

Received March 18, 2021, accepted March 31, 2021, date of publication April 6, 2021, date of current version April 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3071301

Deep Learning in Medical Ultrasound Image Analysis: A Review

YU WANG^{ID1}, XINKE GE², HE MA¹, SHOULIANG QI^{ID1}, GUANJING ZHANG², AND YUDONG YAO^{ID3}, (Fellow, IEEE)

¹College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110819, China

²E-Techco Group, Shenzhen 518102, China

³Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA

Corresponding author: Yudong Yao (yyao@stevens.edu)

ABSTRACT Ultrasound (US) is one of the most widely used imaging modalities in medical diagnosis. It has the advantages of real-time, low cost, noninvasive nature, and easy to operate. However, it also has the unique disadvantages of strong artifacts and noise and high dependence on the experience of doctors. In order to overcome the shortcomings of ultrasound diagnosis and help doctor improve the accuracy and efficiency of diagnosis, many computer aided diagnosis (CAD) systems have been developed. In recent years, deep learning has achieved great success in computer vision with its unique advantages. In the aspect of medical US image analysis, deep learning has also been exploited for its great potential and more and more researchers apply it to CAD systems. In this paper, we first introduce the deep learning models commonly used in medical US image analysis; Second, we review the data preprocessing methods of medical US images, including data augmentation, denoising, and enhancement; Finally, we analyze the applications of deep learning in medical US imaging tasks (such as image classification, object detection, and image reconstruction).

INDEX TERMS Deep learning, medical ultrasound image analysis, ultrasound image preprocessing.

I. INTRODUCTION

Ultrasound (US) is an important part of medical imaging and one of the most commonly used medical diagnostic techniques. It plays an important role in the qualitative and quantitative diagnosis of diseases and clinical examinations. 2D US imaging can be used to observe the morphology and anatomical structures of tissues and organs and detect blood flow and muscle contraction speed; 3D US can observe tissues and organs with complex 3D morphology, such as the heart and fetus (Fig. 1). Compared with other medical imaging technologies, US imaging has the advantages of low cost, convenient, no ionizing radiation, high sensitivity, and real-time imaging. However, compared with X-ray, CT, and MRI, US imaging faces some unique problems. Such as more artifacts and noise, low contrast between tissues causing boundary ambiguities [1] (as shown in Fig. 2), highly subjective diagnosis, and highly dependent on the experience

of doctors. In order to overcome the above problems, it is particularly important to introduce computer-aided diagnosis (CAD) systems. CAD can be used as a supplement to the personal experience and knowledge of doctors, thereby improving the accuracy of US diagnosis.

Deep learning, as a branch of machine learning, is capable of “representation learning” or “feature learning”. After the initial “low-level” feature representation is gradually transformed into “high-level” feature representation through a multi-layer neural network process, classification and other complex tasks can be completed by using a “simple model”, reducing the dependence on physicians [5]. In recent years, deep learning has made breakthroughs in computer vision, speech recognition, natural language processing, and bioinformatics [6], which is hailed as one of the top ten technological breakthroughs in 2013 [7]. Currently, CAD based on deep learning has been applied to various tasks such as diseases classification [8], [9], target detection [10], region of interest (ROI) segmentation [11], and image reconstruction [12] in the analysis of medical US image and other imaging

The associate editor coordinating the review of this manuscript and approving it for publication was Ahmed Farouk^{ID}.

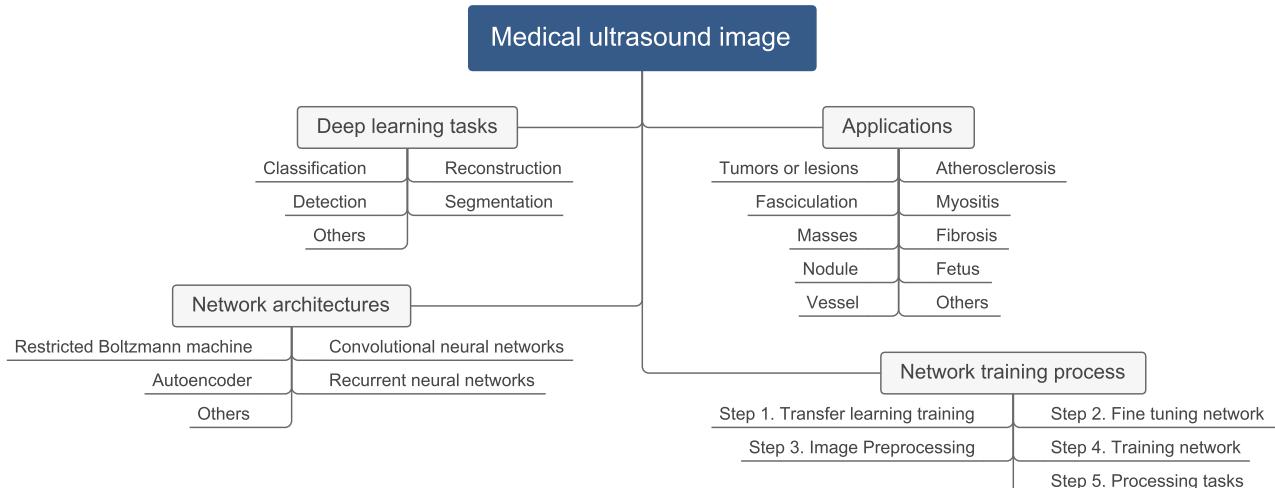


FIGURE 1. Applications of deep learning in medical ultrasound imaging.

TABLE 1. Performance comparison between traditional methods and deep learning methods in classification tasks. PCA: principal component analysis; SVM: support vector machine; HOG: histogram of gradient direction; DPN: deep polynomial network.

Classification	Traditional method				Deep learning method				Paper
	Algorithm	Accuracy	Sensitivity	Specificity	Algorithm	Accuracy	Sensitivity	Specificity	
Tumors or lesions	PCA-SVM	88.80 ± 0.45	89.11 ± 4.21	89.34 ± 4.13	DPN	92.40 ± 1.67	93.84 ± 0.89	90.84 ± 3.15	Liu et al. [2]
Nodules	HOG	0.829	0.699	0.906	CNN	0.931	0.908	0.945	Sun et al. [3]
Fatty liver disease	Multi-SVM	$96.5\% \pm 0.019$	$97.6\% \pm 0.021$	$92.5\% \pm 0.057$	AE	$97.2\% \pm 0.023$	$98\% \pm 0.018$	$95.7\% \pm 0.029$	Hassan et al. [4]

modes analysis. Compared with traditional CAD, which requires manual feature extraction, the deep learning based CAD can automatically extract the low-level and high-level features from the US images based on its deep nonlinear structure, overcoming the limitations of traditional CAD's lack of feature expression capabilities. At present, deep learning has achieved the best performance in many tasks, as shown in Table 1. This paper presents a review of deep learning applications in medical US image analysis, covering deep learning models, data preprocessing methods, and deep learning tasks. We show in Fig. 2 some applications of deep learning in medical ultrasound image analysis.

The rest of the paper is organized as follows. Section II introduces five models of deep learning, including convolutional neural network (CNN), recurrent neural network (RNN), autoencoder (AE), restricted Boltzmann machine (RBM), and transfer learning (TL). The commonly used methods of medical US image preprocessing in deep learning including data augmentation, denoising, and enhancement are presented in Section III. In Section IV, we discuss the specific tasks (classification, detection, segmentation, and reconstruction) of deep learning in medical US image analysis. Finally, conclusions are drawn in Section V.

II. DEEP LEARNING MODELS AND ULTRASOUND DATASETS

A. CONVOLUTIONAL NEURAL NETWORK

CNN was first proposed by LeCun *et al.* [13] in 1989, which was applied to the image recognition of handwritten

characters. In 2012, with AlexNet [14] winning the championship by far surpassing the second place in the ImageNet competition, CNN has advanced rapidly since then, a large number of classic CNNs have emerged. VGGNet [15] focuses on studying the relationship between network depth and performance, combined with a small convolution kernel, and achieved good performance. GoogLeNet [16] explores the problem of multi-scale fusion in convolution calculations to better characterize image information. ResNet [17] introduces the idea of residual in CNN, which alleviates the problem of gradient disappearance caused by deepening of the network. At present, CNN has become the most widely used network structure in deep learning [18], applied in various fields including computer vision. Fig. 3 shows the CNN architectures of GoogLeNet, ResNet, VGGNet.

A typical CNN includes convolutional layers, pooling layers, activation functions, and fully connected layers. Among them, the convolutional layer and pooling layer are responsible for automatically learning the high-level features of an image from the input; the activation function is used to introduce nonlinear transformation into the convolution neural network to solve the problem of insufficient expression ability of the convolution, which is a linear model; the fully connected layer is like a general neural network, which connects each neuron with all the neurons in the upper layer and inputs them to a classifier (such as SoftMax).

In medical US image analysis, CNN is widely used when considering a variety of anatomical organs, such as liver [19],

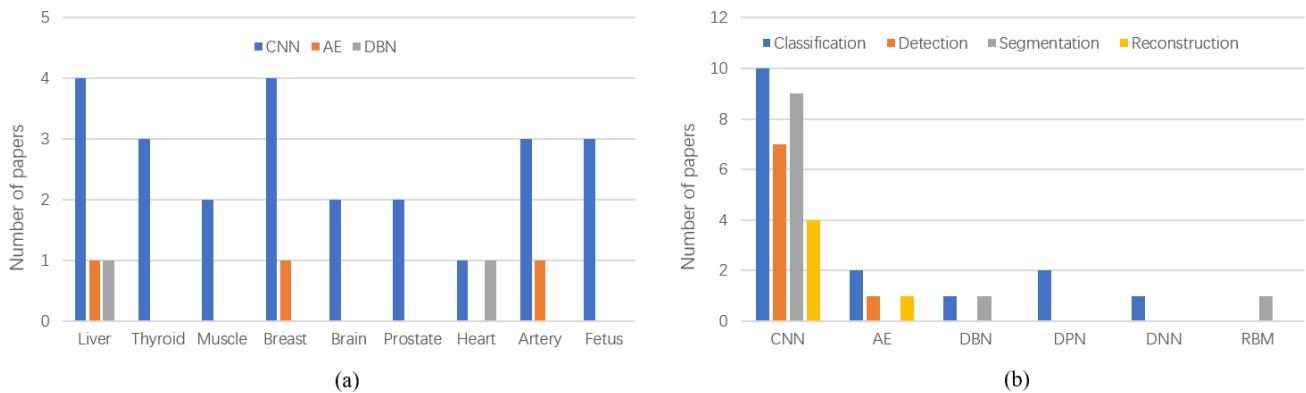


FIGURE 2. (a) Neural networks used in different anatomical structures. (b) Neural networks used in different tasks.

