NEED: Cross-Subject and Cross-Task Generalization for Video and Image Reconstruction from EEG Signals

Shuai Huang · *Huan Luo · Haodong Jing · Qixian Zhang · Litao Chang · Yating Feng · Xiao Lin*

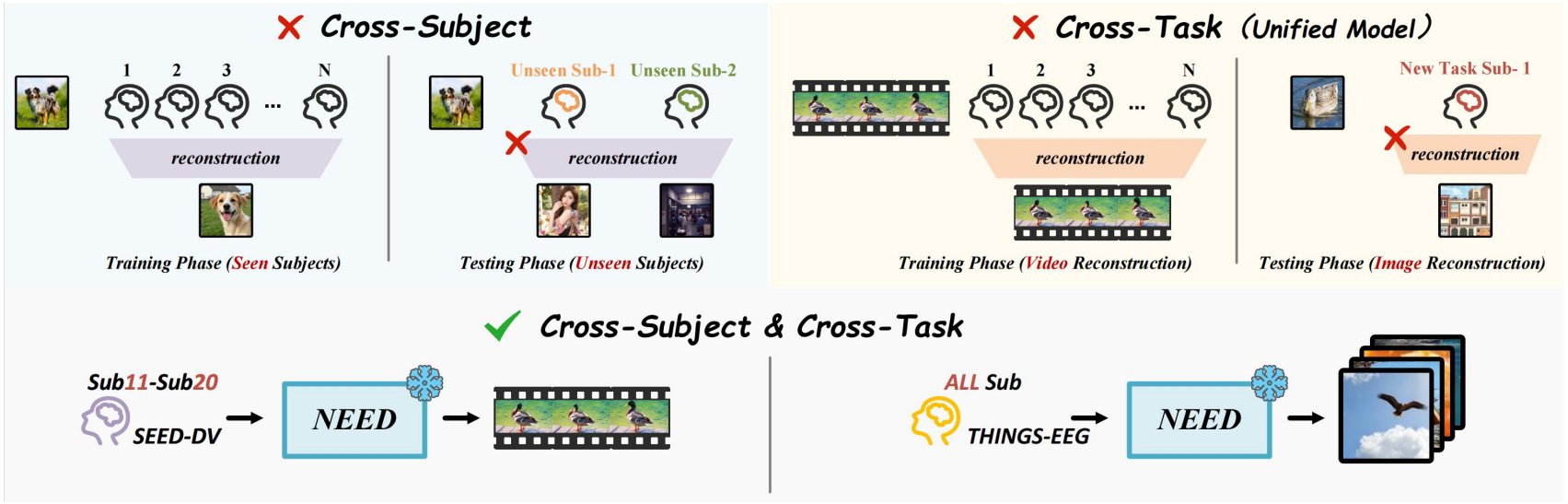
Chendong Qin · Han Chen · Shuwen Jia · Siyi Sun · Yongxiong Wang

Paper *Code* *Web* *WeChat*

NEURAL INFORMATION PROCESSING SYSTEMS

Paper Code Web WeChat

Research Background & Motivation



Why does this matter?

Brain-Computer Interfaces need to work:
→ Across different individuals (not just lab subjects)
→ For diverse visual experiences (not just one stimulus type)

UNIVERSITY OF OXFORD

NEURAL INFORMATION PROCESSING SYSTEMS

Paper Code Web WeChat

Main Contributions

First Unified Framework:

✓ Zero-shot cross-subject

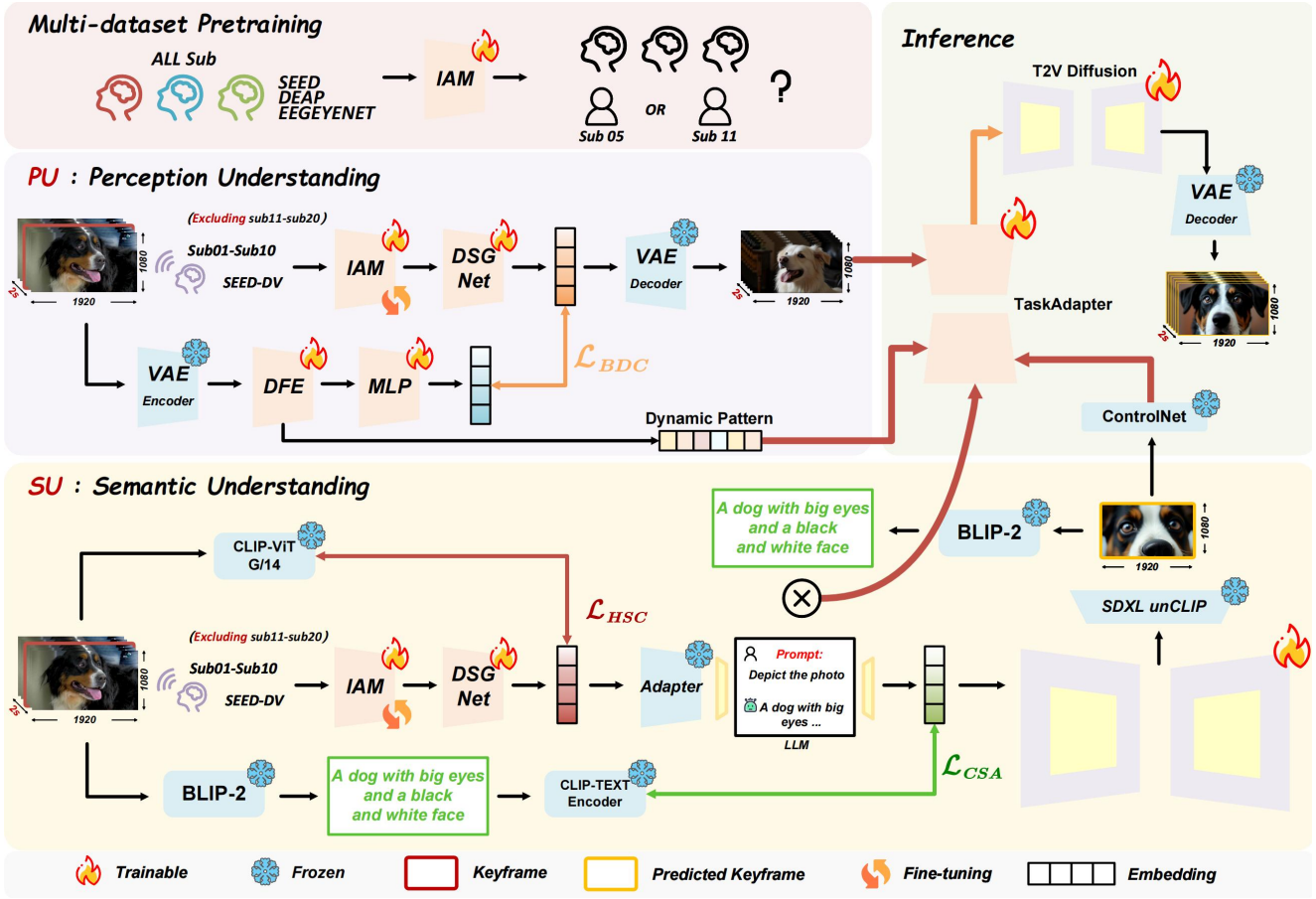
✓ Zero-shot cross-task

DSGNet:

Dual-stream EEG encoder

IAM Module:

