RESHARE: REGION-STRUCTURED LEARNING FOR SHARED NEURAL VISUAL DECODING

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural visual decoding holds promise for brain-computer interfaces, but collecting neural recordings paired with visual stimuli is time-consuming. Many high-performing EEG and MEG decoding methods rely on participant-specific models, while existing shared decoders remain challenged by differences in neural response distributions across participants. These differences hinder the reuse of existing recordings when only limited training data are available from a new participant. We propose ReShare, a region-structured framework for shared visual decoding from EEG and MEG. ReShare uses EEG and MEG sensor layouts to define regional groups for spatial and temporal encoding. It then calibrates the encoded responses using learned, participant-specific scaling and offsets within each region. A network shared across participants integrates these calibrated responses and aligns its outputs with intermediate visual features from a frozen CLIP ViT-L/14 encoder. Fine-tuning the pretrained decoder with 10% of a new EEG participant's training pairs achieves 78.0% Top-1 accuracy. This exceeds the 75.0% achieved by a prior state-of-the-art participant-specific model trained on the same participant's full training set. With one shared model per dataset, ReShare also achieves state-of-the-art 200-way brain-to-image retrieval of unseen images from participants included in training. Top-1 accuracies reach 95.70% on THINGS-EEG2 and 61.88% on THINGS-MEG. These results show that region-structured shared learning supports accurate visual decoding while reducing the paired training data needed to adapt to a new participant.

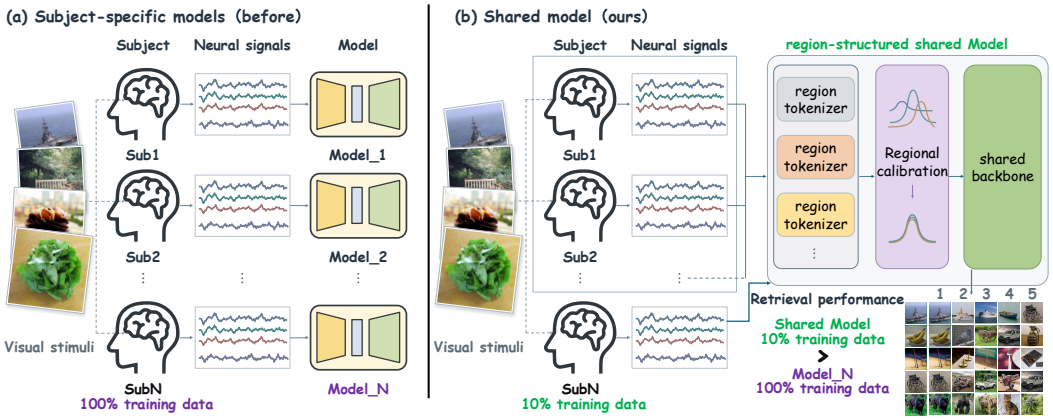


Figure 1: (a) Prior participant-specific decoders; (b) ReShare combines shared regional encoding and participant-specific calibration for decoding and adaptation with limited new-participant data.

1 INTRODUCTION

Visual decoding from non-invasive brain recordings is an important direction in brain-computer interface research (Wolpaw et al., 2002). Electroencephalography (EEG) and magnetoencephalography (MEG) capture visual responses with high temporal resolution (Cichy et al., 2014; Gifford et al.,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

2022; Hebart et al., 2023). Despite expanding natural-image resources (Hebart et al., 2019), paired neural recordings remain scarce, and response differences across participants complicate shared visual decoding (Zhou et al., 2026).

Contrastive learning advances retrieval by aligning neural responses with pretrained visual representations (Song et al., 2024; Li et al., 2024), with recent work refining targets through blur-aware and intermediate features (Liu et al., 2026; Du et al., 2026). Strong retrieval often relies on participant-specific decoders. Shared models reuse supervision across participants (Xia et al., 2024; Li et al., 2025a; Kneeland et al., 2026). In fMRI, shared pretraining already supports adaptation with limited new-participant data (Scotti et al., 2024). Extending this benefit to EEG and MEG requires accounting for individual differences in spatiotemporal responses.

Within an EEG or MEG dataset, participants share a consistent sensor layout but differ in their recorded responses. We use a topographic partition of this layout to give inputs a shared regional organization across participants. Participant-specific calibration adjusts the encoded regional responses before shared integration, reducing the impact of individual response differences on joint learning. Prior region-aware decoders emphasize spatial encoding or local neural–visual alignment (Zhou et al., 2026; Zhu et al., 2026), leaving the role of regional structure in participant-specific calibration less explored. Our design carries the same partition through regional encoding, participant calibration, and shared integration to improve the decoder's ability to learn from multiple participants.

We propose ReShare, a region-structured decoding framework with one model shared across participants within EEG or MEG modality (Figure 1). The EEG and MEG models are trained independently. Region-specific tokenizers shared across participants learn weighted combinations of multichannel signals over the full response window. A temporal encoder shared across regions and participants produces feature tokens, which undergo participant-specific regional calibration before shared integration. The outputs align with intermediate features of a frozen CLIP ViT-L/14 encoder, following prior evidence for intermediate visual supervision (Du et al., 2026).

Our contributions are threefold:

- **We design a region-structured interface for shared neural decoding.** We turn the topographic sensor partition into a persistent interface between neural recordings and the decoder. This interface provides a common organization for pooling supervision from multiple participants and specifies where participant-specific calibration is applied before features enter the shared network.
- **Participant-specific regional calibration for shared decoding.** We learn separate scales and offsets for each sensor region in each participant before features enter the shared network. These adjustments accommodate regional response differences within each participant. The resulting framework achieves state-of-the-art 200-way brain-to-image retrieval on THINGS-EEG2 and THINGS-MEG.
- **Shared pretraining for data-efficient adaptation to new participants.** We pretrain the model across participants. For a new participant, we fine-tune the full model using limited paired training data. Alternatively, we update only the input interface. Both routes outperform training from scratch in our EEG experiments with matched numbers of target training pairs, demonstrating effective reuse of recordings from previous participants.

2 Related Work

Visual supervision for neural decoding. Visual models provide neural decoding targets (Radford et al., 2021; Dosovitskiy et al., 2021) and show hierarchical correspondence with brain responses (Cichy et al., 2016). NICE uses contrastive EEG alignment (Song et al., 2024); Benchetrit et al. (2024) combine contrastive and regression objectives for MEG retrieval and reconstruction. Target design increasingly emphasizes features predictable from neural recordings. UBP and VisualEEG use blur-aware targets (Wu et al., 2025; Liu et al., 2026). Shallow Alignment selects intermediate features, HCF combines layers, and SAMGA learns subject-aware multi-granularity targets (Du et al., 2026; Tang et al., 2026; Jiang et al., 2026). Shared visual targets leave participant variability to the neural encoder. ReShare addresses these differences through regional calibration within a shared decoder trained to predict intermediate visual patch tokens.

Spatial and temporal structure in neural encoders. Predicting visual features requires an encoder that captures the spatial and temporal structure of multichannel recordings. EEGNet separates temporal and spatial filtering, while EEG Conformer adds self-attention to convolutional features (Lawhern et al., 2018; Song et al., 2023). EEGPT and CBraMod model spatial and temporal dependencies separately (Wang et al., 2024; 2025); LUNA compresses multichannel inputs into learned latents (Döner et al., 2025). For visual decoding, ATM combines channel attention with spatiotemporal convolutions (Li et al., 2024), EEGiT groups sensors into patches (Zhou et al., 2026), and RAHS aligns channel aggregates with visual sub-features (Zhu et al., 2026). Spatial encoding alone leaves individual regional responses uncalibrated; ReShare adds participant-specific calibration between temporal encoding and shared integration.

Shared decoding across participants. Functional alignment maps individual fMRI responses into shared representational spaces (Haxby et al., 2011; Chen et al., 2015). BIOT and LaBraM enable cross-dataset EEG pretraining with channel-based tokens (Yang et al., 2023; Jiang et al., 2024). In visual decoding, UMBRAE learns a shared fMRI mapping (Xia et al., 2024), and BrainFLORA couples modality-specific encoders to a universal projector (Li et al., 2025a). NEED incorporates individual adaptation into cross-subject reconstruction (Huang et al., 2025). In ENIGMA, participant-specific calibration follows global spatial mixing, leaving individual regional responses implicit in the compact representation (Kneeland et al., 2026). ReShare retains regional features for participant-specific FiLM (Perez et al., 2018) before integration in a shared latent integration network, using one sensor partition to coordinate encoding, calibration, and latent interactions.

Generalization and reuse for new participants. Transfer to new participants can use unsupervised EEG alignment (He & Wu, 2020) or adaptation with paired data. BENDR and REVE demonstrate transfer of pretrained EEG representations across tasks and recording setups (Kostas et al., 2021; El Ouahidi et al., 2025). Without paired target data, source-side selection and weight averaging limit overfitting (Gulrajani & Lopez-Paz, 2021; Cha et al., 2021). SATTC calibrates retrieval with frozen encoders using unlabeled target geometry (Huang & Zhu, 2026); SCORE estimates coordinate transformations from automatically selected EEG–image landmarks (Cui et al., 2026). When limited paired data are available, MindEye2 demonstrates effective fMRI adaptation after shared pretraining (Scotti et al., 2024), and ENIGMA extends data-efficient fine-tuning to EEG (Kneeland et al., 2026). How much of a pretrained decoder can remain fixed during regional input adaptation is less clear. ReShare measures the benefit of regional reuse by comparing full-decoder fine-tuning with interface-only adaptation that keeps the shared latent network fixed.

3 METHOD

3.1 PROBLEM STATEMENT AND REGIONAL STRUCTURE

We design ReShare to align neural recordings with visual features using a decoding framework shared across participants (Figure 2). For participant s, let $\mathbf{X} \in \mathbb{R}^{C \times T}$ denote a preprocessed response to an image, with C sensors and T time samples. The decoder predicts visual-feature tokens $\mathbf{Y} \in \mathbb{R}^{N \times D}$, where N is the number of tokens and D is their feature dimension. The regional encoder and latent integration network are shared across participants; regional calibration parameters are participant-specific. EEG and MEG models are trained separately with modality-specific sensor partitions and input dimensions.

ReShare organizes shared encoding and participant-specific calibration around a common regional structure, providing an inductive bias for more efficient shared training. Neighboring sensors capture related activity, so we organize the input into regional responses. Within each modality, a consistent sensor layout lets the same regional encoders learn from multiple participants at matched locations. Calibration adjusts these responses before shared latent integration, so participant differences stay local instead of perturbing the whole model.

3.2 REGIONAL TOKENIZATION AND PARTICIPANT CALIBRATION

We partition the sensors into K non-overlapping topographic regions shared across participants within each modality. Region k contains C_k sensors and produces n_k tokens, with $\sum_{k=1}^{K} n_k = N$.

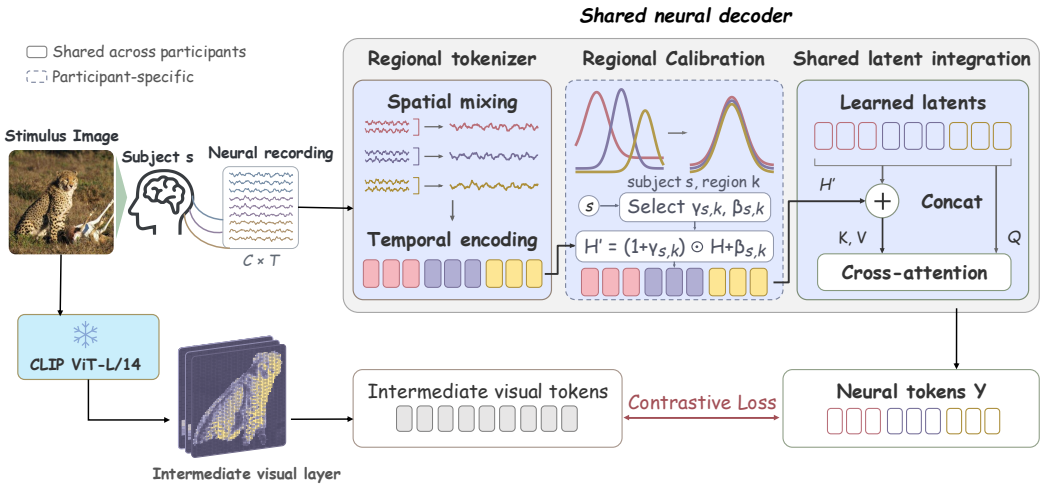


Figure 2: Overview of ReShare. EEG/MEG recordings are partitioned into topographic sensor groups. A region-specific tokenizer forms regional tokens, each retaining a full T-sample temporal trajectory, and participant-specific calibration adjusts the regional features. The calibrated tokens enter a shared latent integration network, where attention integrates information within and across regions. The outputs are aligned with intermediate features of a frozen CLIP ViT-L/14.

Its recording is $\mathbf{X}_k \in \mathbb{R}^{C_k \times T}$. A regional tokenizer forms n_k weighted combinations of these sensor signals through a learned spatial projection $\mathbf{W}_k \in \mathbb{R}^{n_k \times C_k}$,

$$\mathbf{Z}_k = \mathbf{W}_k \mathbf{X}_k. \tag{1}$$

Here $\mathbf{Z}_k \in \mathbb{R}^{n_k \times T}$ contains the regional token trajectories. Each region has its own projection, shared across participants and applied at every time point to preserve each token's full temporal response. The effective projection includes normalized spatial weights and learned token gains; its parameterization and token allocation are specified in Appendices B.1 and B.2.

