Sensors and Materials, Vol. 37, No. 11 (2025) 5123–5139
MYU Tokyo

S & M 4236

Enhanced Noise Reduction in Photoplethysmography Signals Using a Denoising Autoencoder

Shin-Chi Lai,[1] Yao-Feng Liang,[2] Yi-Chang Zhu,[2]
Li-Chuan Hsu,[3] Shang-Sian Wu,[2] and Szu-Ting Wang[4*]

[1]Department of Automation Engineering/Smart Machinery and Intelligent Manufacturing Research Center, National Formosa University, No. 64, Wenhua Rd., Huwei, Yunlin 632301, Taiwan (R.O.C.)
[2]Doctoral Degree Program in Smart Industry Technology Research and Development, National Formosa University, No. 64, Wenhua Rd., Huwei, Yunlin 632301, Taiwan (R.O.C.)
[3]Master's Program, Department of Electrical Engineering, National Yunlin University of Science and Technology, No. 123, Sec. 3, University Road, Douliu, Yunlin 64002, Taiwan (R.O.C.)
[4]Department of Computer Science and Information Engineering Chaoyang University of Technology, Taichung, Taiwan (R.O.C.)

(Received September 5, 2025; accepted November 14, 2025)

Keywords: photoplethysmography (PPG), denoising autoencoder (DAE), convolutional neural network (CNN), deep learning

We propose a dilated denoising autoencoder (DDAE) based on a multilayer one-dimensional convolutional neural network (CNN) to remove motion-induced noise—baseline wander (BW), muscle artifacts (MA), and electrode motion (EM)—from photoplethysmography (PPG) signals in the PPG-DaLiA dataset acquired using the Empatica E4 wrist-worn wearable device. The model integrates dilated convolutions, residual blocks, batch normalization, Leaky ReLU, max-pooling, and skip connections to capture long-range dependences while preserving pulse morphology. Compared with baseline methods such as the deep neural network (DNN), CNN, fully convolutional network (FCN), and convolutional denoising autoencoder (CDA), it offers three advantages: (1) pooling layers and skip connections preserve key features, (2) optimized parameter design reduces computational complexity and improves efficiency, and (3) a lightweight architecture enables efficient signal reconstruction on resource-constrained hardware. Evaluated across signal-to-noise ratio (*SNR*) conditions (−6 to 24 dB), the model significantly improves *SNR*, achieving 36.27 dB at −6 dB [root mean square error (*RMSE*): 0.008271; percentage root-mean-square difference (*PRD*): 13.97%] and 40.40 dB at 24 dB (*RMSE*: 0.005090; *PRD*: 9.00%. Deployed on a mobile device for real-time PPG processing, the model demonstrates superior performance and strong potential for medical and personal health applications.

*Corresponding author: e-mail: stwang@cyut.edu.tw
https://doi.org/10.18494/SAM5927

ISSN 0914-4935
https://myukk.org/

1. Introduction

Photoplethysmography (PPG) is a noninvasive optical technique widely utilized in wearable devices to monitor physiological parameters, such as heart rate and blood oxygen saturation, and respiratory activity.[1–3] However, PPG signals are highly susceptible to various types of noise in real-world applications, especially under dynamic conditions. Among these, baseline wander (BW), muscle artifacts (MA), and electrode motion (EM) are the three most common sources of interference. BW, typically caused by respiration or slow body movement, manifests as low-frequency fluctuations; MA, resulting from muscle contractions, appears as high-frequency random disturbances; and motion-induced artifacts, owing to variations in skin–sensor contact or pressure, generate abrupt signal distortions.[4] The superposition of these noises degrades signal quality and introduces significant errors in downstream physiological estimations such as heart rate variability, oxygen saturation, or blood pressure analysis. Consequently, effective denoising methods are essential for reliable PPG-based monitoring in wearable systems.

Existing denoising techniques include adaptive filtering, constrained independent component analysis (cICA), wavelet decomposition, and deep learning models such as stacked contractive denoising autoencoders (CDAEs), convolutional denoising autoencoders (CDAs), and fully convolutional denoising autoencoders (FCN-DAEs). Peng *et al.* combined cICA with least mean squares (*LMS*) to remove MA from PPG while preserving amplitude, outperforming traditional ICA on synthetic and real motion data from seven subjects.[5] Ahmed *et al.* used fixed wavelet transform (FWT) with a "db10" basis and a feedforward neural network (FFNN) to suppress Gaussian, Poisson, uniform, and salt-and-pepper noise, achieving mean square error (*MSE*) reductions of 56.40–72.36% on datasets.[6] Xiong *et al.* proposed CDAE for ECG denoising with the Frobenius norm penalty, improving the signal-to-noise ratio (*SNR*) by >2.40 dB at 1.25 dB input using MIT-BIH data with the Noise Stress Test Database (NSTDB) noise (BW, MA, and EM).[7] Mohagheghian *et al.* developed CDA for ECG, enhancing *SNR* and reducing the heart rate root mean square error (*RMSE*) to ~6.6 bpm in AF/non-AF cases using PulseWatch data.[8] Chiang *et al.* introduced FCN-DAE with 13 layers, compressing a 1024-sample electrocardiogram (ECG) to 32-D features and outperforming DNN/CNN in *RMSE*, percentage root-mean-square difference (*PRD*), and *SNR* improvement on MIT-BIH with NSTDB noise.[9] Despite progress, these methods often require multichannel inputs, high computation, or lack real-time deployment on wearable devices, underscoring the need for lightweight, single-channel, sensor-native solutions.

Despite these advancements, significant research gaps remain. While methods like wavelet transform and CNN-based models show promise, they struggle with capturing long-range temporal dependences and often require substantial computational resources, limiting their applicability in resource-constrained wearable devices.

To address these challenges, we propose a dilated denoising autoencoder (DDAE) that integrates dilated convolutions, residual blocks (RBs), and skip connections. This architecture effectively captures both short- and long-term dependences while preserving key signal features, offering efficient and robust denoising for real-time PPG processing. Experimental results demonstrate that our model significantly outperforms baseline methods in denoising performance across various noise conditions.

The proposed method is closely related to the field of machine learning, particularly deep-learning-based signal processing. The denoising autoencoder (DAE) utilized in this study is a representative machine learning model that performs unsupervised feature learning from corrupted data. By combining convolutional and dilated convolutional structures with residual learning and skip connections, the DDAE leverages modern machine learning techniques to improve signal reconstruction accuracy and noise robustness. In this research, we extend the application of machine learning to biomedical signal enhancement, demonstrating how data-driven models can effectively address nonlinear and dynamic noise in physiological signals. Therefore, this work contributes to the advancement of related machine learning technologies by providing an efficient and generalizable denoising framework for wearable health monitoring systems.