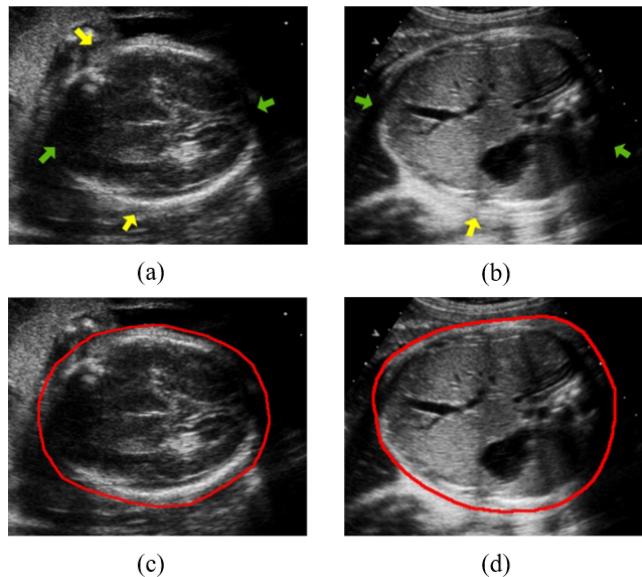


FIGURE 3. (a), (b) represents ultrasound images of fetal head and abdomen respectively. The yellow and green arrows in (a) and (b) denote boundary deficiency and ambiguity, respectively, and the red lines in (c) and (d) denote boundaries.

breast [20], thyroid [21], fetus [22], and artery [23], which achieves impressive diagnosis results.

B. RECURRENT NEURAL NETWORK

In basic neural networks, they process independent input data each time, which implies that they have no memory function. When dealing with problems of data association, we need to introduce the neural network with memory function, which leads to RNN.

RNN introduces a recurrent layer. Each layer in the network can be regarded as a time sequence. The current sequence's state h_t is determined by an input x_t and the previous sequence's state h_{t-1} , while h_{t-1} is determined by x_{t-1} and h_{t-2} . Basically, it shows that h_t is actually determined by all the previous states and input variables, which reflects

the memorability of RNN (see Fig. 4 (a)). However, because RNN considers the impact of all inputs on the time sequence before the current output, the gradient explosion or disappearance problem will occur when RNN is backpropagated, especially when the input sequence is very long. In order to solve this problem, Hochreiter and Schmidhuber [24] proposed a long short-term memory (LSTM) model in 1997. LSTM changes the recurrent layer of RNN, adding three gating units, input gate, forget gate, and output gate. The computation graph of LSTM is shown in Fig. 4 (d). In Fig. 4 (d), σ is Sigmoid function, C_t is the state value at time t , w_i is weight matrix, and b_i is bias. If all the output values of the LSTM are expanded, it will be found that the input value at the earliest moment is not repeatedly multiplied by the w_i , which avoids the problem of difficulty in backpropagation.

In medical US image analysis, we can predict the next sequence from the shape of the previous sequence's tissue structure through RNN, which can be used to optimize the boundary of tissue structure and other tasks. Chen et al. [25] used knowledge transferred RNN (T-RNN) to automatically detect the standard plane of the fetus and achieved an accuracy of 0.908, outperforming state-of-the-art methods. Yang et al. [26] used RNN to predict prostate shape during the study of prostate US image segmentation. LSTM is often used to generate ultrasound image description. Zeng et al. used LSTM to conduct research on ultrasound image caption generation in 2018 [27] and 2020 [28].

C. AUTOENCODER

AE is an unsupervised model, or a self-supervised model. The input data itself is used as a label. AE consists of three parts: input layer, hidden layer, and output layer. The hidden layer is an encoder, whose number of neurons is fewer than that of the input layer. The significance is that less data can be used to represent the input data and realize the feature extraction. The output layer is a decoder, which is responsible for reconstructing the extracted features to be the same as the input data (Fig. 4 (b)). Compared with the traditional

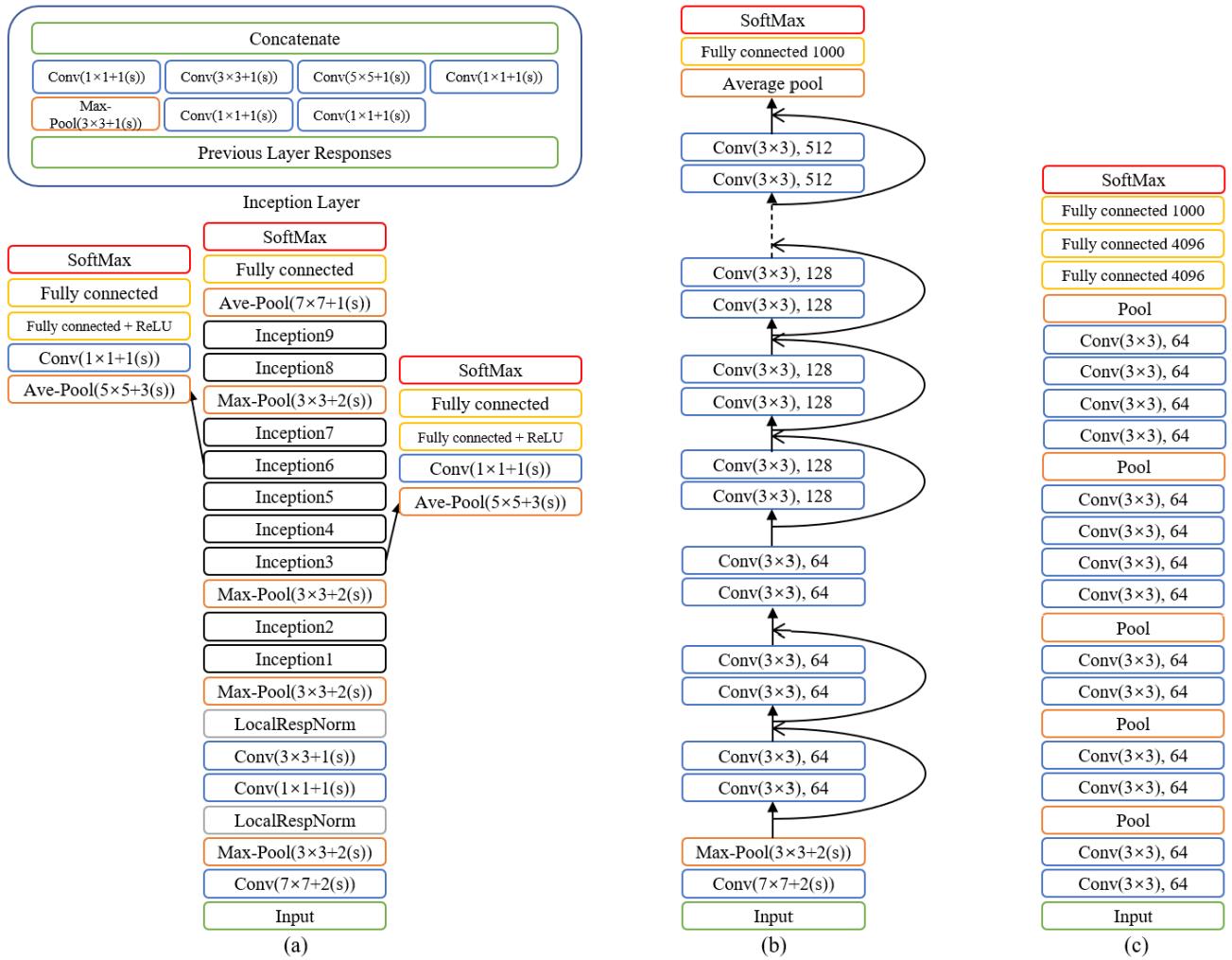


FIGURE 4. CNN architectures of (a) GoogLeNet, (b) ResNet, (c) VGGNet.

principal component analysis (PCA), AE can extract more complex and higher level features. The reason is that PCA is a linear transformation method and AE has the activation function of nonlinear transformation, which can deal with more complex data.

In medical US image analysis, AE and its variants (such as denoising autoencoder [29], sparse autoencoder [30], and the stacked versions of AE) are often used for classification. For example, Cheng *et al.* [31] used CAD based on the stacked deoiling autoencoder (SDAE) to segment the breast lesions. Compared with traditional methods, it significantly improves the performance of CAD. Hassan *et al.* [4] used stacked sparse autoencoder to extract high-level features from the segmented image as the input of classifier SoftMax and the classification accuracy was 97.2%. Combining AE with convolution, convolutional autoencoders (CAE) can be formed [32]. CAE and AE have the same structure, but because of the introduction of convolution, CAE is more suitable for image operation and is often used in image

denoising and feature extraction [33]. Li *et al.* [33] used the improved version of CAE, named denoising convolutional auto-encoder (DCAE), to denoise and extract features from B-mode ultrasound tongue images.

D. RESTRICTED BOLTZMANN MACHINE

RBM is a random neural network whose output of neurons obeys the Boltzmann distribution. The variables in RBM are divided into visible variables and hidden variables. They are all binary variables, that is, only take 0 or 1. The whole RBM is a bipartite graph and the connections between the variables have no direction (Fig. 4 (c)). RBM finds the optimal parameters by maximizing a likelihood function.