A temporal encoder shared across regions and participants maps each trajectory independently to a D-dimensional feature. Let $\mathcal{E}$ denote this row-wise temporal mapping, implemented by a linear projection followed by nonlinear residual refinement and normalization. The regional features $\mathbf{H}_k \in \mathbb{R}^{n_k \times D}$ are

$$\mathbf{H}_k = \mathcal{E}(\mathbf{Z}_k) + \mathbf{P}_k, \tag{2}$$

where $\mathbf{P}_k \in \mathbb{R}^{n_k \times D}$ combines learned region and token-position embeddings. Temporal encoding preserves the regional grouping while mapping all trajectories to a common feature dimension. Before shared integration, a feature-wise affine transformation (Perez et al., 2018) calibrates these features for participant s, yielding $\mathbf{H}'_k$,

$$\mathbf{H}'_k = (1 + \boldsymbol{\gamma}_{s,k}) \odot \mathbf{H}_k + \boldsymbol{\beta}_{s,k}. \tag{3}$$

The vectors $\boldsymbol{\gamma}_{s,k}, \boldsymbol{\beta}_{s,k} \in \mathbb{R}^D$ specify scaling adjustments and offsets for this participant and region. Element-wise multiplication $\odot$ and the offset are broadcast across the region's tokens. These parameters are learned jointly with the shared decoder, allowing calibration to vary by region while preserving a common encoding network. Setting both vectors to zero gives identity calibration.

3.3 Shared Latent Integration

After calibration, regional responses enter a network shared across participants for integration. Concatenating the regional tokens gives $\mathbf{H}' \in \mathbb{R}^{N \times D}$, which supplies fixed neural context throughout the shared network. Each region initializes n_k learned latent vectors with region embeddings. Together, these form a latent sequence that gathers information from the calibrated responses through attention. For a representative cross-attention update, let $\mathbf{L} \in \mathbb{R}^{N \times D}$ denote the current latents and $\mathbf{L}'$ the updated latents,

$$\mathbf{L}' = \mathbf{L} + \mathrm{Attention}\big(\mathbf{L}, [\mathbf{H}'; \mathbf{L}], [\mathbf{H}'; \mathbf{L}]\big). \tag{4}$$

The three arguments provide queries, keys, and values; brackets denote token concatenation, corresponding to the Concat operation in Figure 2. The multi-head attention operator includes separate normalization of the neural and latent token banks and learned projections. Each latent can read from all calibrated regions and the current latent sequence. Successive updates refine the latents while keeping the neural context fixed.

Within each region, self-attention lets the region's latents exchange information. Global self-attention then connects latents across regions, combining regional evidence into a shared representation of the stimulus. Feed-forward transformations refine the integrated features, with a final mixture-of-experts router combining expert outputs according to each token's features (Shazeer et al., 2017). The final latents are normalized and mapped to the predicted visual tokens $\mathbf{Y}$. Each output token can draw on all regions to predict its corresponding image patch. The attention schedule and network dimensions are specified in Appendix B.2.

3.4 Visual Alignment and Model Reuse

Intermediate visual features provide effective alignment targets for neural decoding (Du et al., 2026). A frozen CLIP ViT-L/14 encoder (Radford et al., 2021) supplies an ordered intermediate patch sequence $\mathbf{V}_i \in \mathbb{R}^{N\times D}$ for image $\mathbf{I}_i$. The index i identifies one of B paired examples from the same participant; $\mathbf{Y}_i$ is predicted after regional dropout and input mixing. A learned tokenwise mapping, flattening, and ℓ_2 normalization of $\mathbf{Y}_i$ give the neural embedding $\mathbf{u}_i$; flattened, normalized $\mathbf{V}_i$ gives the image embedding $\mathbf{v}_i$. For neural input i and candidate image j, the cosine-similarity logit with temperature $\tau > 0$ is

$$a_{ij} = \frac{\mathbf{u}_i^\top \mathbf{v}_j}{\tau}. \tag{5}$$

BiMixCo (Scotti et al., 2023) assigns neural input i a target weight q_{ij} for image j. Unmixed inputs assign unit weight to their paired image; mixed inputs weight their two constituent images by the mixing coefficients. The bidirectional loss is

$$\mathcal{L}_{\text{con}} = -\frac{1}{2B}\sum_{i,j} q_{ij}\left(\log\frac{e^{a_{ij}}}{\sum_m e^{a_{im}}} + \log\frac{e^{a_{ij}}}{\sum_m e^{a_{mj}}}\right), \tag{6}$$

where all indices range from 1 to B; the summation index m runs over candidate images in the first denominator and neural inputs in the second.

Token regression compares the unprojected prediction $\mathbf{Y}_i$ with the original, unmixed visual target $\mathbf{V}_i$. The alignment objective is

$$\mathcal{L}_{\text{align}} = \mathcal{L}_{\text{con}} + \frac{\lambda_{\text{mse}}}{BND}\sum_i \|\mathbf{Y}_i - \mathbf{V}_i\|_{\text{F}}^2, \tag{7}$$

where λ_{mse} weights regression and the squared Frobenius norm $\|\cdot\|_{\text{F}}^2$ sums squared errors over token features. Dividing by BND gives mean squared error over B samples, N tokens, and D feature dimensions. Appendix B.3 details augmentation and the complete objective, including regularization of expert usage and sensor contributions. At inference, candidate images are ranked by cosine similarity.

For a new participant without paired data, we use identity regional calibration. With limited paired data, we fine-tune the decoder, retrieval mapping, and new participant's regional parameters while freezing the visual encoder (Scotti et al., 2024). Alternatively, we update only the input interface (regional encoders, input embeddings, and target calibration), keeping the shared latent network and retrieval mapping fixed.

4 Experiments

We first evaluate retrieval accuracy and the contribution of regional components, then test adaptation to new participants with limited EEG–image pairs. Neural-to-neural matching further assesses whether the shared representations preserve stimulus identity across participants. Finally, we vary the repeat count, defined as the number of test responses averaged per image, to assess retrieval with fewer repetitions.

4.1 Experimental Setup and Main Results

Datasets and inputs. THINGS-EEG2 contains ten participants, each with 16,540 training images and 200 test images from disjoint concepts (Gifford et al., 2022). EEG inputs average four training or 80 test responses per image. THINGS-MEG contains four participants (Hebart et al., 2023). Excluding the 200 test concepts leaves 19,848 training images, each presented once per participant; testing averages twelve responses for each of 200 images. Preprocessing is detailed in Appendix A.

Training. We train separate EEG and MEG models from random initialization. Within each modality, participants share the regional encoders and shared latent network, with participant-specific regional calibration. Visual supervision consists of ordered patch tokens from block 12 of frozen CLIP ViT-L/14. Regional dropout precedes within-participant input mixing, and BiMixCo contrastive supervision is used throughout training (Scotti et al., 2023). We optimize with AdamW (Loshchilov & Hutter, 2019), summing losses from two participant batches per update. Joint EEG and MEG models train for 500 and 200 epochs, respectively. Network dimensions, loss weights, optimizer settings, and LOSO checkpoint-selection rules are specified in Appendix B.

Evaluation and comparisons. Intra-subject evaluation uses jointly trained participants and unseen images; most reference methods train one model per participant. Leave-one-subject-out (LOSO) training uses nine EEG or three MEG participants, with identity FiLM and no labeled adaptation for the held-out target. Validation uses concepts held out from the source participants' training data; test responses are excluded from fitting and validation. Tables 1 and 2 report participant-averaged Top-1/Top-5 for bidirectional 200-way retrieval, with neural-to-image as the primary direction. Reference neural-to-image scores are published results under their reported settings; image-to-neural scores marked with a dagger are our checkpoint reevaluations in the corresponding setting, obtained from separate reporting runs. Reference settings, sources, results for each participant, and variability are provided in Appendix C.

Table 1: Participant-averaged bidirectional 200-way retrieval accuracy (%) on THINGS-EEG2, with column-wise best scores in bold.

	Intra-subject				Inter-subject			
	EEG → Image		Image → EEG		EEG → Image		Image → EEG	
Method	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
BraVL (2023)	5.80	17.50	–	–	1.80	7.00	–	–
NICE (2024)	16.10	43.60	–	–	6.20	21.40	–	–
ATM (2024)	27.10	58.10	–	–	11.90	33.80	–	–
Neural-MCRL (2025b)	32.40	64.10	–	–	14.00	34.30	–	–
UBP (2025)	50.90	79.70	–	–	12.40	33.40	–	–
NeuroBridge (2026)	63.20	89.90	–	–	19.00	45.90	–	–
EEGiT (2026)	70.40	95.10	–	–	24.00	55.60	–	–
HVE (2026)	75.70	94.60	74.25†	93.85†	20.00	44.10	24.70†	54.30†
VisualEEG (2026)	80.00	96.90	75.40†	94.80†	20.00	48.00	11.15†	32.95†
Shallow (2026)	82.60	97.70	88.65†	98.95†	21.80	49.40	32.70†	65.30†
ReShare	**95.70**	**99.80**	**98.45**	**100.00**	**26.75**	**56.00**	**39.40**	**70.85**

Main results. Joint training achieves 95.70/99.80% Top-1/Top-5 on EEG and 61.88/84.63% on MEG. Top-1 exceeds the results reported in the main tables of prior state-of-the-art retrieval work by 13.10 and 13.88 percentage points, respectively (Du et al., 2026). Jointly trained Shallow controls reach 80.25% EEG and 32.25% MEG Top-1 (Appendix D.4). LOSO retrieval achieves 26.75/56.00% on EEG and 6.25/18.00% on MEG, with EEG Top-1 exceeding EEGiT by 2.75 percentage points (Zhou et al., 2026).

Prior work has explored label-free test-time calibration for neural retrieval (Huang & Zhu, 2026). We separately evaluate a column-z variant on MEG, standardizing each image's scores across the unlabeled query batch using column means and standard deviations. Without paired labels or weight updates, it raises intra- and inter-subject Top-1 to 71.13% and 8.88%, respectively (Appendix C.3).

Table 2: Participant-averaged bidirectional 200-way retrieval accuracy (%) on THINGS-MEG, with column-wise best scores in bold above the separator and column-z calibration reported separately below.

	Intra-subject				Inter-subject			
	MEG → Image		Image → MEG		MEG → Image		Image → MEG	
Method	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
NICE (2024)	12.80	36.00	–	–	–	–	–	–
NICE-SA (2024)	12.70	35.00	–	–	–	–	–	–
NICE-GA (2024)	14.30	42.30	–	–	–	–	–	–
UBP (2025)	26.70	55.20	–	–	2.20	10.40	–	–
NeuroBridge (2026)	32.20	60.80	–	–	3.40	12.80	–	–
HVE (2026)	33.70	60.50	31.00†	57.88†	5.40	15.20	4.50†	16.38†
VisualEEG (2026)	44.00	72.00	–	–	5.30	15.90	–	–
Shallow (2026)	48.00	74.40	52.53†	77.75†	3.70	13.00	4.23†	14.03†
ReShare	**61.88**	**84.63**	**67.50**	**87.75**	**6.25**	**18.00**	**7.50**	**22.75**
ReShare + column-z	71.13	88.50	67.50	87.75	8.88	25.50	7.50	22.75

4.2 Ablation of Regional Organization

Table 3 evaluates progressive additions of regional tokenization, regional FiLM, regional latent integration, MoE routing, and region dropout to a non-regional baseline (B0–B5; definitions in Appendix D.1). Each increment is conditional on the preceding components. Results average ten THINGS-EEG2 participants, with B0–B5 additionally averaged over three training seeds.

Table 3: Neural-to-image retrieval accuracy (%) for progressive ablations on THINGS-EEG2, with B0–B5 averaged over three seeds and ten participants.

	Cumulative components					Neural → image	
Variant	Tokenizer	FiLM	Attention	MoE	Dropout	Top-1	Top-5
B0	–	–	–	–	–	93.90	99.70
B1	✓	–	–	–	–	94.05	99.70
B2	✓	✓	–	–	–	94.90	99.65
B3	✓	✓	✓	–	–	95.30	**99.85**
B4	✓	✓	✓	✓	–	95.35	99.65
B5 (ReShare)	✓	✓	✓	✓	✓	**95.70**	99.80

Regional calibration and regional attention contributed the largest Top-1 increments in the progressive sequence. In control experiments on the full model, regional FiLM achieved 95.70% Top-1, exceeding NoFiLM (91.85%) and global participant FiLM (93.25%) by 3.85 and 2.45 percentage points, respectively (Appendix D.2). Using unmixed inputs for regression and mixed inputs for contrastive learning yielded 94.00% (Appendix D.3).

4.3 Adapting a Shared Decoder with Limited Target Pairs

We tested limited-data reuse through paired scratch controls for EEG participant 1 and adaptation to all ten participants. Each budget started independently from a model pretrained on the other nine participants; target test pairs were excluded from adaptation. Each budget specifies the fraction of the target participant's paired EEG–image training examples used for adaptation (Appendix E.1). For participant 1, scratch training and adaptation used identical 10/30/50% subsets and matched update counts within each architecture (Appendix E.4). Shallow, VisualEEG, and ATM were evaluated using the same source participants and target subsets (Du et al., 2026; Liu et al., 2026; Li et al., 2024), with their own visual targets and optimization configurations.