2. Methods

2.1 Proposed model for PPG denoising

We propose a network architecture for PPG signal denoising, termed DDAE, as illustrated in Fig. 1. The architecture consists of five encoder layers, a bottleneck layer, five decoder layers, which incorporate dilated convolutions, RBs, and skip connections to enhance feature extraction and reconstruction. Initially, the noisy PPG signal is fed into the input layer. During the encoding phase, the features of the clean PPG signal are extracted through progressively dilated convolutional layers, forming a latent representation z in the bottleneck layer. In the decoding phase, transposed convolutions, residual connections, and skip connections are utilized to reconstruct the denoised PPG signal while preserving essential details. The detailed parameters of the architecture are provided in Table 1.

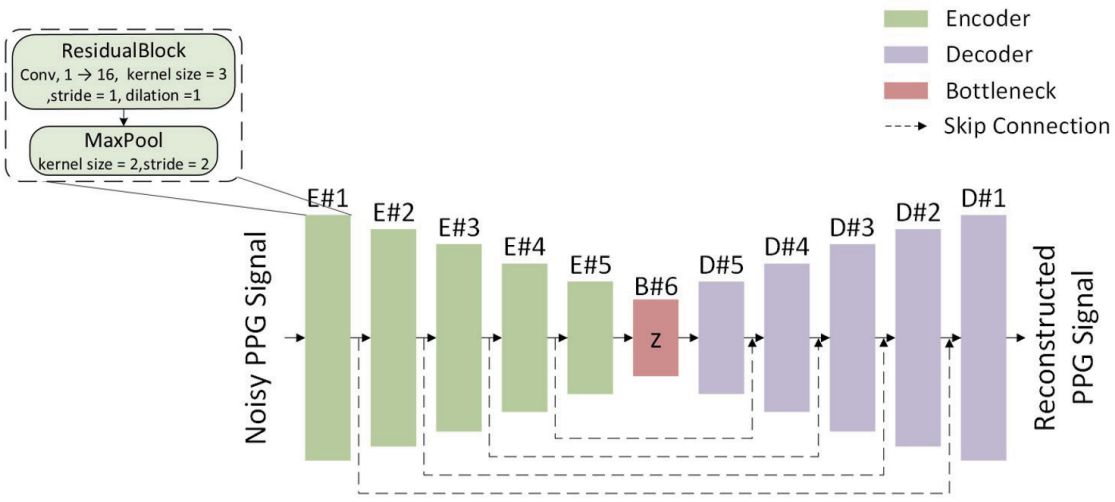

Fig. 1. (Color online) Proposed system architecture diagram.

Table 1
Architecture of the proposed DDEA for 1D signal processing.

Layer	Layer details	Output shape
Encoder #1	RB [Conv, 1 → 16, 3, 1, dilation=1] + MaxPool [2, 2]	[−1, 16, 256]
Encoder #2	RB [Conv, 16 → 32, 3, 1, dilation=2] + MaxPool [2, 2]	[−1, 32, 128]
Encoder #3	RB [Conv, 32 → 64, 3, 1, dilation=4] + MaxPool [2, 2]	[−1, 64, 64]
Encoder #4	RB [Conv, 64 → 128, 3, 1, dilation=8] + MaxPool [2, 2]	[−1, 128, 32]
Encoder #5	RB [Conv, 128 → 256, 3, 1, dilation=16] + MaxPool [2, 2]	[−1, 256, 16]
Bottleneck #6	RB [Conv, 256 → 256, 3, 1, dilation=32]	[−1, 256, 16]
Decoder #5	RB [ConvTranspose, 256 → 128, 2, 2] + RB [Conv, 128 → 128, 3, 1, dilation=16] + Skip	[−1, 128, 32]
Decoder #4	RB [ConvTranspose, 256 → 64, 2, 2] + RB [Conv, 64 → 64, 3, 1, dilation=8] + Skip	[−1, 64, 64]
Decoder #3	RB [ConvTranspose, 128 → 32, 2, 2] + RB [Conv, 32 → 32, 3, 1, dilation=4] + Skip	[−1, 32, 128]
Decoder #2	RB [ConvTranspose, 64 → 16, 2, 2] + RB [Conv, 16 → 16, 3, 1, dilation=2] + Skip	[−1, 16, 256]
Decoder #1	RB [ConvTranspose, 32 → 8, 2, 2] + RB [Conv, 8 → 1, 3, 1, dilation=1] + Sigmoid	[−1, 1, 512]

Note: "RB" denotes a residual block, which includes a 1D convolution (or transposed convolution), batch normalization (*BN*), *LeakyReLU* activation, and a residual connection. "Conv" and "ConvTranspose" denote 1D convolutional and transposed convolutional layers, respectively, with details of [operation, input → output, kernel size, stride, dilation rate]. "MaxPool" and "Skip" denote max pooling and skip connections, respectively. Output shapes are [batch size, channels, length], with a batch size of −1.

2.2 Encoder layer

The encoder assumes that the input PPG signal has been normalized to the range [0, 1] during the data preprocessing stage to enhance the stability of model training. Subsequently, the signal is processed by an encoder composed of five layers of one-dimensional convolutions, which progressively extract multiscale features and compress the temporal dimension. The encoding process begins at the first layer, where a one-dimensional RB (1D Res) expands the number of channels from 1 to 16 using a convolution operation with a kernel size of 3, followed by a max-pooling layer (kernel size = 2, stride = 2), which reduces the temporal dimension from 512 to 256. In the second layer, the number of channels increases to 32, and the temporal dimension is further reduced to 128. The third layer expands the channel number to 64, which compresses the temporal dimension to 64. The fourth layer increases the number of channels to 128, which reduces the temporal dimension to 32. Finally, in the fifth layer, the number of channels expands to 256, with the temporal dimension being compressed to 16, which yields a low-dimensional feature representation (z). Each 1D Res block in the encoder consists of a 1D convolution (kernel size = 3), batch normalization (BatchNorm), and a *LeakyReLU* activation function. Additionally, to capture long-range temporal dependences, the dilation rates progressively increase across the layers, set to 1, 2, 4, 8, and 16. This hierarchical feature extraction framework lays a strong foundation for effective denoising and signal reconstruction in the subsequent decoding stage.

2.3 Bottleneck layer

The bottleneck layer serves as a compact representation of the input signal, which preserves essential features extracted from the encoder while preparing them for reconstruction in the decoder. It consists of a single RB (kernel size = 3, dilation = 32), which maintains 256 channels and a temporal dimension of 16. This layer ensures effective feature retention while enhancing the model's ability to remove noise before upsampling in the decoder.

2.4 Decoder layer

The decoder reconstructs the PPG signal from the low-dimensional feature representation at the bottleneck layer (256 channels, temporal dimension = 16) through five one-dimensional convolutional layers, progressively restoring the temporal dimension to 512 while reducing the number of channels to 1.

