Computer Science and Information Systems 22(1):239–257 https://doi.org/10.2298/CSIS240804005H

PSBD-EWT-EGAN: Heart Sound Denoising Using PSBD-EWT and Enhancement Generative Adversarial Network

Jianqiang Hu[1,2], Lin Chen[1], Miao Yang[1], Shigen Shen[3,⋆], and Xiao-Zhi Gao[4]

[1] School of Computer and Information Engineering, Xiamen University of Technology, 361024 Xiamen, P. R. China
hujianqiang@tsinghua.org.cn
275914270c@gmail.com
yangmiao1133@gmail.com
[2] Software School, Quanzhou University of Information Engineering, Quanzhou, P. R. China
[3] School of Information Engineering, Huzhou University, Huzhou, P. R. China
shigens@zjhu.edu.cn
[4] School of Computing, University of Eastern Finland, Kuopio, Finland
xiao-zhi.gao@uef.fi

Abstract. A heart sound signal (HSS) is sensitive to physiological noise and environmental noise, thereby degrading their quality, which makes the accurate diagnosis of machines or doctors difficult and unreliable. To this end, we present a heart sound denoising method using Parameterless Scale-space Boundary Detection (PSBD)-Empirical Wavelet Transform (EWT) and Enhancement Generative Adversarial Network (EGAN) to remove noises that corrupt HSSs in this paper. First, it introduces PSBD and kurtosis to find boundaries delimiting consecutive EWT modes. And then, it further selects the relevant modes on the Pearson's correlation coefficient between each of empirical modes and the original signal to reconstruct HSSs. Finally, EGAN is proposed to improve PSBD-EWT's generalization capacity with regard to different noises. Experimental validation is carried out on PASCAL, MHSDB and WUPHSD databases. The results show that our proposed method achieves significant improvements over state-of-the-art methods. In the case of white Gaussian noise with Signal Noise Ratio (SNR)=5dB, it obtains the best denoising performance under a SNR of 12.53dB and an Root Mean Square Error (RMSE) of 0.034.

Keywords: heart sound signals, denoising, empirical wavelet transform, heart sound signal enhancement.

1. Introduction

As reported by World Health Statistics 2023, in 2019, it is estimated that 17.9 million (UI: 13.4-22.9 million, 27%) people died of chronic diseases. There are 330 million Cardiovascular Disease (CVD) patients, including 11.39 million with Coronary Heart Disease

⋆ Corresponding author

(CHD), 8.9 million with Heart Failure (HF), 5 million with Pulmonary Heart Disease (PHD), and 4.87 million with Atrial Fibrillation (AF) in China [1]. Fortunately, the analysis of Heart Sound Signals (HSSs) has been the primary option for screening and early diagnosis of heart disease. This is because that heart sound contains a great number of biomedical signals related to cardiac activity, which can reveal many pathological heart conditions, such as HF, arrhythmia, and valvular heart disease (VHD).

Traditionally, heart sound was collected from a doctor with a stethoscope. However, manual auscultation brings about great uncertainty and diagnostic delay because it depends on the auscultation skills and experience. Challenges arise due to the presence of heart sound (below 600Hz) inaudible to the human ear. Therefore, Internet of Things (IoT)-based digital stethoscope, due to its non-invasive and easy-to-use nature, is developing rapidly. IoT-based digital stethoscope can record heart sounds and connect to edge clouds for real-time remote analysis and diagnosis [2]. Under uncontrolled environments, it is impossible to capture noise-free heart sound. Heart sounds are affected by respiratory sound, ambient noise and even signals from environments thereby degrading its quality. There is a significant overlapping frequency between heart sounds varying from 20Hz to 800Hz and respiratory sounds varying from 20Hz to 1600Hz. Besides, when people are walking or running, Motion Artifact (MA) is easily introduced to heart sounds. IoT Devices' circuits also may produce Powerline Interference (PLI) and Additive White Gaussian Noise (AWGN). These factors have seriously affected the outcomes of heart sound diagnosis. Due to the chaotic and non-stationary nature of heart sounds and their variability with changes in human body's physical conditions, heart sound denoising is a very complex problem and has become a research hotspot.

Over the past few decades, various techniques have been exploited to denoise heart sounds. First, various statistical methods, including Non-negative Matrix Decomposition (NMD) [3] and Singular Value Decomposition (SVD) [4], are usually used to distinguish HSSs from respiratory sounds, but these methods are difficult in dealing with the difference between the noise and murmurs [5]. Second, various time-domain methods, including conventional filters and auto-correlation methods are exploited to remove noise. Conventional filters such as Butterworth band-pass filters and Finite Impulse Response (FIR) filters can only remove noises outside the range of HSS frequency [6]. Empirical Mode Decomposition (EMD) [7] has good adaptability for non-linear decomposition of HSSs but suffers from modal aliasing and endpoint effect. Although Variational Modal Decomposition (VMD) can overcome modal aliasing but its decomposition performance decreases with the increase of noise intensity. Finally, various frequency-domain, including Wavelet transform (WT) [8], Fourier transform [9] and Empirical Wavelet Transform (EWT) [10] were exploited to denoise HSSs. Wavelet transform depends on predefined parameters including mother wavelet, threshold and DL. Similarly, Fourier transform needs fixed basis functions. Therefore, they lack self-adaptability and are difficult in suppressing the burst noise of HSSs. Compared to WT and Fourier transform, EWT is an adaptive technique which is more suitable to analyze the non-stationary signals. However, it has no improper segmentations of Fourier Transform spectrum due to noise interference, which directly leads to the failure of EWT decomposition. It is worth mentioning that the effectiveness of the above methods can be verified under the controlled clinical environment or simulated noisy conditions. In practice, there are irregular, unpredictable transient distortions of HSSs under uncontrolled environments. Moreover, in the exist-

ing techniques, there is no unified denoising method that can be suitable for multi-noise scenarios.

Notably, deep learning has attracted much attention in various signals denoising in biomedical engineering such as electrocardiogram (ECG) and electroencephalogram (EEG), but has rarely involved in the field of HSSs. To the best of our knowledge, only two works exist in the literature that have exploited deep learning-based denoising methods to remove noises from HSSs, i.e., Denoising Convolutional Neural Network (DnCNN) [11] and LU-Net [12]. DnCNN predicts the residual noise, which is the difference between the noisy HSSs and potential clean HSSs. A deep encoder-decoder-based denoising architecture, called LU-Net, was utilized to suppress respiratory sound and ambient noise that corrupt the heart sounds. Among them, sparse categorical cross-entropy and Mean Square Error (MSE) can be used as loss of useful information to measure the difference between noise-free and processed signals. In practice, the design of the loss function is difficult because of the requirement to capture as many details as possible in the fluctuations of HSSs.

In this work, we develop a new heart sound denoising method called PSBD-EWT-EGAN, using Parameterless Scale-space Boundary Detection (PSBD)-EWT and Enhancement Generative Adversarial Network (EGAN). First, it introduces PSBD and kurtosis to find boundaries delimiting consecutive EWT modes. And then, it selects the relevant modes on the Pearson's correlation coefficient between each of empirical modes and the original signal to reconstruct HSSs. Finally, EGAN architecture with an optimized loss function is presented to further suppress noises rather than directly outputting denoised HSSs of PSBD-EWT. Our proposed method combines the strengths of PSBD-EWT and EGAN in HSSs denoising. On one hand, PSBD-EWT decomposes HSSs to sub-band signals more discriminately than EWT. On the other hand, EGAN can improve PSBD-EWT's generalization capacity with regard to different noises, especially under low SNR (Signal Noise Ratio) environments.

