Supplementary material

This is the supplementary material for the paper "TrustEMG-Net: Using Representation-Masking Transformer with U-Net for Surface Electromyography Enhancement." This study compared our proposed method, TrustEMG-Net, with several signal-processing- and neural-network (NN)-based sEMG denoising methods. Sections I and II of this document provide the technical details of these comparison sEMG denoising methods for better understanding and reproducibility. Section III presents an additional experiment for validating the effectiveness of the integration method and representation-masking (RM) approach.

I. Signal-processing-based sEMG denoising methods

This section presents the detailed implementation of the signal-processing-based sEMG denoising methods used in this study, including IIR filters, template subtraction (TS), empirical mode decomposition (EMD), complete ensemble EMD with adaptive noise (CEEMDAN), and variational mode decomposition (VMD).

A. IIR filters

The selection of IIR filters for sEMG contaminant removal is shown in Table I. Each IIR filter is applied only if the corresponding contaminant exists in sEMG.

TABLE I
IIR filters for sEMG contaminant removal.

Contaminant	Filter specification
BW	4^{th} order Butterworth high-pass filter, f_c 10 Hz
PLI	Notch filter, f_c 60 Hz and quality factor 5
MOA [1]	4^{th} order Butterworth high-pass filter, f_c 40 Hz
MOA [2]	4^{th} order Butterworth high-pass filter, f_c 20 Hz
ECG	4^{th} order Butterworth high-pass filter, f_c 40 Hz
WGN	4^{th} order Butterworth band-pass filter, f_c 20 and 500 Hz

f_c denotes the cutoff frequency.

B. TS+IIR

TS+IIR applies TS and IIR filters for ECG and non-ECG contaminants, respectively. TS consists of three steps: ECG detection, template extraction, and ECG subtraction. For ECG detection, zero-padding is first performed on signal segments for 0.5 seconds at the front and end of the segments. We then calculate two moving averages (1 s and 0.1 s) and identify ECG-containing segments when these moving averages intersect twice within a specified time (more than 0.14 seconds in this study) [3]. Since the sEMG segments are short (2 s), we choose a filtering approach to create ECG templates, using a 4th-order Butterworth high-pass filter with cutoff frequency 50 Hz [4]. After subtracting ECG artifacts, we apply a 4th-order Butterworth high-pass filter with a cutoff frequency of 40 Hz to obtain the best results [5], [6].

C. EMD-based and CEEMDAN-based methods

The EMD-based and CEEMDAN-based methods adopt EMD and CEEMDAN for signal decomposition, respectively. The maximum number of IMFs is set to 8 for both methods. As for CEEMDAN, the scale for added noise is set to 0.005, and the number of trials is 20.

The contaminant removal algorithms for both methods refer to previous research [7]–[11]. To discern whether certain contaminant types exist in each mode, we use the Fast Fourier Transform to find the frequency with maximum energy in the spectrum, denoted as f_{max}. We then apply the following algorithms for different types of contaminants if included [9].

- BW: modes with f_{max} less than 10 Hz are removed. Other modes remove BW by subtracting passing themselves through a 4th-order Butterworth low-pass filter with a cutoff frequency of 10 Hz.
- PLI: for modes with f_{max} between 50 and 70 Hz, we apply a narrow-band notch filter with center frequency f_{max} and quality factor 20.
- ECG: we calculate the Pearson correlation coefficients of each mode and ECG template to discern modes with more ECG contamination. The ECG template here is extracted from noisy sEMG by applying a 4th-order Butterworth high-pass filter with a cutoff frequency of 40 Hz. For the five modes with higher correlations with the ECG template, we pass them through a 4th-order Butterworth high-pass filter with a cutoff frequency of 30 Hz.
- MOA: modes with f_{max} less than 20 Hz are removed. Other modes are passed through a 4th-order Butterworth high-pass filter with a cutoff frequency of 20 Hz.
- WGN: the first mode is removed. For other modes, whether to remove them is determined based on their standard deviation. If the standard deviation of the mode exceeds the corresponding threshold value, we apply wavelet thresholding to the mode with sym8 as the mother wavelet [7], [8], [11].

D. VMD-based method

The VMD-based method conducts signal decomposition using VMD with a parameter set as follows. The number of IMF is 10, the penalty factor is 1000, and the tolerance is 0.001 [9]. The initialization of center frequency adopts uniform distribution.

