Blind sonar image quality assessment via machine learning: leveraging micro- and macro-scale texture and contour features in the wavelet domain.

TOLIE, H.F., REN, J., CHEN, R., ZHAO, H. and ELYAN, E.

2025

Research paper

Blind sonar image quality assessment via machine learning: Leveraging micro- and macro-scale texture and contour features in the wavelet domain

Hamidreza Farhadi Tolie [a,c], Jinchang Ren [b,a,c,*], Rongjun Chen [b], Huimin Zhao [b], Eyad Elyan [c]

[a] *National Subsea Centre, Robert Gordon University, Aberdeen, AB21 0BH, UK*
[b] *School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, 610556, China*
[c] *School of Computing, Engineering, and Technology, Robert Gordon University, Aberdeen, AB10 7AQ, UK*

ARTICLE INFO

ABSTRACT

In subsea environments, sound navigation and ranging (SONAR) images are widely used for exploring and monitoring infrastructures due to their robustness and insensitivity to low-light conditions. However, their quality can degrade during acquisition and transmission, where standard SONAR image processing techniques can hardly produce high-quality outcomes. An effective image quality assessment (IQA) method can assess their usefulness and aid to develop refinement techniques by identifying the degradation issues, ensuring the reliability of SONAR data. Existing methods often fail to account for degradations from noise, distortion, and resolution changes simultaneously. To address this challenge, we propose a new blind quality assessment method that measures the overall quality of SONAR images by quantifying both the perceptual and utility qualities using the micro- and macro-scale texture and contour features derived from the wavelet domain. By combining the local binary pattern (LBP) micro-scale texture features with the proposed histograms of Schmid Gabor-like edge maps as macro-scale features, a support vector regression model is learned to map from these features to subjective quality scores. Extensive experiments have demonstrated the superiority of our method over existing SONAR IQA techniques on distorted and reconstructed super-resolution side-scan, acoustic lens, and forward-looking SONAR images. Specifically, our method achieves Pearson's and Spearman's correlation metrics of 0.8616 and 0.8541, respectively, for distorted SONAR images, demonstrating improvements of 4.69% and 4.8%. For reconstructed super-resolution SONAR images, our method attains correlation metrics of 0.9415 and 0.9408, reflecting improvements of 0.8% and 1.6% over the second-best method, respectively. To facilitate ease of access, a comprehensive list of key abbreviations and their full names is provided in Table A.9 in the Appendix section. The source code of the proposed method will be shared at https://github.com/hfarhaditolie/BSIQA.

1. Introduction

During the last few years, there has been considerable interest in exploring the underwater world. With the development of marine-related science and technologies, researchers have investigated underwater navigation, exploration, mapping and object tracking through underwater vision (Li et al., 2020). Optical sensors have been widely used various fields such as condition monitoring of pipelines (Liu and Bao, 2022) and crack detection (Liu and Bao, 2023), however, as optical wavelengths poorly penetrate into deep water, this limits their functionality in the subsea (Jaffe, 2014), leading to inaccurate and low-performance vision methods. In contrast, acoustic and sound waves can easily transmit in underwater scenarios (Callow, 2003) to recognize the presence of the various objects and targets. Acoustics waves can be used either individually or complementary to other modalities to provide

more complete and accurate knowledge of the environment, especially in dark and harsh subsea environment (Luo et al., 2019; Chen et al., 2019c).

To collect acoustic images, sound navigation and ranging (SONAR) devices such as side-scan and forward-looking SONARs are usually mounted on ships, remotely operated vehicle (ROV) and autonomous underwater vehicle (AUV). Due to the limited processing capability of edge devices, the captured images are then transmitted through an acoustic channel for further analysis (Chen et al., 2019a; Zhang et al., 2022). Because of the unpredictable and complex underwater environment (Zhang et al., 2020) and considering the limited bandwidth and unstable link of the acoustic channel (Stojanovic and Freitag, 2013), during both the acquisition and transmission processes variety of distortions are introduced to the SONAR images. This results in

degraded and low-quality images (Chen et al., 2017). Thus it is vital to develop reliable, accurate, and efficient SONAR image quality assessment (SIQA) methods to guide both the acquisition and transmission processes toward a quality-assured data collection.

The image quality assessment (IQA) techniques have a wide range of applications across various fields. In camera manufacturing, IQA is used to optimize the image processing pipeline directly in the device, detect low-quality images, and identify artifacts to ensure the usability of the images. Additionally, IQA ensures that images and videos captured by autonomous vehicles are of high quality and usability, which is important for not only saving the costs and lower the risks and emissions but also for improving the accuracy of following tasks e.g. reliable object/target detection and environment mapping (Geiger et al., 2012). Regarding SIQA, these techniques are particularly useful in the development of sonar technology, such as transmission, enhancement, and compression algorithms. They can also assist in optimizing the placement and configuration of sonar equipment and developing super-resolution models to enhance the pixel density in low-resolution sonar images (Chen et al., 2024).

In general, the optical IQA methods can be classified into three categories based on the availability of the reference image including full-reference (FR), reduced-reference (RR), and no-reference (NR) methods. In FR methods, the entire information of the reference image is available so that quality evaluation is adopted by comparison between particular features from both the reference and distorted images. In contrast, NR methods do not have access to any reference information, thus they commonly train conventional machine learning or deep learning methods on these images to learn the difference between distortion types, distortion levels, and content variation in line with the quality assessment. On the other hand, RR methods evaluate the quality by taking partial information of the reference image into account. In the case of underwater SONAR image acquisition and transmission, RR and NR methods would be preferable since the reference image is not often available.

Depending on the type of source image, the quality evaluation differs to better incorporate the image's characteristics and the influence of the individual distortions. To this end, particular IQA methods have been designed for natural scene (Wang et al., 2004; Moorthy and Bovik, 2011a; Mittal et al., 2013), screen content (Gu et al., 2017; Ni et al., 2017; Tolie and Faraji, 2022), tone-mapped (Yue et al., 2019; Fang et al., 2020b), low-light (Zhang et al., 2021), underwater (Yang and Sowmya, 2015; Panetta et al., 2016; Wang et al., 2018) images, etc. In acoustic imaging, the SONAR device sends sound pulses and then records the arrival time of the returning echo signals to form an image (Callow, 2003). Thus the captured SONAR images are not only different from the visual perspective but also from the quality perspective. Due to the use of SONAR images in practical scenarios rather than for human observation, it is important to measure their quality with respect to their utility quality (Chen et al., 2019a).

In recent years, some efforts have been made to effectively and efficiently evaluate the quality of the SONAR images. The developed methods include two FR, two RR, and four NR methods, namely, the FR SONAR image quality predictor (SIQP) (Chen et al., 2019b), the FR SONAR image quality evaluation (SOIQE) (Zhang et al., 2020), the partial-reference SONAR image quality predictor (PSIQP) (Chen et al., 2018), the RR task- and perception-oriented SONAR image quality assessment (TPSIQA) (Chen et al., 2020), the NR contour degradation measurement (NRCDM) (Chen et al., 2019a), the NR dual-path deep neural network (DPDNN) (Zhang et al., 2022), perception-and-cognition-inspired quality assessment method for sonar image super-resolution (PCASS) (Chen et al., 2024), and super-resolution sonar image quality assessment (SRSIQA) (Feng et al., 2024). However, their performance is not ideal enough, especially for the RR and NR methods, due to the following reasons. Firstly, in most of these works, high-level features such as brightness, entropy, and statistical information of the images are extracted, which generally retain the perceptual quality.

Secondly, the structural features are extracted and considered as the utility quality of the images, while they mainly demonstrate the level of informativeness not the effectiveness for object recognition/detection.

From Fig. 1, it can be seen that the occurring degradations during the acquisition and transmission generally affect the texture, edges, and contour information of the objects in the SONAR images. As mentioned earlier the quality of SONAR images is also related to their utility, which means the less the texture, gradients, and contours degrades, the better the performance of vision methods would be. Therefore, considering object detection and recognition tasks performance as the utility quality, the texture and contour continuity can be used to measure the degradation that occurs at the texture level on the continuity of the object's boundaries. Hence, in this work, we propose to represent the SONAR images by texture and contour descriptors in the wavelet domain to benefit from both the frequency and spatial domain information. Specifically, we utilized the well-known uniform local binary pattern (ULBP) descriptor (Ojala et al., 2002) and fused it with a proposed Gabor-like filters-inspired texture descriptor to represent images. The image's quality is then assessed by training a regression model over the generated representations. The major contributions of our work can be highlighted as follows:

- Integrating multi-scale features through wavelet decomposition for effective SONAR image representation;
- Proposing a novel macro-scale contour information descriptor based on the image's gradient maps obtained using the Schmid Gabor-like filter bank (Schmid, 2001);
- Measuring the overall quality of SONAR images by quantifying both perceptual and utility qualities using advanced micro- and macro-scale texture and contour features;
- Predicting the quality of SONAR images by training a support vector regression (SVR) model and verifying the findings through extensive experiments on publicly available datasets.

The rest of the manuscript is organized as follows. Section 2 briefly reviews the related work, Section 3 presents the proposed feature representation, i.e., the micro- and macro-scale texture and contour feature extraction, and describes the quality regression. Section 4 provides analyses over components of the proposed metric and compares the obtained performance with the selected FR, RR, and NR SIQA methods. Finally, Section 5 draws the conclusion.

2. Related work

The quality of the SONAR images can be defined as the combination of both the visual and the utility qualities. Thus, in subjective measurements, the mean opinion score (MOS) and the existence of the target are considered together to represent the visual perception and usefulness of the image, respectively (Chen et al., 2017). As a consequence, various objective IQA methods have been proposed in line with the above-mentioned definition. In this section, previous studies are explained in detail.

2.1. Full-reference SIQA

The SIQP (Chen et al., 2019b), proposed by Chen et al., computes the local entropy and edge map from both the reference and distorted SONAR images to represent their statistical and structural information. Next, it computes the similarity maps between the extracted feature maps from both images and weighs them by the activity map (Saha and Vemuri, 2000) to measure the degradation with respect to statistical and structural perspectives. The overall quality score is then assessed by fusing the extracted features via a quadratic polynomial model. In 2020, Zhang et al. proposed the SOIQE (Zhang et al., 2020) method, in which the structural features of the image are extracted and then weighed by the main target region map to measure the degradation level. The utilized structural features are brightness, contrast, and
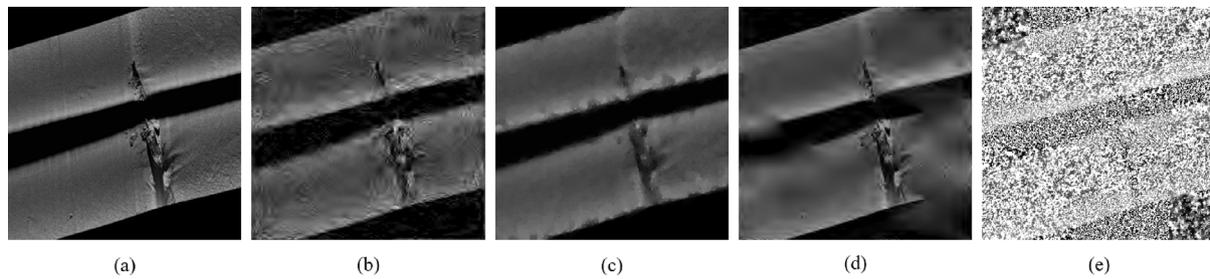
Fig. 1. Sample SONAR image and its distorted versions. (a) reference image (b)–(e) distorted SONAR images, all taken from the SIQD (Chen et al., 2017) database.

sharpness which are quantized by computing images' local mean, local variance, and gradient map, respectively. To obtain the main target regions, Zhang et al. utilized the residual Fourier amplitude spectrum inspired by Hou and Zhang (2007).

2.2. Reduced-reference SIQA

The designed FR SIQA methods can be used in designing the transmission channel and developing compression or transmission methods to compare the quality of the received image with its original version. However, as mentioned earlier, in practical scenarios, e.g. data acquisition and survey missions, it is crucial to evaluate the captured image's quality in real-time. In such scenarios, the reference image is not usually available, therefore the RR and NR methods are the only way of the measurements.

