Decoding 3D Perception via BrainSSD: Synergistic Fusion of EEG Representations from Static and Dynamic Visual Streams

Supplementary Material

1. Experimental details

1.1. Dataset details

EEG-3D dataset. The experimental evaluation is conducted on the EEG-3D dataset. This is the first benchmark dataset specifically designed for decoding 3D visual perception from EEG signals, providing paired EEG recordings and 3D stimulus data. The dataset features high-density EEG recordings from 12 healthy subjects viewing 72 distinct object categories derived from the Objaverse dataset, which are also categorized into six groups based on their dominant color styles. To capture the full spectrum of 3D perception, the visual stimuli for each object are presented in two modalities: a 0.5-second static image rendered from an optimal elevation angle, followed by a 6-second video clip exhibiting a 360-degree azimuthal rotation at 30 Hz. The dataset employs an event-related design where training objects are presented with 2 repetitions and testing objects with 4 repetitions to ensure robust signal quality.

EEG signals were originally acquired using a 64-channel system at a sampling rate of 1000 Hz. Following the standard preprocessing pipeline established in the source study, the continuous raw data were processed using MNE-Python. Specifically, the signals were downsampled to 250 Hz and subjected to a band-pass filter (0.1-100 Hz) alongside a 50 Hz notch filter to eliminate power line noise. The data were subsequently segmented into fixed-length epochs time-locked to the stimulus onset, with baseline correction applied using the pre-stimulus interval. Finally, multivariate noise normalization was utilized to standardize signal amplitude across different channels and subjects.

1.2. Implementation details

Our framework is implemented using PyTorch 2.1.0 and trained on NVIDIA RTX 3090 GPUs. The core EEG encoder employs a hierarchical transformer architecture, configured with $L = 3$ layers and $H = 4$ attention heads. The input sequences are segmented into 250 and 600 time points for static and dynamic streams, respectively. For the Hierarchical Phase-Amplitude Coupling Fusion (HPACF) module, time-frequency features are extracted via Short-Time Fourier Transform (STFT), utilizing an FFT size of 128 for static signals and 64 for dynamic signals, with hop lengths adaptively calculated to produce approximately 26 aligned time windows. The differentiable phase binning mechanism operates with $N_{bins} = 18$ bins. Regarding the frequency band selection for phase-amplitude coupling, we determined the optimal configuration through systematic ablation studies across various phase-amplitude combinations (*e.g.*, $\theta - \gamma$, $\alpha - \gamma$, $\beta - \gamma$) to maximize decoding performance.

The EEG encoder is trained for 100 epochs with a batch size of 64, using the AdamW optimizer with a learning rate of 1×10^{-3}, a weight decay of 5×10^{-4}, and a final projection dimension of 1024. For the generative backend, we employ the SDXL-Turbo with frozen U-Net and VAE weights. To bridge the EEG manifold with the generative latent space, we fine-tune an IP-Adapter module, jointly optimizing the image projection layers and the cross-attention modules injected into the U-Net. This fine-tuning is conducted for 50 epochs using the AdamW optimizer with a learning rate of 1×10^{-4} and a weight decay of 1×10^{-2}. During inference, the generation is purely driven by the EEG embeddings and performed with 4 denoising steps.

2. Results details

2.1. Additional results on discriminative tasks

In this section, we present detailed quantitative evaluations to supplement the results reported in the main text. Table 1, Table 2, and Table 3 report the Top-K accuracies for retrieval, 72-way object classification, and 6-way color classification, respectively. The quantitative analysis demonstrates that BrainSSD achieves consistently superior performance compared to existing baselines. Specifically, our method exhibits a significant performance advantage across all evaluated tasks, confirming its effectiveness in extracting semantic representations from EEG signals. Moreover, the consistent improvements observed over both static-only and dynamic-only baselines further validate the efficacy of our synergistic fusion strategy.

