MSECG: Incorporating Mamba for Robust and Efficient ECG Super-Resolution

Jie Lin[*†], I Chiu[*¶], Kuan-Chen Wang[*‡], Kai-Chun Liu[§], Hsin-Min Wang[¶], Ping-Cheng Yeh[‡], and Yu Tsao[¶]
[†]Department of Information Management, National Taiwan University, Taiwan [¶]Academia Sinica, Taiwan
[‡]Graduate Institute of Communication Engineering, National Taiwan University, Taiwan
[§]College of Information and Computer Sciences, University of Massachusetts, Amherst, USA
Email: b11705048@ntu.edu.tw, d13949002@ntu.edu.tw, d12942016@ntu.edu.tw, kaichunliu@umass.edu,
whmat@iis.sinica.edu.tw, pcyeh@ntu.edu.tw, yu.tsao@citi.sinica.edu.tw

***Abstract*—Electrocardiogram (ECG) signals play a crucial role in diagnosing cardiovascular diseases. To reduce power consumption in wearable or portable devices used for long-term ECG monitoring, super-resolution (SR) techniques have been developed, enabling these devices to collect and transmit signals at a lower sampling rate. In this study, we propose MSECG, a compact neural network model designed for ECG SR. MSECG combines the strength of the recurrent Mamba model with convolutional layers to capture both local and global dependencies in ECG waveforms, allowing for the effective reconstruction of high-resolution signals. We also assess the model's performance in real-world noisy conditions by utilizing ECG data from the PTB-XL database and noise data from the MIT-BIH Noise Stress Test Database. Experimental results show that MSECG outperforms two contemporary ECG SR models under both clean and noisy conditions while using fewer parameters, offering a more powerful and robust solution for long-term ECG monitoring applications.**

***Index Terms*—Electrocardiography, deep learning, Mamba, super-resolution, wearable device**

I. Introduction

Cardiovascular diseases pose significant life-threatening risks, often preventable with early detection of cardiac Arrhythmias (CAs) [1]. Electrocardiogram (ECG) serves as a non-invasive diagnostic tool, measuring the heart's electrical activity and aiding in monitoring cardiac anomalies [2]. To facilitate the detection of CAs in daily life, wearable or portable devices are commonly employed for long-term ECG monitoring [3], [4]. Reducing power consumption in these devices is crucial, and this can be achieved by collecting and transmitting ECG with a low sampling rate [5]. However, lowering the sampling rate risks compromising diagnostic accuracy due to information loss [6], and traditional upsampling methods often produce severe artifacts that further hinder accurate diagnosis [7], [8].

Super-resolution (SR) techniques aim to recover high-resolution (HR) data from low-resolution (LR) versions. Modern SR methods, leveraging neural networks (NNs), have found extensive applications in fields such as computer vision [9], [10] and audio processing [11], [12]. In ECG analysis, NN-based SR methods have emerged, such as SRECG [8], which applies a convolutional neural network (CNN) inspired by SRResNet [10] to reconstruct HR ECG from LR inputs. Similarly, DCAE-SR [13] utilizes a denoising autoencoder structure for ECG SR under noisy conditions. These studies demonstrate the feasibility of using NNs to recover HR ECG signals from low-sampling-rate inputs. However, existing approaches based on convolutional structures may be limited in capturing temporal dependencies inherent in data [14], [15], leading to constrained ECG reconstruction capability.

Mamba [16], a recently developed recurrent neural network (RNN), enhances traditional state-space models (SSMs) by introducing a selection mechanism and an efficient hardware-aware algorithm. Compared to Transformer [14], which also handles global dependencies but with higher time complexity, Mamba offers a more efficient solution. Owning to its advantages, Mamba has been applied across diverse fields like speech processing [17], computer vision [18], [19], and genomics [20], [21]. However, its potential for ECG SR has not yet been explored, as most existing approaches rely on CNNs.