Multi-dataset knowledge transfer

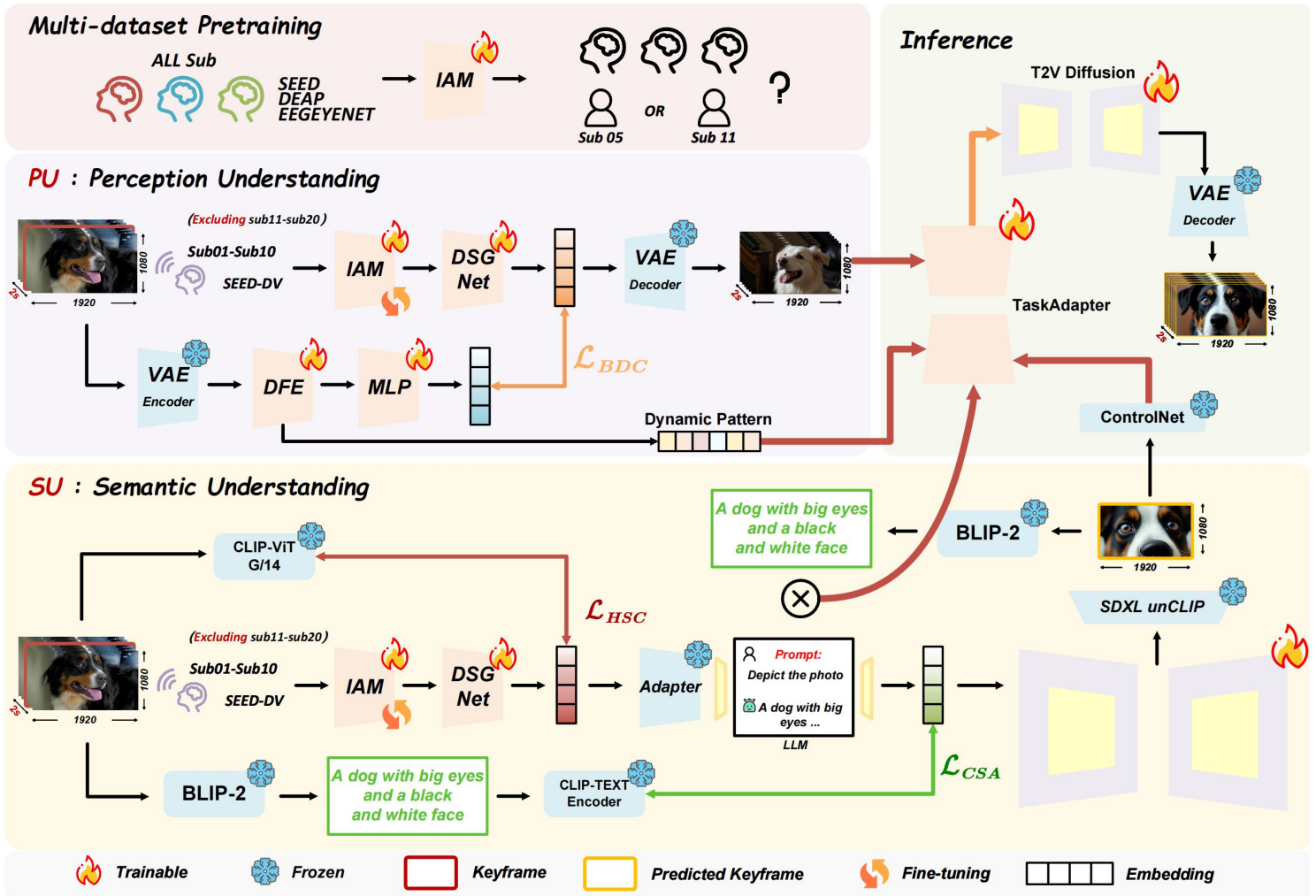


UNIVERSITY OF OXFORD

NEURAL INFORMATION PROCESSING SYSTEMS

Paper Code Web WeChat

Framework Architecture



Algorithm 1 Training and Inference of NEED Framework

Require:

Training: EEG dataset $\mathcal{D}_{train} = \{x_i, v_i, d_i\}$, auxiliary datasets $\mathcal{D}_{aux}$, pretrained models $T2I, B, C$

Inference: Unseen subject EEG x_{test}, task type $\tau_t \in \{\text{video}, \text{image}\}$

Ensure: Trained model components: $IAM, DSGNet, PU, SU$, Inference Mechanism

Training Stage:

Pretrain IAM on multiple datasets: $\mathcal{L}_{pretrain} = \alpha\mathcal{L}_{recon} + \beta\mathcal{L}_{adv} + \gamma D_{KL}$

for each $(x_i, v_i) \in \mathcal{D}_{train}$ **do**

  Process EEG: $Z_i = IAM(x_i)$, $Z_{DSG} = DSGNet(Z_i)$

end for

Train PU with Bidirectional Dynamic Contrastive loss:

$\mathcal{L}_{BDC} = -\sum_{i,j} w_{ij} \log \frac{\exp(sim(Z_0^i, V_1^j)/\tau)}{\sum_k \exp(sim(Z_0^i, V_1^k)/\tau)} + \lambda \|\nabla_t Z_0 - \nabla_t V_1\|_2^2$

Train SU with Hierarchical Semantic Contrastive loss:

$\mathcal{L}_{HSC} = \sum_{l=1}^{L} \alpha_l \cdot \left[-\log \frac{\exp(Z_s^l \cdot C_k^l/\tau)}{\sum_n \exp(Z_s^l \cdot C_n^l/\tau)} \right] + \beta \cdot \|G(Z_s) - G(C_k)\|_F$

Train unified inference with Cross-modal Alignment and Diffusion:

$\mathcal{L}_{CSA} = \mathcal{L}_{cont}(Z_3, T) + \gamma \cdot \mathcal{L}_{KL}(Z_3, T) + \delta \cdot \mathcal{L}_{align}(Z_3, \text{Keyframe})$

Learn task-specific parameters $\gamma_l(\tau_t)$ and $\beta_l(\tau_t)$ for conditional layer modulation

$h_l(x_t, t, c_t, \tau_t) = \gamma_l(\tau_t) \odot \text{Norm}(h_{l-1}) + \beta_l(\tau_t)$

Inference Stage:

Process test EEG: $Z_{test} = IAM(x_{test})$, $Z_{DSG} = DSGNet(Z_{test})$

Extract perception features: $Z_p = PU(Z_{DSG})$

Extract semantic features: $Z_s = SU(Z_{DSG})$

Apply task adaptation: $Z_{task} = \text{TaskAdapter}([Z_p; Z_s], \tau_t)$

Configure conditional weights based on task: $\alpha_{temp} = \sigma(W_\tau \cdot \tau_t)$, $\alpha_{spat} = \sigma(W_\tau \cdot \tau_t)$

Generate integrated conditioning:

$c_t = \alpha_t \cdot \text{ControlNet}(\text{Keyframe}_{pred}) + \beta_t \cdot \text{TextEmbed}(\text{Caption}) + \gamma_t \cdot \text{DynPattern} + \delta_t \cdot Z_{eeg}$

Initialize: $x_T \sim \mathcal{N}(0, I)$

for $t = T$ to 1 **do**

  Apply task-adaptive denoising: $x_{t-1} = \mu_\theta(x_t, t, c_t, \tau_t) + \sigma_t \cdot z$

  Apply conditional layer modulation: $h_l(x_t, t, c_t, \tau_t) = \gamma_l(\tau_t) \odot \text{Norm}(h_{l-1}) + \beta_l(\tau_t)$

end for

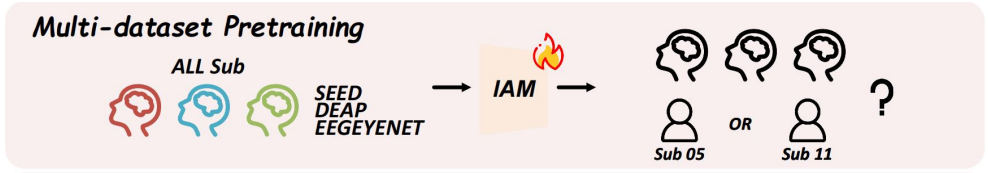
Return: Reconstructed visual content x_0 (video or image based on τ_t)