To this end, Chen et al. proposed the PSIQP (Chen et al., 2018) by taking the statistical information of the distorted image and the structural similarity between the reference and distorted image into account. PSIQP basically extracts the entropy, third- and fourth-statistical moments (i.e., skewness, kurtosis) from the distorted SONAR image and linearly combines them with the extracted structural similarity between the reference and distorted images. It should be noted that as the bandwidth of the transmission channel is limited, it is not always easy to even transfer the computed structural/edge maps as they nearly have the same resolution as the originally captured image. Thus, Chen et al. proposed to divide the edge maps into $m \times m$ blocks and then compute a normalized histogram to represent the distribution of the edges. This reduces the amount of data that needs to be transferred and makes some information about the original image available on the receiver side. PSIQP computes the similarity between the structural histograms of reference and distorted images and weighs it by the received (probably distorted) image's normalized activity map (Saha and Vemuri, 2000) to determine the structural score of the image. The obtained quality indexes are then linearly combined by pre-defined weights to generate the overall quality score. In another attempt, Chen et al. proposed the TPSIQA metric (Chen et al., 2020), which merges the task- and visual perception-derived features to obtain the overall quality score. TPSIQA extracts statistical information (i.e., amount of information or entropy, the energy fluctuation and amplitude magnitude both in the logarithmic domain) from the contourlet coefficients in 10 subbands of the decomposed reference image to reduce the amount of reference information. The received/distorted image is also represented with the same approach. Using the extracted feature vector m learners are then trained over the randomly selected subset of the feature set and the results are linearly combined using identical weights to determine the overall quality score.

2.3. No-reference SIQA

Although the existing RR SIQA methods are effective to reduce the amount of data transferred, extracting the features requires computational resources. This brings up the following challenges: (1) the

introduced distortions during the transmission will also affect the extracted features from the reference image, which makes the quality evaluation unreliable; (2) sender- and receiver-level feature extraction and quality assessment will increase the computational run-time leading to delay in data acquisition. Therefore, NR SIQA methods could be the best alternative to tackle the aforementioned issues.

The blind natural image quality assessment (BNIQA) methods such as the blind image integrity notator using DCT statistics (BLIINDS II) (Saad et al., 2012), blind/referenceless image spatial quality evaluator (BRISQUE) (Mittal et al., 2012), natural image quality evaluator (NIQE) (Mittal et al., 2013), and perception-based image quality evaluator (PIQE) (Venkatanath et al., 2015) generally use the perceptual statistics of the images to obtain a quality score. However, NR SONAR IQA methods mainly try to represent the images using utility-related features. To this end, the NRCDM and DPDNN methods have been recently proposed. NRCDM (Chen et al., 2019a), proposed by Chen et al., measures the contour degradation degree of the SONAR images to address the utility quality. Firstly, NRCDM further degrades the input image by filtering the specific frequency components in the wavelet domain. Secondly, to quantify the contour information, NRCDM represents both the input image and its degraded version in the transformed frequency domains (i.e., discrete cosine transform and Cohen–Daubechies–Feauveau 9/7 wavelet transform) beside their singular value decomposition (SVD) (Kalman, 1996) coefficients computed in the spatial domain. The large coefficients in all three representations indicate the main contour information in the image. In other words, the more the image is degraded the more sparse these coefficients vectors would be. Therefore, the contour degradation is then computed by comparing the extracted features (i.e., sparsity measures: Hoyer measure and Gini Index Hurley and Rickard, 2009) in the transformed domain between the input image and its degraded version. Finally, a bag of SVR models is trained over the contour degradation degree to determine the overall quality score. DPDNN (Zhang et al., 2022), proposed by Zhang et al., trained a dual-path deep neural network to predict the quality of SONAR images. In the first path, microscopic features are extracted by a VGG-Net (Simonyan and Zisserman, 2014)-inspired structure, while in the second path, the well-known skip connection introduced in Res-Net (He et al., 2016) is employed to extract macroscopic features of the image. The extracted features are then fused and pooled to obtain the overall quality score.

Recently with the introduction of the super-resolution reconstructed SONAR images, Chen et al. proposed the PCASS method (Chen et al., 2024), which extracts low-, mid-, and high-level features from SONAR images at different scales and linearly combines them to determine the quality score. Feng et al. proposed a transfer-learning based deep neural network model, named SRSIQA (Feng et al., 2024), which utilized a pyramid-shape neural network to extract multi-level features before fusing them with an adaptive feature weight adjustment block to determine the best features obtained from transfer learning that matched well with SONAR image characteristics.

In this study, we propose a SIQA model using machine learning with hand-crafted features. To contextualize our work, we have reviewed and compared existing IQA techniques from various categories, including conventional image processing (IP) methods and modern artificial
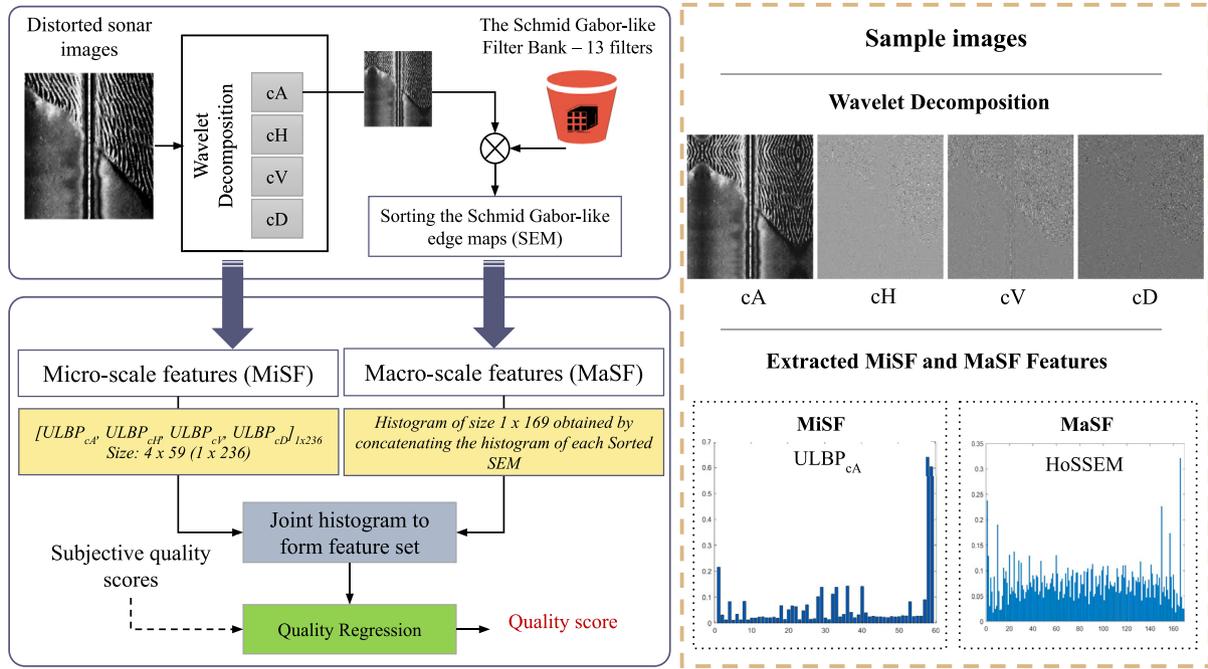
Fig. 2. A general framework of the proposed SIQA method.

Table 1
Information of the compared SIQA methods.

Image category	Method	IQA category	Intelligence
Natural scene images	BLIINDS II	NR	ML
	BRISQUE	NR	ML
	NIQE	NR	IP
	PIQE	NR	ML
Transmission distorted SONAR images	SIQP	FR	ML
	SOIQE	FR	IP
	PSIQP	RR	IP
	TPSIQA	RR	ML
	NRCDM	NR	ML
	DPDNN	NR	DL
Super-resolution reconstructed SONAR images	PCASS	NR	ML + DL
	SRSIQA	NR	DL
	Proposed	NR	ML

intelligence approaches, including both machine learning (ML) and deep learning (DL) methods. Table 1 provides a comparative summary of the relevant methods, including their intelligence category, i.e. IP, ML, or DL based.

3. The proposed method

As mentioned earlier, SONAR images are formed by recording the arrival time of the returning sound pulses at different frequencies. In acoustic imaging devices, to capture the presence of the objects in short- and long-ranges, it is crucial to send sound pulses at different frequencies, i.e. usually lower- and higher-frequencies for objects in the long-range and short-range, respectively (Xie et al., 2022). Consequently, in addition to the spatial degradations, the introduced distortions during the data acquisition and transmission affect the collected frequency information, resulting noise in the formed SONAR images.

Depending on the recorded frequency, i.e. low or high based on the object/target distance, SONAR images may exhibit micro- and macro-scale degradations that distort the texture, gradients, and object contours, respectively (Chen et al., 2019a). In this study, we utilized

micro- and macro-scale texture and contour descriptors to effectively capture these frequency- and spatial-domain degradations. Both feature sets are extracted in the wavelet domain, providing an accurate representation of both frequency and spatial information. Representing such information through feature sets, rather than raw images, facilitates the more effective machine learning based interpretation of SONAR images. The more precisely the features can differentiate between degradation types, such as blurring and noise, and their levels resulting from distortions or quality improvement algorithms, the better the machine learning model can map these features to the associated (ground-truth) quality scores.

The general framework of the proposed method is illustrated in Fig. 2. As seen, benefit from the frequency and spatial illustration of the wavelet transforms, we have extracted microscopic LBP features from both the frequency and spatial domains besides the macro-scale features extracted using the Schmid Gabor-like filter bank from the spatial information. The extracted features are subsequently combined and mapped to the subjective quality scores using an SVR model. For example, as shown on the right side of Fig. 2, a sample image is decomposed into spatial and frequency channels. The extracted MiSF feature from the approximation band and the MaSF features for the illustrated bands are also displayed. A detailed explanation of the wavelet transformation, the feature extraction, and the quality regression are provided in the following subsections.

3.1. Wavelet decomposition

The 2Dimensional (2D) wavelet transformation offers a comprehensive representation of both frequency and spatial information. This imparts robustness against common distortions, enables multi-resolution analysis for examining image features at various scales, and demonstrates effectiveness in localizing distortions, which is crucial for IQA (Gonzales and Wintz, 1987). As a result, it has been widely utilized in various image quality assessment tasks (Wang and Simoncelli, 2005; Moorthy and Bovik, 2011b; Xue et al., 2014; Rezaie et al., 2018; Yu et al., 2022) to capture the response of the human visual system (HVS) to various frequency components. According to Chen et al. (2019a), the low- and high-frequency components are generally relevant to the distortions. For instance, blur primarily impacts low-frequency
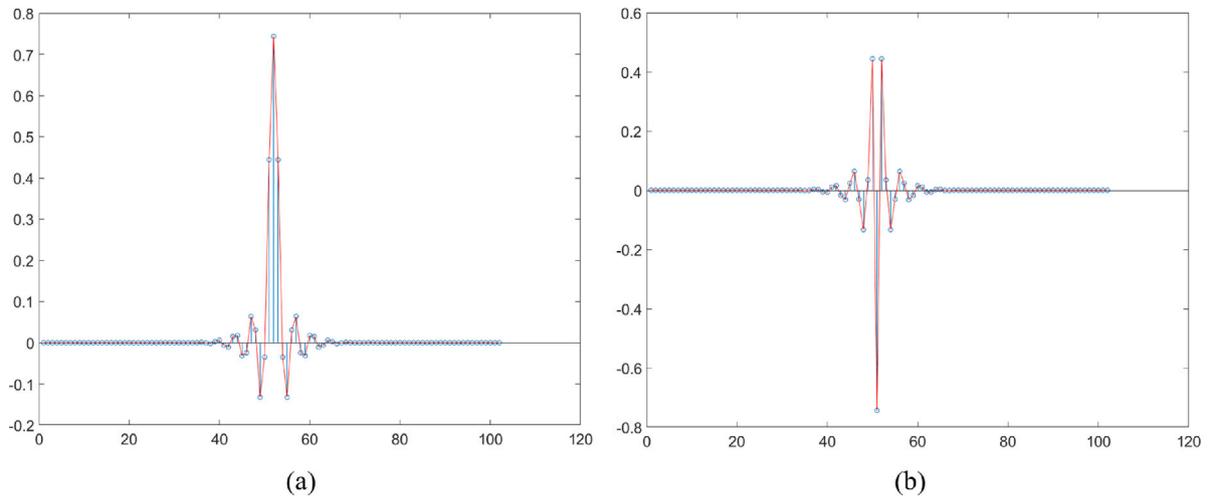
Fig. 3. Frequency spectrum of the (a) low-pass and (b) high-pass dmey filters.

components, leading to a reduction in global structure. On the other hand, high-frequency components are susceptible to noise, like speckles, which can obscure fine details. Additionally, the intermediate frequency represents the contour information of the objects within an image. Therefore, the frequency bands of the image can be decomposed by applying a 2D discrete wavelet transform (2D DWT).

The 2D DWT decomposes an image into the low-pass and high-pass subbands called approximation and detail bands, respectively. The approximation band, i.e., LL, contains the low-frequency information of the image, while the detail level, which consists of 3 bands represents the high-frequencies in horizontal, i.e., LH, vertical, i.e., HL, and diagonal, i.e., HH, orientations. The detail bands are also equivalent to the edge information of the image. Thus, the 2D DWT is highly correlated with HVS's sensitivity to the selectivity of the orientation and spatial frequency (Daugman, 1983; Marĉelja, 1980).