Furthermore, we provide detailed results regarding the frequency band configuration and module ablation. As shown in Table 4, the experimental results identify the theta-alpha/gamma combination as the optimal frequency pair, yielding the highest retrieval accuracy. This observation aligns with neurophysiological principles regarding cognitive processing, suggesting that the synergy of theta and alpha rhythms facilitates the efficient integration of visual features. Additionally, the ablation study presented in Table 5 reveals that the full BrainSSD architecture outperforms all variants where key components (*e.g.*, PAC, Hierarchy, PDA, GSC) are removed, substantiating the essential contribution of each module to the overall decoding capability.

Table 1. Top-K accuracy (%) for each subject on the retrieval task.

Method	Sub 1		Sub 2		Sub 3		Sub 4		Sub 5		Sub 6		Sub 7		Sub 8		Sub 9		Sub 10		Sub 11		Sub 12		Avg	
	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5
DeepNet	4.2	11.8	2.8	6.3	8.3	20.8	4.9	16.0	2.1	9.0	3.5	10.4	4.9	18.8	2.8	6.3	2.8	8.3	3.5	8.3	6.3	16.0	2.8	5.6	4.1	11.5
EEGNet	4.9	10.4	4.2	9.7	4.2	13.2	4.2	6.3	4.2	11.8	4.9	9.0	5.6	14.6	3.5	9.7	2.8	12.5	1.4	8.3	4.2	13.9	3.5	8.3	3.9	10.7
EEGNetv4	5.6	9.0	4.2	11.8	7.6	13.2	4.2	14.6	3.5	10.4	2.8	12.5	2.8	10.4	3.5	10.4	1.4	7.6	4.2	11.8	4.9	12.5	5.6	14.6	4.2	11.6
FBCSP	3.5	11.1	4.2	9.0	2.8	9.7	2.8	11.8	6.3	12.5	2.1	9.0	3.5	8.3	3.5	9.0	3.5	12.5	2.8	9.0	4.2	12.5	3.5	9.7	3.5	10.4
Conformer	5.6	17.4	2.8	10.4	8.3	22.2	4.2	11.8	3.5	8.3	3.5	9.0	5.6	16.7	2.1	9.7	2.8	9.0	4.2	13.2	5.6	13.2	4.2	13.2	4.3	12.9
TSConv	4.9	12.5	4.2	13.2	6.3	19.4	5.6	13.2	2.8	10.4	3.5	13.2	2.8	13.9	2.1	9.0	4.2	11.1	4.9	9.7	6.3	11.1	2.1	8.3	4.1	12.1
BrainAlign	7.1	20.8	4.3	17.4	8.4	17.4	7.7	17.4	4.3	12.5	5.7	16.7	5.7	19.5	4.3	14.6	5.0	15.3	5.0	12.5	6.4	18.8	5.0	14.6	5.7	16.4
Neuro-3D	7.8	21.2	3.5	13.4	12.5	25.1	5.5	19.6	4.7	14.2	3.9	13.4	6.3	20.4	3.1	12.5	4.7	15.8	3.5	9.5	4.7	16.5	4.9	14.2	5.4	16.3
Static-Only	6.9	20.1	2.8	13.9	10.4	26.4	6.3	14.6	2.8	10.4	4.9	12.5	5.6	17.4	2.8	10.4	3.5	13.2	4.9	11.8	5.6	16.0	3.5	10.4	5.0	14.8
Dynamic-Only	4.9	14.6	3.5	12.5	6.9	18.1	5.6	16.0	4.2	12.5	4.2	13.2	6.3	16.7	4.2	11.8	3.5	12.5	3.5	11.8	6.3	18.8	4.9	15.3	4.8	14.5
BrainSSD	8.6	23.6	4.7	15.3	13.9	32.6	7.0	17.4	4.9	13.2	5.5	15.3	8.3	20.1	4.9	13.3	4.2	13.3	4.9	14.6	8.3	21.5	4.7	16.7	6.7	18.1

Table 2. Top-K accuracy (%) for each subject on the 72-way object classification task.

