



Robust Photoplethysmography Signal Denoising via Mamba Networks

I Chiu et al.

2026.04.02



Overview



01 Author & Journal

02 Introduction

03 Background

04 Method

05 Result

Author & Journal

ROBUST PHOTOPLETHYSMOGRAPHY SIGNAL DENOISING VIA MAMBA NETWORKS

I Chiu¹, Yu-Tung Liu², Kuan-Chen Wang³, Hung-Yu Wei⁴, and Yu Tsao²

¹Taiwan International Graduate Program in Artificial Intelligence of Things,
Academia Sinica and National Taiwan University, Taipei, Taiwan

²Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan

³Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan

⁴Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan

Email: d13949002@ntu.edu.tw, yu.tsao@citi.sinica.edu.tw

ABSTRACT

Photoplethysmography (PPG) is widely used in wearable health monitoring, but its reliability is often degraded by noise and motion artifacts, limiting downstream applications such as heart rate (HR) estimation. This paper presents a deep learning framework for PPG denoising with an emphasis on preserving physiological information. In this framework, we propose DPNet, a Mamba-based denoising backbone designed for effective temporal modeling. To further enhance denoising performance, the framework also incorporates a scale-invariant signal-to-distortion ratio (SI-SDR) loss to promote waveform fidelity and an auxiliary HR predictor (HRP) that provides physiological consistency through HR-based supervision. Experiments on the BIDMC dataset show that our method achieves strong robustness against both synthetic noise and real-world motion artifacts, outperforming conventional filtering and existing neural models. Our method can effectively restore PPG signals while maintaining HR accuracy, highlighting the complementary roles of SI-SDR loss and HR-guided supervision. These results demonstrate the potential of our approach for practical deployment in wearable healthcare systems.

Index Terms— Photoplethysmography (PPG), denoising, motion artifact removal, heart rate estimation, neural network

common strategy is the use of bidirectional long short-term memory (BLSTM) networks [6], where stacked BLSTM layers process noisy PPG segments and leverage bidirectional context to recover the underlying clean signals. However, BLSTMs suffer from their recurrent design, which restricts parallelism and results in slow inference. In addition, they have limited capacity to capture long-range dependencies. Transformer-based models [7] offer an alternative, excelling at capturing global temporal context and showing strong potential for PPG denoising. Yet, their quadratic complexity with respect to sequence length results in substantial computational cost [8, 9]. Meanwhile, limited labeled PPG data can make Transformers prone to underfitting, and even well-trained models remain difficult to deploy in real time on resource-constrained wearable devices [10]. In addition, most existing approaches focus primarily on waveform reconstruction using pointwise objective functions (e.g., L1 or L2), with limited attention to downstream tasks such as HR estimation, where subtle distortions may undermine clinical reliability.

To address both the modeling inefficiencies and the lack of physiological awareness in prior works, we introduce a novel deep learning framework for PPG denoising that builds upon the Mamba [11] selective state-space model (SSM). Mamba is an emerging sequence-modeling architecture that maintains linear-time complex-

