

Research Article

CNN-Based Medical Ultrasound Image Quality Assessment

Siyuan Zhang, Yifan Wang , Jiayao Jiang, Jingxian Dong, Weiwei Yi ,
and Wenguang Hou 

College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China

Correspondence should be addressed to Weiwei Yi; 741173912@qq.com and Wenguang Hou; houwenguang99@163.com

Received 26 March 2021; Revised 18 June 2021; Accepted 22 June 2021; Published 2 July 2021

Academic Editor: Hocine Cherifi

Copyright © 2021 Siyuan Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The quality of ultrasound image is a key information in medical related application. It is also an important index in evaluating the performance of ultrasonic imaging equipment and image processing algorithms. Yet, there is still no recognized quantitative standard about medical image quality assessment (IQA) due to the fact that IQA is traditionally regarded as a subjective issue, especially in case of the ultrasound medical images. As such, the medical ultrasound IQA on basis of convolutional neural network (CNN) is quantitatively studied in this paper. Firstly, a dataset with 1063 ultrasound images is established through degenerating a certain number of original high-quality images. Subsequently, some operations are performed for the dataset including scoring and abnormal value screening. Then, 478 ultrasonic images are selected as the training and testing examples. The label of each example is obtained by averaging the scores of different doctors. Afterwards, a deep CNN network and a residuals network are taken to establish the IQA models. Meanwhile, the transfer learning strategy is introduced here to accelerate the training and improve the robustness of the model considering the fact that the ultrasound image samples are not abundant. At last, some tests are taken to evaluate the IQA models. They show that the CNN-based IQA is feasible and effective.

1. Introduction and Motivation

Image quality assessment (IQA) is to quantitatively evaluate the image, which remains a hot topic in image processing field due to the fact that it is regarded as a benchmark for image processing systems and algorithms Tang et al. [1], Kim et al. [2], and Ma et al. [3]. As many pattern recognition problems, IQA tries to simulate the human perception process which is easily influenced by image content, mathematical and psychological effect of the observer, and many other complex factors Krasula et al. [4]. So far, most of IQA methods and research studies focus on optical images rather than medical images. One of the reasons is that the medical image quality is highly related to its specific application. For example, a medical image has overall low contrast and is noisy, but it is still acceptable for the doctor if it is effective in judging the state of a certain tissue. Another is that the medical image including the MR, CT, and US usually contains artifacts, which are caused by the tissue movement in imaging or the scattering of beams Krupa and Bekiesińska-Figatowska [5], Boas and Fleischmann [6], and

Prabhu et al. [7]. Most artifacts cannot be removed. In fact, the doctors can grasp the useful information from the noisy images; that is, noisy medical image does not always mean low quality. By contrast, the noisy optical images are usually seen as images with low quality. Thus, the medical image assessment should be conducted from a different viewpoint.

The traditional IQA can be divided into two kinds: subjective assessment and quantitative assessment Hemmsen et al. [8], Kang et al. [9], Bosse et al. [10], and Kim and Lee [11]. For the former one, the image is scored by observers. According to whether relying on the other image in IQA, the subjective assessment can be divided into single and double excitation cases. In the double excitation assessment, the observers score the image after observing the considered image and its related one with high quality. It means that the considered image should have comparably less quality. The International Telecommunications Union (ITU) provided a standard for double excitation IQA and it has been widely used in the fast MR imaging; Union [12], Shiao et al. [13], Loizou et al. [14], and Hemmsen et al. [8] attempt to apply it in the ultrasound IQA. In the single

excitation assessment, the observers score the image only relying on their experience. It is adaptive to more situations. For an image, the scores given by different observers may have some differences. So, a certain number of observers should take part in the process of assessment. By comparison, the quantitative assessment scores the image through automatically computing some indexes for the considered image.

Generally, the quantitative assessment algorithms can be divided into three kinds: full reference, reduced reference, and no reference IQA Kang et al. [9] and Zhu et al. [15]. The full reference IQA relies on the original high-quality image in evaluating the considered image. The reduced reference IQA relies on the part of the high-quality image. There are not any references for the no reference IQA, which is also named the Blind Image Quality Assessment (BIQA) Kim and Lee [11], Ma et al. [16], and Ma et al. [3]. Generally speaking, the current quantitative image assessment algorithms mostly belong to the full reference and reduced reference assessment. The traditional image quality indexes include peak signal-to-noise ratio (PSNR), mean squared error (MSE), and structural similarity index (SSIM), which all need the original high-quality image Bianco et al. [17]. Besides, some indexes about the image gradient maps are also considered. Human visual system- (HVS-) based methods transfer the image into different space and they simulate the respondents of human visual cortex neuron about the low-quality and high-quality images Litjens et al. [18]. The perceptual difference model (PDM) Daly [19] is widely used in medical image quality assessment, which models the ability of humans to perceive a visual difference between a degraded “fast” MRI image with subsampling of k-space and a “gold standard” image mimicking full acquisition Huo et al. [20]. Mittal et al estimate the possible information loss between the considered image and the original high-quality one on the basis of the normalized intensity coefficient in the spatial space [21]. The full or reduced reference IQA relies on some information to judge the dissimilarity between high-quality and low-quality images. It is effective. However, in most situations, there are not any references in IQA. The former research studies in no reference IQA mainly dealt with the special image deformation such as noise, blur, and image compression. These methods perform feature detection and statistic computation for the considered image, which is computation complex. Woodard evaluates the MR image according to the variances of the considered image and its degraded one. Mortamet et al. evaluated the MR image relying on the part of the atmosphere in the image because 40% of image is the atmosphere in the structure MR brain image Mortamet et al. [22]. Nakhaie and Shokouhi evaluated the image through wavelet transform [23]. Recently, Eck provided a new rule. That is, the good-quality image means that it effectively helps the doctor to detect the changes of tissue Eck et al. [24]. Thus, the standard of IQA is whether the image can make the doctor conduct accurate judgement.

