Bridging the Vision-Brain Gap with an Uncertainty-Aware Blur Prior

Supplementary Material

A. Experimental details

A.1. Datasets details

THINGS-EEG [21] is a large scale EEG dataset included 10 subjects with the Rapid Serial Visual Presentation (RSVP) paradigm [23, 30, 32]. The EEG data are collected using 64-channel EASYCAP equipment with the standard 10-10 system [49]. The training set includes 1654 concepts with each concept 10 images, and each image repeats 4 times (1654 concepts $\times$ 10 images/concept $\times$ 4 trials/image) per subject. The test set includes 200 concepts with each concept 1 image, and each image repeats 80 times (200 concepts $\times$ 1 image/concept $\times$ 80 trials/image) per subject.

For data preprocessing, we follow the method detailed in [60]. Raw EEG data filtered to [0.1, 100] Hz has 63 channels and a sample rate of 1000 Hz. EEG data is epoched into trials ranging from 0 to 1000 ms after stimuli onset with baseline correction using the prior 200 ms average. EEG data is down-sampled to 250 Hz and 17 channels are selected overlying occipital and parietal cortex related to visual[1]. For the purpose of high Signal-to-Noise Ratio (SNR), EEG repetitions are averaged, resulting in total of 16540 training samples and 200 test samples per subject. Additionally, we store EEG data in float16 format to enable faster reading speeds and reduce storage requirements.

THINGS-MEG [26] dataset involves four participants and is characterized by 271 channels. The experimental design incorporates a relatively long stimulus duration of 500 ms, followed by a blank screen with a duration of 1000 ± 200 ms. It consists of 1854 concepts$\times$12 images$\times$1 repetitions in the training stage and 200 concepts$\times$1 image$\times$12 repetitions in the test stage.

We follow the settings described in [60]. During the data processing phase, 200 test concepts are discarded from the training set to construct the zero-shot task, mirroring the procedures in that study. Subsequently, the MEG data are epoched into trials covering the period from 0 to 1000 ms after the stimuli onset. For preprocessing, a band-pass filter within the range of [0.1, 100] Hz is utilized, and baseline correction is carried out after down-sampling the data to 200 Hz. Additionally, we average all MEG repetitions of one image to ensure the signal-to-noise ratio. Additionally, we store EEG data in float16 format to enable faster reading speeds and reduce storage requirements.

[1] P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, O2

A.2. Implementation details

Environment. Our method is implemented with Python 3.8.19, CUDA 12.0, and PyTorch 2.4.1. The required libraries are specified in the `requirements.txt` file provided in the repository. The experiments are performed on a machine equipped with an Intel Xeon Platinum 8352V CPU, four V100 GPUs, and 256 GB of RAM.

Training Configuration. We use a batch size of 1024 and train the model for 50 epochs. The learning rate is set to 1e-4 for intra-subject setting and 1e-5 for inter-subject setting. Gradient updates are performed using the AdamW[42] optimizer with weight decay set to 1e-4. Early stopping is employed to monitor training loss and validation performance, concluding the training process to mitigate overfitting when improvements stabilize. Notably, we use the softplus function instead of the exponential function to ensure that temperature parameter τ remains positive and continuous, as softplus offers a smoother and more stable transition, avoiding the numerical instability of the exponential function. For all above experiments, the hyperparameter r_0 is set to 0.25 and c is set to 10.

Architectures. We use EEGProject as the brain encoder, detailed as follows:

```
Input: Brain Data
       ↓
Linear Layer (Input_dim → Proj_dim)
       ↓
ResidualAdd
 |   +----------------------+
 |   | GELU                 |
 |   | Linear Layer         |
 |   | Dropout (Rate: 0.3)  |
 |   +----------------------+
 +---- ↓
LayerNorm
       ↓
Output: Latent
```

We provided the number of parameters and embedding dimension within different EEG encoders [34, 56, 60], in Tab. 21. Compared to other models, EEGProject achieves its performance through a simple yet effective architecture, while remaining lightweight with 5.154M parameters, especially in comparison to the vision branch.

We also provide the parameter counts for various CLIP vision branch models [29] to offer a comprehensive comparison across architectures in Tab. 22.

B. Results details

B.1. Retrieval Case Analysis

We present the top-5 retrieval results on THINGS-EEG dataset, including both good cases and bad cases, as shown in Fig. 9 and Fig. 10, respectively. Good cases demonstrate the model's capability to effectively align with the target stimuli and retrieve relevant results. An intriguing retrieval result is that the model not only retrieves items with similar materials but also demonstrates **associations with the orientation and quantity of objects**. These observations warrant further investigation in future studies.

Figure 9. Good Cases: Top-5 Retrieval Results for Various Stimuli.

In contrast, bad cases reveal limitations in distinguishing fine-grained features or addressing semantic inconsistencies. It is challenging to distinguish highly similar stimuli due to the limited information contained in brain signals.

