IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 14, 2021, accepted January 31, 2021, date of publication February 12, 2021, date of current version February 23, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3059003

Feature-Guided CNN for Denoising Images From Portable Ultrasound Devices

GUANFANG DONG, YINGNAN MA, AND ANUP BASU, (Senior Member, IEEE)
Department of Computing Science, University of Alberta, Edmonton, AB T6G 2R3, Canada

Corresponding author: Guanfang Dong (guanfang@ualberta.ca)

This work was supported by NSERC Discovery Grant.

ABSTRACT As a non-invasive medical imaging scanning device, ultrasound has greatly increased the efficiency and accuracy of medical diagnosis. In recent years, portable ultrasound is being more widely used for its convenience and lower cost. Patients and physicians can receive the scanned images on their mobile phones at any time via a wireless network with low latency. However, it is difficult for portable ultrasound devices to capture images with the same quality as standard hospital ultrasound image acquisition systems. Usually, the images captured by portable ultrasound equipment have considerable noise. This noise undoubtedly affects the diagnosis of the physician. It is imperative to develop methods to remove the noise while preserving important information in the image. For this reason, we propose a novel denoising neural network model, called Feature-guided Denoising Convolutional Neural Network (FDCNN), to remove noise while retaining important feature information. In order to achieve high-quality denoising results, we employ a hierarchical denoising framework driven by a feature masking layer for medical images. Furthermore, we propose a feature extraction algorithm based on Explainable Artificial Intelligence (XAI) for medical images. Experimental results show that our medical image feature extraction method outperforms previous methods. Combined with the new denoising neural network architecture, portable ultrasound devices can now achieve better diagnostic performance.

INDEX TERMS Biomedical image processing, image denoising, feature extraction, image fusion.

I. INTRODUCTION

Ultrasound devices have several advantages over other medical imaging modalities (e.g., MRI, X-ray and CT). They use an instrumented probe to emit ultrasound signals into the human body [1]. After ultrasound reflection and information processing, tomographic images of the body can be plotted. This process eliminates the need to expose patients and physicians to ionizing radiation [2]. This dramatically reduces the health risks of some of the other medical imaging methods, like X-ray and CT. Also, ultrasound equipment is considerably cheaper and more widely available than other safe methodologies, like MRI.

Unlike other bulky medical imaging systems, handheld portable ultrasound devices can make instant diagnosis possible. Doctors can perform quick tests at a patient's bedside without traveling to a specific medical lab or imaging room. At the same time, portable ultrasound equipment is less expensive to use. For our experiment, the Clarius handheld portable ultrasound device [3] can be purchased for less than $3,000.

Unfortunately, most handheld portable ultrasound devices have poor image acquisition quality. The presence of speckle noise significantly affects the diagnosis. Also, since different brands of portable ultrasound devices have different emphasis on image acquisition, it is difficult to use a simple denoising model that works for all portable ultrasound devices. For the same portable ultrasound device, the noise level may be slightly different depending on the location of the human organ being scanned. All these issues pose several challenges in our work.

Despite the abovementioned difficulties, noise removal is crucial in the field of computer vision, such as for image segmentation [6]. Specifically, organ segmentation can help doctors quickly reconstruct information about human tissues [4], [5]. One of the prerequisites for effective segmentation is having noise free input medical images. Existing methods based on machine learning and interpretable artificial intelligence are already widely used for automatic diagnosis of diseases [7]. If there is noise in the input image, this will inevitably reduce the confidence in the diagnostic

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Liu.

information. Based on the above observations, noise removal from medical images has become an essential step in medical image information processing. As a medical imaging device with quality limitations, portable ultrasound equipment can greatly benefit from the noise removal process.

We present a novel Explainable AI (XAI) based approach for medical image feature extraction. At the same time, we also improve an existing denoising neural network architecture. In the training phase, we adopt a hierarchical noise addition strategy. The noise is added only in regions where no medical feature information is present. After our improvement, this network architecture for normal image denoising achieves excellent performance for medical image denoising. Finally, we combine the medical image feature information with the denoising network. This denoising pipeline ensures that the neural network robustly removes speckle noise while retaining the most meaningful feature information. It is worth mentioning that we subtly utilize the neural network image reconstruction technique to map the medical image feature information. The advantage of this approach is that our proposed feature extraction algorithm can be perfectly adapted to different medical imaging devices. We will describe this innovation in detail in the experimental introduction section.

We call the proposed novel portable ultrasound image denoising network the Feature-guided Denoising Convolutional Neural Network (FDCNN). Experimental results demonstrate that our network outperforms existing denoising methods. For the feature extraction part, compared to traditional methods, our proposed method can accurately detect the important features contained in medical images. This also verifies the superiority of our XAI-based approach over traditional methods for medical image feature extraction. The ultrasound image denoising problem can be solved much better by our proposed method. At the same time, our method has the potential to be adapted for other similar medical devices.

The data used in our experiments is collected with the Clarius handheld ultrasound device. We mainly acquire images of the knuckles and heart regions of the human body. In addition to our proposed novel method, we show the results of existing methods for comparison.

II. REVIEW OF CLASSICAL FILTERING APPROACHES

In this section, we introduce several classical image noise reduction algorithms. These algorithms use different filtering methods to achieve denoising. Overall, different denoising methods have different strengths and weaknesses in terms of removal of different types of noise.

Median filtering is widely considered to be one of the most classical and effective methods for image noise removal [12]. Previous work on noise removal from ultrasound images has also introduced median filtering as a decent method [13]. Since the median value is not affected by the maximum and minimum values of neighboring pixels, median filtering has good performance in removing salt-and-pepper noise. However, when the neighborhood of the median filter is large, the processed image has a strong smoothing effect, with edges not retained properly. Thus, median filtering alone is not sufficient to meet our needs.