The decoding process begins at the fifth layer, where a transposed convolutional RB (kernel size = 2, stride = 2) reduces the number of channels from 256 to 128 and expands the temporal dimension from 16 to 32. A RB with a dilation rate of 16 further refines features before merging with the skip connection from the fourth encoder layer ($skip_4$, with the number of channels adjusted to 128), which restores 256 channels. In the fourth layer, a transposed convolution reduces the number of channels to 64 while increasing the temporal dimension to 64, followed by a RB with a dilation rate of 8. This layer is then merged with the skip connection from the third encoder layer ($skip_3$, with the number of channels adjusted to 64), which results in 128 channels. The third layer continues to adjust the number of channels to 32 while expanding the temporal dimension to 128, with a dilation rate of 4, and merges with the skip connection from the second encoder layer ($skip_2$, with the number of channels adjusted to 32), which yields 64 channels. In the second layer, the number of channels decreases to 16 and the temporal dimension expands to 256, with a dilation rate of 2. This layer then merges with the skip connection from the first encoder layer ($skip_1$, with the number og channels adjusted to 16), which results in 32 channels.

Finally, in the first layer, a transposed convolution reduces the number of channels to 8 and fully restores the temporal dimension to 512, followed by a RB with a dilation rate of 1. To generate the final denoised PPG signal, a one-dimensional convolution and a sigmoid activation function are applied, reducing the number of channels to 1 and mapping the output range to [0,1].

2.5 RB

To mitigate the vanishing gradient problem in deep neural networks, which hinders training efficiency as the number of layers increases, we incorporate RBs as the core component of the denoising autoencoder. Inspired by the residual learning framework proposed by He *et al*.,[10] the RB facilitates gradient flow through skip connections, effectively supporting the training of deep architectures while preserving the critical features necessary for signal reconstruction.

Within the RB, the output is computed by the element-wise addition of the input and the transformed features, as expressed by

$$y = x + f(x), \tag{1}$$

where $x \in R^{C \times L}$ is the input with C channels and temporal length L, and the transformation function is defined by

$$f(x) = LeakyReLU(BN(Conv(x))). \tag{2}$$

Here, $Conv(x)$ denotes a one-dimensional convolution, BN represents batch normalization, and $LeakyReLU$ is the activation function. The skip connection in the RB ensures the effective propagation of input information, mitigating gradient vanishing while improving the reconstruction of high-fidelity PPG signals.

2.6 Dilated convolution

To capture long-range temporal dependences efficiently, in this model, dilated convolutions in the encoder and decoder RBs are employed. By expanding the receptive field without increasing complexity, this approach enables multiscale feature extraction while preserving signal details. The encoder uses dilation rates of 1, 2, 4, 8, and 16, with the bottleneck at 32, while the decoder mirrors this pattern in reverse. This design enhances the model's ability to recognize patterns across extended time ranges in PPG signals. Below is the formula for dilated convolution.[11–13]

Formally, a one-dimensional discrete signal $F : \mathbb{Z} \to \mathbb{R}$ with the dilated convolution at the position p with kernel k is defined as

$$\left(F *_l k\right)\left(p\right) = s + l \cdot t = pF\left(s\right)k\left(t\right), \tag{3}$$

where l is the dilation rate and s and t are the indices of the input and kernel, respectively. When $l = 1$, this operation reduces to a standard convolution.

A key advantage of dilated convolutions is their ability to exponentially expand the receptive field while maintaining a linear increase in the number of parameters. The receptive field size R_i of each element in the i-th layer (starting from index 0) with dilation rate l_i can be computed as

$$R_i = 1 + j = \sum_{j=0}^{i} l_i \cdot (2r), \tag{4}$$

where r is the radius of the convolution kernel (e.g., $r = 1$ for a kernel size of 3). In this model, with dilation rates of {1, 2, 4, 8, 16}, the 4th encoder layer (dilation rate = 16) achieves a receptive

field of 63, which means that each output at this layer can capture dependences spanning 63 time steps. This significantly improves the model's ability to recognize long-range temporal relationships, which is critical for denoising and preserving the morphology of PPG signals.

2.7 Skip connection

Skip connections are incorporated into the model to improve feature propagation and mitigate the vanishing gradient problem. By directly concatenating intermediate features from encoder layers to their corresponding decoder layers along the channel dimension, these connections help retain critical temporal details lost during downsampling. This enhances the model's ability to reconstruct the original signal with greater accuracy. Formally, for an encoder layer E_i with output features $x_i \in R^{C_i \times L_i}$ and its corresponding decoder layer D_i with input features $z_i \in R^{c_i' \times L_i}$, the skip connection is defined as

$$z_i' = Concat(z_i, x_i), \tag{5}$$

where $z_i' \in R^{c_i' \times L_i}$ is the output of the skip connection and $Concat(\cdot,\cdot)$ denotes the concatenation operation along the channel dimension. This mechanism enhances feature reuse, preserves temporal structure, and significantly improves the model's denoising performance.

3. Experimental Results

3.1 Hyperparameter and computation complexity

The model is trained using the AdamW optimizer (with an initial learning rate of 0.001 and a weight decay of 10^{-5}), coupled with a learning rate scheduler (ReduceLROnPlateau) that dynamically adjusts the learning rate on the basis of validation loss, with a minimum learning rate of 10^{-6}. The loss function is a weighted combination of *MSE* and L1 loss (with weights of 0.5 each), designed to balance reconstruction accuracy and smoothness. The model was trained for 250 epochs to ensure sufficient convergence. The final model consists of 626116 parameters and exhibits a computational complexity of 22104064 multiply-accumulate operations (MACs), which demonstrates a favorable balance between computational efficiency and performance.

3.2 Evaluation criteria

To evaluate the effectiveness of the proposed method in denoising PPG signals, we employ three common metrics: *SNR*, *RMSE*, and *PRD*. *SNR* (dB) measures the strength of the signal relative to noise, as shown in Eq. (6) below, with higher values indicating better denoising performance.

$$SNR = 10 \log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \tag{6}$$

RMSE quantifies the error between the denoised and original signals, as shown in Eq. (7), with lower values indicating better restoration.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x[i]-\hat{x}[i])^2} \tag{7}$$

Here, $x[i]$ represents the original clean PPG signal, $\hat{x}[i]$ denotes the denoised PPG signal, and N is the total number of samples in the signal.

PRD indicates the relative difference between the original and denoised signals, as shown in Eq. (8), where a *PRD* below 10% indicates high-quality restoration.

$$PRD = \sqrt{\frac{\sum_{i=1}^{N}(x[i]-\hat{x}[i])^2}{\sum_{I=1}^{N}(x[i])^2}} \times 100 \tag{8}$$

These metrics provide a comprehensive assessment of the model's denoising performance.

3.3 Dataset selection and preprocessing

The PPG-DaLiA dataset[14] is adopted as the benchmark for evaluating the proposed denoising method. This publicly available dataset, hosted by the UCI Machine Learning Repository, contains multimodal physiological signals recorded from 15 healthy subjects (8 females, 7 males; age: 26.3 ± 4.2 years) performing a wide range of daily-life activities under natural conditions. Data were collected using the Empatica E4 wristband, a commercial wearable device equipped with a PPG sensor (green LED at 525 nm, sampling rate: 64 Hz), a 3-axis accelerometer, electrodermal activity (EDA), and temperature sensors. The dataset includes eight standardized activity types: sitting, standing, walking, ascending/descending stairs, cycling, driving, and table soccer. These activities induce diverse motion artifacts—ranging from low-frequency BW during sitting to high-frequency MA during walking and stair climbing—making PPG-DaLiA an ideal real-world testbed for evaluating PPG denoising algorithms under realistic wearable sensing conditions.[15]

For model training and evaluation, the dataset was split into 80% training, 10% validation, and 10% test sets to ensure robust training and assessment. To meet the model's input requirements, the raw PPG signals from the PPG-DaLiA dataset were segmented into fixed-length segments of 512 samples (approximately 8 s). Each segment undergoes three signal quality index (SQI) checks to ensure data reliability. The specific methods for these checks are as follows.