The wavelet decomposition is generally done by applying the low-pass and high-pass filtering on the input data. To this end, various wavelet filters such as *Haar, Daubechies, Biorthogonal wavelets Discrete Meyer, etc.* have been utilized in signal and image processing tasks. In this work, considering the characteristic of the available wavelet filters, we used the *Discrete Meyer* (dmey) filter for decomposition. That is because the dmey wavelet has the widest frequency spectrum, which results in better decomposition of certain frequency bands (Leontiev and Nyurova, 2019) and consequently distinguishes the distortions from the contour information. The low- and high-pass dmey filters are obtained through the following scaling and wavelet functions, respectively.

$$\hat{\phi}(\omega) = \begin{cases} 1 & \text{for } |\omega| \leq \frac{2\pi}{3} \\ \cos\left(\frac{\pi}{2}\nu\left(\frac{3|\omega|}{2\pi}-1\right)\right) & \text{for } \frac{2\pi}{3} < |\omega| < \frac{4\pi}{3} \\ 0 & \text{for } |\omega| \geq \frac{4\pi}{3} \end{cases} \quad (1)$$

$$\hat{\psi}(\omega) = \begin{cases} \sin\left(\frac{\pi}{2}\nu\left(\frac{3|\omega|}{2\pi}-1\right)\right) & \text{if } \frac{2\pi}{3} < |\omega| < \frac{4\pi}{3} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$\nu(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } 0 < x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} \quad (3)$$

where, $\hat{\phi}(\omega)$, $\hat{\psi}(\omega)$, and $\nu(x)$ denote the scaling, wavelet, and transition functions, respectively. The angular frequency ω is obtained via $\omega = 2\pi f$, where f is the frequency. The transition function is used to create a smooth transition between different frequency bands. Note that in practical implementations, these functions are computed through numerical approximations.

Fig. 3 illustrates the frequency spectrum of the low- and high-pass dmey filters. It can be seen that both filters are continuous and they consist of a wider range of frequency levels which helps to capture the

low and high frequencies at different levels. Also, it is apparent that in both filters, the opposite frequency is not fully eliminated, which helps to better represent the scene information of the image.

The wavelet decomposition of an image I is computed using the estimated approximation and detail bands as follows.

$$L_{\text{rows}} = I \otimes \text{LoD} \quad (4)$$

$$H_{\text{rows}} = I \otimes \text{HiD} \quad (5)$$

$$LL = L_{\text{rows}} \otimes \text{LoD}^T \quad (6)$$

$$LH = L_{\text{rows}} \otimes \text{HiD}^T \quad (7)$$

$$HL = H_{\text{rows}} \otimes \text{LoD}^T \quad (8)$$

$$HH = H_{\text{rows}} \otimes \text{HiD}^T \quad (9)$$

where $\otimes$ denotes the convolution operations and LoD and HiD are the low- and high-pass filters obtained by estimating the scaling and wavelet function in Eqs. (1) and (2), respectively. The result of applying these filters on a sample SONAR image is shown in Fig. 4. As seen, the decomposition has effectively captured the high- frequency changes in the three orientations of horizontal, vertical and diagonal.

3.2. Feature extraction

As discussed earlier, the quality of the SONAR images relies on both the perceptual and utility scores. Thus, to take both these quality indexes into account, we proposed to represent images using both micro-scale texture features, i.e. LBP histograms, and macro-scale texture features, i.e. Schmid Gabor-like texture features, in the wavelet domain. The micro-scale texture descriptor enables us to capture the small-scale changes that occur in the image which is relevant to both the perceptual and utility quality as we extract these features from both the approximation and detailed bands of the image in the wavelet domain. For instance, the distortion introduced in Fig. 1. (e) not only impacts the contour information but also adds noise to the image, leading to poor visual quality. These changes can effectively be captured by the LBP texture descriptor as the local structure of the image is changed.

On the other hand, the proposed macro-scale texture descriptor represents the image with respect to its contour and edge information, which can be linked to the utility or usability quality of the image. Degradation of edges and contours makes object detection more difficult, thereby affecting the overall usefulness of the image. As seen in Fig. 1. (d), the introduced distortion degrades the contour information
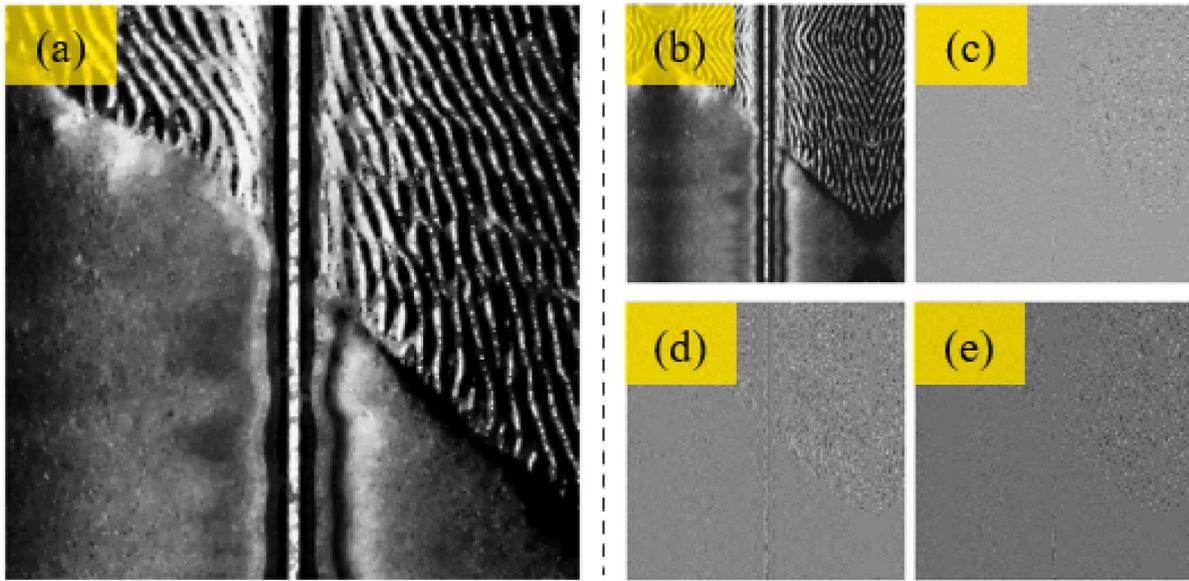
Fig. 4. Illustration of the wavelet decomposition of a sample SONAR image (a). (b), (c), (d), and (e) correspond to the approximation (LL) and horizontal (LH), vertical (HL), and diagonal (HH) detail bands, respectively.

by proportionally blurring the image. Since the resulting blurring occurred on a larger scale, the changes cannot be simply captured using the micro-scale LBP features. Therefore, we also proposed to represent the SONAR images via macro-scale features to capture both the large-scale and contour-level degradations. However, it should be noted that the distinction between the micro- and macro-scale distortions is subtle and certain type of distortion may adversely degrade the image at both levels. This subtlety arises from the complex and interconnected nature of distortions. Even though macro-scale distortions are presumed to impact the structural elements of images on a larger scale, their effects can still be detected through micro-scale features. Hence, we proposed to combine the micro- and macro-scale features in the wavelet domain to represent the images. Note that the wavelet domain-level feature extraction empowers the representation by discriminating the low- and high-frequency details of the image.

3.2.1. Micro-scale LBP texture features in wavelet domain

To capture the structural changes resulting from the quality degradation or the distortions introduced to the image, we proposed to utilize the well-known LBP (Ojala et al., 2002) texture descriptor. Basically, LBP encodes the micro-scale or local structures of the image at each pixel location with respect to its neighbor pixels as follows:

$$LBP(P, R) = \sum_{P=0}^{P-1} S(I_p - I_c) 2^p \tag{10}$$

where P is the angular resolution, R is the spatial resolution, S is the step function, and I_c and I_p are the intensity values of the central and its neighbor pixel, respectively.

It is common to use $P = 8$, $R = 1$, which leads to computing the LBP value of each pixel in a 3×3 block. Fig. 5 shows the 3×3 adjacent block at the pixel location I_c with $P = 8$ and $R = 1$. After identifying the LBP values, images can be represented by computing the histogram of the LBP values. The $LBP(8, 1)$, produces 2^8 different patterns leading to a histogram with 256 bins. However, the computed LBP map is not rotation invariant, which is not aligned with the utility quality. To measure the utility quality of the SONAR images, the representation should be rotation invariant to measure the object's detectability, not the visibility. Therefore, in this study, we represent the images using the uniform LBP (ULBP), which can provide nearly 90% of the image's micro-scale structural pattern (Zhang et al., 2013). A pattern is called uniform when it has a maximum of two transitions from 0 to 1 when
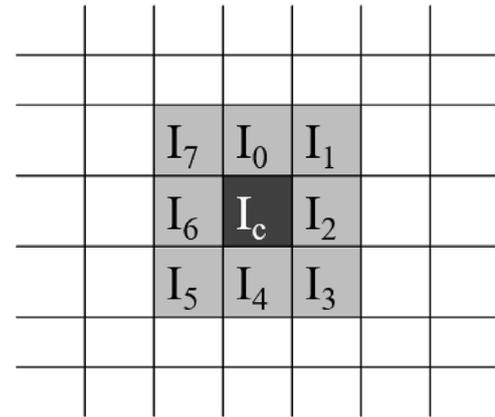
Fig. 5. Illustration of a circularly symmetric neighbor set with $P = 8$ and $R = 1$.

circularly traversed. The utilized $LBP(8, 1)$ produces 256 patterns of which 59 are uniform.

In this work, we use the ULBPs in the wavelet domain to represent the micro-scale features (MiSF). As the wavelet transform decomposes the image into low- and high-frequency bands, this enables us to also capture the micro-structural changes on both bands. Moreover, the high-frequency bands are generated in horizontal, vertical, and diagonal directions, thus the directionality of the high-frequency information is also taken into account to effectively represent the texture information and capture the degradations. Consequently, each SONAR image is represented by the MiSF features as follows:

$$MiSF = \left[ULBP_{cA}, ULBP_{cH}, ULBP_{cV}, ULBP_{cD} \right] \tag{11}$$

where cA, cH, cV, and cD denote the approximation, horizontal, vertical, and diagonal detailed bands coefficients, respectively.

To further enhance the contribution of the extracted ULBPs, we use the square root operation as suggested in Kusumoto et al. (2014) to normalize extracted 236-dimensional MiSF texture features as follows:

$$MiSF = \left[\sqrt{MiSF_1}, \sqrt{MiSF_2}, \dots, \sqrt{MiSF_{236}} \right] \tag{12}$$

The normalized ULBP histograms of a sample SONAR image are shown in Fig. 6. As seen, the occurrence rate of the patterns extracted
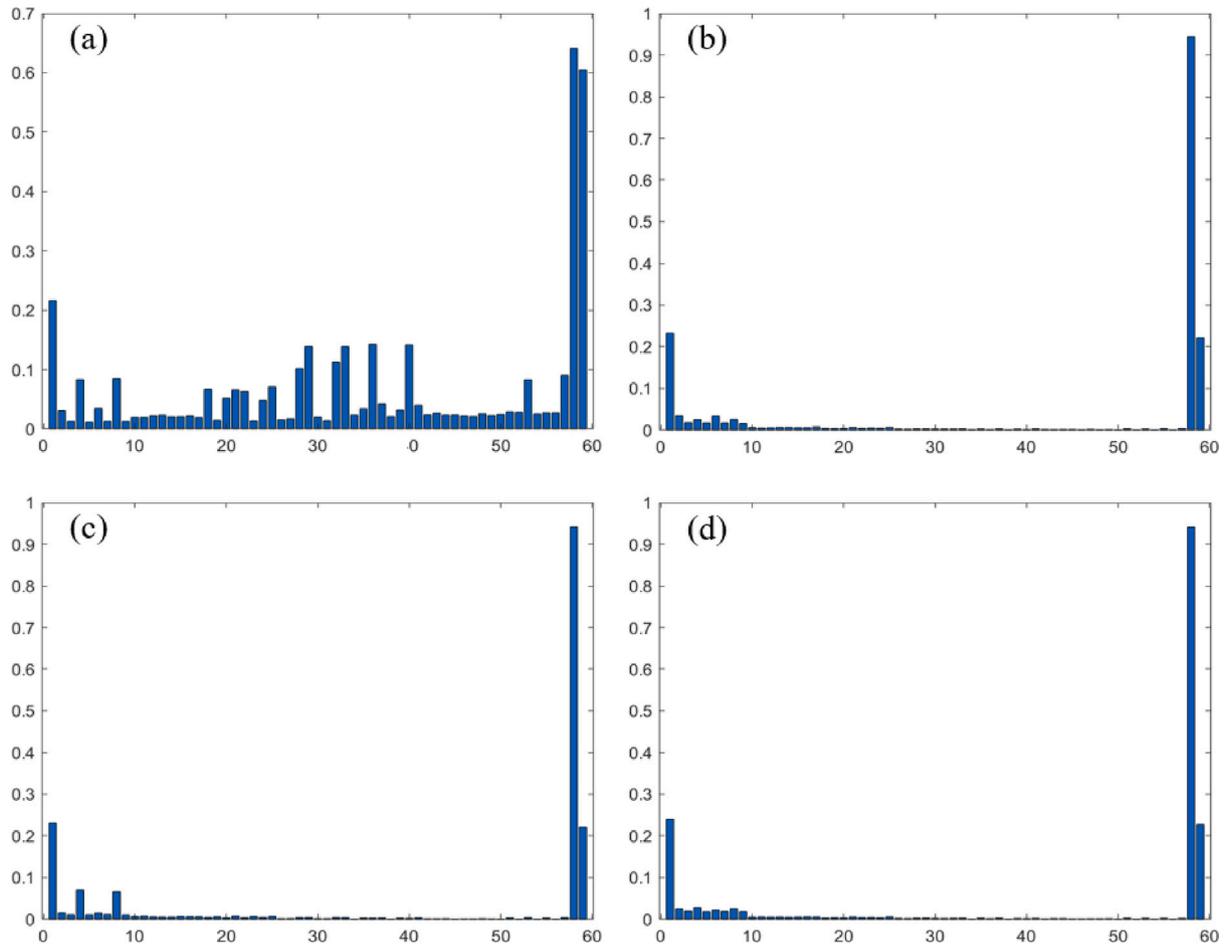
Fig. 6. Normalized ULBP histograms of (a) approximation and (b) horizontal, (c) vertical, and (d) diagonal detailed bands of the sample SONAR image illustrated in Fig. 4.
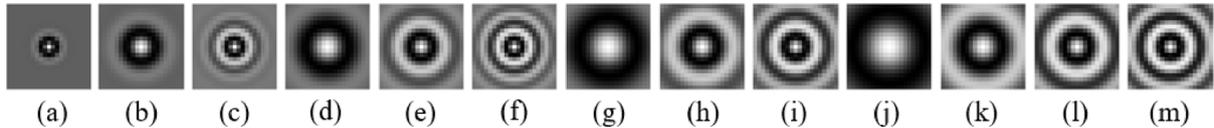
Fig. 7. Illustration of the Schmid Gabor-like filters obtained, from left to right, with the following (σ, τ) pairs: (2, 1), (4, 1), (4, 2), (6, 1), (6, 2), (6, 3), (8, 1), (8, 2), (8, 3), (10, 1), (10, 2), (10, 3), and (10, 4), respectively.

