Enhancing Deep Wavelet ECG Denoising with Squeeze-and-Excitation Blocks

Ngoc-Son Nguyen[1], Thai-Duy Nguyen Huy[1], Minh-Tien Le[1], Huy-Dung Han[1], Tai Le[2], Hung Cao[2,3]
[1]*School of Electrical and Electronic Engineering (Hanoi University of Science and Technology), Hanoi, Vietnam*
[2]*Department of Biomedical Engineering, University of California Irvine, CA, USA*
[3]*Department of Electrical Engineering and Computer Science, University of California Irvine, CA, USA*
Email: son.nn6303@gmail.com , duy.nht213750@sis.hust.edu.vn, tien.lm213778@sis.hust.edu.vn
dung.hanhuy@hust.edu.vn, tail3@uci.edu, hungcao@uci.edu

***Abstract*—ECG denoising is crucial for improving signal quality while preserving clinically important waveform morphology. Recent DWT/IDWT-based deep wavelet convolutional networks have shown promising performance by retaining multi-resolution information during feature compression and reconstruction. However, the learned approximation- and detail-related feature channels may contribute unequally to ECG restoration, while standard convolutional blocks do not explicitly recalibrate their relative importance. To address this limitation, this paper proposes DW-SE, a stage-selective enhancement of DW-CNN that integrates Squeeze-and-Excitation (SE) blocks into selected stages of a wavelet-assisted 1-D encoder–decoder architecture. By combining DWT-based multi-resolution representation, skip connections, and adaptive channel-wise recalibration, DW-SE adjusts the relative contributions of feature channels during reconstruction. Experiments under baseline wander, electrode motion, and muscle artifact noise show that DW-SE consistently outperforms baseline models across all evaluated input SNR levels. Averaged over all settings, DW-SE achieves an RMSE of 0.183 and an output SNR of 10.46 dB, with the best case reaching an RMSE of 0.15 and an output SNR of 12.37 dB. Beyond these controlled experiments, a qualitative case study on real-world BKIC recordings further suggests that DW-SE can suppress non-stationary disturbances while retaining major ECG waveform structures.**

***Index Terms*—ECG denoising, discrete wavelet transform, squeeze-and-excitation, deep learning.**

I. Introduction

Electrocardiogram (ECG) denoising remains challenging because noise suppression must be achieved without distorting diagnostically important waveform structures, including the P wave, QRS complex, and T wave. Traditional wavelet-based methods are well suited to this task because they provide a multi-resolution representation of non-stationary ECG signals, and prior work has shown that different wavelet subbands should not be treated uniformly during denoising [1]. More recently, Jin *et al.* proposed a deep wavelet convolutional neural network (DW-CNN) that replaces conventional pooling and up-sampling with DWT- and IDWT-based operations in an encoder–decoder architecture, thereby reducing information loss during feature compression and reconstruction [2].

However, retaining multi-resolution features alone may not fully exploit the information provided by wavelet decomposition. After each DWT operation, the network preserves feature channels associated with approximation and detail components, which capture different aspects of the ECG waveform and may contribute unequally to reconstruction under different noise conditions. In the original DW-CNN architecture, these channels are processed by standard convolutional blocks without an explicit mechanism to adjust their relative importance. To address this limitation, we introduce Squeeze-and-Excitation (SE) blocks, which provide a lightweight mechanism for modeling inter-channel dependencies and adaptively recalibrating channel responses [3]. The relevance of attention-based feature recalibration to ECG denoising is also supported by recent multi-resolution attention models such as MrSeNet [4]. Based on this motivation, we propose DW-SE, a stage-selective extension of DW-CNN that incorporates SE blocks into selected stages of a DWT/IDWT-based 1-D encoder–decoder architecture.

This study makes three main contributions by extending the DW-CNN architecture with SE blocks at selected encoder, bottleneck, and decoder stages to adaptively recalibrate wavelet-derived feature channels. Second, we evaluate the effect of SE-block placement through an ablation study. Third, we compare the proposed DW-SE model with existing baselines under baseline wander, electrode motion, and muscle artifact noise at multiple input SNR levels, and provide a qualitative case study on real-world BKIC recordings.

II. Methodology

A. Signal Preprocessing and Noise Synthesis

Segmentation: Following the DW-CNN preprocessing protocol [2], the input ECG signal is not subjected to any additional normalization, in order to avoid distorting amplitude information and to preserve the original characteristics of the ECG waveform. The original ECG recordings are uniformly segmented into fixed-length windows of 4096 samples, and these segments are treated as clean ECG signals for the subsequent noise synthesis stage and model training.