In medical US image analysis, multiple RBMs are usually stacked to form a deep Boltzmann machine (DBM) [46]. Due to its multi-layer nature, high-level features of the input data can be extracted, commonly used for classification and segmentation tasks. For example, Zhang *et al.* [47] combined the point-wise gated Boltzmann machine (PGBM) and RBM to

TABLE 2. Ultrasound datasets.

Dataset	Application	Task	Label	Format	Citation
CETUS Challenge	Heart	Segmentation	Defined	mhd/raw	[34]
STACOM	Heart	Track	Defined	DICOM	[35]
ImageCLEF 2015	Various	Classification	Defined	-	[36]
Grand Challenge	Fetus	Measure the fetal head circumference	Defined	CSV	[37]
PICMUS	Image reconstruction	Plane wave imaging	Undefined	-	[38]
P. S. Rodrigues	Breast	Classification	Defined	-	[39]
Dataset B	Breast	Segmentation	Defined	png	[40]
Fatty liver	Liver	Classification	Defined	DICOM	[41]
Thyroid dataset	Thyroid	Classification	Defined	XML	[42]
Muscle dataset	Muscle	Classification	-	-	[43]
BUSI	Breast	Segmentation	Defined	png	[44]
BUSIS	Breast	Segmentation	Defined	png	[45]

classify breast tumors, achieving classification performance with an accuracy of 93.4%; Jaumard-Hakoun *et al.* [48] used RBM to extract the contours of the tongue. No matter RBM or DBM, the neurons are undirected. If the connection direction of the inter layer is specified, DBM can form a deep belief network (DBN).

E. ULTRASOUND DATASETS

US medical datasets are often more difficult to obtain than other datasets. First of all, the annotation of medical images requires significant professional medical knowledge, which makes the annotation very expensive and rare. Secondly, medical data is private and not to be used publicly. Thirdly, different image acquisition devices may lead to different images and medical data parameters. Table 2 shows some publicly available datasets of medical ultrasound images, mainly from competitions and papers. Most datasets exist in CSV format and image (jpg, pgm, png, jpeg and DICOM) format. We selected some datasets for a brief introduction. CETUS Challenge dataset [34] is 3D cardiac US images obtained from 45 patients, the purpose is to segment left ventricular endocardium border. Dataset B [40] uses png format and contains 163 image samples where 106 are benign and 54 are malignant. Thyroid dataset [42] is a set of B-mode ultrasound images containing 329 cases, including several types of lesions such as thyroiditis, cystic nodules, adenomas and thyroid cancers. These lesions accurate contours are provided in XML format. Table 2 shows the details of all datasets.

III. MEDICAL ULTRASOUND IMAGE PREPROCESSING

One of the reasons for the great success of deep learning in various fields is the support of a large number of labeled training samples so that the neural network can obtain good learning performance. However, in medical image analysis, a large number of labeled datasets are difficult to obtain. Relying only on small sample datasets, it is difficult to achieve

satisfactory performance of deep learning in medical image analysis. In order to deal with this problem, the most commonly used method is transfer learning (TL), which has been proved to play an important role in medical image analysis [49] and has been widely used in various tasks. At present, researchers mainly use model transfer to deal with the problem of small amount of labeled data in medical images. A neural network model is pre-trained on other datasets and then applied to medical image analysis tasks. For example, Kermany *et al.* [50] applied TL to develop an artificial intelligence system that can simultaneously diagnose eye diseases and pediatric pneumonia. In addition to model transfer, TL can also be used to reduce the difficulty and costs of data annotation. Zhou *et al.* [51] uses the technology of active learning and TL to implement medical data labeling, which reduces labeling costs by at least half compared with the state-of-the-art methods. In medical US image analysis, TL is frequently used to pre-train neural networks [3], [41], [52]. In addition to TL, feature fusion is also an option to improve the performance of deep learning models [8]. Wang *et al.* [53] used feature fusion in a classification study on COVID-19 to further improve the performance of neural networks. In a study of breast cancer diagnosis, Moon *et al.* [54] fused three types of images and achieved good performance.

Medical US image analysis not only faces the challenge of small amount of labeled data, but also has the disadvantages of a large amount of artifacts and noise. The preprocessing of US images becomes particularly important. Appropriate preprocessing methods can effectively improve the accuracy of the medical US image analysis. The following discusses data preprocessing methods including data augmentation, denoising, and enhancement of medical US images.

A. DATA AUGMENTATION

Data augmentation can partly solve the problem of small amount of medical US image annotation data, thereby

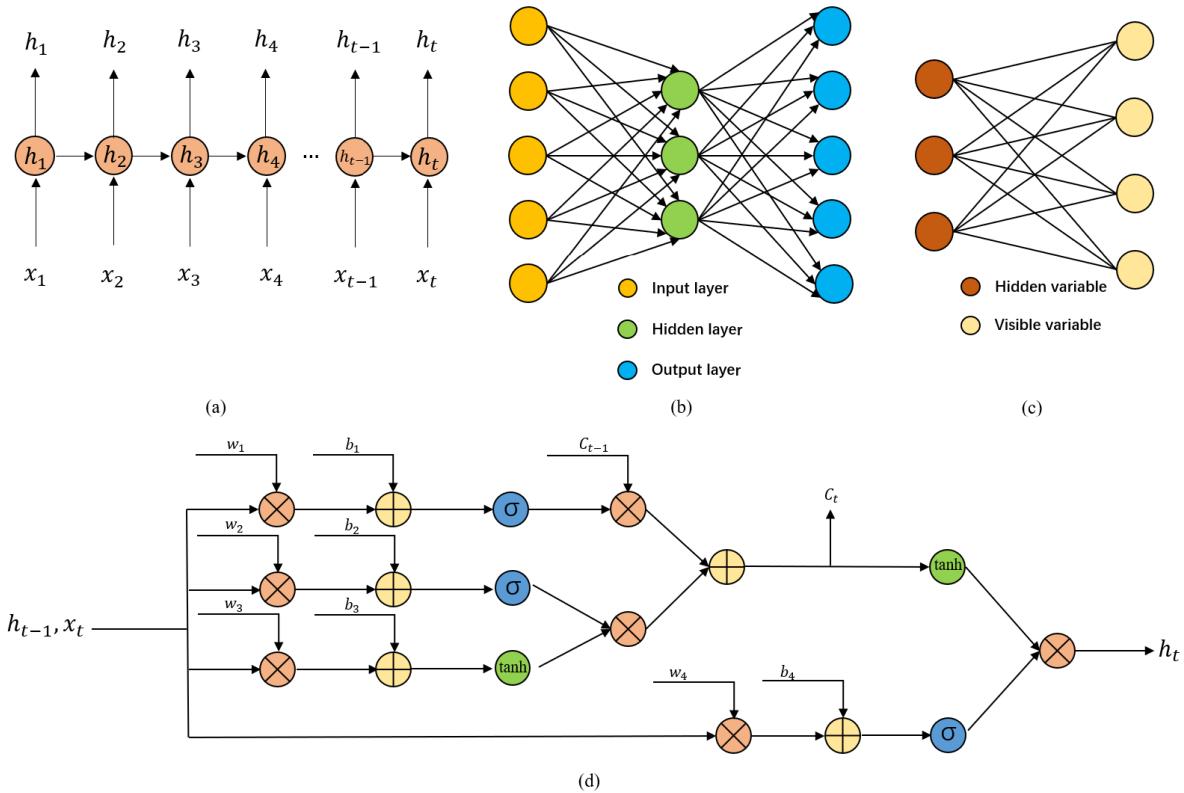


FIGURE 5. (a) The recurrent layer of RNN, (b) AE, (c) RBM, (d) the computation graph of LSTM.

reducing the problem of over fitting and improving the generalization ability of deep learning neural networks. The traditional methods include the use of rotation, random distortion, clipping, and other methods to augment the data. In addition, generative adversarial network (GAN) [55] can be used for data augmentation. Although GAN is more complex than traditional methods, the new data generated by GAN is more abundant and diverse and it can overcome the shortcomings of traditional methods such as the change of image size and position of ROI. In a study of synthetic simulated US images, Tom and Sheet [56] simulated realistic pathological US images by using GAN (Fig. 5).

GAN is composed of a generator and a discriminator. The generator is used to generate samples similar to the real samples. The discriminator is used to determine whether a sample data is a real sample or a generated sample. In the training process of GAN, the samples generated by the generator become more and more similar to the real samples and the judgment ability of the discriminator is also getting stronger and stronger. The generator and the discriminator are constantly confronted and alternately improved. When the system reaches the balance, training ends. The details of the generator and discriminator are introduced below.

For a generator, input a random noise variable z with a probability distribution of $p_z(z)$, learn the probability distribution of z , and generate data $G(z)$ similar to z . The parameters of the generator are represented by θ_g . Use $D(i)$ to

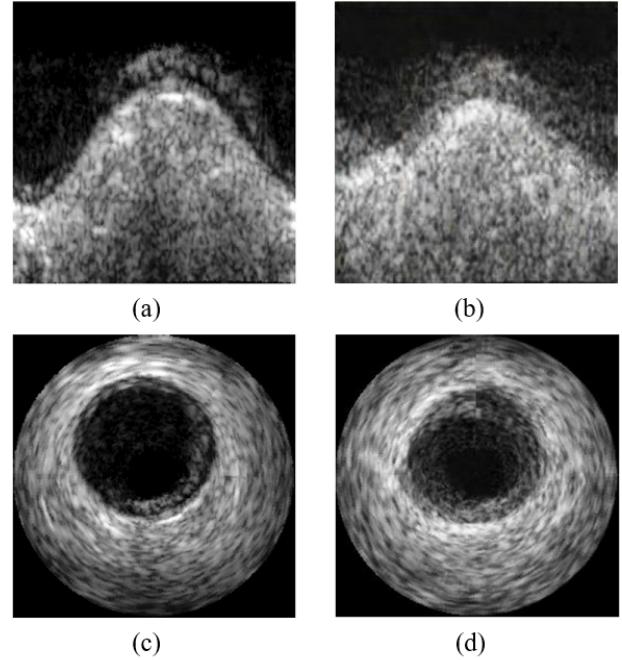


FIGURE 6. (a), (c) represent the original ultrasound images. (b), (d) represent the generated images.

represent the probability of the real sample when the sample i is input. Adjust θ_g so that $D(G(z))$ is closer to 1, that is, minimizing $\ln(1 - D(G(z)))$.

