AFoB: Adaptive Foveated Blur with Dual-Region Preservation for EEG-Image Retrieval

Anonymous authors
Paper under double-blind review

Abstract

Brain-signal-based image retrieval commonly aligns EEG or MEG recordings with pretrained visual representations, making the construction of the image-side target an important design choice. Variability in brain responses and informative content beyond the image center motivate two complementary questions: how much visual detail should be retained, and where should it be preserved? We propose **Adaptive Foveated Blur (AFoB)**, a perception-inspired framework that combines response-conditioned blur with dual-region detail preservation. A joint predictor takes brain-signal and fixed-blur reference-image embeddings and estimates the background-blur strength and concentration parameters controlling the extents of two preserved regions. One region remains centered on the image, while the second is located at the centroid of a low-level saliency map computed from the clean image without a learned localization module. A differentiable soft mask combines the two regions and blends the original image with its blurred counterpart. A frozen CLIP encoder maps the resulting image to a target for contrastive brain–image alignment, through which the brain encoder and foveation predictor are jointly trained. On 200-way zero-shot retrieval, AFoB improves Top-1 / Top-5 accuracy over UBP from 50.2% / 80.0% to 53.8% / 84.3% in the intra-subject setting and from 12.2% / 32.0% to 17.1% / 40.1% in the inter-subject setting on THINGS-EEG, with a proportionally larger gain in the latter, and from 25.6% / 55.6% to 39.6% / 69.8% intra-subject on THINGS-MEG. Analysis of the predicted parameters on THINGS-EEG reveals variation across images within individual subjects, rather than a single transformation per subject. AFoB provides a computational framework for response-conditioned visual abstraction and motivates further research connecting selective information preservation with models of human visual cognition.

1 Introduction

Brain-signal-based image retrieval aims to identify a viewed image from a set of candidates using concurrently recorded brain activity. A common approach maps electroencephalography (EEG) or magnetoencephalography (MEG) recordings into the embedding space of a pretrained vision–language model, such as CLIP (Radford et al., 2021), through contrastive learning (Song et al., 2024; Wu et al., 2025). The encoded stimulus image serves as the alignment target for the brain encoder. Noisy recordings can make some visual details difficult to decode reliably, even when those details are available in the stimulus image (Wu et al., 2025). An important design question is therefore not only how to represent the recorded brain activity, but also which visual information the target should retain.

The Uncertainty-aware Blur Prior (UBP) (Wu et al., 2025) addresses this question by blurring stimulus images before visual encoding. It uses a center-focused foveated transformation and adjusts the blur level according to uncertainty derived from brain–image similarity, improving retrieval over alignment with unblurred images. Its spatial preservation profile nevertheless remains centered on a prescribed location. Building on this approach, we investigate adapting both the amount and the spatial distribution of preserved detail: *how much detail should be retained for a particular brain response, and where should it be preserved?*

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.

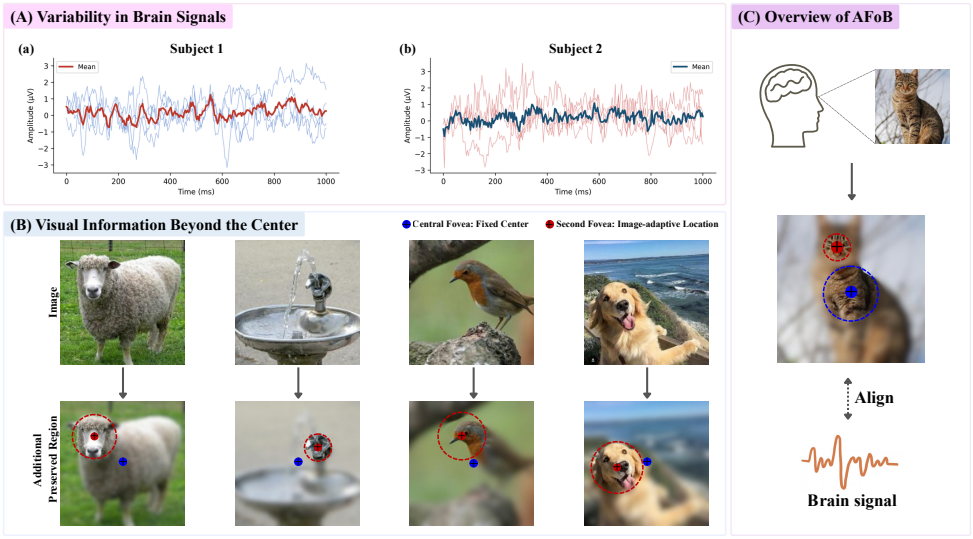

Figure 1: **Motivation of AFoB.** (A) EEG trials (thin) and means (thick) illustrate response variability. (B) Informative content can lie beyond the image center. (C) AFoB combines response-conditioned preservation extent and blur with image-dependent additional preservation. Blue/red markers denote central/additional regions; (B) isolates the latter. The views illustrate transformed alignment targets, not subjective perception.

Figure 1A illustrates response variability across subjects and trials. This motivates conditioning preservation extent and background blur on the paired brain response and image, rather than assigning one setting per subject. Figure 1B highlights informative content beyond the center (Geisler et al., 2006; Pramod et al., 2022), which a center-focused profile may attenuate even when blur strength is adjusted. We therefore retain an additional image-dependent region alongside the central region. Response conditioning controls the amount of detail, while image content guides additional preservation.

We propose **Adaptive Foveated Blur (AFoB)** to combine these choices (Figure 1C). For each brain-signal–image pair, a joint predictor estimates background blur and the extents of two preserved regions from brain and fixed-reference-image embeddings. The first region is central; the second is located by clean-image saliency, independently of the brain signal. A differentiable mask blends clean and blurred images before frozen CLIP encoding and contrastive alignment. Figure 2 details the computation.

Experiments on THINGS-EEG and THINGS-MEG analyze paired-target ranking and parameter distributions. Section 4.1 defines the evaluation inputs and distinguishes the reported diagnostic from reference retrieval. We evaluate AFoB on 200-way paired-target ranking using THINGS-EEG and THINGS-MEG. Compared with our reproduced UBP baseline, which follows its own original, non-adaptive test-time construction, AFoB improves Top-1 / Top-5 ranking from 50.2% / 80.0% to 53.8% / 84.3% in the intra-subject EEG setting and from 12.2% / 32.0% to 17.1% / 40.1% in the inter-subject EEG setting. On THINGS-MEG, intra-subject ranking improves from 25.6% / 55.6% to 39.6% / 69.8%. These results characterize AFoB's design choices under this diagnostic across the evaluated EEG and MEG settings. Our contributions are:

- We propose a response-conditioned predictor jointly controlling blur and the extent of two preserved regions for each brain-signal–image pair.
- We introduce a differentiable dual-fovea transformation combining this central region with an image-dependent, saliency-located second region.
- Controlled ablations on THINGS-EEG and THINGS-MEG show that per-sample blur adaptation and brain-signal dependence are the most consistently significant contributors to AFoB's improvement, replicating across both modalities.