In recent years, CNN has been taken in IQA to simulate the process of human evaluation of optical images inspired by the success of CNN in image processing and pattern

recognition De Angelis et al. [25] and Bosse et al. [10]. Kang firstly proposed no reference IQA on the basis of CNN, in which the CNN is trained using the samples in LIVE IQA Kang et al. [9], Zhang et al. [26], and Zhang et al. [27]. Later, Bianco proposed DeepBIO, which trains a classification CNN using the other kind of data and then transfers the network to the image quality assessment [17]. Kim et al. estimated the quality of an image by adding up the scores of each patch on the basis of the CNN without any reference. After that, they designed a deep CNN IQA model which contains two steps: error map learning and score estimation Kim et al. [2]. These methods attempt to replace the human visual perception process using CNN. A certain number of labeled images are taken to train the CNN. To the best of our knowledge, the related reports are mainly about MR or CT IQA. So far, a few researchers have used deep learning to assess the quality of ultrasound images. In Wu et al.’s work [28], a computerized fetal US image quality assessment scheme is proposed to assist the implementation of US image quality control in the clinical obstetric examination by introducing two deep CNNs. It is adaptive to a special application, and we intend to provide a universal method to evaluate the image quality. The remainder of this paper is organized as follows. Section 2 introduces the way to collect the training samples. The CNN is designed in Section 3, and the results are shown in Section 4. Finally, we conclude our method in Section 5.

2. Dataset

2.1. Optical Image Samples. To train an IQA CNN, we first collect lots of image samples. In optical image quality assessment, LIVE IQA [23] is one of the popular and widely used datasets. It contains 29 original high-resolution reference images. Each original image is degraded into some low-quality images through JPG compression, Gaussian convolution, fast Rayleigh fading, or adding white noise. Then, a total of 982 images including the original ones are obtained. Each of these images is scored at least by 20 observers. The average score for each image is taken as its label. In our algorithm, the images in LIVE are taken to pretrain the IQA CNN. Afterwards, the CNN is finely tuned by some ultrasound images.

2.2. The Ultrasound Samples. The ultrasound image samples in this paper come from two ways. The first is through downloading the images from publicly accessed websites. The other way is by collecting 700 ultrasound images from the Tongji Hospital, affiliated to Huazhong University of Science and Technology. These images are captured in the department of gynaecology, ophthalmology department, and internal medicine and surgical department and strictly screened by some experienced doctors. Then, 95 high-quality ultrasound images remained, which corresponded to the tissues of liver, lymph node, bladder, breast, kidney, heart, blood vessel, pancreas, and uterus. All the images are in 8-bit bmp format. A part of the original high-quality ultrasonic images are shown in Figure 1.

