

Article

SWTT: A Hybrid Stationary Wavelet Transform and TCN-Transformer Network for EEG Artifact Removal

Yonggang Wang , Zheliang Chen, Bingyin Zhou * and Wei Guo *

School of Mathematical Sciences, Hebei Normal University, Shijiazhuang 050024, China;
wangyonggang@hebtu.edu.cn (Y.W.); chenzheliang@stu.hebtu.edu.cn (Z.C.)

* Correspondence: byzhou@hebtu.edu.cn (B.Z.); guowei@hebtu.edu.cn (W.G.)

Abstract

Electroencephalography (EEG) is widely used in clinical diagnosis, cognitive neuroscience, and brain–computer interfaces. During acquisition, it is inevitably affected by artifacts from electrooculography (EOG) and electromyography (EMG), severely degrading signal quality and impacting subsequent analysis. In practice, EEG is often contaminated by multiple artifacts simultaneously, and the degree of contamination can vary among different artifacts. Therefore, artifact removal is an important task for improving the quality of EEG and ensuring the reliability of subsequent analysis. However, most existing methods only consider single artifact removal scenarios. Even when considering mixed artifact scenarios, these methods often fail to effectively utilize the frequency characteristics of different artifacts, thus limiting their ability to accurately separate and suppress artifacts in the frequency domain. This study aims to establish an effective and robust EEG artifact removal method that can suppress both EOG and EMG artifacts while preserving EEG. To this end, we develop the stationary wavelet TCN-Transformer network (SWTT), a deep neural network based on SWT, TCN, and Transformer, designed for removing EEG artifacts. SWTT first uses SWT to decompose the noisy signal into time-aligned multi-scale time-frequency coefficients, which can separate different types of noise to some extent and make it easier to extract effective features. Then, a two-branch network is used for signal recovery in the wavelet domain. Specifically, the TCN branch uses dilated convolution to extract local temporal patterns, while the Transformer branch models cross-band global dependencies through a self-attention mechanism. Finally, the processed coefficients are used to reconstruct the time-domain signal using inverse SWT and then refined using a lightweight time-domain module to obtain the final result. Experiments on benchmark datasets demonstrate that SWTT achieves state-of-the-art performance in both single and mixed artifact removal tasks. Furthermore, because EOG and EMG artifacts exhibit distinct frequency-domain distributions, the network, by learning from the multi-scale time-frequency coefficients provided by SWT, can automatically capture these different time-frequency characteristics, and we observe that the model demonstrates distinct recovery behaviors for different artifact types.



Academic Editor: Jichai Jeong

Received: 13 July 2026

Revised: 31 July 2026

Accepted: 3 August 2026

Published: 4 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: electroencephalography; artifact removal; stationary wavelet transform; temporal convolutional network; Transformer; multi-scale time-frequency analysis

1. Introduction

Electroencephalography (EEG) is a non-invasive neuroimaging technique that records the collective electrical activity of groups of neurons using scalp electrodes. Due to its

high temporal resolution, low cost, and ease of use, EEG has become an important tool for studying brain function and neural states [1,2]. With the rapid development of neuroscience and artificial intelligence, EEG has been widely applied in clinical diagnosis and health monitoring, cognitive neuroscience research, and brain–computer interfaces (BCIs) [3]. In clinical medicine, EEG is used for epileptic seizure detection, Alzheimer’s disease diagnosis, and sleep disorder analysis. In the field of intelligent interaction, EEG is often used in areas such as fatigue driving detection, emotion recognition, and BCIs. All these applications heavily rely on high-quality EEG. Therefore, accurate EEG acquisition and analysis are crucial.

However, EEGs are typically in the microvolt range, making them highly susceptible to non-brain-related interference. Recorded signals contain not only the neural activity of interest but also artifacts and noise originating from ocular activity, muscle activity, poor electrode contact, and body movement. Electrooculography (EOG) artifacts, primarily caused by blinking and eye movements, manifest as high-amplitude low-frequency interference, whereas electromyography (EMG) artifacts arise from facial, neck, or scalp muscle activity and exhibit a much broader spectral range that overlaps with multiple EEG bands [4]. Because the energy of these artifacts often far exceeds that of the EEG itself, they can severely obscure or even distort genuine neural activity, compromising subsequent feature extraction, pattern recognition, and clinical diagnosis. Artifact removal has therefore become a critical step in EEG processing and a fundamental prerequisite for ensuring the reliability of EEG-based applications [5].

Researchers have proposed a variety of methods for removing EEG artifacts, which can be broadly classified into two categories: traditional signal processing methods and deep learning methods. Traditional methods include filtering techniques, wavelet transform (WT) [6], independent component analysis (ICA) [7,8], wavelet-enhanced ICA [9], and empirical mode decomposition (EMD) [10,11]. These methods typically rely on signal decomposition or statistical assumptions to separate artifacts from EEG. Shahbakhti et al. proposed using SWT to eliminate EOG artifacts [12]. They automatically determined the number of decomposition layers by calculating the difference between the absolute skewness values of adjacent scale coefficients, and reconstructed the signal after setting the approximation coefficients containing EOG artifacts to zero. Although this work utilizes the translation invariance advantage of SWT, it essentially still adopts a traditional thresholding strategy and fails to fully exploit the rich information of the multi-scale time-frequency features provided by SWT.

In recent years, deep learning has made significant progress in EEG artifact removal. Convolutional neural networks (CNNs) effectively extract temporal features using local convolutional kernels. Representative works include NovelCNN in EEGdenoiseNet [13] and 1D-ResCNN proposed by Sun et al. [14]. Temporal convolutional networks (TCNs) [15] utilize dilated convolutions and residual connections to expand the receptive field, thereby achieving parallelization and long-range dependency modeling. Representative models include DTP-Net [16] and MHNoiseRegression [17]. The Transformer [18] directly models global dependencies via self-attention mechanisms, such as DeT [19]. To further integrate local details with global context, hybrid architectures such as DenoiseEEG [20] combine U-Net with attention mechanisms, while DHCT-GAN [21] employs a two-branch CNN-Transformer framework. Both aim to achieve fine-grained recovery through complementary local–global modeling. WNOTNet [22] combines the Wavelet Neural Operator (WNO) with the Transformer architecture, using wavelet-based multi-scale time-frequency analysis to capture local details of the signal, while leveraging the Transformer’s self-attention mechanism to model global long-range dependencies. Recently, TF-Denoiser [23] explored a time–frequency representation based on Fourier transform and applied a multi-attention mechanism for joint optimization.

Despite the variety of existing methods, EEG artifact removal remains a challenging problem. First, artifacts are highly non-stationary and vary significantly across different subjects and recording conditions, limiting the generalization ability of most models. For instance, blinking causes sharp transient peaks, while horizontal eye movements produce slow drift potentials, which makes it difficult for fixed-parameter models to adapt. Second, EEG and artifacts exhibit strong overlap in both the time and frequency domains. As mentioned above, EOG artifacts primarily occupy the low-frequency bands that coincide with delta and theta rhythms, while EMG artifacts display broadband characteristics, covering a large portion of the EEG spectrum, rendering simple frequency-based separation methods ineffective. Traditional methods are often constrained by strong assumptions such as linearity and stationarity, which are frequently not met in actual EEG recordings. While deep learning methods alleviate some limitations by learning data-driven representations, in the absence of explicit constraints, these methods cannot distinguish between different artifact types, limiting their ability to effectively separate and suppress artifacts in the frequency domain.

From a modeling perspective, EEG artifact removal requires learning a feature representation that maintains temporal consistency while ensuring the separation of frequency information from noisy observations. This design enables subsequent neural networks to handle aligned multi-scale features, thereby achieving unified modeling of in-band dynamics and cross-frequency relationships within a single framework. To this end, we employ the Stationary Wavelet Transform (SWT) [24] to construct a structured time–frequency representation space. Unlike the standard Discrete Wavelet Transform (DWT), SWT omits the downsampling operation at each decomposition layer, thus maintaining strict temporal alignment across scales and providing translation invariance. These properties make SWT naturally suitable for deep learning-based cross-scale modeling because it provides a temporally consistent and frequency-diverse input space for subsequent modules.

On top of this representation, we further introduce a dual-branch architecture to fully exploit its multi-scale structure. Specifically, a TCN branch is employed to capture local temporal patterns in the wavelet domain, while a Transformer branch is designed to model cross-band global dependencies. This design allows the model to jointly learn fine-grained temporal dynamics and long-range interactions across frequencies. Overall, this leads to the proposed Stationary Wavelet TCN-Transformer Network (SWTT), which integrates SWT-based representation with dual-branch feature modeling to effectively suppress artifacts while preserving task-relevant EEG information.

Furthermore, most existing EEG artifact removal studies construct experimental data based on a single noise source, considering only independent electrooculography (EOG) or electromyography (EMG) contamination [25]. In practical acquisition settings, however, multiple artifact types often simultaneously contaminate the same EEG segment, with their interference intensities varying dynamically over time and across individual states. For example, a subject may blink while simultaneously engaging facial muscles, causing low-frequency EOG artifacts and high-frequency EMG artifacts to jointly contaminate the EEG. Evaluating models solely under single-artifact conditions therefore fails to fully reflect their practical applicability. While a recent CNN-based method has solved the problem of simultaneous EOG and EMG artifacts [26], it does not take into account contamination of varying intensities. To better approximate practical recording scenarios, we design the EOG–EMG mixed-artifact removal task with varying SNR combinations to more comprehensively evaluate the performance and generalization ability of the artifact removal model under mixed-artifact conditions.

Given these challenges, the goal of this work is to develop a deep learning framework that can suppress both EOG and EMG artifacts while achieving accurate recovery of clean EEGs. The main contributions of this work are summarized as follows:

1. We propose to combine stationary wavelet transform with deep learning in the EEG artifact removal task. Since different types of artifacts differ significantly in the frequency domain, SWT can achieve a certain degree of separation in the frequency domain and provide time-aligned multi-scale feature descriptions. Based on this, we consider combining SWT with deep neural networks to fully utilize SWT's structured multi-resolution representation and deep learning's powerful non-linear fitting capabilities to jointly complete the EEG artifact removal task.
2. We propose SWTT, an SWT-based EEG artifact removal model. SWTT first represents EEG as multi-scale time–frequency coefficients using SWT and then uses the TCN branch and the Transformer branch to model the local time patterns in the wavelet domain and the global dependencies across frequency bands, respectively, to achieve effective and interpretable artifact removal.
3. We conducted comprehensive experimental validation and results analysis of the model's behavior and generalization ability. We benchmark SWTT against a range of representative methods, such as CNN, LSTM, MHNoiseRegression, and WNOTNet. Experimental results demonstrate that SWTT achieves state-of-the-art performance across various tasks. Furthermore, we perform ablation study, parameter analyses, and interpretability analyses to evaluate SWTT's capabilities in various aspects.

The remainder of this paper is organized as follows. Section 2 describes the proposed SWTT framework in detail, including the SWT-based time–frequency representation, the dual-branch TCN-Transformer architecture, and the time-domain refinement module. Section 3 presents the experimental setup and reports the results on single-artifact removal, mixed-artifact removal, ablation studies, and interpretability analyses. Section 4 discusses the key findings, limitations of the current work, and potential directions for future research. Finally, Section 5 concludes this paper.

2. Materials and Methods

This section introduces the proposed EEG artifact removal method. First, we introduce the dataset and formulate the problem we aim to solve. We then introduce the SWTT model in detail.

2.1. Dataset and Problem Formulation

First, we introduce the EEGdenoiseNet benchmark dataset used in this study. Then, we elaborate on the EEG denoising problem, define the model's input and output, and propose three tasks for model evaluation, including EOG artifact removal, EMG artifact removal, and EOG–EMG mixed-artifact removal tasks, designed to simulate various contamination scenarios that may occur in real EEG.