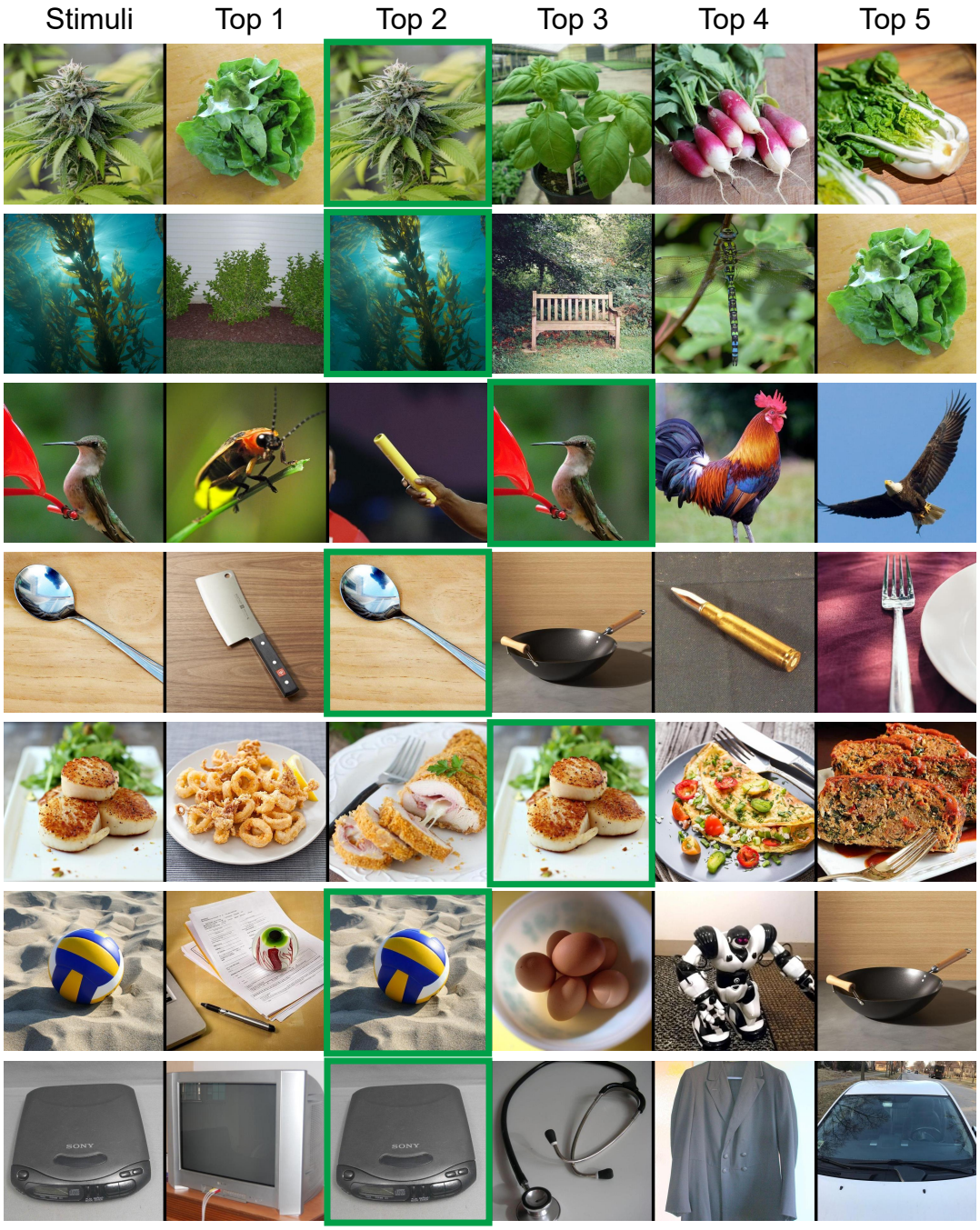
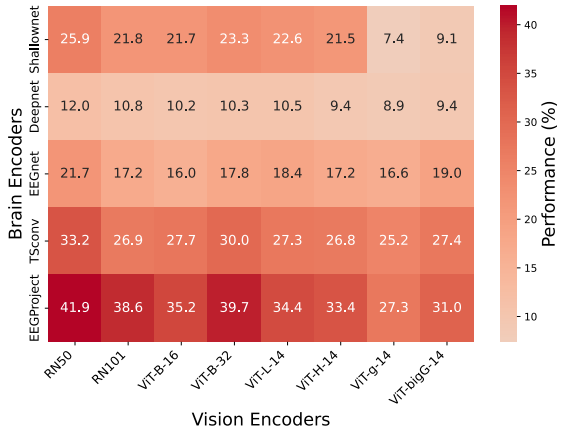
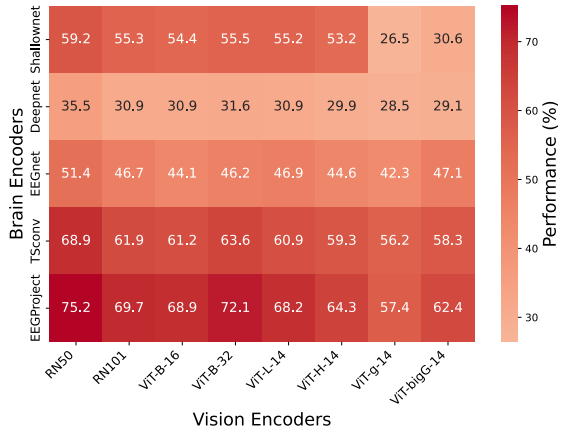


Figure 10. Bad Cases: Top-5 Retrieval Results for Various Stimuli.

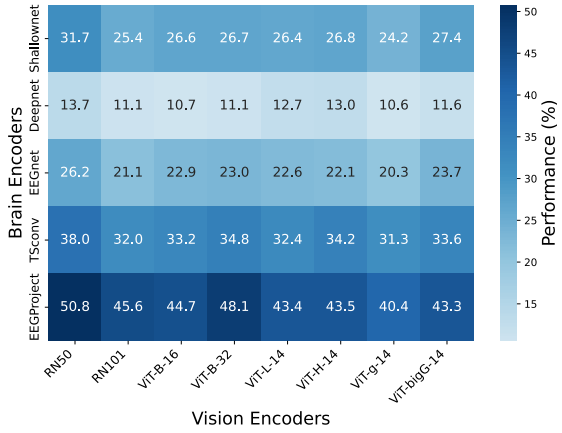
B.2. THINGS-EEG Results



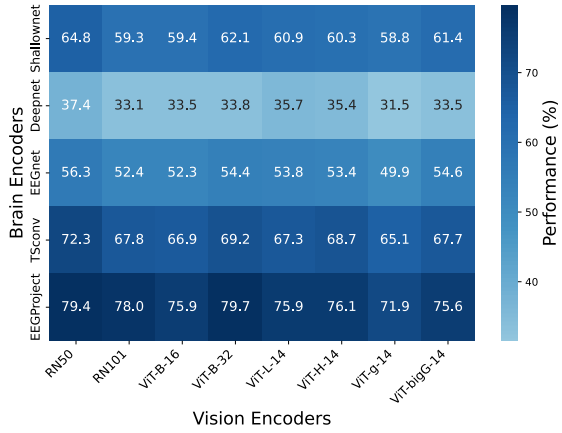
(a) **Top-1** accuracy (%) of **Vanilla** on the THINGS-EEG dataset.



(b) **Top-5** accuracy (%) of **Vanilla** on the THINGS-EEG dataset.

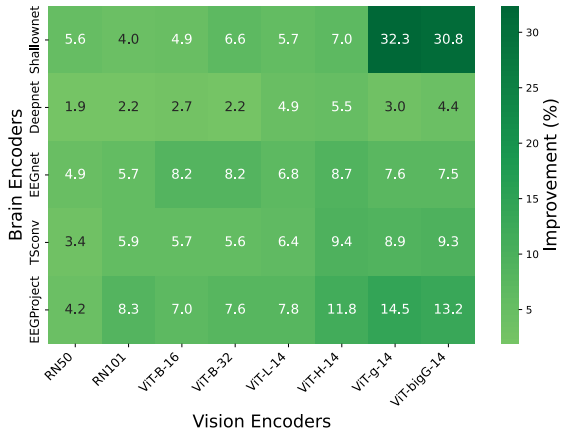


(c) **Top-1** accuracy (%) of **UBP** on the THINGS-EEG dataset.