from the approximation band is almost equal, while in the detailed band, the first patterns occurred more than the others. This is because the approximation band contains more texture information than the detailed bands. In fact, in the detailed bands, only the high-frequency information is present leading to the occurrence of particular patterns. The occurrence of these patterns also changes based on the orientation of the high-frequency information. From Fig. 6, it can be understood that the extracted MiSF feature vector could effectively reflect both the low- and high-frequency texture changes.

3.2.2. Macro-scale Schmid Gabor-like texture features in wavelet domain

In addition to the micro-scale structural texture changes, degradations such as the blurring changes the image structure on a larger-scaler, i.e. macro-scale. These macro-scale degradation mainly affect the contour and edge information rather than the whole image. However, the proposed MiSF features cannot effectively capture these macro-scale changes. Thus, we proposed to address this by introducing a texture descriptor obtained from the image's gradient maps that are computed using the Schmid Gabor-like filters. To mitigate the effect of the micro-scale degradation on the extraction of the macro-scale features, we have used the approximation band of the image in the

wavelet domain. This approach preserves the overall structure and global features of the image, along with their corresponding degradations. Utilization of the Schmid Gabor-like filters is motivated by the fact that representing images using a set of Gabor filters is highly consistent with the human visual system (Ni et al., 2018). Unlike the Gabor filters with a non-zero DC component that makes them highly influenced by the average value of the input signal (Field, 1987), Schmid filters offer a more stable representation. This is crucial for vision applications, as they should not be overly sensitive to minor shifts of the input's mean value. Therefore, we employed the Schmid Gabor-like filters as follows:

$$F(x, y, \tau, \sigma) = F_0(\sigma, \tau) + cos(\frac{\sqrt{x^2 + y^2}\pi\tau}{\sigma})e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (13)$$

where $F_0(\sigma, \tau)$ is added to the Gaussian envelope of the filter (Gabor, 1946) to obtain a zero DC component.

The Schmid filters are generated by 13 pairs of scales σ between 2 and 10 and τ between 1 and 4, i.e., (2, 1), (4, 1), (4, 2), (6, 1), (6, 2), (6, 3), (8, 1), (8, 2), (8, 3), (10, 1), (10, 2), (10, 3), and (10, 4) originally in a 49×49 window. Note that in this study we have empirically set the window size to 21×21, which yields the highest performance
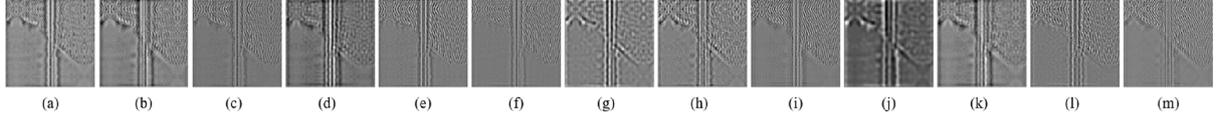
Fig. 8. Illustration of the edge maps (SEM) obtained by applying the Schmid Gabor-like filters shown in Fig. 7 on the approximation band of the SONAR image shown in Fig. 4. (a)–(m) are obtained by convolving the SONAR image by the Schmid filter with the (σ, τ) equal to (2, 1)-(10, 4), respectively.
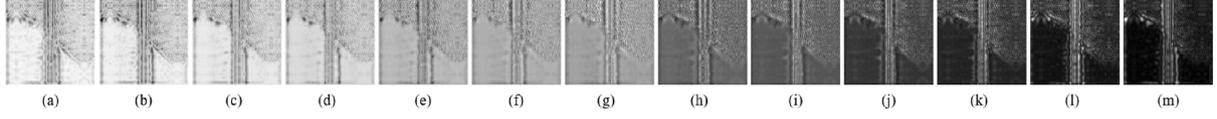
Fig. 9. Sorted SEMs of a sample SONAR image. From left to right each image shows the pixel-wise sorted SEM in an ascending order. (a) and (m) show the lowest and highest edge magnitudes obtained by applying the Schmid filters, respectively.
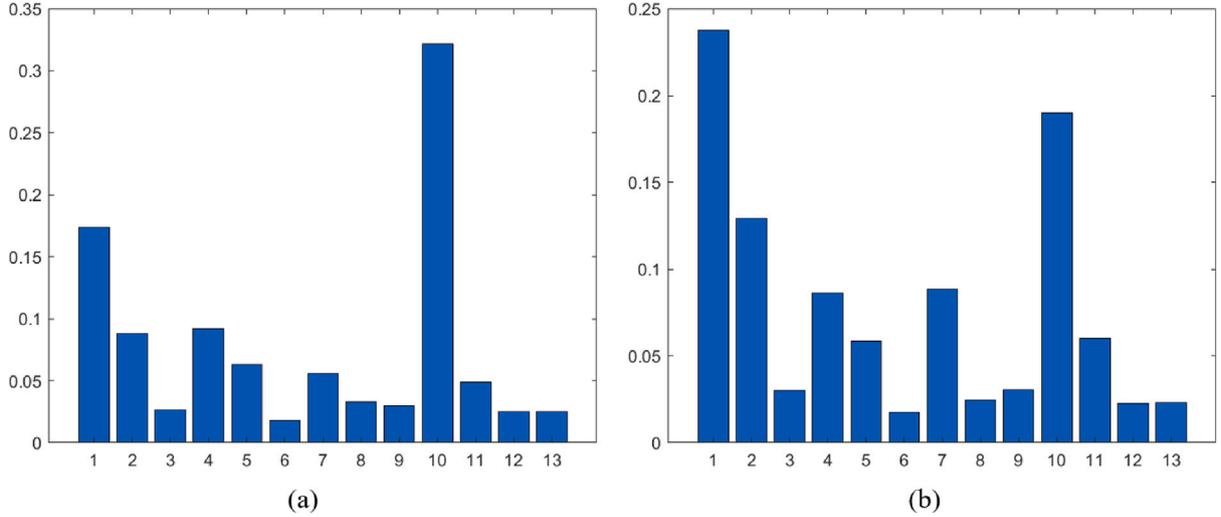
Fig. 10. Illustration of the HoSSEMs for the (a) highest and (b) lowest edge magnitudes. In both HoSSEMs, dominant edge types, i.e., obtained using the 1st and 10th Schmid filters, can be seen. However, the edge types of the high magnitude SSEM, i.e. (a), tend to have less fluctuation than the low magnitude SSEM.

for quality evaluation of the SONAR images based on the conducted experiments explained in . The produced Schmid filters with the size of 21×21 are illustrated in Fig. 7. Using the obtained 21×21 filters we have then computed the image's Schmid Gabor-like edge maps (SEM) in 13 different scales by convolving the image with the Schmid filters as follows:

$$SEM_i = I \otimes F_i(\tau, \sigma), i \in \{1, 2, \ldots, 13\} \tag{14}$$

where I is the input SONAR image and $F_i(\tau, \sigma)$ is the ith Schmid filter shown in Fig. 7.

As seen in Fig. 8, the obtained SEMs captured the small and big structural changes in a macro-scale, i.e. 21×21. To describe the images with the obtained SEMs, we proposed to quantify the frequency of the edge occurrences. To this end, for each pixel of the image, we first sorted the obtained 13 edge magnitudes from Eq. (14) in ascending order. As seen in Fig. 9, this gives us insight into how the structural degradations are locally distributed and how different edge types respond to these degradations. In other words, by sorting the SEMs, we can capture the frequency of edge types, i.e., computed using 13 Schmid Gabor-like filters, in the edge maps with the lowest to highest magnitudes. The frequency of the edge types for each sorted edge map with size $M \times N$ is defined by computing the histogram of the i-th sorted SEMs (HoSSEM) separately as follows:

$$HoSSEM_i(k) = \sum_{m=1}^{M} \sum_{n=1}^{N} \Gamma(SSEM_i(m, n), k), k \in [1, 13]$$

$$\Gamma(val, k) = \begin{cases} \frac{|SSEM_i(m,n)|}{\sum |SSEM_i|}, & argmax(SSEM_i(m, n)) = k \\ 0, & o.w. \end{cases} \tag{15}$$

where, $argmax(SSEM_i(m, n))$ indicates the dominant edge type in the ith SSEM at pixel location (m, n).

For the sample image illustrated in Fig. 4, the HoSSEM of the highest and lowest edge magnitude maps are shown in Fig. 10. As seen, the obtained histograms significantly differ, which demonstrates differences in the impact of the occurred degradation on the lowest and lowest and highest edge magnitude maps.

To form the image's macro-scale features (MaSF), we simply concatenated the HoSSEMs obtained from each SSEM as follows.

$$MaSF = [HoSSEM_1, HoSSEM_2, \ldots, HoSSEM_{13}] \tag{16}$$

Considering each HoSSEM has 13 bins leading to a size of 1×13, the obtained macro-scale feature vector will be a size of 1×169 as there is 13 HoSSEM of size 1×13. Note that similar to the MiSF, MaSF is also normalized to avoid biased inference as follows.

$$MaSF = \left[\sqrt{MaSF_1}, \ldots, \sqrt{MaSF_{236}} \right] \tag{17}$$

The obtained MiSF and MaSF are then concatenated to form the final feature vector. As a result, we obtain a 405-dimensional feature vector to describe the SONAR images using their micro- and macro-scale texture and contour features, where 236 values are for the micro-scale features (i.e., 59 uniform LBP features for each of the approximation and detailed bands leading to a total of 236 (4×59) features) and 169 values are for the macro-scale features (i.e., 13 histogram with 13 bins (13×13 corresponding to the 13 obtained edge maps). The final feature vector is determined as follows.

$$feat = [MiSF, MaSF] \tag{18}$$

Table 2
Comparison of the PLCC, SRCC, and RMSE of the proposed method obtained by four regression models.

Criteria	Linear regression	Ridge regression	AdaBoost	SVR
PLCC ↑	0.7201	0.8409	0.8025	**0.8616**
SRCC ↑	0.7186	0.8351	0.7910	**0.8541**
RMSE ↓	10.9865	7.6166	8.5857s	**7.0503**

3.3. Quality evaluation

Herein we employed machine learning models to learn the optimal mapping between the extracted features and the associated subjective quality scores. To identify the best model, we evaluated four regression techniques, including linear regression, Ridge regression (Hilt and Seegrist, 1977), AdaBoost (Freund and Schapire, 1997; Drucker, 1997), and support vector regression (SVR) (Vapnik et al., 1996), with the LIBSVM package (Chang and Lin, 2011). Among these, SVR achieved the highest correlation with the subjective MOSs, making it the most effective model for our task. A detailed discussion is provided in Section 4.1.