2 Related Work

2.1 Visual Decoding from EEG and MEG

Pretrained visual representations provide a common interface between neural recordings and visual recognition. NICE (Song et al., 2024) learns EEG–image alignment through contrastive learning and evaluates recognition of unseen categories using representations from pretrained visual encoders, including CLIP (Radford et al., 2021). BraVL (Du et al., 2023) takes a multimodal generative approach, learning joint brain–visual–linguistic representations with intra- and inter-modality mutual-information regularization. For MEG, Benchetrit et al. (2024) investigate several pretrained visual representations and combine neural decoding with image generation, showing that the choice of visual representation and the response interval affect decodability. Beyond retrieval, ATM (Li et al., 2024) combines a spatiotemporal EEG encoder with guided diffusion for image reconstruction, and EEG2Video (Liu et al., 2024) extends this to decoding dynamic visual perception across video frames. AVDE (Dai et al., 2026) retains contrastive EEG–image alignment while using a pretrained EEG model and an autoregressive next-scale image decoder. AFoB addresses a different component of this pipeline: the image-side transformation that defines the target for neural alignment, rather than a new neural backbone or generative decoder.

2.2 Image-Side Representations for Neural Alignment

Visual representations can be adapted at the feature level or through transformations of the stimulus image. VE-SDN (Chen et al., 2024) reprojects visual and EEG features into a joint semantic space and separates semantic-related information from modality-specific components. The recent FSDBN preprint (Liu et al., 2026) combines saliency-guided foreground–background separation and semantic feature fusion with dynamic brain-network modeling. These approaches modify the representations used for alignment rather than treating the image embedding as an immutable target. Closest to AFoB, UBP (Wu et al., 2025) applies foveated blur before visual encoding and adapts the training blur level using uncertainty estimated from paired brain–image similarities. Its spatial preservation profile is centered on the image. AFoB builds on this image-side approach, but jointly learns background-blur strength and the preservation extents of two regions from brain-signal and fixed-reference-image features. A separate, non-learned saliency computation locates the additional region, and a differentiable mask composes the alignment target before frozen CLIP encoding. The distinction is therefore joint control of preservation extent and blur together with image-dependent additional preservation, not the introduction of pair-conditioned adaptation itself.

2.3 Foveated Vision and Saliency-Guided Processing

Foveated processing reflects the spatially non-uniform resolution of human vision, while peripheral information can still contribute to visual behavior. Using gaze-contingent displays, Geisler et al. (2006) show that attenuating peripheral resolution affects visual search and eye-movement patterns. Pramod et al. (2022) study object recognition with foveated images and find that recognition depends on the peripheral-blur profile in both neural-network and human experiments. Outside perception research, DeepFovea (Kaplanyan et al., 2019) shows that a learned generative model can reconstruct plausible peripheral video from a sparse, foveated pixel stream for rendering and compression, illustrating that computational foveation is useful even without any brain-signal input. These findings motivate selective preservation without implying that an image's noncentral content can be discarded.

Computational saliency models combine low-level visual cues to identify conspicuous locations (Itti et al., 1998). Recurrent attention models instead learn sequences of local observations for recognition (Mnih et al., 2014). NevaClip (Zanca et al., 2023) connects foveation with CLIP by aligning foveated-image and caption representations to guide scanpath generation. AFoB does not estimate a scanpath: it uses the centroid of a simple edge-and-color saliency map to locate one additional preserved region, while the brain-conditioned predictor controls preservation extent and background blur. Here, the two foveae are computational preservation regions, not two anatomical foveae or measured fixation locations. This separates the amount-of-detail and location questions in Figure 1 while retaining a single image-side alignment target in Figure 2.

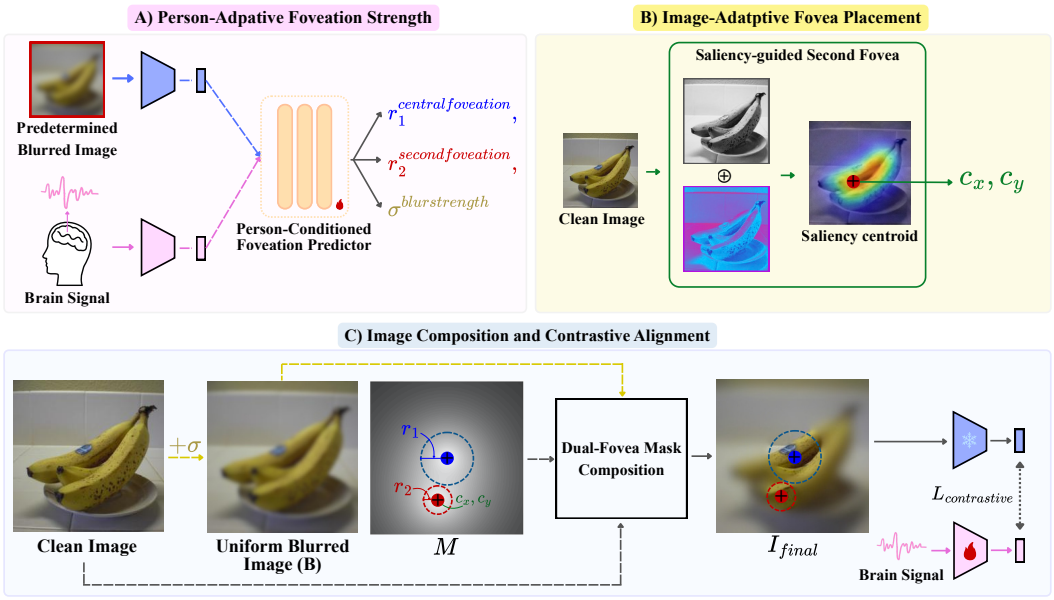

Figure 2: **AFoB training pipeline. (A) Person-conditioned foveation prediction.** A normalized, gradient-detached brain embedding is concatenated with a precomputed reference-image embedding obtained from a fixed foveated-blur transformation. The predictor jointly outputs the concentrations (r_1, r_2) of two preserved regions and the background-blur strength σ. **(B) Saliency-guided second-fovea placement.** A non-learned computation combines grayscale Sobel magnitude with channel-wise color standard deviation, smooths and normalizes the map, and returns its centroid (c_x, c_y). **(C) Image composition and contrastive alignment.** The concentrations and centers define a dual-fovea preservation mask; σ defines a uniformly blurred image. The mask blends the clean and blurred images into an adaptive alignment target. The contrastive loss trains the brain encoder and foveation predictor, while CLIP's weights remain frozen and gradients through the adaptive-image input are retained.