To ensure the quality of input PPG signals, a multistep screening process is applied. Variability detection is first performed to remove overly smooth segments, where the segment length must be $L = 512$ samples (8 s at 64 Hz). The smoothness condition is evaluated using

$$diff_i = \left| x[i] - x[i-1] \right|, \forall \in \{1,2,...,L-1\}. \tag{9}$$

Next, peak detection ensures periodic characteristics by requiring at least two peaks with a minimum interval of 20 points. The relevant definitions are

$$P = len\left(\left\{ p_j \mid x[p_j] > x[p_j - 1] and\, x[p_j] > x[p_j + 1],\, d_j \geq 20 \right\} \right), \tag{12}$$

$$d_j = p_{j+1} - p_j \ , \left(j = 1,2,\ldots,P-1 \right). \tag{13}$$

Here, p_j represents the position of the j-th peak, P is the total number of peaks, and d_j is the distance between adjacent peaks. We require $P \geq 2$ and $d_j \geq 20$ for all j. If the number of peaks is insufficient ($P < 2$) or if any distance between adjacent peaks is less than 20 points ($d_j < 20$), the segment is discarded.

Finally, skewness detection is employed to measure the asymmetry of the signal distribution and further filter segments that meet quality standards. The skewness calculation is given by

$$S = \frac{1}{L} = \sum_{i=1}^{L} \left(\frac{x[i] - \mu^3}{\sigma} \right). \tag{14}$$

Here, μ represents the mean of the signal, ó is the standard deviation, and L = 512. Segments with S < 0.3 are removed. This multistep detection mechanism ensures data reliability and provides a robust framework for assessing the quality of PPG signals, delivering high-quality input data for subsequent denoising and heart rate estimation algorithms. Figure 2 illustrates the clean PPG signals that have passed the three-image quality detection process, showcasing their suitability for further analysis.

To enhance model robustness, training was conducted with a batch size of 32, and data augmentation techniques were applied, including random noise injection to simulate varying *SNR* levels and time warping to mimic temporal distortions in real-world PPG signals. These augmentation strategies improve the model's ability to generalize across diverse noise conditions and dynamic environments.

3.4 Noise simulation

To simulate realistic noise conditions, NSTDB[16] provided by PhysioNet is used. NSTDB contains physiological signal interferences recorded in clinical settings, including (1) MA: high-frequency noise from muscle contractions; (2) BW: low $w_{BW}[i]$-frequency drift caused by respiration or movement; (3) EM: sudden disruptions due to electrode displacement. Since the original sampling frequency of NSTDB is 360 Hz, which differs from the sampling frequency of

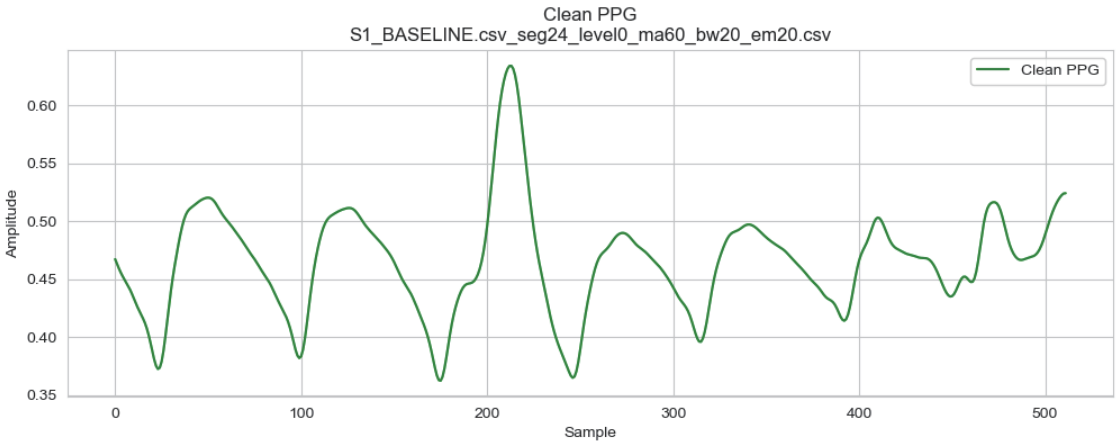

Fig. 2. (Color online) Clean PPG signals that have passed the three-image quality detection process.

the PPG signals in the PPG-DaLiA dataset (64 Hz), the downsampling of the NSTDB noise data is performed to ensure compatibility.

Next, to evaluate the impact of different noise combinations on denoising performance, four noise proportion configurations are defined. In the balanced proportion, the noise consists of 33% MA, 33% BW, and 34% EM. In the MA dominant setting, MA accounts for 60%, while BW and EM each contribute 20%. The BW dominant configuration assigns 60% to BW, with MA and EM each at 20%. Lastly, in the EM dominant setting, EM makes up 60%, while MA and BW each account for 20%.

For each configuration, the noise signals $w_{MA}[i]$ and $w_{EM}[i]$ are tiled to match the length of the PPG signal. A synthetic noise signal $w[i]$ is then generated through weighted summation, as calculated by

$$w[i] = \beta_{MA} \cdot w_{MA}[i] + \beta_{BW} \cdot w_{BW}[i] + \beta_{EM} \cdot w_{EM}[i]. \tag{15}$$

Here, $\beta_{MA} + \beta_{BW} + \beta_{EM} = 1$. Subsequently, six noise levels are defined on the basis of the different *SNR* settings: −6, 0, 6, 12, 18, and 24 dB. The adjustment factor α is calculated using Eq. (16) to control the target *SNR*.

$$\alpha = \sqrt{\frac{P_{signal}}{P_{noise} \times 10^{SNR/10}}} \tag{16}$$

Here, P_{signal} and P_{noise} represent the power of the PPG signal and the original noise, respectively. Subsequently, the adjusted noise is superimposed onto the PPG signal to generate the noisy signal $\tilde{x}[n]$, as shown below.

$$\tilde{x}[i] = x[i] + \alpha \cdot w[i] \tag{17}$$

Finally, the resulting noisy signal is normalized to the range [0, 1], as calculated below.

$$\tilde{x}_{norm}[i] = \frac{\tilde{x}[i] - \min(\tilde{x})}{\max(\tilde{x}) - \min(\tilde{x})} \tag{18}$$

Figure 3 presents the waveform of the clean PPG signal after the addition of noise, highlighting the impact of noise contamination.