A discriminator will receive the real samples and the generated samples, where the real sample x has a label of 1 and the generated sample $G(z)$ has a label of 0. When the input value is x , the value of $D(x)$ should be closer to 1 by adjusting a parameter θ_d , that is, maximizing $\ln(D(x))$; otherwise, if the input value is $G(z)$, the value of $D(G(z))$ should be closer to 0 by adjusting θ_d , that is, maximizing $\ln(1 - D(G(z)))$.

To summarize, the objective function $\min_G \max_D V(D, G)$ of GAN is as follows,

$$\begin{aligned} & \min_G \max_D V(D, G) \\ &= E_{x-p_{\text{data}}(x)}[\ln D(x)] + E_{z \sim p_z}(z)[\ln(1 - D(G(z)))] \end{aligned} \quad (1)$$

where E is the mathematical expectation, \min_G means to minimize $\ln(1 - D(G(z)))$ and \max_D means to maximize $\ln(D(x))$ and $\ln(1 - D(G(z)))$.

B. DENOISING

US images will produce various noises in the process of acquisition, transmission, and analysis, which is not conducive to the further analysis of images by doctors or CAD systems, so it is necessary to denoise the images. The following mainly introduces three denoising methods, non-local mean filtering, anisotropic diffusion, and U-net.

1) NON-LOCAL MEANS FILTER

Non-Local means filter (NLM) was proposed in 2005 by Buades *et al.* [57], [58]. Compared with the traditional neighborhood filtering, NLM fully considers the influence of the global information of the image on the central pixel instead of just around the central pixel. This filtering method can well eliminate noise while retaining edge details. The main idea is to judge whether the gray values of two pixels are similar based on the Gaussian weighted Euclidean distance and then replace the gray values of the central pixel with the average gray values of similar pixels.

Given an image containing noise $v = \{v(i) | i \in I\}$ where I is the image pixel set, the filtered value $NL[v](i)$ is,

$$NL[v](i) = \sum_{j \in I} \omega(i, j)v(j) \quad (2)$$

where $\omega(i, j)$ is the Gaussian weighted Euclidean distance and its value depends on the similarity between the gray level vectors $v(N_i)$ and $v(N_j)$, where N_k represents a fixed size square neighborhood centered on the pixel k . We have

$$\omega(i, j) = \frac{1}{c_{(j)}} e^{-\frac{\|v(N_i) - v(N_j)\|_2^2}{h^2}} \quad (3)$$

where α is the standard deviation of the Gaussian function greater than 0, which usually takes a value between 2–5. h controls the decay speed of the exponential function. $h = \lambda\sigma$, σ is the standard deviation of the noise. λ usually takes a value between 0.8–1.5. $c_{(j)}$ is a normalization constant,

$$c_{(j)} = \sum_j e^{-\frac{\|v(N_i) - v(N_j)\|_2^2}{h^2}} \quad (4)$$

However, because the neighborhood of a central pixel is to be compared with the whole image, the calculation amount of NLM is very large. Therefore, many scholars have improved NLM, such as Mahmoudi and Sapiro [59] and Liu *et al.* [60].

Sudeep *et al.* [61] and Abraham *et al.* [62] used different improved NLM algorithms to denoise medical US images. The results show that the improved NLM algorithms can effectively remove speckle noise while retaining the boundary information of the images.

2) ANISOTROPIC DIFFUSION

An Anisotropic diffusion model was proposed by Perona and Malik [63] in 1990 and it has good edge retention, even enhancement, and excellent denoising capability. It has been widely used and the idea of the anisotropic diffusion is as follows,

$$\begin{cases} \frac{\partial I}{\partial t} = \operatorname{div}(c(\|\nabla I\|) \cdot \nabla I) \\ I(t=0) = I_0 \end{cases} \quad (5)$$

where div is a divergence operator, $c(\|\nabla I\|)$ is the diffusion equation, ∇ is the gradient operator, $\|\cdot\|$ is modular operation, and t is the time of diffusion. Perona and Malik propose the following two expressions for $c(\|\nabla I\|)$,

$$c(\|\nabla I\|) = e^{-\left(\frac{\|\nabla I\|}{k}\right)^2} \quad (6)$$

$$c(\|\nabla I\|) = \frac{1}{1 + \left(\frac{\|\nabla I\|}{k}\right)^2} \quad (7)$$

In the above two equations, two coefficients of t and k need to be determined. The diffusion time t affects the final smoothing effect and the coefficient k is the gradient threshold. If $\|\nabla I\|$ is much larger than k , $c(\|\nabla I\|)$ tends to 0 and the diffusion is suppressed; if $\|\nabla I\|$ is much smaller than k , $c(\|\nabla I\|)$ tends to 1 and the diffusion is enhanced. Therefore, a large value of k will make a processed image smoother.

The discrete iterative expression of anisotropic diffusion applied to digital images can be expressed as,

$$I_p^{t+1} = I_p^t + \frac{\lambda}{|\partial_p|} \sum_{q \in \partial_p} c(I_p^t - I_q^t) \cdot (I_p^t - I_q^t) \quad (8)$$

where I_p^t represents discrete sampling of the current image, p is the coordinates of the sampled pixels, I_q^t is a neighborhood discrete sample of I_p^t , ∂_p represents the neighborhood space of p , $|\partial_p|$ represents the size of the neighborhood space, and λ is a coefficient that controls the overall diffusion intensity.

Although the anisotropic diffusion can mostly remove noise well while retaining the edges, because of the existence of the gradient threshold k , when the gradient of the noise is large, the anisotropic diffusion may not be able to remove the noise. It may even enhance the noise. For this reason, many scholars have improved this diffusion method, such as Catté *et al.* [64] and Ling and Bovik [65]. In the study of the diagnosis of focal liver lesions [4], during the US image

preprocessing stage, an anisotropic diffusion filter is used to enhance the image and remove noise.

3) U-NET

Deep learning has achieved exciting results in many fields such as computer vision. In recent years, image denoising methods based on deep learning have also attracted the attention of many researchers [66]. U-net [67] is a well-known deep learning network, initially known by researchers for medical image segmentation. Due to its unique network structure, many researchers also use U-net and its modified versions as image denoising [68]. U-net is a type of convolutional neural network, which adopts an encoder-decoder structure. U-net includes four times of downsampling and four times of upsampling. At the same time, it uses skip connections to connect the corresponding feature maps of downsampling and upsampling to ensure that the features of the image are not lost.

Lan and Zhang [69] designed a new neural network based on U-net named MARU for denoising medical US images. MARU integrates residual and mixed-attention mechanism into U-net and good denoising performance is achieved in terms of peak signal to noise ratio, structural similarity.

C. ENHANCEMENT

Ultrasound images often have problems such as low contrast, unclear tissue boundaries, blurred edges, and low resolution. When using a CAD system based on deep learning to analyze ultrasound images, the quality of the image is closely related to the performance of the neural network. Therefore, it is particularly important to use appropriate enhancement preprocessing methods for ultrasound images. This section mainly introduces three medical image enhancement methods, histogram equalization, homomorphic filtering and super-resolution generative adversarial network.

1) HISTOGRAM EQUALIZATION

Histogram equalization is the basis of the digital image spatial processing technology. The idea is to transform an image with an uneven gray value distribution into a new image with a uniform probability density distribution. As a result, the dynamic range of pixel values is expanded, so as to enhance the overall contrast of the image [70]. The idea is,

$$P_r(r_k) = \frac{n_k}{MN}, \quad k = 0, 1, 2, \dots, L - 1 \quad (9)$$

where L is the number of possible gray levels in the image, r is the gray value of the image to be processed, and $r \in [0, L-1]$. $r = 0$ means black, $r = L - 1$ means white, n_k is the number of pixels with a gray value of k , MN is the total number of pixels in the image, and $P_r(r_k)$ is the probability of the gray value r_k appearing. The mapping of r_k in the new histogram after equalization is s_k , we have

$$s_k = T(r_k) = < (L - 1) \sum_{j=0}^k P_r(r_j) >, \quad k = 0, 1, 2, \dots, L - 1 \quad (10)$$

where $T(r_k)$ is the histogram equalization transformation, and $< >$ is the nearest integer operation.

However, the traditional histogram equalization is not suitable for images with a two-level gray distribution. In order to address this issue, we can use adaptive histogram equalization [71] or contrast limited adaptive histogram equalization [72]. Han et al. [73] used histogram equalization to enhance image preprocessing in a study on breast lesion classification.

2) HOMOMORPHIC FILTERING

Homomorphic filtering regards an image as the product of illuminance component and reflection component. Since illuminance can be regarded as incident in the environment, with a small relative change, it can be regarded as the low-frequency component of the image; while with a large relative change of reflectivity, it can be regarded as the high-frequency component. The idea is to reduce the low-frequency and increase the high-frequency, so as to reduce the light changes and sharpen the edges or details. The equation of homomorphic filtering is,

$$f(x, y) = f_i(x, y) \cdot f_r(x, y) \quad (11)$$

where $f(x, y)$ is the image to be processed, $f_i(x, y)$ is the illuminance (incident) component, $f_r(x, y)$ is the reflection component. Since the above equation is the product of two functions, we have

$$\ln f(x, y) = \ln f_i(x, y) + \ln f_r(x, y) \quad (12)$$

After using Fourier transform, (12) can be expressed as

$$F(x, y) = F_i(x, y) + F_r(x, y) \quad (13)$$

after filtering the formula (13) by the filtering function $H(u, v)$, we obtain

$$H(u, v)F(u, v) = H(u, v)F_i(u, v) + H(u, v)F_r(u, v) \quad (14)$$

after filtering, the formula (14) is Fourier inverse transformed to get:

$$h_f(x, y) = h_i(x, y) + h_r(x, y) \quad (15)$$

take the exponential transform of the above formula to get the final result:

$$g(x, y) = e^{h_i(x, y)} e^{h_r(x, y)} \quad (16)$$

Generally, the filter function $H(u, v)$ often takes a Gaussian high-pass filter, Butterworth filter, or exponential filter [74]. However, the filtering function needs many parameters to be determined and it often takes multiple experiments to find relatively reasonable parameters. A lot research has been conducted to improve the filtering function. Duong et al. [75], in a study of segmenting alveolar bone from intraoral US images, first used homomorphic filtering to reduce noise and enhance the image and then used U-Net to segment alveolar bone.