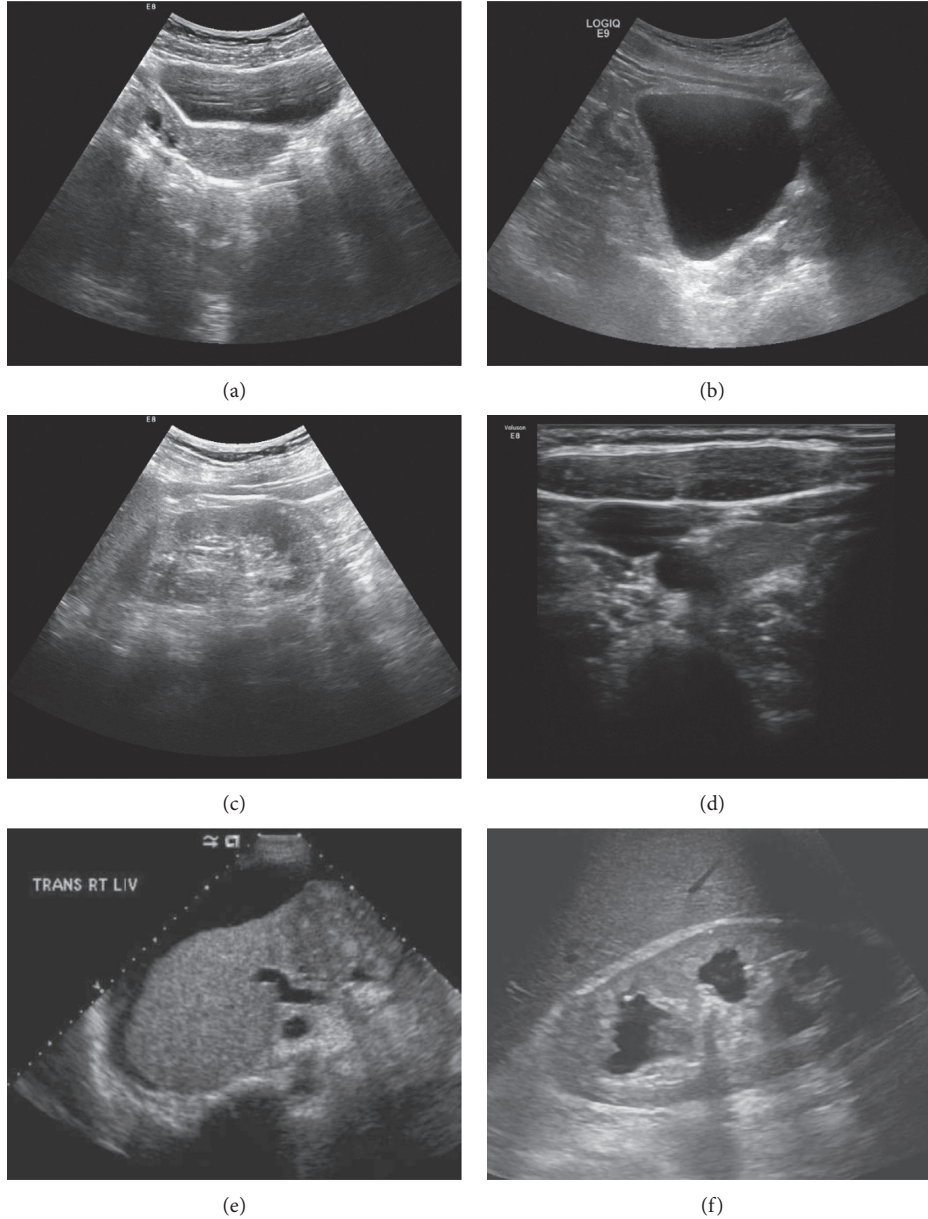


FIGURE 1: Original high-quality ultrasound images.

In order to create an ultrasound image database containing different quality levels, similar to many methods in generating the samples, we degrade the high-quality ultrasound images through JPEG compression, Gaussian blurring, and adding white and speckle noise. The speckle noise is common in ultrasound image; the model in Li et al.'s work [29] is used to generate the speckle noise due to the fact that this model has been validated to effectively describe the speckle noise. Meanwhile, the commonly used degeneration methods are also used to generate the low-resolution images. Specifically, the function “imwrite” in Matlab is taken to perform JPEG compression, in which the quality factor is a random integer between 0 and 100. In adding the white noise, the original image is normalized to 0 and 1. Then, the Gaussian noise with δ standard error $0 < \delta < 2$ is added.

Afterwards, the noisy image is restored to 0 – 255. In Gaussian blurring, a 7×7 template is taken to convolute the original image. The standard error of the Gaussian template is randomly selected between 0 and 1. In generating the speckle noisy images, we used the exponential form noise model in Li et al.'s work [29] as

$$u(x) = v(x) + v^\gamma(x)\eta(x), \quad (1)$$

where $u(x)$ and $v(x)$ are the original and observed images, respectively. $\eta(x) \sim N(0, \delta^2)$ is the Gaussian distributed noise whose average value is zero and the standard error is a random data between 0 and 6. The simulation effect is better when $\gamma = 5$. In generating the speckle noise images, 7 different standard errors are taken. In each of the remaining

three methods, 6 levels of degeneration are conducted for the original images. For each method, we randomly choose half of the 95 images to do degeneration. At last, 1063 ultrasound images are generated.

2.3. Scoring the Ultrasound Samples. Four doctors from Tongji Hospital scored the acquired ultrasound images with a single excitation method. It is out of two considerations: one is that the number of images to be scored is relatively large, and double evaluation is more prone to visual fatigue; the other is that the double excitation only fits limited situations. The four doctors all major in biomedical imaging. The scores are from 0 to 100. It should be mentioned that the images are unorganized before scoring to make the objective scoring. After obtaining the scores of all images, we remove some inconsistent score samples. For each original image, the absolute deviation between the scores and the mean of four observers can be easily computed. When the deviation is greater than a given threshold, it will be regarded as outlier and discarded. For accepted samples, the mean opinion score (MOS) value is taken as its label. It is clear that the different threshold can lead to a different number of samples. The higher the threshold is, the more consistent the samples' scores are, leading to less samples. In our experiment, the threshold is set to be 10, which balances the consistency among samples' scores and samples' numbers. After scoring, 478 images remained.

2.4. Outliers' Screening and Distribution Balance. Based on the simple selection of abnormal samples in Section 2.3, a CNN is used to conduct further outliers' screening for the samples on the basis of the principle of random consistency. The screening process is shown in Figure 2. The network mentioned in the flowchart is shown in Section 3. The basic rule is to first randomly select some images in the database to train the initial evaluation model. Then, we input all data into the model to predict the score value and detect the abnormal samples according to the difference between the prediction score and the actual label. This process (Figure 2, dotted box section) is carried out for several times; then each image can obtain multiple groups of predicted values. If some images show anomalies in multiple predicted values, they will be screened. The above screening process is repeated several times and 78 images are discarded.