We follow [9], [12] and apply the following algorithms for different contaminants. The usage of f_{max} is identical to the CEEMDAN-based methods. Notably, we directly apply the same IIR filter to eliminate ECG artifacts, as there is currently no VMD-based method for ECG contamination.

- BW, PLI, and MOA: identical to the CEEMDAN-based method.
- WGN: the soft iterative interval thresholding method is applied, and the parameter refers to [12]. The only difference is that the threshold value is divided by 4 to yield better denoising results in this study.

II. NN-BASED sEMG DENOISING METHODS

This section demonstrates the structures of NN models used for the ablation study of TrustEMG-Net, including CNN, FCN, and U-Net. Note that all the kernels of the convolutional layers of these models are 1-dimensional.

A. CNN

The CNN consists of four convolutional layers and two fully connected layers. Except for the last fully connected layer, which is a linear layer, all other layers use the ReLU activation function and batch normalization. The parameters for the convolutional layers are shown in Table II.

TABLE II
PARAMETERS OF THE CONVOLUTIONAL LAYERS IN CNN.

Layer	Input channel	Filter number	Filter size	Stride
Convolutional layer 1	1	16	8	2
Convolutional layer 2	16	32	8	2
Convolutional layer 3	32	64	8	2
Convolutional layer 4	64	128	8	2

The dimensions of the two fully connected layers are 400 and 200, respectively. Due to the large number of parameters in the first fully connected layer, dropout regularization with a dropout rate of 50%

Note that the input sEMG signal is segmented into 200 points per segment, as we discover that smaller segments can yield better results on CNN. However, segmentation increases the computation time of the CNN. All the outputs are concatenated after passing through the CNN to form the denoised signal.

B. FCN

The FCN's encoder and decoder each consist of five convolutional layers and five transposed convolutional layers, with detailed parameters shown below. Each layer uses the ReLU activation function and batch normalization except for the last transposed convolutional layer.

C. U-Net

The architecture of the U-Net is the same as the structure of TrustEMG-Net, as shown in Table III. The only difference is that the bottleneck of the U-Net does not employ a Transformer encoder.

TABLE III
PARAMETERS OF THE CONVOLUTIONAL LAYERS IN FCN.

Layer	Input channel	Filter number	Filter size	Stride
Convolutional layer 1	1	64	8	2
Convolutional layer 2	64	128	8	2
Convolutional layer 3	128	256	8	2
Convolutional layer 4	256	512	8	2
Convolutional layer 5	512	1024	8	2
ConvTranspose layer 1	1024	512	8	2
ConvTranspose layer 2	512	256	8	2
ConvTranspose layer 3	256	128	8	2
ConvTranspose layer 4	128	64	8	2
ConvTranspose layer 5	64	1	8	2

III. COMPARISON WITH OTHER INTEGRATION METHODS

This section compares TrustEMG-Net with two other integration approaches featuring similar model architectures. The first integration method, U-Net+5Trans, integrates five one-layer Transformer encoders into all four skip connections and the bottleneck of U-Net [13], as shown in Fig. 1 (a). The Transformer encoders at the skip connections follow TrustEMG-Net's settings, with adjustments made to the embedding dimensions to match the feature dimension of each skip connection. The second integration method, U-Net+LSTM, replaces TrustEMG-Net's Transformer encoder with a one-layer Long Short-Term Memory (LSTM) model [14], which is also known for its ability to capture long-term temporal dependencies [15]. The structure of U-Net+LSTM is shown in Fig. 1 (b). Additionally, both integration methods incorporate direct mapping (DM) and representation masking (RM) approaches, resulting in four distinct model structures. Models using DM are denoted by "(DM)" following their model types.

A. Comparison between integration methods

Table IV presents the overall performance comparison between TrustEMG-Net and other integration methods. It is evident that TrustEMG-Net consistently outperforms all other integration approaches across all five evaluation metrics, with statistically significant differences except when compared with U-Net+5Trans in terms of RMSE of MF. This underscores the effectiveness of TrustEMG-Net's integration approach, which embeds the Transformer encoder at the bottleneck of the U-Net architecture.

We observe that the denoising performance of U-Net+5Trans does not benefit from incorporating more Transformer encoders at skip connections, which can lead to overfitting in this task. Moreover, TrustEMG-Net performs better than U-Net+LSTM. We attribute this outperformance to the parallel processing characteristics of the Transformer encoder, enabling TrustEMG-Net to exhibit better global information extraction capability and computational efficiency compared with U-Net+LSTM.