2.1.1. Dataset

EEGdenoiseNet [13] is one of the most widely used public benchmark datasets for EEG artifact removal. This dataset is constructed by Zhang et al., who collected EEG, EOG, and EMG signals from multiple public databases and performed rigorous preprocessing. The raw signals are first processed with a 50 Hz notch filter and a bandpass filter. The frequency range for EEG is 1–80 Hz, for EOG it is 0.3–10 Hz, and for EMG it is 1–120 Hz. Then, data from different sources are resampled and segmented into 2-second segments. The authors further processed the EEG data using the ICLabel tool (the ICLabel tool is publicly available at <https://github.com/scen/ICLabel>, accessed on 6 July 2026) based on ICA to obtain the cleanest possible representation of brain activity as the ground truth. The final dataset contains 4514 clean EEG segments, 3400 EOG segments, and 5598 EMG segments. We treat these clean signal segments as genuine brain signals and linearly super-

impose them with EOG or EMG artifacts to construct semi-synthetic EEG data with varying degrees of contamination (see Section 2.1.2 for the quantitative definition of contamination levels). To ensure a one-to-one correspondence between clean EEGs and artifacts, we select the first 3400 clean EEG segments and pair them with 3400 EOG segments and 3400 EMG segments, respectively. Each clean EEG segment is combined with one artifact segment to generate contaminated samples. Since the genuine EEGs are known, the artifact removal performance of different models can be objectively and accurately evaluated.

2.1.2. Problem Statement

Let the real signal be denoted as

$$\mathbf{x} = \{x_t\}_{t=1}^T,$$

where T denotes the signal length. Let there be K distinct artifact types, with the j -th artifact signal expressed as

$$\mathbf{n}_j = \{n_t^{(j)}\}_{t=1}^T, \quad j = 1, 2, \dots, K.$$

In practice, recorded EEG signals contain both brain activity and multiple artifact components. We therefore adopt an additive contamination model to construct the noisy signal:

$$\mathbf{y} = \mathbf{x} + \frac{1}{K} \sum_{j=1}^K \lambda_j \mathbf{n}_j, \quad (1)$$

where $\mathbf{y}$ is the noisy signal, and λ_j is the contamination coefficient for the j -th artifact type, controlling its contamination strength.

Stacking the K artifact vectors column-wise into a noise matrix $\mathbf{N} \in \mathbb{R}^{T \times K}$ and arranging the contamination coefficients into a column vector $\boldsymbol{\lambda} \in \mathbb{R}^K$,

$$\mathbf{N} = \begin{pmatrix} \mathbf{n}_1 & \mathbf{n}_2 & \cdots & \mathbf{n}_K \end{pmatrix}, \quad \boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_K \end{pmatrix}. \quad (2)$$

The additive contamination model (1) can then be written as follows:

$$\mathbf{y} = \mathbf{x} + \frac{1}{K} \mathbf{N} \boldsymbol{\lambda}. \quad (3)$$

To evaluate model performance under different contamination strengths, we use the signal-to-noise ratio (SNR) to control the amplitude of each artifact type. Given a target SNR_j (in dB) for the j -th artifact, the contamination coefficient λ_j is defined as

$$\lambda_j = \frac{\text{RMS}(\mathbf{x})}{10^{\text{SNR}_j/20} \text{RMS}(\mathbf{n}_j)}, \quad (4)$$

where

$$\text{RMS}(\mathbf{s}) = \left(\frac{1}{T} \sum_{t=1}^T s_t^2 \right)^{1/2}$$

denotes the root-mean-square value of signal $\mathbf{s}$. According to (4), a lower target SNR_j corresponds to more severe artifact contamination, and vice versa (the specific SNR values used are specified in Section 2.1.3).

Given a noisy signal $\mathbf{y}$, our objective is to learn a non-linear mapping function

$$f_{\theta} : \mathbf{y} \mapsto \hat{\mathbf{x}}, \quad (5)$$

where θ denotes the model parameters, and $\hat{\mathbf{x}}$ is the recovered signal. The recovered signal should approximate the real brain activity as closely as possible:

$$\hat{\mathbf{x}} = f_{\theta}(\mathbf{y}) \approx \mathbf{x}. \quad (6)$$

EEG artifact removal can be formulated as the following optimization problem:

$$\theta^* = \arg \min_{\theta} \frac{1}{2} \|f_{\theta}(\mathbf{y}) - \mathbf{x}\|_2^2. \quad (7)$$

By optimizing (7), the model learns to separate and remove artifact components from noisy EEG while preserving the original brain activity information.

2.1.3. Task Design

We focus on the two most common EEG artifact types: EOG and EMG, hence $K = 2$. The corresponding contamination coefficient vector is

$$\boldsymbol{\lambda} = \begin{pmatrix} \lambda_{\text{eog}} \\ \lambda_{\text{emg}} \end{pmatrix}. \quad (8)$$

To systematically evaluate the model's recovery capabilities under different artifact scenarios, we adopt a training, validation, and test set partitioning strategy. EEG, EOG, and EMG data are randomly shuffled and then divided into training, validation, and test sets in a fixed ratio. The data partitioning method is the same for all tasks, differing only in the sampling strategy for the contamination coefficient vector $\boldsymbol{\lambda}$. We design the following two categories of tasks.

Single-artifact tasks evaluate a model's ability to remove one artifact type at a time, comprising two sub-tasks: EOG artifact removal and EMG artifact removal. For EOG artifact removal, only EOG artifacts are present ($\lambda_{\text{eog}} > 0$, $\lambda_{\text{emg}} = 0$). For EMG artifact removal, only EMG artifacts are present ($\lambda_{\text{eog}} = 0$, $\lambda_{\text{emg}} > 0$). To cover a range of contamination strengths, for each sample we use

$$\text{SNR} \in \{-7, -6, \dots, 1, 2\} \text{ dB}, \quad (9)$$

resulting in 10 contaminated samples for each real signal. The training, validation, and test sets all share the same SNR configuration.

In practical EEG acquisition, EOG and EMG artifacts often coexist, with their contamination strengths varying dynamically over time and across individuals. To simulate this complex scenario, we construct a mixed-artifact removal task in which both EOG and EMG contamination are simultaneously introduced into each sample ($\lambda_{\text{eog}} > 0$, $\lambda_{\text{emg}} > 0$). Unlike single-artifact tasks that use a fixed set of SNR levels, the contamination strengths in the mixed-artifact task are independently sampled from the discrete set

$$\begin{aligned} \text{SNR}_{\text{eog}} &\in \{-7, -6, \dots, 1, 2\} \text{ dB}, \\ \text{SNR}_{\text{emg}} &\in \{-7, -6, \dots, 1, 2\} \text{ dB}. \end{aligned} \quad (10)$$

For each sample, 10 contaminated samples with different intensities are generated by randomly generating independent (SNR_{eog} , SNR_{emg}) pairs. Since the SNRs of EOG and EMG are sampled independently, the model must be able to handle cases with different

primary artifact types and different contamination ratios to more comprehensively evaluate its robustness in complex multi-source artifact scenarios. The training, validation, and test sets all share the same SNR configuration.

Figure 1 shows typical examples of three tasks under different SNR combinations. As the SNR decreases, the deviation between the contaminated signal and the real signal gradually increases in the multi-scale energy distribution, reflecting the growing influence of artifacts.

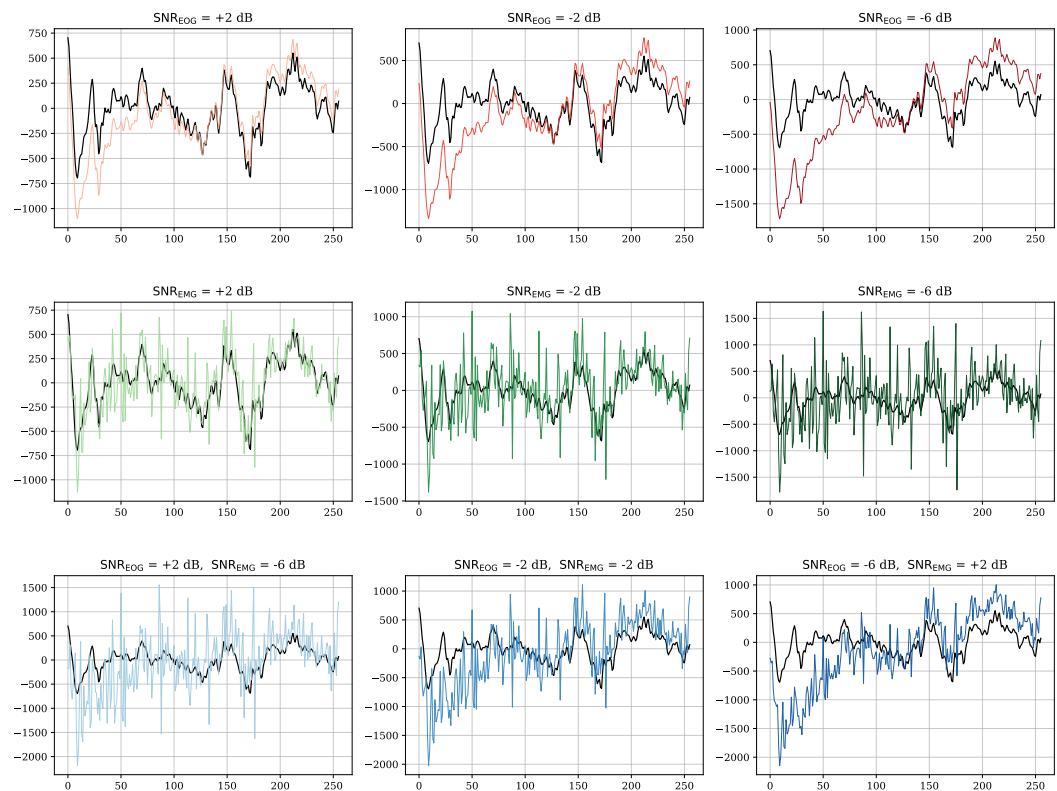


Figure 1. Overview of the artifact removal tasks. The black curves represent clean signals, while the colored curves represent signals contaminated by artifacts of varying intensities. The top row shows EOG contamination at different SNR levels, the middle row shows EMG contamination, and the bottom row presents mixed-artifact scenarios.

2.2. The Proposed SWTT

The architecture of SWTT is presented in Figure 2. First, the input noisy signal is decomposed via the SWT to obtain multi-scale wavelet coefficients of equal temporal length. Because the SWT preserves the same temporal resolution as the original signal at all scales, it enables the construction of a unified, strictly time-aligned time–frequency representation. Meanwhile, because different types of artifacts have different frequency distributions, this module can partially separate various types of artifacts in the mixed-artifact removal task. The stacked wavelet coefficients are then fed into two parallel branches, each starting from an independent convolutional embedding that maps the coefficient channels to a hidden dimension. The TCN branch extracts local temporal patterns in the wavelet domain through stacked dilated convolutions. The Transformer branch applies self-attention across the hidden feature channels, treating the full time series at each channel as a token, to model cross-band global dependencies. The two branches output predicted wavelet coefficients, which are fused by element-wise addition. The fused coefficients are reconstructed back into the time domain via inverse SWT to obtain a preliminary recovered result. Finally, a lightweight time-domain convolutional refinement module performs residual correction on the output.