3 Method

AFoB constructs adaptive image-side targets by controlling how much detail to preserve and where, following Figure 1. Figure 2 comprises a response-conditioned foveation predictor (A), image-only saliency-based placement (B), and image composition with contrastive alignment (C).

3.1 Problem Formulation

For a training pair (e_i, x_i), e_i is a preprocessed brain recording and x_i the clean image. Let g_θ be the trainable brain encoder, f the frozen CLIP image encoder including fixed input normalization, and $\mathcal{N}$ L2 feature normalization. With a fixed foveated-blur transformation $\mathcal{T}_0$, we define the reference image x_i^0, reference embedding z_i^0, and brain embedding z_i^{brain} as

$$x_i^0 = \mathcal{T}_0(x_i), \qquad z_i^0 = \mathcal{N}\big(f(x_i^0)\big), \qquad z_i^{\text{brain}} = g_\theta(e_i). \tag{1}$$

The precomputed, cached reference embedding conditions the predictor; it is not the final alignment target. The clean image supplies saliency estimation and adaptive composition.

3.2 Person-Conditioned Foveation Predictor

The predictor (Figure 2A) jointly estimates the central and second-fovea concentrations $r_{1,i}, r_{2,i}$ and Gaussian blur strength σ_i for each brain-signal–image pair. Larger concentrations narrow the preserved regions, whereas larger σ_i increases background blur. Let sg denote stop-gradient, $[\cdot;\cdot]$ concatenation, and h_ϕ the predictor including bounded output transformations. The conditioning

vector u_i and predictions are

$$u_i = [\mathrm{sg}(\mathcal{N}(z_i^{\mathrm{brain}})); z_i^0], \qquad (r_{1,i}, r_{2,i}, \sigma_i) = h_\phi(u_i). \tag{2}$$

The MLP has hidden widths 256 and 64 with GELU activations. Sigmoid outputs are rescaled logarithmically for the concentrations and linearly for σ_i (Appendix B). Fovea locations are determined separately.

3.3 Saliency-Guided Second-Fovea Placement

For pixel $\mathbf{p} = (u, v)$ on grid Ω, let $\bar{x}_i$ be the channel-wise mean of x_i, K_u, K_v the horizontal and vertical Sobel kernels, $*$ convolution, and std_c the channel-wise standard deviation. The clean-image saliency map S_i (Figure 2B) combines edge magnitude and color dispersion, with $\epsilon = 10^{-8}$:

$$S_i(\mathbf{p}) = 0.6\sqrt{(K_u * \bar{x}_i)^2(\mathbf{p}) + (K_v * \bar{x}_i)^2(\mathbf{p}) + \epsilon} + 0.4\,\mathrm{std}_c(x_i(\mathbf{p})). \tag{3}$$

Three successive 3×3 averaging filters and spatial min–max normalization yield Q_i. The weights P_i and second center $\mathbf{c}_{2,i} = (c_{x,i}, c_{y,i})$ are

$$P_i(\mathbf{p}) = \frac{Q_i(\mathbf{p})}{\sum_{\mathbf{q}\in\Omega} Q_i(\mathbf{q}) + \epsilon}, \qquad \mathbf{c}_{2,i} = \sum_{\mathbf{p}\in\Omega} P_i(\mathbf{p})\,\mathbf{p}. \tag{4}$$

This image-only centroid, rather than a saliency maximum, is computed without learned parameters or gradient tracking and supplies the second mask's center. It need not lie outside the central region.

3.4 Dual-Fovea Mask and Image Composition

Figure 2C combines the predicted parameters and centers. For image height H and width W, define the fixed center $\mathbf{c}_{1,i} = (W/2, H/2)$ and diagonal length $D = \sqrt{H^2 + W^2}$. The local preservation weights $m_{k,i}$ for $k \in \{1, 2\}$ and their soft union M_i are

$$\begin{aligned} m_{k,i}(\mathbf{p}) &= \exp\left(-r_{k,i}\frac{\|\mathbf{p} - \mathbf{c}_{k,i}\|_2}{D/2}\right), \qquad k \in \{1, 2\}, \\ M_i(\mathbf{p}) &= 1 - (1 - m_{1,i}(\mathbf{p}))(1 - m_{2,i}(\mathbf{p})). \end{aligned} \tag{5}$$

Numerical stabilizers in the distances are omitted. The soft union preserves detail from either region with weights in $[0, 1]$. Let G_{σ_i} be a Gaussian kernel of standard deviation σ_i and $\odot$ pixelwise multiplication broadcast over channels. The blurred image B_i and adaptive target image $I_{\mathrm{final},i}$ are

$$B_i = x_i * G_{\sigma_i}, \qquad I_{\mathrm{final},i} = M_i \odot x_i + (1 - M_i) \odot B_i. \tag{6}$$

Larger M_i retains more clean-image detail.

3.5 Contrastive Objective and Gradient Flow

For a minibatch of N matched training pairs, let z_i^{dyn} be the normalized adaptive image embedding and s_{ij} the logit between brain sample i and image target j. A learned scalar a gives the positive logit scale $\alpha = \mathrm{softplus}(a)$:

$$z_i^{\mathrm{dyn}} = \mathcal{N}(f(I_{\mathrm{final},i})), \qquad s_{ij} = \alpha\,(z_i^{\mathrm{brain}})^\top z_j^{\mathrm{dyn}}. \tag{7}$$

No additional L2 normalization is applied to the brain-encoder output in these logits. Each target is constructed from its matched training pair and shared across batch comparisons. With diagonal positives and off-diagonal negatives, the symmetric contrastive loss is

$$\mathcal{L}_{\mathrm{contrastive}} = -\frac{1}{2N}\sum_{i=1}^{N}\left[\log\frac{\exp(s_{ii})}{\sum_{j=1}^{N}\exp(s_{ij})} + \log\frac{\exp(s_{ii})}{\sum_{j=1}^{N}\exp(s_{ji})}\right]. \tag{8}$$

We optimize g_θ, h_ϕ, and a without separate foveation-parameter supervision. Gradients update g_θ through its brain embeddings and h_ϕ through image composition and frozen CLIP. CLIP retains input gradients, while sg blocks the target-construction branch from updating g_θ. Saliency centers are not optimized.

4 EXPERIMENTS

We study the behavior of adaptive image-side targets on THINGS-EEG and THINGS-MEG. The experiments connect the two motivations in Figure 1 to the modules in Figure 2: response-conditioned control of preservation extent and blur, and image-dependent placement of an additional preserved region. The available AFoB results use the paired-target evaluation defined below.