In this study, we propose MSECG, a novel Mamba-based ECG SR model for reconstructing HR ECG with enhanced accuracy. MSECG combines convolutional layers with Mamba to capture both local and global information within data. Additionally, advanced SR techniques, such as pixel shuffle (PS) and skip connection (SC) [22], are incorporated to further boost performance and efficiency. Our experimental results show that MSECG outperforms contemporary methods on the PTB-XL ECG database, even under noisy conditions using data from the MIT-BIH Noise Stress Test Database (NSTDB). These findings underline MSECG's potential as a robust and effective SR solution for real-world ECG applications. To the best of our knowledge, this is the first work to explore Mamba's application in ECG SR.

II. Related Works

A. ECG Super-Resolution

Reducing the power consumption of wearable and portable devices is essential for long-term ECG monitoring. To address this, many studies have focused on developing SR techniques for ECG data, aiming to enable accurate analysis

[*] These authors contributed equally to this work.

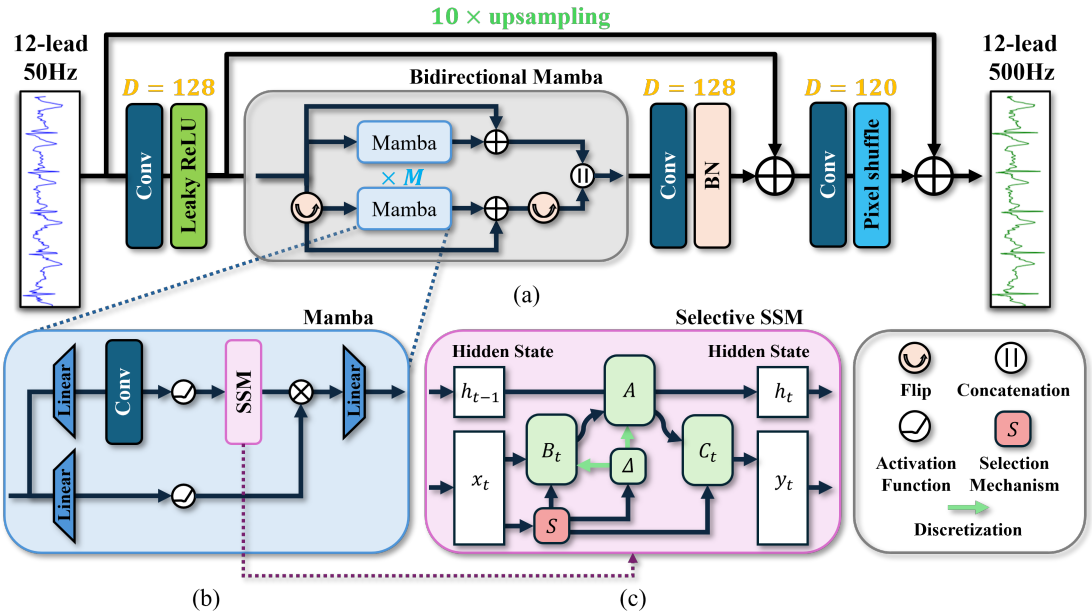

Fig. 1. The architecture of (a) MSECG, (b) Mamba, and (c) selective SSM.

by collecting LR signals. One such approach is SRECG [8], which uses SRResNet [10] as its foundation. It employs residual convolutional blocks to extract features from LR ECG signals, followed by two deconvolutional layers to upsample the signals and generate HR outputs. Another study, DCAE-SR [13], addresses the noise in ECG SR. DCAE-SR consists of an encoder and two decoders, each with four convolutional or deconvolutional layers. The encoder extracts latent features from the LR ECG signals, which are then fed into both decoders. One decoder reconstructs the original noisy LR ECG, while the other generates clean HR ECG signals. In our study, we implement both of these models to compare with our proposed method.