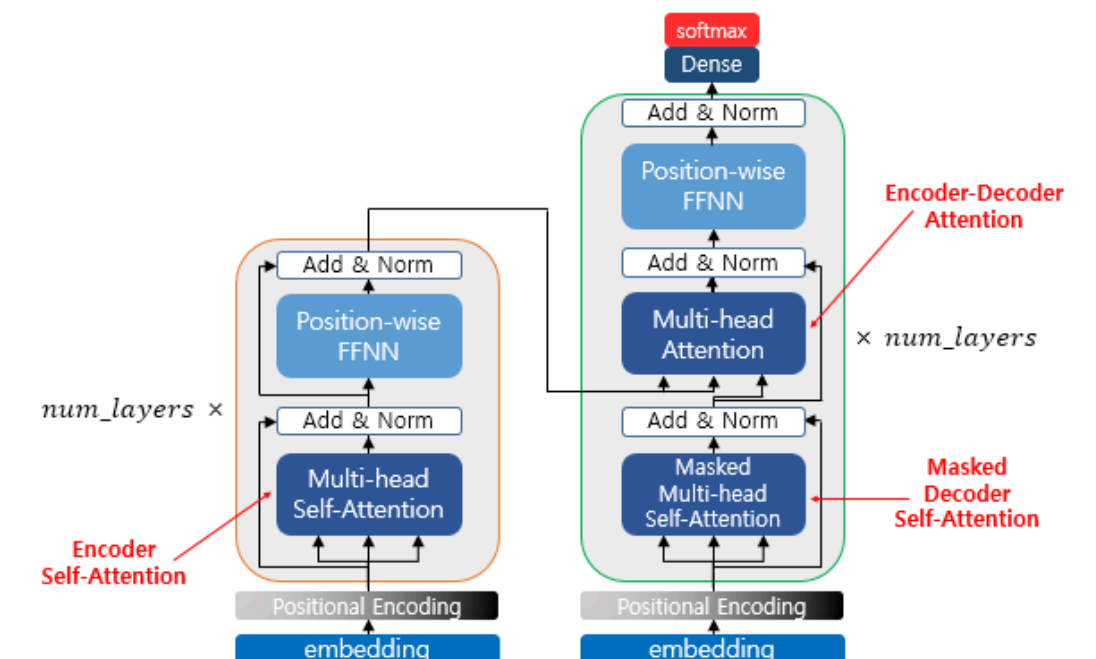
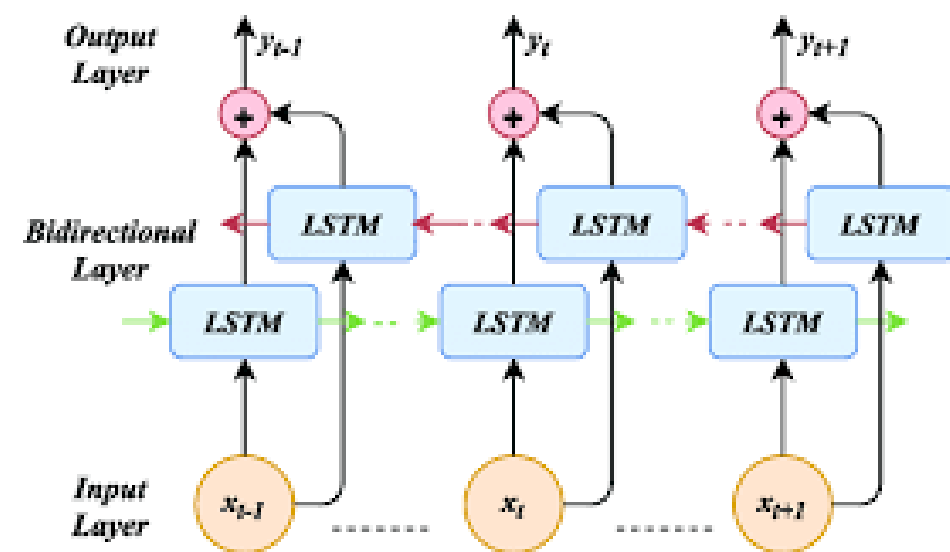
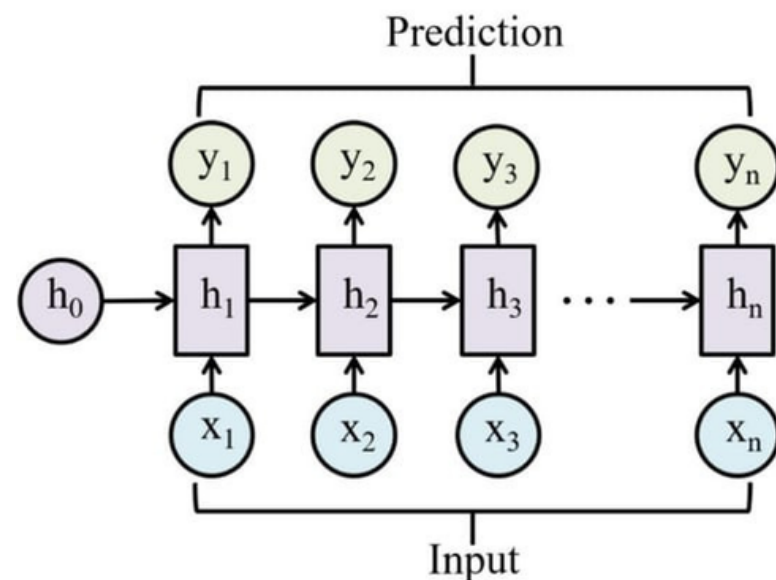
10.11058v1 [cs.LG] 13 Oct 2025