(d) **Top-5** accuracy (%) of **UBP** on the THINGS-EEG dataset.

(e) **Top-1** accuracy **improvement** (%) on the THINGS-EEG dataset.



(f) **Top-5** accuracy **improvement** (%) on the THINGS-EEG dataset.

Figure 11. Results on the THINGS-EEG dataset.

Table 5. Top-1 and Top-5 Accuracy (%) on THINGS-EEG with CLIP **RN50** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Subject 8		Subject 9		Subject 10		Avg	
Backbone	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
ShallowNet	18.0	48.3	24.0	54.8	23.0	58.8	30.0	63.2	20.3	45.7	27.8	62.5	26.0	60.5	32.8	69.7	24.5	59.3	33.0	69.2	25.9	59.2
w/ UBP	24.5	57.5	25.0	60.5	31.2	64.5	35.2	66.8	20.7	49.0	37.3	70.0	33.7	65.3	42.2	78.0	27.0	63.7	40.0	72.7	31.7	64.8
DeepNet	7.8	27.5	11.0	35.8	11.2	32.0	14.5	40.0	8.2	24.0	12.2	38.5	10.5	33.2	16.3	39.0	15.5	40.5	12.5	44.5	12.0	35.5
w/ UBP	10.5	33.2	10.8	35.3	11.7	37.0	17.8	44.2	6.5	23.3	17.0	40.5	14.2	36.2	14.5	42.0	16.0	36.7	17.5	45.7	13.7	37.4
EEGNet	11.0	38.3	18.5	51.7	19.5	51.3	24.0	56.2	16.3	41.8	24.5	57.2	19.5	43.8	30.0	62.5	24.2	50.0	29.5	61.0	21.7	51.4
w/ UBP	18.5	44.0	23.2	53.2	25.7	58.2	33.3	62.0	20.0	45.8	30.2	62.7	22.5	53.2	33.7	66.2	25.0	52.0	29.5	65.5	26.2	56.3
TSConv	29.3	61.3	27.0	63.2	32.0	73.0	35.0	72.5	25.2	59.0	38.0	74.3	31.0	66.5	43.0	76.8	31.5	66.5	40.0	75.5	33.2	68.9
w/ UBP	38.0	69.5	28.7	68.5	38.8	75.7	41.3	74.3	29.7	61.0	42.0	76.7	36.2	70.8	48.5	79.7	33.2	67.3	44.0	79.0	38.0	72.3
EEGProject	30.8	64.2	39.5	71.5	42.2	78.5	42.8	76.5	33.8	69.2	45.5	79.3	40.7	71.8	49.8	80.5	41.7	74.5	52.5	85.8	41.9	75.2
w/ UBP	41.2	71.2	50.7	82.2	50.2	81.7	49.3	77.8	43.3	70.8	58.7	82.5	49.8	79.0	57.8	84.3	44.5	76.7	62.3	87.7	50.8	79.4

Table 6. Top-1 and Top-5 Accuracy (%) on THINGS-EEG with CLIP **RN101** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Subject 8		Subject 9		Subject 10		Avg	
Backbone	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
ShallowNet	17.8	45.7	18.2	53.0	20.7	53.5	28.0	61.0	15.2	43.0	21.8	58.0	20.3	54.8	29.7	65.3	21.0	54.2	25.2	64.2	21.8	55.3
w/ UBP	23.5	55.3	19.5	56.0	26.0	57.7	29.7	63.7	15.5	42.5	29.7	66.0	22.5	56.5	32.8	68.3	22.0	59.5	32.5	67.5	25.4	59.3
DeepNet	8.7	23.5	11.3	29.5	5.7	23.3	15.0	38.0	7.2	21.5	10.2	31.0	10.0	28.7	9.5	33.2	12.5	35.8	17.2	44.5	10.8	30.9
w/ UBP	8.8	27.8	10.0	30.3	8.2	30.5	14.3	39.5	8.5	22.0	12.2	32.8	8.8	30.0	11.2	34.5	13.3	37.7	15.5	46.3	11.1	33.1
EEGNet	12.5	32.8	14.5	42.7	15.3	43.8	22.5	54.7	13.2	40.3	16.5	51.5	16.0	42.7	20.0	56.7	17.7	45.7	24.2	56.2	17.2	46.7
w/ UBP	18.0	43.8	18.8	52.0	19.5	51.7	22.3	55.8	18.3	46.3	26.0	60.5	18.2	43.5	21.5	58.5	21.7	50.0	26.7	62.3	21.1	52.4
TSConv	24.8	56.5	24.8	58.5	23.3	60.5	33.5	67.8	19.0	48.5	27.5	68.2	25.0	58.0	34.3	71.5	25.5	60.5	31.7	69.0	26.9	61.9
w/ UBP	30.8	65.0	24.2	65.2	29.5	65.2	34.0	69.5	24.2	55.0	40.0	77.0	26.5	62.3	41.2	77.8	29.7	66.2	39.5	74.8	32.0	67.8
EEGProject	29.0	59.3	36.0	66.3	40.7	72.5	42.0	73.5	30.5	58.8	39.7	72.0	39.2	70.5	41.0	77.5	37.3	67.0	50.5	79.5	38.6	69.7
w/ UBP	35.2	69.7	45.5	80.8	46.5	78.7	47.7	77.7	40.7	71.3	49.5	83.5	42.7	75.0	55.0	81.0	39.2	75.0	53.5	87.3	45.6	78.0

