Sonar image quality evaluation using deep neural network

Huiqing Zhang1,2,3 | Shuo Li1,2,3 | Donghao Li1,2,3 | Zichen Wang1 | Qixiang Zhou1 | Qixin You1

1 Faculty of Information Technology, Beijing University of Technology, Beijing, China
2 Key Laboratory of Artificial Intelligence, Ministry of Education, Shanghai, China
3 Engineering Research Center of Digital Community, Ministry of Education, Beijing, China

Correspondence
Shuo Li, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China.
Email: shuoli1025@163.com

Funding information
the Major Science and Technology Program for Water Pollution Control and Treatment of China, Grant/Award Number: 2018ZX07111005

Abstract
Sonar technology plays an important role in the development of marine resources and military strategy. Due to the bad quality of underwater acoustics channels, the sonar images collected by sonar technology equipment are easily affected by various kinds of distortions. To obtain high-quality sonar images, the authors devise a novel dual-path deep neural network (DPDNN) to measure the quality of sonar images. In these two paths, the authors use a batch normalization layer to reduce the training time and use the skip operation to speed up the feature extraction. Based on the above two operations, the authors extract the microscopic and macroscopic structures of sonar images, respectively. Finally, a global average pooling layer and a fully connection layer are used to connect the above two paths. Experiments show that the authors’ DPDNN achieves significant improvements in prediction performance and efficiency. The source code will be published in the near future.

1 | INTRODUCTION

With the development of marine resources in various countries, side scan sonar technology is becoming increasingly popular among researchers. Researchers also have established higher requirements for the quality of sonar images. It should be noted that the internal space of underwater navigation equipment is small (as shown in Figure 1), and so it is impossible to carry a large amount of computer equipment. Therefore, the collected sonar images need to be transmitted to ground equipment through an underwater acoustic channel for further analysis and processing. Due to the complexity of the underwater acoustic channel environment, sonar images are often damaged by various kinds of distortion, and this results in degraded image quality [1, 2]. However, low-quality sonar images cannot meet the research requirements, and so it is necessary to design an effective sonar image quality evaluation (IQE) technology to guide the transmission process of sonar image. Moreover, IQE plays an important role in estimating image degradation and optimizing the image compression parameters. According to the amount of information used in reference images, IQE can be divided into three categories: full-reference IQE, reduced-reference IQE and no-reference IQE.

The earliest application objects of IQE were natural scene images [3, 4] and screen content images [5, 6], and the corresponding research is relatively mature. However, there are few studies on sonar IQE. The difference between sonar images and common natural scene images and screen content images is that the quality of the latter two types of images can be judged according to the observer's own perception. Generally, natural scene images are mostly composed of thick lines, rich colours and complex textural content, while screen content images are mostly composed of thin lines, limited colours and simple shapes. Sonar images are mostly greyscale images, in which the pixel value changes infrequently, the contrast is low and the details and content are few; furthermore, they are mostly used for mapping and detection rather than aesthetics, and so the structural information is more of a concern for the observer. Therefore, when evaluating sonar image quality, it is better to use certain professional knowledge rather than use normal aesthetic judgement. The characteristics of the above three types of images are obviously different, and so the evaluation methods are also different.

In recent years, sonar IQE has been widely studied and developed. Chen et al. [7] propose a full-reference sonar image quality predictor based on the similarity of the local entropy...
map and edge map from the perspective of statistical information and structural information. Chen et al. [8] propose a three-stage part-reference sonar image quality predictor based on the definition of the target in the sonar image, the amount of information that can be extracted from the sonar image and the degree of comfort degradation of the sonar image. Chen et al. [9] propose a free-reference sonar image quality index to evaluate image quality by measuring the degree of degradation of the contour after the distorted image is filtered. In addition, some machine learning methods [10–12] are also widely used in the field of IQE.

However, the performance of the above-mentioned sonar IQE methods based on traditional image analysis technology is not ideal [2, 7–9]. Therefore, in this work, we devise a novel dual-path deep neural network (DPDNN) to better evaluate the quality of sonar images, which is inspired by the multi-path neural network [13]. Specifically, in the first path, we connect the convolutional layer and max-pooling layer in turn, and selectively add batch normalization (BN) layers to reduce the training time. In the second path, in addition to the convolutional layers, max-pooling layers and BN layers, we add a skip connection operation to speed up the feature extraction. Finally, the global average pooling (GAP) layer and the fully connected (FC) layer are used to connect the above two paths to get the evaluation quality score of a sonar image. Compared with the traditional sonar IQE method, the DPDNN model achieves significant improvements in the accuracy, monotonicity and consistency of prediction.