4.1 EXPERIMENTAL SETUP

Data and model settings. The EEG experiments use the 10-subject THINGS-EEG dataset with 1,654 training concepts and 200 held-out test concepts. We average repeated recordings of each stimulus, using four repetitions per training image and 80 per test image, and retain the 17 posterior channels specified in our configuration. We report intra-subject and leave-one-subject-out EEG experiments, and intra-subject experiments on the four-subject THINGS-MEG dataset. AFoB refers to the `saliency` configuration throughout: the predictor jointly estimates the two concentrations and background blur strength, while the additional center is computed from clean-image saliency without learned parameters. The image encoder remains frozen, and the brain encoder and foveation predictor are trained with the contrastive objective in Section 3.

Paired-target evaluation. For each candidate stimulus, the existing evaluation constructs an adaptive target using that stimulus's ground-truth matched held-out brain recording. Let e_i denote a query recording, x_j a candidate image, and e_j the recording known to correspond to x_j. Let g_θ and f be the brain and image encoders, $\mathcal{N}$ L2 normalization, and $\mathcal{T}_\phi(e, x)$ the AFoB transformation defined in Section 3. The normalized query b_i, candidate target v_j^{pair}, and score s_{ij}^{pair} are

$$\begin{aligned} b_i &= \mathcal{N}(g_\theta(e_i)), \\ v_j^{\text{pair}} &= \mathcal{N}(f(\mathcal{T}_\phi(e_j, x_j))), \\ s_{ij}^{\text{pair}} &= b_i^\top v_j^{\text{pair}}. \end{aligned} \tag{9}$$

We rank each query against its subject's 200 candidate targets and report Top-1, Top-5, mAP, and mean matched-pair cosine similarity. Here, the known held-out correspondence is used before ranking to construct the candidates; it is not used solely to assess correctness. This is a diagnostic with privileged pairing information, rather than image-gallery retrieval with an unknown query–image correspondence.

Reference results and reporting. The UBP reference values and earlier baseline values compiled by Wu et al. (2025) are retained for context in separate table blocks. The released UBP evaluator uses fixed reference image features at test time, whereas the AFoB rows use Equation equation 9. Their numerical differences therefore do not establish a controlled retrieval improvement under identical evaluation inputs. Top-1 and Top-5 are percentages; ablation-table mAP and similarity are also multiplied by 100. Parenthesized standard deviations summarize variation across subjects, not random seeds. We describe numerical patterns without importing unverified significance tests from other runs or evaluation procedures.

4.2 RESULTS ON THINGS-EEG

Under paired-target evaluation, AFoB obtains mean Top-1 / Top-5 scores of 53.8% / 84.3% intra-subject and 17.1% / 40.1% inter-subject (Table 1). The corresponding UBP reference values are 50.2% / 80.0% and 12.2% / 32.0%, obtained under its distinct evaluation procedure. In the inter-subject setting, AFoB constructs candidate targets using the held-out subject's matched recordings.

4.3 RESULTS ON THINGS-MEG

The intra-subject MEG experiments yield mean Top-1 / Top-5 scores of 39.6% / 69.8% for AFoB under paired-target evaluation (Table 2). The UBP reference row records 25.6% / 55.6% under its own evaluator. The MEG experiments apply the AFoB architecture to a second recording modality using separately fitted models; they do not test transfer of an EEG-trained model to MEG. We report no inter-subject MEG result.

Table 1: Top-1 and Top-5 accuracy (%) for 200-way zero-shot retrieval on THINGS-EEG. All rows except UBP (ours) and AFoB (ours) are as reported in Wu et al. (2025), not independently reproduced by us.

Method	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Subject 8		Subject 9		Subject 10		Avg	
	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5
Intra-subject: train and test on one subject																						
BraVL	6.1	17.9	4.9	14.9	5.6	17.4	5.0	15.1	4.0	13.4	6.0	18.2	6.5	20.4	8.8	23.7	4.3	14.0	7.0	19.7	5.8	17.5
NICE	13.2	39.5	13.5	40.3	14.5	42.7	20.6	52.7	10.1	31.5	16.5	44.0	17.0	42.1	22.9	56.1	15.4	41.6	17.4	45.8	16.1	43.6
NICE-SA	13.3	40.2	12.1	36.1	15.3	39.6	15.9	49.0	9.8	34.4	14.2	42.4	17.9	43.6	18.2	50.2	14.4	38.7	16.0	42.8	14.7	41.7
NICE-GA	15.2	40.1	13.9	40.1	14.7	42.7	17.6	48.9	9.0	29.7	16.4	44.4	14.9	43.1	20.3	52.1	14.1	39.7	19.6	46.7	15.6	42.8
ATM-S	25.6	60.4	22.0	54.5	25.0	62.4	31.4	60.9	12.9	43.0	21.3	51.1	30.5	61.5	38.8	72.0	34.4	51.5	29.1	63.5	28.5	60.4
VE-SDN	32.6	63.7	34.4	69.9	38.7	73.5	39.8	72.0	29.4	58.6	34.5	68.8	34.5	68.3	49.3	79.8	39.0	69.6	39.8	75.3	37.2	69.9
UBP	40.0	71.0	**50.5**	83.0	49.5	81.5	49.0	78.5	44.5	72.5	**59.0**	83.0	47.0	80.0	56.0	86.0	45.0	76.0	61.5	88.5	50.2	80.0
AFoB (ours)	48.5	82.0	48.0	83.5	52.5	84.5	53.0	82.5	47.5	77.5	52.5	87.0	53.0	85.0	65.0	88.0	52.5	82.5	65.5	90.5	**53.8**	**84.3**
Inter-subject: leave one subject out for test																						
BraVL	2.3	8.0	1.5	6.3	1.4	5.9	1.7	6.7	1.5	5.6	1.8	7.2	2.1	8.1	2.2	7.6	1.6	6.4	2.3	8.5	1.8	7.0
NICE	7.6	22.8	5.9	20.5	6.0	22.3	6.3	20.7	4.4	18.3	5.6	22.2	5.6	19.7	6.3	22.0	5.7	17.6	8.4	28.3	6.2	21.4
NICE-SA	7.0	22.6	6.6	23.2	7.5	23.7	5.4	21.4	6.4	22.2	7.5	22.5	3.8	19.1	8.5	24.4	7.4	22.3	9.8	29.6	7.0	23.1
NICE-GA	5.9	21.4	6.4	22.7	5.5	20.1	6.1	21.0	4.7	19.5	6.2	22.5	5.9	19.1	7.3	25.3	4.8	18.3	6.2	26.3	5.9	21.6
ATM-S	10.5	26.8	7.1	24.8	11.9	33.8	14.7	39.4	7.0	23.9	11.1	35.8	16.1	43.5	15.0	40.3	4.9	22.7	20.5	46.5	11.8	33.7
UBP	**12.5**	**29.5**	14.0	38.0	8.5	25.5	13.5	34.0	8.5	31.5	**14.0**	30.0	**9.5**	23.5	12.2	32.2	13.0	32.0	16.5	43.5	12.2	32.0
AFoB (ours)	10.5	29.0	19.5	40.5	13.5	35.5	20.5	44.0	19.0	43.5	12.0	35.0	9.0	28.5	15.0	41.5	24.0	51.5	28.0	51.5	**17.1**	**40.1**

Table 2: Top-1 and Top-5 accuracy (%) for 200-way zero-shot retrieval on THINGS-MEG, intra-subject setting. All rows except UBP (ours) and AFoB (ours) are as reported in Wu et al. (2025), not independently reproduced by us.