Wavelet filtering takes advantage of time-frequency localization and multi-resolution features. These two properties ensure that the wavelet filter can remove noise while retaining signal bursts and image edges. In previous work, Achim *et al.* used a Bayesian estimator to create rules for noise removal from ultrasound images [14]. However, the authors acknowledge that the characteristics of noise vary across signal detection. A perfect model to detect the noise component is difficult to construct using simple statistical models.

Anisotropic diffusion filtering overcomes the drawbacks of Gaussian blur [15]. Broadly speaking, anisotropic diffusion filtering first treats the image as a heat field. As a result, each pixel in the image can be seen as a heat flow. Based on this assumption, if a neighborhood pixel differs significantly from the current pixel, it means that the neighborhood pixel is likely to diffuse. Thus, this neighborhood pixel is most likely a boundary. However, anisotropic diffusion filtering has limitations in medical image processing. First, the edge information of ultrasound images is often fuzzy. As a result, the edge estimation for anisotropic diffusion filtering is often inaccurate. Second, anisotropic diffusion filtering further smoothes the ultrasound image. As a consequence, important medical diagnostic information can be lost.

As a nonlinear filter, bilateral filtering [16] uses a weighted average to represent the intensity of a pixel. Compared to the ordinary Gaussian low-pass filter, which only takes into account the effect of position on the center pixel, bilateral filtering can achieve a smoothing effect with edge preservation and noise reduction. In terms of ultrasound image denoising, bilateral filtering preserves more image edge information than anisotropic diffusion filtering. However, for important diagnostic information, the weighted average method over-smoothes much of the important information. This is unacceptable for medical images.

III. REVIEW OF DEEP LEARNING APPROACHES

In the state-of-the-art, many deep learning approaches have been proposed to remove noise in images. Noise2Self [9] is a typical method, which can be combined with existing neural networks, such as DnCNN, to remove noise in images. Without using the signal prior, the noise estimation and the clean ground truth, it can still remove blind-level noise in images by learning with a single image. The Noise2Self method solves the limitation of clean ground truth and obtains persuasive denoising results.

In addition to the Noise2Self approach, a fast and flexible denoising convolutional neural network (FFDNet) [10] can also remove noise from images. Many existing CNN methods have over-smoothing artifacts during the denoising step, which can make resulting images blurry and result in the loss of important features. Differing from these CNN approaches, FFDNet is a non-blind model. It designs a tunable noise level map, which can control the trade-off between

the noise removal performance and the detail preservation performance. It also involves the down-sampling of sub-images, which can guarantee a good trade-off between inference speed and denoising performance.

In addition to FFDNet, Denoising Convolutional Neural Networks (DnCNN) [11] is another popular CNN architecture for noise removal. DnCNN is modified from the VGG network [18]. More specifically, the size of the convolutional filters in DnCNN is set to 3×3 and pooling layers are removed from the VGG architecture. DnCNN contains three different types of layers: Conv + ReLU, Conv + BN + ReLU and Conv. It takes the Adam algorithm [17] as the gradient-based optimizer, which can guarantee the performance of residual learning.

Similar to DnCNN, Image Restoration Convolutional Neural Network (IRCNN) [30] is also a CNN architecture for noise removal. The network contains seven dilated convolution layers, which can enlarge the receptive field. Batch normalization and residual learning are used to accelerate training, and smaller-size training images are used to solve the problem caused by boundary artifacts. Similar to DnCNN, the IRCNN model also takes the Adam algorithm as the gradient-based optimizer.

IV. THE PROPOSED FEATURE-GUIDED DENOISING CNN MODEL

We propose the Feature-guided Denoising Convolutional Neural Networks (FDCNN) for ultrasound images. As part of its novelty, FDCNN can remove noise hierarchically according to the features with different degrees of importance. The most important features are retained during the fusion step. The remaining features in the original image can also be partly preserved after noise removal, resulting from improvements to the neural network architecture. In particular, we exploit the potential of Guided Backpropagation in feature extraction to retain features. In addition, we utilize feature information to guide noise addition so that the denoising network can achieve better feature preservation. More details are given in the following subsections.

1) FEATURE DETECTION

In the feature extraction phase for ultrasound images, we explore the potential of Guided Backpropagation, as an XAI (Explainable AI) method, for feature extraction. In general, XAI is used to explain the logic behind every decision made by an AI algorithm [21]. This research can be broadly divided into local interpretation and global interpretation. The goal of the Guided Backpropagation algorithm is to map the decision path of a neural network. Thus, the Guided Backpropagation algorithm is a local interpretation method. The Guided Backpropagation algorithm is widely used to explain the classification task of neural networks. In the classification task, key areas of evidence in a picture that are considered for classification are marked out. We have borrowed this special attribute of Guided Backpropagation. That is, if Guided Backpropagation is used in a medical image recovery network, key medical image regions will also be marked during the recovery process. In a broader sense, these marked regions are the feature information for the medical images. With this approach, our feature extraction algorithm outperforms almost all traditional feature extraction methods on ultrasound images.

As opposed to normal backpropagation, Guided Backpropagation limits backpropagation to gradients less than zero [20]. Thus, when the Guided Backpropagation algorithm is used, the partial derivative of the gradient for a particular feature map is given by the following equation.

$$R_i^l = \left(f_i^l > 0\right) \cdot \left(R_i^{l+1} > 0\right) \cdot R_i^{l+1}$$

For comparison, the corresponding formula for ordinary back propagation is:

$$R_i^l = \left(f_i^l > 0\right) \cdot R_i^{l+1}, \text{ where } R_i^{l+1} = \frac{\partial f^{out}}{\partial f_i^{l+1}}$$

The reason for limiting the return to gradients less than 0 is that the portion of the gradients smaller than 0 corresponds to the part that weakens the feature we want to visualize; while our goal is to find the part of the image that maximizes the activation of a feature. Thus, when the above formula is used to pass ReLU, the maximally activated part of the picture can be marked. Compared to traditional feature extraction methods, the neural network feature map captures both low-dimensional and high-dimensional information of a picture. For low-level feature maps, the low-level neurons mainly extract features like edges and stripes. Conversely, for higher-level feature maps, these feature maps mainly capture more abstract, high-dimensional features. By using the Guided Backpropagation algorithm, we can maximize the effect of each layer on the input image to describe features.