3) SUPER-RESOLUTION GENERATIVE ADVERSARIAL NETWORK

When using US to diagnose obese patients, US needs to penetrate a deeper depth. However, with the increase of penetration depth, it is necessary to make a tradeoff between view, frame rate, and scan line density. Since the diagnosis of US is usually real-time imaging, doctors may need to choose to narrow the field of view or reduce the scan line density to increase the depth of penetration, but both options can reduce the quality and resolution of US imaging. Super-resolution generative adversarial network (SRGAN) [76] is a commonly used super-resolution network. Different from ordinary GAN, the generator of SRGAN does not receive random noise variables but low-resolution images.

As mentioned earlier, GAN has two loss functions, one from the generator and the other from the discriminator, and SRGAN modifies the loss functions of GAN. SRGAN defines a new loss function, which can evaluate the generated image in terms of perceptual characteristics. The newly designed perceived loss L_{SR} is the weighted sum of content loss L_X^{SR} and adversarial loss L_{Gen}^{SR} .

L_X^{SR} consists of two parts, namely, the mean square error L_{MSE}^{SR} of the generated super-resolution image and the label, and the mean square error L_{VGG}^{SR} of the feature map of the generated image and the feature map of label. We have

$$L_X^{SR} = L_{MSE}^{SR} + 10^{-6}L_{VGG}^{SR} \quad (17)$$

where L_{MSE}^{SR} and L_{VGG}^{SR} , respectively,

$$L_{MSE}^{SR} = \frac{1}{r^2WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I^{HR} - I^{LR})^2 \quad (18)$$

$$L_{VGG}^{SR} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR}) - \phi_{i,j}(I^{LR}))^2 \quad (19)$$

In 18 and 19, r , W , and H represent the number of pictures, the width, and height of pictures, respectively. I^{HR} represents label, I^{LR} represents the generated image. In 19, $W_{i,j}$, $H_{i,j}$, and $\phi_{i,j}$ refer to the width, height and feature map of convolution of layer j before the i pooling layer, respectively. L_{Gen}^{SR} is

$$L_{Gen}^{SR} = \sum_{n=1}^N -\log D_{\theta_D}(I^{LR}) \quad (20)$$

where $D_{\theta_D}(x)$ is the output of the discriminator to the variable x , and I^{LR} is the generated image by the generator.

We get L_{SR} ,

$$L_{SR} = L_X^{SR} + 10^{-3}L_{Gen}^{SR} \quad (21)$$

IV. APPLICATIONS OF DEEP LEARNING IN MEDICAL US IMAGES

A. CLASSIFICATION

Classification is one of the most fundamental tasks in medical US image analysis, which can provide doctors with diagnosis suggestions, improve the diagnosis efficiency, and reduce

the influence of subjective factors in the diagnosis process. The main classification tasks consider tumor, lesion, nodule, tissue, and organ fibrosis. The traditional CAD classification methods are to extract the morphological features, or use gray-level co-occurrence matrix, wavelet transform, local binary patterns (LBP) and other methods to extract low-level texture features, and then combine the classifier (such as support vector machine, decision tree, naive Bayes, k nearest neighbor) to classification. However the traditional methods are easy to be affected by possible low imaging quality. Compared with traditional methods, deep learning is able to reduce the impact of low quality US image by extracting high-level features.

Among many deep learning structures, CNN has been most widely used in classification. Classic CNN structures (such as GoogLeNet, VGGNet, AlexNet) are still widely used currently with fine-tuning. In 2017, Han *et al.* [73] conducted a study on the classification of breast lesions using a modified GoogLeNet. In this study, GoogLeNet used ImageNet for pre-training and two auxiliary classifiers were removed. At the same time, the input layer was modified to enable it to process grayscale images. Finally, the original 1000 classes of GoogLeNet output are changed to two classes to correspond to benign and malignant lesions. This study has achieved better classification results than the conventional methods. In a study by Liu *et al.* [3] on thyroid nodules, VGG-F was applied. In this study, the team first pre-trained VGG-F with ImageNet and then input ultrasound images into VGG-F to obtain high-level features. At the same time, the gradient direction histogram (HOG) and scale-invariant feature transformation (SIFT), LBP and other traditional low-level features are integrated with the high-level features extracted from VGG-F, which are stitched into a one-dimensional vector, and finally classified by the positive majority voting strategy. The results show that the hybrid method proposed by the team performs better than the traditional single type feature method. Meng *et al.* [77] conducted a study on liver fibrosis classification. This study first used VGGNet pre-trained on the ILSVRC dataset to extract high-level features, and then used FCN to classify the extracted high-level features.

In addition to the classic CNN, CNNs of other structures, such as DBN, deep polynomial network(DPN), deep neural network(DNN), AE, and deep learning software (DLS) are also used in classification tasks and they have achieved good results. For example, Liu *et al.* [2] utilized stacked DPN to classify tumors; Wu *et al.* [83] used DBN to study focal liver lesions; Hassan *et al.* [4] utilized the sparse autoencoder to extract the features of the images in the diagnosis of focal liver lesions and then put the extracted features into a neural network and used the SoftMax classifier.

Various classification studies using deep learning in medical US images are summarized in Table 3.

B. DETECTION

In medical US image analysis, the purpose of detection is mainly to identify and locate region of interest, and to

TABLE 3. Deep learning in classification tasks.

Network architectures	Anatomical structure	Application	Mode	Number of samples	Performance (%)			Paper
					ACC	Sensitivity	Specificity	
CNN	Liver	FLD	B-mode	63 (27 normal, 36 abnormal)	100	100	100	[19]
			B-mode	550	96.30	100	88.20	[41]
		Fibrosis	-	157 (81 normal, 76 abnormal)	90.60	95.00	85.00	[52]
	Thyroid		-	279 (79 normal, 111 late-stage)	-	-	-	[77]
	Nodule	-	1037 (651 benign, 386 malignant)	93.10	90.80	94.50	[3]	
		-	Database1: 428 (357 positive, 71 negative)	98.29	99.10	93.90	[21]	
	Muscle	Myositis	-	Database2: 164 (357 positive, 71 negative)	96.34	82.80	99.30	
		Fasciculation	-	3214	-	-	-	[43]
		Breast	Tumors&lesions	-	100	100	100	[78]
	Brain	Carotid artery	-	7408 (4254 benign, 3154 malignant)	90.00	86.00	96.00	[73]
		Breast	Tumors&lesions	-	501	89.10	89.00	88.00
AE	Breast	Tumors&lesions	-	1400	-	-	-	[31]
			B-mode	110	97.20	98.00	95.70	[4]
DPN	Various	Tumors&lesions	B-mode	200 (100 benign, 100 malignant)	92.40 \pm 1.67	93.84 \pm 0.89	90.84 \pm 3.15	[2]
			B-mode	200 (100 benign, 100 malignant)	92.40 \pm 1.14	92.67 \pm 3.91	91.36 \pm 5.85	[80]
DLS	Breast	Tumors&lesions	-	632 (550 benign, 82 malignant)	-	84.20	80.30	[81]
			-	2686 (1136 phyllodes tumor, 1550 fibroadenoma)	-	78.00	100.00	[82]
DBN	Liver	Tumors or lesions	B-mode	-	86.36	83.33	87.50	[83]
DNN	Thyroid	Nodule	-	297 (172 benign, 125 malignant)	93.00	-	-	[84]

provide assistance for subsequent medical diagnosis, treatment, or image segmentation.

The detection of tumor or lesion is an important task in US image detection. The commonly used methods can be divided into two categories, one is a two-step method combining candidate region and deep learning, which is represented by R-CNN [88] series; the other is a single step method transforming target detection into a regression problem, which is represented by you only look once (YOLO) [89], and single shot detection (SSD) [90]. In 2017, Li *et al.* [10] detected papillary thyroid carcinoma based on the improved Fast R-CNN, which can automatically detect the papillary thyroid cancer area with 93.5%. Cao *et al.* [20] compared four deep learning models in breast cancer detection, Fast R-CNN, Faster R-CNN, YOLO and SSD, and concluded that SSD has the best accuracy and recall.

US detection of the fetus is also of great significance. It can determine the pregnancy cycle, confirm the condition of the fetus, and look for factors that may affect fetal growth and delivery. In fetal ultrasound detection, the acquisition of a fetal standard plane is the prerequisite and key for subsequent biometric measurement and diagnosis [91]. Chen *et al.* [85] and Wu *et al.* [22] implemented pre-trained CNN to detect the standard plane of the fetus.

A number of deep learning detection studies in medical ultrasound images are summarized in Table 4.

C. SEGMENTATION

The image segmentation can extract the region of interest, which is convenient for the analysis and recognition of medical US images. It is also a prerequisite for quantitative analysis of relevant imaging indicators.

In medical US image segmentation, fully convolutional networks (FCN) is one of the most commonly used network structures. Its main contribution is to use deconvolution instead of fully connected layer to restore the output characteristic image to the original image size, so as to achieve pixel-level classification [105]. Chen *et al.* [96] presented an iterative multi-domain regularized FCN to segment the anatomical structure. When the image has artifacts and speckle noise, even when the boundary of the anatomical structure is not clear, the anatomical structure can still be accurately segmented. Zhang *et al.* [94] utilized Coarse-to-Fine Stacked FCN (CFS-FCN) in a study of segmented lymph nodes and compared with U-Net, CUMedNet [106] and other FCN variants. The results show that CFS-FCN is significantly better than other deep learning methods in segmentation performance.