The contributions of this work are summarized as follows: (1) Aiming to address the shortcomings of traditional image analysis techniques [2, 7–9], we applied deep learning technology to sonar IQE and achieved improved results; (2) We designed a novel DPDNN with very few parameters and high execution efficiency; (3) Compared with the state-of-the-art no-reference contour degradation measurement (NRCDM) [9] model, the performance of our DPDNN model is improved by 5.76% on the Pearson linear correlation coefficient (PLCC), 11.01% on the Spearman rank order correlation coefficient (SROCC), 15.90% on the Kendall's rank order correlation coefficient (KROCC), 5.77% on the root mean square error (RMSE) and 8.41% on the mean absolute error (MAE), and the execution efficiency is increased by 15 times.

The rest of this paper is arranged as follows. In Section 2, we introduce the proposed DPDNN model and network training in detail. Section 3 provides the results of the experiments on the new sonar image quality database (SIQD) [14] to demonstrate the execution efficiency and evaluation performance of the DPDNN model. Finally, we summarize this paper in Section 4.

2 PROPOSED DPDNN METHOD

With the development of sonar technology, there has been an increase in technologies used to evaluate sonar image quality [2, 7–9], but these traditional image analysis technologies still cannot achieve satisfactory results. Therefore, we will elaborate the DPDNN model specially designed for sonar IQE in this part, and the structure of our DPDNN model is shown in Figure 2.
In the second section, we first introduce the structure of the DPDNN network, including the establishment of the two paths and how to combine the two paths to evaluate the quality of sonar images. Then, we explain the specific content of the network training.

2.1 | DPDNN structure

First, the establishment of our first path is inspired by the classic VGG-Net network structure [15]. VGG-Net only uses convolutional layers and max-poolings layer for feature extraction, which increases the training time. Therefore, we selectively add a BN layer to each of the last six convolutional layers to reduce the training time of the network and prevent parameter overfitting. Specifically, we connect eight convolution layers, six BN layers and three max-pooling layers to construct a sequence network for extracting the microscopic features of sonar images. Note that the BN layer is not added to the first two convolution layers for better feature protection. The standardization process of each feature map $F_i$ is defined as follows:

$$F_i = (F_i - \bar{F}_i) / (\sigma_i^2 + \xi)^{0.5},$$

where $\bar{F}_i = \frac{1}{N} \sum_{j=0}^{N} F_{ij}$ and $\sigma_i^2 = \frac{1}{N} \sum_{j=0}^{N} (F_{ij} - \bar{F}_i)^2$ are the mean and variance of the meta-batch, respectively; $N$ is the size of the meta-batch; $F_{ij}$ represents the $i$th feature of the $j$th sample in the meta-batch; and $\xi$ is a fitting parameter to improve the stability of the network. However, the representation ability of features may decrease with normalization. Therefore, we introduce two parameters $\alpha$ and $\beta$ to transform the normalization feature in the BN layer to solve this problem:

$$\tilde{F}_i = \alpha F_i + \beta,$$

where $\tilde{F}_i$ is the standardized feature of the transformation, and $\alpha$ and $\beta$ are the optimal values obtained through network training. After going through the above network structure, we express the output $F_{11}$ of the first path as follows:

$$F_{11} = \omega_{11} \ast F_{10} + b_{11},$$

where $\ast$ is the convolution operator; $\omega_{11}$ and $b_{11}$ represent the weight matrix and offset matrix of the 17th layer, respectively; and $F_{10}$ is the feature map of the 16th layer.