Energy-Controlled Noise Injection: From each retained clean ECG segment, we generate paired clean–noisy training data via controlled additive noise injection. A noise sequence $\eta[n]$ of length N is sampled from the MIT-BIH Noise Stress

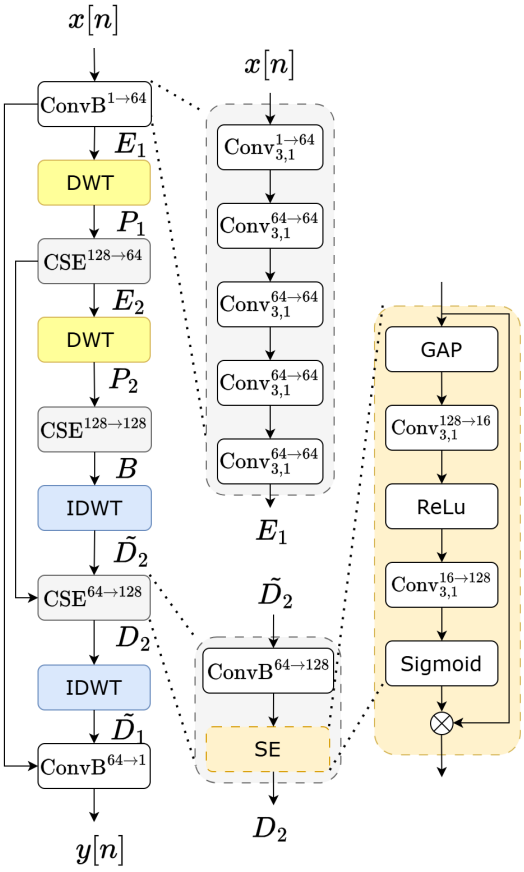

Fig. 1. Overview of the proposed DW-SE architecture with DWT/IDWT-based encoder–decoder stages and SE-enhanced convolutional blocks.

Test Database (NSTDB) [5], and the noisy observation is constructed as

$$x[n] \;=\; x_0[n] + a\,\eta[n] + b. \tag{1}$$

where a is a gain factor controlling the noise strength (thus determining the target SNR) and b is a constant DC offset. The result is discrete-time sequence $x[n]$ of length N, with $n = 0, 1, \ldots, N-1$. Under this construction, each segment corresponds to a single noise environment defined by one noise type and one SNR level. This formulation is consistent with the additive noise model in [6].

B. Proposed DW-SE Architecture

1) Notation and layer definitions: This section introduces the notation and the main building blocks used in the DW-SE model, including 1D convolution, skip connection, and the squeeze-and-excitation mechanism.

1D convolution: We denote a 1D convolution by

$$Q = \mathrm{Conv}_{k,\,s}^{C\to C'}(P), \tag{2}$$

where C and C' are the numbers of input and output channels, k is the kernel size, and s is the stride. Given an input tensor $P \in \mathbb{R}^{C\times N}$, the output is $Q \in \mathbb{R}^{C'\times \lfloor N/s \rfloor}$. In our implementation, all intermediate convolutions use kernel size $k = 3$ and stride $s = 1$, with padding chosen to preserve the temporal length.

Convolution block: Following the DW-CNN design, each convolution block consists of five consecutive 1D convolutions [2]. We denote such a block by

$$Q = \mathrm{ConvB}^{C\to C'}(P), \tag{3}$$

where $P \in \mathbb{R}^{C\times N}$ and $Q \in \mathbb{R}^{C'\times N}$. In the first and second encoder stages, the block outputs 64 channels. In the latent block and the decoder refinement block, the five-layer block is followed by one additional convolution so that the final output is restored to 128 channels, consistent with the original DW-CNN architecture [2].

Squeeze-and-excitation block: Given an input feature vector $X \in \mathbb{R}^{C\times N}$, the squeeze-and-excitation block first compresses the temporal dimension by global average pooling (GAP). For the c-th channel, let $x_c \in \mathbb{R}^{1\times N}$ denote the temporal feature vector of that channel. The pooled descriptor is then computed as

$$z_c = \frac{1}{N}\sum_{n=1}^{N} x_c(n), \qquad c = 1, 2, \ldots, C. \tag{4}$$

By stacking all channel descriptors, we obtain

$$Z = [z_1, z_2, \ldots, z_C]^T \in \mathbb{R}^{C\times 1}. \tag{5}$$

The descriptor Z is then passed through a pointwise convolution to reduce the channel dimension, followed by a ReLU activation:

$$U = \mathrm{ReLU}\Big(\mathrm{Conv}_{1,\,1}^{C\to C'}(Z)\Big), \qquad U \in \mathbb{R}^{C'\times 1}. \tag{6}$$

Next, the compressed representation is mapped back to the original channel dimension by another pointwise convolution, and the resulting channel weights are normalized by a sigmoid function. The final recalibrated output is obtained by channel-wise scaling:

$$Y = X \odot \sigma\Big(\mathrm{Conv}_{1,\,1}^{C'\to C}(U)\Big), \qquad Y \in \mathbb{R}^{C\times N}. \tag{7}$$

This mechanism explicitly models inter-channel dependencies and adaptively emphasizes informative channels while suppressing less useful ones [3].