Method	Subject 1		Subject 2		Subject 3		Subject 4		Avg	
	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5	top-1	top-5
NICE	9.6	27.8	18.5	47.8	14.2	41.6	9.0	26.6	12.8	36.0
NICE-SA	9.8	27.8	18.6	46.4	10.5	38.4	11.7	27.2	12.7	35.0
NICE-GA	8.7	30.5	21.8	56.6	16.5	49.7	10.3	32.3	14.3	42.3
UBP	11.0	35.5	49.0	81.5	24.0	60.0	18.5	45.5	25.6	55.6
AFoB (ours)	24.5	53.5	61.0	87.0	45.0	78.5	28.0	60.0	**39.6**	**69.8**

4.4 Ablation Studies

We examine per-sample adaptation, fovea count, second-fovea placement, and predictor inputs on intra-subject EEG and MEG (Table 3), using the evaluation protocol in Section 4.1.

1. Per-sample adaptation and parameter distributions. We compare the predicted concentration parameters r_1, r_2 and background-blur strength σ with a dual-fovea control that replaces these predictions with their subject-specific means over the test set. Both configurations retain the saliency-based second center, so the control removes per-image variation in preservation extent and background blur without removing image-dependent placement. This is a test-derived subject-mean control, not an optimized fixed-blur baseline. AFoB records Top-1 / Top-5 scores of 53.8% / 84.3% on EEG, compared with 49.9% / 79.6% for the control, a difference of 3.9 / 4.7 percentage points. The corresponding MEG scores are 39.6% / 69.8% and 28.4% / 56.5%, a difference of 11.2 / 13.3 percentage points. These comparisons document the change in the reported scores when per-image predictions are replaced by subject-level constants.

Distribution of predicted parameters. Figure 3 examines the predictions behind this comparison on the THINGS-EEG intra-subject test set. Each box summarizes one subject's per-image predictions, while the dashed lines indicate the overall means: 10.1 for r_1, 9.8 for r_2, and 3.6 for σ. The distributions reveal variation both across images within a subject and across the separately fitted subject models. For example, Subject 3 exhibits a wider distribution of r_1 than Subject 2, whereas Subject 2 has a higher median σ than Subject 3. The two concentration distributions also differ in shape across subjects. The predictor therefore does not reduce to one global parameter tuple or one constant tuple per subject.

These changes have distinct effects on the transformed image: larger r_1 or r_2 concentrates preservation around the corresponding center, whereas larger σ increases background blur. Figure 3 thus

Table 3: Ablation results, mean (std). Left block: THINGS-EEG, intra-subject, over 10 subjects. Right block: THINGS-MEG, intra-subject, over 4 subjects. Each row changes exactly one factor relative to AFoB (top row); all other factors are held fixed as described in Section 4.4. †AFoB's EEG row here is a separate training run under the ablation pipeline and differs marginally from Table 1 due to ordinary run-to-run variation, not a methodological difference.

	THINGS-EEG				THINGS-MEG			
Method	Top-1	Top-5	mAP	Similarity	Top-1	Top-5	mAP	Similarity
AFoB (ours)†	**53.8 (6.4)**	**84.3 (3.6)**	**66.9 (5.1)**	**26.0 (3.6)**	**39.6 (16.8)**	**69.8 (15.6)**	**53.6 (16.4)**	**15.3 (3.2)**
1. Adaptive blur application *(Section 4.4, §1)*								
Fixed blur	49.9 (6.4)	79.6 (4.5)	63.0 (5.5)	23.7 (3.1)	28.4 (14.0)	56.5 (20.8)	41.5 (16.0)	11.9 (2.4)
2. Foveation count *(Section 4.4, §2)*								
Single-fovea	51.8 (7.2)	81.8 (5.0)	65.1 (6.0)	26.1 (4.4)	37.9 (17.6)	69.9 (18.3)	52.3 (16.7)	15.7 (3.1)
3. Second-fovea location method *(Section 4.4, §3)*								
Random	52.8 (8.1)	82.7 (3.4)	66.1 (6.1)	26.3 (3.7)	38.9 (18.0)	69.2 (15.1)	52.5 (16.7)	15.8 (2.7)
Learned	52.4 (8.2)	83.3 (5.5)	65.8 (6.8)	26.4 (3.5)	39.4 (13.8)	71.1 (14.9)	53.6 (13.6)	15.7 (3.0)
4. Foveation Predictor input *(Section 4.4, §4)*								
Image-only	49.0 (7.3)	79.8 (5.1)	62.6 (6.2)	24.6 (3.1)	26.0 (15.2)	54.8 (19.4)	38.9 (16.6)	9.6 (3.9)

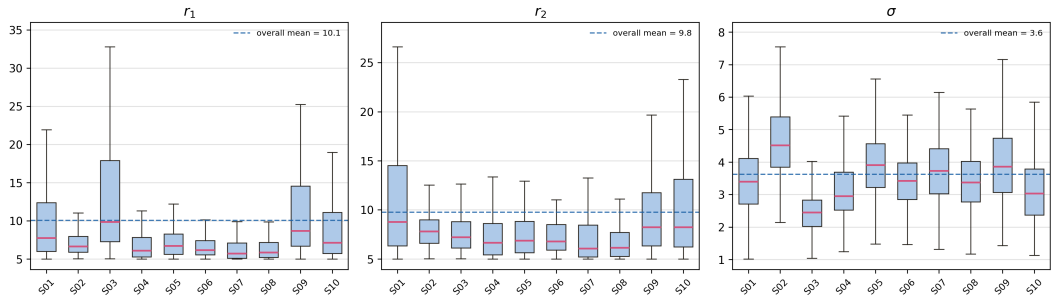

Figure 3: Distribution of AFoB's predicted (r_1, r_2, σ) across the 10 subjects (THINGS-EEG, intra-subject test set). Each box summarizes one subject's per-image predictions; the dashed line marks the overall mean across all subjects. Box height and whisker length reflect variability across images *within* a subject, while differences in box position across subjects reflect variability *between* subjects (Section 4.4, §1).

makes the adaptation in Figure 1A explicit at the parameter level and shows what is removed by subject-mean replacement. The boxplots describe model behavior rather than measured perceptual blur; they do not establish that within-subject variance dominates between-subject variance or that the predicted parameters quantify neural reliability.