Secondly, the construction of our second path benefits from the skip connection operation in Res-Net [16]. Its introduction can avoid the disappearing gradient problem and improve the feature propagation speed. Therefore, we introduce the skip connection operation on the basis of the first path. Specifically, we connect 12 convolution layers, 3 max-pooling layers and 9 BN layers to form a sequence network for extracting macroscopic features. Then, the convolution kernel size of the first convolution layer $F^1$ is 9, which is used to extract more abundant image features. In addition, the BN layer is not added to the sixth convolution layer $F^6$ because this layer is used to reduce the dimensions by properly fusing feature maps. At last, the skip connection operation is used to connect the feature map $F^3$ of the $i$th layer's output with the feature map $F^5$ of the $j$th layer's output. The 1×1 convolution kernel used in the cascade operation is as follows:

$$F^6 = \max(0, \omega^6 \ast [F^1, F^5] + b^6),$$

where $[F^1, F^5]$ means connecting the two 3D feature mapping matrices in the third dimension; $\omega^6$ and $b^6$ represent the weight matrix and offset matrix of the sixth convolution layer, respectively; and $F^6$ is the feature map of the sixth layer. Finally, the output of the second path is expressed as follows:

$$F_{16}^1 = \omega_{16} \ast F_{15} + b_{16},$$

where $\omega_{16}$, $b_{16}$, $F_{15}$ and $F_{16}$ have the same meaning as the corresponding variables in Equation (3). Note that $X^c$ represents the first path and $X_s$ represents the second path, where $X \in (\omega, b, F_i)$ and $x \in (1, 2, \ldots)$.

Finally, we design a feature fusion connection structure to connect the microscopic and macroscopic features. Specifically, the above two paths are fused by a connection operation, and then the convolution layer with a kernel size of one is used to transition features:

$$F_{17} = \max(0, \omega_{17} \ast [F_{11}, F_{16}] + b_{17}),$$

where $\omega_{17}$, $b_{17}$ and $F_{17}$ have the same meaning as the corresponding positions in Equation (4). Then, the GAP layer is used to reduce the number of parameters and alleviate the overfitting problem:

$$g_s = \frac{1}{S} \sum_{i=1}^{S} p_{st} \ s \leq S, (1 \leq j \leq S),$$

where $p_{st}$ represents the $i$th pixel value in the $t$th feature map, $S$ is the total number of pixels in each feature map and $t$ is the number of sequences of the feature map. In the end, 128 GAP values are trained through the FC layer to get the prediction value of the DPDNN for sonar image quality.

2.2 | Network training settings

This paper uses the stochastic gradient descent optimization algorithm in the training process, which has shown good performance in previous research. To make the network converge to better parameters, we use momentum and learning rate attenuation in the optimization algorithm, so that the network can avoid falling into the local optimum and a saddle point [17]. Specifically, the momentum coefficient is 0.9, the initial learning rate is 0.0001, the attenuation rate is 0.00001, the minimum batch size is set to 16 and the number of training epoch is set to 300 [18].
| Path1-layer | Type       | Parameters                                                                 |
|------------|------------|-----------------------------------------------------------------------------|
| $F_1, F_2$ | Convolution| Filter size: 3 × 3, Filter number: 32, Stride: 1 × 1, Padding: Same, Activation function: ReLU |
| $F_3$      | Pooling    | Pooling region size: 3 × 3, Stride: 2 × 2, Padding: Same, Pooling method: Max-pooling |
| $F_4, F_5$ | BN&Convolution | Filter size: 3 × 3, Filter number: 48, Stride: 1 × 1, Padding: Same, Activation function: ReLU |
| $F_6$      | Pooling    | Pooling region size: 3 × 3, Stride: 2 × 2, Padding: Same, Pooling method: Max-pooling |
| $F_7, F_8$ | BN&Convolution | Filter size: 3 × 3, Filter number: 64, Stride: 1 × 1, Padding: Same, Activation function: ReLU |
| $F_9$      | Pooling    | Pooling region size: 3 × 3, Stride: 2 × 2, Padding: Same, Pooling method: Max-pooling |
| $F_{10}, F_{11}$ | BN&Convolution | Filter size: 3 × 3, Filter number: 80, Stride: 1 × 1, Padding: Same, Activation function: ReLU |

