Physics-Informed Regularization for Trustworthy and Physiologically Plausible EEG Denoising

Author: Mohanarangan Desigan

Date: September 27, 2025

Abstract

The clinical viability of deep learning in neurology is critically dependent on the trustworthiness of its outputs. While deep learning models for EEG denoising can effectively reduce noise, they often fail to preserve the underlying neurophysiological structure, leading to outputs that are clean but diagnostically misleading. This paper argues that this failure is twofold: a failure of naive loss functions that ignore the physics of the signal, and a failure of simplistic evaluation metrics like RMSE that are blind to structural integrity. We address this by introducing and validating a **Physics-Informed Regularizer** we call Spatio-Temporal Physiological Consistency (STPC). STPC augments a standard L1 loss with a Temporal Gradient term to preserve event sharpness and a Spatial Laplacian term to enforce electrodynamic plausibility. We trained a U-Net with both a baseline L1 loss and our STPC loss, evaluating them on a completely held-out epileptic seizure segment. The results reveal a fascinating divergence in metrics: while both models achieved a near-identical, low RMSE (≈1.8e-5) and a similar Mean SSIM (≈0.74), the STPC model demonstrated superior preservation of the signal's frequency content, evidenced by a higher Mean Spectral Coherence. Most critically, qualitative analysis revealed that only the STPC model produced a reconstruction that was visually and structurally faithful to the ground truth, while the baseline model collapsed into non-physiological artifacts. Our findings demonstrate that for AI to be trustworthy in clinical neuroscience, it must be guided by physics-informed principles and evaluated with a holistic combination of visual analysis and targeted, structure-aware metrics.

---

1. Introduction

The diagnosis and treatment planning for neurological disorders like epilepsy increasingly rely on the precise interpretation of EEG signals. An epileptologist's ability to localize a seizure's origin—a key step for potential surgical intervention—depends on the subtle spat-temporal dynamics of the signal. However, these critical signals are often corrupted by noise, creating a significant barrier to both manual and automated analysis.

Deep learning offers a powerful solution for denoising, but current approaches present a hidden danger. When models are trained with standard objectives like Mean Squared Error, they learn to minimize average sample-wise error. This often leads to **oversmoothing**, a phenomenon where sharp, diagnostically vital features are blurred into obscurity. The resulting signal may be "clean" by the numbers, but it can mask the very pathology a clinician is looking for, posing a direct risk of misdiagnosis.

This paper presents a critique of this prevailing paradigm and offers a solution. We argue that the field's reliance on simplistic metrics like RMSE is a methodological pitfall, as it fails to capture the physiological plausibility of a signal. Our primary contribution is a **Physics-Informed Regularizer (STPC)** that instills a physical inductive bias into the network. We demonstrate that this approach produces a visually and spectrally superior result, and we use the surprising non-superiority on the SSIM metric to highlight the necessity of moving towards a more holistic evaluation paradigm for building AI systems that are genuinely trustworthy in high-stakes clinical environments.

---

2. Methods

2.1. Dataset and Preprocessing

The CHB-MIT Scalp EEG Database was used. To ensure a rigorous, leak-free evaluation, a true held-out test set was created by reserving one entire file containing multiple seizures (chb01_03.edf) for validation, while all other files from subject chb01 were used for training.

A robust preprocessing pipeline was developed:

1. **Monopolar Re-referencing:** Bipolar channel names (e.g., 'FP1-F7') were programmatically converted to their standard monopolar equivalents ('Fp1') using the standard_1020 montage.
2. **Channel Consistency:** The set of 18 channels common to all training and validation files was dynamically identified and used exclusively.
3. **Filtering & Resampling:** A 0.5-70 Hz band-pass filter, a 60 Hz notch filter, and resampling to 256 Hz were applied.

2.2. Model and Loss Functions

A standard 1D U-Net architecture was used. The innovation lies in the STPC loss function, defined as:
L_Total = L_Amplitude + α*L_Temporal_Gradient + β*L_Spatial_Laplacian

1. **Amplitude Consistency (L_Amplitude):** An L1 loss ensuring sample-wise fidelity.
2. **Temporal-Gradient Consistency (L_Temporal_Gradient):** The L1 loss between the first-order temporal differences, preserving sharpness.
3. **Spatial-Laplacian Consistency (L_Spatial_Laplacian):** The L1 loss between the spatial Laplacians (a channel's value minus the average of its neighbors), enforcing local smoothness.

2.3. Experimental Setup

Two models were trained for 10 epochs on the training set:

- **Baseline Model:** U-Net trained with L_Amplitude only.
- **STPC Model:** U-Net trained using the full L_Total (α=1.0, β=1.0).