B. Mamba

Mamba [16] is an SSM designed for subquadratic-time computation. It addresses the limitations of Transformers [14], whose computational and memory requirements grow quadratically as the input sequence length increases. Mamba introduces a selection mechanism that makes the model input-dependent, allowing it to focus on or ignore specific information as needed. Mamba's architecture combines elements from the H3 model [23] and gated multi-layer perceptron (MLP) blocks [24], enabling it to be stacked homogeneously. This design expands the model's dimensionality, concentrating most parameters in linear projections, which enhances its ability to handle complex tasks.

In Mamba, structured SSMs map an input $\boldsymbol{x}$ to an output $\boldsymbol{y}$ through a higher-dimensional latent state $\boldsymbol{h}$, as described by the following equations:

$$h_n = \bar{\boldsymbol{A}} h_{n-1} + \bar{\boldsymbol{B}} x_n, \tag{1}$$

$$y_n = \boldsymbol{C} h_n, \tag{2}$$

where $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{B}}$ denote discretized state metrices. The discretization process transforms "continuous parameters" $(\boldsymbol{\Delta}, \boldsymbol{A}, \boldsymbol{B})$ into "discrete parameters" $(\bar{\boldsymbol{A}}, \bar{\boldsymbol{B}})$, enabling the model to seamlessly handle discrete-time data.

Mamba also features a hardware-aware algorithm that scales linearly with the input sequence length, enabling faster recurrent computations through efficient scanning. Despite its concise structure, Mamba consistently achieves state-of-the-art performance in various domains, including speech processing [17], computer vision [18], [19], and genomics [20], [21].

III. Proposed Method

A. Model architecture

The architecture of the proposed MSECG is illustrated in Fig. 1 (a). A convolutional layer initially extracts features from the input LR ECG. The features are then processed through a bidirectional Mamba block, retrieving information from both forward and backward directions in sequences. The structure of the Mamba block and selective SSM are provided in Fig. 1 (b) and (c), respectively. We consider that the Mamba blocks can better capture temporal information than the residual convolutional blocks used in SRECG [8]. The notations of D and M refer to the number of output channels in convolutional layers and the number of Mamba layers, respectively.

For upsampling, MSECG employs a one-dimensional PS operation, which is more efficient than traditional deconvolutional layers [25]. Fig. 2 illustrates the PS operation, where every ten channels in the feature maps are sequentially combined into a single output channel. Moreover, since modeling the entire upsampling process is relatively complex, we deploy a SC from the model's input to the output, which is inspired by advanced super-resolution methods in computer vision

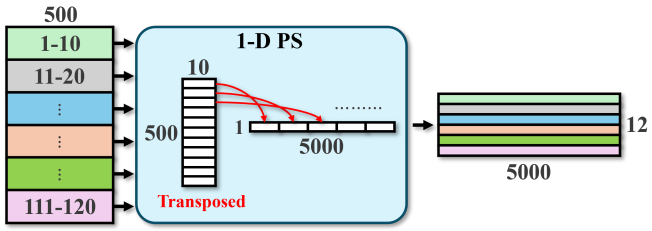

Fig. 2. Illustration of one-dimensional pixel shuffling (PS) operation.

[26]. Specifically, the LR input ECG is upsampled by linear interpolation (LI) with a ratio of 10 and then added directly to the MSECG output as a residual connection. This design can lead to more accurate reconstruction by enabling the model to focus on learning the difference between the LI-upsampled ECG and the target ECG.

IV. Experiments

A. Datasets

In this study, we use the PTB-XL dataset [27] as the source of clean ECG data. PTB-XL contains 21,799 clinical 12-lead ECG signals, each 10 seconds long, from 18,869 patients, with a sampling rate of 500 Hz. The dataset includes metadata for each ECG signal, such as weight, age, and sex, along with manual annotations for signal quality. PTB-XL is ideal for our experiments due to its large amount of high-quality ECG data.