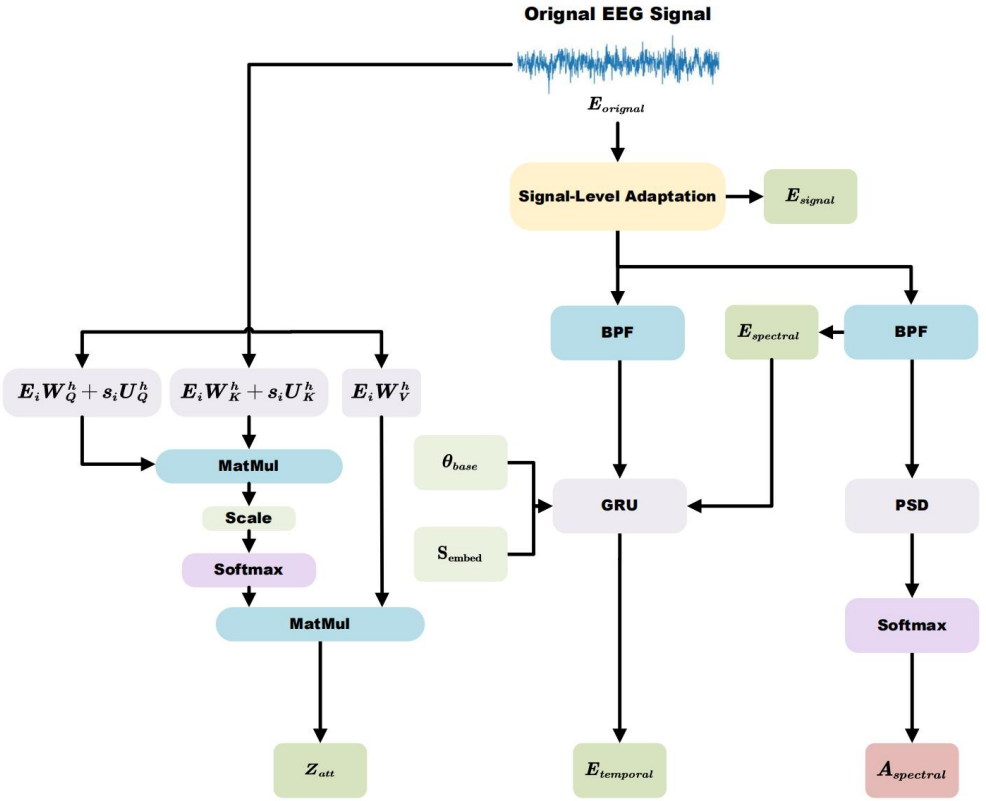
UNIVERSITY OF OXFORD

NEURAL INFORMATION PROCESSING SYSTEMS

Paper Code Web WeChat

Individual Adaptation Module (IAM)





For a given subject i, their EEG representation $E_i \in \mathbb{R}^{T\times d}$ is processed through a multi-head attention mechanism augmented by a learnable subject embedding $s_i \in \mathbb{R}^{d_s}$:

$$Z_{norm} = \underbrace{\sum_{h=1}^{H} W_O^h \left(\sigma \left(\frac{(E_iW_Q^h + s_iU_Q^h)(E_iW_K^h + s_iU_K^h)^T}{\sqrt{d_k}} \right) (E_iW_V^h) \right)}_{\text{Attention output } Z_{attn}} + \underbrace{\Gamma_\phi(E_i, \Omega_S, d_j)}_{\text{Style injection}} \quad (1)$$

where Γ_ϕ applies feature modulation guided by subject-agnostic knowledge Ω_S and dataset-specific embedding d_j to accommodate different experimental paradigms. To enhance cross-subject generalization, IAM incorporates a hierarchical adaptation strategy that progressively normalizes signals from low-level characteristics to high-level patterns with $Z_{final} = \alpha_1 Z_{signal} + \alpha_2 Z_{spectral} + \alpha_3 Z_{temporal} + \alpha_4 Z_{norm}$, where coefficients are learned adaptively.

Multi-Dataset Knowledge Transfer IAM is pretrained on multiple diverse EEG datasets—SEED, DEAP, and EEGEYENET—to capture generalized neural response patterns across subjects with varying cognitive states. During pretraining, we optimize a multi-objective loss function$\mathcal{L}_{pretrain} = \alpha\mathcal{L}_{recon} + \beta\mathcal{L}_{adv}(\theta_D, \theta_G) + \gamma D_{KL}(p(Z_{norm}|S) \parallel p(Z_{norm}))$ where $\mathcal{L}_{recon}$ ensures reconstruction quality, $\mathcal{L}_{adv}$ implements adversarial subject disentanglement, and the KL divergence term minimizes subject-specific information in the representations.

UNIVERSITY OF OXFORD

NEURAL INFORMATION PROCESSING SYSTEMS

Paper Code Web WeChat

Dual-Stream Encoder (DSGNet)

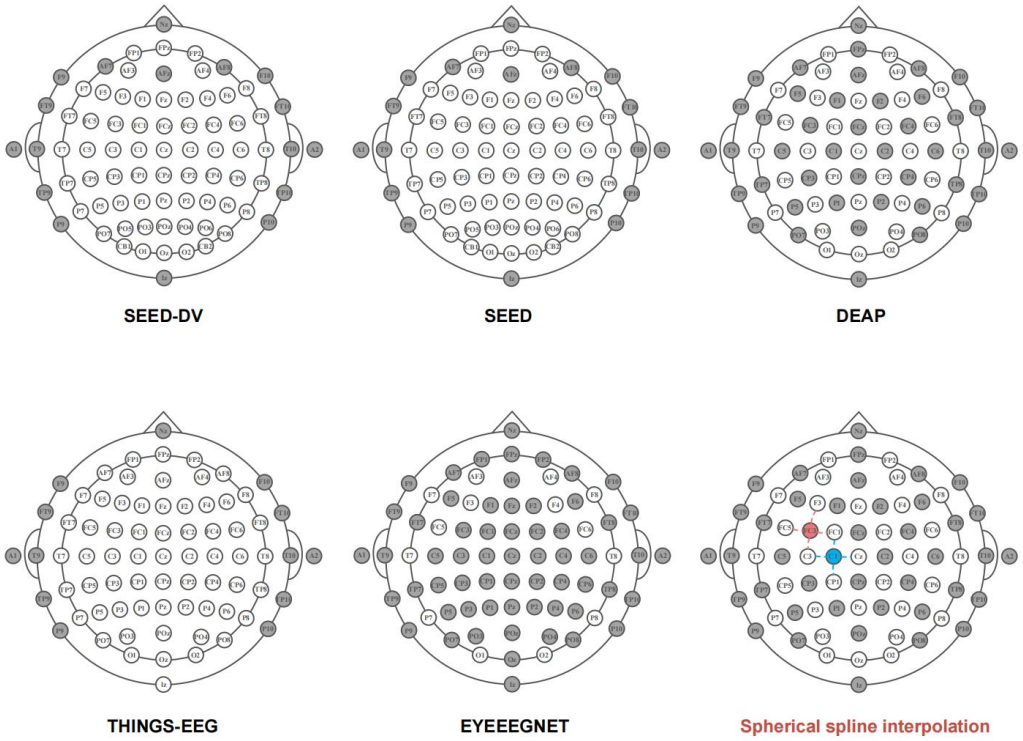
Spatial Stream Processing The spatial stream leverages the inherent topographical structure of EEG recordings by representing electrode positions in a spherical coordinate system. We employ spherical harmonic embeddings to capture spatial patterns $E^{(0)}_{spatial} = \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{l,m} Y_l^m(\theta_i, \phi_i)$ where Y_l^m represent spherical harmonic functions and $c_{l,m}$ are learnable coefficients. These spatial embeddings are further refined through a graph convolutional network that models inter-electrode relationships via $E^{(l+1)}_{spatial} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}E^{(l)}_{spatial}W^{(l)})$, where $\tilde{A} = A + I_N$ incorporates self-connections into the adjacency matrix.