In this work, we apply Guided Backpropagation to a medical image recovery network based on the U-net architecture [22]. U-net does not have fully connected operations in the network. On the left side of the network is a series of downsampling operations consisting of convolution and Max Pooling. These downsampling operations are also known as contracting path. The contracting path consists of 4 blocks. Each block uses three convolutions and one Max Pooling to downsample. After each downsampling, the number of feature maps is divided by two, resulting in a feature map of size 32×32. The right side of the network is called the expansive path, which also consists of 4 blocks. Before passing through each block, the size of the Feature Map is multiplied by two by deconvolution. The U-Net network ensures the coherence and accuracy of feature map learning by symmetric coding and decoding as well as propagation path merging. In our experiments, we take the original image as input. Then, the output of the network is compared to the original image for loss calculation. This way, we can train a medical image recovery network to perform well. The reason for the training is to ensure that the Feature Map learns the details in different dimensions that are needed to

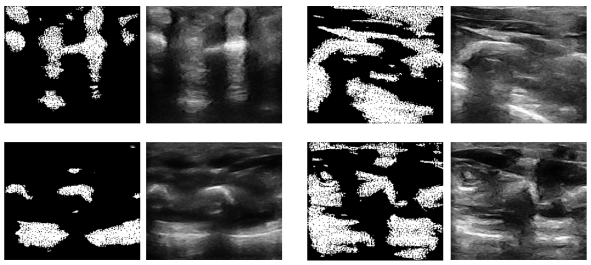

FIGURE 1. Feature detection results based on Guided BackPropagation: (left) the feature mask layer; (right) the original image. We can see that this feature detection method can avoid the undesirable effects of noise.

recover the image. Therefore, the Guided Backpropagation algorithm can accurately capture the feature information of a medical image. Figure 1 shows the results of feature detection for some ultrasound images. We can clearly see that this feature detection method can avoid the effect of noise. At the same time, it preserves the most important part for medical diagnostics.

Here, we explain why the direct application of traditional image feature extraction algorithms, such as FAST (Features from Accelerated Segment Test) [19] or ORB (Oriented FAST and Rotated BRIEF) [23], is less effective. Specifically, two reasons account for this low effectiveness. First, traditional feature extraction methods are less robust in noisy regions. As shown in Figure 2, both the FAST and ORB algorithms mark feature points in the noisy regions of an ultrasound image. This problem is due to the fact that most traditional algorithms only consider local information of the image. For example, the FAST algorithm counts a discretized Bresenham circle with a radius equal to 3 pixels. Then, specific rules are used to decide whether a pixel point is a feature point or not. This undoubtedly leads to some noisy regions with a large area or aggregation being incorrectly identified as feature points. The failure of traditional feature extraction algorithms on noisy ultrasound images also reminds us of the need to devise a feature extraction algorithm that incorporates structure and global information on ultrasound images. Based on this concept, we can maximize the robustness of the algorithm in noisy regions. Second, the feature extraction results of traditional methods are relatively fragmented and independent. This goes against our subsequent denoising framework. The feature extraction results will be used for hierarchical noise addition and image fusion. These two steps require that the feature regions of an ultrasound image are clearly indicated. If the extraction results of a traditional method are used directly, the denoising effect of the neural network will not be improved. Instead, it will be burdened by the addition of incorrect noise addition regions. Thus, using traditional feature extraction methods actually reduces the noise reduction effect. In the experimental results section, we compare the denoising and fusion results for networks trained by different feature extraction algorithms. These comparisons demonstrate the two issues mentioned above. Thus, we propose a feature extraction algorithm for medical images

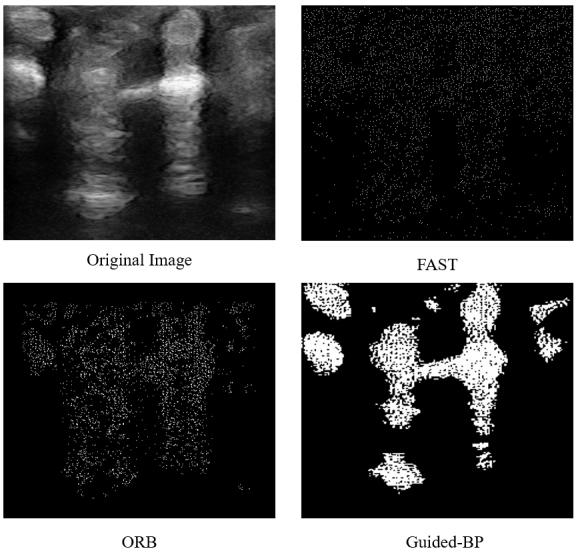

FIGURE 2. Results of different feature extraction algorithms.

based on Guided Backpropagation as one of the important contributions of this experiment [20].

2) NOISE ADDITION

After the features are detected, we need to add Gaussian noise to the original images, guided by the feature information. We implement residual learning [24] so that the dilated convolutional neural network can learn how to remove noise following the differences between the original and the noise-added images. The noise addition step is only involved during training. For testing, FDCNN directly removes the noise in the original ultrasound image without adding noise.