For ECG noise data, we use the noise recordings from the MIT-BIH Noise Stress Test Database (NSTDB) [28] to emulate real-world noisy conditions, in contrast to the simulated noise used in the previous study [13]. The NSTDB includes three common types of noise: baseline wander (BW), muscle artifact (MA), and electrode motion artifact (EM). These recordings were taken from volunteers using two channels at a sampling rate of 360 Hz, lasting approximately 30 minutes.

B. Data pre-processing and preparation

We apply a second-order Butterworth band-pass filter with cutoff frequencies of 1 and 45 Hz to the 500 Hz ECG data from the PTB-XL dataset, which helps extract the most valuable information from the ECG signals [29]. The filtered ECG signals are treated as the ground truth (GT). Following [8], which demonstrates a significant performance drop with a downsampling factor of 10, we downsample the GT signals from 500 Hz to 50 Hz by skipping every 9 sequential time units to generate the LR ECG signals. From the PTB-XL standard, we use Folds 1-8 for training, Fold 9 for validation, and Fold 10 for testing.

To generate noisy ECG data, we downsample the noise signals from the MIT-BIH NSTDB from 360 Hz to 50 Hz to match the sampling rate of the LR ECG signals. Following the framework in [13], we introduce a 50% probability of noise contamination for each ECG segment. If noise is added, each type of noise (BW, MA, and EM) is equally likely to be selected. We then assign a signal-to-noise ratio (SNR) for each noise type, chosen randomly from a reasonable range [30]. Finally, a random noise segment is selected, adjusted in amplitude, and added to the LR ECG signal to generate noisy data with the specified SNR.

C. Evaluation metrics

We use four metrics for signal quality to evaluate the performance of SR approaches [13], [31]. In the equations below, $\boldsymbol{s}$ denotes the predicted ECG waveform, $\boldsymbol{g}$ is the GT waveform, and N is the number of samples in each segment.

• Mean Squared Error (MSE): Measures the average squared difference between $\boldsymbol{s}$ and $\boldsymbol{g}$.

$$\text{MSE} = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{s}[n] - \boldsymbol{g}[n])^2. \tag{3}$$

• Cosine Similarity (CoS): Measures the similarity between two vectors, $\boldsymbol{s}$ and $\boldsymbol{g}$.

$$\text{CoS} = \frac{\boldsymbol{s}\cdot\boldsymbol{g}}{\|\boldsymbol{s}\|\|\boldsymbol{g}\|}, \tag{4}$$

where $\|\cdot\|$ calculates the Euclidean norms of signals.

• SNR: Evaluates how well the signal can be distinguished from the noise by calculating the ratio of the power of $\boldsymbol{g}$ to the noise.

$$\text{SNR} = 10\cdot\log_{10}\left(\frac{\sum_{n=1}^{N}\boldsymbol{g}[n]^2}{\sum_{n=1}^{N}(\boldsymbol{s}[n]-\boldsymbol{g}[n])^2}\right). \tag{5}$$

• Maximum Absolute Distance (MAD): Measures the largest deviation between $\boldsymbol{s}$ and $\boldsymbol{g}$, which is critical for identifying short-period artifacts or noise.

$$\text{MAD} = \max|\boldsymbol{s}[n] - \boldsymbol{g}[n]|, \quad \text{for } 0 \le n \le N. \tag{6}$$

Lower MSE and MAD values reflect better SR performance, and higher SSIM and SNR values indicate better SR performance.

D. Implementation details

We use the Adam optimizer [32] for training. The batch size is set to 64, and we apply the L_2 loss function. Training is done in two stages: the model is first trained for 300 epochs with a learning rate of 1e-4, and the best-performing model on the validation set is saved. In the second stage, the saved model is further trained for 50 epochs with the learning rate reduced to 1e-5.