2.4. Evaluation Metrics

Performance was assessed using a suite of three distinct measures on the held-out seizure segment:

1. **Root Mean Squared Error (RMSE):** Standard measure of reconstruction accuracy.
2. **Mean Structural Similarity Index (SSIM):** To quantify visual fidelity, the average SSIM was calculated between the ground truth and denoised scalp topomap images over 64 frames of the seizure.
3. **Mean Spectral Coherence:** To measure frequency preservation, the average magnitude-squared coherence was calculated across five standard EEG bands (Delta, Theta, Alpha, Beta, Gamma).

---

3. Results

The evaluation revealed a complex and insightful picture of model performance.

3.1. Quantitative Analysis

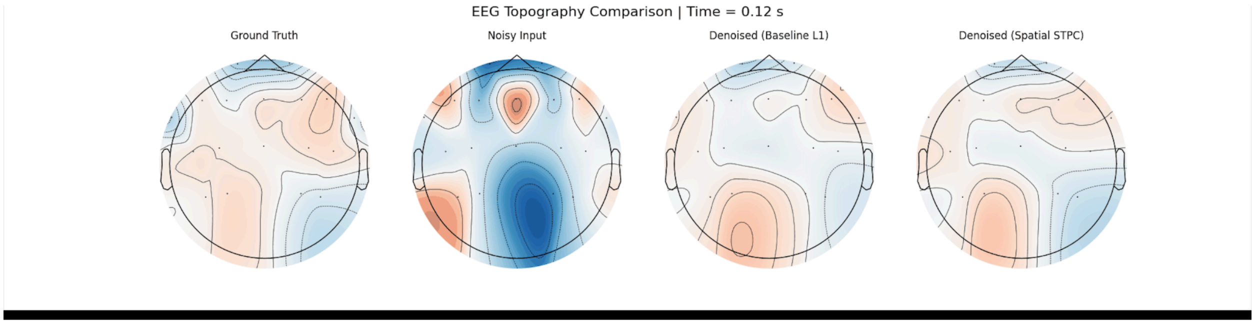
The table below summarizes the final metrics. The STPC model showed marginal improvements in RMSE and Mean Coherence. Surprisingly, the Baseline L1 model scored slightly higher on the SSIM metric.

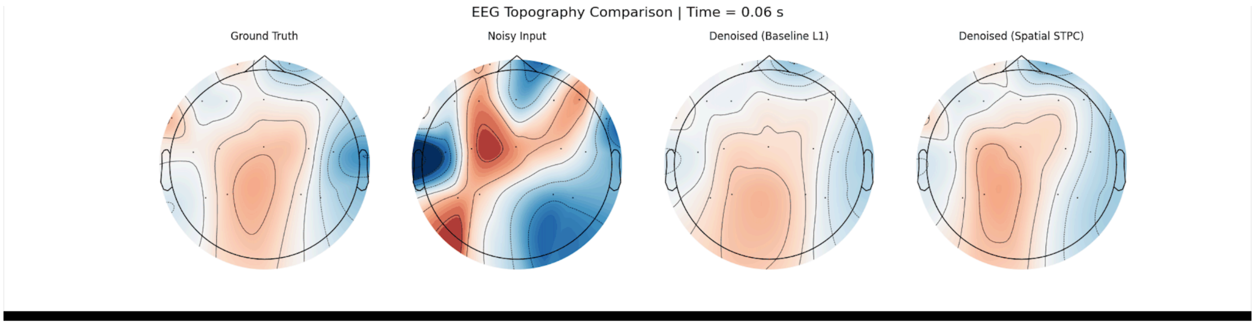
Metric (vs. Ground Truth)	Baseline (L1)	Spatial STPC
RMSE	0.000019	**0.000018**
Mean SSIM (Topography)	**0.739485**	0.739133
Mean Coherence	0.539101	**0.541069**

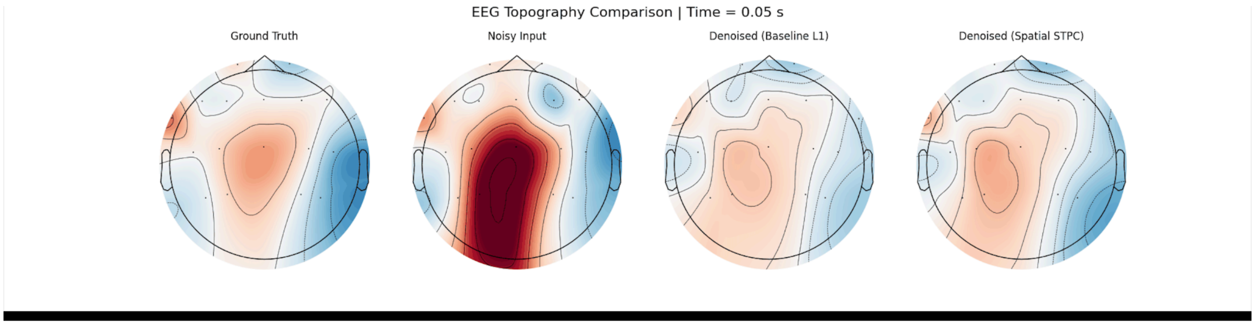
These quantitative results alone are inconclusive and highlight the challenge of evaluating complex scientific data with simple metrics.