In the state-of-the-art, to achieve residual learning, Gaussian noise is randomly added to the original image. According to residual learning, random noise could be added to the important features of the images, which can mislead a neural network into mapping the features as noise. This effect decreases the clarity of the image features and violates the denoising goal for ultrasound images. This situation is not obvious if the neural network is trained with a low noise level. Once we train the network with high noise levels, the probability of adding noise to features will increase. This can make the neural network learn more about feature removal and make the denoised images fuzzy.

To resolve this problem, we propose the feature-guided architecture to achieve noise removal while preserving image features. Instead of adding noise to the entire image, we utilize the detected features to avoid adding noise on the features. Specifically, we generate random Gaussian noise with an adaptive noise level depending on the size of the original ultrasound image. Given the unknown noise level in ultrasound images, we train the neural network with blind noise levels; i.e., we generate noise within a moderate range instead of a fixed level. Specifically, we train the neural

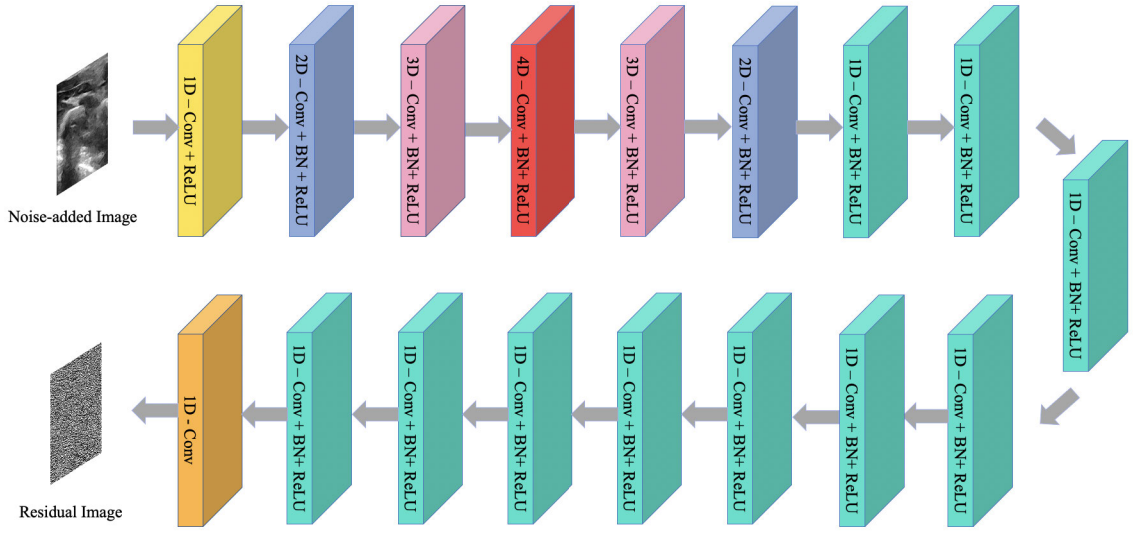

FIGURE 3. The Denoising Network Architecture.

network with a noise level ranging from 15 to 50. Then, we modify the noise array to remove noise that appear on features, so that feature-guided noise can be obtained. After adding feature-guided noise, we utilize the feature-guided noisy images to train the neural network.

It is worth mentioning that the feature guiding step is not only involved in the training and validation steps, it is also utilized during testing. In the testing step, we first use the pre-trained detector model to obtain the important features and save the feature information in a feature array. Then, we remove the noise in the original ultrasound image to obtain the coarse denoised image. To integrate the preserved features with the denoised image, we utilize the Laplacian pyramid to obtain the final feature-guided denoised image. We do not directly use trained models to remove noise in the feature-removed ultrasound image, and add the features back to the resulting image. We abandon this easy approach because the intensity of the resulting image is slightly decreased during noise removal on the feature-removed image. This intensity-decreasing effect is common for denoising techniques, which can make these features stand out after adding the preserved features back to an image. Therefore, we implement the Laplacian pyramid as the image fusion technique, which makes the intensity transition natural. Following this approach, features with different degrees of importance can be preserved. The most important features are retained during the fusion step. The remaining features in the original image can also be partly preserved due to the feature-guided denoising model.

3) FDCNN ARCHITECTURE

The architecture of the FDCNN model is adapted from the VGG network [18] and IRCNN [30]. Compared to the VGG network, we remove max pooling and incorporate dilation in our architecture. The dilated convolution is used to enlarge the receptive field, which can guarantee a larger image region for capturing context information. To make the architecture work well for ultrasound image denoising, we set the depth of the neural network to 17 and the size of convolution filters to 3×3. The number of feature maps for each of the middle layers is set to 64. In addition, batch normalization and residual learning are used to accelerate training, and smaller-size training images are used to solve the problem caused by boundary artifacts. The architecture of the FDCNN denoiser is shown in Fig. 3. In our proposed architecture, FDCNN contains six different types of layers: 1D-Conv + ReLU, 1D-Conv + BN + ReLU, 2D-Conv + BN + ReLU, 3D-conv + BN + ReLU, 4D-conv + BN + ReLU and 1D-Conv. The first layer is a ReLU + Dilated Convolution with a dilation factor of 1, where 64 feature maps are generated by 64 filters with size $3 \times 3 \times 1$. In this case, the receptive field will be increased to 3×3. Following Conv, rectified linear units (ReLU) are involved for non-linearity. To further increase the receptive field, we also increase the dilation factor for the following 5 layers. The second to sixth layers are BN + ReLU + Dilated Convolution with the dilation factors of 2, 3, 4, 3, and 2, respectively. Even though batch normalization can boost the training speed [27], the training time can still be increased significantly due to the increase in the dilation factor. To control the overall training time, we decrease the dilation factor to 1 from the seventh layer with reservation of batch normalization and ReLU. The last layer of the architecture is a Dilated Convolution with a dilation factor of 1, where the output is reconstructed by a $3 \times 3 \times 64$ filter.