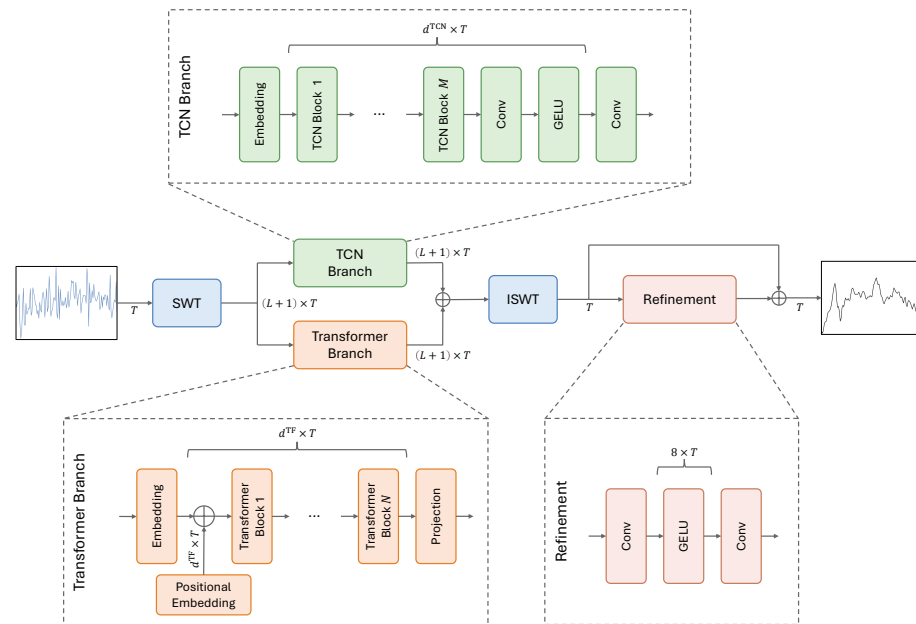


Figure 2. The architecture of the proposed SWTT, which consists of SWT decomposition, parallel TCN and Transformer branches, inverse SWT reconstruction, and a time-domain refinement module.

2.2.1. SWT and Inverse SWT Representation

The SWT [24] is a redundant wavelet representation that omits the downsampling operation of the conventional DWT, thereby producing coefficients of the same length as the original signal at every decomposition level.

Given an input signal $\mathbf{x} = \{x_t\}_{t=1}^T$, a low-pass filter $h[n]$ and a high-pass filter $g[n]$, the l -th SWT level constructs scale-dependent filters via the *à trous* convolution scheme:

$$\begin{aligned} h^{(l)}[n] &= \begin{cases} h[k], & n = k2^{l-1}, \\ 0, & \text{otherwise,} \end{cases} \\ g^{(l)}[n] &= \begin{cases} g[k], & n = k2^{l-1}, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \quad (11)$$

The dilated filters are then convolved with the approximation coefficients from the preceding level:

$$\begin{aligned} \mathbf{c}_A^{(l)} &= \mathbf{c}_A^{(l-1)} * h^{(l)}, \\ \mathbf{c}_D^{(l)} &= \mathbf{c}_A^{(l-1)} * g^{(l)}, \end{aligned} \quad (12)$$

where $*$ denotes the convolution operation and

$$\mathbf{c}_A^{(0)} = \mathbf{x}. \quad (13)$$

After L levels of decomposition, we obtain one set of approximation coefficients and L sets of detail coefficients:

$$\mathbf{C} = [\mathbf{c}_A^{(L)}, \mathbf{c}_D^{(L)}, \dots, \mathbf{c}_D^{(1)}]. \quad (14)$$

Since all coefficients retain a length of T , each time position corresponds to a feature vector composed of multi-scale frequency-band responses, providing a unified representation for subsequent neural network modeling.

Given the fused wavelet coefficients $\hat{\mathbf{C}}$, the inverse SWT reconstructs the time-domain signal through sequential synthesis:

$$\tilde{\mathbf{x}} = \text{ISWT}(\hat{\mathbf{C}}). \quad (15)$$

The inverse SWT likewise employs dilated convolutions, using reconstruction filter banks to upsample and sum the coefficients across scales, ultimately yielding a reconstructed signal of exactly the same length as the original.

2.2.2. TCN Branch

The TCN branch models local patterns in the signal. The multi-scale wavelet coefficients are denoted as $\mathbf{C} \in \mathbb{R}^{(L+1) \times T}$. They are then mapped to a higher-dimensional feature space via 1D convolution:

$$\mathbf{H}_0 = \Phi_{\text{emb}}^{\text{TCN}}(\mathbf{C}), \quad (16)$$

where the embedding module $\Phi_{\text{emb}}^{\text{TCN}}$ consists of a 1D convolution with kernel size 7 that maps the input to d^{TCN} hidden channels, followed by a GELU activation function. The features then pass through M TCN blocks followed by a convolutional block for feature extraction and fusion. A TCN block is defined as

$$\text{TCNBlock}(\mathbf{H}; r) = \mathbf{H} + \Phi_{\text{dil}}(\mathbf{H}; r), \quad (17)$$

where $\mathbf{H}$ denotes the input, r denotes the dilation rate, and Φ_{dil} consists of a dilated convolution with kernel size 7 and dilation r , group normalization, and a GELU activation. To progressively expand the receptive field, the TCN blocks adopt an exponentially increasing dilation schedule. Specifically, we apply $r_m = 2^{m-1}$. Thus, the stacked TCN blocks can be expressed as

$$\mathbf{H}_m = \text{TCNBlock}(\mathbf{H}_{m-1}; r_m), \quad m = 1, \dots, M. \quad (18)$$

Figure 3 illustrates the receptive field expansion mechanism, where the convolution kernel size is set to 3, and the dilation rate increases with network depth. In this structure, output nodes progressively aggregate information from increasingly distant historical time steps through multiple layers of dilated convolutions, allowing the receptive field to grow rapidly with depth. After all TCN blocks, a convolutional block further fuses the multi-scale features, producing the TCN branch output

$$\hat{\mathbf{C}}_{\text{TCN}} = \Phi_{\text{out}}^{\text{TCN}}(\mathbf{H}_M), \quad (19)$$

where $\Phi_{\text{out}}^{\text{TCN}}$ consists of a convolution with kernel size 3 and GELU activation, followed by a convolution with kernel size 1.

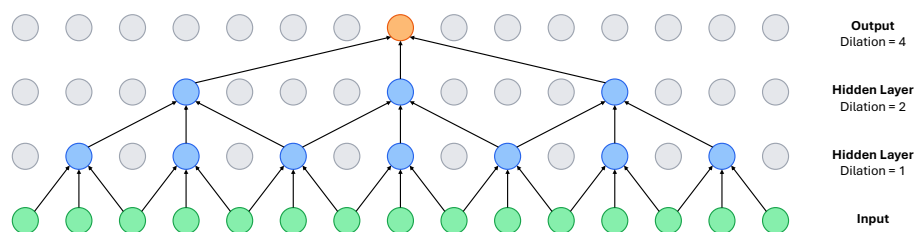


Figure 3. Receptive field expansion mechanism of the TCN with kernel size 3 and 3 layers. The highlighted path shows the receptive field of a representative output node. Green, blue, and orange nodes represent input, hidden, and output features, respectively.

2.2.3. Transformer Branch

To model global dependencies, we introduce a Transformer branch. The stacked wavelet coefficients $\mathbf{C} \in \mathbb{R}^{(L+1) \times T}$ are first mapped into a hidden feature space of dimension d^{TF} through an embedding module:

$$\mathbf{H}_0^{\text{TF}} = \Phi_{\text{emb}}^{\text{TF}}(\mathbf{C}), \quad (20)$$

where $\Phi_{\text{emb}}^{\text{TF}}$ has the same structure as $\Phi_{\text{emb}}^{\text{TCN}}$, consisting of a 1D convolution with kernel size 7 followed by a GELU activation function.

A learnable positional embedding $\mathbf{P} \in \mathbb{R}^{d^{\text{TF}} \times T}$ is added to the embedded features:

$$\mathbf{H}_0^{\text{TF}} = \mathbf{H}_0^{\text{TF}} + \mathbf{P}. \quad (21)$$

The features are then processed by N stacked Transformer encoder layers:

$$\mathbf{H}_n^{\text{TF}} = \text{TFBlock}(\mathbf{H}_{n-1}^{\text{TF}}), \quad n = 1, \dots, N. \quad (22)$$

Each Transformer encoder layer consists of multi-head self-attention, a feed-forward network, residual connections, and layer normalization. The self-attention operation is defined as

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{d^{\text{TF}}}}\right)\mathbf{V}, \quad (23)$$

where $\mathbf{Q}, \mathbf{K}, \mathbf{V} \in \mathbb{R}^{d^{\text{TF}} \times T}$. After the final Transformer encoder layer, the features are projected back to the wavelet coefficient space:

$$\hat{\mathbf{C}}_{\text{TF}} = \Phi_{\text{out}}^{\text{TF}}(\mathbf{H}_N^{\text{TF}}), \quad (24)$$

where $\Phi_{\text{out}}^{\text{TF}}$ consists of layer normalization followed by a linear projection.

2.2.4. Coefficient Fusion

The TCN and Transformer branches predict wavelet coefficients from two complementary perspectives: local pattern extraction and global dependency modeling.

We adopt a simple yet effective element-wise addition for fusion:

$$\hat{\mathbf{C}} = \hat{\mathbf{C}}_{\text{TCN}} + \hat{\mathbf{C}}_{\text{TF}}. \quad (25)$$

The fused coefficients contain complementary information from both branches. They are then reconstructed via the inverse SWT operation described in Section 2.2.1 to obtain the time-domain signal $\tilde{\mathbf{x}}$.

2.2.5. Time-Domain Refinement

Although the SWT-inverse SWT pair provides complete time–frequency domain transformation, reconstructing from wavelet coefficients predicted by a finite-capacity network may still incur local detail errors. We therefore introduce a lightweight convolutional refinement module in the time domain to perform residual correction on the reconstruction.

Let $\tilde{\mathbf{x}}$ denote the recovered signal. The final output is

$$\hat{\mathbf{x}} = \tilde{\mathbf{x}} + \Phi_{\text{ref}}(\tilde{\mathbf{x}}). \quad (26)$$

where Φ_{ref} consists of a convolution with kernel size 7 and GELU activation, followed by another convolution with kernel size 7. This residual structure focuses on learning

local details that are not fully recovered during SWT reconstruction, further improving the fidelity and recovery performance of the signal.

3. Results

This section introduces the experimental evaluation of SWTT. We first define the evaluation metrics and describe the experimental settings, and then we conduct quantitative and qualitative performance evaluation, ablation study, parameter analyses, and interpretability analyses.

3.1. Evaluation Metrics

To comprehensively assess EEG artifact removal performance, we employ four evaluation perspectives: time-domain error, frequency-domain error, SNR improvement, and waveform similarity.

t-RRMSE measures the relative error between the recovered signal and the real signal in the time domain, defined as

$$\text{t-RRMSE} = \frac{\sqrt{\sum_t (\hat{x}_t - x_t)^2}}{\sqrt{\sum_t x_t^2}}, \quad (27)$$

where $\mathbf{x}$ denotes the real signal, and $\hat{\mathbf{x}}$ denotes the recovered signal. A lower t-RRMSE indicates that the recovered result is closer to the real signal.

Band-wise s-RRMSE evaluates the spectral error between the recovered and real signals in different frequency bands. To provide a more comprehensive evaluation of recovery performance under the EOG–EMG mixed-artifact removal task, we introduce the band-wise s-RRMSE metric. Following the classical EEG band subdivision, we compute errors for the delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–50 Hz) bands. Let $P(f)$ and $\hat{P}(f)$ denote the amplitude spectra of the real and recovered signals, respectively. For a frequency band $B = [f_l, f_h)$, the band-wise relative root-mean-square error is defined as

$$\text{s-RRMSE}_B = \frac{\sqrt{\sum_{f \in B} (\hat{P}(f) - P(f))^2}}{\sqrt{\sum_{f \in B} P(f)^2}}, \quad (28)$$

A lower s-RRMSE_B indicates better preservation of spectral information in the corresponding band.