| Path2-layer | Type       | Parameters                                                                 |
|------------|------------|-----------------------------------------------------------------------------|
| $F^1$      | Convolution| Filter size: 9 × 9, Filter number: 16, Stride: 1 × 1, Padding: Same, Activation function: ReLU |
| $F^2$      | Convolution| Filter size: 1 × 1, Filter number: 16, Stride: 1 × 1, Padding: Same, Activation function: ReLU |
| $F^3, F^4, F^5$ | BN&Convolution | Filter size: 3 × 3, Filter number: 32, Stride: 1 × 1, Padding: Same, Activation function: ReLU |
| $F^6$      | Concatenation| $F^1, F^5$                                                                 |
| $F^7$      | Convolution| Filter size: 1 × 1, Filter number: 32, Stride: 1 × 1, Padding: Same, Activation function: ReLU |
| $F^8$      | Pooling    | Pooling region size: 3 × 3, Stride: 2 × 2, Padding: Same, Pooling method: Max-pooling |
| $F^9, F^{10}$ | BN&Convolution | Filter size: 3 × 3, Filter number: 48, Stride: 1 × 1, Padding: Same, Activation function: ReLU |
| $F^{11}$   | Pooling    | Pooling region size: 3 × 3, Stride: 2 × 2, Padding: Same, Pooling method: Max-pooling |
| $F^{12}, F^{13}$ | BN&Convolution | Filter size: 3 × 3, Filter number: 64, Stride: 1 × 1, Padding: Same, Activation function: ReLU |
| $F^{14}$   | Pooling    | Pooling region size: 3 × 3, Stride: 2 × 2, Padding: Same, Pooling method: Max-pooling |
| $F^{15}, F^{16}$ | BN&Convolution | Filter size: 3 × 3, Filter number: 80, Stride: 1 × 1, Padding: Same, Activation function: ReLU |

| Combination | Type       | Parameters                                                                 |
|------------|------------|-----------------------------------------------------------------------------|
| $F^{17}    | Concatenation| $F_11, F^{16}$                                                              |
| $F^{18}    | BN&Convolution| Filter size: 1 × 1, Filter number: 128, Stride: 1 × 1, Padding: Same, Activation function: ReLU |
| $F^{19}    | Pooling    | Pooling method: Global average pooling                                      |
| $F^{20}$   | Full Connection| Neurons number: 1                                                            |

Similar to most regression tasks, the mean square error (MSE) loss function is used to optimize the network during training:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\hat{x}_i - x_i)^2, \quad (8)$$

where $\hat{x}_i$ and $x_i$ represent the true label value and predicted value of the $i$th image, respectively; and $N$ represents the total number of labels. Then, DPDNN is trained in the built environment, so that the network can obtain the best correlation and the best model parameters on the verification set. The network layer structure parameters of the DPDNN are shown in Table 1.

In the neural network training, the amount of data is very important since it determines the robustness and generalization of the network. However, the number of images in the SIQD database is only 840 sonar images, and so we augment the amount of sonar data. Specifically, we will rotate, horizontally fold and vertically fold the sonar images and expand the data to eight times the original of data. Figure 3 is a display diagram.

**Figure 3** Examples of sonar image expansion
3 EXPERIMENTAL RESULTS

In this section, we will demonstrate the performance of the DPDNN model on the newly proposed SIQD database. The SIQD database contains 40 lossless sonar images taken by sonar equipment (as shown in Figure 4, which consists of swimmers, seabed, shipwrecks and underwater creatures) and 800 lossy sonar images generated by four distortion types at four to six distortion levels. Each sonar image in the SIQD database will be scored subjectively by 25 non-professional undergraduate or graduate students, and finally the mean opinion score (MOS) will be obtained through data cleaning. For more details, please refer to reference [14]. In the neural network training task, the amount of data is an important problem to be considered. Therefore, we rotate or flip the images in the SIQD database by $90^\circ$, $180^\circ$, $270^\circ$, horizontally and vertically to get 6720 images. Then the expanded data set is divided into five groups according to the types of reference images, among which three groups form the training set (4032 images), one group is the verification set (1344 images) and the other group is the test set (1344 images). It should be noted that each group needs to be tested, and so we need to train five models and take the final average as the result of the DPDNN.

To better verify the performance, we compare the DPDNN model with 11 state-of-the-art IQE models: BLind Image Integrity Notator using DCT Statistics 2 (BLIINDS2) [19], Blind/Reference-less Image Spatial QUality Evaluator (BRISQUE) [20], Blind image quality assessment using joint statistics of Gradient Magnitude and Laplacian Features (GMLF) [21], No-reference Free Energy based Robust Metric (NFERM) [22], AR-based Image Sharpness Metric (ARISM) [23], Blind Quality Measure for Screen content images (BQMS) [24], Blind Image Quality Assessment method based on High Order Statistics Aggregation (HOSA) [25], Accelerated Screen Image Quality Evaluator (ASIQE) [26], Blind Pseudo Reference Image based quality measure (BPRI) [27], Blind Multiple