3.5 Experimental results and comparison

To evaluate the proposed DDAE, we compared it with four baseline models: DNN, CNN, FCN, and CDA,[7] all with similar parameter scales. DNN adopts a five-layer fully connected structure (311-156-78-156-311 nodes), using *LeakyReLU* and batch normalization, but lacks temporal modeling capability. CNN features a four-layer 1D convolutional encoder–decoder with a bottleneck, which is effective for local feature extraction but limited in capturing long-range dependences. FCN consists of a convolutional encoder–decoder with max pooling, using *LeakyReLU* but without batch normalization or skip connections, which weakens reconstruction. CDA,[7] originally designed for ECG signal denoising, was specifically included in this study owing to its representative and reference-worthy architecture; it incorporates a deeper six-layer encoder–decoder with skip connections and a multiscale design, offering improved feature preservation but with higher complexity. In contrast, the proposed DDAE employs dilated convolutions, RBs, and skip connections to extract multiscale temporal features while preserving the signal structure. Dilated layers expand the receptive field without increasing model size, and residual connections enhance training stability. As shown in Table 2–4, DDAE achieves the highest *SNR* improvement (SNR_{imp}) across all noise levels—from 36.27 dB at –6 dB input *SNR*

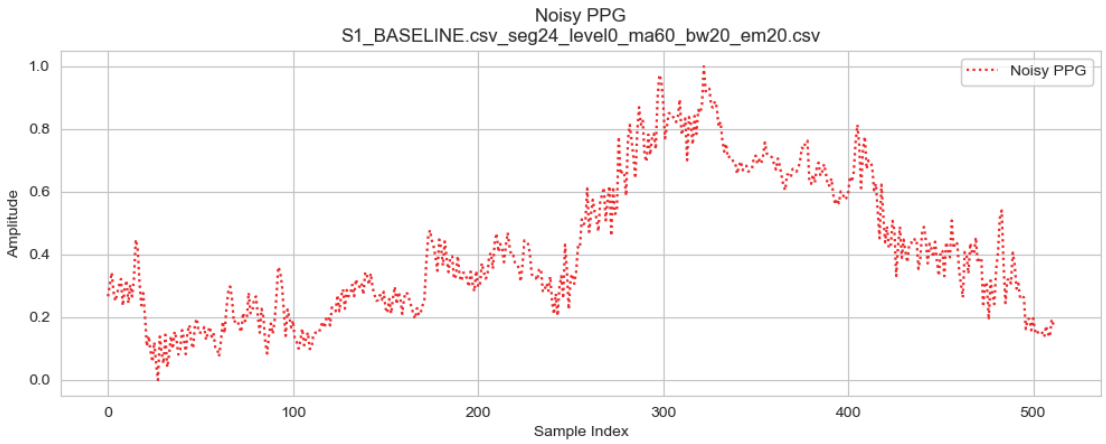

Fig. 3. (Color online) PPG signal after the addition of noise.

Table 2
Performances of the DDAE model and baseline models across different SNR_{in} levels (SNR_{imp}).

DAE model	Average SNR_{imp} (dB)					
	SNR_{in} −6 dB	SNR_{in} 0 dB	SNR_{in} 6 dB	SNR_{in} 12 dB	SNR_{in} 18 dB	SNR_{in} 24 dB
CDA[7]	19.60	21.34	21.54	22.05	22.47	2.61
CNN	26.90	29.91	31.28	31.92	32.12	30.78
DNN	21.53	22.90	23.56	24.06	24.23	24.13
FCN	28.04	30.45	31.13	31.35	31.28	29.78
Proposed	36.27	38.17	39.17	39.97	40.60	40.40

Table 3
Performances of the DDAE model and baseline models across different SNR_{in} levels (*PRD*).

DAE model	Average *PRD* (%)					
	SNR_{in} −6 dB	SNR_{in} 0 dB	SNR_{in} 6 dB	SNR_{in} 12 dB	SNR_{in} 18 dB	SNR_{in} 24 dB
CDA[7]	10.97	8.89	8.65	8.12	7.71	8.44
CNN	4.89	3.34	2.81	2.60	2.53	2.95
DNN	8.74	7.41	6.93	6.61	6.50	6.57
FCN	4.20	3.10	2.85	2.77	2.74	3.31
Proposed	13.97	11.20	9.91	9.18	8.55	9.00

Table 4
Computational complexities and *RMSE* values for DDAE and baseline models across different input SNR_{in} levels.

DAE model	Number of trainable parameters	MACs	Average *RMSE*					
			SNR_{in} −6 dB	SNR_{in} 0 dB	SNR_{in} 6 dB	SNR_{in} 12 dB	SNR_{in} 18 dB	SNR_{in} 24 dB
CDA[7]	790289	99483648	0.0572	0.0465	0.0453	0.0426	0.0410	0.0449
CNN	626779	32546272	0.0251	0.0173	0.0146	0.0135	0.0131	0.0155
DNN	566851	568052	0.0455	0.0386	0.0359	0.0342	0.0336	0.0340
FCN	625939	40292896	0.0215	0.0160	0.0147	0.0144	0.0143	0.0173
Proposed	626116	22104064	0.0082	0.0066	0.0059	0.0053	0.0049	0.0050

(SNR_{in}) to 40.40 dB at 24 dB—alongside the lowest *RMSE* (0.005090) and *PRD* (9.00%). These results confirm DDAE's superior denoising performance under various noise conditions.

Figures 4(a)–4(f) illustrate the performance of PPG signal denoising and reconstruction performance of DDAE in multi-scenario dynamic environments, along with the noisy input (blue scatter points) and the ground truth (green curve) for comparison. The *x*-axis represents time and the *y*-axis represents signal amplitude. Figure 4(c) (BASELINE) demonstrates that DDAE effectively suppresses BW, with the reconstructed signal (orange curve) closely aligning with the ground truth, successfully filtering out low-frequency noise. Figure 4(d) shows that DDAE preserves the pulse wave morphology under noise induced by posture changes, with minimal error between the reconstructed signal and the ground truth. Figures 4(a), 4(e), and 4(f) (WORKING) highlight DDAE's ability to successfully recover PPG features during walking conditions. Despite strong motion artifacts, the reconstructed signal maintains the morphology of peaks and valleys, consistent with the ground truth trend. Figure 4(b) (DRIVING) indicates that DDAE effectively filters out noise caused by hand movements and vibrations in a driving