- 소속: Academia Sinica + NTU (대만)
- 저자: I Chiu(주저자), Yu-Tung Liu, Kuan-Chen Wang, Hung-Yu Wei, Yu Tsao
- arXiv(2025)

Introduction

1) 논문 배경 (Technical Summary)

- 신호 품질 저하 문제: 센서, 피부 움직임, 혈류 변화, 주변 광원 변동 같은 요인들이 파형 왜곡 및 하위 작업의 성능을 저하시키는 아티팩트 발생
- 심층 학습 접근 방식의 제약:
 - **Recurrent Neural Networks (RNNs) 및 BLSTM:**
 - BLSTM(Bidirectional Long Short-Term Memory) 네트워크와 같은 RNN은 병렬 처리가 제한적
 - 추론 속도가 느리며, 장거리 의존성을 포착하는 능력이 부족
 - **Transformer 기반 모델:**
 - Transformer는 전역적인 시간적 맥락을 포착하는데 뛰어나지만, 시퀀스 길이에 대한 2차 복잡도로 인해 상당한 계산 비용 발생
 - 제한된 라벨링된 PPG 데이터로 인해 과소적합되기 쉽고, 웨어러블 장치에 실시간으로 배포하기 어려움
- 하위 작업에 대한 인식 부족:
 - 대부분의 기존 접근 방식은 파형 재구성(예: L1 또는 L2 손실 사용)에 주로 초점
 - HR 추정 등 하위 작업에 대한 고려가 제한적이어서 미묘한 왜곡이 임상적 신뢰성 훼손



Background

1) Mamba

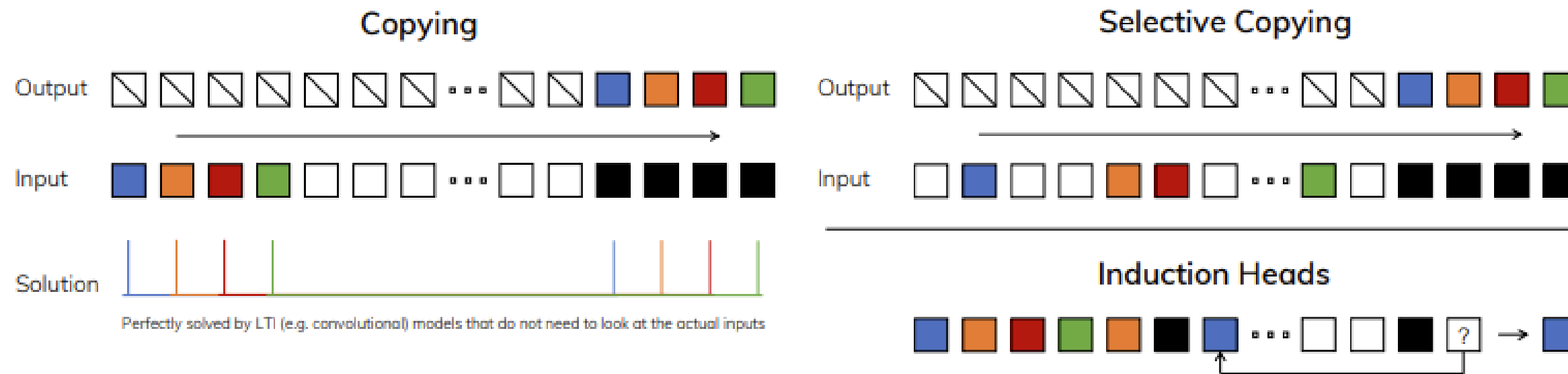


Figure 2: (Left) The standard version of the Copying task involves constant spacing between input and output elements and is easily solved by time-invariant models such as linear recurrences and global convolutions. (Right Top) The Selective Copying task has random spacing in between inputs and requires time-varying models that can *selectively* remember or ignore inputs depending on their content. (Right Bottom) The Induction Heads task is an example of associative recall that requires retrieving an answer based on context, a key ability for LLMs.