| Models          | PLCC  | SROCC | KROCC | RMSE  | MAE   |
|-----------------|-------|-------|-------|-------|-------|
| BLIINDS2 [19]   | 0.4459| 0.4410| 0.3017| 12.517| 10.474|
| BRISQUE [20]    | 0.6142| 0.6072| 0.4240| 11.036| 9.0415|
| GMLF [21]       | 0.3210| 0.0935| 0.0630| 13.244| 10.894|
| NFERM [22]      | 0.4376| 0.4228| 0.2899| 12.574| 10.166|
| ARISM [23]      | 0.4653| 0.4442| 0.3010| 12.378| 10.077|
| BQMS [24]       | 0.4436| 0.3742| 0.2575| 12.533| 10.096|
| HOSA [25]       | 0.5604| 0.5665| 0.3921| 11.582| 9.3752|
| ASIQE [26]      | 0.5093| 0.4309| 0.3010| 12.035| 9.4491|
| BMPRI [27]      | 0.3849| 0.3875| 0.2625| 12.907| 10.364|
| BPRI [28]       | 0.4787| 0.4954| 0.3472| 12.278| 9.9174|
| NRCDM [9]       | 0.7340| 0.7791| 0.5138| 9.4979| 7.4609|
| DPDNN (Prop.)   | 0.7763| 0.7872| 0.5955| 8.9497| 6.8335|

Pseudo Reference Image based quality measure (BPRI) [28] and No-Reference Contour Degradation Measurement (NRCDM) [9]. Among these, the NRCDM model is a recently proposed model for evaluating the quality of sonar images based on the degree of degradation of a filtered sonar image contour. To fairly measure the quality, we use five classic metrics recommended by the video quality expert group [29]: the PLCC is used to measure the model prediction accuracy; the SROCC and KROCC are used to measure the monotonicity of the model predictions and the RMSE and MAE are used to measure the consistency of the model predictions. The values of the PLCC, SROCC and KROCC are closer to 1 and the values of RMSE and MAE are closer to 0, indicating that the performance of the model is better. It should be noted that there may be non-linearity in the subjective scoring process. Therefore, we first use a logistic regression to map the objective quality prediction score of each
image non-linearly:

\[ f(s) = \alpha_1 \left( \frac{1}{2} - \frac{1}{1 + e^{\alpha_2(s-\alpha_3)}} \right) + \alpha_4 s + \alpha_5, \]

(9)

where \( f(s) \) is the evaluation score after the regression; \( s \) represents the sonar image quality prediction score and \( \alpha_{1,2,3,4,5} \) is the five fitting parameters based on the Gauss–Newton method.

The data recorded in Table 2 were obtained using the SIQD database according to the above evaluation metrics. For facilitating identification by readers, we roughen the results of the two models with the best performance. It can be seen from the experimental results that the DPDNN model proposed in this paper achieves the best performance in five test indexes, among which the performance gain are approximately 5.76%, 11.01%, 15.90%, 5.77% and 8.41% in terms of the PLCC, SROCC, KROCC, RMSE and MAE, respectively. In addition, only the performance of our proposed model is higher than 0.75 on the PLCC and SROCC, higher than 0.55 on the KROCC and lower than 9 on the RMSE and MAE, which again show that the DPDNN model makes excellent predictions with respect to accuracy, monotonicity and consistency. Except for the NRCMDM, all other selected methods have achieved good performance for natural images or screen content images, but their performances on sonar images are not as good as our DPDNN model, as shown in Table 2.

In addition to the above four indicators, the execution efficiency is also an important measure. In the experiment, we use MATLAB R2014a to test the selected comparative model, and use TensorFlow [30] and Keras [31] to train and test the DPDNN model. The operating environment is a win10 system server, which has an Intel(R) Xeon(R) CPU E5-2620 v4 at 2.10 GHz with 192.00 GB of RAM and an NVIDIA TITAN Xp GPU. From Table 3, it can be easily found that the proposed DPDNN model only consumes 11.7 ms for each picture, far exceeding the execution efficiency of other models, mainly because the DPDNN model has fewer parameters at only 0.45 MB. Compared with the state-of-the-art NRCMDM, our DPDNN not only achieves excellent prediction performance, but it also improves the execution efficiency by approximately 15 times.