E. Results and discussion

We compare the performance of MSECG with the conventional linear interpolation (LI) method and two NN-based ECG SR methods: SRECG [8] and DCAE-SR [13]. The overall performance of these methods is shown in Table I. MSECG outperforms the other methods across all four quality metrics, demonstrating its superior ability to capture temporal information essential for reconstructing HR ECG signals. In contrast, SRECG and DCAE-SR rely on convolutional structures, which are less effective at leveraging temporal information.

TABLE I
OVERALL PERFORMANCE OF DIFFERENT ECG SR METHODS.

Method	Param. (M)	MSE ($\times 10^{-3}$) ↓	CoS ($\times 10^{-1}$) ↑	SNR (dB) ↑	MAD ↓
LI	-	7.477 ± 12.502	9.021 ± 0.895	8.592 ± 4.285	0.877 ± 0.403
SRECG [8]	3.05	0.422 ± 0.485	9.937 ± 0.047	19.751 ± 2.595	0.371 ± 0.214
DCAE-SR [13]	31.21	4.461 ± 13.472	9.756 ± 0.225	12.591 ± 2.894	0.962 ± 0.802
MSECG (ours)	**1.91**	**0.184 ± 0.335**	**9.975 ± 0.031**	**24.037 ± 2.851**	**0.221 ± 0.187**

Bold represents the best performance.

TABLE II
ABLATION STUDY FOR THE MODEL STRUCTURE OF MSECG.

Model	M	Param. (M)	DConv	PS	SC	MSE ($\times 10^{-3}$) ↓	CoS ($\times 10^{-1}$) ↑	SNR (dB) ↑	MAD ↓
MSECG	5	3.02	✓	-	-	0.206	9.972	23.508	0.228
	5	3.02	✓	-	✓	0.195	9.973	23.796	0.227
	5	1.91	-	✓	-	0.263	9.963	22.292	0.320
	5	1.91	-	✓	✓	**0.184**	**9.975**	**24.037**	**0.221**
	4	1.68	-	✓	✓	0.206	9.971	23.429	0.233
	6	2.14	-	✓	✓	<u>0.192</u>	<u>9.974</u>	<u>23.912</u>	<u>0.222</u>

Bold and <u>underline</u> represent the best and second best performance.

TABLE III
PERFORMANCE UNDER CLEAN AND NOISY SCENARIOS.

Model	Data	MSE ($\times 10^{-3}$) ↓	CoS ↑	SNR (dB) ↑
SRECG	clean	0.400	0.994	19.880
	noisy	0.443	0.994	19.546
DCAE-SR	clean	4.265	0.977	11.304
	noisy	4.648	0.974	10.717
MSECG (ours)	clean	0.165	0.998	24.294
	noisy	0.203	0.997	23.726

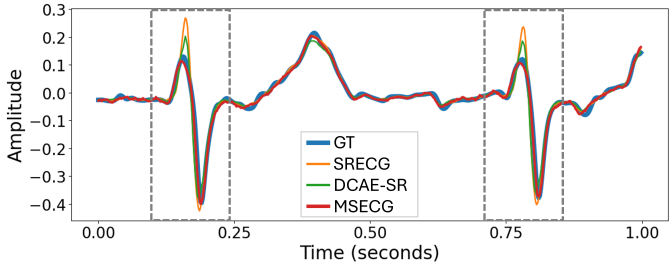

Fig. 3. Reconstructed ECG signals using different SR methods. Dashed boxes highlight areas where the differences are most pronounced.

Table II presents an ablation study that explores the effectiveness of different components within MSECG architecture. The results show that MSECG achieves the best performance by using the PS operation along with SC, which directly combines the linearly upsampled input with the output. Replacing deconvolutional (DConv) layers with PS not only reduces the number of model parameters but also prevents checkerboard artifacts [25] and other issues like ringing and aliasing [7] associated with DConv and LI. The SC further improves the model performance by allowing the main network to focus on learning the residual information. Furthermore, we found that using five Mamba layers ($M = 5$) strikes the best balance between performance and computational cost, which is why this configuration is used in MSECG.

