DS-LCCA: DIFFERENTIALLY SUPERVISED LIGHTNESS–CHROMATICITY COMPLEMENTARY ALIGNMENT FOR NEURAL VISUAL DECODING

Anonymous authors
Paper under double-blind review

ABSTRACT

Decoding visual representations from brain signals helps reveal how the brain represents visual information and enables new forms of brain–computer interaction. However, existing approaches to visual supervision adaptation do not explicitly account for the distinct spatial sensitivities of human vision to Lightness and Chromaticity variations, potentially limiting neural–visual alignment. To address this limitation, we propose Differentially Supervised Lightness-Chromaticity Complementary Alignment (DS-LCCA), which constructs component-specific visual supervision in CIELUV space. The Lightness branch reweights multiscale discrete wavelet coefficients to attenuate spatial detail, whereas the Chromaticity branch applies Gaussian smoothing to suppress fine-scale Chromaticity variations. Each transformation preserves the remaining coordinates before conversion back to RGB. Four independent branches learn neural–visual alignment through symmetric contrastive objectives. Three Residual Projection Encoders align EEG or MEG recordings with CLIP ResNet-50 features of the original and transformed images for retrieval and classification. A dedicated ATM–ViT-H-14 branch learns original-image representations for reconstruction. Its neural embeddings condition a diffusion prior to produce the primary visual condition, while linear projections of the Lightness and Chromaticity embeddings provide auxiliary conditions. Separate IP-Adapters jointly condition a pretrained SDXL generator, and the auxiliary projections are optimized through the generation loss. Extensive experiments on THINGS-EEG and THINGS-MEG demonstrate that DS-LCCA outperforms state-of-the-art methods across retrieval, classification, and reconstruction. The code is available at `https://anonymous.4open.science/r/DS-LCCA-4CB5`.

1 INTRODUCTION

Neural visual decoding aims to identify or reconstruct visual content from brain activity elicited during perception (Kay et al., 2008; Miyawaki et al., 2008). By establishing a correspondence between neural responses and visual representations, it provides a means to investigate perceptual processing and advance brain–computer interfaces. Functional magnetic resonance imaging (fMRI), together with latent diffusion models, has enabled detailed image reconstruction (Takagi & Nishimoto, 2023), yet its slow hemodynamic response limits the temporal resolution required to capture rapidly evolving visual processes (Benchetrit et al., 2024b). In contrast, electroencephalography (EEG) and magnetoencephalography (MEG) offer millisecond-scale temporal resolution, while EEG further benefits from relatively accessible and portable acquisition (Song et al., 2024; Li et al., 2024). Recent advances in pretrained visual representations and generative models have therefore driven growing interest in EEG- and MEG-based visual decoding (Song et al., 2024).

An influential line of work performs visual decoding by aligning neural representations with features derived from the original stimulus images. NICE uses contrastive learning to align EEG responses with image features for zero-shot object recognition (Song et al., 2024). ATM aligns EEG representations with CLIP embeddings and incorporates a diffusion-based generation pipeline for classification, retrieval, and reconstruction (Li et al., 2024). CognitionCapturer further extends this paradigm to image, text, and depth supervision, learning modality-specific EEG representations and combining their conditions for reconstruction (Zhang et al., 2025). Despite these advances in representation

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.

learning and generative conditioning, direct alignment with original-image features may require neural recordings to recover information that is not reliably encoded, resulting in a mismatch between visual supervision and neural representations.

Recent studies have therefore begun to adapt visual supervision itself. The uncertainty-aware blur prior (UBP) adjusts image detail according to estimated cross-modal uncertainty, improving brain-to-image retrieval (Wu et al., 2025). Human-aligned image encoders introduce targets informed by perceptual judgments and improve retrieval across neural encoders and recording modalities (Rajabi et al., 2025). ShrinkAdapter adapts the visual teacher through a task-guided bottleneck to reduce the mismatch between rich visual features and EEG representations (Wu et al., 2026). Together, these studies demonstrate the effectiveness of adapting visual supervision for neural–visual alignment. Psychophysical evidence further shows that spatial sensitivity differs between luminance and chromatic patterns (Campbell & Robson, 1968; Mullen, 1985), motivating separate treatment of Lightness structure and Chromaticity variation when constructing perceptual image targets (Fig. 1). However, existing approaches do not explicitly model this distinction with separate spatial operators for Lightness and chromatic components, leaving their distinct spatial sensitivities unaddressed, thereby weakening neural–visual alignment and limiting decoding performance.