To show the performance of DPDNN model more intuitively, we draw the scatter plots between the objective quality prediction score obtained by the above 11 IQE models and our DPDNN model and the corresponding MOS of in Figure 5. In each scatter diagram, we use different colour marks to distinguish the sampling points of different distortion types. Among them, red circles, green squares, blue diamonds and pink triangles are used to represent ‘Group 1’, ‘Group 2’, ‘Group 3’ and ‘Group 4’, respectively. ‘Group 1’–‘Group 4’ are image groups generated by the four types of distortion. We can easily find two conclusions from Figure 5. (1) Compared with the other 11 IQE models, our DPDNN model has higher monotonicity and linearity. Specifically, the DPDNN model is slen-derer than NRCMDM model (ranking second) which is specially used for sonar IQE. (2) As we all know, an outstanding IQE model should be able to predict each distortion type well. As shown in Figure 5, BLIINDS2, BRISQUE, GMLF, ARISM, BQMS, HOSA and ASIQE cannot accurately evaluate ‘Group 3’; NFERM cannot evaluate ‘Group 1’ and ‘Group 3’ very well and BMPRI and BPRI cannot evaluate ‘Group 2’ properly. However, our DPDNN model is robust to the four distortion types in the SIQD database. In addition, we also show some examples of the DPDNN and NRCMDM model (ranking second) when evaluating sonar image quality. As shown in Figure 6, the evaluation result of the proposed DPDNN network is similar to that of manual scoring, while the prediction results of the NRCMDM have a certain gap with MOS scores, which shows that the DPDNN has good performance.

Finally, we use the \( F \)-test to further test the statistical significance of the DPDNN model. Specifically, we need to compare the variance of two objective methods in the prediction value. If \( F > F_{critical} \), there is a significant difference between the two methods. In our test, \( F_{critical} \) is equal to 0.05. Therefore, we can make a reasonable statistical judgement on the new objective method by analysing the characteristics of the proposed objective method. The statistical significance results are shown in Table 4. The symbol ‘0’ indicates that there is no significant difference between the two objective methods, ‘+1’ indicates that our method is statistically superior to the other method and ‘−1’ indicates that our method is statistically inferior to the method. It can be seen from Table 4 that the DPDNN model we proposed obtains a statistically significant performance improvement.

## 4 Conclusion

The emergence of sonar technology shortens the distance between humans and the ocean, and the clarity of sonar images is an important part of our understanding of the ocean. Considering that sonar images are task-oriented rather than aesthetic images, we design a DPDNN based on microscopic and macroscopic structural information. In our method, BN and skip operations are used to reduce the training time and speed up feature extraction, respectively. Finally, the GAP layer and
FIGURE 5  Scatter plots of MOS versus BLIINDS2 [19], BRISQUE [20], GMLF [21], NFERM [22], ARISM [23], BQMS [24], HOSA [25], ASIQE [26], BMPRI [27], BPRI [28], NRCDM [9] and our DPDNN on SIQD database

TABLE 4  The comparison between DPDNN model and other 11 state-of-the-art IQE models in $F$-test (statistical significance)

| Metrics | BLIINDS2 [19] | BRISQUE [20] | GMLF [21] | NFERM [22] | ARISM [23] | BQMS [24] | HOSA [25] | ASIQE [26] | BMPRI [27] | BPRI [28] | NRCDM [9] |
|---------|----------------|---------------|------------|-------------|-------------|------------|------------|------------|-------------|-------------|-----------|
| Result  | +1             | +1            | +1         | +1          | +1          | +1         | +1         | +1         | +1          | +1          | +1        |
the FC layer are used to integrate the microscopic and macroscopic features extracted by the two paths. The experimental results on the newly established SIQD database show that the proposed DPDNN model achieves a significant improvement in prediction performance and execution efficiency. Compared with the state-of-the-art NRCDM model, our DPDNN model is 15 times more efficient in terms of its execution efficiency, which makes it very suitable for environments with high real-time requirements, such as torpedo detection. In the future research, we will consider applying block processing to sonar imagery to achieve improved performance.