B. Effectiveness of the RM approach

Table IV also highlights the effectiveness of the RM approach in sEMG contaminant removal. Across various integration frameworks (i.e., U-Net+5Trans and U-Net+LSTM), RM-based versions consistently achieve superior denoising results

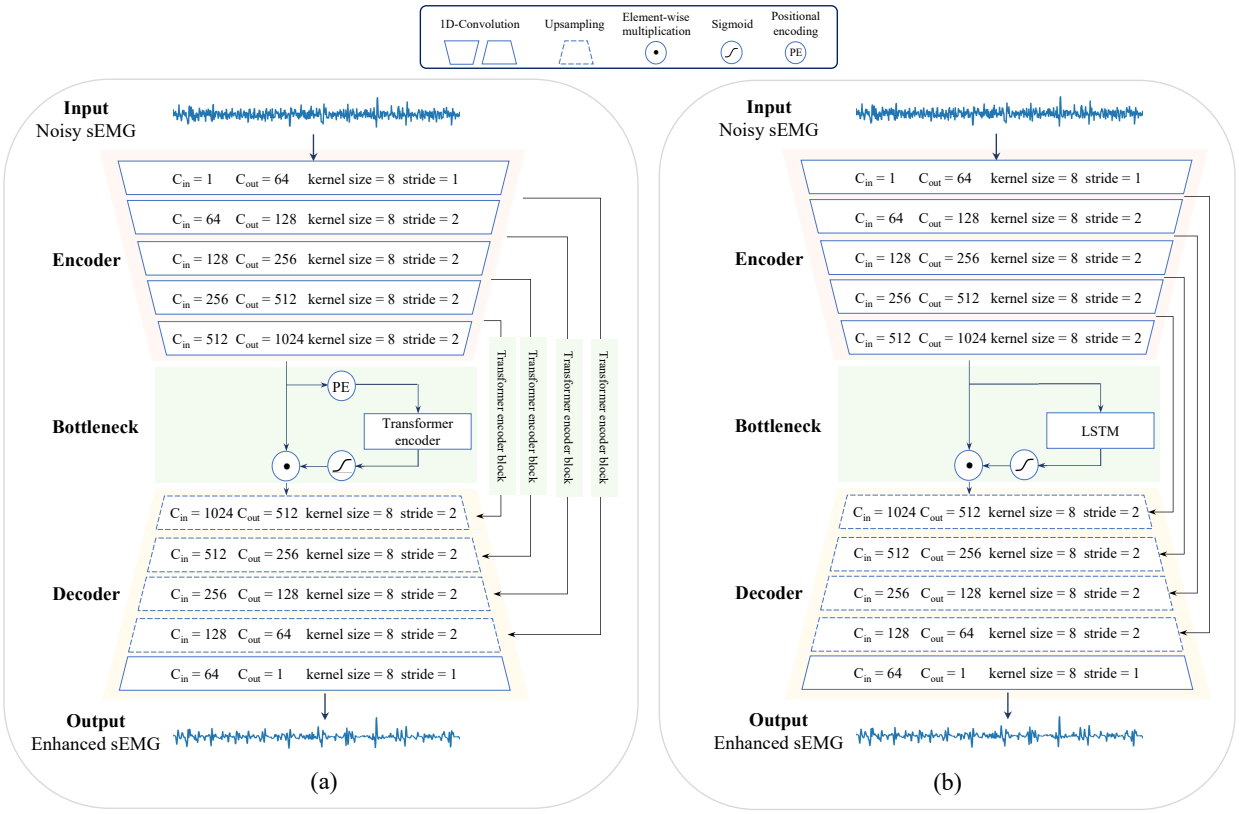

Fig. 1. The architecture of (a) U-Net+5Trans and (b) U-Net+LSTM.

TABLE IV
OVERALL PERFORMANCE OF NN MODELS WITH DIFFERENT INTEGRATION METHODS FOR SEMG CONTAMINANTS REMOVAL.