ΔSNR quantifies the improvement in SNR after recovery:

$$\Delta\text{SNR} = \text{SNR}(\hat{\mathbf{x}}) - \text{SNR}(\mathbf{y}), \quad (29)$$

where $\mathbf{y}$ denotes the noisy signal, and

$$\text{SNR}(\mathbf{s}) = 10 \log_{10} \left(\frac{\sum_t x_t^2}{\sum_t (s_t - x_t)^2} \right). \quad (30)$$

A larger ΔSNR indicates stronger noise suppression capability.

Peak Signal-to-Noise Ratio (PSNR) measures the overall recovery quality, defined as

$$\text{PSNR} = 10 \log_{10} \left(\frac{x_{\max}^2}{\frac{1}{T} \sum_{t=1}^T (\hat{x}_t - x_t)^2} \right), \quad (31)$$

where $x_{\max}$ is the maximum amplitude of the real signal. A higher PSNR indicates better recovery quality.

Pearson Correlation Coefficient (CC) measures the linear correlation between the recovered and real signals, defined as

$$CC = \frac{\sum_{t=1}^T (\hat{x}_t - \bar{\hat{x}})(x_t - \bar{x})}{\sqrt{\sum_{t=1}^T (\hat{x}_t - \bar{\hat{x}})^2} \sqrt{\sum_{t=1}^T (x_t - \bar{x})^2}}, \quad (32)$$

where $\bar{x}$ and $\bar{\hat{x}}$ are the means of the real and recovered signals, respectively. CC ranges in $[-1, 1]$, and the closer the absolute value is to 1, the higher the similarity.

3.2. Experimental Settings

Our experimental settings follow the problem description in Section 2.1.2 and the task design in Section 2.1.3. The noisy signal y is generated by linearly mixing the real signal x with the artifacts, where the mixing ratio is determined according to (1), where the contamination coefficient λ_j is determined by the target SNR using (4). Specific SNR settings for each task are described in Section 2.1.3, and Figure 1 shows some examples at different SNR levels. For each task, the training, validation, and test sets contain 3000, 200, and 200 samples, respectively.

The proposed SWTT uses the sym4 wavelet for an $L = 4$ level SWT decomposition (see Section 2.2.1), yielding a set of approximation coefficients $c_A^{(4)}$ and four sets of detail coefficients $c_D^{(4)}, \dots, c_D^{(1)}$. The TCN branch (see Section 2.2.2) has $M = 6$ TCN blocks, each with $d^{\text{TCN}} = 64$ hidden channels. The Transformer branch (see Section 2.2.3) has $N = 1$ Transformer blocks, each with an embedding dimension of $d^{\text{TF}} = 4$ and two attention heads. The outputs of the two branches are fused by element-wise addition (see Section 2.2.4), and then the final recovered signal $\hat{x}$ is generated through a refinement module (see Section 2.2.5). A dropout rate of 0.1 is applied uniformly across the entire network.

The model is trained using the AdamW optimizer with an initial learning rate of 10^{-3} and a weight decay coefficient of 10^{-4} . Training lasts 100 epochs with a batch size of 64. The StepLR learning rate scheduler decays the learning rate by a factor of 0.3 every 10 epochs. All experiments are implemented in Python 3.9 and PyTorch 2.5.1, combined with CUDA 12.1 and cuDNN 9.1. Training and evaluation are performed on a Linux workstation equipped with an NVIDIA A40 GPU. To ensure reproducibility, the random seed for all experiments is fixed at 2026. All comparison methods used the parameter configurations from their original works.

3.3. Performance Evaluation

For quantitative evaluation, Tables 1–3 present the overall performance of all methods on EOG artifact removal, EMG artifact removal, and EOG–EMG mixed-artifact removal tasks, respectively. SWTT consistently performs best across all four metrics for all three tasks. Its performance advantage is most significant on the EOG artifact removal task, with SWTT achieving a t-RRMSE of 0.247 ± 0.105 , a 28.0% improvement over the next best method. Due to the broadband nature of EMG artifacts, the EMG artifact removal task is more challenging for all methods. SWTT still leads, but the performance gap narrows, indicating that multi-scale decomposition can only provide partial but not complete frequency separation. In the mixed-artifact removal task that most closely resembles real acquisition conditions, SWTT maintains the best overall performance, demonstrating that its multi-scale representation can effectively handle the coexistence of narrow-band EOG artifacts and broadband EMG artifacts.

Table 1. Comparison of different methods on the EOG artifact removal task (mean $\pm$ std). $\downarrow$ indicates lower is better; $\uparrow$ indicates higher is better. The best results are shown in **bold**.

Method	t-RRMSE $\downarrow$	Δ SNR $\uparrow$	PSNR $\uparrow$	CC $\uparrow$
EMD [11]	1.295 $\pm$ 0.330	0.522 $\pm$ 1.080	8.132 $\pm$ 2.482	0.259 $\pm$ 0.155
1D-ResCNN [14]	0.487 $\pm$ 0.089	8.894 $\pm$ 2.755	16.504 $\pm$ 2.025	0.873 $\pm$ 0.050
FCNN [13]	0.506 $\pm$ 0.176	9.012 $\pm$ 4.198	16.622 $\pm$ 3.734	0.850 $\pm$ 0.113
SimpleCNN [13]	0.612 $\pm$ 0.092	6.856 $\pm$ 2.509	14.465 $\pm$ 1.808	0.792 $\pm$ 0.071
NovelCNN [13]	0.375 $\pm$ 0.147	11.648 $\pm$ 3.524	19.258 $\pm$ 3.544	0.920 $\pm$ 0.065
DeepSeparator [13]	0.463 $\pm$ 0.095	9.368 $\pm$ 2.787	16.978 $\pm$ 2.183	0.887 $\pm$ 0.049
LSTM [13]	0.423 $\pm$ 0.122	10.303 $\pm$ 3.159	17.912 $\pm$ 2.685	0.902 $\pm$ 0.059
DeT [19]	0.451 $\pm$ 0.101	9.619 $\pm$ 2.839	17.228 $\pm$ 2.246	0.890 $\pm$ 0.056
DTP-Net [16]	0.343 $\pm$ 0.117	12.264 $\pm$ 3.443	19.874 $\pm$ 3.207	0.934 $\pm$ 0.051
DenoiseEEG [20]	0.442 $\pm$ 0.111	9.861 $\pm$ 3.047	17.471 $\pm$ 2.613	0.893 $\pm$ 0.055
WNOTNet [22]	0.389 $\pm$ 0.121	11.112 $\pm$ 3.380	18.722 $\pm$ 2.949	0.915 $\pm$ 0.057
MHNoiseRegression [17]	0.365 $\pm$ 0.124	11.727 $\pm$ 3.593	19.337 $\pm$ 3.126	0.925 $\pm$ 0.055
SWTT	0.247 $\pm$ 0.105	15.334 $\pm$ 3.834	22.944 $\pm$ 3.779	0.964 $\pm$ 0.036

Table 2. Comparison of different methods on the EMG artifact removal task (mean $\pm$ std). $\downarrow$ indicates lower is better; $\uparrow$ indicates higher is better. The best results are shown in **bold**.

Method	t-RRMSE $\downarrow$	Δ SNR $\uparrow$	PSNR $\uparrow$	CC $\uparrow$
EMD [11]	0.990 $\pm$ 0.041	2.597 $\pm$ 2.767	10.207 $\pm$ 1.344	0.132 $\pm$ 0.164
1D-ResCNN [14]	0.575 $\pm$ 0.131	7.522 $\pm$ 2.009	15.132 $\pm$ 2.278	0.820 $\pm$ 0.087
FCNN [13]	0.558 $\pm$ 0.163	7.995 $\pm$ 3.442	15.605 $\pm$ 3.264	0.824 $\pm$ 0.114
SimpleCNN [13]	0.630 $\pm$ 0.152	6.746 $\pm$ 1.862	14.355 $\pm$ 2.369	0.787 $\pm$ 0.108
NovelCNN [13]	0.570 $\pm$ 0.135	7.613 $\pm$ 2.059	15.222 $\pm$ 2.374	0.820 $\pm$ 0.088
DeepSeparator [13]	0.581 $\pm$ 0.131	7.421 $\pm$ 1.925	15.031 $\pm$ 2.239	0.817 $\pm$ 0.086
LSTM [13]	0.579 $\pm$ 0.133	7.457 $\pm$ 1.951	15.066 $\pm$ 2.239	0.819 $\pm$ 0.086
DeT [19]	0.580 $\pm$ 0.125	7.434 $\pm$ 1.968	15.043 $\pm$ 2.218	0.817 $\pm$ 0.080
DTP-Net [16]	0.563 $\pm$ 0.123	7.695 $\pm$ 2.132	15.305 $\pm$ 2.204	0.826 $\pm$ 0.083
DenoiseEEG [20]	0.554 $\pm$ 0.131	7.883 $\pm$ 2.187	15.492 $\pm$ 2.443	0.828 $\pm$ 0.088
WNOTNet [22]	0.552 $\pm$ 0.132	7.917 $\pm$ 2.209	15.527 $\pm$ 2.405	0.829 $\pm$ 0.090
MHNoiseRegression [17]	0.543 $\pm$ 0.131	8.059 $\pm$ 2.245	15.669 $\pm$ 2.425	0.837 $\pm$ 0.084
SWTT	0.478 $\pm$ 0.134	9.256 $\pm$ 2.333	16.866 $\pm$ 2.882	0.871 $\pm$ 0.077

To assess whether SWTT's numerical advantage over the best-performing comparative method MHNoiseRegression reflects a statistically meaningful difference rather than random variation, we conduct both parametric and non-parametric significance tests on the mixed-artifact task. The paired *t*-test evaluates whether the mean difference between SWTT and MHNoiseRegression is significantly different from zero, assuming approximate normality of the paired differences. The paired Wilcoxon signed-rank test serves as a non-parametric complement that does not rely on distributional assumptions, testing whether the median difference differs from zero. We also report effect sizes because statistical significance alone, particularly with large sample sizes, can flag trivially small differences as significant; effect sizes quantify the practical magnitude of SWTT's improvement. The difference used to calculate the mean/median is obtained by subtracting the SWTT's metric from the MHNoiseRegression's metric. For t-RRMSE where lower values are better, a positive difference indicates that SWTT outperforms MHNoiseRegression. For Δ SNR, PSNR, and CC, where higher values are better, a negative difference indicates that SWTT outperforms MHNoiseRegression. Table 4 reports the results of the paired *t*-test and the paired Wilcoxon signed-rank test. For the paired *t*-test, the 95% confidence intervals exclude zero for all four metrics, and all *p*-values are well below 10^{-160} . The absolute value of Cohen's *d* ranges from 0.676 to 0.765, indicating medium-to-large effect sizes. For the paired Wilcoxon signed-rank test results, all *p*-values are below 10^{-140} , and the absolute value of the rank-biserial correlation r_{rb} ranges from 0.664 to 0.715, confirming large effect

sizes. Collectively, these results confirm that SWTT's superiority over the best competing method is both statistically significant and practically meaningful.

Table 3. Comparison of different methods on the EOG–EMG mixed-artifact removal task (mean $\pm$ std). $\downarrow$ indicates lower is better; $\uparrow$ indicates higher is better. The best results are shown in **bold**.