Temporal Stream Processing The temporal stream captures dynamic neural responses across multiple time scales using a generalized Riemann-Liouville fractional transformation:

$$\mathcal{R}^{\alpha}(\mathbf{X})_t = \frac{1}{\Gamma(\alpha)} \int_0^t \frac{\mathbf{X}(s)}{(t-s)^{1-\alpha}} ds, \quad \alpha \in \{\alpha_1, ..., \alpha_m\} \tag{2}$$

This multi-scale approach with parameters $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$ enables capturing both rapid neural responses and slower contextual dynamics. The transformed features are processed through dilated temporal convolutions with receptive field size $d^{(l)} = 2^{l-1}$ to model dependencies across various temporal ranges.

Cross-Stream Integration We implement a cross-stream attention mechanism that enables dynamic interaction between spatial and temporal representations $Z = \sigma(W_g \cdot [E_{spatial}; E_{temporal}; Z_{cross}] + b_g) \odot \tanh(W_f \cdot [E_{spatial}; E_{temporal}; Z_{cross}] + b_f)$ where Z_{cross} represents cross-attention features computed via $Z_{cross} = \text{MultiHead}(Q = E_{spatial}W^Q, K = E_{temporal}W^K, V = E_{temporal}W^V)$. To enhance robustness against common EEG artifacts, we apply multi-dimensional masking $\tilde{E} = E \odot M_{spatial} \odot M_{temporal} \odot M_{frequency}$.



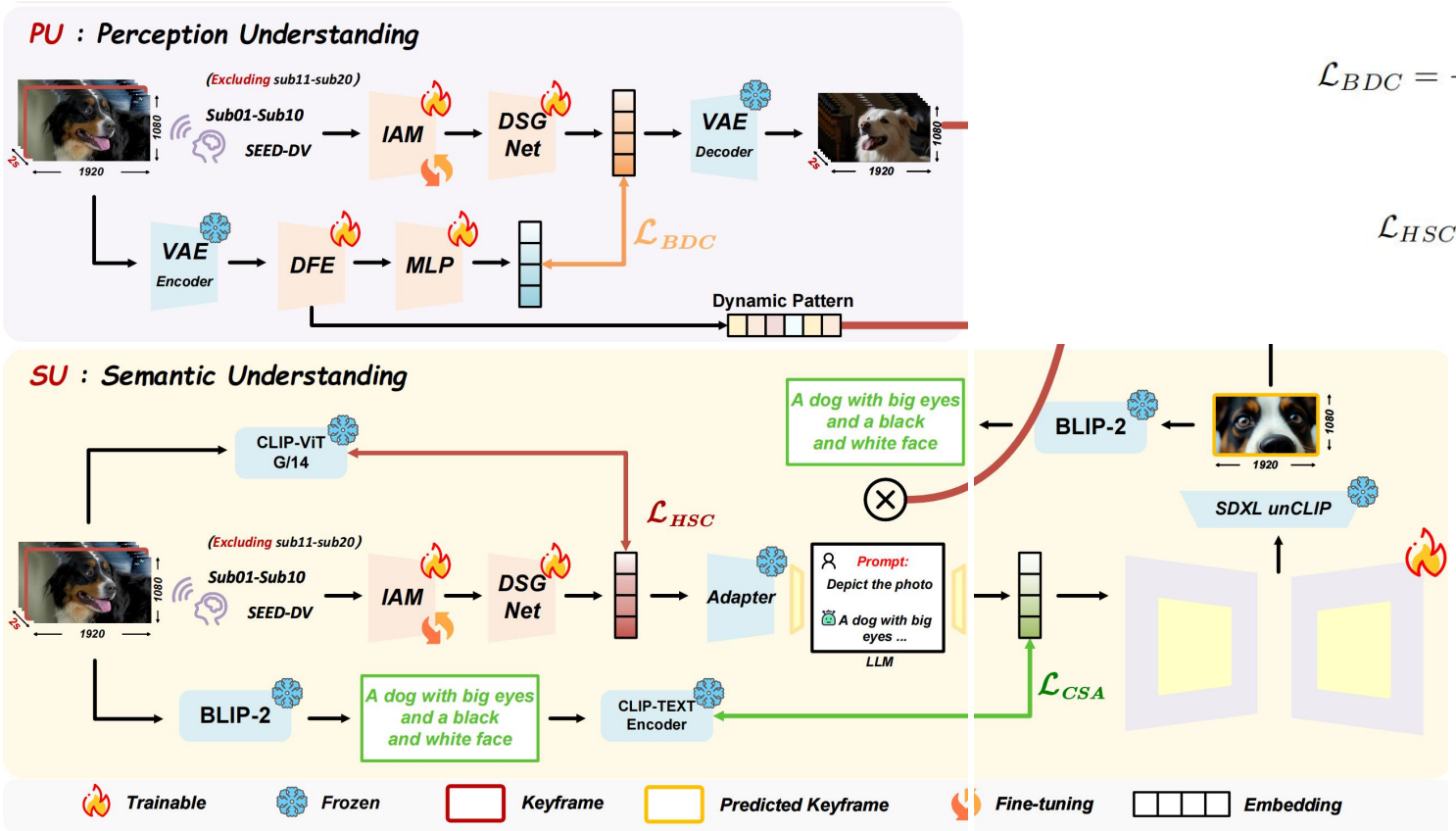
UNIVERSITY OF OXFORD

NEURAL INFORMATION PROCESSING SYSTEMS

Paper Code Web WeChat

Dual-Pathway Understanding

$$\begin{bmatrix} F_m \\ F_s \\ F_t \end{bmatrix} = \begin{bmatrix} \mathcal{F}_{motion}(\mathcal{T}_{3D}(V_0), \Theta_m) \\ \mathcal{F}_{spatial}(\mathcal{T}_{2D}(V_0), \Theta_s) \\ \text{SMARN}(\mathcal{T}_{TD}(V_0), H_{t-1}, \Theta_r) \end{bmatrix}$$



$$\mathcal{L}_{BDC} = -\sum_{i=1}^{N}\sum_{j=1}^{M} w_{ij} \log \frac{\exp(\text{sim}(Z_0^i, V_1^j)/\tau)}{\sum_k \exp(\text{sim}(Z_0^i, V_1^k)/\tau)} + \lambda \|\nabla_t Z_0 - \nabla_t V_1\|_2^2$$