The Ridge regressor extends the linear regressor by adding a penalty term to prevent overfitting. This penalty term, denoted as *alpha*, stabilizes the model's predictions, which was empirically set to 0.1 in our study. As an ensemble learning method, AdaBoost combines multiple weak learners, typically decision trees. SVR, on the other hand, learns to recognize the varying impacts of distortions on the quality-aware features by defining an acceptable error rate. We used the radial basis function (RBF) kernel with default parameters to train the SVR, i.e. *gamma = 1, cost = 128*, and *epsilon = 1*.

Similar to previous studies (Moorthy and Bovik, 2011b; Fang et al., 2018, 2020a; Tolie et al., 2023), we performed a fair comparison by randomly dividing the data into training and testing subsets for 1000 times of iterations. Each iteration involved selecting 80% of the samples with their corresponding subjective quality scores for training, while the remaining 20% were used for testing.

4. Experimental results

The effectiveness of the proposed method is evaluated using the publicly available underwater SONAR IQA dataset known as SONAR image quality dataset (SIQD) (Chen et al., 2017). A total of four advanced BNIQA techniques and six SONAR IQA methods including two full-reference, two reduced-reference, and two no-reference/blind methods are compared using three commonly utilized evaluation metrics: the Pearson's linear correlation coefficient (PLCC), the Spearman's rank-order correlation coefficient (SRCC), and the root mean squared error (RMSE). Comprehensive experiments have demonstrated the superiority of the proposed method over the compared methods. The following subsections provide a detailed analysis of the performance and ablation study.

4.1. Evaluation and comparison of different machine learning models

As mentioned earlier, we have trained the linear regression, Ridge regression, AdaBoost, and SVR on the proposed feature representation extracted from the SIQD dataset to obtain the best regression model for SONAR image quality assessment. The results are reported in Table 2, in which SVR yields the highest performance among all.

4.2. Description of the dataset

We conduct the experiments on the publicly available SIQD (Chen et al., 2017) dataset. SIQD contains 40 reference SONAR images of size 320 × 320 captured by a side-scan sonar. The images in the SIQD

dataset include objects such as swimmers, shipwrecks, underwater creatures, etc. To generate the distorted images, the reference images are compressed using the ComGBR coding (Chen et al., 2016) and SPIHT coding (Said and Pearlman, 1996) algorithms. Overall, 800 distorted images are generated in SIQD by compressing the images at four to six levels to generate compression-related degradations. Moreover, bit errors are also applied to the compressed images to simulate the information loss due to transmission.

In addition to the SIQD dataset, we have also evaluated the performance of the proposed method on a newly published dataset, the super-resolution reconstructed sonar image dataset (SRSID) (Zheng et al., 2022). The SRSID includes 57 reference images, comprising side-scan, acoustic lens, and forward-looking sonar images sourced from the SIQD and Marine Debris (Singh and Valdenegro-Toro, 2021) datasets, along with their corresponding reconstructed versions, totaling 1,026 images. These reconstructed images were generated using six super-resolution algorithms with various scales, and their quality was assessed based on the identification of target objects and the confidence in this identification.

4.3. Evaluation metrics

As suggested in VQEG (2015), we employed three widely used metrics, namely PLCC, SRCC, and RMSE, to assess the accuracy, monotonicity, and consistency of the proposed objective quality assessment method, respectively. PLCC indicates how well the objective scores align with the subjective human ratings (i.e., MOSs). Instead of considering the actual numerical values, SRCC focuses on the relative rankings of the scores. It measures the monotonic relationship between the objective and subjective rankings. Also, RMSE measures the deviation between the objective and subjective scores. In general, the higher PLCC and SRCC values and the lower RMSE value indicate better performance results. Similar to VQEG (2015), Tolie and Faraji (2022), Chen et al. (2018), before reporting the above-mentioned metrics, we applied a non-linear logistic regression between the predicted objective scores and the MOSs to remove the non-linearity caused by the subjective rating process as follows.

$$Q_i = \kappa_1 \left\{ \frac{1}{2} - \frac{1}{1 + \exp\left[\kappa_2(p_i - \kappa_3)\right]} \right\} + \kappa_4 p_i + \kappa_5 \quad (19)$$

where p_i is the assessed quality score of the ith SONAR image computed by the proposed method, Q_i is its corresponding mapped objective score, and κ_{1-5} are the five model parameters in the curve fitting process to minimize the deviation between the objective scores and the MOSs.

4.4. Parameter setting

The proposed method employs Schmid Gabor-like filters to extract macro-scale features (MaSF) from SONAR images. Initially designed at a size of 49 × 49, these filters needed adjustment for optimal SONAR image quality assessment. Given that the available sonar images possess a resolution of 320 × 320, the 49 × 49 filter size may not be suitable. Consequently, we conducted a systematic evaluation to determine the most effective Schmid filter size.

To identify the optimal size, we conducted a series of experiments, varying the Schmid filter size within the range of [3, … , 49]. For each tested filter size, we employed three key evaluation metrics: PLCC, SRCC, and RMSE. The performance of the proposed method was then assessed in terms of these metrics after applying each filter size to generate Schmid Gabor-like filters. Higher PLCC and SRCC values, along with lower RMSE values, indicate stronger correlations between the measured and subjective quality scores.

Fig. 11 presents the median PLCC, SRCC, and RMSE values derived from 1000 runs of the SVR model on the features extracted from SIQD dataset. These features were extracted using the proposed method
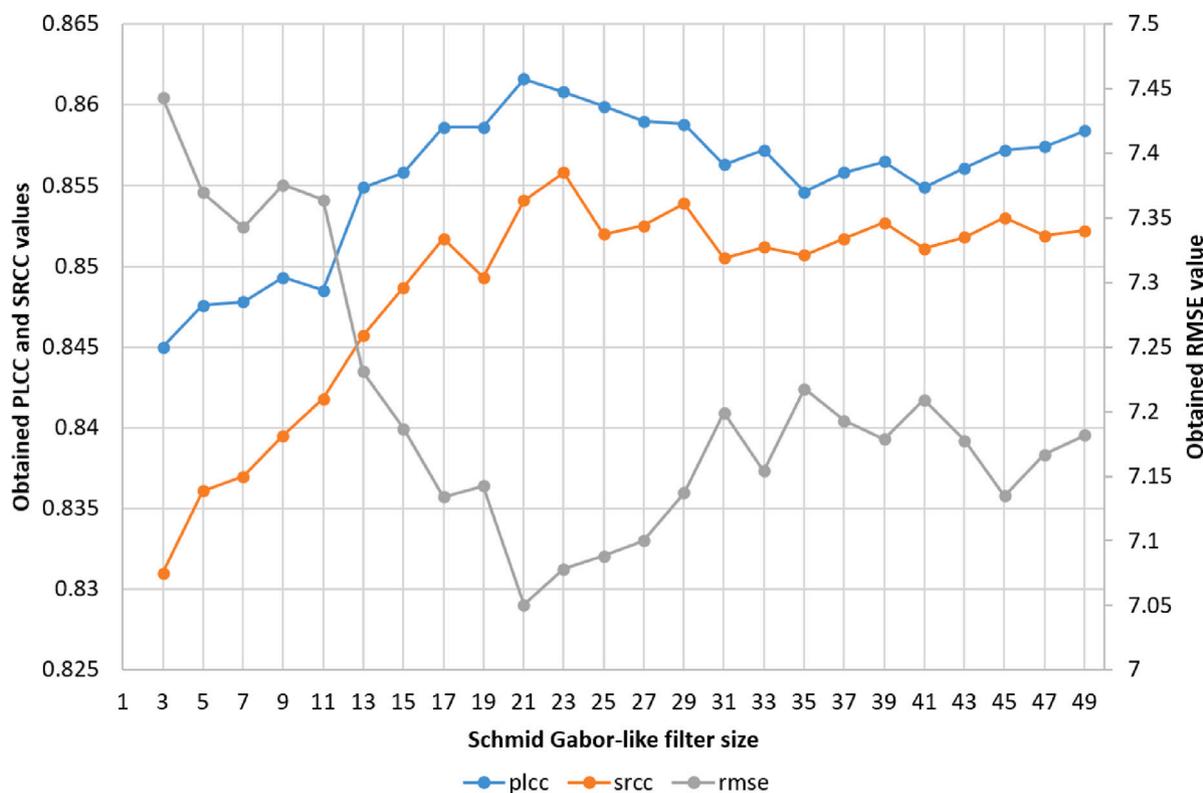
Fig. 11. Scatter plots of the PLCC, SRCC, and RMSE values of our method obtained by applying various size of the Schmid Gabor-like filters.

Table 3
Comparison of the PLCC, SRCC, and RMSE from four BNIQA, eight SONAR IQA, and our proposed methods on the SIQD dataset. The best and second-best results are highlighted in bold and italic, respectively.

Criteria	BNIQA				SONAR IQA methods								Proposed
					FR		RR		NR				
	BLIINDS II	BRISQUE	NIQE	PIQE	SIQP	SOIQE	PSIQP	TPSIQA	NRCDM	DPDNN	PCASS	SRSIQA	
PLCC ↑	0.4459	0.3101	0.2829	0.6161	0.8025	*0.823*	0.7589	0.7689	0.734	0.7763	–	–	**0.8616**
SRCC ↑	0.4410	0.1944	0.3006	0.6079	0.7923	*0.815*	0.7557	0.7560	0.709	0.7872	–	–	**0.8541**
RMSE ↓	12.5170	13.2950	13.4131	11.0148	8.1283	*7.953*	9.1064	8.9417	9.498	8.9497	–	–	**7.0503**

with various tested filter sizes of the Schmid Gabor-like filters. As expected, increasing the filter size from 3 to 21 resulted in elevated PLCC and SRCC values, coupled with a reduction in RMSE. This indicates an enhanced capability in evaluating the quality of SONAR images. This improvement is attributed to the effectiveness of Schmid Gabor-Like filters in representing macro-scale degradations within the image. Consequently, enlarging the filter size helps in capturing higher-level changes, such as edge and contour information degradation.

It is worth noting that further increasing the filter size beyond 21, up to 49, slightly drops the performance of the proposed method. Additionally, larger filter sizes entail greater computational complexity due to increased computation requirements. Thus, while the highest SRCC value was achieved with a filter size of 23, considering the highest PLCC and lower RMSE obtained with a filter size of 21, we opted to set the Schmid Gabor-like filter size at 21. This choice strikes a balance between performance and computational efficiency.

4.5. Performance comparison

To evaluate the performance of the proposed method, we compare the results with four BNIQA, namely, BLIINDS II (Saad et al., 2012), BRISQUE (Mittal et al., 2012), NIQE (Mittal et al., 2013), and PIQE (Venkatanath et al., 2015), and eight SONAR IQA methods including two FR methods, i.e. SIQP (Chen et al., 2019b) and SOIQE (Zhang et al., 2020), two RR methods, i.e. PSIQP (Chen et al., 2018) and

TPSIQA (Chen et al., 2020), and four NR methods, i.e. NRCDM (Chen et al., 2019a), DPDNN (Zhang et al., 2022), PCASS (Chen et al., 2024), and SRSIQA (Feng et al., 2024). The overall performance of each compared method is evaluated based on the median values of the PLCC, SRCC, and RMSE metrics calculated on the testing subsets. Tables 3 and 4 present the comparison results on the SIQD and SRSID datasets, with the best and second-best methods highlighted in bold and italic, respectively. The results for each method were obtained by executing their publicly available source codes, except for SOIQE, DPDNN, PCASS, and SRSIQA, for which the source codes are unavailable. For these methods, therefore, we have reported the results from their respective published papers where available. In both tables, "-" indicates missing results.

From the results in Table 3, it is clear that the proposed method outperforms all compared methods in terms of the PLCC, SRCC, and RMSE metrics. Specifically, the proposed blind/NR method surpasses the second-best method, i.e. SOIQE, by 4.69% and 4.79% in terms of the PLCC and SRCC metrics, respectively. Also, it improves the PLCC and SRCC values of the best NR method, i.e., DPDNN, by 10.98% and 8.49%, respectively. Based on the obtained results, it can be also concluded that the proposed micro- and macro-scale feature-based representation leads to a better quality evaluation compared with the deep learning-based method, i.e. DPDNN.

Moreover, Table 4 reports the performance of the proposed method on the SRSID dataset, highlighting its effectiveness in evaluating the

Table 4
Comparison of the PLCC, SRCC, and RMSE from four BNIQA, eight SONAR IQA, and our proposed methods on the SRSID dataset. The best and second-best results are highlighted in bold and italic, respectively.