Our contributions are summarized as follows:

- A heart sound denoising algorithm using PSBD-EWT is proposed, in which PSBD and kurtosis overcome the problem of noise component interference to sub-band segmentations of EWT, especially under varying temporal-spectral characteristics of environmental noises and physiological sounds.
- An EGAN architecture with an optimized loss function is proposed, which is capable of better preserving meaningful components of heart sounds while removing noise under a low SNR environment. EGAN makes the denoised HSSs as similar to the clean HSSs as possible, which is beneficial to avoid distortion and enhances the quality of HSSs.
- In order to verify the effectiveness of the proposed method, we have performed several experiments on PASCAL [13], MHSDB [14] and WUPHSD [15] databases. The experimental results show that the proposed method has advantages over the state-of-the-art methods. In the case of white Gaussian noise (WGN) with SNR=5dB, our proposed method obtains the best denoising performance under a SNR of 12.53dB and an RMSE (Root Mean Square Error) of 0.034.

In the following, related work of heart sounds denoising is reviewed in Section 2. Subsequent to this, Section 3 describes our proposed method. Later, the denoising performance of our proposed method is compared with several state-of-the-art methods on

publicly available heart sound datasets in Section 4. At the last, Section 5 concludes this paper.

2. Related work

Denoising methods have been explored in diverse branches of biomedical engineering, including ECG, EEG, respiratory sounds, and heart sounds.

(i) ECG Denoising. An adversarial denoising convolutional neural network (ADnCNN) [16] was exploited for a residual signal from noisy ECG to obtain a clean ECG. In a discriminator network, the denoised and clean signals were classified. The network was fed back to ADnCNN model for parameter adjustment. The work presented in paper [17] employed Convolutional Neural Network (CNN) based Generative Adversarial Network (GAN) model for ECG denoising, which was trained end-to-end using the noisy and clean ECG signals. Recently, a modified lightweight U-Net model called LUNet was exploited to handle various noises that corrupt ECG signals, including baseline wander, muscle artifacts, and AWGN [18]. In order to prevent the loss of effective information when the network is compressed, a deep-wave convolutional neural network called DW-CNN [19] replaces the simple complete layer with a convolutional layer to build an encoder and a decoder. Besides, a Cycle-consistent Generative Adversarial Network (Cycle-GANs) [20] was exploited to improve the quality of ECG recordings suffering from various artifacts, especially for an accurate arrhythmia diagnosis.

(ii) EEG Denoising. One Dimensional Residual Convolutional Neural Network (1D-ResCNN) [21] was exploited to remove the EEG artifacts, which included convolutional layers and Inception-ResNet mapping the noisy EEG signals to the clean ones. Because a single-channel approach might not extract specific spatial information, IC-U-Net [22] combined Independent Component (IC) and U-Net architecture to remove EEG artifacts. Furthermore, IC-U-Net is based on U-Net architecture with a loss function ensemble. Similarly, a framework based on GANs, called EEGANet [23], was applied to eliminate the ocular artifacts from EEG signals. In addition, 1D EEG signal denoising network with 2D transformer, called EEGDnet [24], offered a blend of local and nonlocal self-similarity in feed forward blocks to enhance the denoising performance.

(iii) Respiratory Sound Denoising. Respiratory sounds are recorded in noisy environments, which overlap with the different types of noises. A unique method, using EMD, hurst analysis and spectral subtraction, was proposed to denoise the lung sound [25]. Pouyani et al. [26] developed a method based on WT and Artificial Neural Networks (ANN) to remove noises from respiratory sound signals. Besides, singular spectrum analysis, combining with Discrete Cosine Transform (DCT), was applied to distinguishing BV signals from V signals, and finally enhance the quality of lung sounds [27].

(iv) PCG (Phonocardiogram) & HSSs Denoising. Band-pass filters and Butterworth band-pass filters are common methods of HSSs denoising, which can remove high-frequency and low-frequency noises with cut-off frequencies. WT, EMD, Power Law Algorithm (PLA), Hidden Markov Model (HMM) and Short-time Fourier Transform (STFT) [28] are usually used to denoise HSSs. In particular, WT has advantages in the time-frequency representation of HSSs, but selection of mother wavelet, threshold and DL is inevitable and excellent domain expertise is also crucial. These parameters of the wavelet threshold denoising method were optimized for HSSs. In [29], a noise reduction

method, fusing VMD and the wavelet soft threshold algorithm, was presented to suppress the noise contaminants from children's PCGs, especially crying noise. The method might effectively improve the performance of intelligent screening for CHD. In addition, an adaptive denoising algorithm, combining Time Delay Neural Networks (TDNN) and WT, was employed to denoise mobile PCG [30]. An adaptive noise cancellers-based filter was utilized to denoise and recover the PCG signal corrupted by Gaussian and pink noise [19]. In addition, the GAN-based architecture was used to generate synthetic HSSs, while EWT was used to denoise synthetic cardiac signal and decreased the computational cost that GAN requires [31].

Additionally, deep neural networks have achieved significant results in image noise reduction and speech enhancement, such as Residual Dense Generative Adversarial Network [32,33,34], and Speech Enhancement Generative Adversarial Network (SEGAN) [35][36], but rarely in the field of HSSs noise denoising.

Our work first exploits PSBD and kurtosis to find boundaries delimiting consecutive EWT modes. Furthermore, in the training phase, it takes the noisy and clean HSSs as input in a parallel mode instead of only noisy HSSs in an EGAN. From the above discussion, two novel points highlight the contributions to HSSs denoising.

(i) Our proposed method can separate meaningful and informative components from noisy HSSs, especially under varying temporal-spectral characteristics of environmental noises and physiological sounds.

(ii) The robustness of our proposed method will be better than the existing heart sound denoising methods. Quite a few heart sound denoising methods, such as [37,38], can achieve good performances in a strictly constrained environment. These methods only employ simple filters, such as Butterworth bandpass filtering, high-pass filter, etc., in the pre-processing stage, so their denoising capability is insufficient. PSBD-EWT-EGAN makes full use of clean HSSs and an optimized loss function to achieve heart sounds enhancement even under a low SNR environment.

3. Methodology

3.1. PSBD-EWT

EWT is an adaptive decomposition method in which empirical wavelet is used to decompose each signal into its modes. Because each mode revolves around a specific frequency, EWT can divide the normalized spectrum of each signal into N segmentations to extract the empirical modes. Extracting the bank of empirical wavelets is equivalent to finding a set of bandpass filters. In general, EWT determines boundaries of spectral segmentation based on the local maxima. However, Fourier spectrum is prone to noise and non-stationary factors. Once the segmentation boundaries are set incorrectly, it will directly lead to the failure of EWT decomposition.