Table 7. Top-1 and Top-5 Accuracy (%) on THINGS-EEG with CLIP **ViT-B-16** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Subject 8		Subject 9		Subject 10		Avg	
Backbone	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
ShallowNet	15.0	41.0	16.0	48.3	25.3	58.2	28.5	61.8	14.5	41.8	21.2	57.7	21.8	54.8	28.2	63.5	21.5	52.3	25.2	65.0	21.7	54.4
w/ UBP	22.3	51.8	23.5	53.0	24.0	61.8	32.3	65.0	16.8	45.0	28.2	64.0	27.2	60.0	34.7	71.0	23.0	55.3	34.3	67.0	26.6	59.4
DeepNet	9.0	22.5	6.7	29.3	9.0	30.5	13.5	36.8	4.7	17.8	10.0	31.0	10.0	31.2	9.8	33.2	15.3	37.5	14.2	39.0	10.2	30.9
w/ UBP	10.5	26.0	8.7	32.8	9.0	32.5	15.0	42.7	5.5	23.0	12.3	34.0	9.3	31.2	12.0	35.0	13.7	39.5	10.5	38.8	10.7	33.5
EEGNet	12.0	36.0	10.7	40.2	19.8	47.2	19.2	45.5	9.8	33.8	20.5	49.0	14.2	38.8	19.0	52.0	16.3	44.0	18.8	54.2	16.0	44.1
w/ UBP	18.8	41.5	18.5	48.5	23.5	55.0	24.7	57.0	16.5	44.0	27.0	57.2	20.7	49.8	29.3	58.5	23.0	50.2	27.5	61.2	22.9	52.3
TSConv	19.7	49.0	23.5	56.7	27.0	63.7	31.5	67.0	18.8	49.3	34.5	65.3	27.2	60.7	36.5	72.8	27.8	59.5	30.2	67.5	27.7	61.2
w/ UBP	28.7	63.0	28.0	63.0	31.2	66.8	37.0	72.2	22.5	54.5	41.0	74.8	33.7	63.0	42.2	76.7	31.7	63.5	36.3	71.5	33.2	66.9
EEGProject	29.5	59.5	32.8	68.0	36.5	71.8	35.3	68.0	28.2	54.0	37.2	73.3	34.3	70.0	36.8	75.0	37.3	69.2	44.0	79.7	35.2	68.9
w/ UBP	37.3	62.7	46.5	77.5	46.5	76.2	42.8	77.2	34.7	66.8	46.8	79.2	40.3	74.5	51.7	81.2	44.8	75.7	55.5	87.5	44.7	75.9

Table 8. Top-1 and Top-5 Accuracy (%) on THINGS-EEG with CLIP **ViT-B-32** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Subject 8		Subject 9		Subject 10		Avg	
Backbone	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
ShallowNet	15.8	45.0	17.8	49.0	27.3	57.0	28.7	65.5	17.5	41.0	25.2	60.5	20.3	54.7	30.7	63.2	22.7	56.0	27.3	63.0	23.3	55.5
w/ UBP	20.0	53.2	22.0	55.0	26.5	63.0	35.3	69.2	19.5	48.7	27.3	65.7	27.7	62.7	37.0	73.5	22.5	60.8	29.7	69.3	26.7	62.1
DeepNet	8.8	23.8	10.2	30.5	9.8	29.3	12.0	39.0	5.3	19.7	10.7	32.8	11.2	32.0	9.0	31.8	13.5	37.3	12.8	39.7	10.3	31.6
w/ UBP	8.7	29.0	10.2	34.3	9.3	34.3	12.5	41.2	8.5	21.0	11.2	33.5	14.0	33.0	12.0	32.3	12.0	37.0	12.8	42.0	11.1	33.8
EEGNet	14.3	36.0	14.2	43.0	18.3	46.3	24.0	53.0	12.8	36.5	20.0	50.2	15.5	43.8	20.3	53.5	18.8	43.8	20.2	56.0	17.8	46.2
w/ UBP	18.5	40.5	21.7	54.7	20.7	56.0	25.0	58.2	19.2	45.3	30.2	60.5	20.3	48.3	27.0	63.5	22.0	54.0	25.2	63.5	23.0	54.4
TSConv	19.8	55.8	22.7	55.8	33.8	65.0	34.5	72.5	20.0	52.5	33.7	68.3	25.3	60.5	40.3	74.5	33.7	62.5	36.5	69.0	30.0	63.6
w/ UBP	30.8	63.5	27.5	64.0	37.5	70.8	36.8	72.3	25.8	57.7	43.5	76.0	34.8	68.0	41.5	78.7	34.0	69.0	36.2	72.5	34.8	69.2
EEGProject	30.7	59.5	39.2	70.5	43.0	79.5	40.3	73.5	31.2	65.0	42.0	74.3	36.8	68.8	45.7	75.0	41.2	71.3	46.5	83.2	39.7	72.1
w/ UBP	37.5	70.0	46.5	80.5	52.8	85.5	47.2	80.8	37.5	70.7	54.5	81.7	43.8	79.5	58.8	81.5	46.5	77.0	56.0	89.7	48.1	79.7

Table 9. Top-1 and Top-5 Accuracy (%) on THINGS-EEG with CLIP **ViT-L-14** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Subject 8		Subject 9		Subject 10		Avg	
Backbone	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
ShallowNet	16.3	45.3	17.8	51.2	19.7	57.5	27.0	60.2	17.5	45.5	26.3	60.8	20.0	48.3	30.5	63.2	21.8	54.5	29.5	65.7	22.6	55.2
w/ UBP	15.7	47.8	23.3	52.0	24.5	61.2	32.2	66.7	17.2	50.5	33.2	69.2	25.7	61.5	35.5	71.0	23.5	58.8	33.2	70.2	26.4	60.9
DeepNet	8.3	23.0	9.0	29.0	9.0	29.7	12.8	34.7	7.3	20.5	11.5	33.3	8.3	31.7	13.0	36.3	13.5	36.5	12.0	33.7	10.5	30.9
w/ UBP	10.2	28.7	11.8	33.5	12.0	34.3	16.5	41.2	6.7	23.7	14.5	39.2	13.2	37.5	12.5	40.5	15.3	37.5	13.8	41.0	12.7	35.7
EEGNet	12.8	35.8	13.3	42.0	16.3	47.8	20.8	50.0	14.2	34.5	23.8	52.7	14.3	41.8	24.2	58.8	17.5	47.7	27.0	58.5	18.4	46.9
w/ UBP	14.3	40.5	20.7	45.5	21.7	54.0	29.0	59.5	18.0	45.7	27.5	61.3	19.7	53.0	26.3	61.3	21.8	54.0	27.5	63.0	22.6	53.8
TSConv	21.7	55.7	24.0	56.5	25.2	61.0	31.2	65.5	20.0	47.2	31.2	69.0	25.7	58.2	37.7	70.0	21.8	56.0	34.0	70.0	27.3	60.9
w/ UBP	26.5	60.0	26.0	60.5	32.3	66.5	36.3	72.0	24.5	56.0	39.0	75.2	28.2	65.3	42.8	78.5	29.0	63.5	39.0	76.0	32.4	67.3
EEGProject	26.7	55.3	33.5	66.5	35.2	73.3	32.0	70.0	27.5	51.7	36.2	72.2	33.5	72.8	44.5	75.5	29.5	67.0	45.3	77.7	34.4	68.2
w/ UBP	31.2	66.2	39.0	73.8	43.3	77.5	43.8	76.2	31.0	69.5	51.5	83.0	41.0	74.5	56.2	80.5	42.5	74.0	54.8	84.3	43.4	75.9