Convolution-SE block: For notational simplicity, we define

$$\mathrm{CSE}^{C\to C'}(H) = \mathrm{SE}\Big(\mathrm{ConvB}^{C\to C'}(H)\Big). \tag{8}$$

Thus, each intermediate block first extracts features by convolution and then applies channel-wise recalibration.

Wavelet down/up operators: In DW-CNN, the DWT-based pooling layer replaces conventional pooling by decomposing the input into low-frequency and high-frequency components. According to [2], the outputs of a one-level DWT are defined as follows:

$$cA = \mathrm{DWT}_{\mathrm{low}}(H) = f_L \otimes H \downarrow 2, \qquad cA \in \mathbb{R}^{C\times N/2}, \tag{9}$$

$$cD = \mathrm{DWT}_{\mathrm{high}}(H) = f_H \otimes H \downarrow 2, \qquad cD \in \mathbb{R}^{C\times N/2}, \tag{10}$$

where f_L and f_H are the low-pass and high-pass filters, respectively, $\otimes$ denotes convolution, and $\downarrow 2$ denotes downsampling by a factor of 2. In the notation of the original paper,

the low-frequency output corresponds to the approximation coefficients and the high-frequency output corresponds to the detail coefficients.

The corresponding inverse transform reconstructs the higher-resolution representation from the low- and high-frequency branches:

$$H = \mathrm{IDWT}(cA, cD), \qquad H \in \mathbb{R}^{C\times N}. \tag{11}$$

2) Encoder: Let the noisy ECG segment be represented by $x[n] \in \mathbb{R}^{1\times N}$. In the first encoder stage, the signal is processed by a Convolution block,

$$E_1 = \mathrm{ConvB}^{1\to 64}(x[n]), \tag{12}$$

producing a 64-channel feature tensor $E_1 \in \mathbb{R}^{64\times N}$. This tensor is preserved for the first skip connection. A one-level wavelet decomposition is then applied:

$$cA_1 = \mathrm{DWT}_{\mathrm{low}}(E_1), \qquad cA_1 \in \mathbb{R}^{64\times N/2}, \tag{13}$$

$$cD_1 = \mathrm{DWT}_{\mathrm{high}}(E_1), \qquad cD_1 \in \mathbb{R}^{64\times N/2}, \tag{14}$$

where both cA_1 and cD_1 lie in $\mathbb{R}^{64\times N/2}$. Their stacking along the channel dimension forms the pooled representation

$$P_1 = \begin{bmatrix} cA_1 \\ cD_1 \end{bmatrix}, \qquad P_1 \in \mathbb{R}^{128\times N/2}. \tag{15}$$

In the second encoder stage, this pooled representation is processed by

$$E_2 = \mathrm{CSE}^{128\to 64}(P_1), \tag{16}$$

yielding $E_2 \in \mathbb{R}^{64\times N/2}$. This tensor is preserved for the second skip connection. A second wavelet decomposition is then applied:

$$cA_2 = \mathrm{DWT}_{\mathrm{low}}(E_2), \qquad cA_2 \in \mathbb{R}^{64\times N/4}, \tag{17}$$

$$cD_2 = \mathrm{DWT}_{\mathrm{high}}(E_2), \qquad cD_2 \in \mathbb{R}^{64\times N/4}. \tag{18}$$

Their stacking along the channel dimension gives

$$P_2 = \begin{bmatrix} cA_2 \\ cD_2 \end{bmatrix}, \qquad P_2 \in \mathbb{R}^{128\times N/4}. \tag{19}$$

3) Bottleneck: At the bottleneck, the compressed wavelet-informed representation $P_2 \in \mathbb{R}^{128\times N/4}$ is refined by a convolution-SE block whose final output is restored to 128 channels:

$$B = \mathrm{CSE}^{128\to 128}(P_2), \qquad B \in \mathbb{R}^{128\times N/4}. \tag{20}$$

This stage preserves the latent design of DW-CNN while augmenting it with explicit channel-wise recalibration [2], [3].

