SPATIAL CONTRASTIVE PRE-TRAINING OF TRANSFORMER ENCODERS FOR SEEG-BASED SEIZURE ONSET ZONE DETECTION

Zacharie Rodière[1], Pierre Borgnat[1], Paulo Gonçalves[2], Julien Jung[3]

[1] CNRS, ENS de Lyon, LPENSL, UMR5672, Lyon France
[2] Inria, ENS de Lyon, CNRS, UCB Lyon 1, LIP, UMR 5668, Lyon, France
[3] Hôpital Neurologique, HCL ; CRNL, Lyon 1, Inserm UMRS1028, CNRS UMR5292, UJM

ABSTRACT

For the clinical study of epilepsy, we develop a transformer encoder for the detection of Seizure Onset Zone (SOZ) from stereo-EEG. It integrates clinically grounded time-frequency features with spatial contrastive pre-training. While prior spatial transformer approaches analyze learned representations, our method uniquely combines: (1) engineered time-frequency representations (TFRs) encoding epileptic spikes and oscillations, and (2) a contrastive objective leveraging anatomical relationships between the electrode contacts that are in the SOZ and the ones outside the SOZ.

The model processes heterogeneous sEEG records from different patients, using both ictal and interictal data. This contrastive strategy minimizes representational similarity between contact pairs on either side of the SOZ boundary while maximizing intra-SOZ similarity.

Attention heads provide interpretable connectivity patterns, bridging data-driven learning with the study of functional connectivity networks.

Initial experiments demonstrate feasibility, with preliminary evidence of improved generalization between patients. Although full validation is still ongoing, this communication will highlight how domain-informed TFRs combined with contrastive spatial learning advance SOZ detection, and pave the way toward anatomically grounded and data-efficient tools for epilepsy surgery planning.

1. INTRODUCTION

Epilepsy is a prevalent neurological disorder [1]. For drug-resistant focal epilepsy, neurosurgical intervention to resect the epileptogenic zone (EZ) can be the only curative treatment. One aspect of the preparatory localization of the Seizure Onset Zone (SOZ) is to analyze stereo-EEG signals, which are intracranial measures at electrodes (typically between 8 and 12, each with 10 to 12 contact points). These sEEG signals, recorded over a long period of time, help to identify the SOZ considering the connectivity between them [2, 3], the apparition of spikes [4, 5] and their propagation in the epileptogenic network, and/or the existence of high frequency oscillations (HFOs) [6, 7]. They all serve as markers of epilepsy and of the SOZ [5, 8]. In the recent past, we have looked at a graph signal processing approach to find how the connectivity between signals is temporally organized during seizures [9, 10]. However, one needs to be able to consider features related to spikes, ripples, and HFOs – which are best apparent in time-frequency representations of the signals [11, 12] – jointly with the connectivity between the various zones in the brain.

Work supported by the SEIZURE ANR project, ANR-24-CE45-4399.

The approach developed here is a Deep Learning (DL) method for the detection of the SOZ, by coding all these aspects in a transformer that captures connections between contacts by the attention mechanism [13], and which is applied to time-frequency features. The work has common elements with [14] but we develop significant variations by proposing changes to the architecture, and by designing an approach with contrastive pre-training, as it is known to be useful for classification of brain signals [15], and proposing a spatial loss function for this contrastive pre-training. This model has the advantage that we can considerer both ictal (during seizures) and inter-ictal signals, and that we can train it on a heterogeneous group of patient.

2. ELEMENTS OF THE MODEL

2.1. Time-frequency features

The sEEG signals at each electrode are processed by a Daubechies-4 (Db4) Wavelet Packet Transform (WPT) [16]. We use 5 levels of decomposition and this provides localized time-frequency features, useful for detecting spikes, ripples, and HFOs [11, 12].

The compact support of the Db4 wavelet allows sharp temporal resolution for transient spikes (i.e, sudden high-amplitude discharges), while its regularity suppresses smoother background activity [12]. This granularity disentangles overlapping spectral components, allowing DL models to identify ripples as sustained energy in specific sub-bands, and spikes as transient bursts in adjacent bands.