Table 10. Top-1 and Top-5 Accuracy (%) on THINGS-EEG with CLIP **ViT-H-14** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Subject 8		Subject 9		Subject 10		Avg	
Backbone	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
ShallowNet	15.2	42.2	16.8	47.0	25.0	56.5	25.5	59.3	11.5	43.0	23.2	55.0	18.0	51.0	27.8	62.7	22.7	55.5	29.5	60.0	21.5	53.2
w/ UBP	21.0	51.0	21.8	50.3	24.2	62.7	32.0	64.3	17.8	45.7	30.5	69.5	24.2	60.0	31.7	71.8	26.2	59.5	38.3	67.8	26.8	60.3
DeepNet	8.0	24.8	9.0	27.2	11.5	31.0	9.0	30.3	4.5	20.7	9.5	27.8	9.5	30.0	8.8	32.0	11.2	34.0	13.3	41.0	9.4	29.9
w/ UBP	9.3	28.7	14.3	36.5	14.0	37.3	14.0	36.5	8.5	22.2	13.8	33.8	11.7	36.5	14.5	38.0	14.0	38.5	15.7	45.8	13.0	35.4
EEGNet	9.3	30.8	17.3	43.5	18.8	46.0	18.5	47.3	12.5	36.2	21.2	49.0	14.0	40.5	21.2	51.0	17.7	48.5	21.0	53.5	17.2	44.6
w/ UBP	15.0	42.5	19.7	49.0	22.0	54.0	26.0	58.5	20.5	44.0	22.5	58.0	21.2	49.2	26.0	60.0	21.5	54.5	27.0	63.8	22.1	53.4
TSConv	20.5	50.0	22.0	55.8	27.2	60.5	29.5	61.8	19.0	49.8	30.5	61.5	22.7	57.8	35.0	70.0	27.0	59.0	34.7	66.7	26.8	59.3
w/ UBP	28.7	63.3	29.7	65.0	32.8	69.5	37.0	71.0	28.7	59.0	39.5	76.2	33.2	65.3	40.0	77.8	29.7	63.7	43.0	76.5	34.2	68.7
EEGProject	19.7	54.5	31.0	62.0	39.0	70.0	35.0	62.0	26.3	54.0	38.2	68.5	29.3	61.8	39.5	72.2	34.8	62.2	40.7	76.0	33.4	64.3
w/ UBP	32.0	64.0	45.0	75.7	45.8	81.5	42.5	77.5	34.7	66.5	47.0	81.2	40.0	73.5	51.2	82.5	45.7	72.3	51.2	86.5	43.5	76.1

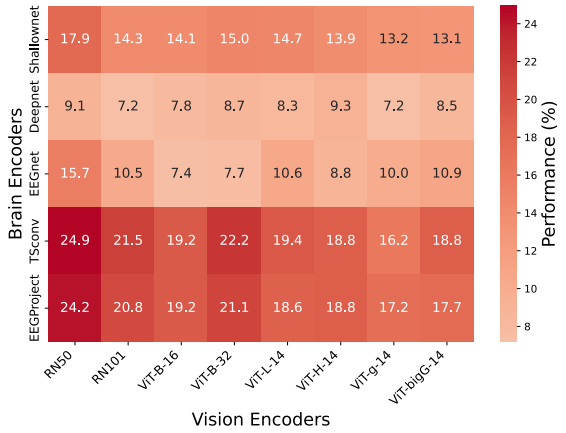
Table 11. Top-1 and Top-5 Accuracy (%) on THINGS-EEG with CLIP **ViT-g-14** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Subject 8		Subject 9		Subject 10		Avg	
Backbone	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
ShallowNet	6.5	23.5	4.7	21.0	6.2	28.8	9.8	31.0	5.0	17.0	7.3	25.5	6.8	24.7	11.3	33.5	5.7	25.0	10.8	34.5	7.4	26.5
w/ UBP	19.5	51.5	16.8	50.2	26.5	61.0	29.0	66.0	17.5	45.0	29.0	62.7	22.5	58.7	27.5	67.5	23.7	58.2	30.2	67.0	24.2	58.8
DeepNet	6.5	23.7	6.8	24.8	10.8	31.5	12.2	32.3	6.5	17.0	9.5	26.2	9.0	29.0	10.5	32.2	9.8	30.8	7.0	37.7	8.9	28.5
w/ UBP	9.5	26.0	8.8	26.5	11.0	33.2	12.8	33.2	6.0	21.5	12.0	33.7	11.0	36.0	9.5	33.7	11.0	32.8	14.2	38.5	10.6	31.5
EEGNet	8.3	29.3	13.3	39.2	19.3	47.0	20.3	48.7	11.5	35.0	19.5	44.0	15.0	36.8	20.0	52.2	16.5	40.5	22.0	50.7	16.6	42.3
w/ UBP	13.3	37.5	16.0	42.7	22.7	52.2	25.0	55.8	14.0	44.2	23.0	52.5	18.8	47.5	25.2	58.0	18.0	50.7	26.8	57.8	20.3	49.9
TSConv	24.5	52.0	20.0	50.5	26.0	57.7	27.5	61.0	20.5	48.8	25.3	58.5	24.5	51.0	31.0	63.0	22.7	55.0	29.7	64.5	25.2	56.2
w/ UBP	31.5	63.5	21.2	54.7	32.8	64.3	32.2	67.0	23.7	55.8	33.7	69.5	32.8	64.0	37.0	77.2	29.3	62.3	39.0	73.3	31.3	65.1
EEGProject	20.3	48.5	23.5	52.3	30.0	59.7	28.5	62.7	23.3	48.3	29.0	60.2	28.2	54.8	30.7	64.0	29.0	57.8	30.5	65.5	27.3	57.4
w/ UBP	29.0	62.0	38.5	67.5	42.2	77.2	41.0	74.8	31.0	59.5	44.3	74.8	37.3	71.0	48.5	81.0	42.0	69.7	50.2	81.2	40.4	71.9