To further assess the feasibility of ECG SR models in the real-world environment, Table III compares model performance in both clean and noisy conditions. Overall, noise degrades the performance across all metrics and models. However, MSECG continues to deliver the best results in both scenarios, demonstrating its robustness and potential for real-world ECG applications.

Fig. 3 shows the reconstructed HR ECG waveforms of different SR models. MSECG generates waveforms that are most similar to the GT signals, while both SRECG and DCAE-SR struggle to reconstruct the R-peaks accurately. Such distortion in ECG waveform and cardiac events could lead to misinterpretations in cardiovascular disease diagnosis [33], [34].

V. CONCLUSION

In this study, we introduce MSECG, a novel ECG SR model that combines Mamba and convolutional layers to effectively capture both local and global information for HR ECG reconstruction. The experimental results demonstrate that MSECG achieves superior performance while using fewer parameters than other methods. Furthermore, its effectiveness is sustained even in noisy environments, highlighting its superiority for real-world ECG applications. In the future, we plan to integrate MSECG with various downstream ECG tasks requiring long-term monitoring, such as CA classification, rhythm detection, and respiratory estimation.

References

[1] S. K. Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, "A Survey on ECG Analysis," *Biomedical Signal Processing and Control*, vol. 43, pp. 216–235, 2018.

[2] J. W. Hurst, "Naming of the Waves in the ECG, with a Brief Account of Their Genesis," *Circulation*, vol. 98, no. 18, pp. 1937–1942, 1998.

[3] N. Huda, S. Khan, R. Abid, S. B. Shuvo, M. M. Labib, and T. Hasan, "A Low-Cost, Low-Energy Wearable ECG System with Cloud-Based Arrhythmia Detection," in *Proc. TENSYMP*, 2020.

[4] Y. Xia et al., "An Automatic Cardiac Arrhythmia Classification System with Wearable Electrocardiogram," *IEEE Access*, vol. 6, pp. 16 529–16 538, 2018.

[5] Y. Nishikawa, S. Izumi, Y. Yano, H. Kawaguchi, and M. Yoshimoto, "Sampling Rate Reduction for Wearable Heart Rate Variability Monitoring," in *Proc. ISCAS*, 2018.

[6] G. P. Pizzuti, S. Cifaldi, and G. Nolfe, "Digital Sampling Rate and ECG Analysis," *Journal of biomedical engineering*, vol. 7, no. 3, pp. 247–250, 1985.

[7] A. S. Krylov, A. V. Nasonov, and A. A. Chernomorets, "Combined Linear Resampling Method with Ringing Control," in *Proc. GraphiCon*, 2009.

[8] T.-M. Chen et al., "SRECG: ECG Signal Super-Resolution Framework for Portable/Wearable Devices in Cardiac Arrhythmias Classification," *IEEE Transactions on Consumer Electronics*, vol. 69, no. 3, pp. 250–260, 2023.

[9] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[10] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using A Generative Adversarial Network," in *Proc. CVPR*, 2017.

[11] R. Yoneyama, R. Yamamoto, and K. Tachibana, "Nonparallel High-Quality Audio Super Resolution with Domain Adaptation And Resampling CycleGANs," in *Proc. ICASSP*, 2023.

[12] C.-Y. Yu, S.-L. Yeh, G. Fazekas, and H. Tang, "Conditioning And Sampling In Variational Diffusion Models For Speech Super-Resolution," in *Proc. ICASSP*, 2023.

[13] U. Lomoio, P. Veltri, P. H. Guzzi, and P. Lio, "DCAE-SR: Design Of A Denoising Convolutional Autoencoder For Reconstructing Electrocardiograms Signals At Super Resolution," *medRxiv preprint medRxiv:2024.04.08.24305091*, 2024.

[14] A. Vaswani et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.