4) Decoder: Before the first inverse wavelet transform, the bottleneck representation $B \in \mathbb{R}^{128\times N/4}$ is interpreted as the channel-wise stacking of the reconstructed low- and high-frequency components at the second wavelet level:

$$B = \begin{bmatrix} c\hat{A}_2 \\ c\hat{D}_2 \end{bmatrix}, \qquad c\hat{A}_2, c\hat{D}_2 \in \mathbb{R}^{64\times N/4}. \tag{21}$$

Here, $\mathrm{c}\hat{A}_2$ and $\mathrm{c}\hat{D}_2$ denote the learned approximation and detail features produced by the bottleneck block, respectively.

The first decoder stage then reconstructs the intermediate-resolution feature tensor by inverse wavelet transform:

$$\tilde{D}_2 = \mathrm{IDWT}(c\hat{A}_2, c\hat{D}_2), \qquad \tilde{D}_2 \in \mathbb{R}^{64\times N/2}. \tag{22}$$

This tensor is fused with the second encoder output by element-wise addition:

$$\hat{D}_2 = \tilde{D}_2 + E_2. \tag{23}$$

The fused representation is then refined by a convolution-SE block, producing

$$D_2 = \mathrm{CSE}^{64\to 128}(\hat{D}_2), \qquad D_2 \in \mathbb{R}^{128\times N/2}. \tag{24}$$

Before the second inverse wavelet transform, the tensor D_2 is again interpreted as the stacking of a low-frequency branch and a high-frequency branch:

$$D_2 = \begin{bmatrix} c\hat{A}_1 \\ c\hat{D}_1 \end{bmatrix}, \qquad c\hat{A}_1, c\hat{D}_1 \in \mathbb{R}^{64\times N/2}. \tag{25}$$

Here, $\mathrm{c}\hat{A}_1$ and $\mathrm{c}\hat{D}_1$ denote the reconstructed approximation and detail features at the first wavelet level.

The second decoder stage restores the original temporal resolution:

$$\tilde{D}_1 = \mathrm{IDWT}(c\hat{A}_1, c\hat{D}_1), \qquad \tilde{D}_1 \in \mathbb{R}^{64\times N}. \tag{26}$$

This output is fused with the first encoder output:

$$\hat{D}_1 = \tilde{D}_1 + E_1, \qquad \hat{D}_1 \in \mathbb{R}^{64\times N}, \tag{27}$$

Finally, the output block maps the fused representation back to the denoised ECG waveform:

$$y[n] = \mathrm{ConvB}^{64\to 1}(\hat{D}_1), \qquad y[n] \in \mathbb{R}^{1\times N}. \tag{28}$$

where $y[n] \in \mathbb{R}^{1\times N}$ is the final denoised signal. Here, $ConvB^{64\to 1}$ denotes the final convolution block, whose last convolution layer projects the 64-channel fused representation to a single output channel corresponding to the denoised ECG signal.

III. Experiments

To ensure the robustness and reproducibility of the proposed model across diverse noise conditions, we aggregated data from two public databases and one locally collected dataset.

A. Data Description

MIT-BIH Arrhythmia Database: The MIT-BIH Arrhythmia Database [7], [8] serves as the primary source of clean ECG signals. In this study, ten specific records (103, 105, 111, 116, 122, 205, 213, 219, 223, and 230) were selected to provide a balanced representation of various cardiac morphologies and rhythms. The MLII lead was extracted for analysis due to its superior morphological consistency compared to V1/V5. All signals were resampled to 360 Hz and processed through a bandpass filter (0.67–100 Hz) to eliminate baseline drift and high-frequency artifacts.

MIT-BIH Noise Stress Test Database (NSTDB): To simulate real-world contamination, we utilized the MIT-BIH

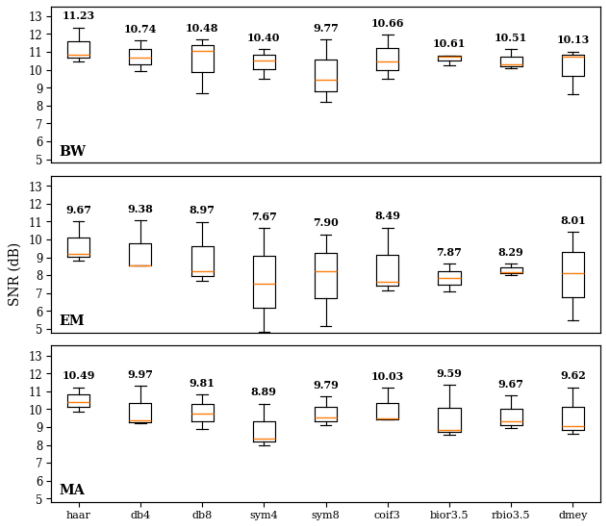

Fig. 2. Output SNR distributions of DW-SE using different wavelet bases under BW, EM, and MA noise, aggregated across input SNR levels of 0, 1.25, and 5 dB. The values above the boxplots denote the mean output SNR.

