Dynamic Convolution and Graph-Coupled Attention for Cross-Subject EEG-Vision Decoding

Tianyu Zhang, Fan Wan†, Kaili Sun, Xingyu Miao, Yueming Sun, Minye Shao, Yang Long†
Durham University · Tongfang Knowledge Network Digital Technology Co., Ltd.

BMVC 2025

Abstract

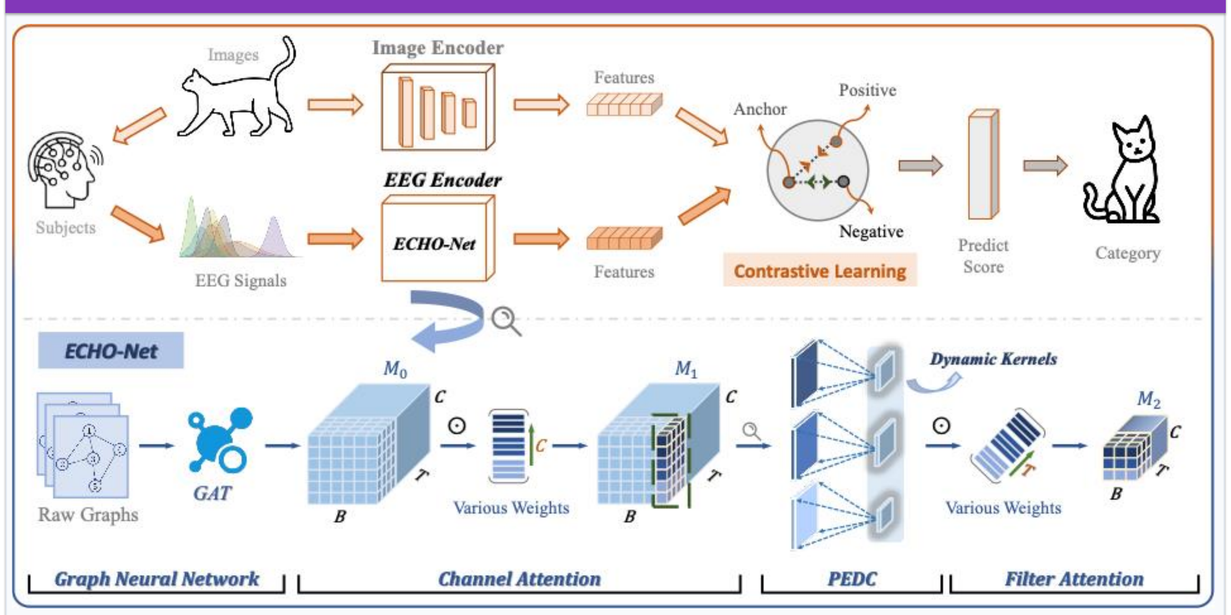
EEG decoding faces non-stationarity, low SNR, and inter-subject variability. We present ECHO-Net, a hierarchical framework that assembles trial-conditioned dynamic convolution (PEDC), channel–filter attention, and an integrated EEG-GAT for spatial reasoning. On the 200-way THINGS-EEG benchmark, ECHO-Net achieves 18.5% Top-1 and 44.1% Top-5 cross-subject accuracy, surpassing prior work by +2.9% Top-1. Zero-shot retrieval aligns EEG with CLIP image embeddings.

Introduction & Motivation

• Challenge: EEG is inherently non-stationary and subject-specific; fixed kernels generalize poorly.
• Uniform attention treats all electrodes/time equally → adds noise.
• Prior GAT modules are decoupled from temporal/semantic branches.

Goal: Jointly model trial-specific dynamics, spatial dependencies, and semantic alignment for robust cross-subject decoding.

Method Overview



Pipeline: Overview of the proposed ECHO-Net framework, integrating dynamic convolution, attention, and contrastive learning. The architecture highlights modular interactions between spatial encoding, dynamic filtering, and multimodal alignment.