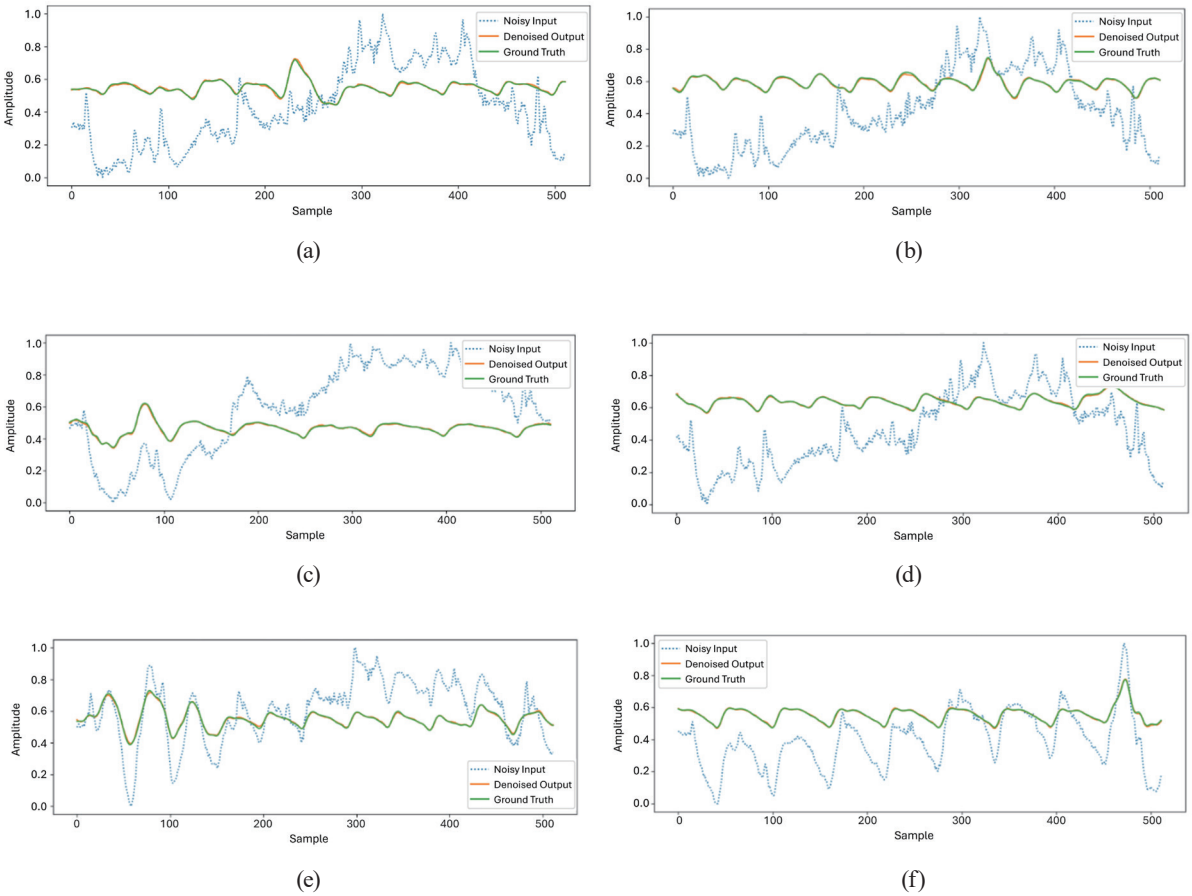

Fig. 4. (Color online) PPG signal denoising and reconstruction performance of DDAE in multi-scenario dynamic environments: (a) WORKING: Computational complexity and RMSE for DDAE and baseline models across different input SNR_{in} of −6. (b) DRIVING: SNR_{in} of 0. (c) BASELINE: SNR_{in} of 6. (d) SOCCER: SNR_{in} of 12. (e) WORKING: SNR_{in} of 18. (f) WORKING: SNR_{in} of 24.

scenario, with the reconstructed signal closely matches the ground truth, demonstrating its robustness and reliability in dynamic environments.

Figures 5–7 show performances of five deep DAE models, DDAE, CNN, CDA, DNN, and FCN, in denoising PPG signals, presented using notched box plots, covering three evaluation metrics: SNR_{imp}, *PRD*, and *RMSE*. These metrics respectively reflect the quality of the denoised signal, the relative magnitude of reconstruction error, and the absolute error level. The data are based on test set results across six different SNR_{in} levels (−6, 0, 6, 12, 18, 24 dB).

The *PRD* box plot in Fig. 5 reveals that DDAE achieves significantly lower *PRD* values across all SNR_{in} levels, with a narrower box range, indicating stable and minimal reconstruction errors. Notably, at mid-to-high SNR_{in} levels (6 to 24 dB), DDAE exhibits fewer outliers, reflecting its high precision in preserving PPG signal details. In contrast, models like FCN and DNN show higher *PRD* medians at certain SNR_{in} levels, highlighting DDAE's superior performance in reducing reconstruction errors.

The *PRD* box plot in Fig. 6 shows that DDAE achieves significantly lower *PRD* values across all SNR_{in} levels, with a narrower box range, indicating stable and minimal reconstruction errors.

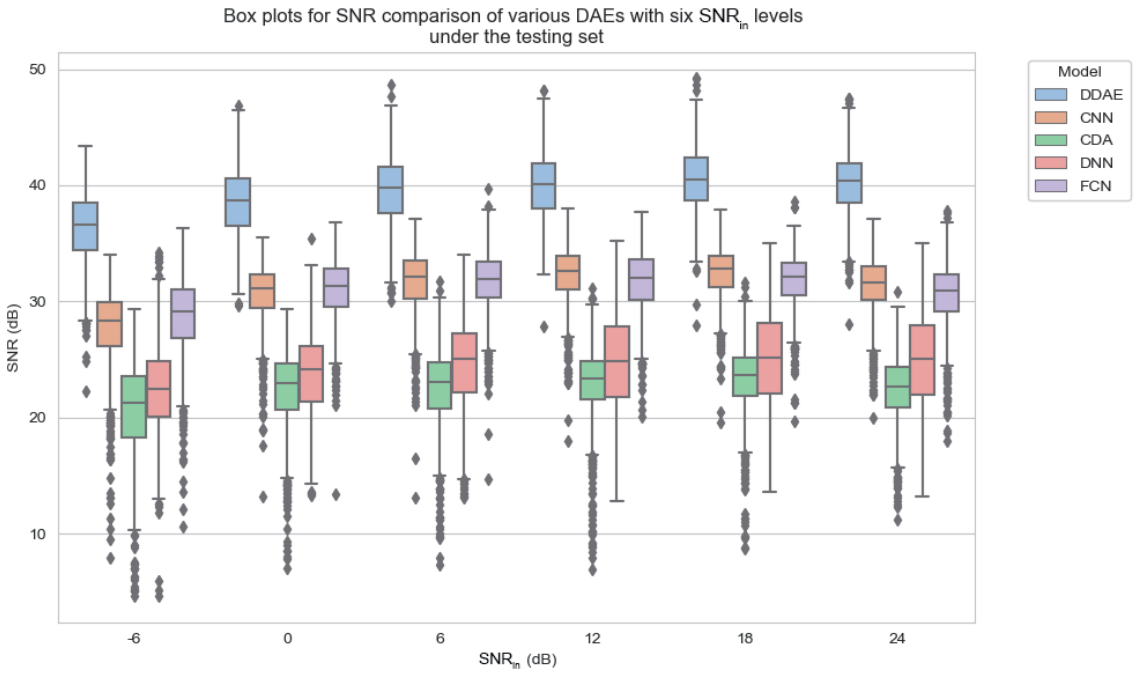

Fig. 5. (Color online) Box plots of *SNR* of various DAEs with six SNR_{in} levels under the testing set.