Metrics	U-Net+5Trans(DM)	U-Net+5Trans	U-Net+LSTM(DM)	U-Net+LSTM	TrustEMG-Net(DM)	TrustEMG-Net
SNR_{imp} (dB)	11.80 ± 0.51*	13.42 ± 0.47*	13.34 ± 0.41*	13.46 ± 0.46*	13.45 ± 0.45*	**13.64 ± 0.38**
RMSE ($\times 10^{-2}$)	2.50 ± 0.34*	2.22 ± 0.30*	2.24 ± 0.30*	2.22 ± 0.31*	2.21 ± 0.30*	**2.18 ± 0.28**
PRD (%)	56.29 ± 2.76*	49.32 ± 2.03*	49.57 ± 1.91*	49.11 ± 1.94*	49.05 ± 2.00*	**48.44 ± 1.70**
RMSE of ARV ($\times 10^{-3}$)	11.89 ± 2.12*	9.16 ± 1.70*	8.98 ± 1.63*	8.93 ± 1.62*	8.90 ± 1.58*	**8.72 ± 1.52**
RMSE of MF (Hz)	18.36 ± 1.86*	16.84 ± 0.81	17.78 ± 0.98*	17.63 ± 1.16*	16.94 ± 0.94*	**16.63 ± 0.91**

*Denotes a significant difference (p-value < 0.05) with the proposed method. **Bold** font indicates the best score for each metric.

compared with their DM-based counterparts, with all performance differences being statistically significant (p-values = 0.000 and 0.0026, respectively.)

REFERENCES

[1] Juliano Machado, Amauri Machado, and Alexandre Balbinot, "Deep learning for surface electromyography artifact contamination type detection," *Biomedical Signal Processing and Control*, vol. 68, pp. 102752, 2021.

[2] George B Moody, W Muldrow, and Roger G Mark, "A noise stress test for arrhythmia detectors," *Computers in cardiology*, vol. 11, no. 3, pp. 381–384, 1984.

[3] José Dilermando Costa Junior et al., "A template subtraction method for reducing electrocardiographic artifacts in emg signals of low intensity," *Biomedical Signal Processing and Control*, vol. 47, pp. 380–386, 2019.

[4] Ryan J Marker and Katrina S Maluf, "Effects of electrocardiography contamination and comparison of ecg removal methods on upper trapezius electromyography recordings," *Journal of Electromyography and Kinesiology*, vol. 24, no. 6, pp. 902–909, 2014.

[5] Kuan-Chen Wang, Kai-Chun Liu, Sheng-Yu Peng, and Yu Tsao, "Ecg artifact removal from single-channel surface emg using fully convolutional networks," in *Proc. ICASSP*, 2023.

[6] Janessa DM Drake and Jack P Callaghan, "Elimination of electrocardiogram contamination from electromyogram signals: An evaluation of currently used removal techniques," *Journal of electromyography and kinesiology*, vol. 16, no. 2, pp. 175–187, 2006.

[7] Yan Zhang, Jijian Lian, and Fang Liu, "An improved filtering method based on eemd and wavelet-threshold for modal parameter identification of hydraulic structure," *Mechanical Systems and Signal Processing*, vol. 68, pp. 316–329, 2016.

[8] Ziyang Sun et al., "Surface electromyography signal denoising via eemd and improved wavelet thresholds," *Mathematical Biosciences and Engineering*, vol. 17, no. 6, pp. 6945–6962, 2020.

[9] Shihan Ma et al., "Emg signal filtering based on variational mode decomposition and sub-band thresholding," *IEEE journal of biomedical and health informatics*, vol. 25, no. 1, pp. 47–58, 2020.

[10] Mohsen Naji, Mohammad Firoozabadi, and Sedighe Kahrizi, "The application of empirical mode decomposition in elimination of ecg contamination from emg signals," in *Proc. ICBME*, 2011.

[11] Guoyan Sun and Kairu Li, "Emg denoising based on ceemdan-pe-wt algorithm," in *Proc. ICIRA*, 2023.

[12] Hassan Ashraf et al., "Variational mode decomposition for surface and intramuscular emg signal denoising," *Biomedical Signal Processing and Control*, vol. 82, pp. 104560, 2023.
[13] Boxiang Yun, Yan Wang, Jieneng Chen, Huiyu Wang, Wei Shen, and Qingli Li, "Spectr: Spectral transformer for hyperspectral pathology image segmentation," *arXiv preprint arXiv:2103.03604*, 2021.
[14] Runnan He, Yang Liu, Kuanquan Wang, Na Zhao, Yongfeng Yuan, Qince Li, and Henggui Zhang, "Automatic detection of qrs complexes using dual channels based on u-net and bidirectional long short-term memory," *IEEE journal of biomedical and health informatics*, vol. 25, no. 4, pp. 1052–1061, 2020.
[15] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.