Criteria	BNIQA				SONAR IQA methods								Proposed
					FR		RR		NR				
	BLIINDS II	BRISQUE	NIQE	PIQE	SIQP	SOIQE	PSIQP	TPSIQA	NRCDM	DPDNN	PCASS	SRSIQA	
PLCC ↑	0.3451	0.5302	0.5341	0.6469	0.3992	–	0.4784	0.3441	0.6314	–	0.7804	*0.9340*	**0.9415**
SRCC ↑	0.3179	0.5054	0.5345	0.6787	0.3662	–	0.4629	0.3958	0.5753	–	0.7636	*0.9260*	**0.9408**
RMSE ↓	12.5560	11.3430	11.3096	10.0022	12.2658	–	11.7478	12.5608	10.3743	–	8.3648	*4.7052*	**4.4898**

Table 5
Distortion-specific comparison of the PLCC, SRCC, and RMSE from four BNIQA, six SONAR IQA, and our proposed methods on the SIQD dataset. The best and second-best results are highlighted in bold and italic, respectively.

Distortion type	Criteria	BNIQA				SONAR IQA methods						Proposed
						FR		RR		NR		
		BLIINDS II	BRISQUE	NIQE	PIQE	SIQP	SOIQE	PSIQP	TPSIQA	NRCDM	DPDNN	
CC	PLCC ↑	0.1310	0.1887	0.5799	0.6915	*0.7799*	–	0.6868	0.7485	0.6721	–	**0.8518**
	SRCC ↑	0.1011	0.1072	0.5829	0.5834	0.7429	–	0.6859	*0.7486*	0.5788	–	**0.8200**
	RMSE ↓	12.7828	12.6622	10.5048	9.2502	*8.0704*	–	9.3721	8.5500	9.5473	–	**6.6730**
CS	PLCC ↑	0.5832	0.4747	0.5033	0.5092	*0.7781*	–	0.7044	0.7677	0.7297	–	**0.8700**
	SRCC ↑	0.5714	0.4560	0.4955	0.4575	*0.7670*	–	0.6888	0.6900	0.6849	–	**0.8359**
	RMSE ↓	9.2267	9.9971	9.8145	9.5812	*7.1351*	–	8.0617	7.2788	7.7660	–	**5.4588**
TC	PLCC ↑	0.5692	0.5701	0.7968	0.6613	*0.8482*	–	0.7624	0.8387	0.8422	–	**0.8632**
	SRCC ↑	0.5307	0.5673	0.7954	0.6096	**0.8443**	–	0.7477	0.8333	*0.8429*	–	0.8337
	RMSE ↓	12.8572	12.8471	9.4491	11.4924	*8.2820*	–	10.1183	8.5166	8.4319	–	**7.7115**
TS	PLCC ↑	0.5529	0.5527	0.1836	0.5187	0.8190	–	*0.8286*	0.7703	0.7059	–	**0.8869**
	SRCC ↑	0.4971	0.5307	0.1756	0.4579	0.8035	–	*0.8223*	0.7623	0.6898	–	**0.8711**
	RMSE ↓	11.5203	11.5218	13.5906	11.6129	7.9340	–	*7.7400*	8.8170	9.7923	–	**6.2559**

quality of the reconstructed super-resolution sonar images. This evaluation includes utility quality assessment as the subjective scores in SRSID dataset are obtained based on the identification of target objects within the images. The results demonstrate the superiority of the proposed method, validating its capability in utility quality assessment. Specifically, the proposed method outperforms the second-best method, SRSIQA, by significant margins: 0.8% in PLCC, 1.6% in SRCC, and 4.58% in RMSE. These improvements clearly demonstrate the robustness and effectiveness of our proposed method in both perceptual and utility quality evaluations of SONAR images.

4.6. Distortion-specific comparison

In addition to the overall performance comparison, we conducted experiments on the SIQD dataset to analyze the performance of compared methods on the images under individual distortion types. Noted that the SIQD contains four types of distortions, namely, (1) compression based on ComGBR coding (CC); (2) compression based on SPIHT coding (CS); (3) transmission based on ComGBR coding (TC); and (4) transmission based on SPIHT coding (TS). Table 5 reports the PLCC, SRCC, and RMSE values of the compared methods on aforementioned distortions in the SIQD dataset. The results of each method are reported by running their publicly available source codes. However, the SOIQE and DPDNN methods do not provide their source codes and did not report distortion-specific comparisons in their respective published papers. Therefore, we used the "-" to reflect these missing results. From the results in Table 5,W the superiority of the proposed method in the quality assessment of the SONAR images under various types of distortions can be clearly seen. This is because the proposed micro- and macro-scale features well represent the impact of each distortion type and they are correlated with the subjective ratings.

4.7. Visual comparison

In addition to the quantitative comparisons, we have illustrated the scatter plots of the MOSs versus predicted quality scores by each

method in Fig. 12(a)–(i) to visually compare the performance of our method and the state-of-the-arts. In each plot, each sample point represents a single test SONAR image, showing its MOS (Mean Opinion Score) versus its predicted quality score. Each sample point is assigned a distinct color and symbol based on whether it represents a reference image, distortion-free, or a distorted one, depending on the type of distortion. This approach allows us to examine the correlation between the predictions and GT values of MOS specifically for images with the same type of distortion. Overall, as seen in Fig. 12, the quality scores assessed by the proposed method have the highest correlation with the subjective MOSs compared with other methods, because most of the data points are gathered around the bisector line and spread along the horizontal axis. Also, as shown in Fig. 12(i), it is apparent that the predicted scores of the images under the TS distortions have the highest correlation, which further validates the results reported in Table 5. Moreover, as the data points representing the same type of distortion are well spread along the horizontal axis, the robustness of the proposed method across the reference images and four distortion types is clearly verified.

Furthermore, Fig. 13 presents SONAR images subjected to four types of distortions (i.e. CC, CS, TC, and TS) categorized into two quality ranks of low and high. Each image is accompanied by its predicted quality score and corresponding MOS. The selected images include both side-scan and forward-looking SONAR images, featuring various object types and noise levels, thereby demonstrating the model's generalization ability across different scenarios. Generally, as MOS values increase, the predicted quality scores also rise, and vice versa. Additionally, it is evident that images with higher levels of distortion consistently receive lower MOS and predicted quality scores, highlighting the models' proficiency in ranking images based on the distortion severity.

4.8. Statistical significance test

For statistical significance analysis of the proposed and the compared BNIQA, i.e. BLIINDS II, BRISQUE, NIQE, and PIQE, and the
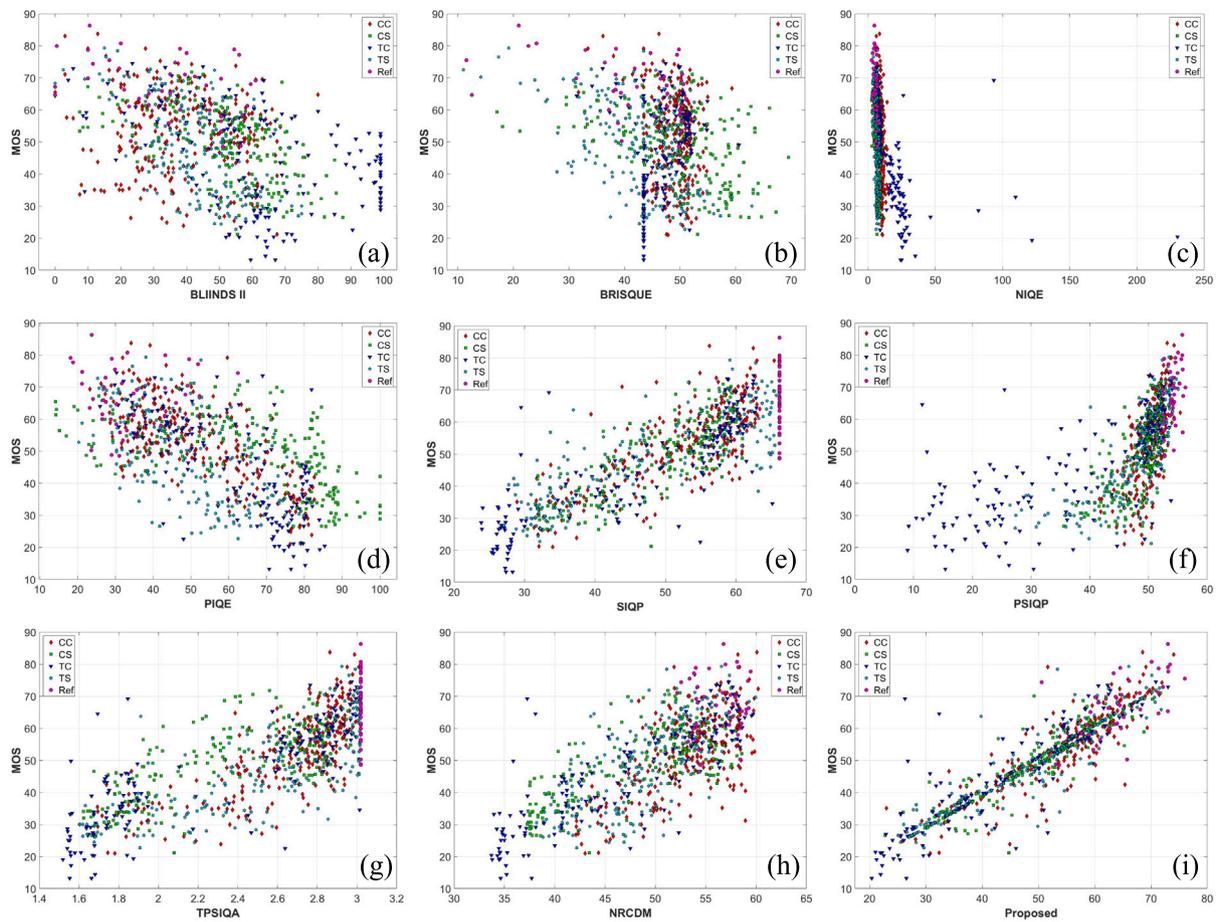
Fig. 12. Scatter plots of the MOSs versus the proposed, BNIQA, and SONAR IQA methods' predicted quality scores. Each data point denotes one test image. (a) BLIINDS II; (b) BRISQUE; (c) NIQE; (d) PIQE; (e) SIQP; (f) PSIQP; (g) TPSIQA; (h) NRCDM; and (i) Proposed.
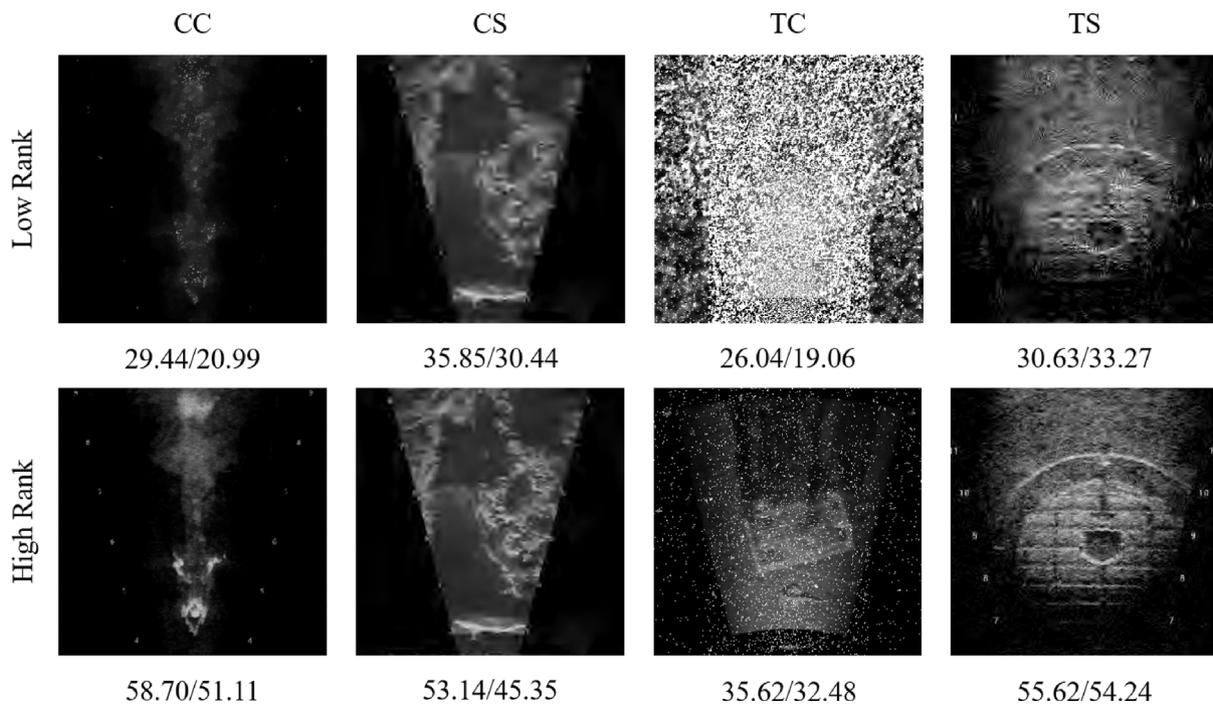
Fig. 13. SONAR images under different distortion types (CC, CS, TC, TS) and quality ranks (low and high), showing predicted quality scores and MOS values.