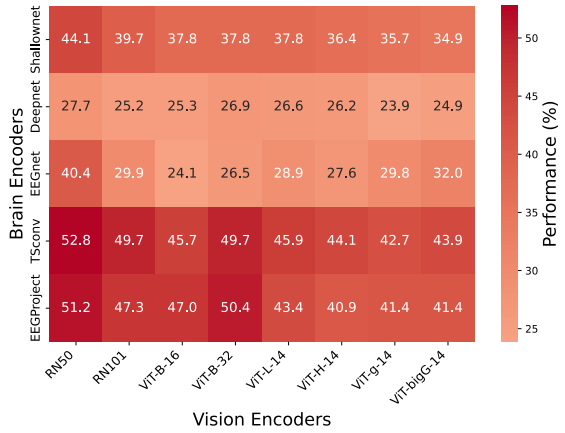
Table 12. Top-1 and Top-5 Accuracy (%) on THINGS-EEG with CLIP **ViT-bigG-14** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Subject 8		Subject 9		Subject 10		Avg	
Backbone	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
ShallowNet	7.8	26.5	6.3	19.0	9.0	33.8	9.8	33.0	6.0	20.0	10.2	29.7	7.8	32.3	12.3	39.0	9.2	32.3	13.0	40.3	9.1	30.6
w/ UBP	19.5	52.8	19.0	57.5	29.7	60.7	34.3	68.3	17.5	46.2	34.0	67.0	28.2	59.7	33.2	71.0	27.2	59.0	31.2	72.0	27.4	61.4
DeepNet	7.3	21.5	8.0	27.8	9.0	31.5	10.5	32.5	7.5	21.7	10.5	29.3	10.8	29.7	10.0	29.7	9.8	32.0	11.0	34.7	9.4	29.1
w/ UBP	7.3	21.5	8.0	27.8	9.0	31.5	10.5	32.5	7.5	21.7	10.5	29.3	10.8	29.7	10.0	29.7	9.8	32.0	11.0	34.7	9.4	29.1
EEGNet	13.0	36.2	13.0	43.0	20.0	47.2	23.2	51.5	17.0	36.7	18.8	50.7	16.8	42.5	25.5	58.0	19.3	50.5	23.8	54.5	19.0	47.1
w/ UBP	16.3	40.7	20.5	53.7	24.0	55.8	28.0	58.8	18.5	45.0	28.0	59.7	23.2	51.2	28.7	62.5	24.5	55.5	25.5	62.7	23.7	54.6
TSConv	22.0	48.0	18.2	57.2	26.7	58.5	31.2	61.0	19.8	46.0	32.5	63.5	24.7	53.7	38.5	68.0	24.8	58.0	35.8	69.5	27.4	58.3
w/ UBP	30.0	62.0	28.5	64.5	32.0	66.2	36.8	70.7	23.7	54.7	43.3	74.3	32.2	67.5	42.0	75.2	30.5	65.5	37.3	76.0	33.6	67.7
EEGProject	22.0	52.5	29.7	58.5	31.5	63.7	29.7	63.7	25.2	53.5	36.2	67.3	32.5	63.0	37.5	68.2	26.3	61.3	39.2	72.0	31.0	62.4
w/ UBP	33.8	63.7	41.8	78.2	43.8	78.2	39.7	73.8	34.0	64.0	55.8	82.7	38.7	74.8	49.0	80.3	43.5	74.3	53.2	85.5	43.3	75.6

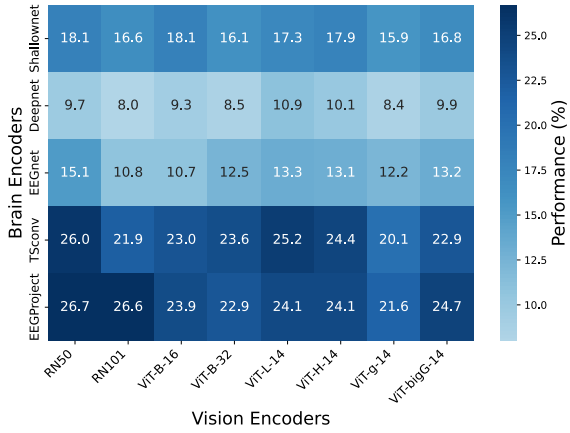
B.3. THINGS-MEG Results



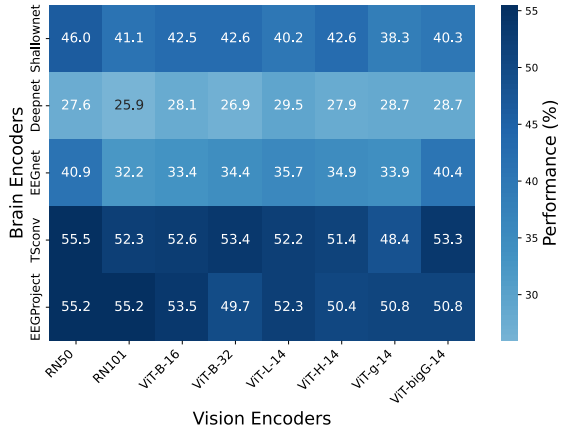
(a) **Top-1** accuracy (%) of **Vanilla** on the THINGS-MEG dataset.



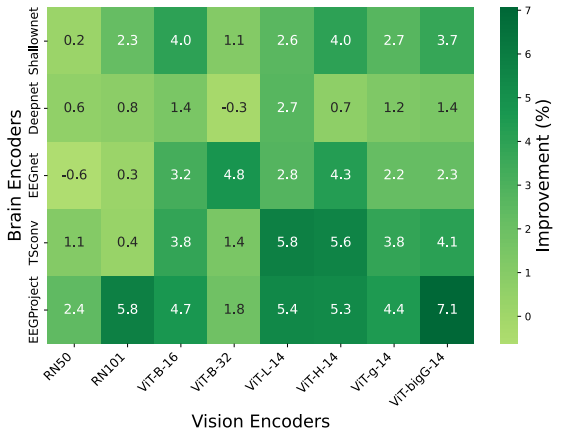
(b) **Top-5** accuracy (%) of **Vanilla** on the THINGS-MEG dataset.



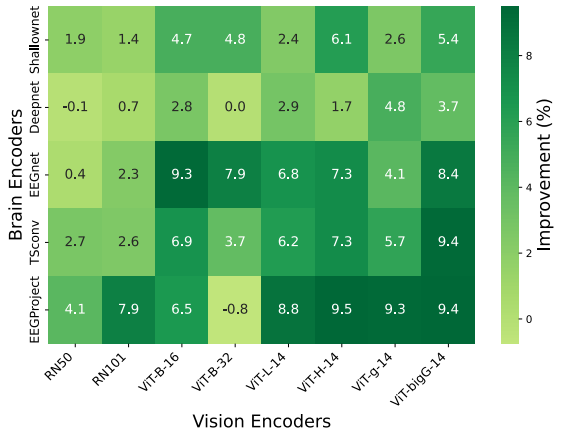
(c) **Top-1** accuracy (%) of **UBP** on the THINGS-MEG dataset.



(d) **Top-5** accuracy (%) of **UBP** on the THINGS-MEG dataset.



(e) **Top-1** accuracy **improvement** (%) on the THINGS-MEG dataset.



(f) **Top-5** accuracy **improvement** (%) on the THINGS-MEG dataset.