After scoring and outliers' screening, the distribution of the average subjective score of the remaining images is not balanced, so a certain distribution balance method is needed to balance the image samples in the database. In this paper, the image expansion is performed on the interval with less samples by rotating the original image at three angles of 90° , 180° , and 270° , and their labels take the corresponding unrotated images' labels. The image dataset after rotation expansion contains 478 ultrasound images. It should be mentioned that the subjective evaluation is conducted in condition of common electric incandescent lamp, the monitor is 4K LED monitor, and the view distance is about sixty centimeters. Simply, the subjective evaluation is conducted under the common office by the experienced doctors.

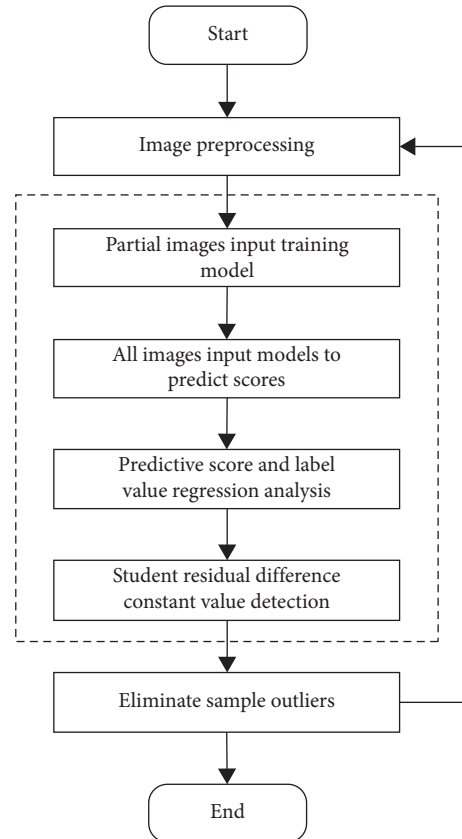


FIGURE 2: The flowchart of outliers' screening.

3. CNN for IQA

3.1. Deep CNN for IQA. Due to the complexity of the medical ultrasound image content, the shallow neural network may not be able to well simulate the perception of HVS in evaluating the image. As such, this project adopts a deep convolutional neural network to assess the quality of medical ultrasound images. The research on HVS shows that human is sensitive to the deformation between images. Therefore, we firstly train a CNN to learn the difference between distorted image and the related undistorted one. After that, the objective scoring for each image is carried out on the basis of the estimated distortion. It tries to simulate the human perceptual processes.

Accordingly, this paper designs a deep CNN to do IQA, which is called DCNN-IQA-14. This network adds six convolution layers to DCNN-IQA-8 in Kim et al.'s work [2] and their structures are shown in Figures 3 and 4. The objective error map is predicted in the first stage. The whole network is a full convolution network; that is, it only contains convolution layer. Zero padding strategy is used in each convolution layer to make the convolution process to retain the pixel information, and two downsampling operations are used to reduce the data dimension. Except for the last layer, each layer has a convolution kernel size of 3×3 , which is activated by ReLU. In the last layer, the 1×1 convolution kernel is used to output the error map prediction. If the error map is directly used in the second stage to do quality assessment, its

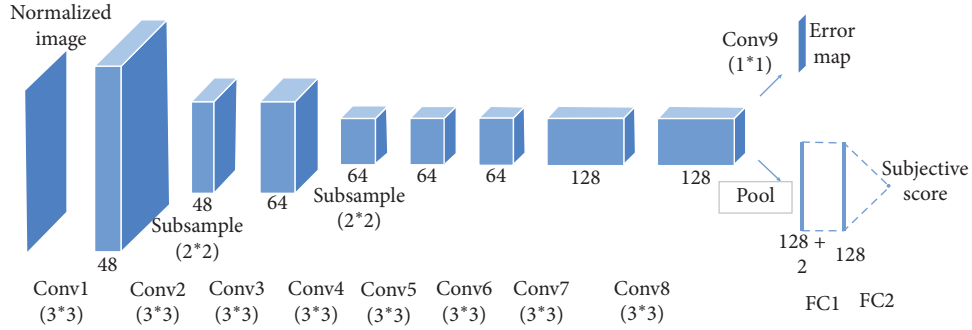


FIGURE 3: The structure of DCNN-IQA-8 proposed by Kim et al. for image quality assessment.