Method	t-RRMSE $\downarrow$	Δ SNR $\uparrow$	PSNR $\uparrow$	CC $\uparrow$
EMD [11]	1.172 $\pm$ 0.241	5.156 $\pm$ 1.909	8.886 $\pm$ 2.081	0.088 $\pm$ 0.139
1D-ResCNN [14]	0.677 $\pm$ 0.088	9.842 $\pm$ 1.767	13.571 $\pm$ 1.641	0.739 $\pm$ 0.078
FCNN [13]	0.627 $\pm$ 0.179	10.845 $\pm$ 3.009	14.575 $\pm$ 3.271	0.762 $\pm$ 0.157
SimpleCNN [13]	0.834 $\pm$ 0.085	8.000 $\pm$ 1.651	11.729 $\pm$ 1.492	0.562 $\pm$ 0.117
NovelCNN [13]	0.650 $\pm$ 0.112	10.249 $\pm$ 1.782	13.979 $\pm$ 1.910	0.763 $\pm$ 0.087
DeepSeparator [13]	0.671 $\pm$ 0.093	9.929 $\pm$ 1.762	13.658 $\pm$ 1.690	0.745 $\pm$ 0.080
LSTM [13]	0.669 $\pm$ 0.098	9.968 $\pm$ 1.767	13.697 $\pm$ 1.714	0.749 $\pm$ 0.082
DeT [19]	0.664 $\pm$ 0.093	10.013 $\pm$ 1.747	13.743 $\pm$ 1.711	0.752 $\pm$ 0.076
DTP-Net [16]	0.650 $\pm$ 0.112	10.245 $\pm$ 1.864	13.974 $\pm$ 1.919	0.760 $\pm$ 0.090
DenoiseEEG [20]	0.651 $\pm$ 0.096	10.203 $\pm$ 1.859	13.933 $\pm$ 1.748	0.758 $\pm$ 0.081
WNOTNet [22]	0.653 $\pm$ 0.104	10.191 $\pm$ 1.831	13.921 $\pm$ 1.814	0.759 $\pm$ 0.089
MHNoiseRegression [17]	0.624 $\pm$ 0.106	10.599 $\pm$ 1.930	14.328 $\pm$ 1.884	0.780 $\pm$ 0.084
SWTT	0.531 $\pm$ 0.137	12.183 $\pm$ 2.326	15.913 $\pm$ 2.727	0.839 $\pm$ 0.093

Table 4. Significance tests between SWTT and the best-performing comparative method MHNoiseRegression on the EOG–EMG mixed-artifact removal task using paired *t*-tests and paired Wilcoxon signed-rank tests. The difference used to calculate the mean/median is obtained by subtracting the SWTT's metric from the MHNoiseRegression's metric. For t-RRMSE, where lower values are better, a positive difference favors SWTT, whereas for Δ SNR, PSNR, and CC, where higher values are better, a negative difference favors SWTT.

Paired <i>t</i> -Test				
Metric	95% CI	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
t-RRMSE	[0.088,0.099]	34.22	2.16×10^{-202}	0.765
Δ SNR	[−1.675,−1.495]	−34.66	1.62×10^{-206}	−0.775
PSNR	[−1.675,−1.495]	−34.66	1.62×10^{-206}	−0.775
CC	[−0.062,−0.055]	−30.23	1.12×10^{-165}	−0.676
Paired Wilcoxon Signed-Rank Test				
Metric	95% CI	<i>z</i>	<i>p</i>	<i>r_{rb}</i>
t-RRMSE	[0.088,0.100]	27.28	7.04×10^{-164}	0.704
Δ SNR	[−1.640,−1.451]	−27.71	6.00×10^{-169}	−0.715
PSNR	[−1.640,−1.451]	−27.71	6.00×10^{-169}	−0.715
CC	[−0.063,−0.056]	−25.73	5.95×10^{-146}	−0.664

For qualitative evaluation, Figures 4–6 show the waveforms of the artifact removal results of six representative methods on the first test sample for each task. In all three tasks, the reconstructed signal generated by SWTT closely matches the corresponding real signal, exhibiting significantly better waveform consistency and fewer residual artifacts compared to the comparison methods. Overall, these results demonstrate that SWTT achieves more accurate signal reconstruction and more effective artifact removal under various contamination scenarios.

Since the artifacts in the EOG–EMG mixed-artifact task are distributed across different frequency bands, we use the s-RRMSE metric for each sub-band to analyze the artifact removal results in more detail, as shown in Table 5. SWTT achieves the lowest error in four of the five frequency bands, while MHNoiseRegression achieves the best result in the alpha band. This indicates that the current SWT configuration demonstrates limited performance in separating the mid-frequency portion of the EEG and EMG artifacts. We

speculate that this phenomenon is mainly caused by the fact that the global temporal optimization objective is not specifically designed for individual frequency bands. Since the model is trained to minimize the overall temporal-domain MSE rather than band-specific losses, the optimization process may prioritize frequency components that contribute more significantly to the overall loss, resulting in relatively limited improvement in some specific bands such as the alpha band.

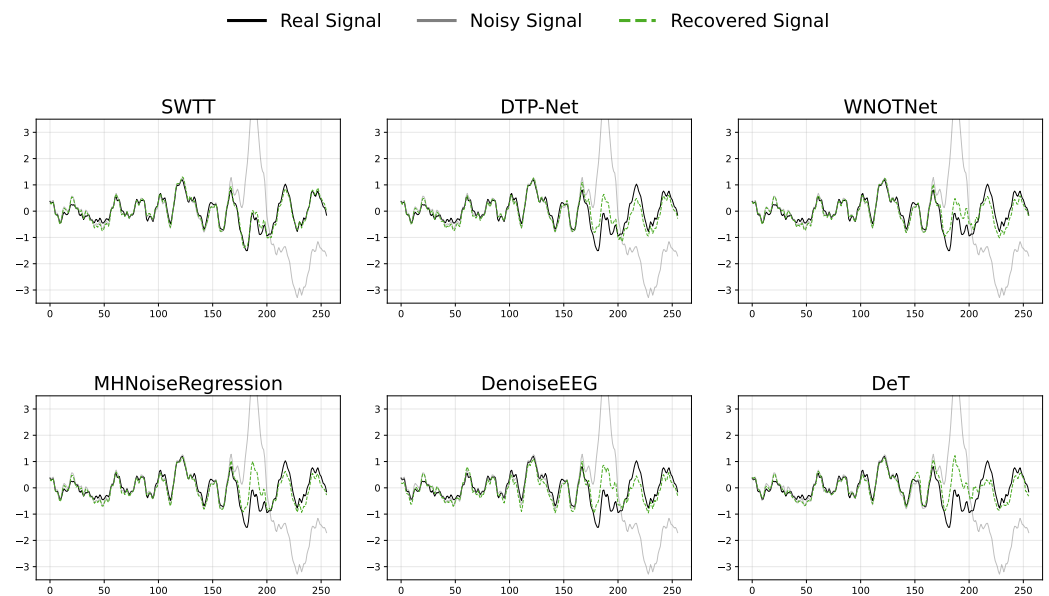


Figure 4. An example of recovery results using the first test sample from the EOG artifact removal task. Black: real; gray: noisy; green dashed: recovered.

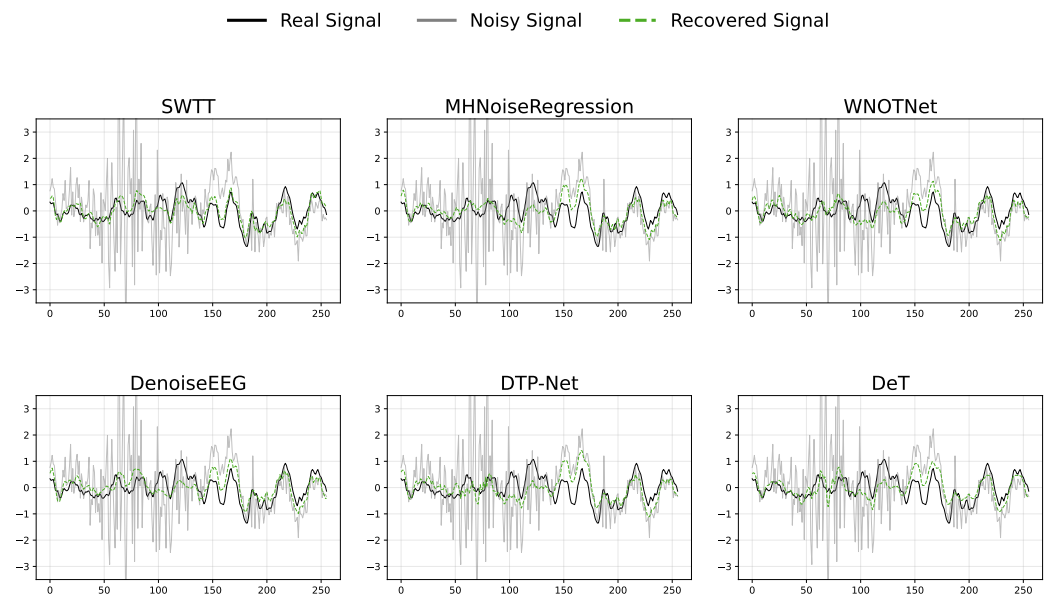


Figure 5. An example of recovery results using the first test sample from the EMG artifact removal task. Black: real; gray: noisy; green dashed: recovered.

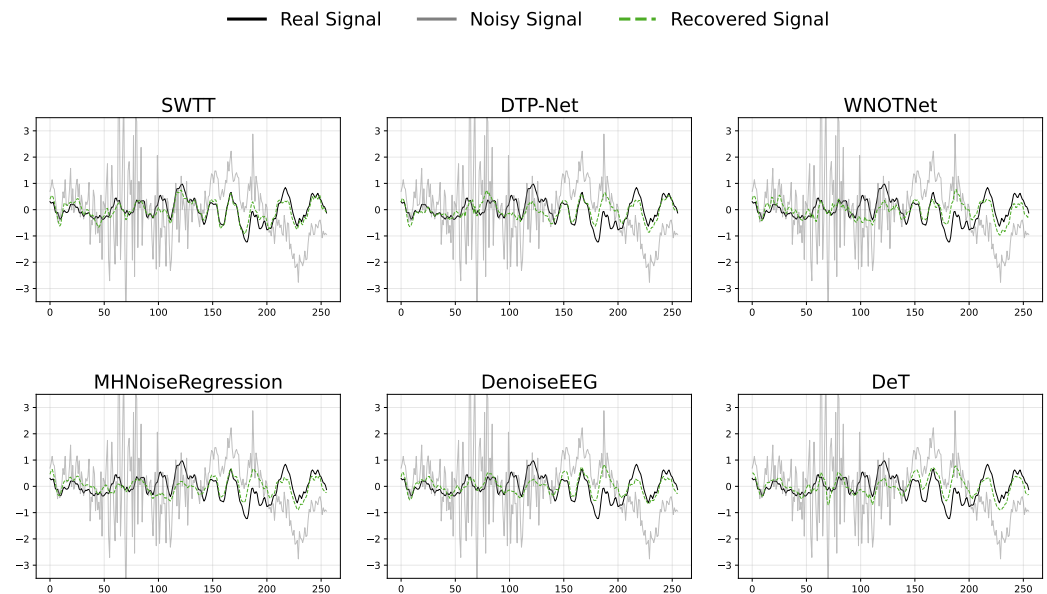


Figure 6. An example of recovery results using the first test sample from the EOG–EMG mixed-artifact task. Black: real; gray: noisy; green dashed: recovered.

Table 5. Band-wise s-RRMSE $\downarrow$ comparison on the EOG–EMG mixed-artifact removal task (mean $\pm$ std). $\downarrow$ indicates lower is better. The best results are shown in **bold**.