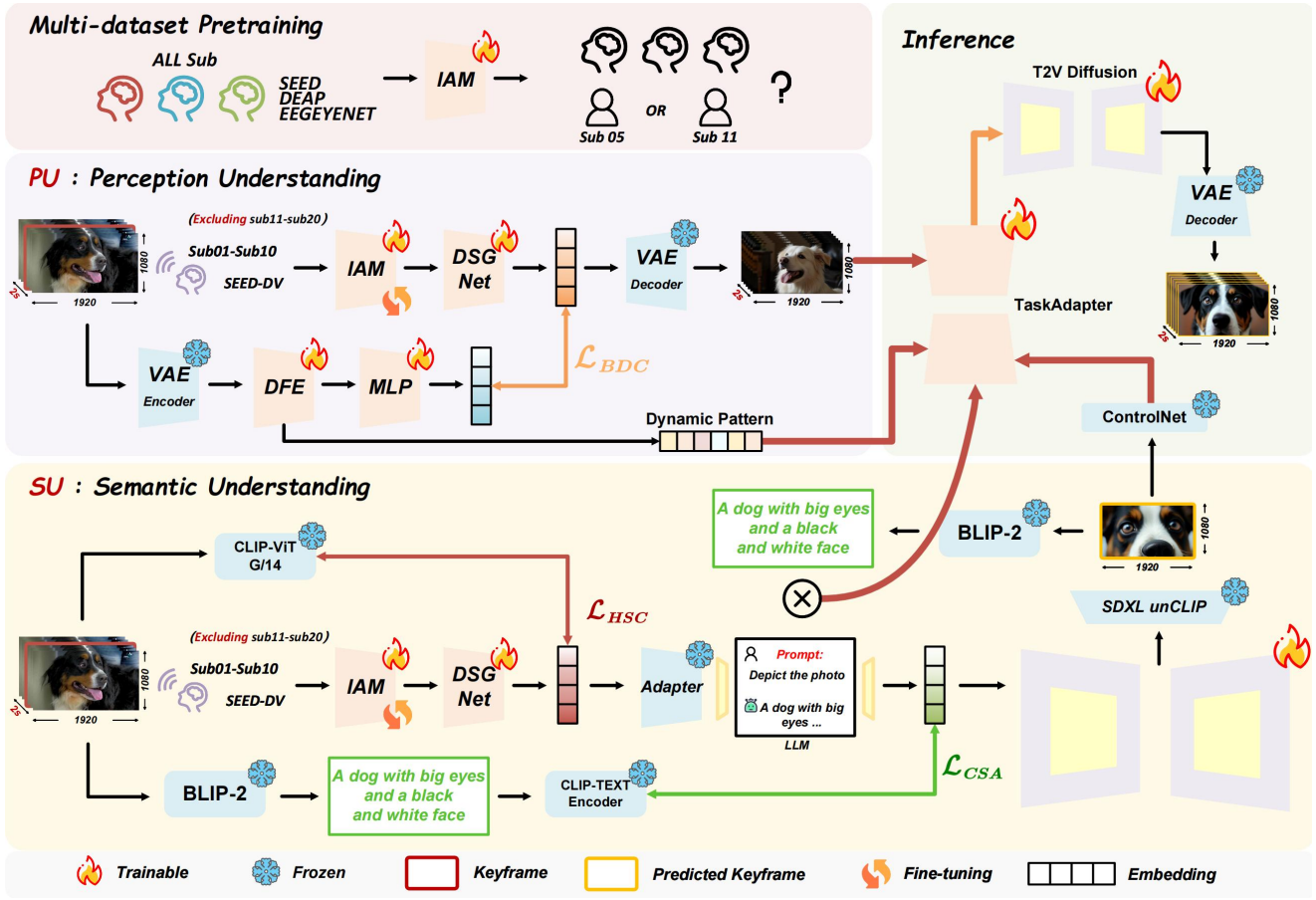
$$\mathcal{L}_{HSC} = \sum_{l=1}^{L} \alpha_l \cdot \left(-\log \frac{\exp(Z_s^l \cdot C_k^l/\tau)}{\sum_n \exp(Z_s^l \cdot C_n^l/\tau)} \right) + \beta \cdot \|G(Z_s) - G(C_k)\|_F$$

UNIVERSITY OF OXFORD

NEURAL INFORMATION PROCESSING SYSTEMS

Paper Code Web WeChat

Unified Inference Mechanism



Adaptive Task Conditioning We implement an adaptive conditioning mechanism that modulates the reconstruction process based on the target task $Z_{task} = \text{TaskAdapter}([Z_p; Z_s], \tau_t)$, where τ_t indicates either video or image reconstruction. This mechanism automatically adjusts the processing pipeline based on the reconstruction target while maintaining a shared parameter space.

Multi-Guidance Integration The integrated conditioning is computed as: $c_t = \alpha_t \cdot \text{ControlNet}(\text{Keyframe}_{\text{pred}}) + \beta_t \cdot \text{TextEmbed}(\text{Caption}_{\text{fusion}}) + \gamma_t \cdot \text{Dyn}_{\text{pattern}} + \delta_t \cdot Z_{\text{eeg}}$. where Caption_{fusion} is a fused textual description combining BLIP-2 generated captions from predicted keyframes and complementary semantic descriptions from EEG embeddings processed through a language model adapter. This dual-source caption fusion enhances semantic richness by incorporating both visual and neural information streams. The adaptive weights $[\alpha_t, \beta_t, \gamma_t, \delta_t] = \text{softmax}(\text{MLP}(z_t))$ dynamically adjust each guidance signal's contribution based on tasks.

Unified Cross-Task Generation Our framework achieves zero-shot cross-task generalization by leveraging shared neural patterns across visual modalities through a task-adaptive diffusion architecture. The diffusion process is defined as $x_T \sim \mathcal{N}(0, I)$, $x_{t-1} = \mu_\theta(x_t, t, c_t, \tau_t) + \sigma_t \cdot z$, where task encoding $\tau_t \in \{\tau_{video}, \tau_{image}\}$ explicitly differentiates between video and image processing, enabling a single model to adapt without structural modifications. For cross-task adaptation, we introduce conditional layer modulation $h_l(x_t, t, c_t, \tau_t) = \gamma_l(\tau_t) \odot \text{Norm}(h_{l-1}) + \beta_l(\tau_t)$, where $\gamma_l(\tau_t)$ and $\beta_l(\tau_t)$ dynamically transform feature spaces according to task requirements. This design creates a shared semantic representation space where $\mathcal{F}_{video}(E_{video}) \approx \mathcal{F}_{image}(E_{image})$. When processing image-related EEG signals, the model automatically configures the diffusion pathway by adjusting conditional signals while keeping all parameters θ fixed. Video reconstruction activates temporal attention modules with weight $\alpha_{temp} = \sigma(W_\tau \cdot \tau_{video})$, while image reconstruction shifts emphasis to spatial detail enhancement with $\alpha_{spat} = \sigma(W_\tau \cdot \tau_{image})$ while effectively zeroing temporal components $\alpha_{temp} \approx 0$. The model adapts feature representation distributions $p(Z|\tau_t)$, enabling conditional transformation $P(x_{t-1}|x_t, \tau_{video}) \neq P(x_{t-1}|x_t, \tau_{image})$ without architectural changes.