Selectivity

- * Problem: 기존 LTI(Linear Time-Invariant, 선형 시간 불변 시스템) 모델의 한계
- **recurrent 관점:** LTI 모델의 constant dynamics (ex. A_{bar} , B_{bar})는 context에서 올바른 정보를 선택하거나 입력 의존적으로 은닉 상태를 변경할 수 없음
- **convolution 관점:** Selective Copying Task에서는 입력과 출력 간 간격이 변화하여, 이를 정적 convolution kernel로 모델링 할 수 없음

* Solution → 각 token마다 B, C, delta가 다를 수 있게 하자.

1) Mamba

SSM: 시스템 내부의 상태 변화를 매개로 하여 출력을 산출하는 모델

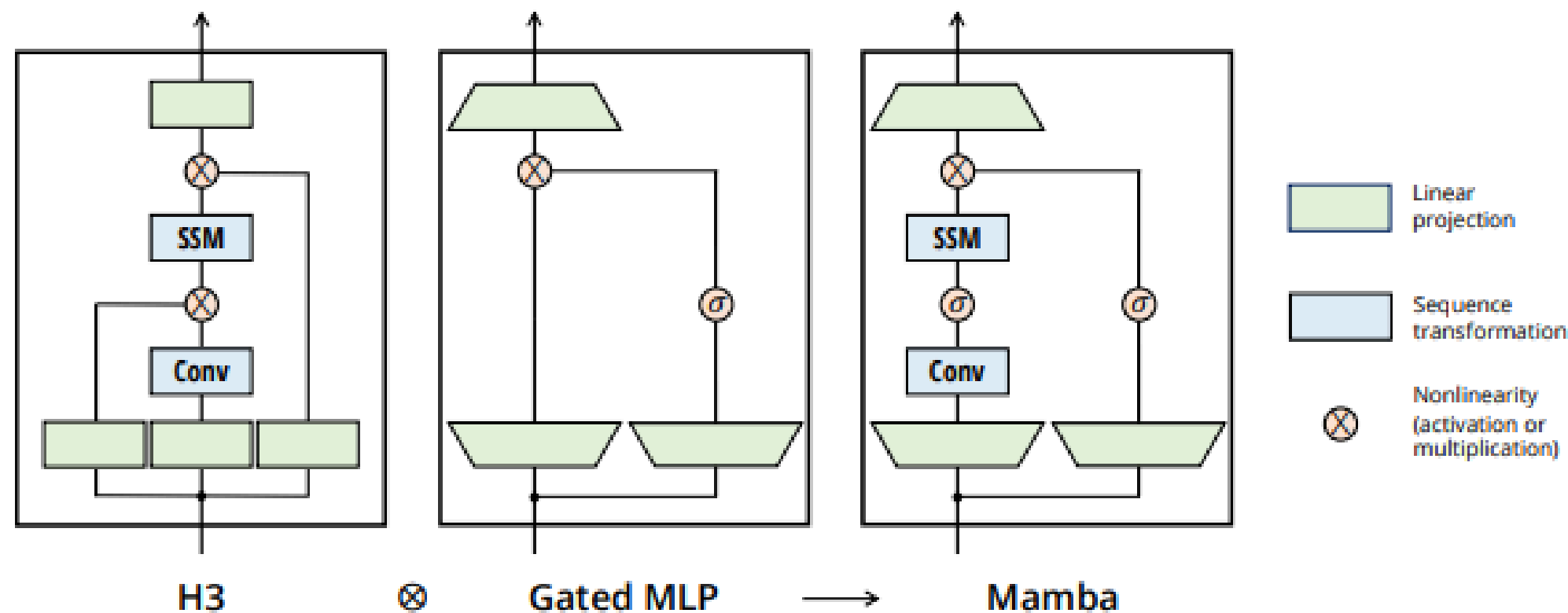


Figure 3: (Architecture.) Our simplified block design combines the H3 block, which is the basis of most SSM architectures, with the ubiquitous MLP block of modern neural networks. Instead of interleaving these two blocks, we simply repeat the Mamba block homogeneously. Compared to the H3 block, Mamba replaces the first multiplicative gate with an activation function. Compared to the MLP block, Mamba adds an SSM to the main branch. For σ we use the SiLU / Swish activation (Hendrycks and Gimpel 2016; Ramachandran, Zoph, and Quoc V Le 2017).