2.2. Model Architecture

Denoting d_{features}, the length of flattened WPT features and C the number of channels for a given patient, time-frequency features $[\boldsymbol{f_1}, \boldsymbol{f_2}, \ldots, \boldsymbol{f_C}] \in \mathbb{R}^{d_{\text{features}} \times C}$ are first projected through a linear layer into a token space of fixed dimension d_{model}, yielding the sequence $[\boldsymbol{t_1}, \boldsymbol{t_2}, \ldots, \boldsymbol{t_C}] \in \mathbb{R}^{d_{\text{model}} \times C}$. This sequence serves as the input to a transformer encoder [13], which produces learned representations $[\boldsymbol{h_1}, \boldsymbol{h_2}, \ldots, \boldsymbol{h_C}] \in \mathbb{R}^{d_{\text{model}} \times C}$, see Fig. 1. These representations are employed, along with the SOZ labels, in the proposed spatial contrastive pre-training task.

In contrast to [14] which adds a [CLS] token to aggregate token-level information, before processing it with an MLP for SOZ probability estimation, we deliberately omit the [CLS] token and rely on spatial contrastive pretraining to ensure that the learned representation of each channel is inherently informative with respect to the SOZ. Finally, to obtain channel-wise probability estimates, we project the encoder outputs through a linear layer and apply a sigmoid activation, producing scores between 0 and 1.

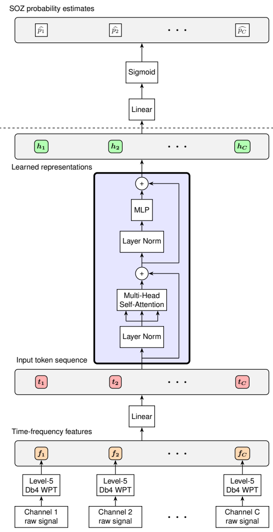

Fig. 1. Model architecture

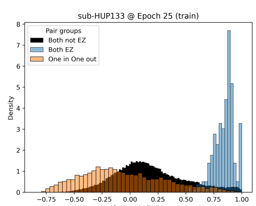

Fig. 2. Similarity distribution of a patient after 25 epochs of training

2.3. Contrastive pre-training

The core of our methodology is a spatial contrastive loss, which differentiates between latent representations based on their spatial origins. Minimizing the loss will minimize the similarity for channel pairs where one signal is in the SOZ, and the other is not, and conversely maximizes the similarity for channel pairs that both originate from the SOZ. The proposed loss reads as follows:

$$L = \frac{1}{N_1} \sum_{\substack{(i,j)\in[\![1,C]\!]^2 \\ i\neq j \\ g_i=g_j=1}} \max(0, m_+ - \mathrm{sim}_{ij}) + \frac{1}{N_2} \sum_{\substack{(i,j)\in[\![1,C]\!]^2 \\ i\neq j \\ g_i\neq g_j}} \max(0, \mathrm{sim}_{ij} - m_-)$$

where,

$$\begin{cases} g_i & \text{is 1 if channel } i \in \text{SOZ, 0 otherwise} \\ m_+ \lesssim 1 & \text{is the positive margin} \\ m_- \gtrsim -1 & \text{is the negative margin} \\ \mathrm{sim}_{ij} = \frac{\boldsymbol{h_i}^\top \boldsymbol{h_j}}{\|\mathbf{h_i}\|\|\mathbf{h_j}\|} & \text{is the cosine similarity between } \boldsymbol{h_i} \text{ and } \boldsymbol{h_j} \\ N_1 = \#\{(i,j) | g_i = g_j = 1\} & \\ N_2 = \#\{(i,j) | g_i \neq g_j\}. & \end{cases}$$

3. NUMERICAL EXPERIMENTS

3.1. Dataset & Training

The model is trained on the HUP iEEG epilepsy dataset [17], using SOZ labels for ictal and interictal recordings. A custom dataset class was implemented in PyTorch [18] to handle multiple patient recordings, output labels, channel names, and patient IDs. This class enables global indexing of all recordings and metadata, providing a unified interface that facilitates training across the entire dataset.