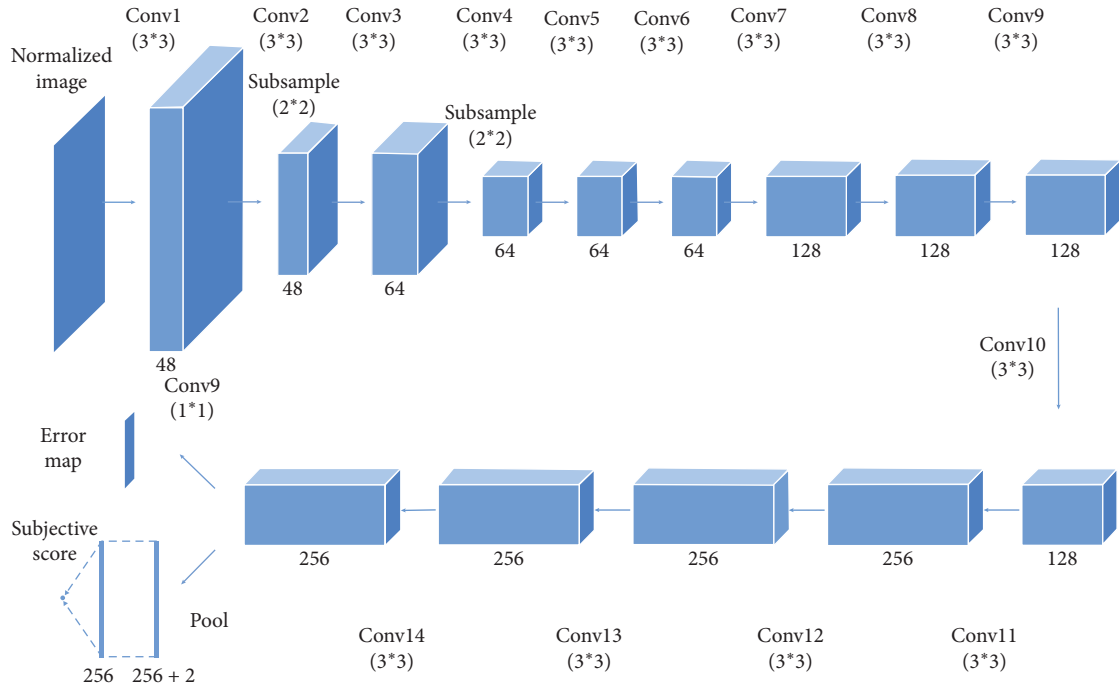


FIGURE 4: The structure of DCNN-IQA-14 for ultrasound image quality assessment.

prediction result is not fine due to the fact that the predicted error map has only one channel and that much information about the difference among images is lost. Therefore, the feature map of the penultimate output is used in the second stage of training. In the second stage, the fourteenth convolution layer of the network will return to the final subjective score through two full-connection layers.

Before the training, all samples should be normalized. The normalization operations include the following steps. The first is to reduce the image to a quarter of the original image size and then enlarge it to the original size. Then, Gaussian Low Pass is taken to do filtering. Subsequently, the filtered image is subtracted from the original image. Finally, we divided them into the nonoverlapped patches with size of 112×112 . Each patch's label is the MOS value of the related image. Let I_d represent the deformation image, I_r represent the corresponding high-quality reference image, I_d^{low} and I_r^{low} , respectively, represent the corresponding image after zooming in and zooming out low-pass filtering, and \hat{I}_d and \hat{I}_r are the

preprocessed images; then the error map calculation formula is given by

$$e_{gt} = |\hat{I}_r - \hat{I}_d|, \quad (2)$$

where e_{gt} is a label for the first training stage. Stochastic gradient descent (SGD) algorithm is taken to minimize the objective function shown in the following equation:

$$\operatorname{argmin}_{\theta} \sum_{i,j} \|f(I_d; \theta)(i, j) - e(i, j) \cdot \hat{r}\|^2, \quad (3)$$

where (i, j) is the gray value of pixels in the image, e is the difference between the distorted image and its corresponding original reference image, and \hat{r} is reliability map obtained by measuring the texture intensity.

In the second training stage, besides the feature map obtained in the first stage, two manually extracted features are added to full-connection layer FC1: $\delta_{I_d}^{\text{low}}$ and μ_r , respectively, the variance of the low-pass distorted image and the mean of the reliability map. By using the subjective

scores of each distorted image as labels and minimizing equation (4), the final score-predictive model can be trained. It should be mentioned that two CNNs are taken to do evaluation, which is just an example to illustrate the effectiveness of the proposed solution. Certainly, there are many other networks which can obtain the same or even better results. As this paper mainly intends to propose this issue and an effective solution, the comparison of different networks is not our main topic.

$$\operatorname{argmin}_{\theta} \sum_{\theta} \|f(I_d; \theta)(i, j) - S\|^2. \quad (4)$$

3.2. Classical Classification Network for IQA. In the experiment, we also used a 34-layer depth residual network (ResNet). The last layer of the network is replaced by a full-connection layer to output the score of 1×1 . This network is called ResNet-IQA and the structure is shown in Figure 5. Except for the first layer, each layer has a convolution kernel size of 3×3 , which is activated by ReLU.

