DECODING BRAINS WITH HFMCA: LEARNING ORTHONORMAL FEATURES FOR ROBUST EEG REPRESENTATIONS

1. SUPPLEMENTARY MATERIAL: DATA AUGMENTATION STRATEGIES FOR EEG SIGNAL PROCESSING

1.1. Data Augmentation Framework

To learn robust representations in a self-supervised manner, we employ a diverse set of data augmentation techniques specifically designed for EEG signals. These augmentations serve to create multiple views of the same input while preserving the underlying neural information. Given an input EEG segment $\mathbf{X} \in R^{C \times T}$, where C represents the number of channels and T represents the time points, we apply the following transformations:

1.1.1. Spatial Augmentations

Channel Permutation. This method randomly permutes the order of EEG channels, disrupting spatial relationships while preserving temporal dynamics. This forces the model to learn features invariant to the absolute spatial positioning of electrodes, focusing instead on their functional signals.

Mathematical formulation:

$$\mathbf{X}_{\text{perm}} = \mathbf{P} \cdot \mathbf{X} \tag{1}$$

where $\mathbf{P} \in \{0,1\}^{C \times C}$ is a random permutation matrix satisfying:

$$\sum_{i=1}^{C} P_{ij} = 1, \quad \forall j \in \{1, \ldots, C\} \tag{2}$$

$$\sum_{j=1}^{C} P_{ij} = 1, \quad \forall i \in \{1, \ldots, C\} \tag{3}$$

Channel Dropout. Entire channels are randomly selected and set to zero, simulating the common real-world issue of noisy or disconnected electrodes. This compels the model to rely on distributed information from the remaining channels.

Mathematical formulation:

$$\mathbf{X}_{\text{dropout}}[c,t] = \mathbf{X}[c,t] \cdot m_c \tag{4}$$

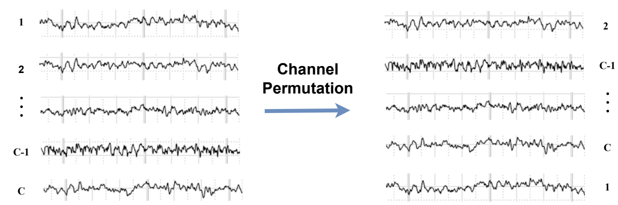

Fig. 1. Channel Permutation

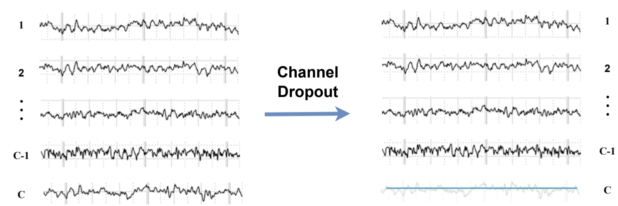

Fig. 2. Channel dropout. where the blue line indicates values set to zero.

where $\mathbf{m} \in \{0,1\}^C$ is a binary mask vector with:

$$m_c \sim \text{Bernoulli}(1 - p_{\text{drop}}) \tag{5}$$

$$\sum_{c=1}^{C} m_c \geq 1 \tag{6}$$

where p_{drop} is the dropout probability, and the constraint ensures at least one channel is preserved.

1.1.2. Temporal Augmentations

Temporal Masking. Random contiguous portions of the time series are masked (set to zero) across all channels simultaneously. This technique encourages the model to develop a robust understanding of temporal contexts and prevents overreliance on any single salient feature.

Mathematical formulation:

$$\mathbf{X}_{\text{masked}}[c,t] = \mathbf{X}[c,t] \cdot w_t \tag{7}$$

where $\mathbf{w} \in \{0,1\}^T$ is a binary window function:

$$w_t = \begin{cases} 0, & \text{if } t \in [t_{\text{start}}, t_{\text{end}}] \\ 1, & \text{otherwise} \end{cases} \tag{8}$$

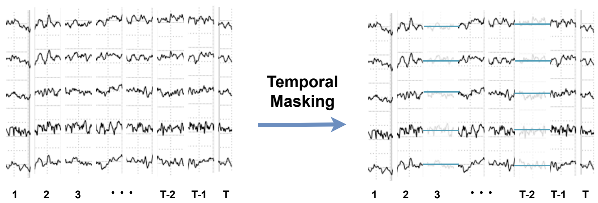

Fig. 3. Temporal Masking, where the blue line indicates values set to zero.

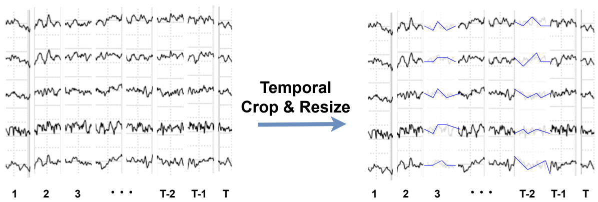

Fig. 4. Temporal Crop and Resize, where the blue line shows interpolation.

with $t_{\text{start}} \sim \text{Uniform}(0, T - L_{\text{mask}})$ and $t_{\text{end}} = t_{\text{start}} + L_{\text{mask}}$, where L_{mask} is the mask length. **Temporal Crop-and-Resize.** A segment of the time series is randomly cropped and then resized back to the original length using interpolation. This augmentation combines warping and cropping, introducing variations in temporal scale and perspective.

Mathematical formulation:

$$\mathbf{X}_{\text{crop}}[c, t] = \text{Interp}\left(\mathbf{X}[c, t_{\text{crop}} : t_{\text{crop}} + L_{\text{crop}}], T\right) \tag{9}$$

where:

$$t_{\text{crop}} \sim \text{Uniform}(0, T - L_{\text{crop}}) \tag{10}$$
$$L_{\text{crop}} \sim \text{Uniform}(\alpha T, \beta T) \tag{11}$$

with $0 < \alpha < \beta < 1$ being hyperparameters controlling the crop ratio range, and $\text{Interp}(\cdot, T)$ denotes an interpolation function (e.g., linear or cubic spline) that resizes the cropped segment back to length T.

1.1.3. Implementation Details

Table 1. Data augmentation hyperparameters.

Augmentation	**Parameters**
Channel Perm.	$p = 0.5$
Channel Drop.	$p = 0.2$, max $= 0.5C$
Temporal Mask	$L \in [0.1T, 0.3T]$, $p = 0.5$
Temp. Crop-Resize	$\alpha = 0.6$, $\beta = 0.95$, $p = 0.5$

These augmentation strategies are designed to:

- Enhance model robustness to electrode placement variations and signal artifacts
- Prevent overfitting to specific spatial or temporal patterns
- Simulate real-world EEG recording conditions (e.g., electrode failures, motion artifacts)
- Create diverse yet semantically consistent views for contrastive learning

The augmentations preserve the essential neural information while introducing controlled variations that improve the model's ability to learn generalizable representations from EEG signals.