Method	Sub 1		Sub 2		Sub 3		Sub 4		Sub 5		Sub 6		Sub 7		Sub 8		Sub 9		Sub 10		Sub 11		Sub 12		Avg	
	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5
DeepNet	4.9	12.5	2.3	6.9	5.6	13.2	4.2	9.7	2.3	11.1	4.2	10.4	6.3	11.8	3.5	5.6	2.1	9.7	3.5	8.3	4.2	12.5	2.1	6.9	3.7	9.9
EEGNet	4.9	11.8	3.5	11.1	2.8	12.5	3.5	7.6	4.2	12.5	4.9	13.2	6.3	10.4	3.5	6.3	4.9	8.3	2.1	4.2	4.2	12.5	1.4	6.3	3.8	9.7
EEGNetv4	3.5	9.7	4.2	11.1	9.0	13.2	3.5	15.3	1.4	11.8	2.8	8.3	4.2	12.5	2.1	11.8	4.2	7.6	4.9	11.1	4.2	13.2	3.5	9.7	3.9	11.3
FBCSP	5.6	13.2	4.2	8.3	6.3	9.0	3.5	12.5	4.2	9.0	4.2	11.1	4.9	11.8	3.5	10.4	4.9	11.1	3.5	9.7	5.6	16.0	3.5	11.1	4.5	11.1
Conformer	4.2	12.5	3.5	6.9	4.8	13.9	2.3	11.1	4.2	10.4	4.8	14.6	7.6	9.7	2.3	8.3	3.5	7.6	4.2	13.2	4.9	10.4	2.8	5.6	4.1	10.3
TSConv	4.2	10.4	2.8	9.7	4.8	12.5	4.2	9.7	4.8	12.5	5.6	11.1	6.3	16.7	2.8	7.6	3.5	9.7	2.8	6.3	4.9	8.3	3.5	7.6	4.1	10.1
BrainAlign	7.8	16.3	5.1	14.7	7.2	18.2	6.5	18.9	5.8	13.3	5.1	15.4	6.5	20.9	4.4	12.6	4.4	14.7	6.5	14.7	7.2	18.2	7.2	18.2	6.1	16.3
Neuro-3D	6.0	20.8	5.5	14.1	11.8	22.7	5.5	16.4	3.9	14.1	5.8	15.6	7.0	23.4	3.1	10.9	3.9	12.5	3.9	11.1	9.7	20.3	4.7	13.3	5.9	16.3
Static-Only	7.6	20.8	4.9	13.2	8.3	28.5	5.6	16.7	2.8	11.1	4.2	12.5	7.6	18.8	3.5	12.5	3.5	13.2	3.5	11.1	6.3	16.7	4.2	13.2	5.2	15.7
Dynamic-Only	5.6	15.3	4.2	12.5	5.6	21.5	5.6	15.3	4.9	11.8	4.2	11.1	6.9	17.4	4.2	10.4	4.2	11.1	4.2	11.8	6.9	19.4	4.9	13.2	5.1	14.2
BrainSSD	9.0	23.6	5.6	14.1	13.2	32.6	6.3	18.8	4.9	14.6	7.0	18.8	9.0	24.3	6.3	13.9	5.5	13.3	5.6	14.6	8.3	22.2	4.9	13.9	7.1	18.7

Table 3. Top-K accuracy (%) for each subject on the 6-way color classification task.