In Consequence, we introduce PSBD [39] to find local maxima of time-frequency spectrum. The PSBD-EWT method exploits maxima on a kind of time-frequency spectrum based on the frequency-band segmentations, and the filter bank of orthogonal wavelet was constructed based on the sub-bands to perform EWT decomposition. The method of determining the boundaries of Fourier spectral segmentation is as follows:

Suppose that the classes of a given Fourier spectrum are represented in L levels, we have the total number of classes $N = n_1 + n_2 + \cdots + n_L$, where n_j is the number

of classes at level j. We adopt Otsu's method [40] to find the optimal threshold T such that the intra-class variances σ^2_{intra} of class A and class B are minimal while inter-class variances σ^2_{inter} are maximal. This is equivalent to maximizing σ^2_{intra} and its expression for intra-class variance is given by

$$\sigma^2_{intra} = w_A w_B (\mu_B - \mu_A)^2 \tag{1}$$

where the probabilities of class occurrence of A and B are $w_A = \frac{1}{N}\sum_{j=1}^{T} n_j$, $w_B = \frac{1}{N}\sum_{j=T+1}^{L} n_j$, respectively, and the class mean levels of A and B are $\mu_A = \frac{\sum_{j=1}^{T} jn_j}{\sum_{j=1}^{T} n_j}$, $\mu_B = \frac{\sum_{j=1}^{L} jn_j - \sum_{j=1}^{T} jn_j}{N(1-w_A)}$, respectively.

Obviously, the optimal threshold T is determined automatically in PSBD on the global characteristics of the given Fourier spectrum. Notably, one of the limitations of PSBD is segmentation boundaries under varying temporal-spectral characteristics of environmental noises and body sounds, such as breathing and intestinal sounds. These noises exhibit some important transient characteristics, including peak levels, peak intervals, and peak durations. Kurtosis [41] is a statistical measure, which can fully account for all the three transient characteristics. Therefore, kurtosis is introduced to PSBD-EWT to enhance the noises suppression under uncontrolled environments.

In order to describe a heart sound signals denoising algorithm using PSBD-EWT, we provides the following definitions.

Definition 1. *Empirical scaling $\theta_i(\varpi)$ and empirical wavelet $\varphi_i(\varpi)$ are expressed as*

$$\theta_i(\varpi) = \begin{cases} 1, & |\varpi| \le (1-\xi)\varpi_i \\ \cos(\frac{\pi}{2}\Phi(\xi,\varpi_i)), & (1-\xi)\varpi_i \le |\varpi| \le (1+\xi)\varpi_i \\ 0, & otherwise \end{cases} \tag{2}$$

$$\varphi_i(\varpi) = \begin{cases} 1, & (1+\xi)\varpi_i \le |\varpi| \le (1-\xi)\varpi_{i+1} \\ \cos(\frac{\pi}{2}\Phi(\xi,\varpi_{i+1})), & (1-\xi)\varpi_{i+1} \le |\varpi| \le (1+\xi)\varpi_{i+1}, \\ \sin(\frac{\pi}{2}\Phi(\xi,\varpi_i)), & (1-\xi)\varpi_i \le |\varpi| \le (1+\xi)\varpi_i \\ 0, & otherwise \end{cases} \tag{3}$$

where $\Phi(\xi,\varpi_i) = \Phi(\frac{|\varpi|-(1-\xi)\varpi_i}{2\xi\varpi_i})$, ξ ensures that there is no overlapping between the empirical scaling functions and empirical wavelets, and ϖ_i denotes i^{th} boundary frequency.

Algorithm 1:Heart sound denoising using PSBD-EWT

Input: An original heart sound signal $x(n)$

Output: A constructed heart sound signal $v(n)$

1. Convert an original heart sound signal x into Fourier spectrum ranging [0,π] by applying fast Fourier transform (FFT).
2. Apply EWT and PSBD to obtain N contiguous segmentations of Fourier spectrum, $\Delta_i = [\omega_{i-1}, \omega_i]$, $\sum_{i=1}^{N} \Delta_i = [0, \pi]$, $i = 1, 2, 3, \ldots, N$. For each ω_i, a transition phase T_i is with width $\tau_i = \xi\omega_i$. Two sequential areas are no overlap if and only if $0 < \xi < \min_i \frac{\Delta_{i+1} - \Delta_i}{\Delta_{i+1} + \Delta_i} < 1$.
3. Apply empirical scaling and wavelet functions over N contiguous segmentations to design bank-pass filters, and obtain detail coefficients D_i and approximation coefficients A_l of sub-bands with scaling function and empirical wavelet function given in Def. 1 [10], $i, l = 1, 2, ..., B$, where B represents the total number of detailed sub-bands.
4. Reconstruct time-domain components of N contiguous segmentations, where detail and approximation signals computed by $x_{D(i)}(n) = \sum_{j=1}^{N_j} D_i(j) W_i(n-j)$, $x_{A(l)}(n) = \sum_{j=1}^{N_l} A_l(j) S_l(n-j)$, where $x_{D(i)}(n)$, $x_{A(l)}(n)$ represents detail and approximation sub-band signals of i^{th} level [10], respectively.
5. Calculate kurtosis given in Def.2 on all the reconstructed time-domain components over N contiguous segmentations, and then find the frequency band with the maximum kurtosis and its overlapped frequency bands which are further divided. Set $i = i+1$, and repeat steps 2 to 4 util $i = N$. Finally, obtain the optimal boundary $\{\varpi'_1, \varpi'_2, ..., \varpi'_u\}$, and corresponding detail coefficients and approximation signals.
6. Obtain empirical modes by convolving the scaling function with the approximation coefficients and empirical wavelet functions with their corresponding detail coefficients, which is express as $e_1 = w_e(1,t) * \theta_1(t)$, $e_j = w_e(j,t) * \varphi_j(t)$, $j = 1, 2, \ldots, M$, M is the number of empirical modes.
7. Calculate the Pearson's correlation coefficient given in Def.3 on all the empirical modes, and select the n^{th} empirical modes e_n if and only if $C_{xe} > C_0$, where C_0 is an experimental parameter.
8. Reconstruct and obtain the denoised signal $v(n)$ using the sum of the denoised empirical modes.

Definition 2. *The expressions for Kurtosis are given by*

$$K[|x_h(n)|]^2 = \frac{E(|x_h(n)|^4)}{[E(|x_h(n)|^2)]^2} \tag{4}$$

$$E(|x_h(n)|^4) = \frac{1}{L_{\Delta(i)}} \sum_{n=t}^{t+L_{\Delta(i)}-1} ||x_h(n)\Delta(n-t)||^4 \tag{5}$$

$$[E(|x_h(n)|^2)]^2 = \{E[SE(x_h(n))]\}^2 \tag{6}$$

where $x(n)$ denotes the zero mean and amplitude normalized signal and $x_h(n)$ denotes its Hilbert transform; $SE(x_h(n))$ and $E(x_h(n))$ denotes absolute squared value and expectation of the signal $x(n)$, respectively, and $L_{\Delta(i)}$ denotes the length of the i^{th} contiguous segmentation $\Delta(i)$ of Fourier spectrum.

Definition 3. *For a sample of N values, the Person's correlation coefficient between the original signal and the empirical mode is defined as*

$$C_{xe} = \left| \frac{\sum_{n=1}^{N} (x(n) - \overline{x})(e_m(n) - \overline{e_m})}{\sqrt{\sum_{n=1}^{N} (x(n) - \overline{x})^2} \sqrt{\sum_{n=1}^{N} (e_m(n) - \overline{e_m})^2}} \right| \tag{7}$$

where $\overline{x}$ and $\overline{e_m}$ denote the average of heart sound signal $x(n)$ and the average of the m^{th} empirical mode $e_m(n)$, respectively.