4) IMAGE FUSION

Our proposed noise reduction network can accurately remove the noise present in an image. However, the denoised image is usually accompanied by a brightness change in the diagnostic

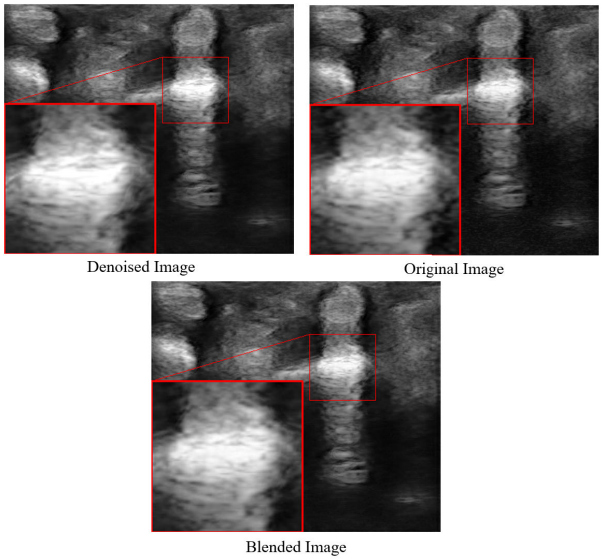

FIGURE 4. In the denoised image, the joint of the hand has some feature information clearly removed. Also, the brightness is increased compared to the original image. In addition, there is a suspicious dashed line. These are some troublesome issues. Our image fusion strategy solves these problems perfectly.

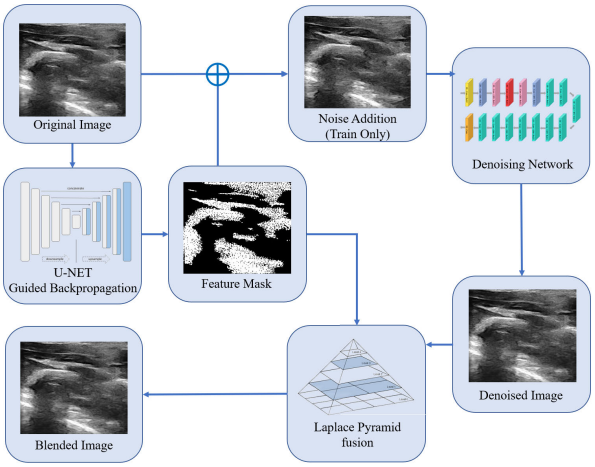

FIGURE 5. This figure explains the overall workflow of FDCNN. First, the original image is extracted with a feature mask layer via a U-net network based on Guided Backpropagation. Then, we add noise to the featureless areas using the mask layer. After that, we feed the image into the noise reduction network and perform residual learning. Then, we merge the feature information and the denoised images by a Laplacian fusion algorithm [26]. Finally, our result can remove noise while preserving feature information in an ultrasound image.

information, an erroneous removal of some important diagnostic information, and an increase of unknown information. Figure 4 shows some of the common problems resulting from deep learning based noise reduction networks. For medical images, these errors can be unacceptable. A small change can result in a wrong diagnosis by a doctor. Based on this consideration, we introduce image fusion for feature information with noise reduced images. Zhang *et al.* devised a method to enhance images through pyramidal fusion [25]. We adopted their design concept to address the above-mentioned common problems.

Figure 5 show the overall workflow of FDCNN. Among all the steps, we focus on the image fusion process. First, the feature extraction image is not a binary map. The brightness of each pixel actually represents the importance of the feature at that location. However, for image fusion, we want the features to have a consistent level of importance. So, we binarized the image first. All pixels greater than 0 are considered as feature regions; this creates a feature mask layer. This mask layer represents the regions that are used for fusion in the original image. For these regions, we use Laplacian pyramids to fuse them into the noise reduced image. The result is an image that preserves diagnostic features while removing noise.

We briefly introduce the Laplacian pyramid fusion algorithm [26]. First, each image involved in the fusion is decomposed into a multi-scale pyramid. The low-resolution images are in the upper layers, while the high-resolution images are in the lower layers. A synthetic pyramid is obtained by fusing the pyramids of all the images at the corresponding layers with certain rules. The synthetic pyramid is then reconstructed according to the inverse process of pyramid generation to obtain the fusion pyramid. For the Laplacian pyramid, we first construct a Gaussian pyramid. Constructing a Gaussian pyramid requires first blurring the image of the next layer with a Gaussian blur. Then, even rows and columns of the blurred image are removed to reduce the image size. For the Laplacian pyramid, the image at each layer is the image of the Gaussian pyramid at the same layer minus the previous layer. Then, the image for each layer will also be Gaussian blurred to obtain the final result. The formula is given below.

$$L_i = G_i - \mathrm{UP}\left(G_{i+1}\right) \otimes \mathcal{G}_{5\times 5}$$

After the above step, we pass in the mask layer. We sum the two images of the Laplacian pyramid according to the mask layer. The result of the summation is a new pyramid. At the same time, we apply the same operation to the Gaussian pyramids of the two images. Each layer of the new pyramid is then up-sampled and summed with the previous layer. The final result is the Laplacian Pyramid Fusion image. Because Gaussian blur is used in the process of building the pyramids, the edges of the fusion are smooth.

V. EXPERIMENTAL RESULTS

In this section, we provide details on how we collect and arrange our data to work on the proposed architecture. Furthermore, we demonstrate the improvements attained by the proposed approach on ultrasound images by comparing our approach to the performance of other networks driven by different feature extraction algorithms. To illustrate the denoising performance, we also compare with five classical filtering methods; namely, median filter, wavelet filter, non-local means filter, anisotropic diffusion filter and bilateral filter. Deep learning approaches, such as DnCNN,

IRCNN, Noise2Self and FFDNet, are also included in the comparisons. To evaluate the denoising performance, we consider both peak signal to noise ratio (PSNR) [28] and structural similarity index (SSIM) [29] for objective evaluation.