Dynamic Convolution (PEDC):
• Split EEG into temporal patches; assemble kernel $K(\mathbf{X}) = \sum_{i=1}^{N} \alpha_i(\mathbf{X}) K_i,$ via input-conditioned weights.
• 2D kernels over channel×time capture trial-specific patterns without per-user fine-tuning.

$$\mathbf{X}_{\text{patch}} \in \mathbb{R}^{C\times P\times T'} = \text{PatchSplit}(\mathbf{X}), \quad \mathbf{Z} \in \mathbb{R}^{C\times D} = \text{Proj}(\text{Pool}_{T'}(\mathbf{X}_{\text{patch}})),$$
$$K(\mathbf{X}) = \sum_{i=1}^{N} \alpha_i(\mathbf{X}) K_i, \quad \text{s.t.} \quad \sum_{i=1}^{N} \alpha_i(\mathbf{X}) = 1, \quad \alpha_i(\mathbf{X}) = \text{softmax}(\mathbf{W}_i(\text{GAP}(\mathbf{X})) + b_i),$$

Channel–Filter Attention:
• Channel attention highlights informative electrodes (global pooling + MLP gating).
• Filter attention emphasizes salient temporal segments (scaled dot-product).

$$\boldsymbol{\alpha}_c = \sigma(W_2\,\delta(W_1\,\text{GAP}(\mathbf{Z}_{\text{EEG}}))), \quad \mathbf{Z}'_{\text{EEG}} = \mathbf{Z}_{\text{EEG}} \odot \boldsymbol{\alpha}_c,$$
$$\boldsymbol{\alpha}_f = \text{softmax}\left(\frac{\mathbf{Q}(\mathbf{Z}_{\text{EEG}})\cdot\mathbf{K}(\mathbf{Z}_{\text{EEG}})^\top}{\sqrt{d}}\right), \quad \mathbf{Z}'_{\text{EEG}} = \mathbf{Z}_{\text{EEG}} \odot \boldsymbol{\alpha}_f,$$

$$\mathcal{L} = -\frac{1}{N}\sum_i \log \frac{\exp(\text{sim}(g_{\text{EEG}}(\mathbf{Z}^{(i)}_{\text{EEG}}), g_{\text{IMG}}(\mathbf{Z}^{(i)}_{\text{IMG}}))/\tau)}{\sum_j \exp(\text{sim}(g_{\text{EEG}}(\mathbf{Z}^{(i)}_{\text{EEG}}), g_{\text{IMG}}(\mathbf{Z}^{(j)}_{\text{IMG}}))/\tau)},$$

EEG-GAT Integration:
• Fully connected 64-node electrode graph; attention-based message passing interleaved with backbone.
• Joint spatio-temporal-semantic modeling → spatial consistency across subjects.

$$\mathbf{x}'_i = \sigma\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij}\mathbf{W}\mathbf{x}_j\right), \quad \alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^T[\mathbf{W}\mathbf{x}_i || \mathbf{W}\mathbf{x}_j]\right)\right)}{\sum_{k\in\mathcal{N}(i)}\exp\left(\text{LeakyReLU}\left(\mathbf{a}^T[\mathbf{W}\mathbf{x}_i || \mathbf{W}\mathbf{x}_k]\right)\right)}.$$

Experimental Setup

Datasets: THINGS-EEG (10 subjects, 200-way test); THINGS-MEG (4 subjects, 200-way test).
Training: PyTorch; Adam 2e-4; cosine annealing; batch 512; 200 epochs; seed=42.
Setting: Zero-shot—align EEG to CLIP image features; Top-k retrieval.
Metrics: Top-1/Top-5; paired t-tests for significance ($p<0.05/0.01/0.001$).