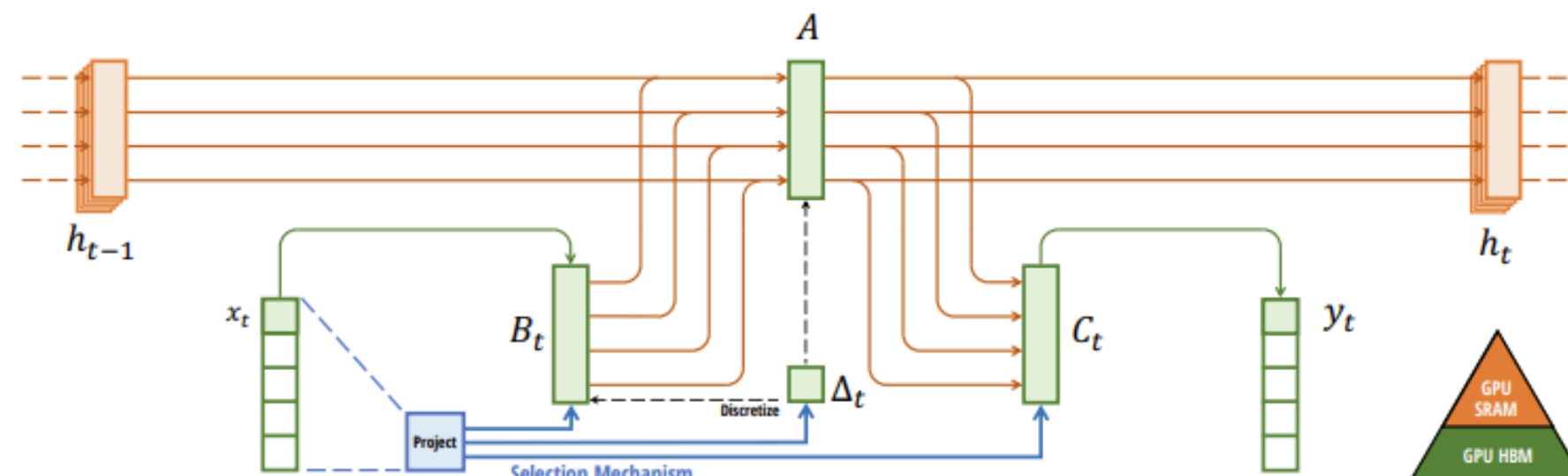
$$x'(t) = Ax(t) + Bu(t)$$

$$y(t) = Cx(t) + Du(t)$$

- $x'(t)$: 시간에 따른 상태 $x(t)$ 의 변화율
- A**: 상태 전이 행렬 (State Transition Matrix)
- B**: 입력 행렬 (Input Matrix)
- $u(t)$: 외부 입력 (Input)
- $y(t)$: 출력 (Output)
- C**: 출력 행렬 (Output Matrix)
- D**: 입력-출력 간 직접적인 관계를 나타내는 행렬

1) Mamba

Selective State Space Model
with Hardware-aware State Expansion



Algorithm 1 SSM (S4)

Input: $x : (B, L, D)$

Output: $y : (B, L, D)$

- 1: $A : (D, N) \leftarrow$ Parameter
 ▶ Represents structured $N \times N$ matrix
- 2: $B : (D, N) \leftarrow$ Parameter
- 3: $C : (D, N) \leftarrow$ Parameter
- 4: $\Delta : (D) \leftarrow \tau_{\Delta}(\text{Parameter})$
- 5: $\bar{A}, \bar{B} : (D, N) \leftarrow \text{discretize}(\Delta, A, B)$
- 6: $y \leftarrow \text{SSM}(\bar{A}, \bar{B}, C)(x)$
 ▶ Time-invariant: recurrence or convolution
- 7: **return** y

Algorithm 2 SSM + Selection (S6)

Input: $x : (B, L, D)$

Output: $y : (B, L, D)$

- 1: $A : (D, N) \leftarrow$ Parameter
 ▶ Represents structured $N \times N$ matrix
- 2: $B : (B, L, N) \leftarrow s_B(x)$
- 3: $C : (B, L, N) \leftarrow s_C(x)$
- 4: $\Delta : (B, L, D) \leftarrow \tau_{\Delta}(\text{Parameter} + s_{\Delta}(x))$
- 5: $\bar{A}, \bar{B} : (B, L, D, N) \leftarrow \text{discretize}(\Delta, A, B)$
- 6: $y \leftarrow \text{SSM}(\bar{A}, \bar{B}, C)(x)$
 ▶ **Time-varying:** recurrence (*scan*) only
- 7: **return** y