Table 6
Statistical significance comparison of the proposed, BNIQA, and SONAR IQA methods with the F-Test on the SIQD dataset.

	BLIINDS II	BRISQUE	NIQE	PIQE	SIQP	PSIQP	TPSIQA	NRCDM	Proposed
BLIIINDS II	-	green	gray	gray	red	red	red	red	red
BRISQUE	green	-	gray	red	red	red	red	red	red
NIQE	gray	gray	-	gray	red	red	red	red	red
PIQE	gray	green	gray	-	red	red	red	red	red
SIQP	green	green	green	green	-	gray	gray	green	red
PSIQP	green	green	green	green	red	-	red	gray	red
TPSIQA	green	green	green	green	green	red	-	green	red
NRCDM	green	green	green	green	red	gray	red	-	red
Proposed	green	green	green	green	green	green	green	green	-

Table 7
Ablation study of the proposed feature representation on the SIQD dataset.

Method	PLCC	SRCC	RMSE
Only MiSF$_{cA}$	0.8015	0.7849	8.3216
Only MiSF$_{cH}$	0.6873	0.6306	10.4966
Only MiSF$_{cV}$	0.6210	0.6055	10.9102
Only MiSF$_{cD}$	0.6089	0.5716	11.0395
All MiSF	0.8422	0.8284	7.5374
MaSF	0.8076	0.8009	8.1854
MiSF$_{cA}$ + MaSF	0.8487	0.8450	7.3186
MiSF$_{cH}$ + MaSF	0.8473	0.8382	7.4142
MiSF$_{cV}$ + MaSF	0.8415	0.8343	7.5171
MiSF$_{cD}$ + MaSF	0.8420	0.8317	7.5335
Proposed (MiSF$_{gray}$ + MaSF$_{gray}$)	0.8475	0.8413	7.4203
Proposed (All MiSF + MaSF)	0.8616	0.8541	7.0503

Table 8
Computational run-time comparison of the proposed and compared methods on the SIQD dataset.

IQA type	Method	Run-time (s)
BNIQA	BLIINDS II	5.4727
	BRISQUE	0.0095
	NIQE	0.0124
	PIQE	0.0119
SONAR IQA	SIQP	0.0668
	PSIQP	0.0264
	TPSIQA	0.0243
	NRCDM	0.0449
	Proposed	0.0523

SONAR IQA, i.e. SIQP, PSIQP, TPSIQA, NRCDM, which made their source codes publicly available, we have used the widely-utilized F-test. Similar to the Xue et al. (2013), Chen et al. (2019a), the ratio between the residual variances of the compared methods is computed and used for analysis as follows. The computed ratio greater than a confidence level, i.e. 95%, shows there is a significant difference between the two compared methods. The results are tabulated in Table 6, where an element with the *green* color means the method in the row is significantly better than the method in the column. The *red* elements indicate that the method in the column is better than the method in the row. And the *gray* elements show there is no significant difference between the compared methods. According to the results, the proposed method performs significantly better than all other compared methods.

4.9. Ablation study

To analyze the contribution of each component of the proposed feature representation, ablation studies were conducted on the SIQD dataset. We separately trained the SVR model using each of the micro- and macro-scale features and reported the results in Table 7. For the micro-scale features, we also studied the effectiveness of features extracted from the approximation (cA) and detailed (cH, cV, cD) bands of the SONAR images. The results indicate that the micro-scale feature set contributes the most to overall performance, showing a high correlation of the ULBP features with subjective ratings.

Additionally, micro-scale features from the approximation band correlate better with subjective ratings than those from the detailed bands. However, while the detailed bands' micro-scale features have a lower correlation, combining them with the approximation band's micro-scale features increases the PLCC and SRCC values from 0.8015 and 0.7849 to 0.8422 and 0.8284, respectively. To assess the impact of the macro-scale features, we combined them with each individual micro-scale feature extracted from the wavelet bands. The results show that this combination consistently improves performance, even though the micro-scale features from the detailed bands perform worse than those from the approximation band. Moreover, combining the macro-scale features with all micro-scale features increases performance from 0.8422 to 0.8616 (PLCC) and from 0.8284 to 0.8541 (SRCC), representing improvements of 2.30% and 3.10%, respectively.

To examine the effectiveness of feature extraction in the wavelet domain, we also reported the performance of extracting both micro- and macro-scale features from the original grayscale image. The results in Table 7 show that the performance is close to when features are extracted from the approximation channel, which is 1.6% and 1.5% lower than in the wavelet domain in terms of PLCC and SRCC, suggesting that converting the image to the wavelet domain further enhances the effectiveness of feature extraction in this context.

Moreover, to verify the effectiveness of the proposed micro- and macro-scale feature representation in distinguishing the SONAR image-related distortions, inspired by Fang et al. (2020a), Tolie et al. (2023), we perform the t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten and Hinton, 2008) on the extracted features. The t-SNE is a dimension-reduction technique used for data visualization. It maps the high-dimensional data to two or three dimensions by keeping the similar samples closer and the dissimilar ones in distant. Therefore, it is used to reveal the patterns in data, e.g. clusters. In this study, we used the t-SNE method to map the extracted 405-dimensional feature vectors into 2-dimensions. The same technique is also applied to the NRCDM method to visualize their proposed feature representation.

Fig. 14. (a) and (b) illustrate the scatter plots obtained by applying the t-SNE method on both the NRCDM and the proposed methods, respectively. The samples in each figure are colored using their distortion type. As shown in Fig. 14. (b), the images distorted using the SPIHT coding, i.e. CS and TS, are well separated from the images distorted by the ComGBR coding, i.e. CC and TC. Also, based on the results reported in Table 5, the proposed method achieves the best results for the images under CS and TS, which further confirms the distinguishability of the proposed feature representation.

4.10. Computational run-time comparison

Apart from ensuring accuracy, consistency, and monotonicity, IQA methods are also required to possess reasonable computational complexity. The average execution time of the proposed and compared methods is presented in Table 8 on the SIQD dataset. To ensure a fair evaluation, the experiments were conducted on a computer equipped with an Intel(R) Core(TM) i9-10885H CPU @ 2.40 GHz and 32 GB memory, utilizing MATLAB R2022a. The results in Table 8 have indicated that the proposed method demonstrates a reasonable run-time performance of 0.0523 s, allowing it to process nearly 20 frames per second with an average size of 320 × 320 pixels.
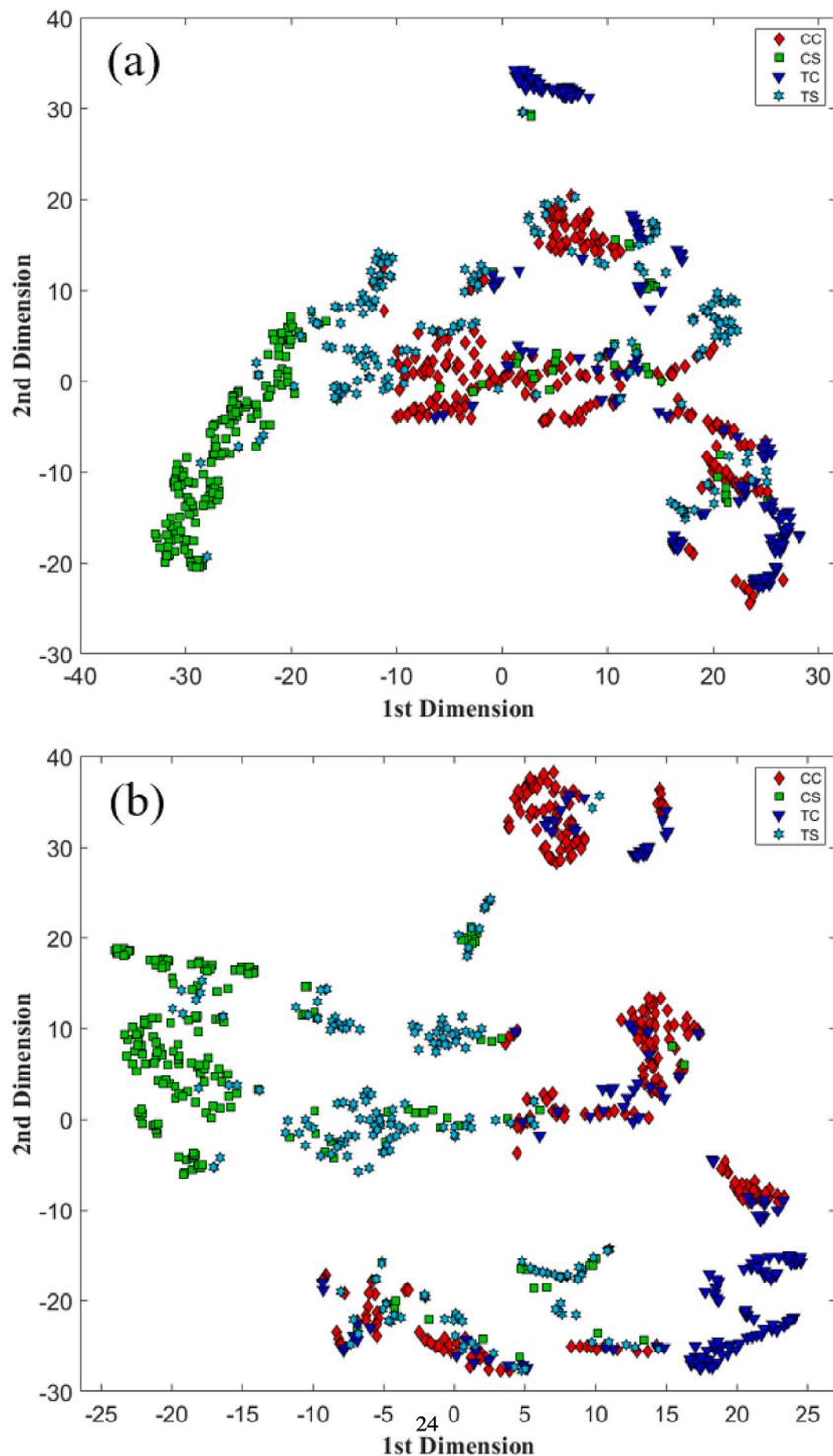
Fig. 14. t-SNE scatter plots of the NRCDM (a) and proposed (b) feature representation on the SIQD dataset. Data points are colored by their distortion types, i.e. CC, CD, TC, and TS.

5. Conclusion

In conclusion, we have introduced a novel blind image quality assessment method that quantifies both the perceptual and utility quality indices of SONAR images. Our approach decomposes the SONAR image into the wavelet bands of the approximation and details to measure the amount of degradation in both the low- and high-frequency components. It is found that micro-scale ULBP texture features are particularly useful for measuring information loss in these frequency components, highlighting the emphasis on the perceptual quality aspect. Addition-

ally, our new edge-based macro-scale contour information descriptor is proven effective for evaluating the utility quality, which is crucial for following-on target and object identification. By integrating these features and training a support vector regression model, our method has demonstrated superior performance. Extensive experiments on two publicly available datasets, i.e. SIQD and SRSID, have fully validated the effectiveness of our method, as evidenced by superior performance in SRCC, PLCC, and RMSE indices when it is applied to distorted and reconstructed SONAR images via super-resolution algorithms, compared to a few state-of-the-art SIQA methods.

Table A.9
The list of key abbreviations and their full names.

Abbreviation	Full name	Abbreviation	Full name
IQA	image quality assessment	SONAR	sound navigation and ranging
FR	full-reference	RR	reduced-reference
NR	no-reference	SVR	support vector regression
ULBP	uniform local binary pattern	SEM	Schmid Gabor-like edge maps
HoSSEM	histogram of the ith sorted SEM	MiSF	micro-scale features
MaSF	macro-scale features		

In future work, we could explore leveraging deep learning-based pixel differencing modules (Li et al., 2024) to enhance the quantification of micro- and macro-scale features. Additionally, integrating probabilistic mixture modeling could provide a more effective framework for understanding the underlying distributions of these features. Furthermore, employing attention mechanisms (Hu et al., 2018; Li et al., 2023) to adaptively fuse these features may lead to more robust and generalized quality assessment procedures.

CRediT authorship contribution statement

Hamidreza Farhadi Tolie: Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jinchang Ren:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Rongjun Chen:** Writing – review & editing, Methodology, Investigation, Formal analysis. **Huimin Zhao:** Writing – review & editing, Resources, Investigation, Conceptualization. **Eyad Elyan:** Writing – review & editing, Visualization, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix. Abbreviations

See Table A.9.

Data availability

Data will be made available on request.

References

Callow, H.J., 2003. Signal Processing for Synthetic Aperture Sonar Image Enhancement. University of Canterbury. Electrical and Electronic Engineering.

Chang, C.C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2 (3), 1–27.

Chen, W., Cai, B., Zheng, S., Zhao, T., Gu, K., 2024. Perception-and-cognition-inspired quality assessment for sonar image super-resolution. IEEE Trans. Multimed..