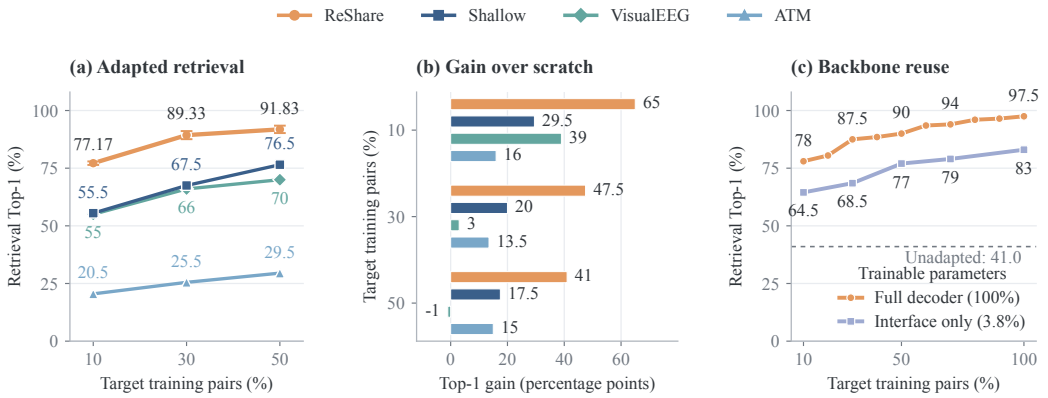


Figure 3: EEG participant 1 retrieval after shared pretraining (a), paired gains over training from scratch (b), and full-decoder versus interface-only adaptation (c), with mean ± SD across three runs for ReShare in (a).

Across three training seeds and subset draws, ReShare reached 77.17 ± 0.76%, 89.33 ± 1.76%, and 91.83 ± 1.61% Top-1 at 10/30/50% target data, respectively (Figure 3a; Appendix E.3). In the original paired controls, scratch training reached 13.0%, 40.0%, and 49.0% Top-1 using 10/30/50% of the target training pairs, respectively. Adaptation on the same subsets reached 78.0%, 87.5%, and 90.0%, yielding gains of 65.0, 47.5, and 41.0 percentage points (Figure 3b). At 10%, adapted Shallow, VisualEEG, and ATM achieved 55.5%, 55.0%, and 20.5% Top-1. Their gains over scratch were 29.5, 39.0, and 16.0 percentage points, respectively, compared with 65.0 for ReShare. Across all three budgets, ReShare achieved the highest adapted accuracy and largest gains over scratch, showing more effective reuse of multi-participant supervision in these comparisons.

The adaptation benefit extended across all ten held-out EEG participants. Using 10% of each target's training pairs increased mean Top-1 from 26.75% to 73.15 ± 6.24% and Top-5 from 56.00% to 95.35 ± 2.93%. Every participant improved, with a mean Top-1 gain of 46.40 percentage points (results for each participant in Appendix E.2). With matched numbers of target training pairs and updates, scratch training reached only 14.15% mean Top-1 and 37.20% Top-5 across the same ten targets. Adaptation outperformed scratch training for every participant, with a mean Top-1 gain of 59.00 percentage points (Appendix E.2).

Adaptation with a fixed shared latent network. We froze the shared latent integration network and retrieval projector, updating only the spatial and temporal encoders, embeddings, and target regional FiLM. This input interface contains approximately 3.8% of the parameters available to full adaptation. Interface-only updates achieved 64.5/77.0/83.0% Top-1 at 10/50/100% budgets, compared with 78.0/90.0/97.5% for full adaptation (Figure 3c). With 10% of the target training pairs, interface-only adaptation outperformed models trained from scratch on 30%, including ReShare (40.0%), Shallow (47.5%), VisualEEG (63.0%), and ATM (12.0%). The frozen shared latent integration network thus provided a reusable visual mapping, allowing interface-only adaptation to achieve higher accuracy with 10% of target training pairs than scratch training with 30%.

4.4 Shared Representations across Participants

To examine the shared representation beyond image retrieval, we tested whether unseen stimuli remained identifiable across different participants' recordings. We used *neural-to-neural matching*, without image features in the similarity computation. For each directed participant pair, 200 responses from one person queried 200 responses from the other; a correct match corresponded to the same image. Scores were averaged over all $10 \times 9 = 90$ EEG and $4 \times 3 = 12$ MEG directed participant pairs. The participants were included in joint training, whereas the images were unseen; this evaluation measures representation consistency, not LOSO transfer.

ReShare achieved 81.00% EEG Top-1, compared with 64.98% for shared Shallow, 49.14% for shared ATM, and 35.56% for shared VisualEEG (Figure 4a). These correspond to gains of 16.02, 31.86, and 45.44 percentage points, respectively. On MEG, it reached 30.42% versus 11.54% for shared Shallow, a gain of 18.88 percentage points (Figure 4b). These margins show more consistent preservation of stimulus identity across participants in both modalities.

Replacing regional FiLM with an identity transformation at otherwise fixed weights reduced Top-1 to 72.34% on EEG and 23.88% on MEG (Figure 4c). In EEG-to-image retrieval, FiLM parameters from another participant yielded 81.69% Top-1, compared with 95.70% for the correct participant and 93.30% with identity FiLM (Appendix D.5, Table 17). These interventions show that the fitted decoder uses individual calibration to preserve stimulus information across participants. They measure dependence on the learned calibration; Section 4.2 instead tests adding regional FiLM during training.

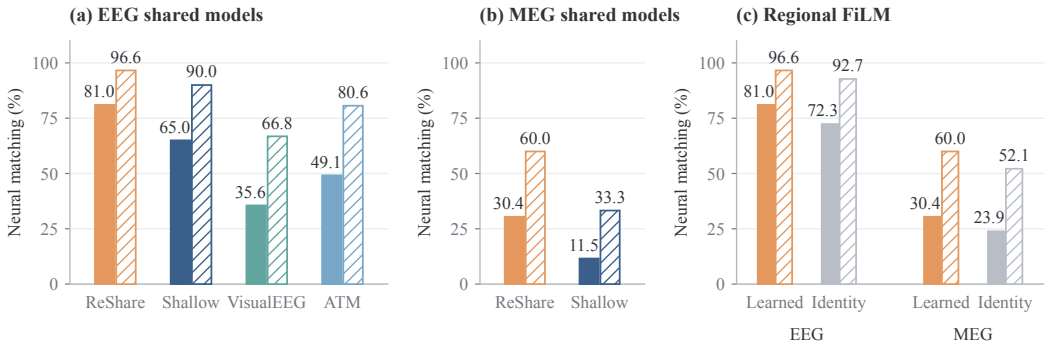


Figure 4: Neural-to-neural matching across EEG (a) and MEG (b) participants, and learned versus identity regional calibration (c), with solid and hatched bars showing correct matches ranked first (Top-1) and within the top five (Top-5), respectively.

Dependence on repeated measurements. We assessed retrieval at lower repeat counts with fixed weights. Single EEG responses yielded 20.70% Top-1, compared with 11.32% for individual Shallow and 14.51% for VisualEEG. At eight repetitions, the corresponding scores were 78.62%, 52.99%, and 55.33%. Single-response MEG retrieval reached 25.95%, versus 13.48% for individual Shallow (Appendix E.5). ReShare maintained the highest Top-1 accuracy at every evaluated repeat count on both datasets, extending its advantage from single responses to fully averaged inputs.

5 Conclusion

We presented ReShare, a region-structured framework for shared neural visual decoding across participants. A common sensor partition connects regional encoding, participant-specific calibration, and shared latent integration, allowing the model to accommodate individual responses while learning from pooled supervision. Experiments on THINGS-EEG2 and THINGS-MEG demonstrate accurate retrieval of unseen images from jointly trained participants and more consistent stimulus representations across their recordings. Regional ablations further support calibrating participant responses at the regional level before shared integration. Beyond joint decoding, pretraining provides a reusable initialization for new participants. Across ten held-out EEG participants, adaptation with 10% of each target's training pairs achieved 73.15% mean Top-1. Training from scratch reached 14.15% with matched numbers of target training pairs and updates. For EEG participant 1, interface-only adaptation improved retrieval with the pretrained shared latent integration network fixed. It updated approximately 3.8% of the parameters available to full adaptation. These findings support regional organization as a basis for reusing multi-participant supervision with limited target data. Future work will extend this framework across recording setups with different sensor layouts and acquisition conditions. We will also investigate training strategies that improve single-trial retrieval under noisy recordings and reduce dependence on repeated measurements.

AI USE STATEMENT

I used generative AI tools to assist with literature retrieval, translation and polishing of my drafts, and research execution. Research assistance included code implementation and debugging, and interpretation of experimental outputs. I determined the research objectives and made the final methodological and editorial decisions. Reported numerical results were obtained from the experiments described in this paper. I take responsibility for the final content, including AI-assisted text, code, analyses, and claims.

REFERENCES

Yohann Benchetrit, Hubert Banville, and Jean-Remi King. Brain decoding: Toward real-time reconstruction of visual perception. In *International Conference on Learning Representations*, 2024.

Junbum Cha, Sanghyuk Chun, Kyungjae Lee, Han-Cheol Cho, Seunghyun Park, Yunsung Lee, and Sungrae Park. SWAD: Domain generalization by seeking flat minima. In *Advances in Neural Information Processing Systems*, volume 34, pp. 22405–22418. Curran Associates, Inc., 2021.

Po-Hsuan (Cameron) Chen, Janice Chen, Yaara Yeshurun, Uri Hasson, James Haxby, and Peter J. Ramadge. A reduced-dimension fMRI shared response model. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

Radoslaw Martin Cichy, Dimitrios Pantazis, and Aude Oliva. Resolving human object recognition in space and time. *Nature Neuroscience*, 17(3):455–462, 2014.

Radoslaw Martin Cichy, Aditya Khosla, Dimitrios Pantazis, Antonio Torralba, and Aude Oliva. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, 6:27755, 2016.

Zhenyao Cui, Siyuan Kan, Siyang Li, Ziwei Wang, and Dongrui Wu. SCORE: Subject coordinate recovery for label-free cross-subject EEG-to-image retrieval. arXiv preprint arXiv:2608.19134, 2026.

Berkay Döner, Thorir Mar Ingolfsson, Luca Benini, and Yawei Li. LUNA: Efficient and topology-agnostic foundation model for EEG signal analysis. In *Advances in Neural Information Processing Systems*, volume 38, pp. 70682–70708. Curran Associates, Inc., 2025.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

Changde Du, Kaicheng Fu, Jinpeng Li, and Huiguang He. Decoding visual neural representations by multimodal learning of brain-visual-linguistic features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10760–10777, 2023.

Yang Du, Siyuan Dai, Yonghao Song, Paul M. Thompson, Haoteng Tang, and Liang Zhan. Deep models, shallow alignment: Uncovering the granularity mismatch in neural decoding. *Transactions on Machine Learning Research*, 2026.

Yassine El Ouahidi, Jonathan Lys, Philipp Thölke, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon, Karim Jerbi, and Giulia Lioi. REVE: A foundation model for EEG - adapting to any setup with large-scale pretraining on 25,000 subjects. In *Advances in Neural Information Processing Systems*, volume 38, pp. 22541–22577. Curran Associates, Inc., 2025.

Alessandro T. Gifford, Kshitij Dwivedi, Gemma Roig, and Radoslaw M. Cichy. A large and rich EEG dataset for modeling human visual object recognition. *NeuroImage*, 264:119754, 2022.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. Bootstrap your own latent - a new approach to self-supervised learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21271–21284. Curran Associates, Inc., 2020.

Ishaan Gulrajani and David Lopez-Paz. In search of lost domain generalization. In *International Conference on Learning Representations*, 2021.

James V. Haxby, J. Swaroop Guntupalli, Andrew C. Connolly, Yaroslav O. Halchenko, Bryan R. Conroy, M. Ida Gobbini, Michael Hanke, and Peter J. Ramadge. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72(2):404–416, 2011.

He He and Dongrui Wu. Transfer learning for brain–computer interfaces: A euclidean space data alignment approach. *IEEE Transactions on Biomedical Engineering*, 67(2):399–410, 2020.

Martin N. Hebart, Adam H. Dickter, Alexis Kidder, Wan Y. Kwok, Anna Corriveau, Caitlin Van Wicklin, and Chris I. Baker. THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLOS ONE*, 14(10):e0223792, 2019.

Martin N. Hebart, Oliver Contier, Lina Teichmann, Adam H. Rockter, Charles Y. Zheng, Alexis Kidder, Anna Corriveau, Maryam Vaziri-Pashkam, and Chris I. Baker. THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. *eLife*, 12:e82580, 2023.

Qunjie Huang and Weina Zhu. SATTC: Structure-aware label-free test-time calibration for cross-subject EEG-to-image retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16887–16896, 2026.

Shuai Huang, Huan Luo, Haodong Jing, Qixian Zhang, Litao Chang, Yating Feng, Xiao Lin, Chendong Qin, Han Chen, Shuwen Jia, Siyi Sun, and Yongxiong Wang. NEED: Cross-subject and cross-task generalization for video and image reconstruction from EEG signals. In *Advances in Neural Information Processing Systems*, volume 38, pp. 173134–173173. Curran Associates, Inc., 2025.

Pavel Izmailov, Dmitrii Podoprikhin, Timur Garipov, Dmitry Vetrov, and Andrew Gordon Wilson. Averaging weights leads to wider optima and better generalization. In *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence*, 2018.

Andrew Jaegle, Felix Gimeno, Andy Brock, Oriol Vinyals, Andrew Zisserman, and Joao Carreira. Perceiver: General perception with iterative attention. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 4651–4664. PMLR, 2021.

Lin Jiang, Qingshan She, Jiale Xu, Haiqi Xu, Duanpo Wu, and Zhenzhong Kuang. Subject-aware multi-granularity alignment for zero-shot EEG-to-image retrieval. arXiv preprint arXiv:2604.17782v2, 2026.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous EEG data in BCI. In *International Conference on Learning Representations*, pp. 16405–16426, 2024.

Reese Kneeland, Wangshu Jiang, Ugo Bruzadin Nunes, Paul Steven Scotti, Arnaud Delorme, and Jonathan Xu. ENIGMA: EEG-to-image in 15 minutes using less than 1% of the parameters. arXiv preprint arXiv:2602.10361, 2026.

Demetres Kostas, Stéphane Aroca-Ouellette, and Frank Rudzicz. BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data. *Frontiers in Human Neuroscience*, 15:653659, 2021.

Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung, and Brent J. Lance. EEGNet: A compact convolutional network for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018.

Dongyang Li, Chen Wei, Shiying Li, Jiachen Zou, and Quanying Liu. Visual decoding and reconstruction via EEG embeddings with guided diffusion. In *Advances in Neural Information Processing Systems*, volume 37, pp. 102822–102864. Curran Associates, Inc., 2024.

Dongyang Li, Haoyang Qin, Mingyang Wu, Chen Wei, and Quanying Liu. BrainFLORA: Uncovering brain concept representation via multimodal neural embeddings. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 5577–5586. Association for Computing Machinery, 2025a.

Yueyang Li, Zijian Kang, Shengyu Gong, Wenhao Dong, Weiming Zeng, Hongjie Yan, Wai Ting Siok, and Nizhuan Wang. Neural-MCRL: Neural multimodal contrastive representation learning for EEG-based visual decoding. In *IEEE International Conference on Multimedia and Expo*, pp. 1–6, 2025b.

Wenchao Liu, Hongwei Li, Zhouyang Xu, Lin Ma, and Haifeng Li. Leveraging visual blur perception characteristics for EEG decoding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(21):17580–17588, 2026.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1):3942–3951, 2018.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 2021.

Paul S. Scotti, Atmadeep Banerjee, Jimmie Goode, Stepan Shabalin, Alex Nguyen, Ethan Cohen, Aidan J. Dempster, Nathalie Verlinde, Elad Yundler, David Weisberg, Kenneth A. Norman, and Tanishq Mathew Abraham. Reconstructing the mind's eye: fMRI-to-image with contrastive learning and diffusion priors. In *Advances in Neural Information Processing Systems*, volume 36, pp. 24705–24728. Curran Associates, Inc., 2023.

Paul S. Scotti, Mihir Tripathy, Cesar Kadir Torrico Villanueva, Reese Kneeland, Tong Chen, Ashutosh Narang, Charan Santhirasegaran, Jonathan Xu, Thomas Naselaris, Kenneth A. Norman, and Tanishq Mathew Abraham. MindEye2: Shared-subject models enable fMRI-to-image with 1 hour of data. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 44038–44059. PMLR, 2024.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*, 2017.

Leslie N. Smith and Nicholay Topin. Super-convergence: Very fast training of neural networks using large learning rates. arXiv preprint arXiv:1708.07120, 2017.

Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. EEG conformer: Convolutional transformer for EEG decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2023.

Yonghao Song, Bingchuan Liu, Xiang Li, Nanlin Shi, Yijun Wang, and Xiaorong Gao. Decoding natural images from EEG for object recognition. In *International Conference on Learning Representations*, 2024.

Jingyi Tang, Shuai Jiang, Fei Su, and Zhicheng Zhao. Aligning what EEG can see: Structural representations for brain-vision matching. arXiv preprint arXiv:2603.07077, 2026.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Guagnyu Wang, Wenchao Liu, Yuhong He, Cong Xu, Lin Ma, and Haifeng Li. EEGPT: Pretrained transformer for universal and reliable representation of EEG signals. In *Advances in Neural Information Processing Systems*, volume 37, pp. 39249–39280. Curran Associates, Inc., 2024.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. CBraMod: A criss-cross brain foundation model for EEG decoding. In *International Conference on Learning Representations*, pp. 75310–75346, 2025.

Jonathan R. Wolpaw, Niels Birbaumer, Dennis J. McFarland, Gert Pfurtscheller, and Theresa M. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, 2002.

Haitao Wu, Qing Li, Changqing Zhang, Zhen He, and Xiaomin Ying. Bridging the vision-brain gap with an uncertainty-aware blur prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2246–2257, 2025.

Weihao Xia, Raoul de Charette, Cengiz Öztireli, and Jing-Hao Xue. UMBRAE: Unified multimodal brain decoding. In *European Conference on Computer Vision*, pp. 242–259, 2024.

Chaoqi Yang, M. Brandon Westover, and Jimeng Sun. BIOT: Biosignal transformer for cross-data learning in the wild. In *Advances in Neural Information Processing Systems*, volume 36, pp. 78240–78260. Curran Associates, Inc., 2023.

Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations*, 2018.

Wenjiang Zhang, Sifeng Wang, Yuwei Su, Xinyu Li, Chen Zhang, and Suyu Zhong. NeuroBridge: Bio-inspired self-supervised EEG-to-image decoding via cognitive priors and bidirectional semantic alignment. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(21):18028–18036, 2026.

Jiawen Zheng, Haonan Jia, Ming Li, Yuhui Zheng, Yufeng Zeng, Yang Gao, and Chen Liang. Learning brain representation with hierarchical visual embeddings. In *International Conference on Learning Representations*, pp. 121474–121496, 2026.

Jiahao Zhou, Chenghao Xu, Wei Wang, Erkun Yang, and Cheng Deng. EEGiT: Teaching vision transformers to understand the EEG signal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 40438–40447, 2026.

Yanan Zhu, Ziwei Xiang, Jiamin Wu, Jinyang Guo, Hongyuan Zhang, Chunfeng Song, Hongjian Fang, Yufei Guo, and Xianglong Liu. Region-aware hierarchical sub-feature alignment for robust EEG-based visual decoding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Findings*, pp. 6667–6676, 2026.

A Data and Evaluation

A.1 THINGS-EEG2

We extract stimulus events from the released continuous recordings, exclude catch events coded 99999, and retain the 63 channels in Table 4. Epochs span -0.2 to 1.0 s, with the prestimulus mean subtracted separately for each trial and channel. Resampling to 250 Hz and retaining the last 250 samples yields the grid $0, 0.004, \ldots, 0.996$ s. Our importer applies no additional ICA, notch or band-pass filtering, or amplitude-based trial rejection.

Each participant completes four sessions containing 8,270 training images with two repetitions per session. Each training image occurs in two sessions, giving 16,540 distinct images with four responses each. The 200 test images have twenty responses per session, totaling eighty per image. Image identities determine cross-session trial grouping; within-session repeat ordering uses seed 20200220. Training averages four responses per image, yielding 16,540 examples per participant, and testing averages eighty.

Baseline correction and multivariate noise normalization (MVNN) follow VisualEEG (Liu et al., 2026). MVNN is fitted separately for each participant and session using training responses. For each cropped response $\mathbf{U}_{i,r} \in \mathbb{R}^{T\times C}$, we estimate its channel covariance $\widehat{\mathbf{\Sigma}}_{i,r}$ from time samples with automatic shrinkage. The estimator is scikit-learn's `_cov` function in

`discriminant_analysis`. For I training images with R responses per image within a session, we compute

$$\mathbf{\Sigma} = \frac{1}{I}\sum_{i=1}^{I}\frac{1}{R}\sum_{r=1}^{R}\widehat{\mathbf{\Sigma}}_{i,r}, \qquad \mathbf{U}'_{i,r} = \mathbf{U}_{i,r}\mathbf{\Sigma}^{-1/2}. \tag{8}$$

The estimator uses individual responses without subtracting the across-repeat evoked mean. The inverse square root is computed in float64 and the transformed responses are stored in float32. Each session's test responses use its training-fitted transform. Repeat averaging precedes model augmentation; the primary EEG models use no further z-score.

MVNN is fitted to each participant's full released training pool before LOSO and source-validation splits. Consequently, held-out participants contribute unlabeled training responses to their preprocessing, and source validation shares the training-pool transform. Test responses never fit the covariance. Limited-data adaptation therefore measures the budget of labeled EEG–image pairs under this preprocessing, not a fraction of the total recording budget. Train and test image identities, concepts, image hashes, and EEG response fingerprints were checked for disjointness; cached visual features were checked against the corresponding image order.

A.2 THINGS-MEG

We use the official preprocessed FIF release, OpenNeuro ds004212, version 2.0.1. Its pipeline includes 0.1–40 Hz filtering, recording-dropout and bad-sensor handling, photodiode-corrected timing, and epochs spanning -100 to 1300 ms. Each trial and sensor is standardized by its prestimulus mean and standard deviation before resampling to 200 Hz. Following the VisualEEG crop, we retain 0–1000 ms inclusively, giving 271 sensors and 201 samples in recorded channel order.

Excluding the 200 test concepts from the 1,854-concept main set leaves $1654 \times 12 = 19848$ distinct training images, each with one response. The test set contains 200 images with twelve responses each, averaged before evaluation. Image identities follow the official trial metadata. Joint training uses these inputs without additional training-pool z-scoring or sample-RMS normalization.

MEG LOSO applies per-example sensor RMS scaling after repeat averaging. Each sensor trajectory is divided by its RMS, clamped below at 10^{-6}, without temporal mean subtraction or fitted population statistics. Training uses one response per image; testing averages twelve before scaling. The Shallow controls retain their native 84-sensor, 200-sample derivative and participant-specific training-pool z-score.

A.3 Frozen Visual Features

Both primary models use `openai/clip-vit-large-patch14`. RGB images are bicubically resized to a shorter side of 224 pixels and center-cropped to 224×224. Channel means are 0.48145466, 0.4578275 and 0.40821073; standard deviations are 0.26862954, 0.26130258 and 0.27577711. We apply this normalization without random image augmentation. The frozen encoder provides `hidden_states[12][:,1:]`, the output of transformer block 12 with the class token removed. Here `hidden_states[0]` is the input to the transformer blocks. No post-layer normalization is applied to the target. Features are cached in float16 and cast to float32 for regression and retrieval normalization.

A.4 Retrieval Evaluation

Evaluation forms $\mathbf{S} = \mathbf{U}\mathbf{V}^\top \in \mathbb{R}^{200\times200}$, where $\mathbf{U}$ and $\mathbf{V}$ stack normalized neural and image embeddings, respectively. Row-wise ranking measures neural-to-image retrieval; ranking $\mathbf{S}^\top$ measures image-to-neural retrieval. Each gallery contains one image per unseen test concept. Candidate images provide frozen features, while their neural pairings are used only to score retrieval. Accuracies are averaged within each participant and then across participants. Chance Top-1/Top-5 accuracy is 0.50/2.50% in either retrieval direction. Both directions for ReShare use identical checkpoints and the same 200 paired image and neural candidates. Joint evaluation uses participants included in supervised training; LOSO excludes the target participant from supervised source training and uses identity FiLM at inference. EEG and MEG models have separate weights. Model development

was informed by benchmark evaluations; the reported weights follow the fixed or source-validation selection rules below.

B Model and Training Details

B.1 Sensor Groups and Token Allocation

The regional prior partitions recorded sensors by scalp or helmet topology. These groups specify sensor coordinates rather than localized cortical sources. Each EEG channel belongs to one group, with fixed token budgets listed in Table 4. Group-specific spatial expanders map 63 channels to 256 trajectories, concatenated in the listed order.

Table 4: EEG sensor membership and spatial-token allocation in concatenation order.

Group	Channel names	Sensors	Tokens
Occipital	PO7, PO3, POz, PO4, PO8, O1, Oz, O2	8	48
Parietal	P7, P5, P3, P1, Pz, P2, P4, P6, P8	9	36
Centroparietal	CP5, CP3, CP1, CPz, CP2, CP4, CP6	7	28
Posterior temporal	TP9, TP7, TP8, TP10	4	20
Central	C5, C3, C1, Cz, C2, C4, C6	7	24
Anterior temporal	FT9, FT7, T7, T8, FT8, FT10	6	24
Frontocentral	FC5, FC3, FC1, FCz, FC2, FC4, FC6	7	24
Frontal	F7, F5, F3, F1, F2, F4, F6, F8	8	28
Prefrontal	Fp1, Fp2, AF7, AF3, AFz, AF4, AF8	7	24
Total		63	256

MEG groups use prefixes MLO/MRO, MLP/MRP, MLT/MRT, MLC/MRC, MLF/MRF, and MZ for left/right occipital, parietal, temporal, central, frontal, and midline sensors. After assigning one token to each group, the remaining 245 are allocated proportionally to sensor counts by the largest-remainder rule, with ties resolved by group order.

Table 5: MEG sensor groups and spatial-token allocation.

Group	Sensors	Tokens
left occipital	19	18
right occipital	17	16
left parietal	22	21
right parietal	22	21
left temporal	34	32
right temporal	34	32
left central	24	23
right central	24	22
left frontal	32	30
right frontal	32	30
midline	11	11
Total	271	256

B.2 Network Operations and Regularization

For group k with C_k sensors and n_k tokens, the spatial expander learns raw weights $\mathbf{R}_k \in \mathbb{R}^{n_k \times C_k}$ and gains $\mathbf{g}_k \in \mathbb{R}^{n_k}$. For local token n, each forward pass normalizes the corresponding row as $\mathbf{A}_{k,n,:} = \mathbf{R}_{k,n,:} / \max(\|\mathbf{R}_{k,n,:}\|_2, \epsilon)$, with $\epsilon = 10^{-6}$. The colon selects all sensors and $\|\cdot\|_2$ is the Euclidean norm. The effective projection in Equation 1 is $\mathbf{W}_k = \mathrm{diag}(\mathbf{g}_k)\mathbf{A}_k$, where diag places the gains on a diagonal matrix. The weights remain signed after normalization. Raw weights are initialized as small Gaussian perturbations of cycling one-hot sensor assignments, and gains start at

one. The spatial regularizer balances absolute channel contributions within each group. For sensor c in group k, define $a_{k,c} = n_k^{-1} \sum_n |A_{k,n,c}|$. We use

$$\mathcal{L}_{\text{spatial}} = \sum_k \frac{1}{C_k} \sum_c \left(\frac{a_{k,c}}{\max(\text{mean}_{c'}\, a_{k,c'}, \epsilon)} - 1 \right)^2 . \tag{9}$$

The temporal encoder processes each trajectory $\mathbf{z}_n \in \mathbb{R}^T$ independently:

$$\mathbf{h}_n = \text{GELU}(\mathbf{W}_\text{t}\mathbf{z}_n + \mathbf{b}_\text{t} + \mathbf{W}_\text{m}\mathbf{m}), \qquad \mathbf{E}_n = \text{LN}\big(\mathbf{h}_n + \mathcal{F}_\text{t}(\text{LN}(\mathbf{h}_n))\big). \tag{10}$$

Here $\mathbf{W}_\text{t}, \mathbf{W}_\text{m} \in \mathbb{R}^{1024\times T}$ are the signal and mask projections, and $\mathbf{b}_\text{t} \in \mathbb{R}^{1024}$ is the signal-projection bias. Each entry of the temporal validity mask $\mathbf{m} \in \{0,1\}^T$ indicates whether that time sample is present. All input windows used here are complete, so $\mathbf{m} = \mathbf{1}_T$. The mask projection is bias-free. The residual MLP $\mathcal{F}_\text{t}$ uses bias-free linear maps $1024 \to 2048 \to 1024$, GELU after the first map, and dropout 0.1 after GELU and the second map. Region and token-position embeddings give $\mathbf{H}_n = \mathbf{E}_n + \mathbf{e}_{k(n)} + \mathbf{p}_n$ before FiLM, where $k(n)$ identifies token n's region in the concatenated sequence. Here $\mathbf{e}_{k(n)}$ and $\mathbf{p}_n$ are learned D-dimensional region and token-position vectors, respectively. Within each region, their sums form the rows of $\mathbf{P}_k$ in Equation 2.