We use one-minute sEEG clips resampled at 500Hz (the HUP dataset recordings are sampled at 500Hz, 512Hz, or 1024Hz), resulting in 30,000 time samples per clip, which reduce to 29,920 after applying a level-5 Wavelet Packet Transform. The transformer encoder consists of 6 layers, each with 8 attention heads and a model dimension of $d_{\text{model}} = 512$, with 30% dropout applied after each multi-head attention and feedforward block. Training is performed over 100 epochs with a batch size of 16, using Adam with a linear warmup cosine annealing learning rate scheduler that ramps from 10^{-6} to 2.10^{-5} over 10 epochs and then decays back to 10^{-6}. The model is trained on an Nvidia Titan RTX using the ENSL server resources at CBP operated by SIDUS[19].

Fig. 2 shows an example of the distributions of cosine similarities of representations for a patient (in the training set) after pre-training. The success of the contrastive training can be seen in the differences between the distributions, according to the position of the contacts w.r.t. the SOZ.

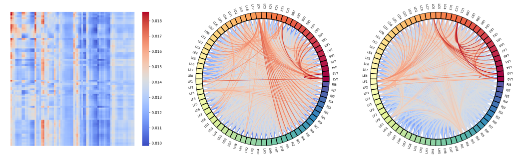

Fig. 3. Attention map seen as a directed graph. This is the first head of the first layer of our pre-trained model on a one minute clip. The center graph shows the upper triangle of the attention matrix, while the rightmost graph shows the lower triangle. Attention flows from queries (line indices) to keys (column indices).

3.2. Supervised fine-tuning

After pre-training, we keep the weights from the first linear layer and the transformer encoder, then we introduce an additional linear layer to project the learned representations to a single value per channel. Sigmoid activation is applied to produce SOZ scores ranging from 0 to 1. To fine-tune the model, we employ a binary cross-entropy loss. While our primary focus has been on pre-training, the fine-tuning stage is still in progress.

3.3. Attention maps seen as directed connectivity graphs

Functional connectivity networks have been widely used by neurologists to understand dependencies between brain regions[2, 3, 4, 8]. Here, we interpret attention maps from the transformer-based model as directed connectivity graphs, drawing a parallel to usual representations in neuroscience. Specifically, the attention maps $A = \mathrm{softmax}\left(\frac{QK^\top}{\sqrt{d_K}}\right)$ of the multi-head attention blocks are viewed as digraphs for each input clip. Given that each input produces multiple such graphs –spanning layers and attention heads– we explore their structure and significance. As an example, we visualize in Fig. 3 the first attention map from the initial layer for a randomly selected clip, highlighting how the model inherently produces graph-like representations. This perspective strengthens the connection between transformer-based models and established methods in brain signal analysis. In future work, the relation between them and the TF features, as signals on these graphs, will be studied.

4. REFERENCES

[1] Ali A et al. Asadi-Pooya, "Adult epilepsy," *The Lancet*, vol. 402, no. 10399, pp. 412 – 424, 2023.

[2] Fabrice Bartolomei, Stanislas Lagarde, Fabrice Wendling, Aileen McGonigal, Viktor Jirsa, Maxime Guye, and Christian Bénar, "Defining epileptogenic networks: Contribution of seeg and signal analysis," *Epilepsia*, vol. 58, no. 7, pp. 1131–1147, 2017.

[3] Pieter van Mierlo, Yvonne Höller, Niels K. Focke, and Serge Vulliemoz, "Network perspectives on epilepsy using eeg/meg source connectivity," *Frontiers in Neurology*, vol. 10, 2019.

[4] Margherita A G Matarrese, Alessandro Loppini, Lorenzo Fabbri, Eleonora Tamilia, M Scott Perry, Joseph R Madsen, Jeffrey Bolton, Scellig S D Stone, Phillip L Pearl, Simonetta Filippi, and Christos Papadelis, "Spike propagation mapping reveals effective connectivity and predicts surgical outcome in epilepsy," *Brain*, vol. 146, no. 9, pp. 3898–3912, 04 2023.