Method

1) 데이터셋

1. BIDMC PPG 및 호흡 클린 데이터셋 (BIDMC PPG and Respiration Dataset)

- ppg 및 호흡 동시에 기록된 생리적 파형 (125 Hz로 샘플링)
- HR 및 호흡 수(RR)같은 파생된 생리적 매개변수 (1 Hz로 샘플링)
- 특징:
 - 총 53명의 피험자
 - 각 피험자의 기록: 약 8분
 - 통제된 환경에서 수집된 기록, 노이즈가 적고 클린 신호

2. 운동 중 손목 PPG 데이터셋 (Wrist PPG During Exercise dataset)

- 러닝 및 사이클링과 같은 신체 활동 중에 기록된 18개의 손목 착용 ppg 신호
- 모션 동역학을 포착하기 위한 가속도계 및 자이로스코프 측정값
- 특징:
 - 총 8명의 피험자
 - 각 피험자 기록: 약 10분
 - 움직임으로 인한 아티팩트가 다수 포함

2) DPNet (Denoising Photoplethysmography Network) 아키텍처

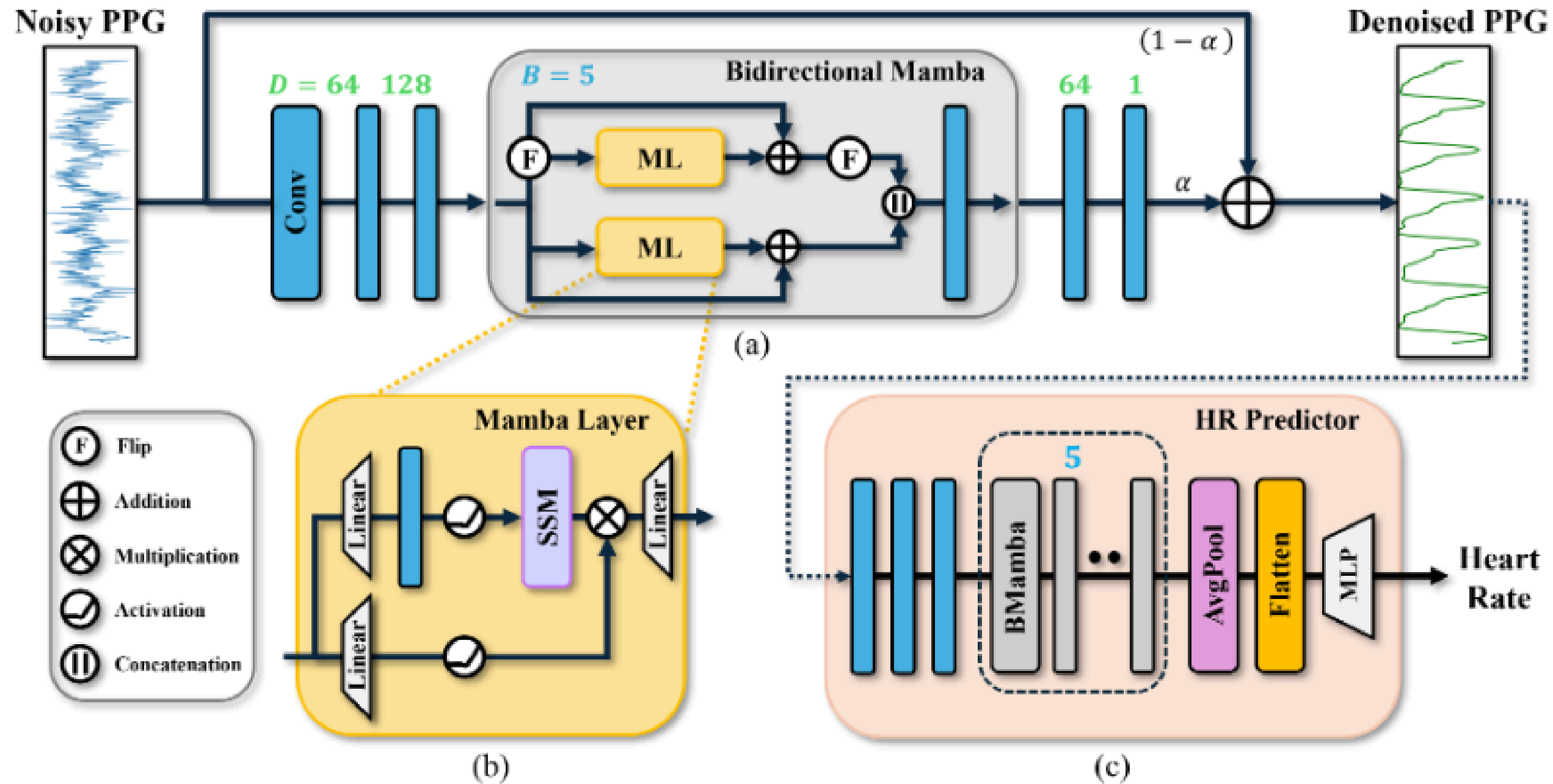


Fig. 1. Architectures of (a) DPNet, (b) Mamba layer, and (c) HRP.

3) DPNet (Denoising Photoplethysmography Network) 수식

$$h_n = \bar{A}h_{n-1} + \bar{B}x_n, \quad (1)$$

$$y_n = Ch_n, \quad (2)$$

- **Mean Squared Error (MSE):** Quantifies the average squared difference between \mathbf{g} and \mathbf{d} . It is defined as:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (g[i] - d[i])^2, \quad (3)$$