For S participants, feature-wise linear modulation (FiLM) learns zero-initialized scale and offset tables $\boldsymbol{\gamma}, \boldsymbol{\beta} \in \mathbb{R}^{S\times K\times D}$, totaling $2SKD$ parameters. For ten EEG participants and nine regions, this is 184,320 parameters. These tables provide participant conditioning; no additional global participant embedding is used. Input dropout at rate 0.1 and normalization follow FiLM, supplying fixed neural input tokens $\mathbf{M}$ to the shared backbone from the concatenated calibrated features $\mathbf{H}'$.

The shared latent backbone comprises six layers that combine cross-attention, within-region self-attention, and global self-attention. Regional latent banks and region-type embeddings initialize the latent sequence $\mathbf{L}^0 \in \mathbb{R}^{N\times D}$. Learned latents read the inputs through cross-attention, as in Perceiver (Jaegle et al., 2021). Table 6 gives the attention sequence and feed-forward update for each layer, with layer indices starting at zero. Each cross-attention block separately normalizes neural input tokens and latents, with 16 heads of width 96 and a projection from 1,536 to 1,024. Regional and global latent attention use 16 heads of width 64 and attention dropout 0.1. Cross-attention has no attention dropout. Dense feed-forward networks (FFNs) use LayerNorm and bias-free maps $1024 \to 4096 \to 1024$ with GELU and no dropout. The output projection is linear $1024 \to 1024$. For the cross-attention update in Equation 4, let $\mathbf{L}^\ell$ denote the latents entering layer ℓ. Separate layer-normalization modules produce $\hat{\mathbf{L}}^\ell$ from these latents and $\hat{\mathbf{M}}^\ell$ from the fixed neural input tokens. The residual update is

$$\mathbf{L}^\ell_\text{read} = \mathbf{L}^\ell + \text{MHA}\big(\hat{\mathbf{L}}^\ell, [\hat{\mathbf{M}}^\ell; \hat{\mathbf{L}}^\ell], [\hat{\mathbf{M}}^\ell; \hat{\mathbf{L}}^\ell]\big), \tag{11}$$

where $\mathbf{L}^\ell_\text{read}$ is the updated latent sequence and MHA denotes multi-head attention (Vaswani et al., 2017) with query, key, and value as its arguments.

Table 6: Attention and feed-forward operations in the shared latent backbone.

Layer	Attention sequence	Feed-forward update
0	Cross-attention	Dense FFN
1	Cross-attention; within-region self-attention; global self-attention	Dense FFN
2	Cross-attention	Dense FFN
3	Within-region self-attention; global self-attention	Dense FFN
4	Cross-attention	Dense FFN
5	Cross-attention	Three experts, top-two routing

The tokenwise retrieval projector uses a BYOL-style MLP (Grill et al., 2020) with widths $1024 \to 2048 \to 2048 \to 1024$; LayerNorm and GELU precede each linear layer. Training uses the regression and contrastive objectives below, with frozen visual targets and no image-side projector or momentum image teacher.

The MoE router (Shazeer et al., 2017) selects two of three logits per token and softmaxes them to weight the selected expert outputs. All three experts are evaluated in this implementation. For a batch of B recordings with N tokens each, let $r_{i,n,e}$ be the router probability for recording i, token n, and expert e. With $E = 3$ experts, we define their mean probabilities p_e and top-choice frequencies f_e, giving the balancing loss

$$p_e = \frac{1}{BN}\sum_{i,n} r_{i,n,e}, \qquad f_e = \frac{1}{BN}\sum_{i,n} \mathbb{1}[e = \arg\max_j r_{i,n,j}], \qquad \mathcal{L}_{\text{MoE}} = E\sum_e p_e f_e. \tag{12}$$

Gradients pass through p_e, while discrete assignments f_e determine the load statistic. Expert identities are learned without anatomical labels.

B.3 Augmentation and Alignment Objective

We restore participant and image indices to specify the training objective. Let $\bar{\mathbf{X}}_{s,i}$ denote participant s's preprocessed response to image $\mathbf{I}_i$. The neural decoder predicts

$$\mathbf{Y}_{s,i} = \mathcal{D}_{\boldsymbol{\theta},\boldsymbol{\phi}_s}(\bar{\mathbf{X}}_{s,i}), \tag{13}$$

where $\mathcal{D}$ is the decoding mapping, $\boldsymbol{\theta}$ parameterizes the shared regional encoder and latent integration backbone, and $\boldsymbol{\phi}_s$ contains the participant's regional calibration parameters. The frozen visual target $\mathbf{V}_i \in \mathbb{R}^{N\times D}$ is the ordered patch sequence from CLIP block 12 for image $\mathbf{I}_i$. The tokenwise projector $\mathcal{P}_{\boldsymbol{\psi}}$, with learned parameters $\boldsymbol{\psi}$, maps neural outputs to retrieval features. The normalized neural and image embeddings are

$$\mathbf{u}_{s,i} = \frac{\text{vec}(\mathcal{P}_{\boldsymbol{\psi}}(\mathbf{Y}_{s,i}))}{\|\text{vec}(\mathcal{P}_{\boldsymbol{\psi}}(\mathbf{Y}_{s,i}))\|_2}, \qquad \mathbf{v}_i = \frac{\text{vec}(\mathbf{V}_i)}{\|\text{vec}(\mathbf{V}_i)\|_2}, \tag{14}$$

where vec flattens the token sequence. The inner product $\mathbf{u}_{s,i}^\top \mathbf{v}_j$ is the cosine similarity to candidate image $\mathbf{I}_j$. The projector architecture is specified in Appendix B.2.

Region dropout selects each example with probability 0.5, then zeros the channels of one or two randomly chosen groups without replacement. It operates before temporal encoding and MixCo; encoded tokens receive no additional region mask. This augmentation is separate from the valid-time mask in Equation 10.

Within participant s's batch of B responses, let $\widetilde{\mathbf{X}}_{s,i}$ denote a response after region dropout and let π be a random within-batch permutation. MixCo selects examples with probability 0.5 and forms

$$\hat{\mathbf{X}}_{s,i} = \lambda_i \widetilde{\mathbf{X}}_{s,i} + (1-\lambda_i)\widetilde{\mathbf{X}}_{s,\pi(i)}, \qquad q_{ij} = \lambda_i \mathbb{1}[j=i] + (1-\lambda_i)\mathbb{1}[j=\pi(i)]. \tag{15}$$

Selected examples use $\lambda_i \sim \text{Beta}(0.15, 0.15)$; otherwise $\lambda_i = 1$. Both alignment terms use $\hat{\mathbf{Y}}_{s,i} = \mathcal{D}_{\boldsymbol{\theta},\boldsymbol{\phi}_s}(\hat{\mathbf{X}}_{s,i})$ from the same forward pass. Section 3.4 writes this augmented prediction as $\mathbf{Y}_i$, omitting the participant index and augmentation mark for readability. Substituting $\hat{\mathbf{Y}}_{s,i}$ into Equation 14 gives $\hat{\mathbf{u}}_{s,i}$. For logits $a_{s,ij} = \hat{\mathbf{u}}_{s,i}^\top \mathbf{v}_j/\tau$ with $\tau = 0.006$, the bidirectional contrastive loss is

$$\mathcal{L}_{\text{con}} = -\frac{1}{2B}\sum_{i,j} q_{ij}\left(\log\frac{e^{a_{s,ij}}}{\sum_k e^{a_{s,ik}}} + \log\frac{e^{a_{s,ij}}}{\sum_k e^{a_{s,kj}}}\right). \tag{16}$$

Here q_{ij} assigns contrastive target mass to the two constituent images, and τ is the contrastive temperature. The inputs and target weights use linear mixing as in mixup (Zhang et al., 2018). The reverse term uses the transposed soft targets without row renormalization. BiMixCo (Scotti et al., 2023) remains active throughout training. Token regression uses $\mathcal{L}_{\text{mse}} = (BND)^{-1}\sum_i \left\|\hat{\mathbf{Y}}_{s,i} - \mathbf{V}_i\right\|_{\text{F}}^2$, retaining the unmixed target $\mathbf{V}_i$ for the mixed neural input. Thus contrastive supervision assigns mass to both mixture constituents, whereas regression anchors the prediction to image i. No second forward pass is performed on the inputs after region dropout and before MixCo mixing. The squared Frobenius norm sums errors over all token features. The combined objective for participant s is

$$\mathcal{L}_s = \mathcal{L}_{\text{con}} + \lambda_{\text{mse}}\mathcal{L}_{\text{mse}} + \lambda_{\text{moe}}\mathcal{L}_{\text{MoE}} + \lambda_{\text{spatial}}\mathcal{L}_{\text{spatial}}. \tag{17}$$

The auxiliary terms balance expert usage and absolute sensor contributions, as defined in Equations 12 and 9. The three weights are $\lambda_{\text{mse}} = 1$, $\lambda_{\text{moe}} = 0.01$, and $\lambda_{\text{spatial}} = 0.001$.

B.4 Optimization and Source-Based Selection

AdamW (Loshchilov & Hutter, 2019) uses the settings in Table 7. Joint training and EEG LOSO select two participants per epoch and traverse each selected participant's training set once. MEG LOSO cycles through all three source-participant pairs in a randomized order every three updates, so each source contributes twice per cycle. Each source has a separately shuffled image stream that continues across epochs and reshuffles after exhaustion. Each update sums the losses computed independently for the two participant batches. Final partial batches are retained; full joint EEG training therefore uses $\lceil 16540/128 \rceil = 130$ updates per epoch. Joint models and EEG LOSO use cosine OneCycle schedules (Smith & Topin, 2017), starting at 1.2×10^{-5} and ending near 1.2×10^{-8}. Bias, normalization, embedding and FiLM parameters are excluded from weight decay.

Table 7: Optimization and input settings for joint and LOSO training.

Setting	EEG joint	MEG joint	EEG LOSO	MEG LOSO
Supervised participants	10	4	9	3
Seed	42	42	42	42
Training images / source	16,540	19,848	15,710	18,852
Validation images / source	None	None	830	996
Epochs	500	200	450	60
Steps / epoch	130	156	123	148
Optimizer updates	65,000	31,200	55,350	8,880
Participants / step	2	2	2	2
Batch / participant	128	128	128	128
Maximum learning rate	$3\cdot10^{-4}$	$3\cdot10^{-4}$	$3\cdot10^{-4}$	$3\cdot10^{-4}$
Warmup epochs	25	10	25	10
Weight decay	0.01	0.01	0.01	0.01
LR after warmup	OneCycle	OneCycle	OneCycle	Source-validation loss plateau
Gradient clipping norm	1.0	1.0	1.0	1.0
Precision	BF16	BF16	BF16	BF16
Contrastive temperature	0.006	0.006	0.006	0.006
Token-MSE weight	1	1	1	1
MoE / spatial weights	.01 / .001	.01 / .001	.01 / .001	.01 / .001
Train / test repeat means	4 / 80	1 / 12	4 / 80	1 / 12
Extra sensor RMS scaling	No	No	No	Yes
Reported checkpoint	Final epoch	Final epoch	Final SWAD	Source-selected SWAD

LOSO source validation withholds 83 of 1,654 training concepts using seed 20260828. EEG supplies ten images per concept, giving 830 validation images; MEG supplies twelve, giving 996. Validation uses ten or twelve 83-way galleries, respectively, with one image per held-out concept in each gallery. Inputs use the training repeat count and modality-specific normalization. Identity FiLM is used for validation, with symmetric contrastive loss averaged across source participants and galleries.

In LOSO training, after warmup, we maintain an exponential moving average (EMA) of the model weights with decay 0.999 and a separate SWAD-style weight average (Cha et al., 2021). The latter extends stochastic weight averaging (Izmailov et al., 2018) with source-validation criteria. For SWAD, weights are averaged over optimizer steps within successive five-epoch intervals. Each interval is assigned the source-validation loss of the unaveraged model at its end. Averaging across intervals begins when the first loss in a window of three consecutive intervals is no greater than either of the next two. The first interval is included immediately, and the acceptance threshold is fixed at 1.2 times the mean loss of these three intervals. Subsequent intervals are held until six are available for consideration. If all six losses exceed the threshold, averaging stops and those intervals are discarded; otherwise, the earliest interval is included and the window advances by one. Included intervals are weighted by their numbers of optimizer steps. At training completion, if averaging has not stopped, the remaining intervals are included in order until all remaining losses exceed the threshold. Before the three-interval criterion is met, SWAD evaluation uses the most recent interval average. EEG LOSO reports this final SWAD average after 450 epochs in every fold.