Method	$\delta \downarrow$	$\theta \downarrow$	$\alpha \downarrow$	$\beta \downarrow$	$\gamma \downarrow$
EMD [11]	0.859 ± 0.478	0.764 ± 0.165	0.905 ± 0.081	0.891 ± 0.097	0.895 ± 0.108
1D-ResCNN [14]	0.699 ± 0.087	0.497 ± 0.104	0.245 ± 0.089	0.477 ± 0.100	0.693 ± 0.075
FCNN [13]	0.505 ± 0.178	0.525 ± 0.172	0.335 ± 0.119	0.629 ± 0.158	0.709 ± 0.131
SimpleCNN [13]	0.606 ± 0.136	0.652 ± 0.074	0.658 ± 0.056	0.553 ± 0.084	0.712 ± 0.081
NovelCNN [13]	0.537 ± 0.145	0.420 ± 0.127	0.231 ± 0.092	0.475 ± 0.081	0.739 ± 0.045
DeepSeparator [13]	0.682 ± 0.094	0.434 ± 0.144	0.237 ± 0.092	0.471 ± 0.083	0.742 ± 0.047
LSTM [13]	0.638 ± 0.109	0.443 ± 0.138	0.241 ± 0.089	0.492 ± 0.081	0.771 ± 0.051
DeT [19]	0.608 ± 0.114	0.443 ± 0.133	0.238 ± 0.095	0.502 ± 0.071	0.771 ± 0.050
DTP-Net [16]	0.541 ± 0.145	0.439 ± 0.117	0.235 ± 0.088	0.449 ± 0.094	0.691 ± 0.044
DenoiseEEG [20]	0.663 ± 0.105	0.430 ± 0.131	0.234 ± 0.100	0.444 ± 0.103	0.701 ± 0.094
WNotNet [22]	0.605 ± 0.117	0.428 ± 0.150	0.262 ± 0.085	0.447 ± 0.094	0.680 ± 0.087
MHNoiseRegression [17]	0.563 ± 0.128	0.424 ± 0.126	0.225 ± 0.093	0.453 ± 0.099	0.702 ± 0.088
SWTT	0.412 ± 0.176	0.376 ± 0.138	0.230 ± 0.099	0.396 ± 0.105	0.598 ± 0.092

We further evaluate the computational efficiency of all methods on a mixed-artifact task, as summarized in Table 6. As can be seen, although our model outperforms some comparative methods in terms of the number of trainable parameters, floating point operations per second (FLOPs), inference time, and GPU memory consumption, this is mainly due to the additional parameter overhead brought by the self-attention mechanism and feed-forward network in the Transformer branch. This branch can establish global dependencies across the entire time series, thereby enhancing the model's ability to model long-range temporal features of EEG signals and further improving denoising performance. Therefore, the increased model complexity is an inevitable cost of achieving better performance, reflecting the trade-off between model size and performance.

Furthermore, to evaluate artifacts of different contamination intensities, Figure 7 shows the performance metrics of each model at different SNR levels in the EOG and EMG artifact removal tasks. SWTT, with its cross-band global receptive field provided by the Transformer and its ability to learn local temporal patterns through TCN, achieves optimal performance across all contamination levels. Meanwhile, in the EMG artifact removal task, we note that FCNN outperforms other comparative methods at high contamination intensities, but its performance degrades rapidly at low contamination intensities. We

speculate that this is because FCNN effectively suppresses high-intensity artifacts within its global receptive field, but due to its lack of explicit local temporal modeling capabilities, it leads to excessive signal smoothing at low contamination intensities, resulting in the misclassification of signal as noise.

Table 6. Computational efficiency on the EOG–EMG mixed-artifact removal task, including number of trainable parameters, floating-point operations (FLOPs), inference time per sample, and GPU memory consumption during inference.

Method	Params (M)	FLOPs (M)	Time (ms)	GPU Mem. (MiB)
EMD [11]	0.00	N/A	5.46	8.9
1D-ResCNN [14]	0.04	17.4 M	1.61	12.5
FCNN [13]	1.05	1.0 M	0.20	17.2
SimpleCNN [13]	0.04	19.0 M	0.56	10.7
NovelCNN [13]	46.15	551.4 M	3.12	415.8
DeepSeparator [13]	0.03	16.0 M	2.30	1.5
LSTM [13]	0.03	0.0 M	0.71	29.1
DeT [19]	0.14	0.1 M	1.55	11.5
DTP-Net [16]	0.57	37.1 M	1.58	15.2
DenoiseEEG [20]	1.00	270.1 M	4.44	40.0
WNOTNet [22]	0.08	25.5 M	2.22	16.8
MHNoiseRegression [17]	0.03	17.4 M	1.05	11.3
SWTT	1.77	96.3 M	3.05	28.6

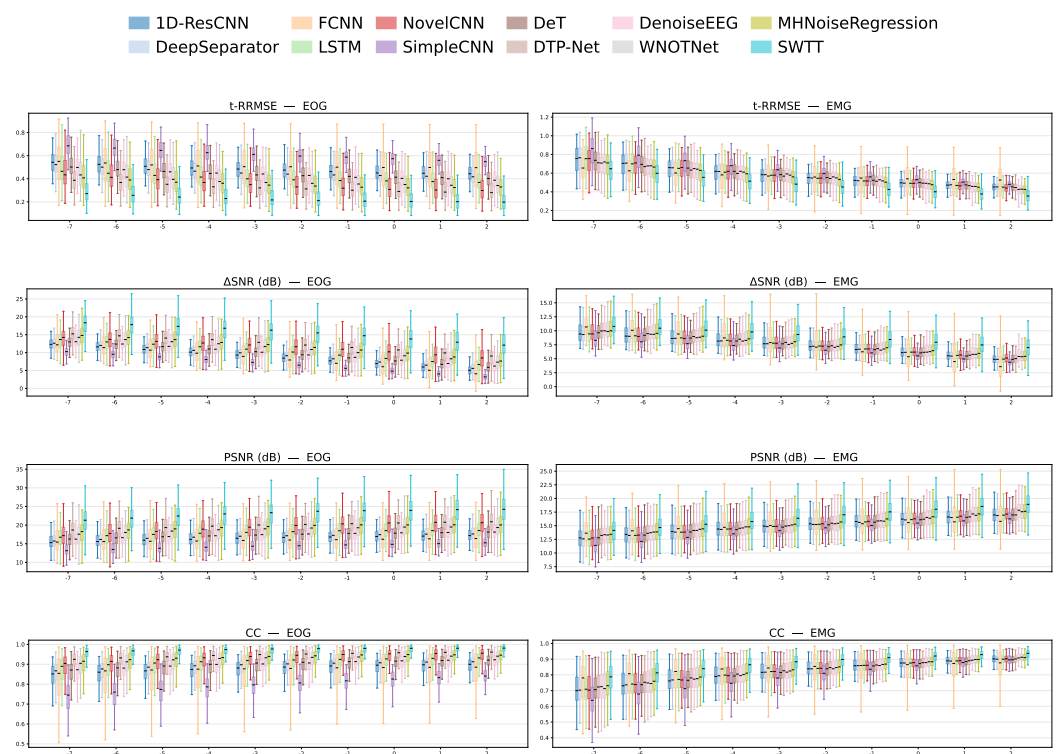


Figure 7. The performance metrics of each model at different SNR levels in the EOG and EMG artifact removal tasks.

3.4. Ablation Study

We evaluate each branch by comparing the complete SWTT model with two ablation variants: SWTT_{TCN} and SWTT_{TF}, which retain only the TCN and Transformer branches, respectively. In these variants, the fusion module is removed, and each branch is used independently for signal recovery. All experiments are performed on the EOG–EMG mixed-artifact removal task to evaluate the contributions of the TCN and Transformer modules.

Table 7 shows the overall metrics, and Table 8 shows the s-RRMSE for each frequency band. SWTT_{TF} outperforms SWTT_{TCN} in overall metrics, indicating that global frequency modeling plays a dominant role in mixed-artifact scenarios. However, the complete SWTT combining both branches achieves the best results across all metrics and frequency bands, suggesting that the Transformer's global cross-band attention mechanism complements the TCN's local pattern extraction capabilities. Notably, the TCN branch contributes more in the alpha band, while the Transformer provides a relatively uniform gain across the entire spectrum, further validating that local and global modeling address different aspects of the artifact removal problem.

Table 7. Ablation study of SWTT branch contributions on the EOG–EMG mixed-artifact removal task (mean $\pm$ std). SWTT_{TCN} and SWTT_{TF} denote the TCN-only and Transformer-only variants, respectively. $\downarrow$ indicates lower is better; $\uparrow$ indicates higher is better. The best results are shown in **bold**.

Variant	t-RRMSE $\downarrow$	Δ SNR $\uparrow$	PSNR $\uparrow$	CC $\uparrow$
SWTT _{TCN}	0.629 $\pm$ 0.116	10.557 $\pm$ 1.956	14.286 $\pm$ 2.028	0.780 $\pm$ 0.086
SWTT _{TF}	0.553 $\pm$ 0.165	11.939 $\pm$ 2.676	15.668 $\pm$ 3.116	0.820 $\pm$ 0.114
SWTT	0.531 $\pm$ 0.137	12.183 $\pm$ 2.326	15.913 $\pm$ 2.727	0.839 $\pm$ 0.093

Table 8. Band-wise s-RRMSE $\downarrow$ ablation results on the EOG–EMG mixed-artifact removal task (mean $\pm$ std). $\downarrow$ indicates lower is better. The best results are shown in **bold**.

Variant	δ $\downarrow$	θ $\downarrow$	α $\downarrow$	β $\downarrow$	γ $\downarrow$
SWTT _{TCN}	0.514 $\pm$ 0.142	0.406 $\pm$ 0.146	0.225 $\pm$ 0.101	0.448 $\pm$ 0.098	0.712 $\pm$ 0.094
SWTT _{TF}	0.425 $\pm$ 0.194	0.396 $\pm$ 0.154	0.240 $\pm$ 0.100	0.407 $\pm$ 0.111	0.631 $\pm$ 0.075
SWTT	0.412 $\pm$ 0.176	0.376 $\pm$ 0.138	0.230 $\pm$ 0.099	0.396 $\pm$ 0.105	0.598 $\pm$ 0.092

3.5. Parameter Analyses

SWTT contains several configurable parameters that may affect both the model performance and computational complexity. In this subsection, we investigate the effects of different parameter configurations in the SWT decomposition and Transformer branches. Based on these analyses, we select a configuration that achieves a favorable trade-off between performance and model complexity.

Due to the large number of possible combinations of wavelet families and decomposition levels, an exhaustive search is computationally inefficient. Therefore, we first adopted an empirical configuration with $L = 4$ and the sym4 wavelet as the baseline, and then we investigated the influence of different wavelet families and decomposition levels. The results are reported in Tables 9 and 10. Among the evaluated wavelet families with $L = 4$, sym4 achieved the lowest t-RRMSE, demonstrating its suitability for EEG artifact removal. Using sym4, increasing the decomposition level from 1 to 5 gradually improved the performance, with $L = 5$ achieving the best overall results. However, the improvement from $L = 4$ to $L = 5$ was marginal, while the latter introduced additional sub-bands and increased computational cost. Further increasing the level to $L = 6$ degraded the performance, likely due to over-decomposition. Therefore, we selected the sym4 wavelet with $L = 4$ as the final configuration, achieving a favorable balance between performance and efficiency.

We also examine the Transformer branch configuration by varying the number of attention heads and the embedding dimension d^{TF} . The results are presented in Tables 11 and 12. In most cases, increasing the number of attention heads improves the model performance; however, the improvement is relatively limited. Increasing d^{TF} from 4 to 16 achieves the highest Δ SNR, but it also introduces a substantial increase in the number of parameters in the Transformer branch. Therefore, considering the trade-off between model complexity and performance, we adopt two attention heads with $d^{\text{TF}} = 4$ in our final model.