Divide the ultrasound images into the nonoverlapped patches with size of 112×112 as the input. Each patch's label is the MOS value of the related image and each patch is normalized according to the following formulas:

$$\hat{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\delta(i, j) + C}, \quad (5)$$

$$\begin{aligned} \mu(i, j) &= \frac{\sum_{p=-P}^{p=P} \sum_{q=-Q}^{q=Q} I(i+p, j+q)}{4PQ}, \\ \delta(i, j) &= \frac{\sqrt{\sum_{p=-P}^{p=P} \sum_{q=-Q}^{q=Q} (I(i+p, j+q) - \mu(i, j))^2}}{4PQ}. \end{aligned} \quad (6)$$

In equation (5), $I(i, j)$ represents the original gray value of point (i, j) and $\hat{I}(i, j)$ is the related one after normalization. $\mu(i, j)$ and $\delta(i, j)$ are the mean value and the variance of the window centered at the considered point, respectively. P and Q are the width and height of the window. They are generally assigned to be 3. C is a constant. In training the CNN, the samples are divided into three kinds: the training data, the validation data, and the testing samples. Their ratios are 0.6, 0.2, and 0.2, respectively, and the loss function is shown as follows:

$$L = \frac{1}{N} \sum_{n=1}^N \|f(x_n; w) - y_n\|_1, \quad (7)$$

where N denotes the number of image patches, x_n is the input patch, w represents the weight, $f(x_n; w)$ denotes the score computed by the network, and y_n is the label of the input patch. The SGD is taken to train the CNN to compute the optimized w by minimizing the difference between $f(x_n; w)$ and y_n .

4. Experiments and Analysis

4.1. DCNN-IQA of Different Convolution Layers Trained from Scratch. In this experiment, we used ultrasound images to train the DCNN-IQA with 8 convolution layers and 14

convolution layers, respectively. Ultrasound images are from the database established in Section 2, with a total of 478 images, and all the images are divided into three kinds: 60% training data, 20% validation data, and 20% testing data. It should be mentioned that there is no repetition among three kinds of data. After that, we use the linear correlation coefficient (LCC) and Spearman's rank-order correlation coefficient (SROCC) to evaluate the accuracy of IQA results.

$$\text{LCC} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}, \quad (8)$$

$$\text{SROCC} = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}. \quad (9)$$

In equations (8) and (9), X and Y denote the subjective score and the predicted score by the DCNN. n is the number of the testing numbers and d_i denotes the difference between the numbers of each image after ranking the computed and original scores.

The results are shown in Figure 6. It can be known that the increase of network layers has no significant effect on the final fitting effect of the model. However, when the number of training epochs is small, the deeper network can learn and acquire image information faster. In the first 40 epochs, the fitting effect of DCNN-IQA-14 is better than that of DCNN-IQA-8, which is consistent with the results in most cases. But due to the increase of computation, the training time of each round is longer in the deeper network. In addition, because of the limited amount of data, the more network layers are, the easier the phenomenon of fitting will appear. Based on these factors, we choose the 8-layer DCNN-IQA network, DCNN-IQA-8, in follow-up studies.

4.2. DCNN-IQA Trained by Transfer Learning. As mentioned above, lots of samples are needed to train the DCNN. However, the medical image samples are usually rare. Moreover, some results in our experiments show that the DCNN may lead to overfitting if there are fewer samples. Inspired by the transfer learning in many other applications, we introduce transfer learning to our application. Specifically, the DCNN-IQA is firstly trained using the labeled optical images in LIVE dataset. During the training, we still randomly take 60% data as the training data, 20% as validation data, and the remaining 20% as the testing data. After that, the second training stage of DCNN is finely tuned by the ultrasound images as mentioned in Section 4.1.

The results are shown in Figure 7. From them, it can be known that transfer learning significantly accelerates the training process. Transfer learning for 20 epochs has reached the LCC value of 100 epochs for direct learning. In transfer learning, the ROCC value not only improves the learning speed significantly but also achieves the result that direct learning cannot achieve. These results all indicate that the learning of natural image quality assessment has a significant improvement on the final fitting effect of ultrasonic image quality evaluation. Transfer learning can effectively solve the

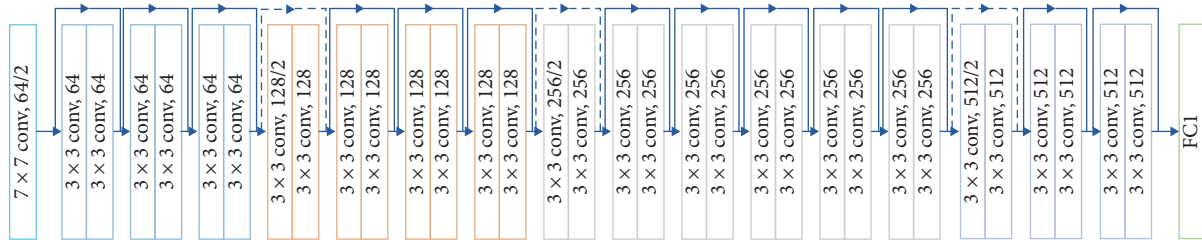


FIGURE 5: The structure of ResNet-IQA.

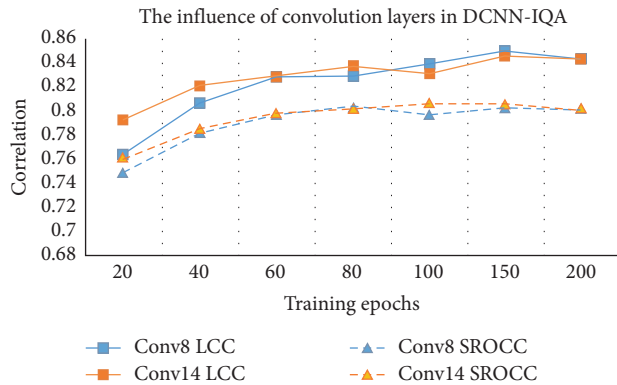


FIGURE 6: The influence of convolutional layers in DCNN-IQA.