MEG LOSO uses ten epochs of cosine warmup from 1.2×10^{-5} to 3×10^{-4}. Thereafter, the source-validation loss of the model with unaveraged weights is evaluated each epoch and controls learning-rate reduction with factor 0.2, patience 1, cooldown 2, and minimum 10^{-6}. The AdamW first-moment coefficient continues to follow the OneCycle schedule. The SWAD-averaged model is separately evaluated on source validation data every five epochs. We select the checkpoint with the lowest loss, retaining the earliest checkpoint when losses are equal. Retrieval scores on the target participant's test set do not select the reported EEG or MEG weights.

C Complete Retrieval Results

C.1 Reference Settings and Participant-Level Scores

Tables 8 and 9 report available participant-level results, separating independent fitting, joint fitting, and LOSO. The main tables combine published neural-to-image results with dagger-marked image-to-neural scores from our checkpoint reevaluations under the corresponding intra- or inter-subject setting. Reevaluated HVE models use a fixed training schedule rather than model selection using the target participant's test results. Literature methods retain their reported visual targets, sensor selections, and training settings. The shared-training controls used for representation and adaptation analysis are specified separately in Appendices D.4 and E.4.

EEG values for BraVL, NICE, ATM, Neural-MCRL, UBP, and NeuroBridge follow NeuroBridge Table 1 (Zhang et al., 2026). VisualEEG, EEGiT, and HVE use their own main tables (Liu et al., 2026; Zhou et al., 2026; Zheng et al., 2026). For MEG, NICE variants and VisualEEG follow VisualEEG Table 3; UBP, NeuroBridge, and Shallow follow Shallow Table 2 (Du et al., 2026). HVE uses the MEG result reported in its main table, 33.7/60.5%. Shallow's reference scores are its main single-layer results. Reported means are retained at their published precision; unavailable participant scores remain unfilled. The reproduced Shallow MEG results average five seeds per participant using its native inputs.

C.2 Variation across Participants

Table 10 reports the mean and sample standard deviation across ten EEG or four MEG participants. For accuracies $a_1, \ldots, a_S$, the deviation is $\sqrt{\sum_s (a_s - \bar{a})^2/(S-1)}$. It measures between-participant variation at the reported checkpoints; LOSO folds share source participants and are not independent training replicates. Joint models use weights at the final epoch of the fixed training schedule. EEG LOSO uses final SWAD across all ten folds, while MEG uses source-selected SWAD across all four folds.

C.3 Label-Free Score Calibration

Given the higher uncalibrated retrieval accuracy on EEG in our experiments, we focus this supplementary analysis on MEG to explore the benefit of label-free score calibration. We do not evaluate column-z calibration on EEG.

MEG column calibration uses the same joint or LOSO weights as raw retrieval. Given $S_{ij} = \mathbf{u}_i^\top \mathbf{v}_j$, with 200 neural queries in rows and candidate images in columns, we compute

$$\mu_j = \frac{1}{200}\sum_i S_{ij}, \quad \sigma_j = \sqrt{\frac{1}{200}\sum_i (S_{ij} - \mu_j)^2}, \quad S_{ij}^{\mathrm{CZ}} = \frac{S_{ij} - \mu_j}{\max(\sigma_j, 10^{-6})}. \tag{18}$$

This adjusts each candidate's score distribution across the unlabeled query batch. It uses candidate-image features through $\mathbf{S}$, with no pairing labels, pseudo-labels, or parameter updates. Statistics are fitted separately for each participant, making this a transductive batch-inference setting. Unlike SATTC's feature whitening and neighborhood-based refinement (Huang & Zhu, 2026), column calibration requires only the column means and standard deviations of the existing similarity matrix.

Joint Top-1 increases from 61.88% to 71.13%, and Top-5 from 84.63% to 88.50%. Top-1 for each participant changes from 40.5/87.5/72.5/47.0% to 50.5/95.0/82.5/56.5%, respectively. Table 11 gives the corresponding LOSO comparison. Image-to-neural accuracy is unchanged because positive affine column transforms preserve within-column rankings.

Table 8: EEG-to-image retrieval accuracy (%) for each participant under the indicated training settings.

Method	Metric	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Avg.
Intra-subject: independent model per participant												
NICE	Top-1	13.2	13.5	14.5	20.6	10.1	16.5	17.0	22.9	15.4	17.4	16.10
	Top-5	39.5	40.3	42.7	52.7	31.5	44.0	42.1	56.1	41.6	45.8	43.60
ATM	Top-1	25.6	22.0	25.0	31.4	12.9	21.3	30.5	38.8	34.4	29.1	27.10
	Top-5	60.4	54.5	62.4	60.9	43.0	51.1	61.5	72.0	51.5	63.5	58.10
Neural-MCRL	Top-1	27.5	28.5	37.0	35.0	22.5	31.5	31.5	42.0	30.5	37.5	32.40
	Top-5	64.0	61.5	69.0	66.0	51.5	61.0	62.5	74.5	59.5	71.0	64.10
UBP	Top-1	41.2	51.2	51.2	51.1	42.2	57.5	49.0	58.6	45.1	61.5	50.90
	Top-5	70.5	80.9	82.0	76.9	72.8	83.5	79.9	85.8	76.2	88.2	79.70
NeuroBridge	Top-1	50.0	63.2	61.6	61.4	54.8	69.7	62.7	71.2	64.0	73.6	63.20
	Top-5	77.6	90.6	91.1	90.0	85.0	92.9	88.8	95.1	91.0	97.1	89.90
Shallow	Top-1	75.0	87.5	83.2	79.5	74.6	89.9	78.5	86.9	81.3	89.3	82.60
	Top-5	94.3	98.9	98.2	96.1	96.4	99.3	97.3	99.4	97.8	99.1	97.70
VisualEEG	Top-1	81.9	81.7	78.3	76.9	71.4	84.3	78.2	84.5	78.7	84.1	80.00
	Top-5	96.8	96.8	96.0	97.2	93.7	98.7	96.7	98.1	96.8	98.1	96.90
Intra-subject: joint seen-subject training (10 participants, one model)												
ReShare	Top-1	97.5	95.0	95.5	91.0	95.5	96.5	95.5	99.0	94.5	97.0	95.70
	Top-5	100.0	100.0	100.0	99.5	99.5	100.0	99.5	100.0	99.5	100.0	99.80
Inter-subject: leave one subject out for test												
NICE	Top-1	7.6	5.9	6.0	6.3	4.4	5.6	5.6	6.3	5.7	8.4	6.20
	Top-5	22.8	20.5	22.3	20.7	18.3	22.2	19.7	22.0	17.6	28.3	21.40
ATM	Top-1	10.5	7.1	11.9	14.7	7.0	11.1	16.1	15.0	4.9	20.5	11.90
	Top-5	26.8	24.8	33.8	39.4	23.9	35.8	43.5	40.3	22.7	46.5	33.80
Neural-MCRL	Top-1	13.0	12.0	14.5	12.5	11.5	13.5	14.0	18.5	13.5	17.0	14.00
	Top-5	31.5	30.5	35.5	35.5	29.0	35.5	36.0	38.5	32.5	39.0	34.30
UBP	Top-1	11.5	15.5	9.8	13.0	8.8	11.7	10.2	12.2	15.5	16.0	12.40
	Top-5	29.7	40.0	27.0	32.3	33.8	31.0	23.8	32.2	40.5	43.5	33.40
NeuroBridge	Top-1	23.2	21.2	13.2	17.0	14.5	25.0	15.3	20.1	13.7	27.2	19.00
	Top-5	52.4	49.3	36.5	45.3	37.7	55.0	45.1	44.9	36.5	56.3	45.90
Shallow	Top-1	23.5	30.6	10.0	19.5	18.1	22.7	18.6	17.3	23.0	34.4	21.80
	Top-5	53.2	60.4	28.1	48.3	45.2	49.8	46.0	46.1	54.8	62.0	49.40
ReShare	Top-1	41.0	26.5	20.0	24.5	20.5	24.5	26.5	21.5	32.0	30.5	26.75
	Top-5	68.0	58.5	49.5	54.5	52.0	49.5	58.0	53.5	62.5	54.0	56.00

D Regional Structure and Shared Representations

D.1 Progressive Component Study

Table 3 evaluates successive additions to a shared decoder. B0 uses unrestricted sensor mixing and a non-regional latent backbone, without participant FiLM, MoE, or region dropout. B1 introduces within-region sensor mixing and input region embeddings. B2 adds participant-specific regional FiLM; B3 adds regional latent identities and attention scopes. B4 replaces the final dense FFN with MoE, and B5 adds region dropout. Each increment is evaluated in the context of the preceding components.

All variants use 500 epochs of joint training and 200 test queries per participant. B0–B5 report neural-to-image accuracy averaged over three training seeds and ten participants, using the weights at the final epoch of each run. Training and loss definitions follow Appendix B with the corresponding components enabled.

Table 9: MEG-to-image retrieval accuracy (%) for each participant, comparing published references, reproduced Shallow models, and ReShare.

Method	Metric	S1	S2	S3	S4	Avg.
Intra-subject: independent model per participant						
UBP	Top-1	15.0	46.0	27.3	18.5	26.70
	Top-5	38.0	80.5	59.0	43.5	55.20
NeuroBridge	Top-1	16.5	53.7	40.4	18.1	32.20
	Top-5	41.6	85.3	73.2	43.1	60.80
Shallow	Top-1	25.5	81.9	56.0	28.6	48.00
	Top-5	54.5	97.4	87.5	58.3	74.40
Shallow (our reproduction)	Top-1	25.8	81.5	57.1	30.4	48.70
	Top-5	53.3	97.0	86.2	57.2	73.43
Intra-subject: joint seen-subject training (4 participants, one model)						
ReShare	Top-1	40.5	87.5	72.5	47.0	61.88
	Top-5	68.0	100.0	95.0	75.5	84.63
Intra-subject: unlabeled test-batch score calibration, same weights						
ReShare + column-z	Top-1	50.5	95.0	82.5	56.5	71.13
	Top-5	77.5	100.0	96.5	80.0	88.50
Inter-subject: leave one subject out for test						
UBP	Top-1	2.0	1.5	2.7	2.5	2.20
	Top-5	5.7	17.2	10.5	8.0	10.40
NeuroBridge	Top-1	4.3	3.6	3.0	2.5	3.40
	Top-5	13.1	15.6	11.2	11.3	12.80
Shallow	Top-1	1.3	6.6	5.4	1.5	3.70
	Top-5	6.9	18.5	18.5	7.9	13.00
ReShare	Top-1	5.0	8.5	7.5	4.0	6.25
	Top-5	12.5	22.5	27.5	9.5	18.00
Inter-subject: unlabeled test-batch score calibration, same weights						
ReShare + column-z	Top-1	5.5	14.5	12.0	3.5	8.88
	Top-5	14.5	31.5	37.5	18.5	25.50

Table 10: Participant-averaged retrieval and between-participant sample standard deviation for ReShare.

Dataset	Setting	Participants	Top-1 (%)	Top-5 (%)
THINGS-EEG2	Joint seen-participant	10	95.70 ± 2.12	99.80 ± 0.26
THINGS-EEG2	LOSO	10	26.75 ± 6.39	56.00 ± 5.86
THINGS-MEG	Joint seen-participant	4	61.88 ± 21.97	84.63 ± 15.32
THINGS-MEG	LOSO	4	6.25 ± 2.10	18.00 ± 8.43

Table 11: MEG LOSO retrieval accuracy (%) before and after column calibration at identical source-selected weights.

	Uncalibrated		Column-z	
Participant	Top-1	Top-5	Top-1	Top-5
1	5.00	12.50	5.50	14.50
2	8.50	22.50	14.50	31.50
3	7.50	27.50	12.00	37.50
4	4.00	9.50	3.50	18.50
Mean	6.25	18.00	8.88	25.50

D.2 Participant Calibration Granularity

We compared regional FiLM with global participant calibration and a decoder trained without FiLM. The global variant shares one scale and offset across all tokens of each participant, while regional FiLM assigns a separate pair to each sensor group. All three retain regional tokenization, regional latent attention, MoE and region dropout, with the same 500-epoch joint-training schedule and visual targets. The no-FiLM control removes calibration throughout training; it is distinct from replacing a trained model's FiLM with identity at inference. Each model starts from random initialization with seed 42 and is evaluated at its final epoch, with 80 test responses averaged per image. Global FiLM has 20,480 parameters; regional FiLM has 184,320, adding 0.13% to the global variant's 130.13M total parameters. Table 12 reports both retrieval directions; Table 13 shows neural-to-image Top-1 for each participant.

Table 12: Retrieval accuracy (%) for complete decoders trained without FiLM, with global participant FiLM, or with regional FiLM.

	Neural → image		Image → neural	
FiLM resolution	Top-1	Top-5	Top-1	Top-5
No FiLM	91.85	99.10	96.85	99.95
Global participant	93.25	99.45	97.60	99.95
Participant-region	95.70	99.80	98.45	100.00

Table 13: Neural-to-image Top-1 accuracy (%) for each participant under the three calibration settings, with participant sample SD.