The algorithm of heart sound signals denoising using PSBD-EWT is shown in Algorithm 1. It inspired from PSBD and kurtosis. On one hand, PSBD further enhances EWT's adaptive abilities to overcome noise interference. On the other hand, kurtosis improves EWT's robustness due to removal of noises with varying temporal-spectral characteristics, especially in the case of small fluctuations in places where heart sounds are not present. On basis of the above characteristics, PSBD-EWT can effectively decompose heart sounds under the noisy HSSs to the maximum extent.

3.2. Enhancement Generative Adversarial Network (EGAN)

Although PSBD-EWT can obtain the shape of HSSs well by decomposing the raw HSSs, its anti-aliasing ability are still insufficient. Especially under low SNR environments, the denoised HSSs of PSBD-EWT should be avoided distortion. To overcome such limitations, we here introduce EGAN to PSBD-EWT. It can further reduce the influence of different EWT mode selections.

The end-to-end architecture of EGAN is composed of a generator (G) and a discriminator (D). G network utilizes an encoder-decoder fully-convolutional structure. To obtain better performance of HSSs denoising, the encoder includes k convolutional layers with a kernel size 1×31 and a stride of 4, and Parameteric Rectified Linear Units (PReLUs). Correspondingly, the decoder has an anti-symmetric structure of the encoder. As the number of layers in the network is too deep, the features are easily lost. Consequently, a skip connection was needed between the encoder and decoder which can retain some important features and convey the details of HSSs. Moreover, it can avoid the vanishing gradients. The output of each layer of the encoder and the corresponding feature of the decoder are concatenated as input to the next layer of the decoder. The G network architecture of EGAN is shown in Fig. 1.

In EGAN, the input of G consists of a noisy HSS $\tilde{v}$, a clean HSS v and a random representation z from a distribution $P_z(z)$. The output of G is the denoised HSS $G(z, v)$. The received input of discriminator D is a pair of signals including a clean HSS and a noisy HSS $(x, \tilde{v})$, or a denoised HSS and a noisy HSS $(G(z, v), \tilde{v})$, respectively, and the output is 0 or 1. Actually, GANs chooses the sigmoid cross entropy as a loss function,

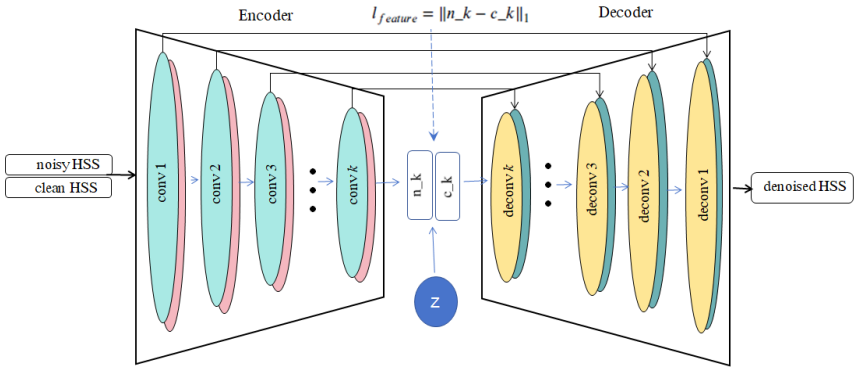


Fig. 1. The G network architecture of EGAN

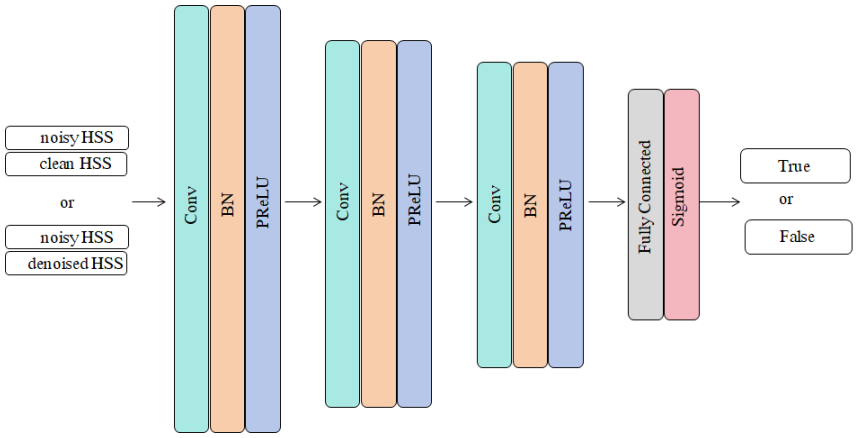


Fig. 2. The D network architecture of EGAN

which leads to vanishing gradients problem. Therefore, $L_2 - distance$ between a clean HSS v and a denoised HSS $G(z, \tilde{v})$ is introduced to the loss function, which can be defined as

$$l_{distance} = \sqrt{\sum_{j=1}^{k} |v - G(z, \tilde{v})|^2}. \tag{8}$$

In addition, the $L_1 - distance$ between a clean HSS v and a denoised HSS $G(z, \tilde{v})$ should be further introduced to the loss function, thereby generating the more realistic samples. It is expressed as

$$l_{samples} = \|v - G(z, \tilde{v})\|_1. \tag{9}$$

In low SNR environments, it is easy to bring about an incomplete feature representation due to difficulties in suppressing noise contaminations. Therefore, a clean HSS is needed to provide an important reference for the corresponding noisy HSS in the training

stage. Inspired by SEGAN [35], the clean HSSs and the noisy HSSs are compressed simultaneously and independently. They share the same network parameters, which can enhance the subsequence effect of reconstructed HSS. In practice, the noisy HSSs and clean HSSs are spliced at the same width and input into the network as two-dimensional vectors. Through layer-to-layer convolution operations, G network obtain a low-dimensional feature n_k of the noisy HSS and a low-dimensional feature c_k of the clean HSS from the decoder output. In order to force G to learn feature representations of HSSs, it is obvious that two features need to be as close to each other as possible. Therefore, the $l_{feature} - distance$ between n_k and c_k needs to be introduced to the loss function. It is expressed as

$$l_{feature} = \|n\_k - c\_k\|_1. \tag{10}$$

In this way, a new total loss function of G in EGAN can be represented as

$$\begin{aligned} \min_G V_{EGAN}(G) = \frac{1}{2} E_{z\sim P_z(z), v_c\sim Pdata(v_c)}[(D(G(z,\tilde{v}),\tilde{v})-1)^2] \\ +\lambda l_{samples} + \mu l_{feature} + \eta l_{distance} \end{aligned} \tag{11}$$

where $P_z(z)$ denotes the prior distribution of input variable z; $Pdata(v_c)$ denotes the distribution of the real data v_c; λ, μ and η denote the weight factors of $l_{samples}$, $l_{feature}$ and $l_{distance}$, respectively.

To obtain better denoising performance of HSSs, the network architecture for D is shown in Fig.2. D consists of 3 convolutional layers and 1 fully connected (FC) layer. Moreover, it employs a batch normalization (BN) and a PReLU activation function after each convolutional layer. Besides, D uses a sigmod activation function after FC to perform classification. The loss function of D in EGAN is expressed as

$$\begin{aligned} \min_D V_{EGAN}(G) = \frac{1}{2} E_{z\sim P_z(z), \tilde{v}\sim Pdata(\tilde{v})}[(D(G(z,\tilde{v}),\tilde{v}))^2] \\ +\frac{1}{2} E_{v,\tilde{v}\sim Pdata(\tilde{v})}[(D(v,\tilde{v})-1)^2] \end{aligned} \tag{12}$$

where $P_z(z)$ denotes the prior distribution of input variable z and $Pdata(\tilde{v})$ denotes the distribution of noisy HSS $\tilde{v}$.