As one of the most classic segmentation models, U-Net is widely used in US image segmentation and a variety of derivative models have been developed, such as Viksit *et al.* [107] used Multi U-net to effectively segment the breast masses, achieving a mean Dice of 0.82.

TABLE 4. Deep learning in detection tasks.

Network architectures	Application	Anatomical structure	Mode	Number of samples	Performance	Paper
CNN	Tumors&lesions	Thyroid	-	4670	TPR: 0.935 FPR: 0.185 TNR: 0.815 FNR: 0.065	[10]
		Breast	-	1043 (579 benign, 464 malignant)	-	[20]
			-	Dataset A (60 malignant, 246 benign) Dataset B (53 malignant, 110 benign)	TPF: 0.99, F-measure: 0.92 TPF: 0.93, F-measure: 0.88	[40]
	Fetus	Standard plane	B-mode	300 videos	Accuracy: 0.904, Recall: 0.995 Precision: 0.908, F1: 0.950	[85]
				492 videos	-	[22]
	Masses	Brain	B-mode	SNUBH (3891 benign, 1533 malignant) UDIAT (110 benign, 53 malignant)	-	[86]
	Vessel	Various	-	12840	-	[87]
AE	Atherosclerosis	Artery	-	67	-	[23]

TABLE 5. Deep learning in segmentation tasks.

Network architectures	Anatomical structure	Segmented regions	Mode	Number of samples	Performance	Paper
CNN	Fetus	Standard plane	B-mode	Fetal head: 1136 Fetal abdomen: 1193	Fetal head: Dice: 98.43, Jaccard: 96.90 Fetal abdomen: Dice: 96.18, Jaccard: 96.28	[1]
		Volume	3D ultrasound	3,064	Dice: 73.00, Hausdorff distance: 27 mm	[92]
	Prostate	Contour	B-mode	3892	Dice: 93.00, Hausdorff distance: 3 mm	[93]
			B-mode	-	Dice: 93.90 ± 3.5%, Hausdorff distance: 2.7 ± 2.3 mm	[11]
	Lymph	Various	B-mode	80	Mean IU: 86.00, F1: 85.80	[94]
			-	-	Mean IU: 87.90, F1: 87.40	[95]
	Heart	Left ventricle	-	42894	-	[96]
	Artery	Intima-media thickness	B-mode	396	-	[97]
	Brain	ROI	-	162	-	[98]
DBN	Heart	Left ventricle	-	480	TPR: 95.00	[99]
RBM	Tongue	Contour	-	-	Mean Sum of Distances: 0.8 mm	[48]

TABLE 6. Deep learning in reconstruction tasks.

Network architectures	Objective	Mode	Number of samples	Performance	Paper
CNN	3D reconstruction	B-mode	-	-	[100]
	High quality imaging	B-mode	30360	Contrast-to-noise ratio: 1.85 Peak-signal-to-noise ratio: 22.83 dB Structure similarity: 0.86 Reconstruction time: 1	[101]
		-	-	Decorrelation: 0.956	[102]
		-	-	Structural similarity: 0.935	
		Simulation ultrasound	6826	-	[12]
		-	-	-	[103]
AE	High quality imaging	B-mode	-	-	[104]

A number of deep learning segmentation studies in medical US images are summarized in Table 5.

D. RECONSTRUCTION

Improving the quality of US imaging can effectively improve the accuracy and efficiency of diagnosis. The traditional

approach to improve the quality of US images is to use compressed sensing (CS) approaches. Although CS approaches can improve the quality of US imaging, the improvement is limited. CS approaches usually need to replace the hardware of US imaging machines, which adds additional costs. Yoon *et al.* [101] introduced CNN into the study of US image

reconstruction and the results show that, compared with the traditional CS approaches, this method can improve the US image quality better, and does not require any hardware replacement, and can be applied to any B-mode US system or transducer. Perdios *et al.* [104] used stacked denoising autoencoder (SDAE) to compress and recovery US images. This SDAE consists of four layers. The first layer compresses the ultrasound signal and the next three layers are used for signal reconstruction. The results show that this method is superior compared with other US image reconstruction techniques based on the CS method in terms of reconstruction quality and time consumption.

A number of deep learning reconstruction studies in medical US images are summarized in Table 6.

V. CONCLUSION

Cadieu *et al.* [108] confirmed in a study that deep learning can achieve the same performance as the primate visual inferotemporal cortex, which explains that deep learning has advantages over traditional methods in computer vision tasks. With a simple network structure and automatic learning mode, deep learning can complete various complex medical US image analysis tasks. However, deep learning still faces many problems in medical US image analysis. First of all, compared with other natural image datasets, such as ImageNet, the available datasets in the field of US are very limited, which makes the training of deep learning models difficult. Although traditional methods or TL can be used to alleviate the problem of small sample datasets, it is still not the best way to address this issue. Therefore, it is very important for researchers in the field of deep learning and hospitals to cooperate to find more efficient methods of medical data annotation. Secondly, the generalization ability of deep learning is poor. A model can obtain good performance on the images collected by a specific device. However when different devices are used, the performance will generally decline, which is also an urgent problem to be solved. Third, the innovation of neural network models has slowed down. Many researches have improved performance by simply modifying or superimposing existing neural network models. Fourth, deep learning has poor interpretability. In the field of medical artificial intelligence, one of the major reasons why deep learning is not currently applied on a large scale is its “black-box” effect. Although the neural network is constructed by several very simple mathematical formulas, its output is very complicated, and it is difficult for us to know how the neural network works internally. This will cause the doctor-patient community to have doubts about the diagnosis. Also because of the “black box” effect, it is difficult for researchers to adjust the parameters of the neural network effectively. It usually takes many different attempts to determine a better parameter.

Although deep learning still has many problems in the analysis of medical ultrasound images, its value has long been widely recognized. Deep learning technology can improve the medical environment in areas with underdeveloped

medical resources, improve the diagnostic efficiency of medical institutions. It can also provide doctors with auxiliary diagnosis information before, during, and after the disease treatment. To better integrate the artificial intelligence technology represented by deep learning into ultrasound image analysis, and to deal with the shortcomings of deep learning in practical applications, we need to make the following efforts. First, it is necessary to strengthen the interdisciplinary cooperation between deep learning researchers and medical institutions to reduce the difficulty of obtaining medical data and the difficulty of labeling. At the same time, deep learning researchers can better understand the medical artificial intelligence application needs, and continuously improve the accuracy and ease of use of deep learning technologies. Second, establish a standardized and open large-scale database. Deep learning requires high data quantity and quality. At present, there are relatively few open databases in the field of medical ultrasound data, with different formats and uneven quality, which is not beneficial to the development of data-driven deep learning in the medical ultrasound field.

Generally speaking, deep learning is still in the development stage in the medical ultrasound images. With the advancement of the research in deep learning, the hardware improvement of computer and image acquisition equipment, and the enhanced interdisciplinary cooperation, deep learning will achieve better performance in medical ultrasound image analysis tasks.

REFERENCES

- [1] L. Wu, Y. Xin, S. Li, T. Wang, P.-A. Heng, and D. Ni, “Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation,” in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 663–666.
- [2] X. Liu, J. Shi, and Q. Zhang, “Tumor classification by deep polynomial network and multiple kernel learning on small ultrasound image dataset,” in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Munich, Germany: Springer, 2015, pp. 313–320.
- [3] T. Liu, S. Xie, J. Yu, L. Niu, and W. Sun, “Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 919–923.
- [4] T. M. Hassan, M. Elmogy, and E.-S. Sallam, “Diagnosis of focal liver diseases based on deep learning technique for ultrasound images,” *Arabian J. Sci. Eng.*, vol. 42, no. 8, pp. 3127–3140, Aug. 2017.
- [5] Z. Akkus, J. Cai, A. Boonrod, A. Zeinoddini, A. D. Weston, K. A. Philbrick, and B. J. Erickson, “A survey of deep-learning applications in ultrasound: Artificial intelligence-Powered ultrasound for improving clinical workflow,” *J. Amer. College Radiol.*, vol. 16, no. 9, pp. 1318–1328, Sep. 2019.
- [6] L. Deng and D. Yu, “Deep learning: Methods and applications,” *Found. Trends Signal Process.*, vol. 7, nos. 3–4, pp. 197–387, Jun. 2014.
- [7] G. Wang, “A perspective on deep imaging,” *IEEE Access*, vol. 4, pp. 8914–8924, 2016.
- [8] S.-H. Wang, D. R. Nayak, D. S. Guttry, X. Zhang, and Y.-D. Zhang, “COVID-19 classification by CCSHNet with deep fusion using transfer learning and discriminant correlation analysis,” *Inf. Fusion*, vol. 68, pp. 131–148, Apr. 2021.
- [9] S.-H. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang, and H. Cheng, “Classification of Alzheimer’s disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling,” *J. Med. Syst.*, vol. 42, no. 5, p. 85, May 2018.
- [10] H. Li, J. Weng, Y. Shi, W. Gu, Y. Mao, Y. Wang, W. Liu, and J. Zhang, “An improved deep learning approach for detection of thyroid papillary cancer in ultrasound images,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–12, Dec. 2018.