Participant	No FiLM	Global FiLM	Regional FiLM
1	93.50	96.50	97.50
2	89.00	90.50	95.00
3	91.50	91.50	95.50
4	84.50	88.50	91.00
5	88.50	92.00	95.50
6	92.50	95.50	96.50
7	96.50	91.50	95.50
8	95.50	95.00	99.00
9	91.50	93.50	94.50
10	95.50	98.00	97.00
Mean ± SD	91.85 ± 3.73	93.25 ± 2.97	95.70 ± 2.12

Regional calibration improved mean Top-1 by 3.85 percentage points over no FiLM and 2.45 points over global FiLM.

D.3 Regression Input Control

We tested whether dense regression benefits from a separate unmixed EEG input while retaining the complete regional decoder. Both models used the same 500-epoch joint-training schedule, seed 42, visual targets, loss weights and evaluation at the final epoch.

Let $\widetilde{\mathbf{X}}_{s,i}$ denote the response after region dropout and $\hat{\mathbf{X}}_{s,i}$ its within-batch MixCo mixture, as defined in Appendix B.3. The reference regresses $\mathcal{D}_{\boldsymbol{\theta},\boldsymbol{\phi}_s}(\hat{\mathbf{X}}_{s,i})$ to the original visual target $\mathbf{V}_i$; the control instead regresses $\mathcal{D}_{\boldsymbol{\theta},\boldsymbol{\phi}_s}(\widetilde{\mathbf{X}}_{s,i})$ to $\mathbf{V}_i$. Thus, both regression inputs retain region dropout, and BiMixCo continues to use mixed EEG in both settings. The control adds a second forward pass on the same inputs after region dropout, without MixCo mixing. This pass remains differentiable and uses independently sampled network dropout. Contrastive and auxiliary losses remain on the mixed branch, with all gradients accumulated before each optimizer update. Table 14 reports both retrieval directions with sample SD across the ten participants.

Table 14: Retrieval accuracy (%) with mixed or unmixed EEG regression inputs, reported as mean and sample SD across ten participants.

	Neural → image		Image → neural	
Regression input	Top-1	Top-5	Top-1	Top-5
Mixed (reference)	95.70 ± 2.12	99.80 ± 0.26	98.45 ± 1.07	100.00 ± 0.00
Unmixed	94.00 ± 2.52	99.65 ± 0.67	97.95 ± 1.64	99.95 ± 0.16

Regression using unmixed inputs reached 94.00% neural-to-image Top-1, compared with 95.70% for the reference, a difference of 1.70 percentage points. This comparison supports retaining regression using mixed inputs under the evaluated training conditions.

D.4 Cross-Participant Neural Matching

We assess whether jointly trained models produce compatible representations for the same unseen image across participants. For each ordered pair (s, t), normalized neural embeddings form $S_{ij}^{s,t} = \langle \mathbf{z}_{s,i}, \mathbf{z}_{t,j} \rangle$ over 200 image conditions. The shared image ordering supplies scoring labels; the matching computation uses neural embeddings alone, without image features or fitted score calibration. The joint checkpoints include all evaluated participants in training. There are $S(S-1)$ directed pairs: ninety for EEG and twelve for MEG.

ReShare uses flattened retrieval-projector outputs, Shallow its native 512-dimensional embedding, and VisualEEG and ATM their 1,024-dimensional embeddings. The EEG Shallow control is jointly trained for fifty epochs; the MEG control uses 100 epochs, batch 1,024, and 8,000 updates. For 200-way neural-to-image retrieval, these Shallow controls achieve Top-1 accuracies of 80.25% on EEG and 32.25% on MEG. VisualEEG and ATM use the source-participant pretraining settings in Appendix E.4, extended to all ten participants and all 16,540 training conditions. Their MEG extensions are not included. All controls retain native visual targets and input processing. Table 15 summarizes matching, and Figures 5–6 show every directed pair. Pairs sharing participants are correlated observations.

Table 15: Neural-to-neural matching accuracy (%) averaged over directed participant pairs for unseen images within jointly trained cohorts.

Dataset	Neural embedding	Top-1	Top-5
EEG	ReShare	81.00	96.59
EEG	Identity FiLM	72.34	92.74
EEG	Shared Shallow	64.98	89.99
EEG	Shared VisualEEG	35.56	66.80
EEG	Shared ATM	49.14	80.64
MEG	ReShare	30.42	59.96
MEG	Identity FiLM	23.88	52.13
MEG	Shared Shallow	11.54	33.25

D.5 Representation Stages and Calibration Identity

At the fixed joint EEG checkpoint, we compare regional tokens before and after FiLM, backbone outputs, and retrieval-projector outputs. Each 256×1024 representation is flattened and ℓ_2-normalized for the same matching test. Table 16 shows how cross-participant matching changes along the trained network. Gallery shuffling retains the scoring labels while breaking the pairing; its near-chance scores verify the dependence on matched image conditions. The immediate pre/post-FiLM change describes representation geometry, while the following intervention measures its downstream use.

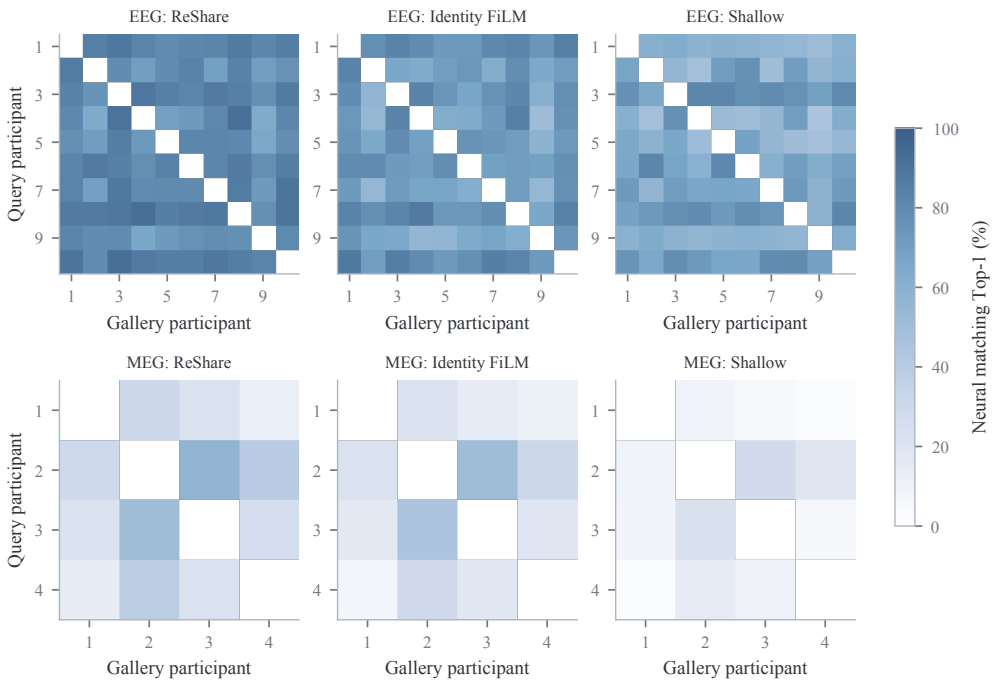


Figure 5: Neural-to-neural Top-1 accuracy (%) across distinct EEG and MEG participant pairs, with diagonal entries excluded.

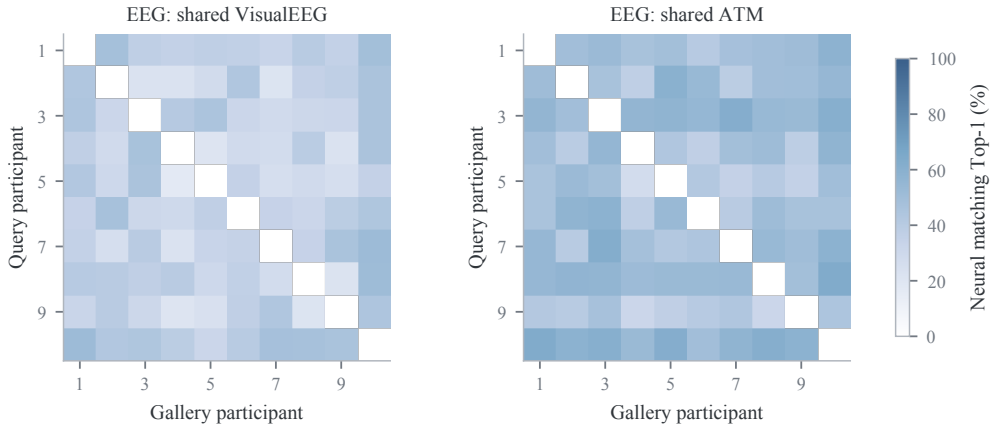


Figure 6: Neural-to-neural Top-1 accuracy (%) across all ninety EEG participant pairs for shared VisualEEG and ATM models.

Table 16: EEG neural-to-neural Top-1 / Top-5 (%) at successive representations of the fixed decoder with matched or shuffled galleries.

Representation	Matched gallery	Shuffled gallery
Before regional FiLM	9.96 / 26.52	0.46 / 2.34
After regional FiLM	9.97 / 26.82	0.49 / 2.37
Shared backbone output	50.30 / 79.33	0.51 / 2.33
Retrieval projector output	81.00 / 96.59	0.58 / 2.45

We then replace the trained participant's FiLM with identity modulation or another participant's parameters, keeping all other weights and inputs fixed. For each participant, we evaluate each alternative set of FiLM parameters separately and average the retrieval scores over the nine alternatives for EEG or the three for MEG. Table 17 measures neural-to-image retrieval under these interventions. The trained calibration's functional role is complemented by the training comparisons in Appendices D.1 and D.2.

Table 17: Neural-to-image Top-1 / Top-5 (%) using the correct participant's FiLM, identity FiLM, or another participant's FiLM, with the jointly trained weights otherwise fixed.

Dataset	Correct participant	Identity FiLM	Other-participant FiLM
EEG	95.70 / 99.80	93.30 / 99.50	81.69 / 96.27
MEG	61.88 / 84.63	60.63 / 82.75	51.08 / 76.42

E Model Reuse and Additional Analyses

E.1 Target Adaptation Settings

Adaptation starts from a LOSO model pretrained on the other nine EEG participants. For participant 1, this is the final SWAD checkpoint trained on participants 2–10, with identity target FiLM. Following shared-model reuse in MindEye2 (Scotti et al., 2024), each target experiment initializes from the pretrained neural model and uses a fresh optimizer. Within each of 1,654 training concepts, we permute ten image conditions and select a budget-dependent prefix. The 10–100% budgets select one through ten images per concept, with four responses averaged per image. These nested subsets vary the number of training images while retaining coverage of all training concepts.

Full adaptation updates the spatial and temporal encoders, embeddings, latent backbone, retrieval projector, and target regional FiLM. Other participants' FiLM rows remain fixed, as do CLIP and its cached targets. Interface-only adaptation updates the spatial expanders, temporal encoder, region/token embeddings, and target FiLM, totaling 5,003,328 parameters versus 130,123,840 for full adaptation. The input normalization, latent backbone and retrieval projector remain frozen. Gradients pass through these frozen modules to the interface, with training-time dropout retained.

Each experiment runs for 100 epochs with batch 128, BF16, gradient clipping at 1, and the final partial batch retained. AdamW uses weight decay 0.01, excluding bias, normalization, embedding and FiLM parameters. Peak learning rates are 3×10^{-5} for shared parameters and 3×10^{-4} for target FiLM. A fresh OneCycle schedule warms up for two epochs, with initial/final division factors of 25/1,000. Region dropout, BiMixCo, regression against unmixed visual targets and auxiliary losses match source training; target adaptation adds no EMA or SWAD. Evaluation uses fixed epoch-100 weights and raw retrieval scores after averaging 80 test responses per image.

Table 18 gives the participant-1 budget sweep with training seed 42 and subset seed 20260911. Full and interface-only runs use identical image subsets and update counts at each common budget. Source pretraining supplies additional supervision before these target updates. Appendix A specifies the preprocessing and how target training pairs are counted.

E.2 Ten-Participant Adaptation and Matched Scratch Controls

We adapted each of the ten EEG participants from its LOSO model trained on the other nine participants, using 1,654 paired training conditions (10%). Each source model excluded its target participant from supervised training and used identity FiLM before adaptation. Matched scratch controls used random initialization and the same target subsets, architecture, visual targets, losses and preprocessing. Training seed 101 and subset seed 2026092001 were fixed across targets; both initializations used 100 epochs, batch size 128 and 1,300 updates. Peak shared learning rates were 3×10^{-4} from scratch and 3×10^{-5} for adaptation; target FiLM used 3×10^{-4} in both. All scores use final-epoch weights and raw 200-way retrieval, with 80 test responses averaged per image. This seed-101 series supplements the seed-42 participant-1 comparison in Figure 3b.

Table 18: Participant-1 retrieval accuracy (%) across budgets of target training pairs for full-decoder and interface-only adaptation.

			Full decoder		Interface only	
Pairs	Images	Updates	Top-1	Top-5	Top-1	Top-5
0%	0	0	41.0	68.0	41.0	68.0
10%	1,654	1,300	78.0	94.0	64.5	92.0
20%	3,308	2,600	80.5	98.5	–	–
30%	4,962	3,900	87.5	100.0	68.5	95.0
40%	6,616	5,200	88.5	99.5	–	–
50%	8,270	6,500	90.0	99.0	77.0	99.0
60%	9,924	7,800	93.5	100.0	–	–
70%	11,578	9,100	94.0	100.0	79.0	99.5
80%	13,232	10,400	96.0	100.0	–	–
90%	14,886	11,700	96.5	100.0	–	–
100%	16,540	13,000	97.5	100.0	83.0	99.5

Table 19: Retrieval accuracy (%) for matched scratch and adaptation controls using 10% of target training pairs across all ten held-out EEG participants, with participant sample SD.