Method	Sub 1		Sub 2		Sub 3		Sub 4		Sub 5		Sub 6		Sub 7		Sub 8		Sub 9		Sub 10		Sub 11		Sub 12		Avg	
	t-1	t-2	t-1	t-2	t-1	t-2	t-1	t-2	t-1	t-2	t-1	t-2	t-1	t-2	t-1	t-2	t-1	t-2	t-1	t-2	t-1	t-2	t-1	t-2	t-1	t-2
DeepNet	20.8	41.7	20.6	41.7	22.2	58.7	19.4	57.1	20.1	40.3	19.4	66.9	21.5	40.0	24.3	63.4	19.4	40.5	22.2	57.8	24.3	47.4	18.7	40.5	21.0	49.7
EEGNet	12.5	46.6	15.3	43.8	19.4	42.0	13.2	35.7	16.6	44.0	38.2	61.4	21.5	48.2	27.1	55.1	14.6	42.0	10.4	43.3	11.1	44.7	21.3	51.0	18.4	46.5
EEGNetv4	31.3	56.9	27.8	50.0	27.8	55.6	18.8	41.0	26.4	47.2	31.3	53.5	23.6	50.7	22.9	48.6	20.1	40.3	20.8	47.2	17.4	41.7	21.5	48.6	24.1	48.4
FBCSP	23.6	53.5	11.1	28.5	15.3	15.3	18.8	47.2	15.3	36.8	18.1	25.0	9.7	28.5	9.7	15.3	25.0	50.0	16.0	16.7	32.6	58.3	10.4	36.8	17.2	34.3
Conformer	19.8	30.4	19.1	31.1	17.7	41.9	21.9	35.3	16.7	33.9	19.1	36.7	18.4	38.1	18.1	35.3	15.3	37.1	18.4	36.0	16.7	36.7	17.7	37.5	18.3	35.8
TSConv	33.1	61.9	27.2	55.7	31.0	59.2	28.9	57.8	33.8	61.9	33.1	61.3	35.2	67.5	28.9	55.7	31.0	59.2	27.6	59.2	28.9	56.4	34.5	58.3	31.1	59.5
BrainAlign	36.9	69.5	40.2	60.4	42.1	67.4	44.0	47.9	40.2	63.2	36.9	65.3	43.0	68.8	41.6	63.9	38.8	61.1	41.3	62.5	41.3	63.9	42.1	69.5	40.7	63.6
Neuro-3D	49.8	64.2	38.9	58.0	42.0	64.2	35.7	58.0	38.9	65.8	38.1	61.9	39.6	58.0	39.6	61.1	37.3	59.5	38.9	61.9	36.5	62.4	43.6	62.4	39.9	61.4
Static-Only	38.9	63.9	32.6	55.6	40.3	59.0	27.1	50.7	34.7	53.5	32.6	51.4	37.5	61.8	36.1	56.3	32.6	52.8	39.6	61.8	30.6	54.2	39.6	66.7	35.2	57.3
Dynamic-Only	29.9	54.2	29.9	52.1	32.6	54.9	34.7	56.9	34.0	58.3	36.8	56.9	35.4	58.3	34.0	54.9	35.4	61.8	32.6	57.6	38.9	66.0	43.1	54.9	34.8	57.2
BrainSSD	47.2	69.5	43.1	65.3	48.4	73.6	38.9	63.9	41.0	62.5	45.1	63.2	42.4	68.8	44.4	69.4	39.6	70.1	41.0	69.4	42.4	72.2	43.1	68.8	43.0	68.1

Table 4. Top-K retrieval accuracy (%) for each subject using different PAC frequency band combinations.