In this way, EGAN has the following characteristics:

(i) In uncontrolled environments, especially in low SNR environments, it is difficult to obtain a complete feature representation from the noisy HSSs alone. EGAN can force the G network to learn the feature representations from the clean HSSs, thereby compensating for the loss of important components.

(ii) Through multi-layer convolution operations, the total loss function makes the denoised HSSs as similar to the clean HSSs as possible. More importantly, the noise contained in higher-level features is significantly eliminated in low-level features, which is more conducive to enhancing heart sounds in low SNR environments.

4. Experiment

4.1. Data Resources

Publicly available heart sound datasets are used to validate the effect of our proposed method. (i) PASCAL dataset. There are 176 files (Dataset A) and 656 files (Dataset B) collected from general public and clinical trials, respectively. Dataset A and B are gathered by iStethoscope Pro-iPhone app and digital stethoscope, DigiScope, respectively. (ii) Michigan Heart Sound and Murmur Database (MHSDB). There are 23 heart sound recordings with a total duration of 1496.8 seconds (iii) Washington University public heart sound dataset (WUPHSD). Heart sound records are collected from 50 patients, aged from 6 to 85 years. In addition, we utilize PyTorch 1.7 to implement deep learning architectures. The models are trained on the environments with Intel CPU, Nvidia GPU (RTX 2070Super), Ubuntu Server 22.04.

4.2. Performance metrics

SNR and RMSE are exploited to evaluate the noise reduction. SNR and RMSE are expressed as

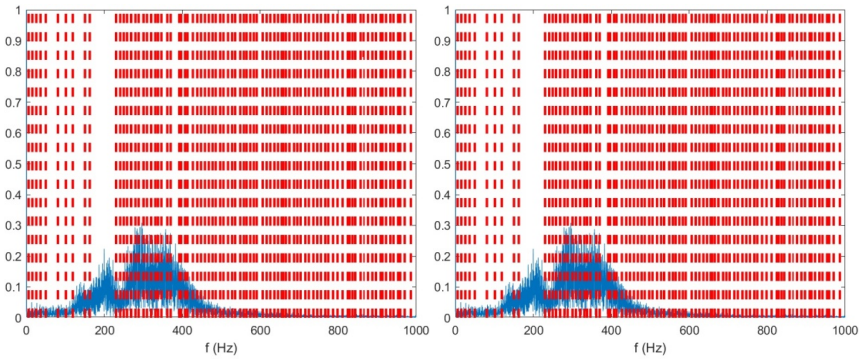
$$SNR = 10 \lg \frac{\sum_{n=1}^{N} (v(n))^2}{\sum_{n=1}^{N} (v(n) - \hat{v}(n))^2}, \tag{13}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (v(n) - \hat{v}(n))^2}, \tag{14}$$

respectively. Here, N denotes the number of HSS samples, $v(n)$ denotes the original HSS, $\hat{v}(n)$ denotes the denoised HSS.

4.3. Denoising Results of PSBD-EWT-EGAN

To verify whether our proposed method can perform noise reduction on HSSs in real-world scenarios, we use noisy HSSs from PASCAL dataset, which are polluted by various unavoidable entities. Take 201102201230.aif from PASCAL/A as an example, we first convert heart sound signal into Fourier spectrums and then apply EWT to obtain 97 contiguous segmentations. The resample points are set to 2000. Then, we exploit PSBD and kurtosis to merge the boundaries of contiguous segmentations and obtain 87 new segmentations. Finally, we exploit PSBD-EWT-EGAN to obtain the final constructed heart sound. Fig.3 show the segmentations of 201102201230.aif using EWT, PSBD and kurtosis, respectively. Some empirical modes are shown in Fig.4. Fig.5 clearly demonstrates the comparison the original HSS and the reconstructed HSS of PSBD-EWT-EGAN. It can be clearly observed that the noisy portions of the noisy HSS are effectively suppressed. Hence, PSBD-EWT-EGAN can denoise HSSs effectively under real-world noisy environments.



(a) The segmentations using EWT **(b)** The segmentations using PSBD kurtosis

Fig. 3. The segmentations of 201102201230.aif from PASCAL/A

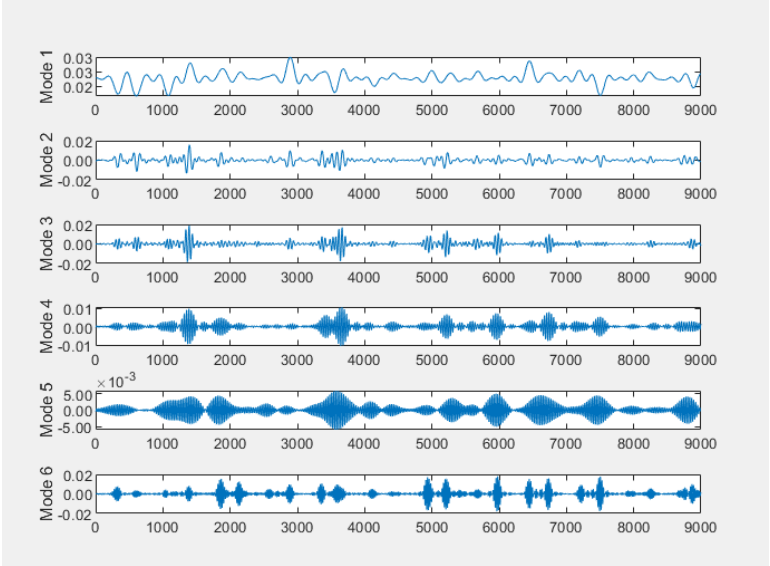


Fig. 4. The empirical modes of 201102201230.aif from PASCAL/A

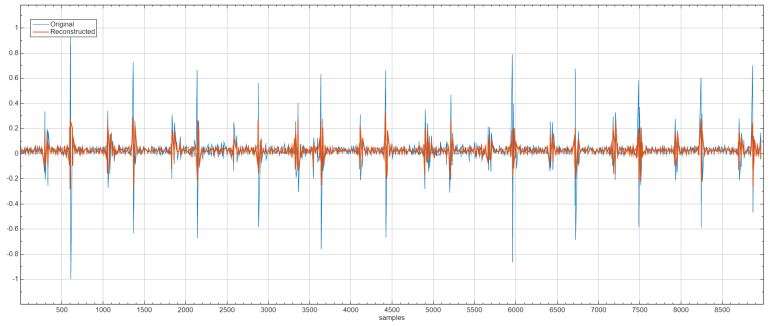


Fig. 5. The original HSSs and the reconstructed HSSs of PSBD-EWT-EGAN

4.4. Denoising Performance Comparisons

Under the same conditions, we evaluate noise reduction results of PSBD-EWT-EGAN compared to WT (constant threshold), EMD and AE-CGAN (autoencoder-Conditional Generative Adversarial Networks) [42]. WT and EMD are two frequently used tools for non-stationary signals. Compared with the EGAN, the generator network of AE-CGA only introduces $L_1 - distance$ to its loss function. The database of clean HSSs consists of the recordings from MHSDB and WUPHSD. The recordings were randomly divided two sets: the testing set and the training set, in a 1:4 ratio. The clean HSSs are mixed with WGN and pink noise at different SNRs (-5dB, -2.5dB, 0dB, 2.5dB, and 5dB) to get the noisy HSSs. All the recordings are partitioned into 3-second clips using the non-overlapping windows.