	From scratch				Adapted	
	Neural → image		Image → neural		Neural → image	
Participant	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
1	17.50	37.00	17.50	36.50	76.50	98.00
2	12.00	29.00	10.00	30.50	67.00	91.00
3	9.00	30.00	9.50	25.50	66.50	92.50
4	12.50	39.00	20.50	42.50	63.50	92.00
5	9.50	25.50	7.00	23.50	70.00	93.50
6	7.00	32.00	9.50	30.50	78.50	97.00
7	19.00	43.00	16.00	39.00	80.50	96.50
8	23.50	60.00	29.50	60.00	76.00	98.00
9	8.50	22.50	7.50	24.50	72.00	95.50
10	23.00	54.00	21.00	56.50	81.00	99.50
Mean ± SD	14.15 ± 6.14	37.20 ± 12.19	14.80 ± 7.37	36.90 ± 12.91	73.15 ± 6.24	95.35 ± 2.93

Adaptation improved neural-to-image Top-1 for every participant, with a mean gain of 59.00 percentage points over scratch training. The corresponding mean Top-5 gain was 58.15 percentage points.

E.3 Repeated Limited-Data Adaptation

We evaluated full-decoder adaptation to EEG participant 1 across three runs using 10/30/50% of target training pairs. All runs started from the same fixed source checkpoint and used the 100-epoch configuration in Appendix E.1. The three runs used training seed 42 with subset seed 20260911, and training seeds 101 and 102 with subset seeds 2026092001 and 2026092002. Each draw independently permuted the ten images within every training concept; the selected one, three, or five images formed nested budgets within that draw. Table 20 reports all three runs through their mean and sample standard deviation, capturing training randomness and subset selection at a fixed source model. Figure 3a summarizes the three runs; panels (b,c) show the seed-42 paired scratch and interface controls.

Table 20: Mean and sample standard deviation across three full-decoder adaptations to EEG participant 1.

Target pairs	Image conditions	Top-1 (%)	Top-5 (%)
10%	1,654	77.17 ± 0.76	96.00 ± 2.00
30%	4,962	89.33 ± 1.76	100.00 ± 0.00
50%	8,270	91.83 ± 1.61	99.67 ± 0.58

E.4 Adaptation across Models after Shared Pretraining

We compare ReShare, Shallow, VisualEEG and ATM on matched nested 10/30/50% target subsets for EEG participant 1. Each adapted model starts from joint pretraining on participants 2–10; its scratch control uses random initialization. Within each model, both conditions retain the architecture and use matched numbers of target training pairs and updates. Each run uses 100 target epochs, training seed 42 and final-epoch weights. Target training is separate for each budget. The ReShare scratch control uses a peak shared learning rate of 3×10^{-4}, with all other settings from Appendix E.1 retained.

Shallow. We evaluate two Shallow configurations, identified by their visual encoders in Table 21. The InternViT-6B configuration uses seventeen occipital/parietal channels and pooled block-28 features (Du et al., 2026). Its EEG means are rounded to half precision and then cast to float32. The CLIP ViT-L/14 configuration uses 63 channels, float32 means, and globally averaged block-12 features. Both use 512-dimensional alignment, dropout 0.3, symmetric contrastive loss and a fixed softplus logit scale. Training normalizes image features; evaluation normalizes both modalities. Source training uses 100 epochs, batch 1,024 and 14,400 updates. AdamW uses weight decay 10^{-4} and constant learning rates of 10^{-4} for source/scratch training and 3×10^{-5} for adaptation. Table 21 includes two target batch sizes for the InternViT-6B configuration and the CLIP ViT-L/14 control under the same budgets.

Table 21: Participant-1 Top-1 / Top-5 (%) from scratch and after shared pretraining on source participants for ReShare and Shallow controls.

Model	Batch	Target pairs	Updates	Scratch	Adapted
ReShare	128	10%	1,300	13.00 / 33.00	78.00 / 94.00
ReShare	128	30%	3,900	40.00 / 67.50	87.50 / 100.00
ReShare	128	50%	6,500	49.00 / 79.50	90.00 / 99.00
Shallow (InternViT-6B)	128	10%	1,300	23.50 / 60.50	56.50 / 84.00
Shallow (InternViT-6B)	128	30%	3,900	40.00 / 65.00	61.50 / 87.50
Shallow (InternViT-6B)	128	50%	6,500	56.50 / 83.00	61.00 / 90.00
Shallow (InternViT-6B)	1024	10%	200	26.00 / 57.50	55.50 / 83.00
Shallow (InternViT-6B)	1024	30%	500	47.50 / 80.50	67.50 / 89.50
Shallow (InternViT-6B)	1024	50%	900	59.00 / 88.00	76.50 / 93.50
Shallow (CLIP ViT-L/14)	128	10%	1,300	8.50 / 28.00	22.00 / 45.50
Shallow (CLIP ViT-L/14)	128	30%	3,900	11.00 / 35.50	28.00 / 55.50
Shallow (CLIP ViT-L/14)	128	50%	6,500	20.00 / 47.50	29.00 / 59.50

VisualEEG and ATM. VisualEEG retains its spatial and temporal EEG encoder and visual feature fusion module (Liu et al., 2026). The visual module combines ResNet-50 image features at twelve blur levels with four learned feature vectors through a learned weighted sum, followed by a feature adapter. Training uses symmetric cross-entropy with unit logit scale. Its EEG means are rounded to half precision and then cast to float32. Training normalizes image features, while evaluation normalizes both modalities. ATM uses the participant-conditioned EEG encoder from its released joint-training implementation (Li et al., 2024), together with normalized ViT-H/14 image and text features and a learned logit scale. The image and text alignment losses have weights 0.99 and 0.01,

respectively. Each participant has a dedicated input projection and embedding. The target participant's projection weights, biases, and embedding are initialized by averaging the corresponding parameters across source participants. The released visual features are paired with EEG conditions by sorting concept folder names and then image filenames within each folder lexicographically, following the ordering used by the released implementation.

VisualEEG source training uses 150 epochs and batch 1,024; ATM uses forty epochs and batch 64. Each source epoch traverses all source participants through interleaved batches, each containing recordings from a single participant. Both use AdamW with weight decay 0.01. Constant source/scratch learning rates are 10^{-3} for VisualEEG and 3×10^{-4} for ATM; adaptation uses 10^{-4} and 3×10^{-5}, respectively. Table 22 reports the paired results.

Table 22: Participant-1 Top-1 / Top-5 (%) from scratch and after shared pretraining on source participants for VisualEEG and ATM.

Model	Batch	Target pairs	Updates	Scratch	Adapted
VisualEEG	1,024	10%	200	16.00 / 42.50	55.00 / 86.50
VisualEEG	1,024	30%	500	63.00 / 84.50	66.00 / 93.00
VisualEEG	1,024	50%	900	71.00 / 91.00	70.00 / 94.50
ATM	64	10%	2,600	4.50 / 13.00	20.50 / 49.00
ATM	64	30%	7,800	12.00 / 33.50	25.50 / 55.50
ATM	64	50%	13,000	14.50 / 43.00	29.50 / 62.00

All three baseline families use 15,710 source training images and 830 validation images from 83 disjoint concepts. Source test images are excluded, and final source weights initialize adaptation. These shared-training extensions retain native visual targets and model-specific optimization; numbers of target training pairs and updates are matched within each model. Figure 7 complements the main Top-1 analysis with Top-5 results.

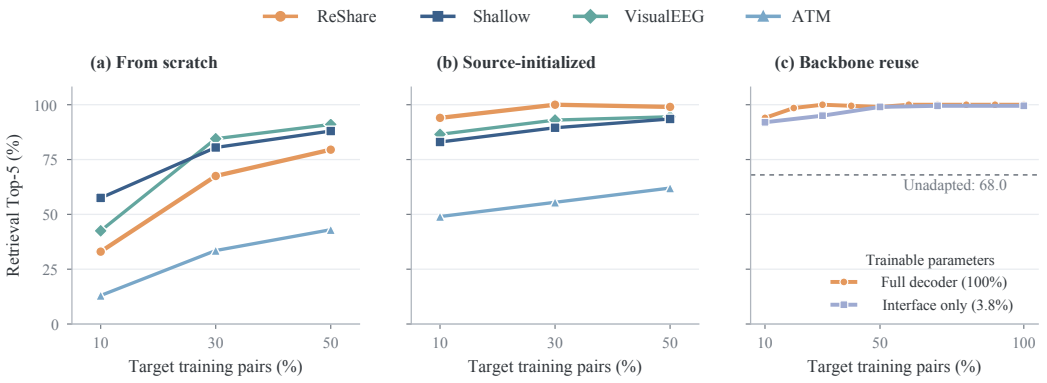


Figure 7: Participant-1 Top-5 retrieval from scratch (a), after shared pretraining (b), and with full-decoder or interface-only adaptation (c).

E.5 Sensitivity to Test Repetitions

We evaluate retrieval at reduced repeat counts with fixed model weights, where repeat count is the number of test responses averaged per image. For each participant and image, five uniform permutations of the available responses provide nested repeat subsets, shared across methods without session stratification. Scores at each reduced repeat count are averaged across draws and participants; averaging all available responses gives identical inputs across draws. These draws measure variation across sampled subsets of test responses. No new pooled normalization or score calibration is fitted.

ReShare uses final-epoch weights from its jointly trained EEG and MEG models after 500 and 200 epochs, respectively. Independent and shared Shallow EEG controls use fixed epoch-50 weights; independent Shallow MEG uses epoch-50, seed-24 models. VisualEEG uses models trained with seed 25 and selected using source validation (Liu et al., 2026). Each method retains its native channels, time window, precision and visual targets. These checkpoint reevaluations provide the curves in Figure 8, separately from published multi-seed averages reported in the main tables. Tables 23 and 24 give the corresponding values.

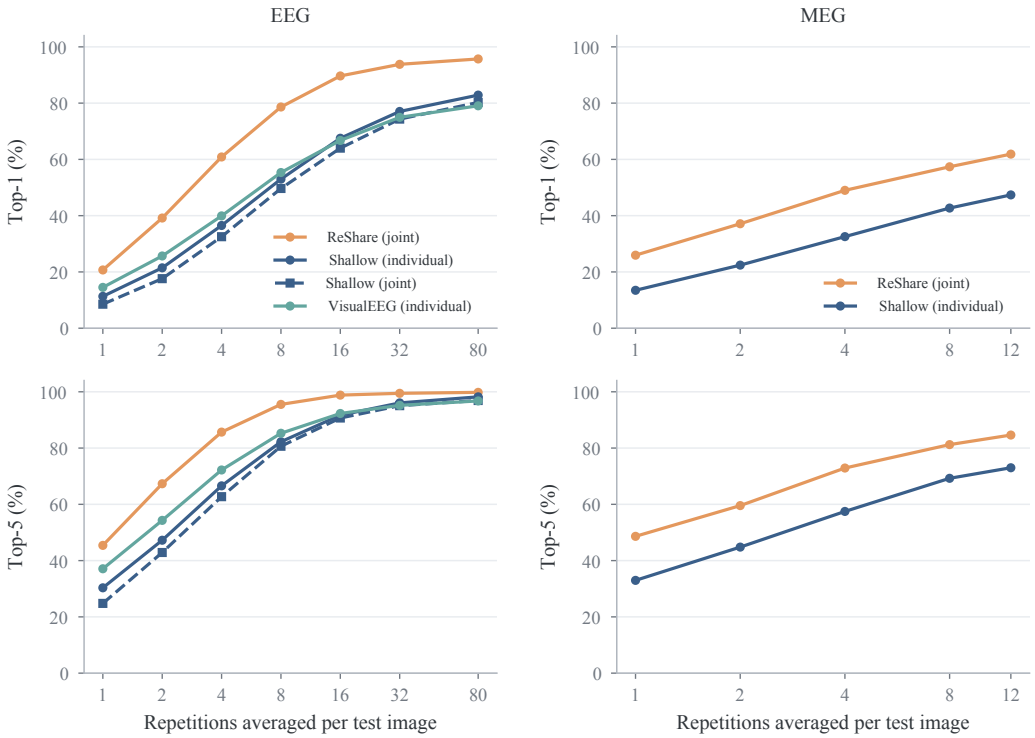


Figure 8: Retrieval accuracy versus the number of averaged test responses at fixed model weights and native model inputs.

Table 23: EEG-to-image Top-1 / Top-5 (%) across repeat counts.

Repeats	ReShare	Shallow individual	Shallow joint	VisualEEG individual
1	20.70 / 45.41	11.32 / 30.33	8.53 / 24.79	14.51 / 37.10
2	39.16 / 67.32	21.46 / 47.24	17.61 / 42.88	25.67 / 54.30
4	60.86 / 85.67	36.51 / 66.57	32.52 / 62.72	39.92 / 72.21
8	78.62 / 95.52	52.99 / 82.24	49.66 / 80.64	55.33 / 85.26
16	89.66 / 98.82	67.52 / 91.68	64.00 / 90.68	66.73 / 92.29
32	93.79 / 99.46	77.04 / 96.05	74.32 / 95.07	74.95 / 95.18
80	95.70 / 99.80	82.85 / 98.20	80.25 / 96.90	79.05 / 96.75

With four EEG responses, ReShare reaches 60.86/85.67%, compared with 36.51/66.57% for individual Shallow and 39.92/72.21% for VisualEEG. One MEG response yields 25.95/48.63%, compared with 13.48/32.98% for individual Shallow. The retrieval advantage persists under reduced averaging, while all comparisons retain the corresponding input budgets.

Table 24: MEG-to-image Top-1 / Top-5 (%) across repeat counts.

Repeats	ReShare	Shallow individual
1	25.95 / 48.63	13.48 / 32.98
2	37.13 / 59.55	22.43 / 44.80
4	49.00 / 72.90	32.55 / 57.48
8	57.38 / 81.23	42.70 / 69.25
12	61.88 / 84.63	47.38 / 73.00