REFERENCES

1. Zhang, W., et al.: Underwater pipeline leakage detection via multi-beam sonar imagery. J. Acoust. Soc. Am. 141, 3917–3917 (2017)
2. Chen, W., et al.: Subjective and objective quality evaluation of sonar images for underwater acoustic transmission. In: IEEE Int. Conference on Image Process., pp. 17–20 (2018)
3. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Trans. Image Process. 15, 430–444 (2006)
4. Gu, K., et al.: A fast reliable image quality predictor by fusing micro- and macro-structures. IEEE Trans. Ind. Electron. 64, 3903–3912 (2017)
5. Gu, K., et al.: Saliency-guided quality assessment of screen content images. IEEE Trans. Multimedia 18, 1–13 (2016)
6. Fang, Y., et al.: Objective quality assessment of screen content images by uncertainty weighting. IEEE Trans. Image Process. 26, 2016–2027 (2017)
7. Chen, W., et al.: Statistical and structural information backed full-reference quality measure of compressed sonar images. IEEE Trans. Circuits Syst. Video Technol. 30, 334–348 (2020)
8. Chen, W., et al.: Partial-reference sonar image quality assessment for underwater transmission. IEEE Trans. Aerosp. Electron. Syst. 54, 2776–2787 (2018)
9. Chen, W., et al.: Reference-free quality assessment of sonar images via contour degradation measurement. IEEE Trans. Image Process. 28, 5336–5351 (2019)
10. Bi, J., et al.: SGW-SCN: An integrated machine learning approach for workload forecasting in geo-distributed cloud data centers. Inf. Sci. 481, 57–68 (2019)
11. Bi, J., et al.: Time-dependent cloud workload prediction via multi-task learning. IEEE Rob. Autom. Lett. 4, 2401–2406 (2019)
12. Wang, G.M., et al.: TL-GDBN: Growing deep belief network with transfer learning. IEEE Trans. Autom. Sci. Eng. 16, 874–885 (2019)
13. Gu, K., et al.: Deep dual-channel neural network for image-based smoke detection. IEEE Trans. Multimedia 22, 311–323 (2020)
14. Chen, W., et al.: Subjective quality assessment of sonar images based on underwater acoustic transmission. To be published in Signal Image Video Process. [Online]. http://arxiv.org/abs/2104.02661
15. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014
16. He, K., et al.: Deep residual learning for image recognition. In: Proc. IEEE Conf. Comp. Vis. and Pattern Recognit., pp. 770–778 (2016)
17. Bengio, Y.: Practical recommendations for gradient-based training of deep architectures. In: Neural Networks: Tricks of the Trade, pp. 437–478. Springer, Berlin, Heidelberg (2012)
18. Lecun, Y., et al.: Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324 (1998)
19. Saad, M.A., Bovik, A.C., Charrier, C.: Blind image quality assessment: A natural scene statistics approach in the DCT domain. IEEE Trans. Image Process. 21, 3339–3352 (2012)
20. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. 21, 4695–4708 (2012)
21. Xue, W., et al.: Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features. IEEE Trans. Image Process. 21, 4850–4862 (2014)
22. Gu, K., et al.: Using free energy principle for blind image quality assessment. IEEE Trans. Multimedia 17, 50–63 (2015)
23. Gu, K., et al.: No-reference image sharpness assessment in autoregressive parameter space. IEEE Trans. Image Process. 17, 3218–3231 (2015)
24. Gu, K., et al.: Learning a blind quality evaluation engine of screen content images. Neurocomputing 171, 140–149 (2016)
25. Xu, J., et al.: Blind image quality assessment based on high order statistics aggregation. IEEE Trans. Image Process. 25, 4444–4457 (2016)
26. Gu, K., et al.: No-reference quality assessment of screen content pictures. IEEE Trans. Image Process. 26, 4005–4018 (2017)
27. Min, X., et al.: Blind image quality estimation via distortion aggravation. IEEE Trans. Broadcast. 26, 508–517 (2018)
28. Min, X., et al.: Blind quality assessment based on pseudo-reference image. IEEE Trans. Multimedia 20, 2049–2062 (2018)
29. VQEG: Final report from the video quality experts group on the validation of objective models of video quality assessment, [EB/OL]. 2006. http://www.vqeg.org/
30. Abadi, M., et al.: TensorFlow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467, 2016
31. Chollet, F., Keras, 2015. [Online]. http://keras.io

How to cite this article: Zhang H, Li S, Li D, Wang Z, Zhou Q, You Q. Sonar image quality evaluation using deep neural network. IET Image Process. 2021;1–8. https://doi.org/10.1049/ip-r2.12199