NSTDB [5], which contains three representative noise types: baseline wander (BW), muscle artifact (MA), and electrode motion (EM). These noise profiles were additively injected with the clean ECG segments to generate synthetic noisy inputs at SNR levels $\{0, 1.25, 5\}$ dB

BKIC Dataset: The BKIC dataset [9], collected at the BKIC Laboratory, Hanoi University of Science and Technology, was employed for external validation. This dataset consists of single-lead (MLII) ECG recordings captured under non-stationary real-world conditions, including motion and respiration artifacts. The signals were collected simultaneously with other physiological modalities such as phonocardiogram (PCG) and photoplethysmogram (PPG) using a custom non-invasive wearable acquisition system with non-contact electrodes.

B. Dataset Splitting

Following the partitioning strategy in [2], the MIT-BIH-derived dataset was split at the segment level within each noisy record. For each noise configuration, 80% of the segments were used for training, 10% for validation, and the remaining 10% for internal testing. The same input SNR levels of $0, 1.25, 5$ dB were used across all subsets. The BKIC dataset was excluded from training and validation and was used only as a qualitative case study to illustrate model behavior under independently collected real-world noisy conditions.

C. Training setup

The proposed DW-SE model was implemented in PyTorch and trained on an NVIDIA GPU using supervised learning with mean squared error (MSE) loss. Training was performed for 100 epochs with a batch size of 512 using the Adam optimizer and an initial learning rate of 1×10^{-3}. Each input ECG segment contained 4096 samples at a sampling rate

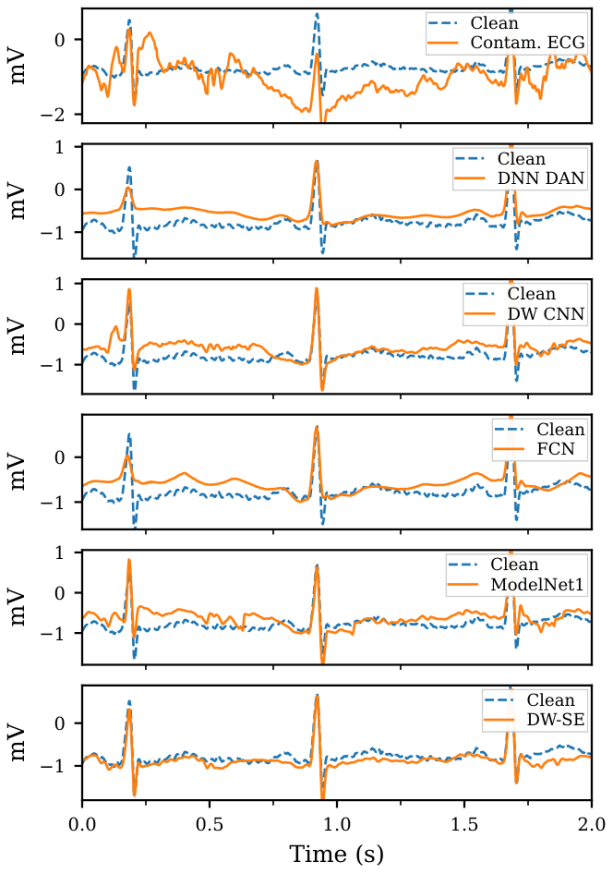

Fig. 3. Qualitative comparison of ECG waveforms denoised by different models under EM noise at 0 dB input SNR, compared with the clean reference signal.

of 360 Hz. Within each SE module, the channel descriptors are reduced by a factor of 8 before being expanded back to the original number of feature channels. To examine the influence of the wavelet basis, DW-SE was independently trained and evaluated under the same protocol using nine wavelet functions: sym4, bior3.5, dmey, db8, rbio3.5, db4, coif3, sym8, and haar.

D. Evaluation setup

The proposed DW-SE model was compared with four baselines: DW-CNN [2], ModelNet1 [2], DNN-DAN [10], and FCN [11]. The internal test set consisted of 10% of the segments generated from the selected MIT-BIH records and was excluded from training. Model performance was evaluated under BW, EM, and MA noise synthesized using NSTDB recordings at input SNR levels of $0, 1.25, 5$ dB. The influence of the wavelet basis on DW-SE was examined using nine independently trained wavelet configurations, while the main baseline comparison and SE-block placement ablation were reported using the Haar wavelet. For latency measurement, five warm-up iterations were followed by 100 timed runs with CUDA synchronization. The reported latency is the average per-sample inference time at a batch size of 512 on an NVIDIA GeForce RTX 3090 Ti in FP32 mode.