UNIVERSITY OF OXFORD

NEURAL INFORMATION PROCESSING SYSTEMS

Paper Code Web WeChat

Experimental Results

Methods		Multi-class Classification					Binary Classification			
		40-c top-1	40-c top-5	9-c top-1	9-c top-3	Color	Fast/Slow	Numbers	Human Face	Human
Chance level		2.50	12.50	11.11	33.33	20.57	50.00	65.64	62.25	71.43
EEG Features										
SVM (PSD)[58]		5.19/2.81*	-	19.02/3.27*	-	21.31/2.97	53.56/1.11*	64.15/1.22	58.94/2.21	70.91/1.84
MLP (PSD)		6.20/3.02*	18.91/5.94*	21.59/3.00*	49.86/3.78*	22.02/3.27	55.15/1.20*	64.48/0.92	63.94/1.13	71.74/1.76
GLMNet (PSD)[26]	*[NIPS'24]*	6.23/2.91*	18.98/5.62*	21.69/3.20*	50.03/4.10*	26.40/2.99*	55.42/1.32*	64.68/0.92	64.22/1.43	72.27/1.57
SVM (DE)[58]		4.82/2.80*	-	19.05/3.39*	-	21.07/2.88	53.34/1.25*	63.62/1.73	57.82/3.50	0.25/1.94
MLP (DE)		6.12/3.08*	19.02/5.71*	21.17/3.24*	49.40/4.94*	25.91/3.27*	54.76/1.25*	64.10/0.70	63.41/1.57	71.74/1.76
GLMNet (DE)[26]	*[NIPS'24]*	6.16/3.18*	19.12/6.07*	21.34/3.34*	49.55/4.57*	26.15/3.24*	55.06/1.20*	64.25/0.74	63.63/1.80	72.27/1.58
Raw EEG Signals										
ShallowNet[6]	*[HBM'17]*	5.59/2.27*	16.93/4.66*	21.40/1.96*	49.62/2.34*	27.00/2.09*	56.62/1.77*	66.15/0.89	64.87/1.54	73.21/1.52
DeepNet[6]	*[HBM'17]*	4.56/1.52*	14.30/3.25*	20.27/1.25*	48.06/1.59*	26.37/1.95*	55.42/0.59*	65.71/0.24	61.58/3.93	72.86/0.40
EEGNet[10]	*[JNE'18]*	4.64/0.86*	14.25/1.87*	19.63/0.81*	47.04/1.45*	25.46/1.31*	51.99/2.00	64.67/0.60	61.37/1.31	72.38/0.98
TSCNet[8]	*[SMC'20]*	6.84/2.15*	18.26/3.95*	22.18/1.76*	50.62/2.38*	27.85/1.63*	56.98/1.54*	66.54/0.87	65.42/1.35	73.51/1.28
Conformer[48]	*[TNSRE'22]*	4.93/1.57*	15.36/4.44*	20.92/0.98*	49.25/1.49*	27.53/1.37*	55.02/0.83*	65.73/0.26	64.96/1.14	73.00/0.85
BraVL[2]	*[TPAMI'23]*	7.05/2.23*	19.05/3.75*	22.53/1.84*	51.07/2.41*	28.12/1.52*	57.88/1.64*	66.92/0.83	65.76/1.38	73.65/1.23
EEGPT[13]	*[NIPS'24]*	6.95/2.63*	18.85/4.11*	22.10/1.96*	50.55/2.33*	27.95/1.73*	57.10/1.85*	66.75/0.94	65.35/1.42	73.25/1.33
TSConv[19]	*[ICLR'24]*	4.92/0.99*	15.05/2.31*	20.00/1.01*	47.76/1.51*	26.89/1.83*	55.32/0.99*	65.39/0.41	64.39/1.47	72.68/0.67
GLMNet[26]	*[NIPS'24]*	6.20/3.02*	17.75/4.24*	21.93/1.87*	50.01/2.52*	27.33/1.45*	57.35/1.98*	66.21/0.91	65.10/1.45	73.34/1.31
DSGNet (Ours)	*[This work]*	**14.23/3.12***	**25.75/4.44***	**29.93/2.07***	**58.01/2.72***	**35.33/1.65***	**65.35/2.18***	**74.21/1.11**	**73.10/1.65**	**81.34/1.51**
Relative Improvement of DSGNet over best baseline (%)										
		+101.8%	**+35.2%**	**+32.8%**	**+13.6%**	**+25.6%**	**+12.9%**	**+10.9%**	**+11.2%**	**+10.4%**

UNIVERSITY OF OXFORD

NEURAL INFORMATION PROCESSING SYSTEMS

Paper Code Web WeChat

Experimental Results

Model Variant	Video Reconstruction					Image Reconstruction				
	Video-based		Frame-based			Low-level		High-level		
	2-way↑	40-way↑	2-way↑	40-way↑	SSIM↑	PixCorr↑	SSIM↑	CLIP↑	Inception↑	SwAV↓
EEG2video (EEG)[26]	0.852±0.02	0.340±0.01	0.798±0.03	0.232±0.02	0.300±0.03	-	-	-	-	-
Wang (fMRI)[47]	0.773±0.03	-	0.713±0.04	-	0.118±0.08	-	-	-	-	-
Kupershmidt (fMRI)[46]	0.771±0.03	-	0.764±0.03	0.179±0.02 (50-way)	0.135±0.08	-	-	-	-	-
MinD-Video (fMRI)[30]	0.839±0.03	0.197±0.02 (50-way)	0.796±0.03	0.174±0.03 (50-way)	0.171±0.08	-	-	-	-	-
NeuroClips (fMRI)[45]	0.834±0.03	0.220±0.01 (50-way)	0.806±0.03	0.203±0.01 (50-way)	**0.390±0.08**	-	-	-	-	-
ATM-S (EEG)[45]	-	-	-	-	-	0.155	0.330	0.786	0.730	0.582
CognitionCapturer (EEG)[44]	-	-	-	-	-	0.150	0.347	0.715	0.669	0.580
Mind's Eye (fMRI)[25]	-	-	-	-	-	**0.295**	0.317	**0.918**	**0.929**	**0.382**
META-MEG (MEG)[40]	-	-	-	-	-	0.072	0.320	0.683	0.702	0.615
Full Model (Ours) (EEG)	**0.898±0.02**	**0.405±0.02**	**0.839±0.03**	**0.293±0.02**	0.356±0.02	0.162	**0.352**	0.795	0.742	0.575
w/o IAM	0.642±0.03	0.104±0.02	0.587±0.04	0.085±0.02	0.171±0.02	0.054	0.063	0.421	0.398	0.732
w/o PU	0.723±0.02	0.293±0.02	0.678±0.03	0.184±0.03	0.201±0.03	0.092	0.116	0.563	0.526	0.694
w/o SU	0.804±0.03	0.146±0.03	0.756±0.03	0.102±0.02	0.278±0.02	0.127	0.153	0.654	0.613	0.621
w/o Task Cond.	0.845±0.03	0.336±0.02	0.786±0.02	0.228±0.03	0.295±0.04	0.043	0.063	0.389	0.352	0.793
w/o DynPattern	0.836±0.02	0.327±0.02	0.743±0.04	0.217±0.03	0.284±0.03	0.109	0.124	0.623	0.575	0.659
w/o ControlNet	0.825±0.03	0.312±0.03	0.767±0.02	0.193±0.02	0.263±0.03	0.123	0.152	0.642	0.592	0.625
w/o Caption	0.831±0.02	0.298±0.02	0.775±0.03	0.205±0.03	0.271±0.02	0.115	0.146	0.607	0.584	0.638

Training Strategy	Video-based				Frame-based		
	Semantic-level		Temporal		Semantic-level		Pixel-level
	2-way Acc.↑	40-way Acc.↑	FVD↓	CLIP-pcc↑	2-way Acc.↑	40-way Acc.↑	SSIM↑
Within-subject (20/20)	0.898±0.02	0.405±0.02	175.3±8.6	0.738±0.04	0.839±0.03	0.293±0.02	0.356±0.02
1/20 subject	0.231±0.06	0.042±0.03	584.7±21.3	0.128±0.08	0.213±0.07	0.036±0.03	0.076±0.05
2/20 subjects	0.312±0.05	0.067±0.04	536.1±19.8	0.195±0.07	0.287±0.06	0.058±0.04	0.102±0.05
5/20 subjects	0.485±0.05	0.136±0.04	412.5±17.2	0.324±0.07	0.452±0.05	0.107±0.04	0.167±0.05
10/20 subjects	0.685±0.05	0.252±0.05	254.6±12.8	0.487±0.07	0.643±0.05	0.162±0.05	0.241±0.05
15/20 subjects	0.782±0.04	0.321±0.04	215.3±11.5	0.602±0.06	0.731±0.04	0.204±0.05	0.293±0.04
19/20 subjects	0.841±0.03	0.372±0.03	197.5±10.2	0.682±0.05	0.786±0.03	0.245±0.04	0.329±0.03
Performance retention (19/20)	93.7%	91.9%	88.8%	92.4%	93.7%	83.6%	92.4%
Performance retention (10/20)	76.3%	62.2%	68.9%	66.0%	76.6%	55.3%	67.7%
Performance retention (5/20)	54.0%	33.6%	42.5%	43.9%	53.9%	36.5%	46.9%
Performance retention (1/20)	25.7%	10.4%	30.0%	17.3%	25.4%	12.3%	21.3%