- **Cosine Similarity (CoS):** Reflects how similar the directions of \mathbf{g} and \mathbf{d} are in a multi-dimensional space:

$$\text{CoS} = \frac{\mathbf{g} \cdot \mathbf{d}}{\|\mathbf{g}\| \|\mathbf{d}\|}, \quad (4)$$

- **Signal-to-Noise Ratio improvement (SNR_{imp}):** Measures the relative improvement in SNR achieved by the denoised signal compared to the noisy input:

$$\text{SNR}_{imp} = \text{SNR}(\mathbf{d}, \mathbf{g}) - \text{SNR}(\mathbf{n}, \mathbf{g}). \quad (5)$$

- **HR-MAE:** Computes the Mean Absolute Error (MAE) of HR between \mathbf{g} and \mathbf{d} :

$$\text{HR-MAE} = |\text{HR}(\mathbf{g}) - \text{HR}(\mathbf{d})|, \quad (6)$$

4) DPNet (Denoising Photoplethysmography Network) 손실함수

$$\mathcal{L} = \begin{cases} \mathcal{L}_{\text{MSE}} + \lambda_1 \cdot \mathcal{L}_{\text{SI-SDR}}, & \text{if } E < E_w, \\ \mathcal{L}_{\text{MSE}} + \lambda_1 \cdot \mathcal{L}_{\text{SI-SDR}} + \lambda_2 \cdot \mathcal{L}_{\text{MAE}}, & \text{if } E \geq E_w, \end{cases} \quad (7)$$

