
KeyNet	7	1.0000	0.1447	7	1.0000	0.0313	7	1.0000	0.0243

Table A11. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 14.

Layers Detectors	Grayscale Layer *			Gradient Layer (Scharr)			PC Layer		
	N	P	D	N	P	D	N	P	D
AKAZE	12	1.0000	0.8286	18	1.0000	0.7376	10	1.0000	0.9262
BRISK *	82	1.0000	0.0019	319	0.9785	0.0003	83	1.0000	0.0007
FAST	58	0.8056	0.0523	387	0.6355	0.9953	104	0.7820	0.4632
Harris	17	0.3864	0.7894	1	1.0000	0.9713	14	0.7368	0.9557
ORB	55	1.0000	0.0000	91	1.0000	0.0000	49	1.0000	0.0000
Shi-Tomasi	31	0.5536	0.9953	30	0.5357	0.1682	23	0.5897	0.9782
SIFT	80	0.7692	0.8076	72	0.6486	0.0002	40	0.9756	0.0253
SURF	52	0.9811	0.8918	108	0.9818	0.9223	74	0.9610	0.2070
SAR-SIFT	41	0.9318	0.0000	1	0.5000	0.9713	69	0.4157	0.0000
KeyNet	51	0.9273	0.0497	58	0.8788	0.9141	44	0.9362	0.0040

Table A12. Comparison of Number of keypoints (*N*), Precision (*P*), and Distribution (*D*) of image 15.

Layers Detectors	Grayscale Layer			Gradient layer (Sobel) *			PC Layer		
	N	P	D	N	P	D	N	P	D
AKAZE	20	1.0000	0.0000	56	0.9180	0.0000	326	0.7342	0.2802
BRISK	39	0.9070	0.0001	344	0.8190	0.0000	1177	0.6386	0.3195
FAST	6	0.8571	0.0862	152	0.7415	0.0000	624	0.5714	0.8864
Harris	24	0.7500	0.1796	62	0.7126	0.0663	203	0.5867	0.9995
ORB *	249	0.8384	0.0000	398	0.8964	0.0000	373	0.7460	0.0000
Shi-Tomasi	73	0.7300	0.8915	71	0.7100	0.6492	46	0.4600	0.9765
SIFT	171	0.8953	0.0000	280	0.7609	0.0000	1021	0.6713	0.4981
SURF	290	0.7160	0.2040	639	0.6968	0.0041	547	0.5295	0.4551
SAR-SIFT	57	0.1926	0.0000	386	0.2338	0.0000	544	0.2981	0.0000
KeyNet	317	0.6967	0.0000	359	0.5583	0.0079	451	0.4713	0.2504

Table A13. Comparison of Number of keypoints (N), Precision (P), and Distribution (D) of image 16.

Detectors \ Layers	Grayscale Layer			Gradient layer (Scharr) *			PC Layer		
	N	P	D	N	P	D	N	P	D
AKAZE	7	0.8750	0.4935	373	0.2087	0.0000	26	0.2185	0.0244
BRISK *	64	0.2560	0.0000	2882	0.2230	0.0000	265	0.2221	0.0000
FAST	33	0.2025	0.0000	2216	0.2014	0.0000	189	0.2043	0.0000
Harris	65	0.2038	0.0000	46	0.2738	0.0000	52	0.2023	0.0123
ORB	91	0.2907	0.0000	68	0.1360	0.0000	113	0.2369	0.0000
Shi-Tomasi	21	0.2100	0.0045	17	0.1700	0.8357	17	0.1700	0.0764
SIFT	33	0.6735	0.0002	446	0.1739	0.0256	65	0.2273	0.0000
SURF	135	0.3890	0.0000	673	0.2165	0.1812	141	0.2117	0.0000
SAR-SIFT	29	1.0000	0.0026	123	0.2808	0.0000	583	0.1462	0.0000
KeyNet	210	0.1888	0.0308	409	0.2205	0.9114	215	0.1955	0.0000