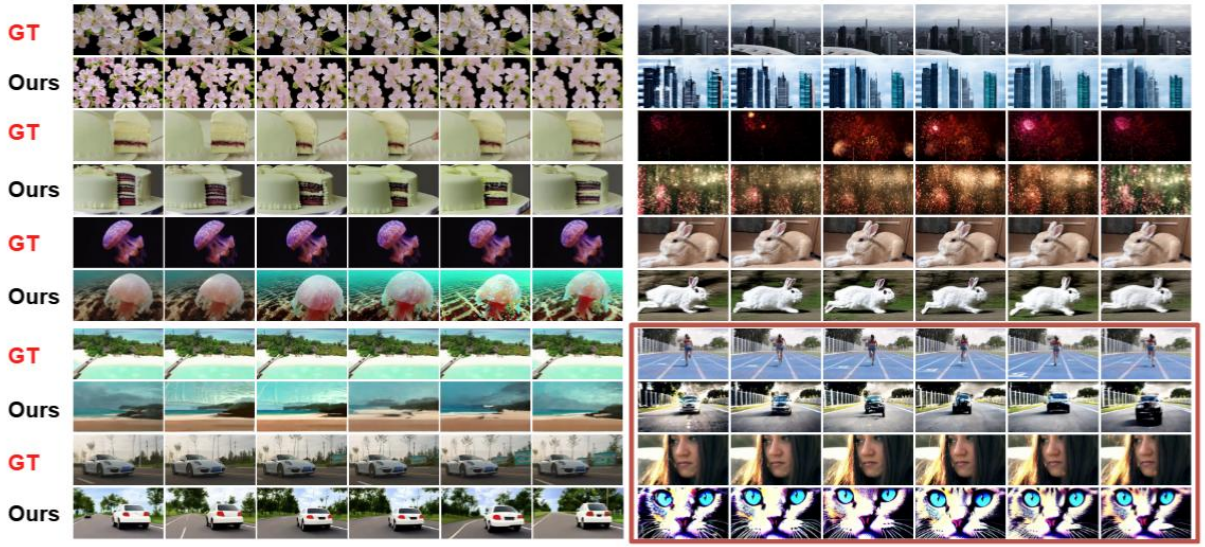
Training Subject	Video-based				Frame-based		
	Semantic-level		Temporal		Semantic-level		Pixel-level
	2-way Acc.	40-way Acc.	FVD	CLIP-pcc	2-way Acc.	40-way Acc.	SSIM
Subject 01	0.231	0.042	584.7	0.128	0.213	0.036	0.076
Subject 02	0.217	0.039	602.3	0.113	0.196	0.032	0.068
Subject 03	0.242	0.045	573.6	0.135	0.221	0.038	0.082
Subject 04	0.228	0.041	591.8	0.122	0.205	0.034	0.073
Subject 05	0.235	0.044	578.2	0.131	0.217	0.037	0.079
Subject 06	0.253	0.049	567.5	0.142	0.232	0.040	0.085
Subject 07	0.219	0.040	598.4	0.116	0.199	0.033	0.070
Subject 08	0.224	0.041	593.7	0.120	0.208	0.035	0.072
Subject 09	0.236	0.044	576.9	0.132	0.218	0.038	0.080
Subject 10	0.246	0.047	571.2	0.138	0.226	0.039	0.083
Subject 11	0.229	0.042	589.5	0.124	0.210	0.035	0.074
Subject 12	0.238	0.045	574.8	0.133	0.220	0.038	0.081
Subject 13	0.251	0.048	569.3	0.140	0.230	0.040	0.084
Subject 14	0.221	0.040	596.2	0.118	0.202	0.034	0.071
Subject 15	0.243	0.046	572.4	0.136	0.223	0.039	0.082
Subject 16	0.233	0.043	581.5	0.129	0.215	0.037	0.078
Subject 17	0.248	0.047	570.6	0.139	0.228	0.039	0.083
Subject 18	0.240	0.045	573.8	0.134	0.221	0.038	0.081
Subject 19	0.227	0.041	590.6	0.123	0.206	0.035	0.073
Average	0.235	0.044	582.5	0.129	0.215	0.037	0.077