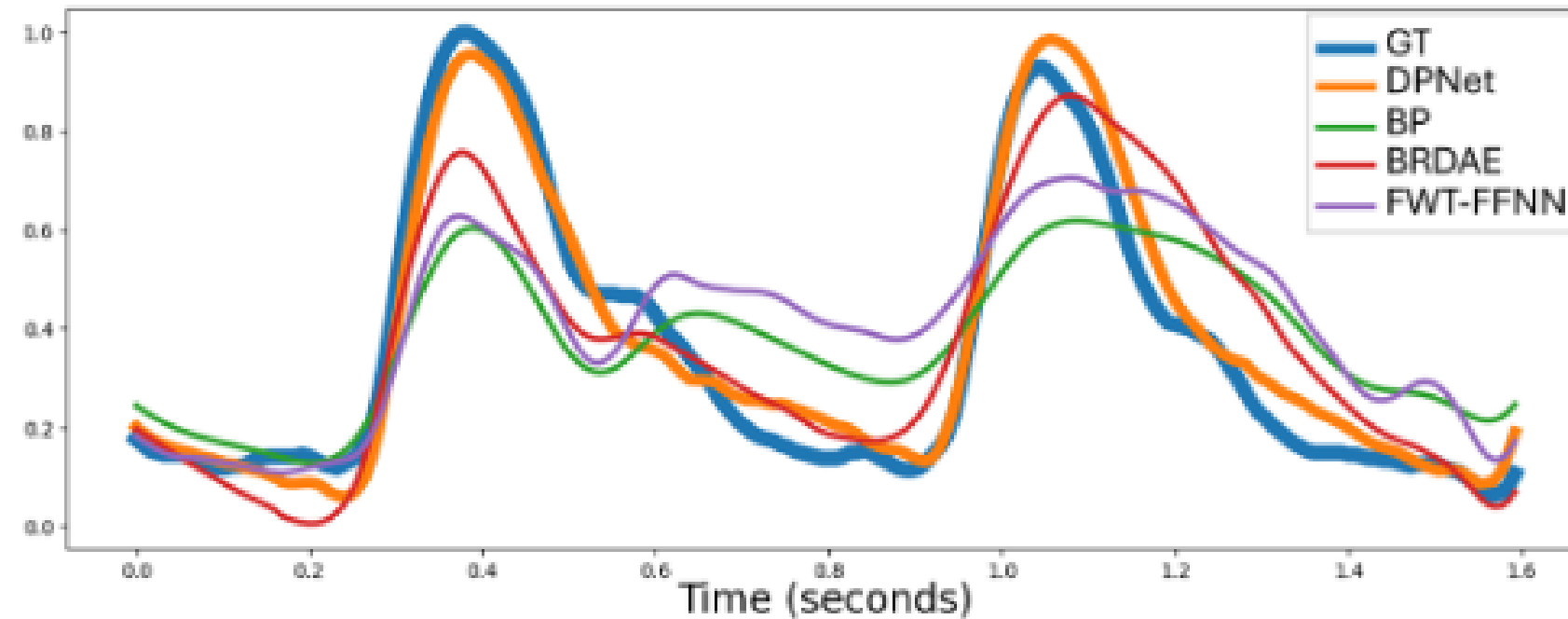
1. 워밍업 에포크(E_w) 동안 모델은 파형 충실도를 안정화하고 복구하기 위해 MSE(Mean Squared Error) 손실과 SI-SDR (Scale-Invariant Signal-to-Distortion Ratio) 손실로 최적화
2. 주요 훈련 단계에선 E_w 이후에는 GT와 HRP-추정 BPM 간의 추가 MAE (Mean Absolute Error) 손실이 도입되어 잡음 제거가 생리적 일관성을 유지하도록 유도
 - λ_1 과 λ_2 는 손실의 크기 균형을 맞추기 위한 가중치

Result

Table 1. Quantitative performance comparison of different methods on the PPG denoising task.

Method	MSE ($\times 10^{-3}$) \downarrow	CoS \uparrow	SNR _{imp} (dB) \uparrow	HR-MAE \downarrow
Noisy	324.029 \pm 723.622	0.726 \pm 0.272	-	109.410 \pm 324.619
BP filters [15]	29.531 \pm 25.088	0.823 \pm 0.144	-0.206 \pm 4.744	5.168 \pm 16.556
BRDAE [6]	19.229 \pm 18.578	0.881 \pm 0.126	2.054 \pm 3.574	2.896 \pm 21.060
FWT-FFNN [16]	31.562 \pm 36.412	0.802 \pm 0.237	1.517 \pm 2.726	11.510 \pm 36.174
DPNet (The proposed)	6.663\pm9.845	0.961\pm0.069	8.323\pm4.789	1.025\pm4.869

Bold represents the best performance.

**Fig. 2.** Denoised PPG signals using different methods.**Table 2.** Effect of different training loss used to train DPNet.

Loss combinations	MSE ($\times 10^{-3}$) \downarrow	HR-MAE \downarrow
\mathcal{L}_{MSE}	7.657	1.979
$\mathcal{L}_{\text{MSE}} + \mathcal{L}_{\text{SI-SDR}}$	7.255	1.856
$\mathcal{L}_{\text{MSE}} + \mathcal{L}_{\text{SI-SDR}} + \mathcal{L}_{\text{MAE}}$	6.663	1.025

Bold represents the best performance.

Table 3. Comparison of Temporal Modeling Architectures.

Block	MSE ($\times 10^{-3}$) \downarrow	HR-MAE \downarrow
Transformer	10.017	4.862
BLSTM	9.828	1.149
DPNet (BMamba)	6.663	1.025

Bold represents the best performance.



Thank You



2026.04.02

BrainLAB Journal Club
국방인공지능응용학과 이정인