A. EXPERIMENTAL SETTINGS

1) DATA

Given the lack of public ultrasound image datasets, we use our Clarius handheld ultrasound device to collect ultrasound images for training and testing. Our data focuses on two parts of the human body: knuckles and the heart. Our data includes 500 ultrasound images, which includes 400 ultrasound images of size 180×180 for training, 46 ultrasound images of size 321×481 for validation and 54 ultrasound images of size 440×380 for testing.

2) ARRANGEMENT

For the experiments, we use separate datasets for training, validation and testing steps. For the filtering methods, we involve the training and validation sets to determine the parameter chosen, and use the testing set to test the denoising performance of the five filtering methods. For existing deep learning methods and our proposed method, we use three datasets separately for training, validation and testing. Following this arrangement, all the approaches are tested on the same testing set so that the evaluation can be quantitative and qualitative.

B. EVALUATION

In this section, we demonstrate the effectiveness of our approach in improving the quality of ultrasound images captured by our handheld device. In order to test the denoising performance, state-of-the-art approaches add Gaussian noise to the original clean image and test the denoising performance on the noise-corrupted image. However, this testing method can only prove the denoising performance on noise-corrupted images instead of real noisy images. In our experiments, we tested our proposed approach on the original ultrasound images. These original ultrasound images are real noisy images directly captured by the device. We calculate the PSNR and SSIM values for the denoised and original images to evaluate the denoising performance. Also, we perform a user study to demonstrate performance based on subjective human evaluations.

1) IMAGE ANALYSIS

In order to illustrate the denoising performance of our proposed method, we first test our feature-guided architecture with two other feature extraction methods. In particular, we compare the performance of FDCNN driven by Guided-BP with networks driven by FAST and ORB. The denoising results of FDCNN driven by different feature extraction algorithms are shown in Fig. 6 and Table. 1. To evaluate the denoising performance on ultrasound images,

TABLE 1. PSNR and SSIM results of FDCNN driven by different feature extraction algorithms.

Methods	Guided-BP	ORB	FAST
PSNR	**0.9801**	0.9698	0.9652
SSIM	**40.3113**	36.1567	35.7888

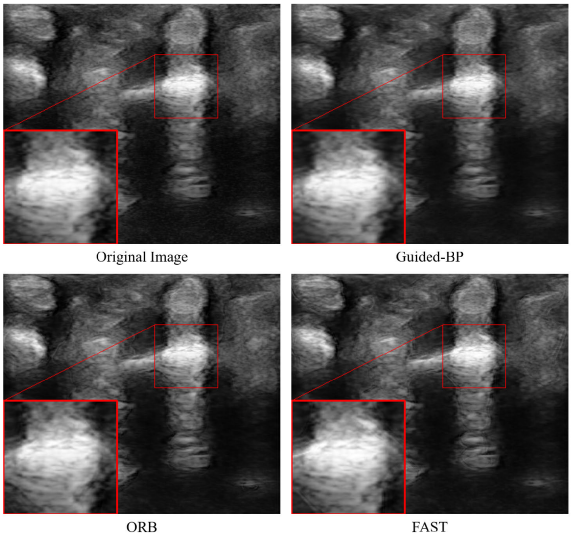

FIGURE 6. Denoising results of FDCNN driven by different feature extraction algorithms.

we need to pay attention to the noise removal performance and feature preservation in the ultrasound images. From Fig. 6, we can see that our proposed architecture (driven by Guided-BP) can successfully remove the noise appearing in the magnified dark area. In addition, the features are clear without over-smoothing artifacts. Compared to our proposed architecture, the networks driven by FAST and ORB cannot achieve promising results. From Fig. 6, we can find that the results of ORB and FAST generate negative artifacts near the bottom left corner and the intensity of the important features are decreased compared to the original image. The reason for these artifacts is that ORB and FAST can only detect limited features. When we utilize these features to guide the noise addition, the possibility that the noise is generated on the undetected features is still high. This situation can mislead the network into mapping the features as noise, which can result in over-smoothing artifacts and lower intensity of features.

In order to demonstrate the performance of our proposed method, we also compare our FDCNN with five traditional filtering methods and the four deep learning methods mentioned in Section 2. The comparison results can be seen in Fig. 7. From Fig. 7 we can see that FDCNN can preserve features quite well. The feature details are preserved, which are very similar to the original image. We can also see that FDCNN successfully removes noise in the magnified dark area. Compared to FDCNN, DnCNN makes the resulting images a little blurred, which can result in loss