2. Central versus dual-region preservation. The single-fovea variant retains adaptive control of the central concentration and background blur but removes the second preserved region. On EEG, its Top-1 / Top-5 scores are 51.8% / 81.8%, compared with 53.8% / 84.3% for AFoB, a numerical difference of 2.0 / 2.5 percentage points. On MEG, the scores change from 37.9% / 69.9% to 39.6% / 69.8%: Top-1 is higher, while Top-5 is essentially unchanged and numerically lower by 0.1 percentage points. The additional region is therefore associated with higher EEG scores, but its benefit is not uniform across datasets and metrics. This comparison addresses the motivation in Figure 1B by retaining additional detail beyond the central region, rather than moving or replacing the central fovea. Figure 4 contrasts the resulting single- and dual-region transformations.

3. Second-fovea placement. We compare the proposed non-learned saliency centroid with random placement and a learned spatial localization module, retaining the joint concentration/blur predictor and dual-fovea mask design. On EEG, random and learned placement record Top-1 / Top-5 scores of 52.8% / 82.7% and 52.4% / 83.3%, respectively, compared with 53.8% / 84.3% for saliency

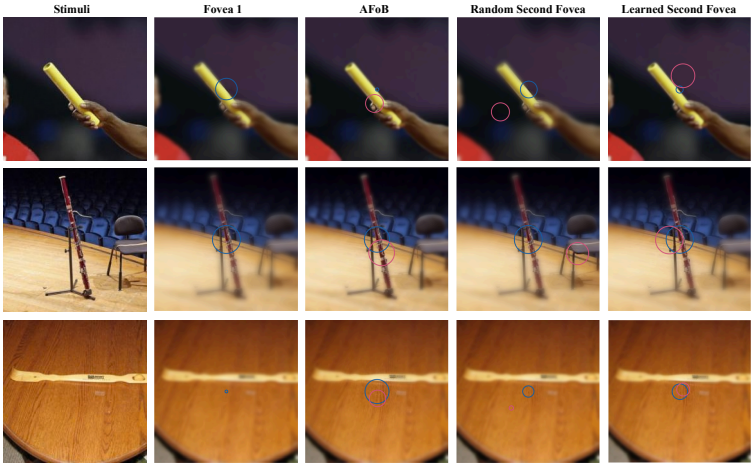

Figure 4: Qualitative comparison of the ablated foveation variants for three example test images (top two rows: Subject 3; bottom row: Subject 10). Blue circles mark the central fovea, present in every condition; pink circles mark the second fovea, absent in *Fovea 1* and set by saliency (*AFoB*), a random draw (*Random Second Fovea*), or a learned predictor (*Learned Second Fovea*). Radii and location are predicted independently by each condition's own trained model.

placement. The corresponding MEG scores are 38.9% / 69.2%, 39.4% / 71.1%, and 39.6% / 69.8%. Saliency placement gives the highest EEG Top-1 and Top-5 values among these configurations without a trainable localization module, whereas learned placement gives the highest MEG Top-5 value. The results do not establish a consistent ranking across both datasets. Figure 4 illustrates how the placement mechanisms change the location of additional detail preservation in selected examples; these examples are not a measurement of gaze or a dataset-wide localization assessment.

4. Brain-signal conditioning. The image-only variant removes the brain embedding from the foveation predictor while retaining the reference-image input. It records EEG Top-1 / Top-5 scores of 49.0% / 79.8% and MEG scores of 26.0% / 54.8%, compared with AFoB's 53.8% / 84.3% and 39.6% / 69.8%, respectively. On EEG, mAP and mean similarity also change from 62.6 and 24.6 to 66.9 and 26.0 on the table's $\times 100$ scale. Removing brain input changes both training-time conditioning and access to matched held-out recordings during target construction. The difference therefore does not isolate a training-time decoding benefit.

5 Conclusion

We introduced Adaptive Foveated Blur (AFoB), a perception-inspired approach to constructing image-side targets for brain–image alignment. AFoB jointly predicts background-blur strength and the preservation extents of two regions from brain-signal and reference-image features, while a non-learned saliency centroid locates the additional region. This design addresses both how much visual detail to retain and where to preserve it. In the reported EEG ablation, the dual-fovea configuration records Top-1 / Top-5 scores of 53.8% / 84.3%, compared with 51.8% / 81.8% for the single-fovea variant, a numerical increase of 2.0 / 2.5 percentage points. The parameter analysis further shows that AFoB produces image-varying transformations within each subject rather than a single subject-level setting. We hope this work encourages further research on models of human visual cognition that connect individual variability, selective attention, and non-uniform detail preservation with neural representations. An important direction is to test such perception-inspired modeling assumptions against both brain recordings and behavioral observations.

AI USE STATEMENT

We used generative AI tools, including ChatGPT, to assist with drafting sections of the manuscript, polishing language and organization, and retrieving, identifying, and summarizing relevant literature. AI assistance was also used to explore preliminary figure layouts, draft and revise code, and obtain feedback on implementation, evaluation procedures, and the interpretation of results. The authors retain responsibility for verifying the cited sources, implementation, and scientific claims, and take full responsibility for the final manuscript and reported results.

REPRODUCIBILITY STATEMENT

The AFoB architecture and training objective are described in Section 3, and the datasets, experimental settings, and evaluation procedures are specified in Section 4.1. Upon acceptance, we will publicly release the source code, training and evaluation scripts, and experimental configurations with the camera-ready version to facilitate reproduction of the reported experiments.

REFERENCES

Yohann Benchetrit, Hubert Banville, and Jean-Rémi King. Brain decoding: Toward real-time reconstruction of visual perception. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/1fb0a4de9c14f5557eeea886e22569cd-Abstract-Conference.html.

Hongzhou Chen, Lianghua He, Yihang Liu, and Longzhen Yang. Visual neural decoding via improved visual-EEG semantic consistency. *arXiv preprint arXiv:2408.06788*, 2024. URL https://arxiv.org/html/2408.06788v1. arXiv preprint at time of writing; cited only for the Table **??** comparison numbers quoted from Wu et al. (2025), not discussed as related work.

Sicheng Dai, Hongwang Xiao, Shan Yu, and Qiwei Ye. Autoregressive visual decoding from EEG signals. In *International Conference on Learning Representations (ICLR)*, 2026.

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.

Wilson S. Geisler, Jeffrey S. Perry, and Jiri Najemnik. Visual search: The role of peripheral information measured using gaze-contingent displays. *Journal of Vision*, 6(9):858–873, 2006. doi: 10.1167/6.9.1. URL https://svi.cps.utexas.edu/Geisler-2006-jov-6-9-1.pdf.

Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.