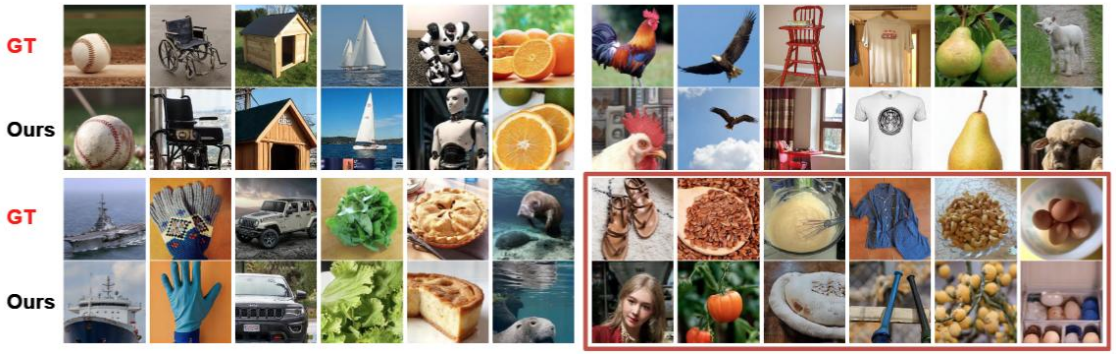
UNIVERSITY OF OXFORD

NEURAL INFORMATION PROCESSING SYSTEMS

Paper Code Web WeChat

Experimental Results



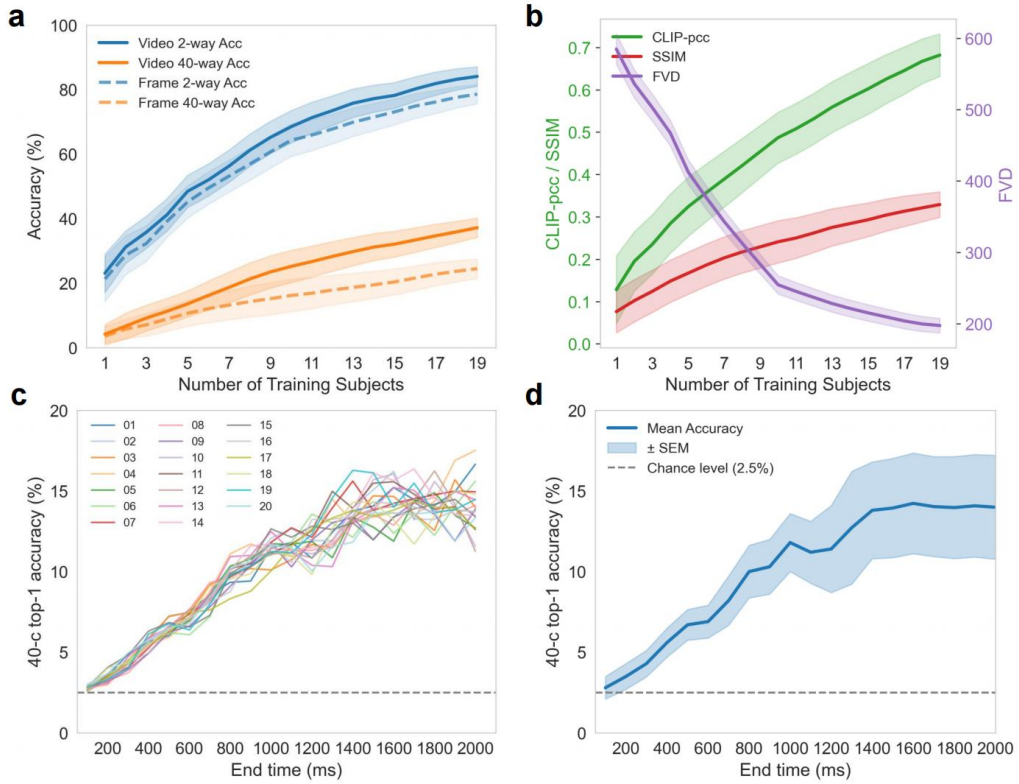


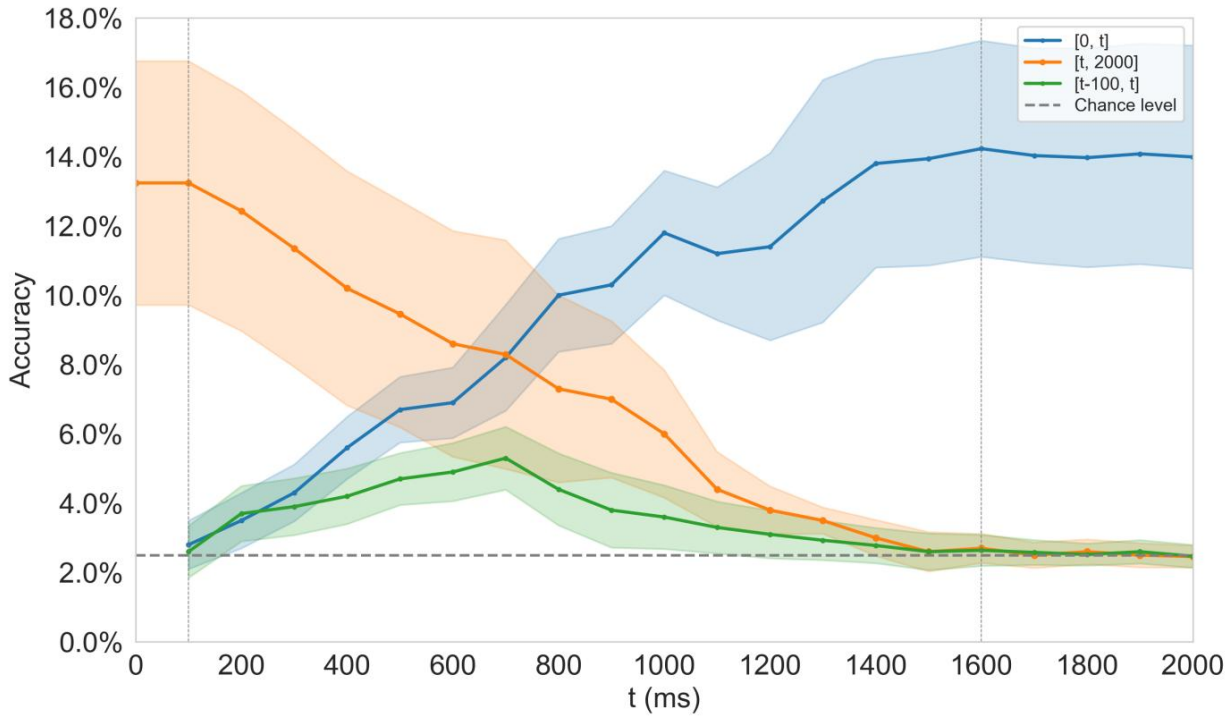
UNIVERSITY OF OXFORD

NEURAL INFORMATION PROCESSING SYSTEMS

Paper Code Web WeChat

Experimental Results



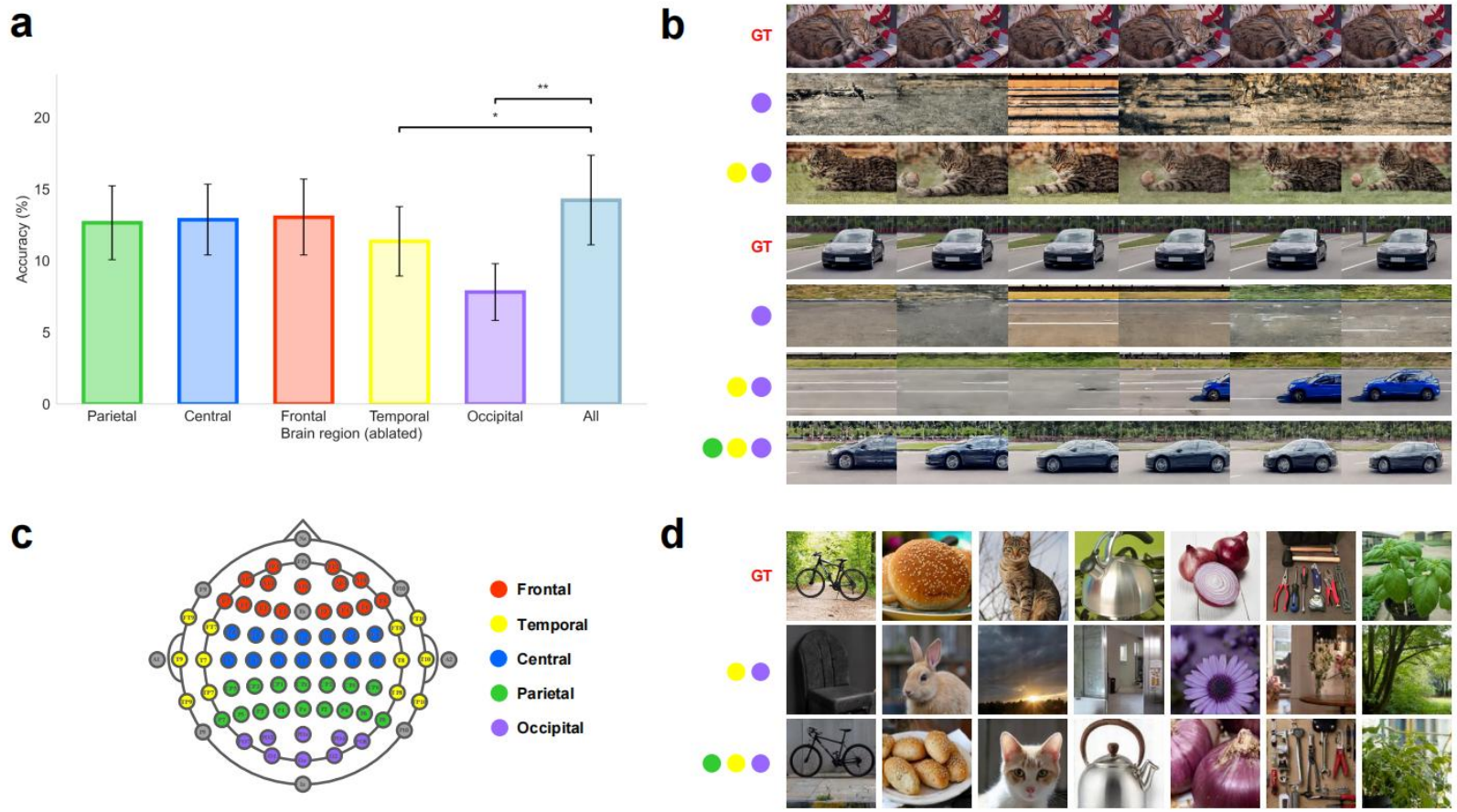


UNIVERSITY OF OXFORD

NEURAL INFORMATION PROCESSING SYSTEMS

Paper Code Web WeChat

Experimental Results



UNIVERSITY OF OXFORD

NEURAL INFORMATION PROCESSING SYSTEMS

Paper Code Web WeChat

Conclusion & Future Work

NEED Framework: Breaking the Boundaries

[Before]	[After]
Subject-Specific	Cross-Subject
Task-Specific	Cross-Task
Limited Scalability	Unified & Scalable

Achievements

- 93.7% cross-subject performance retention
- 92.4% cross-task quality retention
- Zero-shot generalization capability

Current Challenges

- Complex scene details
- High-density EEG requirement

Next Steps

- Paired static-dynamic dataset
- Enhanced reconstruction quality
- Portable device adaptation
- Real-world BCI applications

"Towards truly generalizable brain-computer interfaces"

UNIVERSITY OF OXFORD