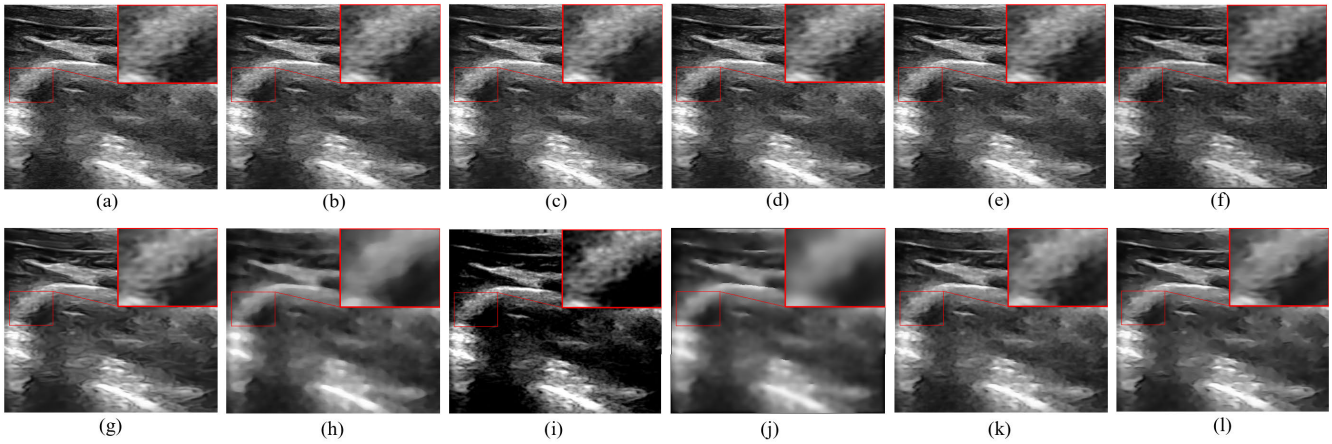

FIGURE 7. Denoising results for different methods: (a) Original image; (b) FDCNN with fusion; (c) FDCNN without fusion; (d) DnCNN; (e) IRCNN; (f) Noise2Self; (g) FFDNet; (h) Median Filter; (i) Wavelet Filter; (j) Anisotropic Diffusion Filter; (k) Bilateral Filter; and (l) Non-local Means Filter.

TABLE 2. PSNR and SSIM results of different methods on testing set.

Methods	FDCNN (with fusion)	FDCNN (without fusion)	DnCNN	IRCNN	Noise2Self	FFDNet	Median Filter	Wavelet Filter	Anisotropic Filter	Bilateral Filter	NLM Filter
PSNR	**40.3113**	**36.1387**	33.5115	33.3894	29.4837	34.1458	32.3649	28.2906	31.3950	34.6056	33.8018
SSIM	**0.9801**	**0.9503**	0.9093	0.9067	0.9086	0.8434	0.6955	0.1973	0.6702	0.8674	0.8183

TABLE 3. Subjective user evaluation results for different methods.

Methods	FDCNN (with fusion)	FDCNN (without fusion)	DnCNN	IRCNN	Noise2Self	FFDNet	Median Filter	Wavelet Filter	Anisotropic Filter	Bilateral Filter	NLM Filter
Overall	**87**	**79**	72	75	62	73	23	37	15	20	47

of details. Compared to DnCNN, IRCNN can preserve the image features better and reduce noise in the dark target area. Compared to our proposed architecture, Noise2Self and FFDNet can also remove the noise quite well. However, these two approaches have over-smoothing artifacts on ultrasound images.

Compared to the filtering methods, our proposed architecture achieves much better denoising performance. From Fig. 7 we can see that the median filter and anisotropic diffusion filter cannot achieve good denoising performance on ultrasound images. Their denoised results are quite blurry, making it difficult to even recognize detailed textures. Compared to the median and anisotropic diffusion filters, the result from the wavelet filter is much clearer, with a legible boundary. However, the wavelet filter also removes considerable feature information near the boundary, which runs counter to the goal of feature preservation. Compared to other filtering methods, the bilateral filter performs the best among the four filtering methods, and provides acceptable denoising performance. From Fig. 7 we can see that the noise in the dark area is reduced. Even if the image feature in the demo image is a little blurry, it is still clearer than the median and anisotropic diffusion filters. Non-local means filter achieves similar denoising performance. The result is slightly blurry compared to the bilateral filter, but it is still better than the median filter. Compared to the five filtering methods and four deep learning methods, our proposed approach has superior denoising performance based on an analysis of the results.

2) STATISTICAL ANALYSIS

In addition to analyzing the result images, we also evaluate the denoising performance by conducting a statistical analysis. From Table 1 we can see that our FDCNN driven by Guided-BP can achieve better denoising performance than the networks driven by ORB and FAST. From Table 2, we can see that our proposed FDCNN also has top rankings. Specifically, performance of FDCNN with fusion slightly exceeds FDCNN without fusion. In addition, DnCNN and IRCNN also provide reliable results with SSIM over 0.9 and PSNR over 33. The Noise2Self method returns higher SSIM than IRCNN with lower PSNR value. FFDNet has opposite performance, producing higher PSNR and lower SSIM. For the filtering methods, Bilateral filter and non-local means filter perform best with PSNR values around 34 and SSIM values around 0.8. From the above statistical analysis, we can see that the proposed feature-guided denoising CNN architectures can achieve excellent denoising performance.

3) SUBJECTIVE USER EVALUATION

In addition to the PSNR and SSIM metrics, we also introduce a study based on human subjective evaluation. For

this evaluation, we select 10 participants. Five of them are from the University of Alberta's Department of Computing Science and have a background in denoising techniques. The remaining five participants are from their families and have no experience with denoising techniques. We present the results of 11 deep learning and traditional filter-based approaches to the participants without telling them the exact method. After viewing all the test images, they are asked to evaluate the results for each method. During the evaluation, we allow subjects to compare with the original image. The comparison scores lie in a range of 0 to 10, where 0 is the worst and 10 is the best. Table 3 shows the results of this user evaluation. Clearly, we can see that FDCNN achieves the leading score without fusing the feature images. After fusion, the final results beat all the methods we tested.

VI. CONCLUSION

We presented a feature-preserving denoising approach for portable ultrasound images. In our algorithm, we first developed a novel feature extraction technique for medical images. This feature extraction technique adapted ideas from interpretable artificial intelligence and the U-net image reconstruction network. Benefiting from the application of guided back-propagation path tracking, we detected the specific location of the features accurately. Following this, we utilized an optimized neural network for strong denoising. We then combined the previously obtained feature images with the Laplacian Pyramid Fusion method. Based on the above methods, we were able to retain the original medical features as much as possible, while strongly denoising the images. At the same time, Laplacian Pyramid Fusion also avoids the problem of sharp edges that can occur during fusion. In the end, we can conclude that noise from portable ultrasound images can be almost perfectly removed by our method. Our research can allow portable ultrasound images to approach medical diagnostic capabilities as large ultrasound devices in medical labs and hospitals.