TABLE I
Performance comparison of denoising models using the Haar wavelet based on RMSE and output SNR under various noise conditions.

Metric	SNRin	Model	BW	EM	MA
RMSE	0	DW-CNN	0.28 ± 0.11	0.30 ± 0.10	0.30 ± 0.11
		ModelNet1	0.30 ± 0.11	0.31 ± 0.10	0.30 ± 0.11
		DNN_DAN	0.23 ± 0.10	0.29 ± 0.09	0.25 ± 0.12
		FCN	0.24 ± 0.09	0.30 ± 0.09	0.25 ± 0.12
		DW-SE	$\mathbf{0.19 \pm 0.10}$	$\mathbf{0.21 \pm 0.08}$	$\mathbf{0.20 \pm 0.11}$
	1.25	DW-CNN	0.28 ± 0.10	0.28 ± 0.09	0.28 ± 0.10
		ModelNet1	0.30 ± 0.11	0.29 ± 0.09	0.29 ± 0.12
		DNN_DAN	0.20 ± 0.10	0.25 ± 0.09	0.24 ± 0.10
		FCN	0.19 ± 0.09	0.25 ± 0.08	0.26 ± 0.11
		DW-SE	$\mathbf{0.18 \pm 0.10}$	$\mathbf{0.20 \pm 0.08}$	$\mathbf{0.19 \pm 0.11}$
	5	DW-CNN	0.23 ± 0.11	0.23 ± 0.06	0.23 ± 0.09
		ModelNet1	0.25 ± 0.11	0.23 ± 0.06	0.23 ± 0.09
		DNN_DAN	0.22 ± 0.09	0.21 ± 0.07	0.20 ± 0.09
		FCN	0.22 ± 0.10	0.21 ± 0.06	0.22 ± 0.09
		DW-SE	$\mathbf{0.15 \pm 0.09}$	$\mathbf{0.16 \pm 0.06}$	$\mathbf{0.17 \pm 0.08}$
SNR	0	DW-CNN	6.39 ± 3.50	5.65 ± 3.12	5.98 ± 3.50
		ModelNet1	5.84 ± 3.79	5.43 ± 2.91	5.86 ± 3.56
		DNN_DAN	8.41 ± 3.43	5.94 ± 2.88	7.60 ± 3.11
		FCN	7.77 ± 4.02	5.50 ± 3.10	7.50 ± 3.16
		DW-SE	$\mathbf{10.47 \pm 4.40}$	$\mathbf{8.82 \pm 3.74}$	$\mathbf{9.85 \pm 4.19}$
	1.25	DW-CNN	6.32 ± 3.90	6.36 ± 2.77	6.26 ± 3.81
		ModelNet1	5.94 ± 4.25	5.87 ± 2.99	6.37 ± 3.33
		DNN_DAN	9.46 ± 3.78	7.31 ± 3.02	7.88 ± 3.58
		FCN	10.05 ± 3.81	7.28 ± 3.18	7.22 ± 3.24
		DW-SE	$\mathbf{10.84 \pm 4.62}$	$\mathbf{9.19 \pm 3.54}$	$\mathbf{10.39 \pm 4.44}$
	5	DW-CNN	8.25 ± 4.37	7.85 ± 3.37	8.33 ± 3.50
		ModelNet1	7.78 ± 4.61	7.73 ± 3.34	8.23 ± 3.48
		DNN_DAN	8.70 ± 4.24	8.79 ± 3.04	9.54 ± 3.48
		FCN	8.75 ± 4.02	8.59 ± 3.22	8.57 ± 3.58
		DW-SE	$\mathbf{12.37 \pm 4.77}$	$\mathbf{11.01 \pm 3.69}$	$\mathbf{11.23 \pm 4.01}$

E. Evaluation Metrics

To quantitatively assess the proposed DW-SE model, RMSE and SNR are used as objective metrics to evaluate denoising quality and signal fidelity [2]. To further examine computational efficiency and real-time applicability, we report the End-to-End time ($E2E$), including host-to-device data transfer:

$$E2E = \frac{t_{finish} - t_{start}}{Batch\_Size} \times 1000\,(ms) \tag{29}$$

IV. Results

As shown in Fig. 2, the choice of wavelet basis affects the denoising performance of DW-SE. Among the evaluated wavelets, Haar achieves the highest mean output SNR across all three noise types, reaching 11.23 dB for BW, 9.67 dB for EM, and 10.49 dB for MA. EM generally produces lower output SNR values than BW and MA, indicating that it is the most challenging evaluated noise condition. Based on this post-hoc comparison, the subsequent baseline comparison and SE-block placement ablation are reported using the Haar wavelet.