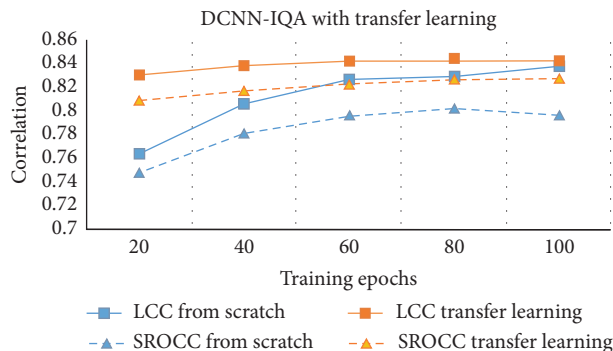


FIGURE 7: The influence of convolutional layers in DCNN-IQA.

network training overfitting problem caused by the limited amount of ultrasound images and improve network performance to a certain extent.

4.3. Comparison with Other Assessment Metrics. In this experiment, we use the ResNet-IQA mentioned in Section 3.2 to do ultrasound image quality assessment. We randomly generate 10 groups' training data. In each group, there are also 478 images from the database, and all the images are divided into three kinds: 60% training data, 20% validation data, and 20% testing data. The results are shown in Table 1.

LCC and SROCC are two error metrics to evaluate the IQA model by comparison with the scores of the experienced doctors. According to the above experiments, our IQA conforms to the doctor's evaluation. Meanwhile, the PSNR and SSIM are usually used in image processing. As such, they

are used to evaluate the quality of ultrasound image in this paper. Afterwards, we compare the parameters of PSNR and SSIM with the doctor's scores. Their correlations are evaluated in metrics of LCC and SROCC. The comparisons are shown in Table 2. Finally, for the 10 groups of experiments, the average LCC is 0.832 and the average SROCC is 0.797. In the table, *w/o* indicates no transfer learning and *w* indicates transfer learning. It should be mentioned that the larger the score of the image in the database is, the worse the image quality is. Therefore, the calculated PSNR and SSIM values are negatively correlated with subjective score. According to the data in the table, the traditional assessment methods of PSNR and SSIM have poor correlation with the subjective assessment of doctors; and the consistency between the CNN-based assessment method and the subjective quality score is far better than the traditional method, while the deep CNN assessment model based on transfer learning further improves the accuracy of the assessment on the basis of the

TABLE 1: 10 groups' test results of ResNet-IQA.

Group	1	2	3	4	5	6	7	8	9	10
LCC	0.770	0.867	0.864	0.826	0.711	0.884	0.863	0.809	0.863	0.860
SROCC	0.749	0.780	0.833	0.812	0.679	0.873	0.828	0.764	0.815	0.836

TABLE 2: Comparison of CNN-based and traditional IQA methods.

Index	PNSR	SSIM	DCNN-IQA		ResNet-IQA
			w/o	w	
LCC	-0.615	-0.517	0.754	0.787	0.832
SROCC	-0.609	-0.608	0.732	0.797	0.797

ordinary deep CNN model. ResNet-IQA model performs best in linear correlation, and it is the same as DCNN-IQA in transfer learning on SROCC.

5. Conclusion

Ultrasound image plays a vital role in medical related applications. How to quantitatively assess the quality of ultrasound image remains an untouched issue. In this paper, CNN is used to evaluate the quality of ultrasound image. Based on the study of optical IQA, we introduce the CNN into the assessment of ultrasound image. We collected ultrasound images from hospitals and websites and established a medical ultrasound image database with subjective score tags. The ultrasound images are captured by different types of equipment. These images are scored by four experienced doctors and the average score is used as the gold standard of image in IQA. Deep CNN is applied to ultrasound IQA task, and the network adjustment and training strategy design are carried out for ultrasound images. The transfer learning strategy is borrowed here to overcome the obstacle of the scarcity of labeled ultrasound samples. Transfer learning can also speed up training and improve IQA accuracy. Meanwhile, we modified the classic classification network for ultrasound IQA. These methods are compared with traditional evaluation methods. The results show that the method based on deep CNN is more reliable than the traditional metrics, and the results of transfer learning and ResNet are better than that of deep CNN.

There is still a long way to go. First, the ultrasound images should be further collected from more channels. Moreover, more experienced doctors will join in scoring the images to obtain the gold standard. Another future research should design a more applicable CNN.

Data Availability

All data will be provided upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the National Key Research and Development Program of China (Grant no.

2017YFBI1303100) and Wuhan Science and Technology Project (Grant no. 2016060101010056).

