EEGiT: Teaching Vision Transformers to Understand the EEG signal

Supplementary Material

A. Brain regionalization

based To construct EEG patches, we perform linear interpolation of EEG channels based to their respective brain regions. Due to differences in data acquisition systems between the THINGS-EEG and EEG-3D datasets, the spatial organization of brain regions differs slightly across the two datasets. The specific brain region divisions are as follows:

THINGS-EEG dataset:

Frontal: Fp1, Fp2, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, F2, F4, F6,F8.

Central: FC5, FC3, FC1, FCz, FC2, FC4, FC6, C5, C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CPz, CP2, CP4, CP6.

Temporal: FT9, FT7, FT8, FT10, T7, T8, TP9, TP7, TP8, TP10.

Parietal: P7, P5, P3, P1, P2, P4, P6, P8, Pz.

Occipital: PO7, PO3, POz, PO4, PO8, O1, Oz, O2.

EEG-3D dataset:

Frontal: FP1, FPZ, FP2, AF3, AF4, F7, F5, F3, F1, FZ, F2, F4, F6, F8.

Central: FC5, FC3, FC1, FCZ, FC2, FC4, FC6, C5, C3, C1, CZ, C2, C4, C6.

Temporal: FT7, FT8, T7, T8, TP7, TP8, M1, M2.

Parietal: CP5, CP3, CP1, CPZ, CP2, CP4, CP6, P7, P5, P3, P1, PZ, P2, P4, P6, P8.

Occipital: PO7, PO5, PO3, POZ, PO4, PO6, PO8, O1, OZ, O2, CB1, CB2.

Notably, in the THINGS-EEG dataset, the CP channels are categorized as central, whereas in the EEG-3D dataset they are assigned to the parietal region. This difference arises because the CP channels represent a transitional area between the central and parietal regions. To avoid an imbalance where the central region contains too many channels and the parietal region too few, which could introduce interpolation errors, we made this adjustment. The M1 and M2 channels in the EEG-3D dataset correspond to mastoid electrodes rather than cortical EEG channels. Following common practice in neuroscience for computational and visualization convenience, we classify them as part of the temporal region. The CB channel corresponds to the cerebellar region, which does not belong to the standard cortical areas. Since its spatial location is closer to the occipital region, we assign it to the occipital region.

B. More experimental results

Table 1 presents the results of five experimental runs of EEGiT on the THINGS-EEG dataset across different subjects and experimental settings. To evaluate the model's robustness and ensure the completeness of the experimental results, we also computed the corresponding standard deviations. As shown in Table 4, EEGiT demonstrates consistent performance across multiple random initializations, indicating stable behavior and further validating the effectiveness and reliability of our approach.

Table 4. Standard deviations of EEGiT retrieval accuracy across subjects under inter-subject and intra-subject settings on the THINGS-EEG dataset.

Method	S1		S2		S3		S4		S5	
	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5
Inter	3.6	5.1	3.3	4.3	1.0	3.5	1.8	2.8	2.2	3.9
Intra	4.5	1.3	3.0	1.9	3.5	2.7	4.5	3.5	4.8	3.5
Method	S6		S7		S8		S9		S10	
	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5
Inter	1.5	2.1	2.0	2.4	2.2	1.9	3.2	4.7	2.4	4.0
Intra	3.2	1.7	3.0	1.2	2.1	0.8	3.1	1.9	2.9	0.4
Inter: 0.9 (top-1), 1.1 (top-5)						**Intra: 1.3 (top-1), 0.8 (top-5)**				