References

- Neira, J.; Sequeiros, C.; Huamani, R.; Machaca, E.; Fonseca, P. Review on unmanned underwater robotics, structure designs, materials, sensors, actuators, and navigation control. *J. Robot.* **2021**, *1*, 22–30. [\[CrossRef\]](#)
- Ma, J.; Jiang, X.; Fan, A.; Jiang, J.; Yan, J. Image matching from handcrafted to deep features: A survey. *Int. J. Comput. Vis.* **2021**, *129*, 23–79. [\[CrossRef\]](#)
- Jiang, X.; Ma, J.; Xiao, G.; Shao, Z.; Guo, X. A review of multimodal image matching: Methods and applications. *Inf. Fus.* **2021**, *73*, 22–71. [\[CrossRef\]](#)
- Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [\[CrossRef\]](#)
- Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up robust features. In Proceedings of the 2006 European Conference on Computer Vision, Berlin, Germany, 18–22 September 2006.
- Vandrish, P.; Vardy, A.; Walker, D.; Dobro, O. Side-scan sonar image registration for AUV navigation. In Proceedings of the 2011 IEEE Symposium on Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies, Tokyo, Japan, 5–8 April 2011.
- King, P.; Anstey, B.; Vardy, A. Comparison of feature detection techniques for AUV navigation along a trained route. In Proceedings of the 2013 OCEANS-San Diego, San Diego, CA, USA, 23–27 September 2013.
- Daniel, S.; Le Léannec, F.; Roux, C.; Soliman, B.; Maillard, E. Side-scan sonar image matching. *IEEE J. Ocean. Eng.* **1998**, *23*, 245–259. [\[CrossRef\]](#)
- Khater, H.A.; Gad, A.S.; Omran, E.A.; Abdel-Fattah, A. Enhancement matching algorithms using fusion of multiple similarity metrics for sonar images. *World Appl. Sci. J.* **2009**, *6*, 759–763.
- Zhang, W.; Zhou, T.; Xu, C.; Liu, M. A SIFT-Like Feature Detector and Descriptor for Multibeam Sonar Imaging. *J. Sens.* **2021**, *2021*, 8845814. [\[CrossRef\]](#)
- Tao, W.; Liu, Y. Combined imaging matching method of side scan sonar images with prior position knowledge. *IET Image Process.* **2018**, *12*, 194–199. [\[CrossRef\]](#)
- Shang, X.; Zhao, J.; Zhang, H. Automatic overlapping area determination and segmentation for multiple side scan sonar images mosaic. *IEEE J.-Stars* **2021**, *14*, 2886–2900. [\[CrossRef\]](#)
- Tueller, P.; Kastner, R.; Diamant, R. A comparison of feature detectors for underwater sonar imagery. In Proceedings of the OCEANS 2018 MTS/IEEE Charleston, Charleston, SC, USA, 22–25 October 2018.
- Ansari, S. A review on SIFT and SURF for underwater image feature detection and matching. In Proceedings of the 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 20–22 February 2019.
- Zhou, X.; Yu, C.; Yuan, X.; Luo, C. Matching Underwater Sonar Images by the Learned Descriptor Based on Style Transfer Method. In Proceedings of the 2021 2nd International Conference on Mechatronics Technology and Intelligent Manufacturing (ICMTIM 2021), Hangzhou, China, 13–15 August 2021.
- Oliveira, A.J.; Ferreira, B.M.; Cruz, N. A Performance Analysis of Feature Extraction Algorithms for Acoustic Image-Based Underwater Navigation. *J. Mar. Sci. Eng.* **2021**, *9*, 361. [\[CrossRef\]](#)
- Oppenheim, A.V.; Lim, J.S. The importance of phase in signals. *Proc. IEEE* **1981**, *69*, 529–541. [\[CrossRef\]](#)
- Ma, W.; Wu, Y.; Liu, S.; Su, Q.; Zhong, Y. Remote sensing image registration based on phase congruency feature detection and spatial constraint matching. *IEEE Access* **2018**, *6*, 77554–77567. [\[CrossRef\]](#)
- Li, J.; Hu, Q.; Ai, M. RIFT: Multi-modal image matching based on radiation-variation insensitive feature transform. *IEEE Trans. Image Process* **2019**, *29*, 3296–3310. [\[CrossRef\]](#)
- Ye, Y.; Shan, J.; Hao, S.; Bruzzone, L.; Qin, Y. A local phase based invariant feature for remote sensing image matching. *Remote Sens.* **2018**, *142*, 205–221. [\[CrossRef\]](#)
- Kovesi, P. Phase congruency detects corners and edges. In Proceedings of the Australian Pattern Recognition Society Conference: DICTA, Sydney, Australia, 10–12 December 2003.
- Bradski, G. The OpenCV library. *Dr Dobb's J. Softw. Tools* **2020**, *120*, 122–125.

23. Barroso-Laguna, A.; Riba, E.; Ponsa, D.; Mikolajczyk, K. Key. Net: Keypoint detection by handcrafted and learned CNN filters. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019.
24. Riba, E.; Mishkin, D.; Ponsa, D.; Rublee, E.; Bradski, G. Kornia: An Open Source Differentiable Computer Vision Library for PyTorch. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass, CO, USA, 14 May 2020.
25. Alcantarilla, P.F.; Solutions, T. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 1281–1298.
26. Leutenegger, S.; Chli, M.; Siegwart, R.Y. BRISK: Binary robust invariant scalable keypoints. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 26 January 2011.
27. Rosten, E.; Porter, R.; Drummond, T. Faster and better: A machine learning approach to corner detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *32*, 105–119. [[CrossRef](#)] [[PubMed](#)]
28. Harris, C.; Stephens, M. A combined corner and edge detector. In Proceedings of the Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988.
29. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011.
30. Shi, J. Good features to track. In Proceedings of the 1994 IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 6 August 1994.
31. Dellinger, F.; Delon, J.; Gousseau, Y.; Michel, J.; Tupin, F. SAR-SIFT: A SIFT-like algorithm for SAR images. *IEEE Trans. Geosci. Remote* **2014**, *53*, 453–466. [[CrossRef](#)]
32. T-SEA Marine Technology Company. Available online: <http://www.t-sea.com/> (accessed on 6 September 2022).
33. Pengcheng Laboratory. Available online: <https://www.pcl.ac.cn/> (accessed on 6 September 2022).
34. Valdenegro-Toro, M. Improving sonar image patch matching via deep learning. In Proceedings of the 2017 European Conference on Mobile Robots (ECMR), Paris, France, 6–8 September 2017.
35. Sound Metrics. Available online: <http://www.soundmetrics.com/> (accessed on 6 September 2022).
36. DeepVision AB Company. Available online: <http://deepvision.se/> (accessed on 6 September 2022).
37. Zhang, T.; Liu, S.; He, X.; Huang, H.; Hao, K. Underwater Target Tracking Using Forward-Looking Sonar for Autonomous Underwater Vehicles. *Sensors* **2020**, *20*, 102. [[CrossRef](#)] [[PubMed](#)]
38. Kazimierski, W.; Zaniwicz, G. Determination of Process Noise for Underwater Target Tracking with Forward Looking Sonar. *Remote Sens.* **2021**, *13*, 1014. [[CrossRef](#)]
39. Wawrzyniak, N.; Stateczny, A. MSIS Image Positioning in Port Areas with the Aid of Comparative Navigation Methods. *Pol. Marit. Res.* **2017**, *24*, 32–41. [[CrossRef](#)]