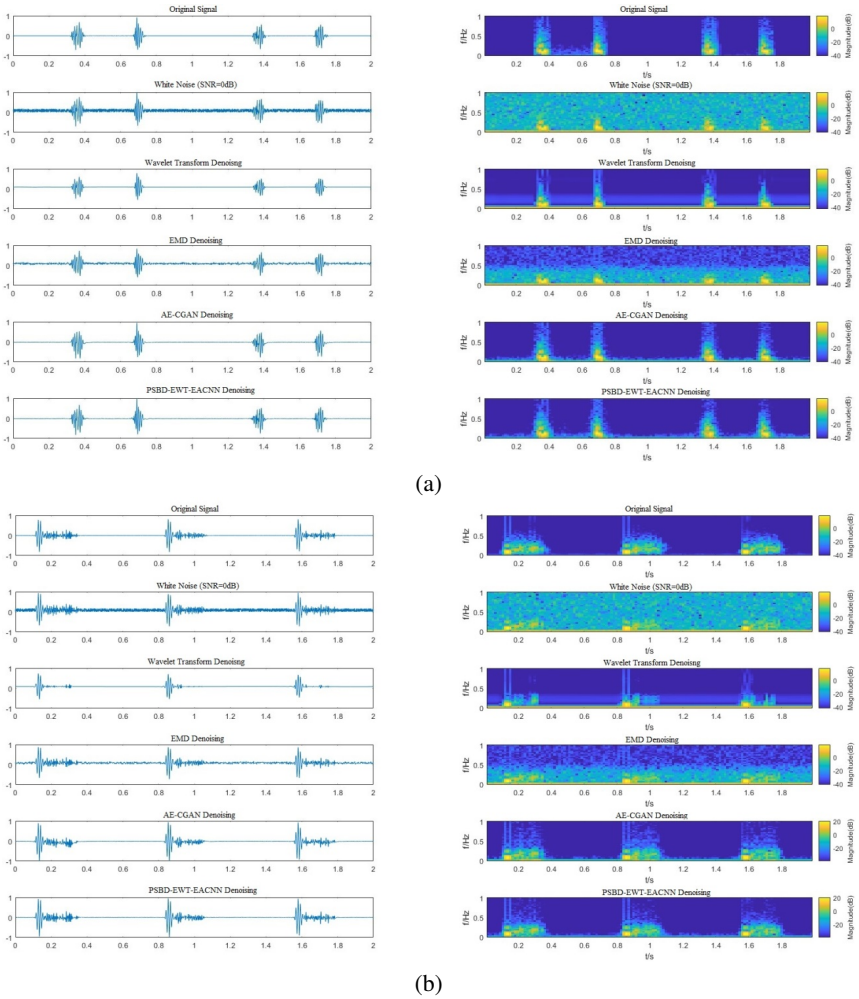
Figures 6 and 7 showcases the effects of noise reduction of normal HSSs from WUPHSD and aortic stenosis HSSs from MHSDB by four models when the HSSs are mixed with WGN noise and pink noise at SNRs=0dB, respectively. The resample points are set to 2000. The time-domain diagram is on the left and the time-frequency diagram is on the right. It can be seen that the WT denoising can not only remove most of the noise, but also eliminate meaningful components of the original HSSs, resulting in the waveform distortion of the denoised HSSs. Although EMD retains most of the original HSSs, it also retains components of the noise, and the denoised effect is not ideal. The main reasons for the above phenomenon are as follows:(i) The basis function of WT denoising is pre-selected and the parameters are fixed manually; (ii) The noise variance of EMD denoising is contained in the interval threshold, and even seriously destroys the structure of HSSs. Therefore, WT and EMD denoising of HSSs are difficult to accurately distinguish heart sounds and noises. Compared with WT and EMD, both AE-CGAN and PSBD-EWT-EGAN can retain more meaningful components of the original HSSs while removing noises in a low SNR environment. Notably, compared to AE-CGAN, PSBD-EWT-EGAN is more capable of preventing excessive distortion.

Table 1. Denoising performance comparison among WT, EMD, AE-CGAN and PSBD-EWT-EGAN

Noise Type	Input SNR(dB)	SNR(dB)				RMSE			
		WT	EMD	AE-CGAN	PSBD-EWT-EGAN	WT	EMD	AE-CGAN	PSBD-EWT-EGAN
White	-5	-4.08	-4.03	7.77	9.26	0.179	0.178	0.052	0.044
	-2.5	-1.69	-1.61	8.19	10.14	0.136	0.134	0.050	0.041
	0	0.68	0.85	8.63	11.24	0.103	0.101	0.049	0.037
	2.5	3.00	3.30	8.63	11.24	0.079	0.076	0.049	0.037
	5	5.28	5.75	9.04	12.53	0.061	0.058	0.047	0.034
Pink	-5	2.5	-3.25	5.63	6.46	0.084	0.162	0.067	0.063
	-2.5	3.64	-0.79	6.83	7.95	0.074	0.122	0.058	0.051
	0	4.89	1.62	7.64	9.10	0.064	0.093	0.053	0.046
	2.5	6.18	4.11	8.49	10.45	0.055	0.070	0.049	0.040
	5	7.56	6.51	8.83	11.54	0.047	0.053	0.048	0.037

Table 1 shows performance of different methods for different noise types and the input SNR levels, including WT, EMD, AE-CGAN and PSBD-EWT-EGAN. Compared with WT and EMD, our proposed method is advantageous in all aspects. In the case of WGN

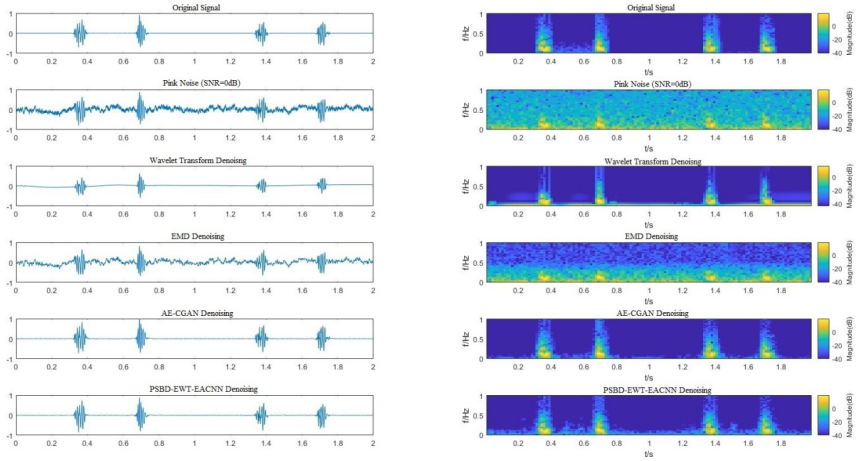
SNR=5dB, our proposed method gains a best output SNR (dB) (12.53 vs.5.28 vs. 5.75) and an RMSE (0.034 vs. 0.061 vs. 0.058). In the case of pink noise SNR=5dB, PSBD-EWT-EGAN provides a best output SNR (dB) (11.54 vs. 7.56 vs. 6.51) and an RMSE (0.037 vs. 0.047 vs. 0.053). It demonstrates the superiority of exploiting PSBD-EWT-EGAN to enhance the quality of the HSSs. Besides, compared with AE-CGAN, PSBD-EWT-EGAN consistently outperforms AE-CGAN in output SNR and RMSE across all input SNRs. Overall, PSBD-EWT-EGAN has the highest anti-noise ability compared with WT, EMD and AE-CGAN. The main reason for the above results is that we replaced the cross-entropy loss with the least square loss. Moreover, the an optimized loss function of G combining $L_1 - distance$, $L_2 - distance$ and $l_{feature} - distance$ between the denoised HSS and the clean HSS is very effective.