[5] Roehri N, Pizzo F, Lagarde S, Lambert I, Nica A, McGonigal A, Giusiano B, Bartolomei F, and Bénar CG, "High-frequency oscillations are not better biomarkers of epileptogenic tissues than spikes," *Ann Neurol*, vol. 83, no. 1, pp. 84–97, Jan. 2018.

[6] Maeike Zijlmans, Premysl Jiruska, Rina Zelmann, Frans S.S. Leijten, John G.R. Jefferys, and Jean Gotman, "High-frequency oscillations as a new biomarker in epilepsy," vol. 71, no. 2, pp. 169–178.

[7] Nicolas Roehri and Fabrice Bartolomei, "Are high-frequency oscillations better biomarkers of the epileptogenic zone than spikes?," *Current Opinion in Neurology*, vol. 32, no. 2, pp. 213–219, Apr. 2019.

[8] Azeem A, von Ellenrieder N, Hall J, Dubeau F, Frauscher B, and Gotman J, "Interictal spike networks predict surgical outcome in patients with drug-resistant focal epilepsy," *Ann Clin Transl Neurol*, vol. 8, no. 6, pp. 1212–1223, June 2021.

[9] Gaëtan Frusque, Julien Jung, Pierre Borgnat, and Paulo Gonçalves, "Multiplex network inference with sparse tensor decomposition for functional connectivity," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, pp. 316–328, 2020.

[10] Gaëtan Frusque, Pierre Borgnat, Paulo Gonçalves, and Julien Jung, "Semi-automatic extraction of functional dynamic networks describing patient's epileptic seizures," *Frontiers in Neurology*, vol. 11, 2020.

[11] Nicolas Roehri, Jean-Marc Lina, John C. Mosher, Fabrice Bartolomei, and Christian-George Bénar, "Time-frequency strategies for increasing high-frequency oscillation detectability in intracerebral eeg," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 12, pp. 2595–2606, 2016.

[12] Olesya Grinenko, Jian Li, John C Mosher, Irene Z Wang, Juan C Bulacio, Jorge Gonzalez-Martinez, Dileep Nair, Imad Najm, Richard M Leahy, and Patrick Chauvel, "A fingerprint of the epileptogenic zone in human epilepsies," *Brain*, vol. 141, no. 1, pp. 117–131, 12 2017.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. 2017, vol. 30, Curran Associates, Inc.

[14] Jamie Norris, Aswin Chari, Dorien van Blooijs, Gerald Cooray, Karl Friston, Martin Tisdall, and Richard Rosch, "Localising the seizure onset zone from single-pulse electrical stimulation responses with a cnn transformer," ArXiv cs.LG preprint: 2403.20324, Mar. 2024.

[15] Mostafa Neo Mohsenvand, Mohammad Rasool Izadi, and Pattie Maes, "Contrastive representation learning for electroencephalogram classification," in *Proceedings of the Machine Learning for Health NeurIPS Workshop*, Emily Alsentzer, Matthew B. A. McDermott, Fabian Falck, Suproteem K. Sarkar, Subhrajit Roy, and Stephanie L. Hyland, Eds. 11 Dec 2020, vol. 136 of *Proceedings of Machine Learning Research*, pp. 238–253, PMLR.

[16] Ingrid Daubechies, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, 1992.

[17] John M. Bernabei, Adam Li, Andrew Y. Revell, Rachel J. Smith, Kristin M. Gunnarsdottir, Ian Z. Ong, Kathryn A. Davis, Nishant Sinha, Sridevi Sarma, and Brian Litt, ""hup ieeg epilepsy dataset"," 2023.

[18] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, Eds. 2019, vol. 32, Curran Associates, Inc.

[19] Emmanuel Quemener and Marianne Corvellec, "Sidus—the solution for extreme deduplication of an operating system," *Linux Journal*, vol. 2013, no. 235, pp. 3, 2013.