Frequency Pairs	Sub 1		Sub 2		Sub 3		Sub 4		Sub 5		Sub 6		Sub 7		Sub 8		Sub 9		Sub 10		Sub 11		Sub 12		Avg	
	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5
θ/γ	6.3	19.4	3.1	9.7	10.2	29.2	4.2	12.5	4.9	9.0	4.2	13.2	6.9	19.4	2.8	12.5	4.2	9.0	4.2	12.5	7.0	15.3	2.8	10.4	5.1	14.4
α/γ	6.3	8.3	4.9	9.0	11.1	27.8	3.5	12.5	5.6	9.0	4.9	16.0	6.3	19.4	4.2	11.8	3.5	11.1	3.5	8.3	4.2	14.6	4.9	11.1	5.2	13.3
β/γ	7.6	20.1	2.1	7.6	10.4	27.8	3.5	11.8	1.4	7.6	4.9	15.3	6.3	18.8	2.1	11.1	4.9	11.1	4.2	12.5	4.2	16.0	3.5	11.1	4.6	14.2
θ-α/γ	8.6	23.6	4.7	15.3	13.9	32.6	7.0	17.4	4.9	13.2	5.5	15.3	8.3	20.1	4.9	13.3	4.2	13.3	4.9	14.6	8.3	21.5	4.7	16.7	6.7	18.1
α-β/γ	1.4	8.3	2.3	9.0	11.1	31.3	3.5	12.5	2.1	7.6	4.2	13.9	6.3	18.1	3.5	11.1	3.5	11.1	4.2	11.8	4.9	15.3	3.5	11.1	4.2	13.4
θ-α-β/γ	6.3	18.1	2.8	11.1	10.2	11.8	3.5	12.5	1.4	7.6	3.5	13.2	5.6	14.6	2.8	11.1	2.1	9.0	4.9	13.2	4.2	16.0	3.5	10.4	4.2	12.4

Table 5. Top-K retrieval accuracy (%) for each subject under different ablation settings of BrainSSD.

Model Variants	Sub 1		Sub 2		Sub 3		Sub 4		Sub 5		Sub 6		Sub 7		Sub 8		Sub 9		Sub 10		Sub 11		Sub 12		Avg	
	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5	t-1	t-5
w/o PAC	4.2	14.6	3.5	11.8	9.0	25.7	4.2	13.2	3.5	11.1	3.5	10.4	6.3	20.1	3.5	11.1	3.5	11.8	3.5	11.8	6.3	20.8	6.3	15.3	4.8	14.8
w/o Hierarchy	6.9	20.1	3.5	13.2	12.5	28.5	4.2	15.3	3.5	11.8	4.9	15.3	6.9	17.4	2.8	8.3	3.5	13.9	3.5	13.9	4.9	14.6	4.9	10.4	5.2	15.2
w/o PAC & Hierarchy	6.3	19.4	4.2	12.5	8.3	23.6	3.5	12.5	3.5	10.4	3.5	11.1	6.9	20.8	2.1	7.6	3.5	11.1	3.5	9.0	3.5	13.2	3.5	12.5	4.3	13.7
w/o PDA	6.3	18.8	2.1	11.8	8.3	24.3	4.9	10.4	8.3	13.2	2.1	9.7	5.6	17.4	4.2	12.5	2.8	10.4	2.8	13.9	3.5	13.9	4.9	12.5	4.6	14.1
w/o GSC	7.0	20.1	2.3	13.3	8.3	25.7	3.9	13.3	3.1	11.8	5.5	13.2	6.3	18.8	4.2	11.1	3.5	9.7	3.5	11.8	4.9	13.9	4.2	11.1	4.7	14.5
w/o PDA & GSC	6.9	17.4	2.1	8.3	6.3	23.6	3.5	11.8	3.5	10.4	4.2	13.9	6.3	18.8	3.5	9.7	3.5	8.3	4.9	12.5	4.9	14.6	3.5	11.8	4.4	13.4
BrainSSD (Full)	8.6	23.6	4.7	15.3	13.9	32.6	7.0	17.4	4.9	13.2	5.5	15.3	8.3	20.1	4.9	13.3	4.2	13.3	4.9	14.6	8.3	21.5	4.7	16.7	6.7	18.1