Chen, W., Gu, K., Lin, W., Xia, Z., Le Callet, P., Cheng, E., 2019a. Reference-free quality assessment of sonar images via contour degradation measurement. IEEE Trans. Image Process. 28 (11), 5336–5351.

Chen, W., Gu, K., Lin, W., Yuan, F., Cheng, E., 2019b. Statistical and structural information backed full-reference quality measure of compressed sonar images. IEEE Trans. Circuits Syst. Video Technol. 30 (2), 334–348.

Chen, W., Gu, K., Min, X., Yuan, F., Cheng, E., Zhang, W., 2018. Partial-reference sonar image quality assessment for underwater transmission. IEEE Trans. Aerosp. Electron. Syst. 54 (6), 2776–2787.

Chen, W., Gu, K., Zhao, T., Jiang, G., Le Callet, P., 2020. Semi-reference sonar image quality assessment based on task and visual perception. IEEE Trans. Multimed. 23, 1008–1020.

Chen, F., Luo, Z., Xu, Y., Ke, D., 2019c. Complementary fusion of multi-features and multi-modalities in sentiment analysis. arXiv preprint arXiv:1904.08138.

Chen, W., Yuan, F., Cheng, E., 2016. Adaptive underwater image compression with high robust based on compressed sensing. In: Proceedings of the IEEE Int. Conf. Signal Process. Commun. Comput.. ICSPCC, IEEE, pp. 1–6.

Chen, W., Yuan, F., Cheng, E., Lin, W., 2017. Subjective and objective quality evaluation of sonar images for underwater acoustic transmission. In: Proceedings of the IEEE Int. Conf. Image Process.. ICIP, IEEE, pp. 176–180.

Daugman, J.G., 1983. Six formal properties of two-dimensional anisotropie visual filters: Structural principles and frequency/orientation selectivity. IEEE Trans. Syst. Man Cybern. (5), 882–887.

Drucker, H., 1997. Improving regressors using boosting techniques. In: Icml, vol. 97, Citeseer, pp. 107–115.

Fang, Y., Du, R., Zuo, Y., Wen, W., Li, L., 2020a. Perceptual quality assessment for screen content images by spatial continuity. IEEE Trans. Circuits Syst. Video Technol. 30, 4050–4063.

Fang, Y., Yan, J., Du, R., Zuo, Y., Wen, W., Zeng, Y., Li, L., 2020b. Blind quality assessment for tone-mapped images by analysis of gradient and chromatic statistics. IEEE Trans. Multimed. 23, 955–966.

Fang, Y., Yan, J., Li, L., Wu, J., Lin, W., 2018. No reference quality assessment for screen content images with both local and global feature representation. IEEE Trans. Image Process. 27 (4), 1600–1610.

Feng, Q., Zheng, S., Zhang, K., Wei, H., 2024. A brain-inspired quality assessment model for sonar image super-resolution. Displays 82, 102620.

Field, D.J., 1987. Relations between the statistics of natural images and the response properties of cortical cells. Josa A 4 (12), 2379–2394.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55 (1), 119–139.

Gabor, D., 1946. Theory of communication. Part 1: The analysis of information. J. Inst. Electr. Eng.-Part III: Radio Commun. Eng. 93 (26), 429–441.

Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 3354–3361.

Gonzales, R.C., Wintz, P., 1987. Digital Image Processing. Addison-Wesley Longman Publishing Co. Inc..

Gu, K., Zhou, J., Qiao, J.-F., Zhai, G., Lin, W., Bovik, A.C., 2017. No-reference quality assessment of screen content pictures. IEEE Trans. Image Process. 26 (8), 4005–4018.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit.. pp. 770–778.

Hilt, D.E., Seegrist, D.W., 1977. Ridge, a Computer Program for Calculating Ridge Regression Estimates. Department of Agriculture, Forest Service, Northeastern Forest Experiment ….

Hou, X., Zhang, L., 2007. Saliency detection: A spectral residual approach. In: Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit.. Ieee, pp. 1–8.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7132–7141.

Hurley, N., Rickard, S., 2009. Comparing measures of sparsity. IEEE Trans. Inform. Theory 55 (10), 4723–4741.

Jaffe, J.S., 2014. Underwater optical imaging: the past, the present, and the prospects. IEEE J. Ocean. Eng. 40 (3), 683–700.

Kalman, D., 1996. A singularly valuable decomposition: the SVD of a matrix. College Math. J. 27 (1), 2–23.

Kusumoto, R., Han, X., Chen, Y.-W., 2014. Hybrid aggregation of sparse coded descriptors for food recognition. In: Proceedings of the 22nd Int. Conf. Pattern Recog.. IEEE, pp. 1490–1495.

Leontiev, N., Nyurova, A., 2019. The use of discrete meyer wavelet for speech segmentation. In: Proceedings of the Int. Multi-Con. Ind. Eng. Mod. Tech.. FarEastCon, IEEE, pp. 1–3.

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., Tao, D., 2020. An underwater image enhancement benchmark dataset and beyond. IEEE Trans. Image Process. 29, 4376–4389. http://dx.doi.org/10.1109/TIP.2019.2955241.

Li, Y., Ren, J., Yan, Y., Liu, Q., Ma, P., Petrovski, A., Sun, H., 2023. CBANet: An end-to-end cross-band 2-D attention network for hyperspectral change detection in remote sensing. IEEE Trans. Geosci. Remote Sens. 61, 1–11. http://dx.doi.org/10.1109/TGRS.2023.3276589.

Li, Y., Ren, J., Yan, Y., Ma, P., Assaad, M., Gao, Z., 2024. ABBD: Accumulated band-wise binary distancing for unsupervised parameter-free hyperspectral change detection. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 17, 9880–9893. http://dx.doi.org/10.1109/JSTARS.2024.3407212.

Liu, Y., Bao, Y., 2022. Review on automated condition assessment of pipelines with machine learning. Adv. Eng. Inform. 53, 101687.

Liu, Y., Bao, Y., 2023. Intelligent monitoring of spatially-distributed cracks using distributed fiber optic sensors assisted by deep learning. Measurement 220, 113418.

Luo, Z., Xu, H., Chen, F., 2019. Audio sentiment analysis by heterogeneous signal features learned from utterance-based parallel neural network. In: AffCon@ AAAI. Shanghai, China, pp. 80–87.

Marĉelja, S., 1980. Mathematical description of the responses of simple cortical cells. JOSA 70 (11), 1297–1300.

Mittal, A., Moorthy, A.K., Bovik, A.C., 2012. No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. 21 (12), 4695–4708.

Mittal, A., Soundararajan, R., Bovik, A.C., 2013. Making a "completely blind" image quality analyzer. IEEE Signal Process. Lett. 20 (3), 209–212.

Moorthy, A.K., Bovik, A.C., 2011a. Blind image quality assessment: From natural scene statistics to perceptual quality. IEEE Trans. Image Process. 20 (12), 3350–3364. http://dx.doi.org/10.1109/TIP.2011.2147325.

Moorthy, A.K., Bovik, A.C., 2011b. Blind image quality assessment: From natural scene statistics to perceptual quality. IEEE Trans. Image Process. 20 (12), 3350–3364.

Ni, Z., Ma, L., Zeng, H., Chen, J., Cai, C., Ma, K.-K., 2017. ESIM: Edge similarity for screen content image quality assessment. IEEE Trans. Image Process. 26 (10), 4818–4831.

Ni, Z., Zeng, H., Ma, L., Hou, J., Chen, J., Ma, K.-K., 2018. A gabor feature-based quality assessment model for the screen content images. IEEE Trans. Image Process. 27 (9), 4516–4528.

Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24 (7), 971–987.

Panetta, K., Gao, C., Agaian, S., 2016. Human-visual-system-inspired underwater image quality measures. IEEE J. Ocean. Eng. 41 (3), 541–551. http://dx.doi.org/10.1109/JOE.2015.2469915.

Rezaie, F., Helfroush, M.S., Danyali, H., 2018. No-reference image quality assessment using local binary pattern in the wavelet domain. Multimedia Tools Appl. 77, 2529–2541.

Saad, M.A., Bovik, A.C., Charrier, C., 2012. Blind image quality assessment: A natural scene statistics approach in the DCT domain. IEEE Trans. Image Process. 21 (8), 3339–3352.

Saha, S., Vemuri, R., 2000. An analysis on the effect of image activity on lossy coding performance. In: Proceedings of the IEEE Int. Symp. Circuits Syst.. ISCAS, vol. 3, IEEE, pp. 295–298.

Said, A., Pearlman, W.A., 1996. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. IEEE Trans. Circuits Syst. Video Technol. 6 (3), 243–250.

Schmid, C., 2001. Constructing models for content-based image retrieval. In: Proceedings of the IEEE CS Conf. Comput. Vis. Pattern Recog. CVPR 2001, vol. 2, IEEE, II–II.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Singh, D., Valdenegro-Toro, M., 2021. The marine debris dataset for forward-looking sonar semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3741–3749.

Stojanovic, M., Freitag, L., 2013. Recent trends in underwater acoustic communications. Mar. Technol. Soc. J 47 (5), 45–50.

Tolie, H.F., Faraji, M.R., 2022. Screen content image quality assessment using distortion-based directional edge and gradient similarity maps. Signal Process., Image Commun. 101, 116562. http://dx.doi.org/10.1016/j.image.2021.116562.

Tolie, H.F., Faraji, M.R., Qi, X., 2023. Blind quality assessment of screen content images via edge histogram descriptor and statistical moments. Vis. Comput. 1–16.

Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9 (11).

Vapnik, V., Golowich, S., Smola, A., 1996. Support vector method for function approximation, regression estimation and signal processing. In: Advances in Neural Information Processing Systems, vol. 9.

Venkatanath, N., Praneeth, D., Bh, M.C., Channappayya, S.S., Medasani, S.S., 2015. Blind image quality evaluation using perception based features. In: Proceedings of the 21st Natl. Conf. Commun.. NCC, IEEE, pp. 1–6.

VQEG, 2015. Final report from the video quality experts group on the validation of objective models of video quality assessment. http://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx.

Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612. http://dx.doi.org/10.1109/TIP.2003.819861.

Wang, Y., Li, N., Li, Z., Gu, Z., Zheng, H., Zheng, B., Sun, M., 2018. An imaging-inspired no-reference underwater color image quality assessment metric. Comput. Electr. Eng. 70, 904–913. http://dx.doi.org/10.1016/j.compeleceng.2017.12.006.

Wang, Z., Simoncelli, E.P., 2005. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In: Proceedings of the Human Vision and Electronic Imaging X, vol. 5666, SPIE, pp. 149–159.

Xie, K., Yang, J., Qiu, K., 2022. A dataset with multibeam forward-looking sonar for underwater object detection. Sci. Data 9 (1), 739.

Xue, W., Mou, X., Zhang, L., Bovik, A.C., Feng, X., 2014. Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features. IEEE Trans. Image Process. 23 (11), 4850–4862.

Xue, W., Zhang, L., Mou, X., Bovik, A.C., 2013. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. IEEE Trans. Image Process. 23 (2), 684–695.

Yang, M., Sowmya, A., 2015. An underwater color image quality evaluation metric. IEEE Trans. Image Process. 24 (12), 6062–6071. http://dx.doi.org/10.1109/TIP.2015.2491020.

Yu, M., Tang, Z., Zhang, X., Zhong, B., Zhang, X., 2022. Perceptual hashing with complementary color wavelet transform and compressed sensing for reduced-reference image quality assessment. IEEE Trans. Circuits Syst. Video Technol. 32 (11), 7559–7574.

Yue, G., Hou, C., Zhou, T., 2019. Blind quality assessment of tone-mapped images considering colorfulness, naturalness, and structure. IEEE Trans. Ind. Electron. 66 (5), 3784–3793.

Zhang, H., Li, D., Li, S., Li, Y., 2020. Using multiscale structural fusion for sonar image quality evaluation. In: Proceedings of the 5th Int. Conf. Mech. Control Comput. Eng.. ICMCCE, IEEE, pp. 2331–2335.

Zhang, H., Li, S., Li, D., Wang, Z., Zhou, Q., You, Q., 2022. Sonar image quality evaluation using deep neural network. IET Image Process. 16 (4), 992–999.

Zhang, Z., Sun, W., Min, X., Zhu, W., Wang, T., Lu, W., Zhai, G., 2021. A no-reference evaluation metric for low-light image enhancement. In: Proceedings of the IEEE Int. Conf. Multimedia Expo. ICME, IEEE, pp. 1–6.

Zhang, M., Xie, J., Zhou, X., Fujita, H., 2013. No reference image quality assessment based on local binary pattern statistics. In: Proceedings of the Vis. Commun. Image Process.. VCIP, IEEE, pp. 1–6.

Zheng, S., Chen, W., Zhao, T., Wei, H., Lin, L., 2022. Utility-oriented quality assessment of sonar image super-resolution. In: OCEANS 2022, Hampton Roads. IEEE, pp. 1–5.