Spectrum-Aware Masking for EEG Signal Pretraining

Zihang Wang[1], Zheng Chen[2], Peng Chen[3], Mohamed Wahib[3], Youhua Shi[1]

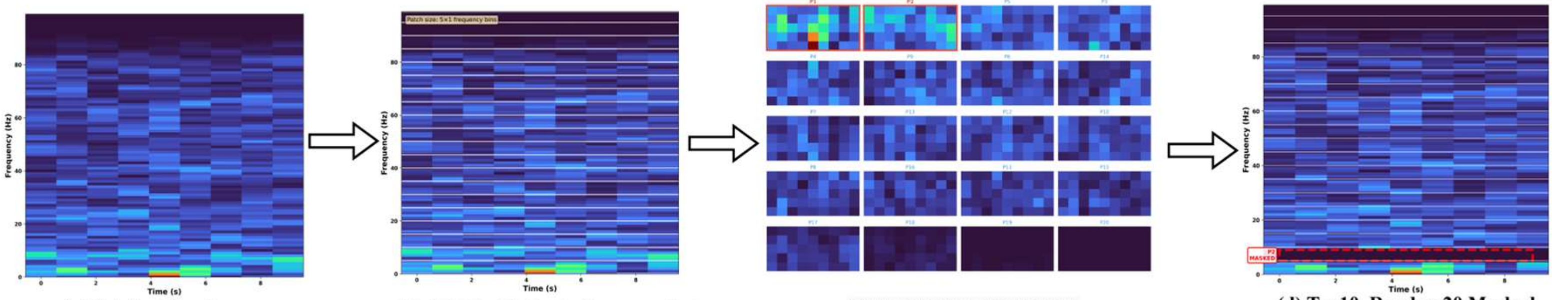
1. Waseda University, Japan, {wang.zihang@akane.waseda.jp} { shi@waseda.jp}
2. SANKEN, The University of Osaka Japan {chenz@sanken.osaka-u.ac.jp }
3. RIKEN Center for Computational Science, Japan, {peng.chen@a.riken.jp, mohamed.attia@riken.jp }

① Introduction:

Electroencephalography (EEG) enables analysis of brain activity for tasks such as seizure detection and sleep staging. However, noise, inter-subject variability, and complex frequency characteristics pose challenges for deep learning models. Existing methods often rely on random masking or fixed pre-processing, neglecting the importance of frequency regions in representation learning [1]. We propose a frequency-aware masking framework that ranks spectrogram patches by intensity and selectively masks mid-frequency regions, leading to improved robustness and generalization. This improvement comes at the cost of increased computational overhead, resulting in longer training time and higher memory usage.

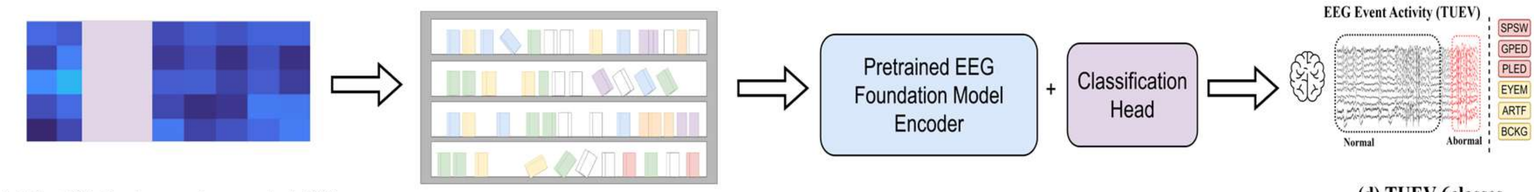
② Method:

We propose a **frequency-aware pretraining framework** to analyze the role of different spectral regions in EEG representation learning. Raw EEG signals are converted into STFT-based spectrograms and partitioned into frequency-oriented patches, which are ranked by spectral intensity to reflect their relative importance. Based on this ranking, multiple frequency-aware masking strategies are designed by selectively masking patches from different intensity ranges during pretraining. As a representative example, 20% of patches are randomly masked from the top 10% highest-intensity regions. The encoder is pretrained [2] in a self-supervised manner and fine-tuned for downstream EEG event classification on the six-class TUEV dataset[3].