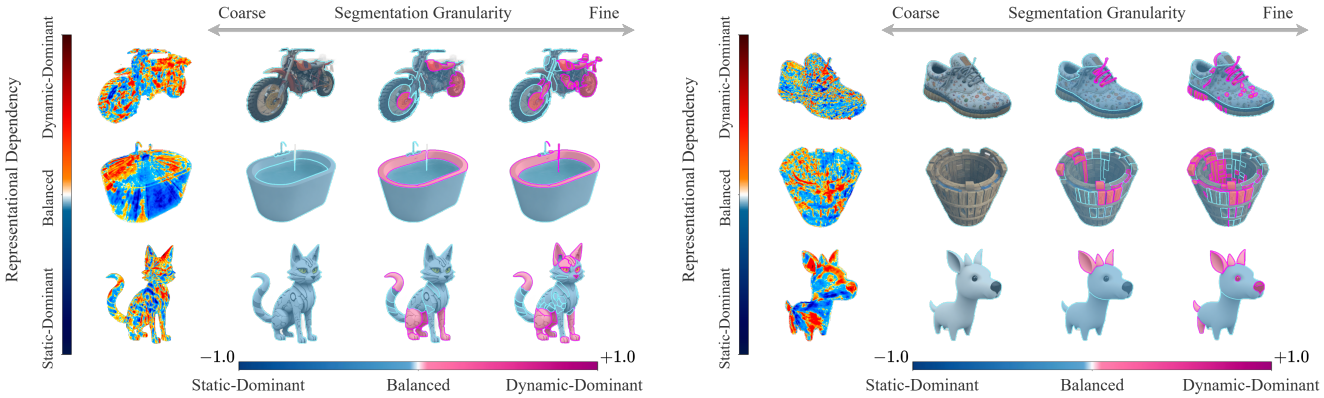

Figure 1. Extended visualization of representational dependency across diverse object categories and segmentation granularities. The first column of each block displays the RDI maps, where red denotes dynamic-stream dominance and blue denotes static-stream dominance. The subsequent columns visualize the relative dependency on static versus dynamic streams across varying segmentation levels. Consistent with the main text, regions corresponding to the object's main body or smooth surfaces (*e.g.*, the bathtub's wall) show a reliance on static signals, whereas intricate geometric details (*e.g.*, the motorcycle's engine, the shoe's laces, or the cat's ears) are dominated by dynamic signals.

2.2. Additional results on generative reconstruction

Additional reconstructed results alongside the corresponding Ground Truth and baseline comparisons are presented in Fig. 2. The proposed BrainSSD framework exhibits robust performance across a wide range of diverse object categories. As illustrated in the visual comparisons, our method effectively integrates complementary cues from static and dynamic streams, thereby capturing correct semantic categories, intricate geometric details, and consistent 3D structures, significantly outperforming the single-stream baselines.

2.3. Additional interpretation analysis

Figure 1 presents extended visualizations of the Representational Dependency Index (RDI) accompanied by a multi-granularity segmentation analysis. Consistent with the findings in the main paper, these results confirm that dynamic stream dominance correlates strongly with the geometric complexity of the visual stimulus. For instance, the *Motorcycle* and *Shoe*, featuring intricate geometric details like mechanical structures and laces, exhibit a pronounced Dynamic-Dominant pattern (visualized in red). Conversely, the *Bathtub*, which consists primarily of smooth surfaces and a simple holistic form, shows a widespread Static-Dominant pattern (visualized in blue). This functional specialization is further elucidated by the segmentation analysis. By isolating specific object parts, we observe that the static stream consistently dominates regions corresponding to the main body and global contour. Meanwhile, the dynamic stream asserts strong dominance in localized regions rich in geometric details, exemplified by the tire treads of the *Motorcycle*. These results reinforce our conclusion that the static stream efficiently encodes the object's holistic form, while the dynamic stream is crucial for resolving fine-grained structural information.

Figure 2. Additional qualitative results of the generative reconstruction pipeline. We present a comprehensive comparison between the proposed BrainSSD (Fused) and single-stream baselines (Static-Only, Dynamic-Only) alongside the Ground Truth. For each object, both the reconstructed 3D object images (left) and the corresponding point clouds (right) are displayed.