- [11] D. Karimi, Q. Zeng, P. Mathur, A. Avinash, S. Mahdavi, I. Spadiner, P. Abolmaesumi, and S. E. Salcudean, "Accurate and robust deep learning-based segmentation of the prostate clinical target volume in ultrasound images," *Med. Image Anal.*, vol. 57, pp. 186–196, Oct. 2019.
- [12] M. Feigin, D. Freedman, and B. W. Anthony, "A deep learning framework for single-sided sound speed inversion in medical ultrasound," 2018, *arXiv:1810.00322*. [Online]. Available: <http://arxiv.org/abs/1810.00322>
- [13] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989.
- [14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [18] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, and T. Wang, "Deep learning in medical ultrasound analysis: A review," *Engineering*, vol. 5, no. 2, pp. 261–275, Apr. 2019.
- [19] M. Biswas, V. Kuppili, D. R. Edla, H. S. Suri, L. Saba, R. T. Marinhoe, J. M. Sanches, and J. S. Suri, "Symtosis: A liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm," *Comput. Methods Programs Biomed.*, vol. 155, pp. 165–177, Mar. 2018.
- [20] Z. Cao, L. Duan, G. Yang, T. Yue, Q. Chen, H. Fu, and Y. Xu, "Breast tumor detection in ultrasound images using deep learning," in *Proc. Int. Workshop Patch-Based Techn. Med. Imag.* Quebec City, QC, Canada: Springer, 2017, pp. 121–128.
- [21] J. Chi, E. Walia, P. Babyn, J. Wang, G. Groot, and M. Eramian, "Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network," *J. Digit. Imag.*, vol. 30, no. 4, pp. 477–486, Aug. 2017.
- [22] 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 Trans. Cybern.*, vol. 47, no. 5, pp. 1336–1349, May 2017.
- [23] R.-M. Menchón-Lara, J.-L. Sancho-Gómez, and A. Bueno-Crespo, "Early-stage atherosclerosis detection using deep learning over carotid ultrasound images," *Appl. Soft Comput.*, vol. 49, pp. 616–628, Dec. 2016.
- [24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [25] H. Chen, Q. Dou, D. Ni, J.-Z. Cheng, J. Qin, S. Li, and P.-A. Heng, "Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Munich, Germany: Springer, 2015, pp. 507–514.
- [26] X. Yang, L. Yu, L. Wu, Y. Wang, D. Ni, J. Qin, and P.-A. Heng, "Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 1–7.
- [27] X.-H. Zeng, B.-G. Liu, and M. Zhou, "Understanding and generating ultrasound image description," *J. Comput. Sci. Technol.*, vol. 33, no. 5, pp. 1086–1100, Sep. 2018.
- [28] X. Zeng, L. Wen, B. Liu, and X. Qi, "Deep learning for ultrasound image caption generation based on object detection," *Neurocomputing*, vol. 392, pp. 132–141, Jun. 2020.
- [29] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn. (ICML)*, 2008, pp. 1096–1103.
- [30] Y. Luo and Y. Wan, "A novel efficient method for training sparse auto-encoders," in *Proc. 6th Int. Congr. Image Signal Process. (CISP)*, vol. 2, Dec. 2013, pp. 1019–1023.
- [31] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen, "Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in US images and pulmonary nodules in CT scans," *Sci. Rep.*, vol. 6, no. 1, pp. 1–13, Apr. 2016.
- [32] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proc. Int. Conf. Artif. Neural Netw.* Toronto, ON, Canada: Springer, 2011, pp. 52–59.
- [33] B. Li, K. Xu, D. Feng, H. Mi, H. Wang, and J. Zhu, "Denoising convolutional autoencoder based B-mode ultrasound tongue image feature extraction," in *Proc. ICASSP-IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 7130–7134.
- [34] O. Bernard et al., "Standardized evaluation system for left ventricular segmentation algorithms in 3D echocardiography," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 967–977, Apr. 2016.
- [35] C. Tobon-Gomez et al., "Benchmarking framework for myocardial tracking and deformation algorithms: An open access database," *Med. Image Anal.*, vol. 17, no. 6, pp. 632–648, Aug. 2013.
- [36] J. Kalpathy-Cramer, A. G. S. D. Herrera, D. Demner-Fushman, S. Antani, S. Bedrick, and H. Müller, "Evaluating performance of biomedical image retrieval systems—An overview of the medical image retrieval task at ImageCLEF 2004–2013," *Computerized Med. Imag. Graph.*, vol. 39, pp. 55–61, Jan. 2015.
- [37] T. L. A. van den Heuvel, D. D. Bruijn, C. L. D. Korte, and B. V. Ginneken, "Automated measurement of fetal head circumference using 2D ultrasound images," *PLoS ONE*, vol. 13, no. 8, Aug. 2018, Art. no. e0200412.
- [38] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J. A. Jensen, and O. Bernard, "Plane-wave imaging challenge in medical ultrasound," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2016, pp. 1–4.
- [39] P. S. Rodrigues. (2017). *Breast Ultrasound Image*, Mendeley Data. [Online]. Available: <http://dx.doi.org/10.17632/wmy84gznvgw.1>
- [40] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentis, R. Zwiggelaar, A. K. Davison, and R. Martí, "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 4, pp. 1218–1226, Jul. 2018.
- [41] M. Byra, G. Styczynski, C. Szmigielski, P. Kalinowski, Ł. Michałowski, R. Paluszakiewicz, B. Ziarkiewicz-Wróblewska, K. Zieniewicz, P. Sobieraj, and A. Nowicki, "Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 12, pp. 1895–1903, Dec. 2018.
- [42] L. Pedraza, C. Vargas, F. Narváez, O. Durán, E. Munoz, and E. Romero, "An open access thyroid ultrasound image database," *Proc. SPIE*, vol. 9287, Jan. 2015, Art. no. 92870W.
- [43] P. Burlina, S. Billings, N. Joshi, and J. Albayda, "Automated diagnosis of myositis from muscle ultrasound: Exploring the use of machine learning and deep learning methods," *PLoS ONE*, vol. 12, no. 8, Aug. 2017, Art. no. e0184059.
- [44] W. Al-Dhabayani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, Feb. 2020, Art. no. 104863.
- [45] M. Xian, Y. Zhang, H.-D. Cheng, F. Xu, K. Huang, B. Zhang, J. Ding, C. Ning, and Y. Wang, *A Benchmark for Breast Ultrasound Image Segmentation (BUSIS)*. Infinite Study, 2018. [Online]. Available: <https://books.google.com>
- [46] R. Salakhutdinov and G. Hinton, "Deep Boltzmann machines," in *Proc. Artif. Intell. Statist.*, 2009, pp. 448–455.
- [47] Q. Zhang, Y. Xiao, W. Dai, J. Suo, C. Wang, J. Shi, and H. Zheng, "Deep learning based classification of breast tumors with shear-wave elastography," *Ultrasoundics*, vol. 72, pp. 150–157, Dec. 2016.
- [48] A. Jaumard-Hakoun, K. Xu, P. Roussel-Ragot, G. Dreyfus, and B. Denby, "Tongue contour extraction from ultrasound images based on deep neural network," 2016, *arXiv:1605.05912*. [Online]. Available: <http://arxiv.org/abs/1605.05912>
- [49] A. Menegola, M. Fornaciari, R. Pires, S. Avila, and E. Valle, "Towards automated melanoma screening: Exploring transfer learning schemes," 2016, *arXiv:1609.01228*. [Online]. Available: <http://arxiv.org/abs/1609.01228>
- [50] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.
- [51] Z. Zhou, J. Y. Shin, S. R. Gurudu, M. B. Gotway, and J. Liang, "Active, continual fine tuning of convolutional neural networks for reducing annotation efforts," 2018, *arXiv:1802.00912*. [Online]. Available: <http://arxiv.org/abs/1802.00912>
- [52] D. S. Reddy, R. Bharath, and P. Rajalakshmi, "A novel computer-aided diagnosis framework using deep learning for classification of fatty liver disease in ultrasound imaging," in *Proc. IEEE 20th Int. Conf. E-Health Netw. Appl. Services (Healthcom)*, Sep. 2018, pp. 1–5.