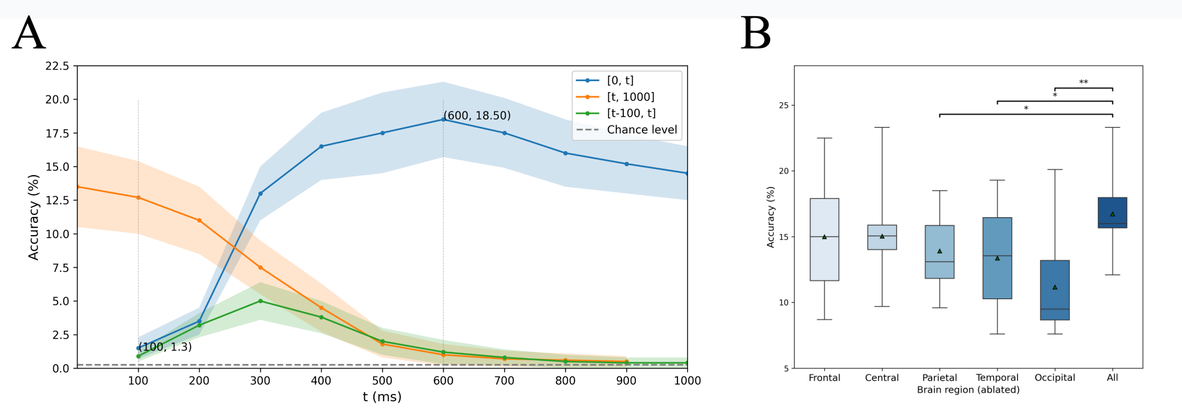
Key Results

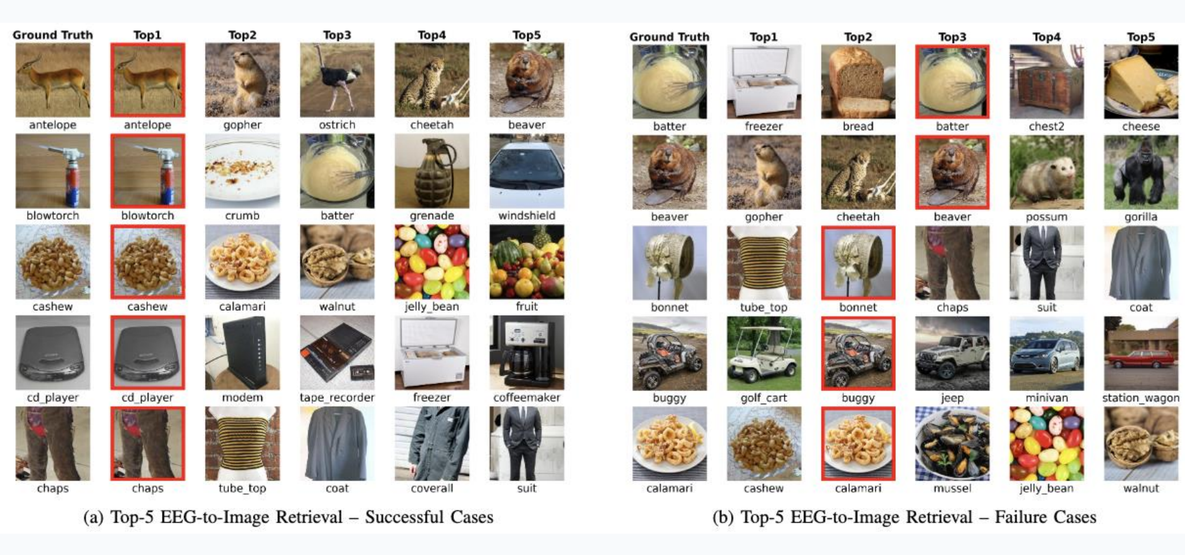
Overall Results (THINGS-EEG, subject-dependent):
• ECHO-Net: 18.5% Top-1 / 44.1% Top-5 (best).
• +2.9% Top-1 over NICE-GA; significant ($p<0.01$).

Subject-independent (LOSO):
• ECHO-Net: 9.1% Top-1 / 25.3% Top-5; above NICE (6.2/21.4) and NICE-GA (5.9/21.6).