Figure 12. Results on the THINGS-EEG dataset.

Table 13. Top-1 and Top-5 Accuracy (%) on THINGS-MEG with CLIP **RN50** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Avg	
Backbone	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5
ShallowNet	9.8	32.0	27.3	58.5	25.2	56.2	9.3	29.7	17.9	44.1
w/ UBP	10.7	36.0	27.0	65.3	22.2	51.8	12.5	31.0	18.1	46.0
DeepNet	4.3	17.0	13.3	40.0	12.8	34.7	6.2	19.2	9.1	27.7
w/ UBP	4.5	14.8	16.5	40.5	11.3	36.0	6.7	19.2	9.7	27.6
EEGNet	10.5	26.8	25.2	59.3	19.3	47.5	8.0	28.2	15.7	40.4
w/ UBP	7.5	24.5	29.7	66.0	14.5	46.8	8.7	26.3	15.1	40.9
TSconv	14.8	40.5	39.2	73.5	30.2	60.8	15.5	36.2	24.9	52.8
w/ UBP	18.3	45.0	40.7	76.5	28.0	59.0	17.0	41.5	26.0	55.5
EEGProject	13.2	34.5	43.3	74.5	27.0	57.0	13.5	38.8	24.2	51.2
w/ UBP	15.0	38.0	46.0	80.5	27.3	59.0	18.5	43.5	26.7	55.2

Table 14. Top-1 and Top-5 Accuracy (%) on THINGS-MEG with CLIP **RN101** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Avg	
Backbone	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5
ShallowNet	8.5	27.8	20.0	53.0	21.0	49.7	7.7	28.2	14.3	39.7
w/ UBP	8.7	30.0	25.0	55.3	21.5	46.3	11.0	33.0	16.6	41.1
DeepNet	4.8	15.0	11.5	36.2	8.3	30.5	4.3	19.3	7.2	25.2
w/ UBP	3.5	17.0	12.3	38.5	9.5	27.5	6.8	20.7	8.0	25.9
EEGNet	6.5	18.0	17.2	47.2	14.2	36.5	4.0	17.7	10.5	29.9
w/ UBP	6.0	20.7	19.3	49.8	11.8	36.5	6.0	21.8	10.8	32.2
TSconv	15.0	37.7	32.0	67.2	24.7	55.5	14.3	38.5	21.5	49.7
w/ UBP	13.5	40.5	31.5	68.8	25.5	58.3	17.0	41.7	21.9	52.3
EEGProject	10.7	33.2	37.3	69.5	26.2	55.5	9.0	31.0	20.8	47.3
w/ UBP	15.3	37.7	44.8	76.5	31.0	63.5	15.5	43.3	26.6	55.2

Table 15. Top-1 and Top-5 Accuracy (%) on THINGS-MEG with CLIP **ViT-B-16** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Avg	
Backbone	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5
ShallowNet	8.2	28.5	21.5	49.8	20.5	45.0	6.2	27.8	14.1	37.8
w/ UBP	9.8	28.8	25.0	59.7	25.0	52.8	12.8	28.7	18.1	42.5
DeepNet	4.0	15.3	14.8	35.3	9.0	32.2	3.5	18.3	7.8	25.3
w/ UBP	6.8	19.0	14.8	35.3	8.0	35.8	7.5	22.3	9.3	28.1
EEGNet	3.0	12.8	14.5	40.3	9.5	29.5	2.7	13.8	7.4	24.1
w/ UBP	5.3	20.8	18.5	54.5	13.7	38.5	5.3	19.7	10.7	33.4
TSconv	10.8	33.3	29.3	62.3	26.0	54.8	10.7	32.5	19.2	45.7
w/ UBP	13.5	40.3	35.3	70.3	29.0	62.3	14.3	37.5	23.0	52.6
EEGProject	14.2	35.8	27.8	65.3	23.7	52.0	11.0	35.0	19.2	47.0
w/ UBP	16.7	39.2	38.2	72.5	23.7	62.3	16.8	40.0	23.9	53.5

Table 16. Top-1 and Top-5 Accuracy (%) on THINGS-MEG with CLIP **ViT-B-32** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Avg	
Backbone	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5
ShallowNet	8.2	24.2	24.0	54.8	18.3	46.0	9.5	26.2	15.0	37.8
w/ UBP	7.5	30.0	23.3	54.8	21.8	55.8	11.7	29.7	16.1	42.6
DeepNet	5.0	18.5	12.5	38.5	12.2	32.2	5.3	18.2	8.7	26.9
w/ UBP	4.7	18.3	13.3	37.5	10.7	32.8	5.2	19.0	8.5	26.9
EEGNet	4.5	15.7	15.7	45.5	8.5	31.2	2.2	13.5	7.7	26.5
w/ UBP	8.0	22.5	24.5	58.0	13.3	41.0	4.2	16.3	12.5	34.4
TSconv	16.3	38.5	34.3	68.5	24.2	54.0	14.0	37.7	22.2	49.7
w/ UBP	14.8	42.2	37.3	69.2	27.8	62.0	14.8	40.2	23.6	53.4
EEGProject	10.0	35.8	34.7	70.3	25.5	54.5	14.0	41.3	21.1	50.4
w/ UBP	15.0	38.0	34.0	66.5	29.7	63.3	12.7	31.0	22.9	49.7

Table 17. Top-1 and Top-5 Accuracy (%) on THINGS-MEG with CLIP **ViT-L-14** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Avg	
Backbone	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5
ShallowNet	6.5	26.0	23.0	51.7	19.0	44.0	10.2	29.5	14.7	37.8
w/ UBP	8.7	27.5	30.8	57.7	17.5	48.0	12.0	27.5	17.3	40.2
DeepNet	4.5	18.0	12.3	37.3	11.3	34.3	5.0	16.8	8.3	26.6
w/ UBP	6.2	18.8	18.5	44.0	12.8	34.5	6.2	20.7	10.9	29.5
EEGNet	6.2	18.5	18.2	45.0	11.7	33.8	6.0	18.5	10.6	28.9
w/ UBP	5.5	24.7	27.5	55.5	14.7	42.2	5.5	20.5	13.3	35.7
TSconv	15.0	34.8	28.8	64.0	21.5	50.0	12.5	35.0	19.4	45.9
w/ UBP	15.5	41.2	39.5	69.0	28.2	57.2	17.8	41.2	25.2	52.2
EEGProject	12.2	31.2	32.5	64.0	16.8	47.0	13.0	31.5	18.6	43.4
w/ UBP	14.8	37.5	39.0	72.5	24.5	56.2	18.0	42.8	24.1	52.3