Anton S. Kaplanyan, Anton Sochenov, Thomas Leimkühler, Mikhail Okunev, Todd Goodall, and Gizem Rufo. DeepFovea: Neural reconstruction for foveated rendering and video compression using learned statistics of natural videos. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 38(6):212:1–212:13, 2019.

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 (NeurIPS)*, volume 37, pp. 102822–102864, 2024.

Xuan-Hao Liu, Yan-Kai Liu, Yansen Wang, Kan Ren, Hanwen Shi, Zilong Wang, Dongsheng Li, Bao-Liang Lu, and Wei-Long Zheng. EEG2Video: Towards decoding dynamic visual perception from EEG signals. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 72245–72273, 2024.

Yiheng Liu, Chuhang Zheng, Peiliang Gong, Jingtao Liu, Daoqiang Zhang, and Qi Zhu. FS-DBN: Foreground-aware EEG-visual alignment via dynamic brain networks. *arXiv preprint arXiv:2607.18344*, 2026. URL https://arxiv.org/html/2607.18344v2.

Volodymyr Mnih, Nicolas Heess, Alex Graves, and Koray Kavukcuoglu. Recurrent models of visual attention. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

R. T. Pramod, Harish Katti, and S. P. Arun. Human peripheral blur is optimal for object recognition. *Vision Research*, 200:108083, 2022. doi: 10.1016/j.visres.2022.108083. URL https://www.sciencedirect.com/science/article/pii/S004269892200089X.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, pp. 8748–8763, 2021.

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 (ICLR)*, 2024.

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 (CVPR)*, 2025.

Dario Zanca, Andrea Zugarini, Simon Dietz, Thomas R. Altstidl, Mark A. Turban Ndjeuha, Leo Schwinn, and Bjoern Eskofier. Contrastive language-image pretrained models are zero-shot human scanpath predictors. *arXiv preprint arXiv:2305.12380*, 2023. URL https://arxiv.org/abs/2305.12380.

A Appendix

B AFoB Implementation Details

Fixed reference and clean-image inputs. The EEG input pipeline provides two image representations for each sample. A fixed reference image is produced by `FoveaBlur` with height and width 224, kernel size 51, an exponential curve, and `system_g`=3. The resulting image is normalized with CLIP's input statistics, encoded with frozen CLIP weights, and L2-normalized. These reference embeddings are computed once per unique stimulus and cached. They condition the foveation predictor; they are not the adaptive targets used by the contrastive loss. With `uncertainty_aware=False`, the loader uses a single configured reference transformation rather than selecting among low-, medium-, and high-blur reference features. This flag does not disable AFoB's learned, per-pair transformation. The clean image is loaded separately, resized to 224×224, and converted to a tensor in $[0, 1]$ for saliency computation and image composition.

Parameterization. Let d be the embedding dimension. The foveation predictor uses layers $2d \to 256 \to 64 \to 3$, with GELU after each hidden layer. Let ρ denote the sigmoid function, and let $(a_{1,i}, a_{2,i}, a_{\sigma,i})$ be its raw outputs for conditioning vector u_i. Let $r_{\min}, r_{\max}$ and $\sigma_{\min}, \sigma_{\max}$ denote the configured concentration and blur bounds. The bounded parameters are

$$r_{k,i} = \exp\left[\log r_{\min} + \rho(a_{k,i}) \log \frac{r_{\max}}{r_{\min}}\right], \qquad k \in \{1, 2\}, \tag{10}$$

$$\sigma_i = \sigma_{\min} + (\sigma_{\max} - \sigma_{\min})\rho(a_{\sigma,i}). \tag{11}$$

Output biases are calibrated to the configured nominal initial values. The output weights retain their initialization, so this does not force identical initial predictions for all samples.

Saliency preprocessing. The Sobel convolutions and smoothing filters use one-pixel zero padding. The color term is the standard deviation across RGB channels, not variance. Let $\mathcal{A}$ denote a 3×3 averaging filter and let $\widetilde{S}_i = \mathcal{A}^3(S_i)$ be the smoothed saliency map. Writing $S_i^{\min}$ and $S_i^{\max}$ for its spatial extrema, the map Q_i in Equation equation 4 is

$$Q_i(\mathbf{p}) = \frac{\widetilde{S}_i(\mathbf{p}) - S_i^{\min}}{S_i^{\max} - S_i^{\min} + 10^{-8}}. \tag{12}$$

Table 4: Per-subject mAP and mean matched-pair cosine similarity ($\times 100$) for UBP and AFoB on THINGS-EEG. Avg is the unweighted mean over ten subjects. UBP reference results and AFoB paired-target results follow the respective evaluation procedures in Section 4.1.

Method	Metric	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Avg
Intra-subject: train and test on one subject												
UBP	mAP	53.9	63.7	63.9	62.3	57.7	70.6	61.6	68.9	59.2	73.1	63.5
	Similarity	15.3	20.7	19.9	17.4	19.1	21.8	19.5	18.7	21.5	23.7	19.8
AFoB (ours)	mAP	62.9	62.8	66.2	65.8	60.9	66.8	66.3	75.0	65.9	76.5	66.9
	Similarity	18.6	27.3	22.3	26.0	27.2	29.2	25.2	24.5	29.1	30.9	26.0
Inter-subject: leave one subject out for test												
UBP	mAP	22.4	26.8	18.0	24.3	20.5	23.3	18.9	22.6	28.8	30.1	23.6
	Similarity	8.5	9.8	7.8	9.0	8.6	7.9	8.2	7.9	9.7	9.2	8.7
AFoB (ours)	mAP	20.6	30.0	25.4	32.1	31.2	23.9	20.2	28.2	38.6	38.9	28.9
	Similarity	14.7	16.4	13.9	16.1	16.0	14.8	14.8	14.1	17.0	15.8	15.4

Pixel coordinates range from 0 to $W-1$ horizontally and from 0 to $H-1$ vertically. The coordinate average is not rounded or constrained to be outside the central region. The entire location branch is evaluated without gradient tracking.

Adaptive blur. The adaptive background blur uses a Gaussian kernel whose standard deviation is the predicted σ_i. Let R be the configured kernel half-width and $t \in \{-R, \ldots, R\}$ an integer offset. The normalized one-dimensional Gaussian weights are proportional to $\exp(-t^2/(2\sigma_i^2))$. Horizontal and vertical separable convolutions apply these weights with reflection padding. The resulting kernel width is $2R+1$. This adaptive kernel setting is separate from the reference transform's kernel size 51. The mask implementation adds 10^{-6} inside each squared-distance square root and 10^{-8} to the distance-normalization denominator.

Loss and optimization. The contrastive criterion computes row-wise cross-entropy in both directions with diagonal labels and averages the two mean losses. It adds no feature normalization, ranking weights, or auxiliary objective. In distributed training, the contrastive loss uses device-local batches without cross-device feature gathering. We use AdamW with weight decay 10^{-4}. The foveation predictor receives the base learning rate multiplied by the configured predictor multiplier; the remaining trainable parameters receive the base rate.