Table I shows that the proposed DW-SE model consistently achieves the lowest RMSE and highest output SNR across all evaluated noise types (BW, EM, and MA) and input SNR levels ($\mathrm{SNR}_{in} = 0$, 1.25, and 5 dB). Averaged over the nine noise configurations, DW-SE obtains a mean RMSE of 0.183 and an average output SNR of 10.46 dB, outperforming the strongest baseline, DNN-DAN, by approximately 21.1% in RMSE and 2.28 dB in SNR. Its performance also improves consistently as the input SNR increases: the average RMSE decreases from 0.20 at 0 dB to 0.19 at 1.25 dB and 0.16 at 5 dB, while the corresponding output SNR increases from 9.71 dB to 10.14 dB and 11.54 dB, respectively. Among the evaluated noise types, EM remains the most challenging, producing higher RMSE and lower output SNR values across the compared models. Nevertheless, DW-SE maintains a clear performance advantage under all evaluated conditions.

As illustrated in Fig. 3, the waveform comparison under EM noise at $\mathrm{SNR}_{in} = 0$ dB shows that DW-SE produces a denoised signal that more closely follows the clean reference than the compared baselines. In particular, the model reduces residual fluctuations while retaining the major waveform structures, including the sharp transitions associated with the QRS complexes. The reconstructed waveform also avoids excessive smoothing, suggesting a favorable trade-off between noise suppression and morphology preservation under this representative condition.

As indicated by Table I and Fig. 2, EM is the most challenging evaluated noise type. Therefore, the SE-block placement ablation in Table II is conducted under EM noise at $\mathrm{SNR}_{in} = 0$ dB. In DW-SE, SE blocks are incorporated at three stages: the encoder, bottleneck, and decoder. Adding an SE block at any individual stage improves performance over the original DW-CNN, with bottleneck placement providing the strongest single-stage result. The complete DW-SE configuration achieves the best performance, reducing RMSE from 0.300 to 0.210 and increasing output SNR from 5.650 dB to 8.820 dB. These results indicate that multi-stage channel recalibration provides complementary benefits under challenging noise conditions.

Fig. 4 presents denoising results on representative real ECG signals from the BKIC dataset. Under BW noise, model differences are most noticeable in the final 2–3 s segment where weak ECG activity is affected by baseline drift. ModelNet1 and DW-CNN retain residual fluctuations, while DW-SE produces a more stable signal with improved baseline suppression.

The end-to-end efficiency comparison is shown in Fig. 5. DNN-DAN is the fastest model with an average latency of 2.11 ms and 0.23M parameters, while FCN requires 2.36 ms despite having the largest parameter count (0.74M). The proposed DW-SE model contains 0.37M parameters and requires 7.17 ms for inference, which is the highest latency among the compared models. Although not the most computationally efficient, DW-SE provides the best reconstruction quality.

V. Discussion and Conclusion

The results indicate that DW-SE improves ECG denoising under the evaluated conditions, achieving the lowest RMSE and highest output SNR among the compared models. Qualitative comparisons also suggest improved preservation of major waveform structures. This improvement can be attributed to the combination of DWT/IDWT-based multi-resolution recon-

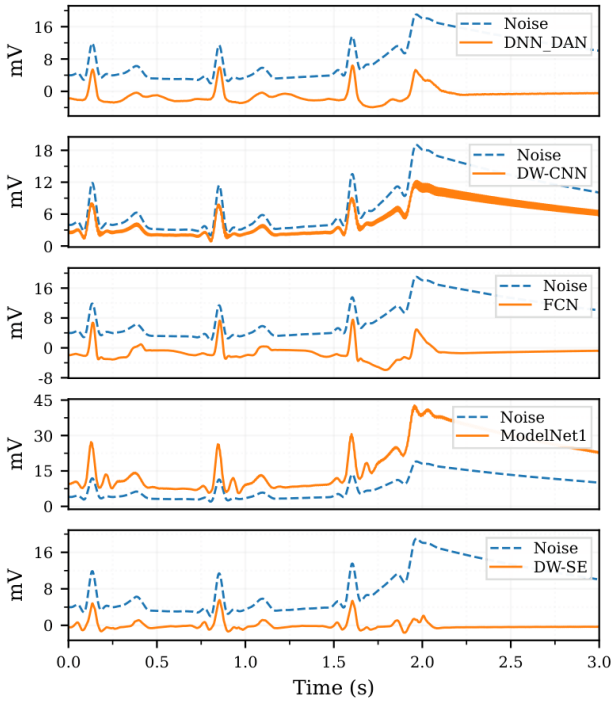

Fig. 4. Qualitative comparison of ECG denoising results on the BKIC dataset. Dashed lines denote noisy inputs, while solid lines denote the corresponding denoised outputs.