Table 18. Top-1 and Top-5 Accuracy (%) on THINGS-MEG with CLIP **ViT-H-14** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Avg	
Backbone	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5
ShallowNet	7.5	26.0	24.5	50.5	14.5	43.8	9.0	25.5	13.9	36.4
w/ UBP	11.0	34.5	28.7	55.8	20.0	53.7	11.7	26.2	17.9	42.6
DeepNet	5.3	18.3	16.0	39.2	11.0	30.0	5.0	17.2	9.3	26.2
w/ UBP	4.3	19.5	18.0	41.8	12.7	33.2	5.2	17.0	10.1	27.9
EEGNet	5.7	18.3	16.7	45.0	8.2	31.2	4.5	15.7	8.8	27.6
w/ UBP	7.5	25.0	23.2	53.7	16.0	41.5	5.5	19.2	13.1	34.9
TSconv	12.0	31.5	30.0	58.7	20.3	51.5	13.0	34.8	18.8	44.1
w/ UBP	14.5	39.5	39.0	70.0	30.5	58.0	13.7	38.0	24.4	51.4
EEGProject	11.0	28.0	30.7	62.3	22.7	44.8	10.8	28.7	18.8	40.9
w/ UBP	15.5	37.3	41.0	74.8	24.8	55.3	15.3	34.5	24.1	50.4

Table 19. Top-1 and Top-5 Accuracy (%) on THINGS-MEG with CLIP **ViT-g-14** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Avg	
Backbone	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5
ShallowNet	9.5	25.7	17.5	47.0	17.3	42.7	8.5	27.3	13.2	35.7
w/ UBP	10.5	28.2	25.5	51.0	19.2	46.5	8.3	27.5	15.9	38.3
DeepNet	5.0	16.0	11.5	33.7	9.8	29.7	2.5	16.0	7.2	23.9
w/ UBP	4.3	17.2	14.2	42.5	9.0	33.8	6.2	21.2	8.4	28.7
EEGNet	6.5	20.3	17.0	46.3	12.0	32.5	4.5	20.3	10.0	29.8
w/ UBP	6.2	24.0	23.3	53.8	13.7	37.5	5.5	20.2	12.2	33.9
TSconv	10.8	33.8	26.0	59.3	18.2	48.5	10.0	29.3	16.2	42.7
w/ UBP	11.5	34.7	32.0	65.0	24.2	56.5	12.5	37.3	20.1	48.4
EEGProject	11.0	28.7	27.5	59.5	19.5	44.3	10.7	33.2	17.2	41.4
w/ UBP	13.5	40.0	35.8	69.8	23.8	55.0	13.5	38.3	21.6	50.8

Table 20. Top-1 and Top-5 Accuracy (%) on THINGS-MEG with CLIP **ViT-bigG-14** With/Without UBP.

	Subject 1		Subject 2		Subject 3		Subject 4		Avg	
Backbone	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5
ShallowNet	9.0	25.8	18.5	47.5	16.3	42.2	8.5	24.0	13.1	34.9
w/ UBP	6.5	25.5	25.5	55.8	25.5	51.2	9.8	28.7	16.8	40.3
DeepNet	4.2	18.2	13.0	36.0	11.3	29.0	5.5	16.5	8.5	24.9
w/ UBP	4.7	16.3	16.0	43.0	12.0	35.3	7.0	20.2	9.9	28.7
EEGNet	5.7	23.0	19.0	49.2	13.3	38.2	5.7	17.5	10.9	32.0
w/ UBP	8.2	28.5	23.7	61.0	16.3	50.2	4.5	22.0	13.2	40.4
TSconv	11.2	34.5	28.5	59.5	23.7	49.8	11.7	32.0	18.8	43.9
w/ UBP	13.2	40.5	36.5	68.5	28.7	65.7	13.0	38.5	22.9	53.3
EEGProject	9.0	30.0	32.8	59.5	20.3	46.5	8.7	29.5	17.7	41.4
w/ UBP	13.5	33.3	40.3	70.0	26.7	63.2	18.5	36.8	24.7	50.8

Table 21. Details of different EEG encoders with Emb dimension of 1024

Brain Encoder	Params
ShallowNet	2.56 M
DeepNet	2.76 M
EEGNet	2.34 M
TSConv	2.56 M
EEGProject	5.40 M

Table 22. Details of different Vision encoders

Vision Encoder	Params	Emb dim
RN50	38.32 M	1024
RN101	56.26 M	512
ViT-B-16	86.19 M	512
ViT-B-32	87.85 M	512
ViT-L-14	303.97 M	768
ViT-H-14	632.08 M	1024
ViT-g-14	1012.65 M	1024
ViT-bigG-14	1844.91 M	1280

B.4. EEG Feature Selection Analysis

Unless otherwise specified, EEG data spanning 1000 ms were selected, focusing on 17 visual-related (O+P) channels out of a total of 63, following previous work [60], where ablation studies are conducted on both channels and epochs. To verify the influence of the selection of EEG features, we conducted relevant ablation experiments. Tab. 23 and Fig. 13 show consistent improvements across various settings.

Table 23. Top-1 ACC (%) compariation with different channels.

Method	Occipital	Parietal	O+P (Our)	Others	All
Vanilla	21.0	33.7	42.0	10.2	35.6
UBP	26.9 (+5.9)	40.7 (+7.0)	50.9 (+8.9)	12.1 (+1.9)	42.2 (+6.6)

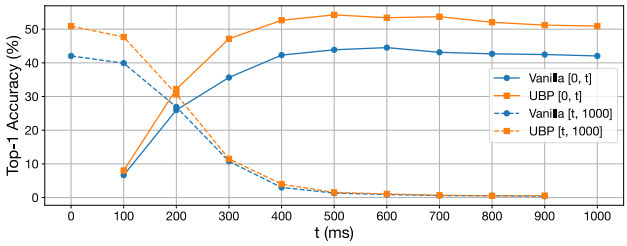


Figure 13. Top-1 ACC (%) compariation with different epoching.

B.5. More Evaluation Metrics

For a more comprehensive assessment of the method's performance, we provided additional metric values to validate UBP's effectiveness, specifically two new metrics: **mAP** and **Similarity Score** of paired samples. We present the performance averaged over 10 subjects. Tab. 24 demonstrates the results of the proposed method evaluated across different metrics. The calculation of metrics can be found in the repo.

Table 24. Retrieval Performance on THINGS-EEG.

Method	Top-1 Acc	Top-5 Acc	mAP*	Similarity*
Vanilla	42.0	75.2	56.6	0.160
UBP	50.9 (+8.9)	79.7 (+4.5)	63.8 (+7.2)	0.199 (+0.039)