[15] K.-C. Wang, K.-C. Liu, P.-C. Yeh, S.-Y. Peng, and Y. Tsao, "TrustEMG-Net: Using Representation-Masking Transformer with U-Net for Surface Electromyography Enhancement," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[16] A. Gu and T. Dao, "Mamba: Linear-Time Sequence Modeling with Selective State Spaces," *arXiv preprint arXiv:2312.00752*, 2023.

[17] R. Chao, W.-H. Cheng, M. L. Quatra, S. M. Siniscalchi, C.-H. H. Yang, S.-W. Fu, and Y. Tsao, "An Investigation of Incorporating Mamba for Speech Enhancement," *arXiv preprint arXiv:2405.06573*, 2024.

[18] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, "Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model," *arXiv preprint arXiv:2401.09417*, 2024.

[19] J. Ma, F. Li, and B. Wang, "U-Mamba: Enhancing Long-Range Dependency for Biomedical Image Segmentation," *arXiv preprint arXiv:2401.04722*, 2024.

[20] J. Zhang, C. Song, T. Cui, C. Li, and J. Ma, "ChiMamba: Predicting Chromatin Interactions Based on Mamba," in *Proc. ICIC*, 2024.

[21] V. Thoutam and D. Ellsworth, "MSAMamba: Adapting Subquadratic Sequence Models to Long-Context DNA MSA Analysis," in *Proc. ICIST*, 2024.

[22] W. Shi et al., "Real-Time Single Image And Video Super-Resolution Using An Efficient Sub-Pixel Convolutional Neural Network," in *Proc. CVPR*, 2016.

[23] D. Y. Fu, T. Dao, K. K. Saab, A. W. Thomas, A. Rudra, and C. Ré, "Hungry Hungry Hippos: Towards Language Modeling with State Space Models," in *Proc. ICLR*, 2023.

[24] H. Liu, Z. Dai, D. So, and Q. V. Le, "Pay Attention to MLPs," *Advances in Neural Information Processing Systems*, vol. 34, pp. 9204–9215, 2021.

[25] Y. Sugawara, S. Shiota, and H. Kiya, "Super-Resolution Using Convolutional Neural Networks without any Checkerboard Artifacts," in *Proc. ICIP*, 2018.

[26] M. V. Conde et al., "Deep RAW Image Super-Resolution. A NTIRE 2024 Challenge Survey," in *Proc. CVPR*, 2024.

[27] P. Wagner et al., "PTB-XL, a Large Publicly Available Electrocardiography Dataset," *Scientific data*, vol. 7, no. 1, pp. 1–15, 2020.

[28] G. B. Moody, W. E. Muldrow, and R. G. Mark, "A Noise Stress Test for Arrhythmia Detectors," *Computers in cardiology*, vol. 11, no. 3, pp. 381–384, 1984.

[29] T. Cao, N. Tran, L. Nguyen, H. Nguyen, and H. Pham, "IncepSE: Leveraging InceptionTime's Performance with Squeeze and Excitaion Mechanism in ECG Analysis," in *Proc. SOICT*, 2023.

[30] L. Hu, W. Cai, Z. Chen, and M. Wang, "A Lightweight U-Net Model for Denoising and Noise Localization of ECG Signals," *Biomedical Signal Processing and Control*, vol. 88, p. 105504, 2024.

[31] H. Li, G. Ditzler, J. Roveda, and A. Li, "DeScoD-ECG: Deep Score-Based Diffusion Model for ECG Baseline Wander and Noise Removal," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 9, pp. 5081–5091, 2023.

[32] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. ICLR*, 2015.

[33] L. Meng, K. Ge, Y. Song, D. Yang, and Z. Lin, "Long-Term Wearable Electrocardiogram Signal Monitoring and Analysis Based on Convolutional Neural Network," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[34] F. Liu et al., "Performance Analysis of Ten Common QRS Detectors on Different ECG Application Cases," *Journal of healthcare engineering*, vol. 2018, no. 1, p. 9050812, 2018.