References

- [1] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality, IEEE," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, USA, June 2011.
- [2] J. Kim, A.-D. Nguyen, and S. Lee, "Deep cnn-based blind image quality predictor," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 11–24, 2019.
- [3] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1202–1213, 2018.
- [4] L. Krasula, P. Le Callet, K. Fliegel, and M. Klima, "Quality assessment of sharpened images: challenges, methodology, and objective metrics," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1496–1508, 2017.
- [5] K. Krupa and M. Bekiesińska-Figatowska, "Artifacts in magnetic resonance imaging," *Polish Journal of Radiology*, vol. 80, pp. 93–106, 2015.
- [6] F. E. Boas and D. Fleischmann, "Ct artifacts: causes and reduction techniques," *Imaging in Medicine*, vol. 4, no. 2, pp. 229–240, 2012.
- [7] S. J. Prabhu, K. Kanal, P. Bhargava, S. Vaidya, and M. K. Dighe, "Ultrasound artifacts," *Ultrasound Quarterly*, vol. 30, no. 2, pp. 145–157, 2014.
- [8] M. C. Hemmsen, M. M. Petersen, S. I. Nikolov, M. B. Nielsen, and J. A. Jensen, "Ultrasound image quality assessment: a framework for evaluation of clinical image quality," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 7629, pp. 76290C–76290C–12, 2010.
- [9] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1733–1740, Columbus, OH, USA, June 2014.
- [10] S. Bosse, D. Maniry, K.-R. Muller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, 2018.
- [11] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 206–220, 2017.
- [12] I. T. Union, *Recommendation 500-10: Methodology for the Subjective Assessment of the Quality of Television Pictures*, ITU-R Rec. BT.500-11, Geneva, Switzerland, 2000.
- [13] Y.-H. Shiao, T.-J. Chen, K.-S. Chuang, C.-H. Lin, and C.-C. Chuang, "Quality of compressed medical images," *Journal of Digital Imaging*, vol. 20, no. 2, pp. 149–159, 2007.
- [14] C. P. Loizou, C. S. Pattichis, M. Pantziaris, T. Tyllis, and A. Nicolaidis, "Quality evaluation of ultrasound imaging in the carotid artery based on normalization and speckle reduction filtering," *Medical & Biological Engineering & Computing*, vol. 44, no. 5, pp. 414–426, 2006.

- [15] H. Zhu, L. Li, J. Wu, W. Dong, and G. Shi, *Metaaiqa: Deep Meta-Learning for No-Reference Image Quality Assessment*, pp. 14131–14140, 2020.
- [16] K. Ma, X. Liu, Y. Fang, and E. Simoncelli, “Blind image quality assessment by learning from multiple annotators. 2019 IEEE International Conference on Image Processing (ICIP),” *Proceedings*, vol. 8, p. 2344, 2019.
- [17] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, “On the use of deep learning for blind image quality assessment,” *Signal, Image and Video Processing*, vol. 12, no. 2, pp. 355–362, 2018.
- [18] G. Litjens, T. Kooi, B. E. Bejnordi et al., “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [19] S. J. Daly, *Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity*, MIT Press, Cambridge, MA, USA, 1992.
- [20] D. Huo, K. A. Salem, Y. Jiang, and D. L. Wilson, “Optimization of spiral mri using a perceptual difference model,” *International Journal of Biomedical Imaging*, vol. 2006, no. 4, p. 35290, 2014.
- [21] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.
- [22] B. Mortamet, M. A. Bernstein, Jack Jr. et al., “Automatic quality assessment in structural brain magnetic resonance imaging,” *Magnetic Resonance in Medicine*, vol. 62, no. 2, pp. 365–372, 2009.
- [23] A. A. Nakhaie and S. B. Shokouhi, “No reference medical image quality measurement based on spread spectrum and discrete wavelet transform using roi processing,” in *Proceedings of the 24th Canadian Conference on Electrical and Computer Engineering (CCECE), Niagara Falls*, pp. 121–125, Canada, May 2011.
- [24] B. L. Eck, R. Fahmi, K. M. Brown et al., “Computational and human observer image quality evaluation of low dose, knowledge-based ct iterative reconstruction,” *Medical Physics*, vol. 42, no. 10, pp. 6098–6111, 2015.
- [25] A. De Angelis, A. Moschitta, F. Russo, and P. Carbone, “Image quality assessment: an overview and some metrological considerations,” in *Proceedings of the 2007 IEEE International Workshop on Advanced Methods for Uncertainty Estimation in Measurement*, p. 47+, July 2007.
- [26] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, “Blind image quality assessment using a deep bilinear convolutional neural network,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 1, pp. 36–47, 2020.
- [27] W. Zhang, K. Ma, G. Zhai, and X. Yang, “Uncertainty-aware blind image quality assessment in the laboratory and wild,” *IEEE Transactions on Image Processing*, vol. 30, pp. 3474–3486, 2021.
- [28] L. Wu, J.-Z. Cheng, S. Li, B. Lei, T. Wang, and D. Ni, “FUIQA: fetal ultrasound image quality assessment with deep convolutional networks,” *IEEE Transactions on Cybernetics*, vol. 47, no. 5, pp. 1336–1349, 2017.
- [29] L. Li, W. Hou, X. Zhang, and M. Ding, “GPU-based block-wise nonlocal means denoising for 3d ultrasound images,” *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 921303, 10 pages, 2013.