ACKNOWLEDGMENT

(Guanfang Dong and Yingnan Ma contributed equally to this work.)

REFERENCES

[1] A. Ninomiya *et al.*, "Ultrasonic probe," U.S. Patent Appl. 29/061 038, Jan. 27, 1998.
[2] K. Sakhel, C. B. Benson, L. D. Platt, S. R. Goldstein, and B. R. Benacerraf, "Begin with the basics: Role of 3-dimensional sonography as a first-line imaging technique in the cost-effective evaluation of gynecologic pelvic disease," *J. Ultrasound Med.*, vol. 32, no. 3, pp. 381–388, 2013.
[3] (Dec. 8, 2020). *Portable Pocket Handheld Ultrasound Scanners Clarius (2020)*. [Online]. Available: https://clarius.com/
[4] J. Yang, M. Faraji, and A. Basu, "Robust segmentation of arterial walls in intravascular ultrasound images using dual path U-Net," *Ultrasonics*, vol. 96, pp. 24–33, Jul. 2019.
[5] J. Yang, L. Tong, M. Faraji, and A. Basu, "IVUS-Net: An intravascular ultrasound segmentation network," in *Proc. Int. Conf. Smart Multimedia*. Cham, Switzerland: Springer, Aug. 2018, pp. 367–377.
[6] Y. Ma, G. Dong, C. Zhao, A. Basu, and Z. Wu, "Background subtraction based on principal motion for a freely moving camera," in *Proc. Int. Conf. Smart Multimedia*. Cham, Switzerland: Springer, Dec. 2019, pp. 67–78.
[7] A. R. Hareendranathan, D. Zonoobi, M. Mabee, D. Cobzas, K. Punithakumar, M. Noga, and J. L. Jaremko, "Toward automatic diagnosis of hip dysplasia from 2D ultrasound," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 982–985.
[8] 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.
[9] J. Batson and L. Royer, "Noise2Self: Blind denoising by self-supervision," 2019, *arXiv:1901.11365*. [Online]. Available: http://arxiv.org/abs/1901.11365
[10] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.
[11] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[12] B. I. Justusson, "Median filtering: Statistical properties," in *Two-Dimensional Digital Signal Prcessing II*. Berlin, Germany: Springer, 1981, pp. 161–196.
[13] T. Loupas, W. N. McDicken, and P. L. Allan, "An adaptive weighted median filter for speckle suppression in medical ultrasonic images," *IEEE Trans. Circuits Syst.*, vol. 36, no. 1, pp. 129–135, Jan. 1989.
[14] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Med. Imag.*, vol. 20, no. 8, pp. 772–783, Aug. 2001.
[15] J. Weickert, *Anisotropic Diffusion in Image Processing*, vol. 1. Stuttgart, Germany: Teubner, 1998, pp. 59–60.
[16] M. Elad, "On the origin of the bilateral filter and ways to improve it," *IEEE Trans. Image Process.*, vol. 11, no. 10, pp. 1141–1151, Oct. 2002.
[17] S. Boyd, N. Parikh, and E. Chu, *Distributed Optimization and Statistical Learning Via the Alternating Direction Method of Multipliers*. Boston, MA, USA: Now, 2011.
[18] 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
[19] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, May 2006, pp. 430–443.
[20] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," 2014, *arXiv:1412.6806*. [Online]. Available: http://arxiv.org/abs/1412.6806
[21] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.
[22] 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.* Cham, Switzerland: Springer, Oct. 2015, pp. 234–241.
[23] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571.
[24] W. Jifara, F. Jiang, S. Rho, M. Cheng, and S. Liu, "Medical image denoising using convolutional neural network: A residual learning approach," *J. Supercomput.*, vol. 75, no. 2, pp. 704–718, Feb. 2019.
[25] S. Zhang, C. Euler, and A. Basu, "Image dynamic range enhancement based on fusion pyramid," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2020, pp. 1–4.
[26] A. Toet, "Image fusion by a ratio of low-pass pyramid," *Pattern Recognit. Lett.*, vol. 9, no. 4, pp. 245–253, May 1989.
[27] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 2483–2493.
[28] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, Jun. 2008.
[29] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.
[30] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3929–3938.

GUANFANG DONG received the M.Sc. degree from the Department of Engineering, Hong Kong University of Science and Technology, in 2020. He is currently pursuing the Ph.D. degree in computing science with the University of Alberta, under the supervision of Prof. A. Basu. His research interests include explainable AI, noise removal, and background subtraction.

YINGNAN MA received the B.Sc. degree in computing science from University of Alberta, Edmonton, Canada, in 2019, where he is currently pursuing the M.Sc. degree in computer science, under the supervision of Prof. A. Basu. His research interests include medical image processing and segmentation.

ANUP BASU (Senior Member, IEEE) received the Ph.D. degree in CS from the University of Maryland, College Park, USA. He originated the use of foveation for image, video, stereo, and graphics communication in the early 1990s; an approach that is now widely used in industrial standards. He pioneered the active camera calibration method emulating the way the human eyes work and showed that this method is far superior to any other camera calibration method. He pioneered a single camera panoramic stereo, and several new approaches merging foveation and stereo with application to 3D TV visualization and better depth estimation. His current research applications include multi-dimensional image processing and visualization for medical, consumer and remote sensing applications, multimedia in education and games, and robust wireless 3D multimedia transmission. He has been a Professor with the CS department at UofA since 1999. He has also held the following positions: Visiting Professor, University of California, Riverside, 2003–2004; Guest Professor, Technical University of Austria, Graz, in 1996; Director, Hewlett-Packard Imaging Systems Instructional Laboratory, UofA, 1997 to 2000. He has also been an NSERC, iCORE and Castle Rock Research Chair, and is a Fellow of the American Neurological Association.