(a) Original Spectrogram (b) STFT with Patch Segmentation (c) Patch Intensity Ranked (d) Top10_Random20 Masked

Fig.1. Overview of the proposed pretraining framework



(a) Top10% Patches random masked 20% (b) Token Vocabulary (c) Pretrained Foundation Model (d) TUEV 6classes

Fig.2. Randomly masking 20% of the Top10% intensity patches.

③ Results and Discussion:

Table 1. Accuracy evaluation on the TUEV dataset

Masking Strategy	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Test Acc
EEG Foundation Model	0.0874	0.5314	0.5274	0.6475	0.6633	0.7976	0.7116
Top10% Patches–random 20%	0	**0.5942**	0.2955	0	0.6141	**0.9102**	**0.7527**
Middle 30–60% Patches	0	**0.5960**	0.3473	0	**0.7226**	**0.8895**	**0.7552**
Middle 40–70% Patches	0.0699	0.3821	**0.5990**	0.4800	0.6531	**0.8959**	**0.7516**
Bottom 30% Patches	0.0499	0.4182	0.2627	0.5886	0.6725	**0.8605**	**0.7158**

Table 2. Computational cost comparison

Masking Strategy	GPUs	Batch Size	Epochs	Training Time
Random Masking (50%)	8×A100	128	100	~4.50 h
Top10% Random20%	8×A100	128	100	~5.10 h
Middle 30–60%	8×A100	128	100	~5.30 h
Middle 40–70%	8×A100	128	100	~5.36 h
Bottom 30%	8×A100	128	100	~4.90 h

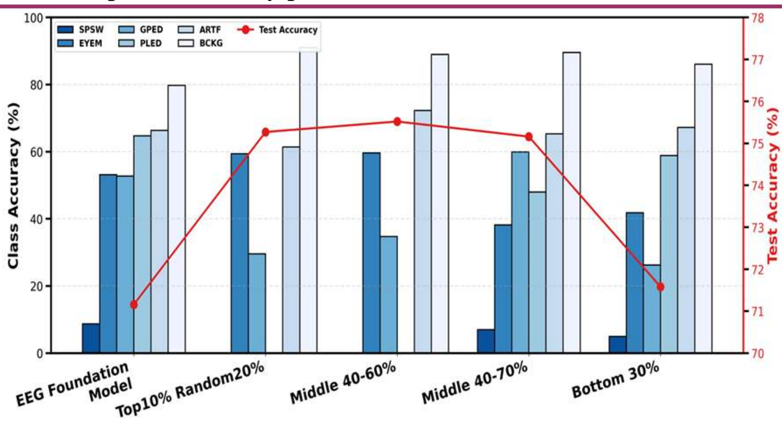


Fig.3. Performance comparison across classes and masking strategies.

- Frequency-domain modeling improves performance with increased computational cost.
- Effective frequency learning is reflected by sensitivity to spectral masking.
- Current EEG foundation models are limited by frequency feature extraction rather than model scale.

References

[1] W. Ma, Y. Zheng, T. Li, et al. A comprehensive review of deep learning in eeg based emotionrecognition: classifications, trends, and practical implications. PeerJ Computer Science, 10:e2065, 2024.
[2] J. Pradeepkumar et al. Tokenizing single-channel eeg with time-frequency motif learning. In NeurIPS 2025 Workshop on Learning from Time Series for Health, 2025.
[3] J. Obeid and J. Picone. Temple university hospital eeg corpus: Tuh eeg events (tuev), 2015. Version 2.0.1, Accessed: 2025-10-12.