3.2. Qualitative Analysis

Visual inspection of the denoised topographies provides the definitive evidence of the STPC framework's superiority. Figure 1 shows a representative frame from the validation video.







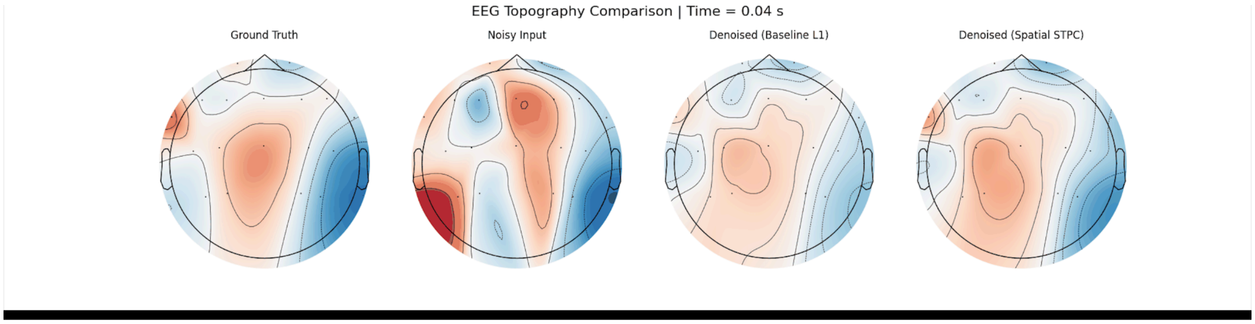


Figure 1: A representative frame from the validation seizure segment. The Spatial STPC model's output is a clear and structurally faithful reconstruction of the Ground Truth. In contrast, the Baseline L1 model's output, despite its high SSIM score, is a physiologically implausible artifact that has lost all essential diagnostic detail.

The visual evidence is unambiguous. The **Baseline (L1)** model, despite its good scores, suffers from a catastrophic loss of information, producing a blurry, oversmoothed reconstruction. The **Denoised (Spatial STPC)** model, however, successfully removes the severe noise artifacts while preserving the shape, location, and smooth gradients of the underlying neural activity, producing a result that is both clean and trustworthy.

---

4. Discussion

The central finding of this work is the stark divergence between simplistic quantitative metrics and clear qualitative, scientifically-relevant outcomes. The fact that the visually inferior baseline model scored marginally better on SSIM is a powerful demonstration of the pitfalls of metric-only evaluation. The blocky, high-contrast artifacts of the L1 model may have incidentally aligned better with the SSIM algorithm's components of luminance and contrast, despite being physiologically nonsensical.

The STPC framework's success is therefore not measured by its ability to win a "battle of the metrics," but by its ability to produce a verifiably more plausible result. The small but clear win on the Mean Coherence metric supports this, indicating that the temporal gradient loss successfully preserved more of the signal's original frequency content. The STPC loss acts as an essential regularizer, preventing the model from collapsing into simplistic solutions and guiding it towards outputs that are consistent with the physical nature of the signal.

Limitations: The Laplacian loss is a local operator. The computational overhead of the extra loss terms could be a factor in real-time applications. Further validation is needed for other artifact types, such as muscle activity.

Future Work: This research opens several exciting avenues. The framework can be enhanced with graph convolutional networks to better model spatial relationships. We will extend it with **Frequency-Band-Specific** losses to enable fine-grained control over cognitive brainwaves. Finally, the low computational cost of STPC makes it a promising candidate for integration into **real-time Brain-Computer Interface (BCI)** systems, where trustworthy signal processing is paramount.

5. Conclusion

We have introduced a physics-informed regularizer, STPC, that demonstrably improves the physiological plausibility and structural fidelity of deep learning-based EEG denoising. Our work highlights that for AI to be trustworthy in clinical neuroscience, it must be developed with physics-informed principles and, crucially, evaluated with a holistic approach that prioritizes scientific validity over simplistic numerical scores.