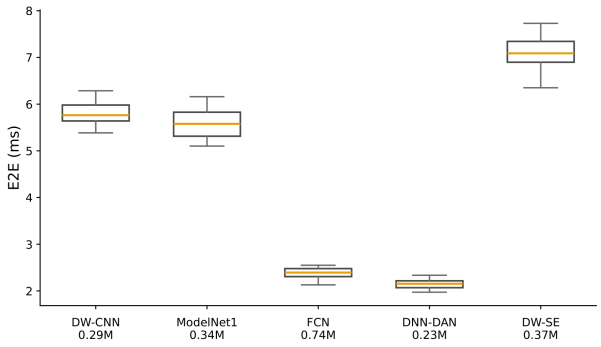

Fig. 5. Comparison of the evaluated models in terms of parameter count and end-to-end inference time. The boxplots summarize the distribution of per-sample latency across repeated runs.

struction, skip connections, and SE-based channel recalibration. Since wavelet-related features may contribute unequally to denoising [1], SE modules help adjust their relative importance by modeling inter-channel dependencies [3]. However, the improved reconstruction quality comes at the cost of higher inference latency.

Several limitations remain, as the quantitative evaluation is based mainly on synthetic NSTDB noise, while the BKIC dataset is used only as a qualitative case study because clean reference signals are unavailable. The study also does not include downstream clinical evaluation or embedded-device benchmarking. In conclusion, DW-SE provides an effective stage-selective extension of DW-CNN for improving ECG reconstruction quality. Future work will focus on broader clinical validation and lightweight implementations for real-time deployment.

TABLE II
Ablation study of SE-block placement at the encoder, bottleneck, and decoder stages under EM noise at an input SNR of 0 dB.

Configuration	Encd.	Bott.	Dec.	RMSE ↓	SNR (dB) ↑
DW-CNN	–	–	–	0.300 ± 0.100	5.650 ± 3.120
Case 1	✓	–	–	0.245 ± 0.095	7.539 ± 3.715
Case 2	–	✓	–	0.229 ± 0.093	8.165 ± 3.785
Case 3	–	–	✓	0.233 ± 0.093	7.990 ± 3.760
DW-SE	✓	✓	✓	$\mathbf{0.210 \pm 0.080}$	$\mathbf{8.820 \pm 3.740}$

REFERENCES

[1] S. Poornachandra and N. Kumaravel, "Subband-adaptive shrinkage for denoising of ecg signals," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, p. 081236, 2006.

[2] Y. Jin, C. Qin, J. Liu, Y. Liu, Z. Li, and C. Liu, "A novel deep wavelet convolutional neural network for actual ecg signal denoising," *Biomedical Signal Processing and Control*, vol. 87, p. 105480, 2024.

[3] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.

[4] Z. Wang, H. Xie, Y. Liu, H. Zhu, H. Zhang, Z. Wang, and Y. Pan, "Mrsenet: Electrocardiogram signal denoising based on multi-resolution residual attention network," *Journal of Electrocardiology*, vol. 89, p. 153858, 2025.

[5] G. B. Moody and R. G. Mark, "The mit-bih noise stress test database (nstdb)." https://physionet.org/content/nstdb/1.0.0/, 1995. Accessed: January 2026.

[6] E. Liu, Y. Liu, Z. Wang, H. Zhou, and X. Wang, "A noise prediction method for paper-based grayscale ecg denoising," *Biomedical Signal Processing and Control*, vol. 111, p. 108408, 2026.

[7] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[8] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[9] D.-A. Truong, K.-T. Hoang, H.-H. Nguyen, M.-T. Nguyen, H.-D. Han, and L. Pham-Nguyen, "Baseline wander removal in biomedical signal acquisition system for heart rate monitoring," in *2025 2nd International Conference on Health Science and Technology (ICHST)*, pp. 1–6, IEEE, 2025.

[10] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103, 2008.

[11] H.-T. Chiang, Y.-Y. Hsieh, S.-W. Fu, K.-H. Hung, Y. Tsao, and S.-Y. Chien, "Noise reduction in ecg signals using fully convolutional denoising autoencoders," *Ieee Access*, vol. 7, pp. 60806–60813, 2019.