THINGS-MEG:
• Best subject-dependent Top-1 16.2% / Top-5 45.4%.

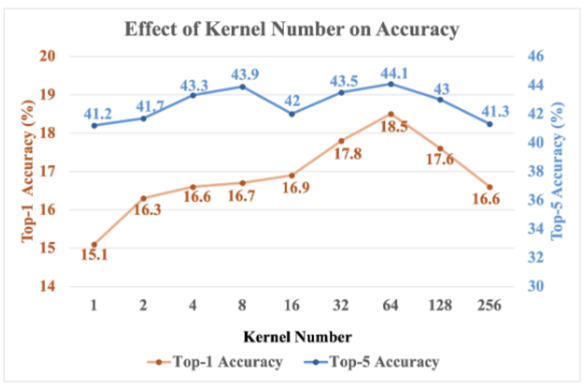
Method	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Subject 8		Subject 9		Subject 10		Avg	
	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5
Subject dependent - train and test on one subject																						
BraVL [8]	6.1	17.9	4.9	14.9	5.6	17.4	5.0	15.1	4.0	13.4	6.0	18.2	6.5	20.4	8.8	23.7	4.3	14.0	7.0	19.7	5.8	17.5
NICE [26]	12.3	36.6	10.4	33.9	13.1	39.0	16.4	47.0	8.0	26.9	14.1	40.6	15.2	42.1	20.0	49.9	13.3	37.1	14.9	41.9	13.8	39.5
NICE-SA [26]	13.3	40.2	12.1	36.1	15.3	39.6	15.9	49.0	9.8	34.4	14.2	42.4	17.9	43.6	18.2	50.2	14.4	38.7	16.0	42.8	14.7	41.7
NICE-GA [26]	15.2	40.1	13.9	40.1	14.7	42.7	17.6	48.9	9.0	29.7	16.4	44.4	19.1	43.1	20.3	52.1	14.1	39.7	19.6	46.7	15.6	42.8
ECHO-Net	17.0	43.0	20.5	41.0	16.5	45.0	19.0	49.5	14.0	32.5	20.0	44.0	20.5	47.5	24.5	51.5	14.5	38.5	18.0	48.5	18.5	44.1
Subject independent - leave one subject out for test																						
BraVL	2.3	8.0	1.5	6.3	1.4	5.9	1.7	6.7	1.5	5.6	1.8	7.2	2.1	8.1	2.2	7.6	1.6	6.4	2.3	8.5	1.8	7.0
NICE	7.6	22.8	5.9	20.5	6.0	22.3	6.3	20.7	4.4	18.3	5.6	22.2	5.6	19.7	6.3	22.0	5.7	17.6	8.4	28.3	6.2	21.4
NICE-SA	7.0	22.6	6.6	23.2	7.5	23.7	5.4	21.4	6.4	22.2	7.5	22.5	3.8	19.1	8.5	24.4	7.4	22.3	9.8	29.6	7.0	23.1
NICE-GA	5.9	21.4	6.4	22.7	5.5	20.1	6.1	21.0	4.7	19.5	6.2	22.5	5.9	19.1	7.3	25.3	4.8	18.3	6.2	26.3	5.9	21.6
ECHO-Net	7.3	24.7	8.0	24.9	8.7	26.4	6.6	23.8	7.6	23.9	9.0	24.3	5.0	21.1	9.9	26.2	8.8	24.8	11.5	32.4	9.1	25.3





(a) Top-5 EEG-to-Image Retrieval – Successful Cases (b) Top-5 EEG-to-Image Retrieval – Failure Cases

Ablation Study



• Dynamic conv improves over static (+1.0 Top-1).
• Each attention helps; Graph Attention yields largest Top-5 gains.
• Best with Channel + Filter + Graph → 18.5/44.1.
• Kernel count: gains up to 64; too many → overfitting.

Conclusion

Dynamic convolution + multimodal attention + EEG-GAT deliver robust cross-subject EEG decoding.
Future: cross-dataset transfer, efficiency, real-time deployment.