- [53] S.-H. Wang, V. V. Govindaraj, J. M. Górriz, X. Zhang, and Y.-D. Zhang, “COVID-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network,” *Inf. Fusion*, vol. 67, pp. 208–229, Mar. 2021.
- [54] W. K. Moon, Y.-W. Lee, H.-H. Ke, S. H. Lee, C.-S. Huang, and R.-F. Chang, “Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks,” *Comput. Methods Programs Biomed.*, vol. 190, Jul. 2020, Art. no. 105361.
- [55] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [56] F. Tom and D. Sheet, “Simulating patho-realistic ultrasound images using deep generative networks with adversarial learning,” in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1174–1177.
- [57] A. Buades, B. Coll, and J. M. Morel, “A review of image denoising algorithms, with a new one,” *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, Jan. 2005.
- [58] A. Buades, B. Coll, and J.-M. Morel, “Nonlocal image and movie denoising,” *Int. J. Comput. Vis.*, vol. 76, no. 2, pp. 123–139, Feb. 2008.
- [59] M. Mahmoudi and G. Sapiro, “Fast image and video denoising via nonlocal means of similar neighborhoods,” *IEEE Signal Process. Lett.*, vol. 12, no. 12, pp. 839–842, Dec. 2005.
- [60] Y.-L. Liu, J. Wang, X. Chen, Y.-W. Guo, and Q.-S. Peng, “A robust and fast non-local means algorithm for image denoising,” *J. Comput. Sci. Technol.*, vol. 23, no. 2, pp. 270–279, Mar. 2008.
- [61] P. V. Sudeep, P. Palanisamy, J. Rajan, H. Baradaran, L. Saba, A. Gupta, and J. S. Suri, “Speckle reduction in medical ultrasound images using an unbiased non-local means method,” *Biomed. Signal Process. Control*, vol. 28, pp. 1–8, Jul. 2016.
- [62] B. A. Abrahim, Z. A. Mustafa, and Y. M. Kadah, “Modified non-local means filter for effective speckle reduction in ultrasound images,” in *Proc. 28th Nat. Radio Sci. Conf. (NRSC)*, Apr. 2011, pp. 1–8.
- [63] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
- [64] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll, “Image selective smoothing and edge detection by nonlinear diffusion,” *SIAM J. Numer. Anal.*, vol. 29, no. 1, pp. 182–193, Feb. 1992.
- [65] H. Ling and A. C. Bovik, “Smoothing low-SNR molecular images via anisotropic median-diffusion,” *IEEE Trans. Med. Imag.*, vol. 21, no. 4, pp. 377–384, Apr. 2002.
- [66] H. Wu, R. Wang, G. Zhao, H. Xiao, J. Liang, D. Wang, X. Tian, L. Cheng, and X. Zhang, “Deep-learning denoising computational ghost imaging,” *Opt. Lasers Eng.*, vol. 134, Nov. 2020, Art. no. 106183. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0143816620301809>
- [67] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Munich, Germany: Springer, 2015, pp. 234–241.
- [68] M. P. Heinrich, M. Stille, and T. M. Buzug, “Residual U-Net convolutional neural network architecture for low-dose CT denoising,” *Current Directions Biomed. Eng.*, vol. 4, no. 1, pp. 297–300, Sep. 2018. [Online]. Available: <https://www.degruyter.com/view/journals/cdbme/4/1/article-p297.xml>
- [69] Y. Lan and X. Zhang, “Real-time ultrasound image despeckling using mixed-attention mechanism based residual UNet,” *IEEE Access*, vol. 8, pp. 195327–195340, 2020.
- [70] J. Li and J. Wang, “Novel algorithm for image enhancement with histogram enhancement with histogram equalization and MSRCR,” *J. Xidian Univ.*, vol. 41, no. 3, pp. 103–109, 2014.
- [71] A. M. Reza, “Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement,” *J. VLSI Signal Process.-Syst. Signal, Image, Video Technol.*, vol. 38, no. 1, pp. 35–44, Aug. 2004.
- [72] K. Zuiderveld, *Contrast Limited Adaptive Histogram Equalization*. Palo Alto, CA, USA: Graphics Gems, 1994.
- [73] S. Han, H.-K. Kang, J.-Y. Jeong, M.-H. Park, W. Kim, W.-C. Bang, and Y.-K. Seong, “A deep learning framework for supporting the classification of breast lesions in ultrasound images,” *Phys. Med. Biol.*, vol. 62, no. 19, p. 7714, 2017.
- [74] Y. Zhang and M. Xie, “Colour image enhancement algorithm based on HSI and local homomorphic filtering,” *Comput. Appl. Softw.*, 2013.
- [75] D. Q. Duong, K.-C.-T. Nguyen, N. R. Kaipatur, E. H. M. Lou, M. Noga, P. W. Major, K. Punithakumar, and L. H. Le, “Fully automated segmentation of alveolar bone using deep convolutional neural networks from intraoral ultrasound images,” in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 6632–6635.
- [76] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.
- [77] D. Meng, L. Zhang, G. Cao, W. Cao, G. Zhang, and B. Hu, “Liver fibrosis classification based on transfer learning and FCNet for ultrasound images,” *IEEE Access*, vol. 5, pp. 5804–5810, 2017.
- [78] H. Nodera, N. Takamatsu, H. Yamazaki, R. Satomi, Y. Osaki, A. Mori, Y. Izumi, and R. Kaji, “Deep learning for identification of fasciculation from muscle ultrasound images,” *Neurol. Clin. Neurosci.*, vol. 7, no. 5, pp. 267–275, Sep. 2019.
- [79] S. Savaş, N. Topaloğlu, O. Kacıcı, and P. N. Koşar, “Classification of carotid artery intima media thickness ultrasound images with deep learning,” *J. Med. Syst.*, vol. 43, no. 8, p. 273, Aug. 2019.
- [80] J. Shi, S. Zhou, X. Liu, Q. Zhang, M. Lu, and T. Wang, “Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset,” *Neurocomputing*, vol. 194, pp. 87–94, Jun. 2016.
- [81] A. S. Becker, M. Mueller, E. Stoffel, M. Marcon, S. Ghafoor, and A. Boss, “Classification of breast cancer from ultrasound imaging using a generic deep learning analysis software: A pilot study,” *Brit. J. Radiol.*, vol. 91, Dec. 2017, Art. no. 20170576.
- [82] E. Stoffel, A. S. Becker, M. C. Wurnig, M. Marcon, S. Ghafoor, N. Berger, and A. Boss, “Distinction between phyllodes tumor and fibroadenoma in breast ultrasound using deep learning image analysis,” *Eur. J. Radiol. Open*, vol. 5, pp. 165–170, 2018.
- [83] K. Wu, X. Chen, and M. Ding, “Deep learning based classification of focal liver lesions with contrast-enhanced ultrasound,” *Optik*, vol. 125, no. 15, pp. 4057–4063, Aug. 2014.
- [84] D. Chen, J. Niu, Q. Pan, Y. Li, and M. Wang, “A deep-learning based ultrasound text classifier for predicting benign and malignant thyroid nodules,” in *Proc. Int. Conf. Green Informat. (ICGI)*, Aug. 2017, pp. 199–204.
- [85] H. Chen, D. Ni, J. Qin, S. Li, X. Yang, T. Wang, and P. A. Heng, “Standard plane localization in fetal ultrasound via domain transferred deep neural networks,” *IEEE J. Biomed. Health Inform.*, vol. 19, no. 5, pp. 1627–1636, Sep. 2015.
- [86] S. Y. Shin, S. Lee, I. D. Yun, S. M. Kim, and K. M. Lee, “Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images,” *IEEE Trans. Med. Imag.*, vol. 38, no. 3, pp. 762–774, Mar. 2019.
- [87] E. Smistad and L. Løvstakken, “Vessel detection in ultrasound images using deep convolutional neural networks,” in *Deep Learning and Data Labeling for Medical Applications*. Athens, Greece: Springer, 2016, pp. 30–38.
- [88] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
- [89] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
- [90] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “SSD: Single shot multibox detector,” in *Proc. Eur. Conf. Comput. Vis. Amsterdam*, The Netherlands: Springer, 2016, pp. 21–37.
- [91] D. Ni, T. Li, X. Yang, J. Qin, S. Li, C.-T. Chin, S. Ouyang, T. Wang, and S. Chen, “Selective search and sequential detection for standard plane localization in ultrasound,” in *Proc. Int. MICCAI Workshop Comput. Clin. Challenges Abdominal Imag.* Nagoya, Japan: Springer, 2013, pp. 203–211.
- [92] P. Looney, G. N. Stevenson, K. H. Nicolaides, W. Plasencia, M. Molloholl, S. Natsis, and S. L. Collins, “Automatic 3D ultrasound segmentation of the first trimester placenta using deep learning,” in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 279–282.
- [93] E. M. A. Anas, P. Mousavi, and P. Abolmaesumi, “A deep learning approach for real time prostate segmentation in freehand ultrasound guided biopsy,” *Med. Image Anal.*, vol. 48, pp. 107–116, Aug. 2018.

- [94] Y. Zhang, M. T. C. Ying, L. Yang, A. T. Ahuja, and D. Z. Chen, "Coarse-to-fine stacked fully convolutional nets for lymph node segmentation in ultrasound images," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 443–448.
- [95] L. Yang, Y. Zhang, J. Chen, S. Zhang, and D. Z. Chen, "Suggestive annotation: A deep active learning framework for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Quebec City, QC, Canada: Springer, 2017, pp. 399–407.
- [96] H. Chen, Y. Zheng, J.-H. Park, P.-A. Heng, and S. K. Zhou, "Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Athens, Greece: Springer, 2016, pp. 487–495.
- [97] M. Biswas, V. Kuppili, T. Araki, D. R. Edla, E. C. Godia, L. Saba, H. S. Suri, T. Omerzu, J. R. Laird, N. N. Khanna, A. Nicolaides, and J. S. Suri, "Deep learning strategy for accurate carotid intima-media thickness measurement: An ultrasound study on Japanese diabetic cohort," *Comput. Biol. Med.*, vol. 98, pp. 100–117, Jul. 2018.
- [98] F. Milletari, S.-A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, J. Levin, O. Dietrich, B. Ertl-Wagner, K. Bötzeli, and N. Navab, "Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound," *Comput. Vis. Image Understand.*, vol. 164, pp. 92–102, Nov. 2017.
- [99] G. Carneiro, J. C. Nascimento, and A. Freitas, "The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 968–982, Mar. 2012.
- [100] R. Prevost, M. Salehi, S. Jagoda, N. Kumar, J. Sprung, A. Ladikos, R. Bauer, O. Zettinig, and W. Wein, "3D freehand ultrasound without external tracking using deep learning," *Med. Image Anal.*, vol. 48, pp. 187–202, Aug. 2018.
- [101] Y. H. Yoon, S. Khan, J. Huh, and J. C. Ye, "Efficient B-mode ultrasound image reconstruction from sub-sampled RF data using deep learning," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 325–336, Feb. 2019.
- [102] O. Senouf, S. Vedula, G. Zurakhov, A. Bronstein, M. Zibulevsky, O. Michailovich, D. Adam, and D. Blondheim, "High frame-rate cardiac ultrasound imaging with deep learning," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Granada, Spain: Springer, 2018, pp. 126–134.
- [103] R. J. G. van Sloun, O. Solomon, M. Bruce, Z. Z. Khaing, Y. C. Eldar, and M. Mischi, "Deep learning for super-resolution vascular ultrasound imaging," in *Proc. ICASSP-IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1055–1059.
- [104] D. Perdios, A. Besson, M. Ardit, and J.-P. Thiran, "A deep learning approach to ultrasound image recovery," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2017, pp. 1–4.
- [105] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
- [106] H. Chen, X. J. Qi, J. Z. Cheng, and P. A. Heng, "Deep contextual networks for neuronal structure segmentation," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 1–7.
- [107] V. Kumar, J. M. Webb, A. Gregory, M. Denis, D. D. Meixner, M. Bayat, D. H. Whaley, M. Fatemi, and A. Alizad, "Automated and real-time segmentation of suspicious breast masses using convolutional neural network," *PLoS ONE*, vol. 13, no. 5, May 2018, Art. no. e0195816.
- [108] C. F. Cadieu, H. Hong, D. L. K. Yamins, N. Pinto, D. Ardila, E. A. Solomon, N. J. Majaj, and J. J. DiCarlo, "Deep neural networks rival the representation of primate IT cortex for core visual object recognition," *PLoS Comput. Biol.*, vol. 10, no. 12, Dec. 2014, Art. no. e1003963.

• • •