EEG inputs. For EEG, we select 17 posterior channels and the first 250 stored temporal samples. The interval `[0,250]` is an array slice and is not specified in milliseconds by the loader. With both `train_avg` and `test_avg` enabled, the loader averages the repetition axis before constructing samples. The EEG configuration specifies a training batch size of 1024 and validation/test batch sizes of 200.

B.1 Per-Subject mAP and Similarity (THINGS-EEG)

Table 4 supplements the per-subject Top-1 / Top-5 results in Table 1 with mAP and mean matched-pair cosine similarity from our UBP and AFoB runs. Subject ordering follows the main table. Both metrics are multiplied by 100, and Avg is the unweighted arithmetic mean over the ten subjects.

Intra-subject results. AFoB records average mAP and similarity values of 66.9 and 26.0, respectively, compared with the UBP reference values of 63.5 and 19.8. AFoB's mAP is numerically higher for eight of the ten subjects, with Subjects 2 and 6 as exceptions. Its matched-pair similarity is higher for all ten subjects.

Inter-subject results. The corresponding AFoB averages are 28.9 for mAP and 15.4 for similarity, while UBP records 23.6 and 8.7. AFoB's mAP is higher for nine subjects, with Subject 1 as the exception, and similarity is higher for every subject.

Ranking and matched-pair similarity. The metrics capture different properties: mAP depends on the correct candidate's rank relative to alternatives, whereas similarity measures the cosine score

Table 5: Per-subject Top-1 and Top-5 ablation scores (%) on THINGS-EEG in the intra-subject setting. Avg denotes the arithmetic mean over the 10 subjects. The evaluation protocol is described in Section 4.1.

	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Subject 8		Subject 9		Subject 10		Avg	
Method	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
1. Adaptive blur application																						
Fixed blur	41.0	74.0	46.0	80.0	51.5	78.5	50.0	79.0	42.5	74.0	59.5	83.5	49.5	77.0	57.5	85.0	45.0	77.5	56.0	87.5	49.85	79.60
2. Foveation count																						
Single-fovea	42.5	74.0	50.5	82.5	49.0	80.5	48.0	78.5	44.5	76.5	55.5	85.5	48.5	80.5	55.5	89.5	56.5	81.5	67.5	88.5	51.80	81.75
3. Second-fovea location method																						
Random	47.0	79.5	50.0	82.0	45.5	81.5	52.0	81.5	42.5	78.0	53.0	83.0	52.0	82.5	62.5	89.0	53.5	82.0	70.0	88.0	52.80	82.70
Learned	39.5	77.0	48.5	78.5	49.5	84.0	53.5	85.0	44.5	75.0	51.0	87.5	49.5	82.0	64.0	91.0	58.0	82.5	65.5	91.0	52.35	83.35
4. Foveation Predictor input																						
Image-only	42.5	76.5	43.5	79.0	45.5	80.0	50.0	77.5	39.0	72.5	60.5	83.5	50.0	77.0	57.0	88.5	44.5	76.0	57.5	87.0	49.00	79.75

Table 6: Per-subject mAP and mean cosine similarity ($\times 100$) for the THINGS-EEG intra-subject ablations. Avg denotes the arithmetic mean over the 10 subjects. The evaluation protocol is described in Section 4.1.

	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Subject 8		Subject 9		Subject 10		Avg	
Method	mAP	Sim.	mAP	Sim.	mAP	Sim.	mAP	Sim.	mAP	Sim.	mAP	Sim.	mAP	Sim.	mAP	Sim.	mAP	Sim.	mAP	Sim.	mAP	Sim.
1. Adaptive blur application																						
Fixed blur	55.4	18.9	60.8	24.2	63.8	22.8	62.4	20.5	56.9	23.2	70.4	26.1	62.9	24.5	70.1	21.0	58.1	27.1	69.4	29.1	63.02	23.74
2. Foveation count																						
Single-fovea	57.5	16.7	64.4	28.4	62.7	23.0	61.3	25.5	58.8	27.9	68.5	30.4	62.0	24.4	70.4	23.8	67.8	30.3	77.4	31.0	65.08	26.14
3. Second-fovea location method																						
Random	60.6	17.8	63.9	27.8	62.0	25.1	65.5	25.5	58.1	27.6	66.3	29.4	66.0	25.0	74.1	24.4	65.9	29.3	78.5	31.0	66.09	26.29
Learned	55.7	18.6	61.8	27.9	64.4	26.6	66.9	26.8	58.2	27.4	66.2	29.6	63.2	22.9	75.1	24.8	68.7	29.8	77.3	29.8	65.75	26.42
4. Foveation Predictor input																						
Image-only	57.5	19.0	59.4	24.5	61.0	24.4	62.3	21.7	54.1	25.1	71.3	28.5	61.8	24.7	70.8	22.4	57.5	27.0	70.4	28.9	62.61	24.62

of the matched pair. For example, intra-subject Subject 2 has higher similarity with AFoB (27.3 versus 20.7) but lower mAP (62.8 versus 63.7). Higher matched-pair similarity therefore need not imply better ranking. These numerical patterns are interpreted under the method-specific evaluation procedures in Section 4.1, not as a matched-protocol performance comparison.

B.2 Per-Subject Ablation Results

Tables 5 and 6 provide the THINGS-EEG intra-subject results for the five ablation variants summarized in Table 3. The variants cover four design choices: adaptive blur application (Fixed blur), fovea count (Single-fovea), second-fovea placement (Random and Learned), and predictor input (Image-only). AFoB denotes the proposed saliency-based method; Learned is a placement ablation, not the proposed configuration. The variants and evaluation procedure are defined in Sections 4.4 and 4.1, respectively.

Subject columns follow the same Subject 1–10 ordering as Table 1. Top-1 and Top-5 are percentages; mAP and mean matched-pair cosine similarity are multiplied by 100. Avg is the unweighted arithmetic mean of the ten reported subject values. Subject values are shown to one decimal place, and averages to two decimal places to retain the precision of this breakdown.

Subject-level patterns. The breakdown exposes variation hidden by aggregate means. Subject 6 has the highest Top-1 score within the Fixed blur and Image-only variants, whereas Subject 10 has the highest score within Single-fovea, Random, and Learned. For the placement comparison, Learned exceeds Random in Top-1 for five subjects, and Random exceeds Learned for the other five. Random has higher mean Top-1 (52.80% versus 52.35%) and mAP (66.09 versus 65.75), while Learned has higher mean Top-5 (83.35% versus 82.70%) and similarity (26.42 versus 26.29). Neither placement alternative therefore dominates across all subjects and metrics. These are descriptive comparisons of the reported results, not statistical-significance claims.