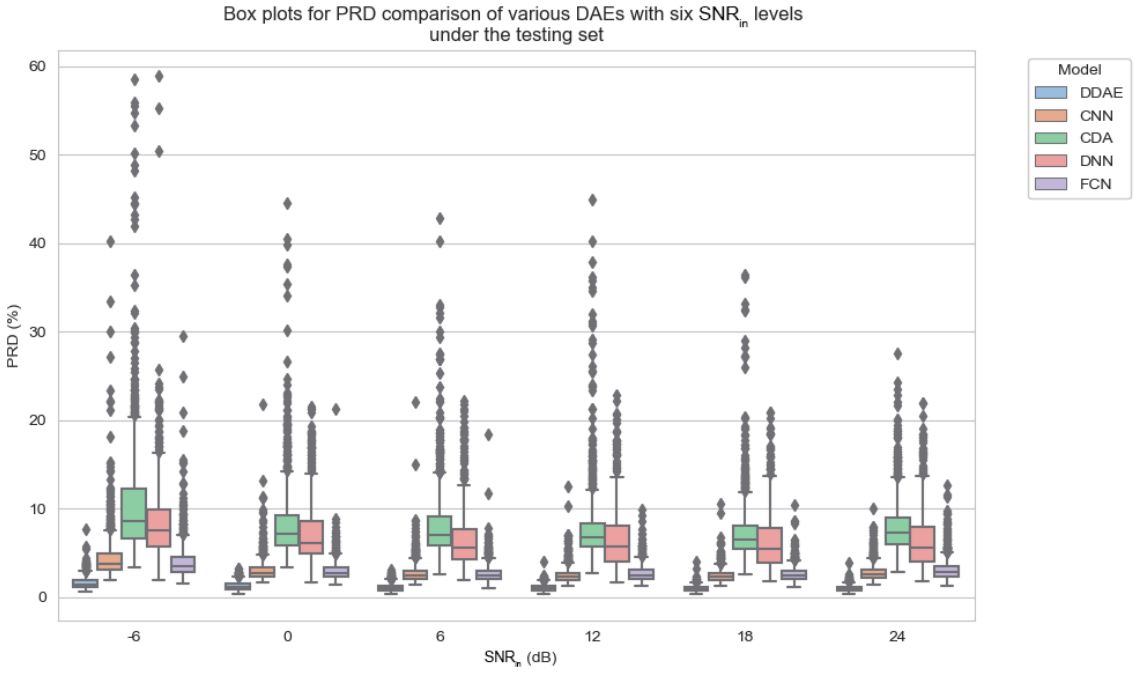

Fig. 6. (Color online) Box plots of *PRD* of various DAEs with six SNR_{in} levels under the testing set.

Notably, at mid-to-high SNR_{in} levels (6 to 24 dB), DDAE exhibits fewer outliers, reflecting its high precision in preserving PPG signal details. In contrast, models like FCN and DNN show higher *PRD* medians at certain SNR_{in} levels, highlighting DDAE's superior performance in reducing reconstruction errors.

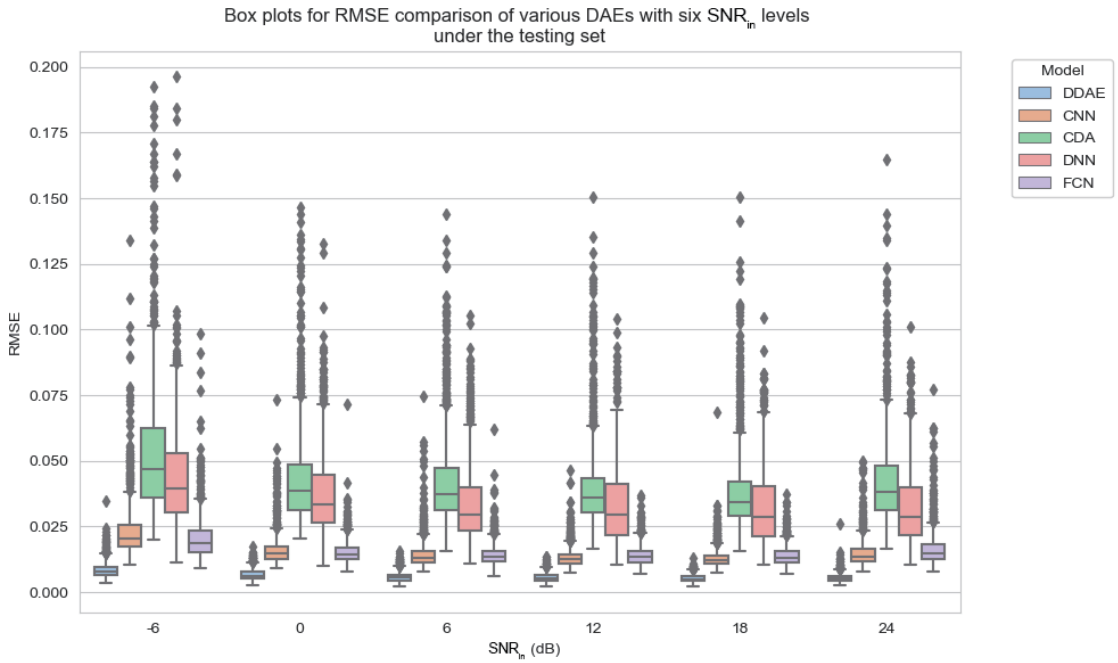

Fig. 7. (Color online) Box plots of *RMSE* of various DAEs with six SNR_{in} levels under the testing set.

The *RMSE* box plot in Fig. 7 further confirms DDAE's advantage, with lower median values across all SNR_{in} levels, particularly within the −6 to 12 dB range, where its error distribution displays greater concentration and lower variability. This suggests that DDAE accurately reconstructs PPG signals with reduced absolute errors, especially under high-noise conditions, outperforming CNN and CDA and demonstrating strong adaptability.

Overall, DDAE exhibits multifaceted superiority in PPG signal denoising tasks, with high SNR_{imp}, low *PRD*, and low *RMSE* indicating its ability to enhance signal quality, reduce reconstruction errors, and maintain stability, particularly in low *SNR* ratio environments. These characteristics make DDAE an ideal choice for processing noisy PPG data, offering competitive denoising performance compared with CNN, CDA, DNN, and FCN, and making it well suited to applications in medical and signal processing fields where high signal quality is critical.