Table 9. Effect of wavelet basis and decomposition level on the EOG–EMG mixed-artifact removal task (mean $\pm$ std). $\downarrow$ indicates lower is better; $\uparrow$ indicates higher is better. The best results are shown in **bold**.

L	Wavelet	t-RRMSE $\downarrow$	Δ SNR $\uparrow$	PSNR $\uparrow$	CC $\uparrow$
1	sym4	0.578 $\pm$ 0.139	11.405 $\pm$ 2.070	15.134 $\pm$ 2.610	0.809 $\pm$ 0.100
2	sym4	0.585 $\pm$ 0.142	11.292 $\pm$ 2.062	15.022 $\pm$ 2.601	0.804 $\pm$ 0.107
3	sym4	0.556 $\pm$ 0.142	11.762 $\pm$ 2.215	15.491 $\pm$ 2.712	0.822 $\pm$ 0.101
4	sym4	0.531 $\pm$ 0.137	12.183 $\pm$ 2.326	15.913 $\pm$ 2.727	0.839 $\pm$ 0.093
5	sym4	0.521 $\pm$ 0.144	12.378 $\pm$ 2.444	16.107 $\pm$ 2.803	0.844 $\pm$ 0.101
6	sym4	0.543 $\pm$ 0.143	11.993 $\pm$ 2.312	15.722 $\pm$ 2.750	0.831 $\pm$ 0.099
4	bior4.4	0.534 $\pm$ 0.144	12.141 $\pm$ 2.363	15.871 $\pm$ 2.802	0.835 $\pm$ 0.099
4	db4	0.536 $\pm$ 0.145	12.113 $\pm$ 2.389	15.843 $\pm$ 2.778	0.834 $\pm$ 0.101
4	haar	0.534 $\pm$ 0.144	12.145 $\pm$ 2.355	15.874 $\pm$ 2.808	0.836 $\pm$ 0.100

Table 10. Band-wise s-RRMSE $\downarrow$ for different wavelet bases and decomposition levels on the EOG–EMG mixed-artifact removal task (mean $\pm$ std). $\downarrow$ indicates lower is better. The best results are shown in **bold**.

L	Wavelet	δ $\downarrow$	θ $\downarrow$	α $\downarrow$	β $\downarrow$	γ $\downarrow$
1	sym4	0.451 $\pm$ 0.161	0.401 $\pm$ 0.136	0.245 $\pm$ 0.097	0.418 $\pm$ 0.109	0.604 $\pm$ 0.081
2	sym4	0.449 $\pm$ 0.171	0.401 $\pm$ 0.149	0.249 $\pm$ 0.099	0.423 $\pm$ 0.108	0.641 $\pm$ 0.088
3	sym4	0.412 $\pm$ 0.166	0.386 $\pm$ 0.135	0.234 $\pm$ 0.091	0.411 $\pm$ 0.105	0.614 $\pm$ 0.082
4	sym4	0.412 $\pm$ 0.176	0.376 $\pm$ 0.138	0.230 $\pm$ 0.099	0.396 $\pm$ 0.105	0.598 $\pm$ 0.092
5	sym4	0.391 $\pm$ 0.180	0.367 $\pm$ 0.137	0.231 $\pm$ 0.091	0.397 $\pm$ 0.104	0.582 $\pm$ 0.088
6	sym4	0.400 $\pm$ 0.161	0.370 $\pm$ 0.144	0.234 $\pm$ 0.092	0.404 $\pm$ 0.106	0.589 $\pm$ 0.092
4	bior4.4	0.403 $\pm$ 0.174	0.384 $\pm$ 0.145	0.230 $\pm$ 0.096	0.386 $\pm$ 0.105	0.598 $\pm$ 0.082
4	db4	0.397 $\pm$ 0.171	0.374 $\pm$ 0.133	0.230 $\pm$ 0.095	0.398 $\pm$ 0.105	0.609 $\pm$ 0.090
4	haar	0.405 $\pm$ 0.177	0.382 $\pm$ 0.140	0.233 $\pm$ 0.098	0.393 $\pm$ 0.107	0.609 $\pm$ 0.087

Table 11. Effect of Transformer configuration on the EOG–EMG mixed-artifact removal task (mean $\pm$ std). $\downarrow$ indicates lower is better; $\uparrow$ indicates higher is better. The best results are shown in **bold**.

Heads	d^{TF}	t-RRMSE $\downarrow$	Δ SNR $\uparrow$	PSNR $\uparrow$	CC $\uparrow$
2	4	0.531 $\pm$ 0.137	12.183 $\pm$ 2.326	15.913 $\pm$ 2.727	0.839 $\pm$ 0.093
2	8	0.531 $\pm$ 0.160	12.269 $\pm$ 2.613	15.998 $\pm$ 3.043	0.835 $\pm$ 0.112
2	16	0.584 $\pm$ 0.141	11.313 $\pm$ 2.074	15.042 $\pm$ 2.562	0.811 $\pm$ 0.101
4	4	0.525 $\pm$ 0.144	12.308 $\pm$ 2.412	16.038 $\pm$ 2.848	0.840 $\pm$ 0.099
4	8	0.534 $\pm$ 0.157	12.199 $\pm$ 2.499	15.928 $\pm$ 2.963	0.836 $\pm$ 0.109
4	16	0.547 $\pm$ 0.165	12.017 $\pm$ 2.556	15.746 $\pm$ 3.058	0.827 $\pm$ 0.117
8	4	0.522 $\pm$ 0.147	12.379 $\pm$ 2.468	16.108 $\pm$ 2.914	0.842 $\pm$ 0.100
8	8	0.524 $\pm$ 0.152	12.355 $\pm$ 2.497	16.084 $\pm$ 2.982	0.840 $\pm$ 0.105
8	16	0.521 $\pm$ 0.174	12.529 $\pm$ 2.842	16.258 $\pm$ 3.303	0.841 $\pm$ 0.116

Table 12. Band-wise s-RRMSE $\downarrow$ for different Transformer configurations on the EOG–EMG mixed-artifact removal task (mean $\pm$ std). $\downarrow$ indicates lower is better. The best results are shown in **bold**.

Heads	d^{TF}	δ $\downarrow$	θ $\downarrow$	α $\downarrow$	β $\downarrow$	γ $\downarrow$
2	4	0.412 $\pm$ 0.176	0.376 $\pm$ 0.138	0.230 $\pm$ 0.099	0.396 $\pm$ 0.105	0.598 $\pm$ 0.092
2	8	0.388 $\pm$ 0.173	0.370 $\pm$ 0.146	0.231 $\pm$ 0.093	0.390 $\pm$ 0.111	0.554 $\pm$ 0.089
2	16	0.438 $\pm$ 0.171	0.405 $\pm$ 0.136	0.233 $\pm$ 0.095	0.418 $\pm$ 0.100	0.613 $\pm$ 0.076
4	4	0.409 $\pm$ 0.181	0.367 $\pm$ 0.143	0.226 $\pm$ 0.091	0.392 $\pm$ 0.108	0.594 $\pm$ 0.092
4	8	0.389 $\pm$ 0.173	0.375 $\pm$ 0.150	0.234 $\pm$ 0.099	0.388 $\pm$ 0.110	0.574 $\pm$ 0.086
4	16	0.400 $\pm$ 0.186	0.384 $\pm$ 0.153	0.243 $\pm$ 0.096	0.394 $\pm$ 0.112	0.534 $\pm$ 0.097
8	4	0.400 $\pm$ 0.184	0.372 $\pm$ 0.146	0.227 $\pm$ 0.092	0.388 $\pm$ 0.107	0.587 $\pm$ 0.097
8	8	0.388 $\pm$ 0.181	0.378 $\pm$ 0.140	0.230 $\pm$ 0.093	0.392 $\pm$ 0.111	0.552 $\pm$ 0.097
8	16	0.375 $\pm$ 0.180	0.371 $\pm$ 0.148	0.239 $\pm$ 0.102	0.380 $\pm$ 0.118	0.511 $\pm$ 0.103

3.6. Interpretability

Here we further demonstrate the interpretability of SWTT based on the analysis of the wavelet coefficients. First, we examine the artifact energy distribution across SWT sub-bands. Let

c_k^{noisy} and c_k^{real} denote the SWT coefficients of the noisy and real signals, respectively. The artifact component is $c_k^{\text{art}} = c_k^{\text{noisy}} - c_k^{\text{real}}$. The noise ratio R_k for the k -th sub-band is defined as

$$R_k = \frac{\sum_t (c_{k,t}^{\text{art}})^2}{\sum_t (c_{k,t}^{\text{real}})^2 + \sum_t (c_{k,t}^{\text{art}})^2}, \quad (33)$$

where a larger R_k indicates that the sub-band is more severely contaminated by artifacts.

Figure 8 shows the average R_k values over the test sets. For EOG, the artifact energy is highly concentrated in approximation coefficients $c_A^{(4)}$, while for EMG, significant contamination is observed across all detail sub-bands $c_D^{(4)}$ through $c_D^{(1)}$.

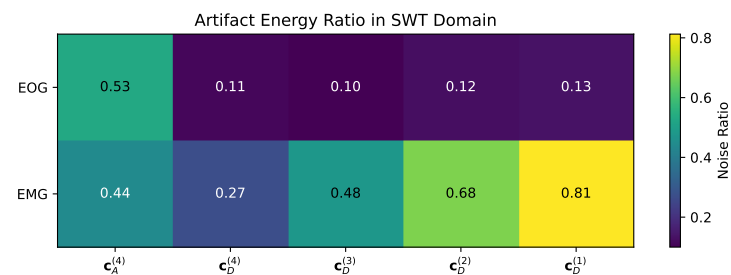


Figure 8. Average artifact energy ratio R_k across SWT sub-bands, computed over the EOG and EMG test sets. Larger values indicate higher artifact contamination in the corresponding sub-band.

Furthermore, Figure 9 shows the recovery results of the wavelet coefficients at selected SWT decomposition levels using the first test samples of the EOG and EMG test sets. In the EOG scenario, the model's correction is most significant on $c_A^{(4)}$, which is consistent with the fact that the energy of EOG artifacts is concentrated in the approximation coefficients. In the EMG scenario, significant corrections are distributed across all detail coefficients, reflecting the broadband characteristics of EMG artifacts. This experiment further demonstrates that the correction behavior of SWTT is consistent with the distribution characteristics of artifacts in the frequency domain: the model mainly adjusts the approximation coefficients for EOG artifacts, while applying corrections distributed across detail coefficients for EMG artifacts.

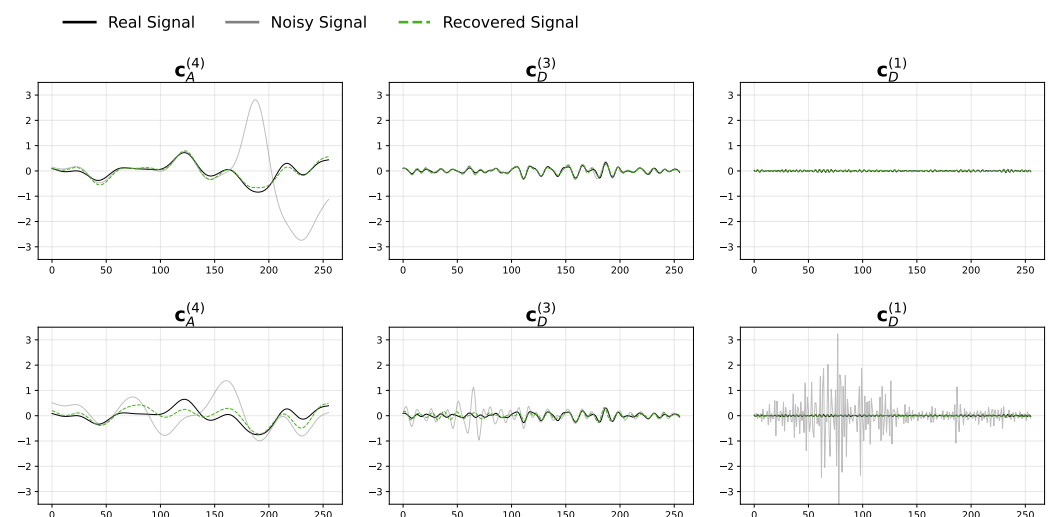


Figure 9. An example of wavelet coefficient recovery results at each SWT decomposition level using the first test sample. The top row shows the EOG results, and the bottom row shows the EMG results. Black: real; gray: noisy; green dashed: recovered.