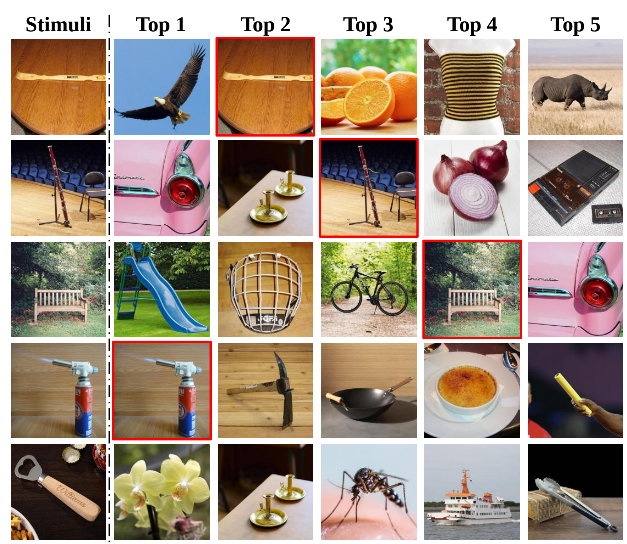

Figure 9. Qualitative examples of top-5 EEG-to-image retrieval results under the inter-subject setting.

Fig. 4 presents qualitative retrieval results of EEGiT under the intra-subject setting. We further visualize retrieval results in the inter-subject setting to assess cross-subject

Method	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Subject 8		Subject 9		Subject 10		Avg	
	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5
Occipital	71.3	95.9	66.3	94.4	61.0	91.0	48.4	86.2	59.2	87.1	71.5	91.5	72.7	75.0	86.3	95.9	56.4	83.7	73.9	98.2	66.7	89.9
Early-stage	71.8	96.6	56.8	91.3	58.1	90.6	54.3	85.0	51.5	82.7	71.9	94.7	62.5	93.2	85.5	96.4	59.7	85.4	76.4	97.3	64.9	91.3
Occ&Early	69.8	91.0	57.1	88.9	52.6	88.3	44.5	72.3	47.1	84.3	65.9	91.3	68.5	92.1	72.1	92.3	50.3	83.6	72.4	95.3	60.0	87.9
EEGiT	**82.3**	**97.6**	**73.7**	**97.9**	**67.5**	**95.7**	**64.9**	**93.5**	**55.9**	**91.3**	**66.8**	**92.0**	**70.2**	**98.1**	**84.8**	**98.6**	**51.7**	**87.3**	**86.5**	**99.4**	**70.4**	**95.1**

Table 5. Model performance on individual subjects from the THINGS-EEG dataset under the intra-subject setting using different EEG signal configurations. "Occ&Early" denotes the use of early-stage EEG signals from the occipital region.

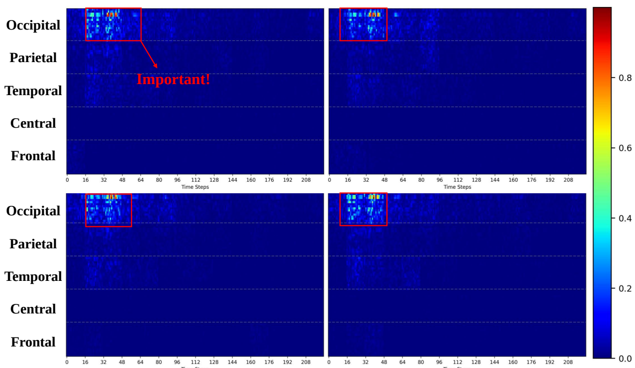

Figure 10. Spatiotemporal activation patterns of EEG signals across brain regions under the inter-subject setting.

generalization. Due to the substantial variability across subjects, the inter-subject scenario remains highly challenging. As shown in Fig. 9, EEGiT exhibits noticeably reduced retrieval accuracy in this setting and struggles to extract stable semantic information from EEG signals. These observations highlight the need for further improvements in cross-subject robustness.

C. Further Spatial and Temporal Analysis

In Section 4.5, we investigate how different brain regions and temporal segments of EEG signals influence model performance in the EEG-to-image retrieval task. Our results underscore the critical role of the occipital region and early-stage neural responses. Fig. 10 further visualizes the spatiotemporal activation patterns across brain regions under the inter-subject setting, revealing trends that are largely consistent with those observed in the intra-subject scenario. As shown in Table 5, we report model performance across subjects under three conditions: (i) using only occipital-region signals, (ii) using only early-stage signals, and (iii) using only early-stage signals from the occipital region. Notably, when restricted to early occipital signals, the model maintains competitive performance while substantially reducing computational cost. These findings offer practical insights for future EEG data acquisition protocols and the design of efficient decoding models.