To further validate the practical deployment of the proposed DDAE model on resource-constrained hardware, we conducted additional real-world testing using PPG signals acquired with the MAX30102 sensor (a low-cost, reflectance-based pulse oximeter module widely used in wearable prototypes). These signals were collected from healthy volunteers during controlled daily activities and processed in real time on a Samsung Galaxy A33 5G mobile device. The Samsung Galaxy A33 5G is equipped with an Exynos 1280 chipset featuring an octa-core CPU (2x 2.4 GHz ARM Cortex-A78 and 6x 2.0 GHz ARM Cortex-A55), 6 GB or 8 GB RAM, and a Mali-G68 GPU, making it suitable for testing the model's performance on resource-constrained hardware. Figures 8 illustrates the denoising performance of the DDAE model on PPG signals processed on this device. Figures 8 is divided into two sections: the upper half presents the unfiltered PPG signal (blue curve), which exhibits significant noise interference, while the lower half displays the denoised output (red curve) processed by the DDAE model. The filtered signal

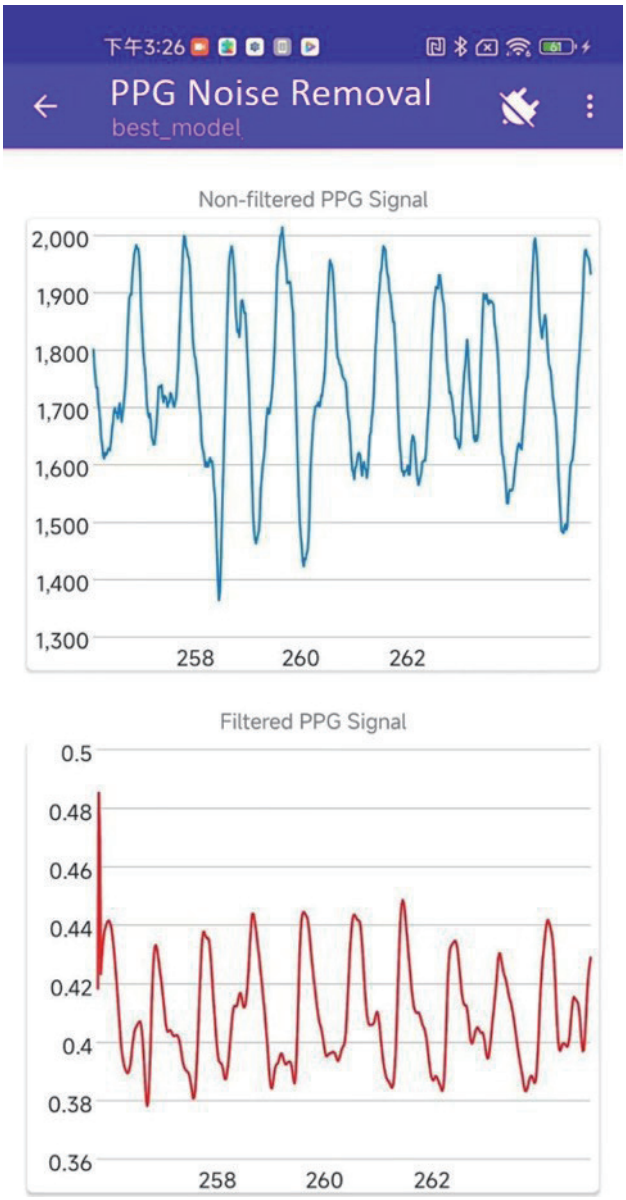

Fig. 8. (Color online) Real-time PPG signal denoising of the DDAE model on a mobile device.

shows a notable reduction in noise, with the periodic pulse wave characteristics becoming more distinct. This demonstrates the model's effectiveness in suppressing noise while preserving physiological signal integrity on a mid-range mobile device.

4. Conclusions

In this study, we proposed a denoising autoencoder-based approach for enhancing the quality of PPG signals contaminated by various types of noise. The proposed model, built upon a CNN architecture, demonstrated effective noise reduction capabilities while preserving the morphological features critical for downstream biomedical analysis. Through both qualitative

waveform comparisons and quantitative performance metrics, the results confirmed that our method outperforms other existing models, particularly in scenarios involving motion artifacts and BW. DDAE's performance may degrade at *SNR* levels below −6 dB, where noise dominates signal features. Additionally, its reliance on SQI preprocessing may limit applicability to raw signals with severe artifacts. Our enhancement paves the way for more accurate physiological monitoring in wearable health devices. In future work, we will explore the integration of the model with real-time embedded systems, further evaluate its robustness across diverse populations and acquisition conditions, and address the above limitations by incorporating adaptive preprocessing and testing on additional datasets.

Acknowledgments

This work was supported in part by National Science and Technology Council, Taiwan, under Grant NSTC 113-2221-E-150-002, in part by Smart Machinery and Intelligent Manufacturing Research Center, National Formosa University, Yunlin, Taiwan.

References

1 P. H. Charlton, P. A. Kyriacou, J. Mant, V. Marozas, P. Chowienczyk, and J. Alastruey: Proc. IEEE **110** (2022) 355. https://doi.org/10.1109/JPROC.2022.3149785
2 J. Allen: Physiol. Meas. **28** (2007) R01. https://doi.org/10.1088/0967-3334/28/3/R01
3 M. Nardelli and R. Bailón: Sensors **23** (2023) 7064. https://doi.org/10.3390/s23167064
4 E. Mejía-Mejía and P. A. Kyriacou: Biomed. Signal Process. Control **80** (2023) 104291. https://doi.org/10.1016/j.bspc.2022.104291
5 F. Peng, Z. Zhang, X. Gou, H. Liu, and W. Wang: Biomed. Eng. Online **13** (2014) 1. https://doi.org/10.1186/1475-925X-13-50.
6 R. Ahmed, A. Mehmood, M. M. U. Rahman, and O. A. Dobre: IEEE Sens. Lett. **7** (2023) 3285135. https://doi.org/10.1109/LSENS.2023.3285135
7 P. Xiong, H. Wang, M. Liu, F. Lin, Z. Hou, and X. Liu: Physiol. Meas. **37** (2016) 2214. https://doi.org/10.1088/0967-3334/37/12/2214
8 F. Mohagheghian, S. Tewari, R. Rajan, S. Jha, R. Santhanam, and A. S. Pandharipande: IEEE Trans. Biomed. Eng. **71** (2024) 456. https://doi.org/10.1109/TBME.2023.3307400
9 H. T. Chiang, Y. Y. Hsieh, S. W. Fu, K. H. Hung, Y. Tsao, and S. Y. Chien: IEEE Access **7** (2019) 60806. https://doi.org/10.1109/ACCESS.2019.2912036
10 K. He, X. Zhang, S. Ren, and J. Sun: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (IEEE, 2015) 770. https://doi.org/10.1109/CVPR.2016.90.
11 F. Yu and V. Koltun: 4th Int. Conf. Learn. Representations (2016). https://arxiv.org/pdf/1511.07122 (accessed Jun. 2025).
12 L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille: IEEE Trans. Pattern Anal. Mach. Intell. **40** (2016) 834. https://doi.org/10.1109/TPAMI.2017.2699184
13 L. C. Chen, G. Papandreou, F. Schroff, and H. Adam: arXiv:1706.05587 (2017). https://doi.org/10.48550/arXiv.1706.05587.
14 PPG-DaLiA - UCI Machine Learning Repository. https://archive.ics.uci.edu/dataset/495/ppg+dalia (accessed Mar. 13, 2025).
15 A. Reiss, I. Indlekofer, P. Schmidt, and K. Van Laerhoven: Sensors **19** (2019) 3079. https://doi.org/10.3390/s19143079
16 MIT-BIH Noise Stress Test Database v1.0.0. https://physionet.org/content/nstdb/1.0.0/ (accessed May 27, 2025).