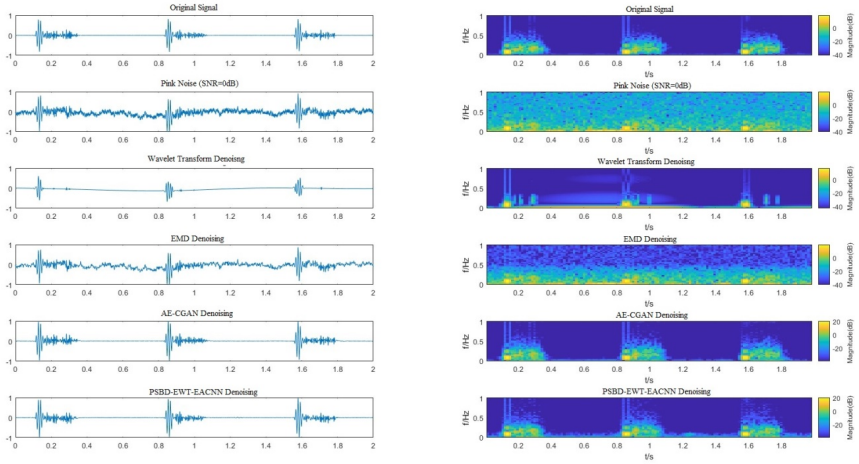
(a)

(b)

Fig. 6. The effects of four denoising methods on white noise SNR=0dB



(c)



(d)

Fig. 7. The effects of four denoising methods on pink noise SNR=0dB

5. Conclusion

Under uncontrolled environments, various uncertain noises inherently corrupt HSSs. In this work, we take advantage of the interpretability of PSBD-EWT for HSSs denoising. Furthermore, we design EGAN architecture with an optimized loss function to improve PSBD-EWT's generalization capability with regard to different noises. Therefore, PSBD-EWT-EGAN has both stronger interpretability and higher generalization capability. Besides, we have validated it on PASCAL, MHSDB and WUPHSD databases, and both healthy and pathological recordings. Both normal and aortic stenosis HSSs contaminated with AWGN or pink noise at different levels of SNR, our approach achieves significant

improvements over state-of-the-art methods. In the case of WGN SNR=5dB, our proposed method obtains the best denoising performance, with a SNR of 12.53dB and an RMSE of 0.034.

PSBD-EWT-EGAN can play a vital role in resisting the noise interference for screening cardiac diseases in uncontrolled environments. The good effect of heart sound denoising lays the foundation for heart sound diagnosis, especially for non-invasive digital stethoscopes, which is conducive to accelerating the application and promotion of heart sound diagnosis in home healthcare monitoring.

Acknowledgments. This work was supported by Fujian Provincial Natural Science Foundation of China No. 2023J011426 and No. 2023J011428; Engineering Research Center for Big Data Application in Private Health Medicine of Fujian Universities, Putian University, Putian, Fujian 351100, China (MKF202408).

References

1. Organization, W.H., et al.: World health statistics overview 2019: monitoring health for the sdgs, sustainable development goals. Tech. rep., World Health Organization (2019)
2. Hadiyoso, S., Mardiyah, D.R., Ramadan, D.N., Ibrahim, A.: Implementation of electronic stethoscope for online remote monitoring with mobile application. Bulletin of Electrical Engineering and Informatics 9(4), 1595–1603 (2020)
3. Cañadas-Quesada, F.J., Ruiz-Reyes, N., Carabias-Orti, J., Vera-Candeas, P., Fuertes-Garcia, J.: A non-negative matrix factorization approach based on spectro-temporal clustering to extract heart sounds. Applied Acoustics 125, 7–19 (2017)
4. Zheng, Y., Guo, X., Jiang, H., Zhou, B.: An innovative multi-level singular value decomposition and compressed sensing based framework for noise removal from heart sounds. Biomedical Signal Processing and Control 38, 34–43 (2017)
5. Ali, A.M., Ghobashy, A.A., Sultan, A.A., Elkhodary, K.I., El-Morsi, M.: A 3d scaling law for supravalvular aortic stenosis suited for stethoscopic auscultations. Heliyon (2024)
6. Centracchio, J., Parlato, S., Esposito, D., Andreozzi, E.: Accurate localization of first and second heart sounds via template matching in forcecardiography signals. Sensors 24(5), 1525 (2024)
7. Sangeetha, B., Periyasamy, R.: Heart sound noise separation from lung sound based on enhanced variational mode decomposition for diagnosing pulmonary diseases. Biomedical Engineering: Applications, Basis and Communications 36(01), 2350035 (2024)
8. Hu, J., Hu, Q., Liang, M.: Heart sounds classification using adaptive wavelet threshold and 1d ldcnn. Computer Science and Information Systems 20(4), 1483–1501 (2023)
9. Xiao, F., Liu, H., Lu, J.: A new approach based on a 1d+ 2d convolutional neural network and evolving fuzzy system for the diagnosis of cardiovascular disease from heart sound signals. Applied Acoustics 216, 109723 (2024)
10. Gilles, J.: Empirical wavelet transform. IEEE transactions on signal processing 61(16), 3999–4010 (2013)
11. Sharan, T.S., Bhattacharjee, R., Sharma, S., Sharma, N.: Evaluation of deep learning methods (dncnn and u-net) for denoising of heart auscultation signals. In: 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA). pp. 151–155. IEEE (2020)
12. Ali, S.N., Shuvo, S.B., Al-Manzo, M.I.S., Hasan, A., Hasan, T.: An end-to-end deep learning framework for real-time denoising of heart sounds for cardiac disease detection in unseen noise. IEEE Access 11, 87887–87901 (2023)