4. Discussion

The experimental results in Section 3 demonstrate that SWTT outperforms twelve comparison methods across all three tasks and four evaluation metrics. We believe one of the key reasons for this advantage is the temporally aligned multi-scale representation provided by SWT. Most existing deep artifact removal methods either operate directly in the original temporal domain or employ frequency decomposition at the expense of temporal resolution. Pure temporal models, such as FCNN and LSTM [13], lack explicit multi-scale structures, while DTP-Net [16] introduces a learnable filter bank, but its downsampling in the decomposition leads to temporal resolution mismatches between sub-bands. In contrast, SWT in SWTT provides multi-scale signal representation while maintaining temporal alignment, making it more conducive for modeling cross-band artifact structures. Apart from SWTT, other methods typically achieve excellent performance only on certain tasks presented in this paper, with limited capabilities on others. For example, DTP-Net achieves the second-best performance in the EOG artifact removal task, but its performance degrades in the EMG artifact removal task. SWTT performs relatively stably under different artifact types, indicating that it can indeed adapt to various scenarios. Interpretability analysis shows that SWTT has significantly different processing strategies for EOG and EMG artifacts, indicating that the model does not merely learn the specific statistical regularities of the dataset, but also truly utilizes the distribution properties of the artifacts, which is difficult to verify with other models.

This study also has some limitations. First, the frequency band analysis results show that SWTT's performance in the alpha band is not optimal, indicating that the current four-level decomposition of the sym4 wavelet may not provide sufficient frequency resolution to adequately separate artifacts from the underlying EEG. Second, this study only considers EOG and EMG artifacts, while actual recordings may contain other interference sources, such as electrocardiogram (ECG) artifacts. Third, the semi-synthetic data used in this study are constructed by linearly superimposing EOG and EMG artifacts onto clean EEG segments. Although this approach provides controllable ground truth for quantitative evaluation, it cannot fully reproduce the complex and non-stationary noise structure present in real clinical EEG recordings, and the additive model may not capture the non-linear interactions between different artifact sources, which may limit the direct transferability of the reported performance to clinical settings. Fourth, the current evaluation is conducted entirely on a single benchmark dataset. The reported performance may not directly translate to real clinical EEG recordings or datasets acquired under different experimental protocols. Fifth, the current fusion of TCN and Transformer branches uses simple element-wise addition. While this method is parameter-free and efficient, it may not be optimal for integrating complementary local and global representations. Exploring more complex fusion mechanisms is expected to further improve performance.

From a computational perspective, the proposed architecture requires additional memory and computational resources to maintain multi-resolution representations compared to pure temporal models due to the additional computational cost introduced by multi-scale SWT decomposition. However, SWT has no trainable parameters and can be efficiently implemented through parallel filtering operations, and its improved artifact separation capability may offset these additional costs in applications. In clinical applications, reliable EEG artifact removal is crucial for subsequent analysis. SWTT, by better recovering EEG while suppressing physiological artifacts, can serve as a preprocessing module to improve the accuracy and reliability of subsequent EEG processing, such as emotion recognition and depression detection.

Future research will focus on improving the robustness, efficiency, and clinical applicability of SWTT. First, we will investigate its generalization ability on large-scale, multi-center

clinical EEG datasets and recordings containing various artifact sources. Second, we will explore adaptive wavelet decomposition strategies to achieve more flexible frequency resolution and improve the separation of individual bands. Third, we will investigate better feature fusion strategies to more effectively integrate the complementary representations from the TCN and Transformer branches. Finally, further validation on real-world clinical EEG recordings is required to assess the potential of SWTT as a preprocessing component for clinical EEG applications.

5. Conclusions

We propose SWTT, a deep learning framework for EEG artifact removal that integrates SWT, TCN, and Transformer models. SWT provides time-aligned multi-scale representations; the TCN branch captures local temporal patterns in the wavelet domain; and the Transformer models cross-band global dependencies. Experiments on EEGdenoiseNet across EOG artifact removal, EMG artifact removal, and EOG–EMG mixed-artifact removal tasks demonstrate that SWTT consistently outperforms twelve comparison methods on all reconstruction metrics. In sub-band analysis, SWTT achieves the lowest error in four of the five frequency bands. Interpretability analysis shows that SWTT can adapt its denoising strategy based on the physical properties of the artifacts. Nevertheless, SWTT still faces challenges in generalization to diverse clinical scenarios, adaptive frequency decomposition, and efficient feature fusion. Future work will investigate these aspects to further enhance its robustness and clinical applicability.

Author Contributions: Y.W.: Writing—Original Draft, Writing—Review and Editing, Conceptualization, Methodology, Analysis, Validation, Software. Z.C.: Writing—Original Draft, Methodology, Analysis, Validation, Software. B.Z.: Writing—Review and Editing, Methodology, Analysis, Supervision. W.G.: Writing—Review and Editing, Conceptualization, Methodology, Analysis, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Science and Technology Project of Hebei Education Department under Grant CXZX2025038 and QN2023141; Ministry of Education Industry-University Collaborative Project under Grant 241002531081031; Science Foundation of Hebei Normal University under Grant L2024ZD12 and L2024K01.

Data Availability Statement: All original contributions of this work are contained within this article. For further information, please contact the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Niedermeyer, E.; da Silva, F.L. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2005.
2. Sanei, S.; Chambers, J.A. *EEG Signal Processing*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
3. Craik, A.; He, Y.; Contreras-Vidal, J.L. Deep learning for electroencephalogram (EEG) classification tasks: A review. *J. Neural Eng.* **2019**, *16*, 031001. [[CrossRef](#)] [[PubMed](#)]
4. Fatourehchi, M.; Bashashati, A.; Ward, R.K.; Birch, G.E. EMG and EOG artifacts in brain computer interface systems: A survey. *Clin. Neurophysiol.* **2007**, *118*, 480–494. [[CrossRef](#)] [[PubMed](#)]
5. Urigüen, J.A.; Garcia-Zapirain, B. EEG artifact removal—State-of-the-art and guidelines. *J. Neural Eng.* **2015**, *12*, 031001. [[CrossRef](#)] [[PubMed](#)]
6. Donoho, D.L. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory* **1995**, *41*, 613–627. [[CrossRef](#)]
7. Makeig, S.; Bell, A.; Jung, T.P.; Sejnowski, T.J. Independent component analysis of electroencephalographic data. *Adv. Neural Inf. Process. Syst.* **1995**, *8*, 145–151.
8. Jung, T.P.; Makeig, S.; Humphries, C.; Lee, T.W.; Mckeown, M.J.; Iragui, V.; Sejnowski, T.J. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* **2000**, *37*, 163–178. [[CrossRef](#)]

9. Issa, M.F.; Juhasz, Z. Improved EOG Artifact Removal Using Wavelet Enhanced Independent Component Analysis. *Brain Sci.* **2019**, *9*, 355. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [\[CrossRef\]](#)
11. Elouaham, S.; Nassiri, B.; El Abbadi, R.; El Melhaoui, O.; Said, S.; El Khadiri, K.; Dliou, A.; Latif, R.; Dlimi, S.; Zougagh, H. Hybridization Denoising Method for EMG Signals Using EWT and EMD Techniques. *Int. J. Eng. Appl.* **2025**, *13*, 574–581. [\[CrossRef\]](#)
12. Shahbakhti, M.; Maugeon, M.; Beiramvand, M.; Marozas, V. Low complexity automatic stationary wavelet transform for elimination of eye blinks from EEG. *Brain Sci.* **2019**, *9*, 352. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Zhang, H.; Zhao, M.; Wei, C.; Mantini, D.; Li, Z.; Liu, Q. EEGdenoiseNet: A benchmark dataset for deep learning solutions of EEG denoising. *J. Neural Eng.* **2021**, *18*, 056057. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Sun, W.; Su, Y.; Wu, X.; Wu, X. A novel end-to-end 1D-ResCNN model to remove artifact from EEG signals. *Neurocomputing* **2020**, *404*, 108–121. [\[CrossRef\]](#)
15. Lea, C.; Flynn, M.D.; Vidal, R.; Reiter, A.; Hager, G.D. Temporal convolutional networks for action segmentation and detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 156–165.
16. Pei, Y.; Xu, J.; Chen, Q.; Wang, C.; Yu, F.; Zhang, L.; Luo, W. DTP-Net: Learning to reconstruct EEG signals in time-frequency domain by multi-scale feature reuse. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 2662–2673. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Shaikh, U.Q.; Kalra, A.M.; Lowe, A.; Niazi, I.K. Multi-head noise regression for single-channel EEG: Estimating ocular and muscle contamination to guide artifact removal. *J. Neural Eng.* **2026**, *23*, 026024. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 4–9 December 2017; Volume 30.
19. Pu, X.; Yi, P.; Chen, K.; Ma, Z.; Zhao, D.; Ren, Y. EEGDnet: Fusing non-local and local self-similarity for EEG signal denoising with transformer. *Comput. Biol. Med.* **2022**, *151*, 106248. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Chen, W.; Yu, C.; Zhao, Z.; Zheng, N.; Li, H.; Li, Y. Self-Supervised EEG Denoising via Dual-Branch Consistency Learning with Masked Reconstruction. *Knowl.-Based Syst.* **2025**, *330*, 114703. [\[CrossRef\]](#)
21. Cai, Y.; Meng, Z.; Huang, D. DHCT-GAN: Improving EEG signal quality with a dual-branch hybrid CNN–transformer network. *Sensors* **2025**, *25*, 231. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Jiang, X.; Sun, X.; Hao, Y.; Cai, Q.; Hou, H.; Xia, L.; Li, H.; Li, G.; Gao, Z. WNOTNet: A hybrid wavelet neural operator and transformer framework for enhanced EEG denoising. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 2535715. [\[CrossRef\]](#)
23. Meng, Y.; Yuan, C.; Feng, W.; Li, D.; Nan, J.; Xia, Y.; Zhu, F.; Song, J. TF-Denoiser: A Time-Frequency Domain Joint Method for EEG Artifact Removal. *Electronics* **2026**, *15*, 132. [\[CrossRef\]](#)
24. Nason, G.P.; Silverman, B.W. The stationary wavelet transform and some statistical applications. In *Wavelets and Statistics*; Springer: Berlin/Heidelberg, Germany, 1995; pp. 281–299.
25. Roy, Y.; Banville, H.; Albuquerque, I.; Gramfort, A.; Falk, T.H.; Faubert, J. Deep learning-based electroencephalography analysis: A systematic review. *J. Neural Eng.* **2019**, *16*, 051001. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Azhar, M.; Shafique, T.; Amjad, A. A Convolutional Neural Network for the Removal of Simultaneous Ocular and Myogenic Artifacts from EEG Signals. *Electronics* **2024**, *13*, 4576. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.