13. Everingham, M., Eslami, S.A., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes challenge: A retrospective. International journal of computer vision 111, 98–136 (2015)
14. Liu, C., Springer, D., Li, Q., Moody, B., Juan, R.A., Chorro, F.J., Castells, F., Roig, J.M., Silva, I., Johnson, A.E., et al.: An open access database for the evaluation of heart sound algorithms. Physiological measurement 37(12), 2181 (2016)
15. Grailu, H.: Compression of high-sampling-rate heart sound signals based on downsampling and pattern matching. Multimedia Tools and Applications 83(1), 201–226 (2024)
16. Hou, Y., Liu, R., Shu, M., Chen, C.: An ecg denoising method based on adversarial denoising convolutional neural network. Biomedical Signal Processing and Control 84, 104964 (2023)
17. Singh, P., Pradhan, G.: A new ecg denoising framework using generative adversarial network. IEEE/ACM transactions on computational biology and bioinformatics 18(2), 759–764 (2020)
18. Hu, L., Cai, W., Chen, Z., Wang, M.: A lightweight u-net model for denoising and noise localization of ecg signals. Biomedical Signal Processing and Control 88, 105504 (2024)
19. Jin, Y., Qin, C., Liu, J., Liu, Y., Li, Z., Liu, C.: A novel deep wavelet convolutional neural network for actual ecg signal denoising. Biomedical Signal Processing and Control 87, 105480 (2024)
20. Kiranyaz, S., Devecioglu, O.C., Ince, T., Malik, J., Chowdhury, M., Hamid, T., Mazhar, R., Khandakar, A., Tahir, A., Rahman, T., et al.: Blind ecg restoration by operational cycle-gans. IEEE Transactions on Biomedical Engineering 69(12), 3572–3581 (2022)
21. Sun, W., Su, Y., Wu, X., Wu, X.: A novel end-to-end 1d-rescnn model to remove artifact from eeg signals. Neurocomputing 404, 108–121 (2020)
22. Chuang, C.H., Chang, K.Y., Huang, C.S., Jung, T.P.: Ic-u-net: a u-net-based denoising autoencoder using mixtures of independent components for automatic eeg artifact removal. NeuroImage 263, 119586 (2022)
23. Sawangjai, P., Trakulruangroj, M., Boonnag, C., Piriyajitakonkij, M., Tripathy, R.K., Sudhawiyangkul, T., Wilaiprasitporn, T.: Eeganet: Removal of ocular artifacts from the eeg signal using generative adversarial networks. IEEE Journal of Biomedical and Health Informatics 26(10), 4913–4924 (2021)
24. Pu, X., Yi, P., Chen, K., Ma, Z., Zhao, D., Ren, Y.: Eegdnet: Fusing non-local and local self-similarity for eeg signal denoising with transformer. Computers in Biology and Medicine 151, 106248 (2022)
25. Haider, N.S.: Respiratory sound denoising using empirical mode decomposition, hurst analysis and spectral subtraction. Biomedical Signal Processing and Control 64, 102313 (2021)
26. Pouyani, M.F., Vali, M., Ghasemi, M.A.: Lung sound signal denoising using discrete wavelet transform and artificial neural network. Biomedical Signal Processing and Control 72, 103329 (2022)
27. Baharanchi, S.A., Vali, M., Modaresi, M.: Noise reduction of lung sounds based on singular spectrum analysis combined with discrete cosine transform. Applied Acoustics 199, 109005 (2022)
28. Yang, Y., Guo, X.M., Wang, H., Zheng, Y.N.: Deep learning-based heart sound analysis for left ventricular diastolic dysfunction diagnosis. Diagnostics 11(12), 2349 (2021)
29. Zhang, A., Wang, J., Qu, F., He, Z.: Classification of children's heart sounds with noise reduction based on variational modal decomposition. Frontiers in Medical Technology 4, 854382 (2022)
30. Gradolewski, D., Magenes, G., Johansson, S., Kulesza, W.J.: A wavelet transform-based neural network denoising algorithm for mobile phonocardiography. Sensors 19(4), 957 (2019)
31. Narváez, P., Percybrooks, W.S.: Synthesis of normal heart sounds using generative adversarial networks and empirical wavelet transform. Applied Sciences 10(19), 7003 (2020)
32. Gastineau, A., Aujol, J.F., Berthoumieu, Y., Germain, C.: A residual dense generative adversarial network for pansharpening with geometrical constraints. In: 2020 IEEE International Conference on Image Processing (ICIP). pp. 493–497. IEEE (2020)

33. Shen, S., Xie, L., Zhang, Y., Wu, G., Zhang, H., Yu, S.: Joint differential game and double deep q-networks for suppressing malware spread in industrial internet of things. IEEE Transactions on Information Forensics and Security 18, 5302–5315 (Aug 2023)
34. Hu, J., Wu, K., Liang, W.: An ipv6-based framework for fog-assisted healthcare monitoring. Advances in Mechanical Engineering 11(1), 1687814018819515 (2019)
35. Yang, F., Wang, Z., Li, J., Xia, R., Yan, Y.: Improving generative adversarial networks for speech enhancement through regularization of latent representations. Speech Communication 118, 1–9 (2020)
36. Feng, S., Zhao, L., Shi, H., Wang, M., Shen, S., Wang, W.: One-dimensional vggnet for high-dimensional data. Applied Soft Computing 135, 110035 (2023)
37. Li, J., Ke, L., Du, Q.: Classification of heart sounds based on the wavelet fractal and twin support vector machine. Entropy 21(5), 472 (2019)
38. Li, S., Li, F., Tang, S., Luo, F.: Heart sounds classification based on feature fusion using lightweight neural networks. IEEE Transactions on instrumentation and measurement 70, 1–9 (2021)
39. Gilles, J., Heal, K.: A parameterless scale-space approach to find meaningful modes in histograms—application to image and spectrum segmentation. International Journal of Wavelets, Multiresolution and Information Processing 12(06), 1450044 (2014)
40. Otsu, N., et al.: A threshold selection method from gray-level histograms. Automatica 11(285-296), 23–27 (1975)
41. Qiu, W., Murphy, W.J., Suter, A.: Kurtosis: a new tool for noise analysis. Acoust Today 16(4), 39–47 (2020)
42. Ge, F.: Brief review of recent researches in speech enhancement from filters to neural networks. In: 2020 International Conference on Computing and Data Science (CDS). pp. 260–264. IEEE (2020)

Jianqiang Hu is an associate professor in school of computer and information engineering, Xiamen University of Technology, China. He once worked as a postdoctoral researcher at Tsinghua University. He received his Ph.D. degree in computer science and engineering from National University of Defense Technology, China, in 2005. He is the author of more than 60 articles, and more than 8 inventions. His current research interests include Edge Computing, Biomedical Signal Processing, and Big Data Analytics.

Lin Chen is a master student at school of computer and information engineering, Xiamen University of Technology, China. His interest include Edge Computing and Biomedical Signal Processing.

Miao Yang received the B.S. degree from Jilin University, Changchun, China, in 2015, and the Ph.D. degree from the Shanghai Advanced Research Institute, Shanghai, and the University of Chinese Academy of Sciences, Beijing, China. He is currently with the Xiamen University of Technology, as a Lecturer. His current research interests include signal processing, federated learning, and edge AI.

Shigen Shen received the B.S. degree in fundamental mathematics from Zhejiang Normal University, Jinhua, China, in 1995, the M.S. degree in computer science and technology from Zhejiang University, Hangzhou, China, in 2005, and the Ph.D. degree in pattern recognition and intelligent systems from Donghua University, Shanghai, China, in 2013. He is a Professor with the School of Information Engineering, Huzhou University,

Huzhou, China. He has published more than 100 technical papers, including respected journals such as IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and IEEE INTERNET OF THINGS JOURNAL. His current research interests include Internet of Things, cyber security, edge computing, and game theory.

Xiao-Zhi Gao received his B.Sc. and M.Sc. degrees from the Harbin Institute of Technology, China in 1993 and 1996, respectively. He obtained his D.Sc. (Tech.) degree from the Helsinki University of Technology (now Aalto University), Finland in 1999. He has been working as a professor at the University of Eastern Finland, Finland since 2018. He has published more than 450 technical papers in refereed journals and international conferences. His research interests are nature-inspired computing methods with their applications in optimization, data mining, machine learning, control, signal processing, and industrial electronics. Prof. Gao is currently serving on a number of prestigious editorial boards, such as Information Sciences, Applied Soft Computing, International Journal of Machine Learning and Cybernetics, Neural Computing and Applications, and Journal of Intelligent and Fuzzy Systems.

Received: August 04, 2024; Accepted: December 14, 2024.