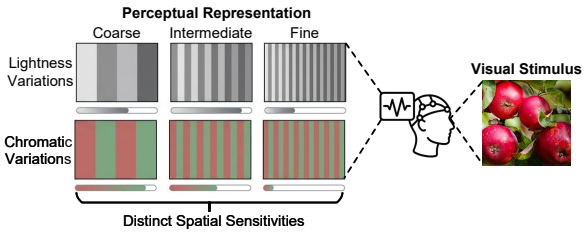

Figure 1: Distinct spatial sensitivities to Lightness and Chromaticity variations shape visual perception.

To address this limitation, we propose Differentially Supervised Lightness-Chromaticity Complementary Alignment (DS-LCCA), a component-specific visual supervision framework for EEG and MEG decoding. DS-LCCA converts each stimulus to CIELUV and constructs complementary RGB targets through multiscale lightness reweighting and Gaussian chromaticity smoothing. Three independently trained Residual Projection Encoders (RPEs) align neural recordings with CLIP ResNet-50 features of the original, Lightness-processed, and Chromaticity-processed images. A fourth branch uses ATM and ViT-H-14 to learn original-image representations for generation. A diffusion prior converts this branch's neural embeddings into the primary condition, while generation-trained linear projections provide auxiliary conditions from the lightness and chromaticity branches. The resulting conditions jointly guide SDXL through separate IP-Adapters. We evaluate the framework on THINGS-EEG and THINGS-MEG across retrieval, classification, and reconstruction. Our contributions are threefold.

- We are the first to explicitly distinguish Lightness structure from Chromaticity variation when constructing image-side supervision for neural visual decoding, translating their distinct spatial sensitivities into separate perceptual targets.
- We propose DS-LCCA, which constructs differential perceptual supervision through wavelet-based lightness reweighting and Gaussian chromaticity smoothing, and couples independent neural–visual alignment with complementary diffusion conditioning for decoding and reconstruction.
- Extensive experiments on THINGS-EEG and THINGS-MEG demonstrate that DS-LCCA achieves state-of-the-art performance across retrieval, classification, and reconstruction.

2 Related Work

2.1 Neural visual decoding and visual supervision

Early fMRI studies established the feasibility of identifying natural images from brain activity and reconstructing visual patterns through multiscale local decoding (Kay et al., 2008; Miyawaki et al., 2008). Diffusion-based methods later extended this capability to detailed natural-image reconstruction (Takagi & Nishimoto, 2023). For EEG, NICE (Song et al., 2024), ATM (Li et al., 2024), and CognitionCapturer (Zhang et al., 2025) advance contrastive representation learning, diffusion-based reconstruction, and multimodal conditioning, respectively. These methods commonly align neural representations with features extracted from the original stimulus images. For MEG,

Benchetrit et al. (2024b) combine contrastive and regression objectives to map neural recordings into pretrained image-representation spaces and condition an image generator. Their analysis further shows that decoding depends on both the visual representation and the timing of neural responses, underscoring the importance of matching image features to temporally resolved neural information. More recent studies have addressed this relationship by adapting visual supervision itself. UBP (Wu et al., 2025), human-aligned image encoders (Rajabi et al., 2025), and ShrinkAdapter (Wu et al., 2026) explore uncertainty-guided image processing, perceptually informed representations, and task-adaptive visual teachers, respectively. Collectively, these approaches revise supervision either by altering the spatial detail presented to the image encoder or by adapting the feature space matched by neural representations. Their improvements indicate that the original stimulus representation is not necessarily the most suitable decoding target. However, existing strategies do not explicitly distinguish the spatial processing of lightness and chromatic components, leaving component-specific supervision insufficiently explored.

2.2 Spatial sensitivity and perceptual color coordinates

Human vision exhibits different spatial sensitivity profiles for luminance and chromatic patterns. Using gratings with different waveforms, Campbell & Robson (1968) showed that detection thresholds can be explained by mechanisms tuned to restricted spatial-frequency ranges. Mullen (1985) further measured red–green and blue–yellow chromatic contrast sensitivity and observed low-pass profiles under controlled viewing conditions, in contrast to those for luminance patterns. Together, these findings support differentiated spatial processing across visual components. At the physiological level, Derrington et al. (1984) characterized cone-opponent responses in the macaque lateral geniculate nucleus, providing additional evidence for distinct chromatic and achromatic response properties. CIELUV offers a practical coordinate system for separating Lightness and chromatic attributes in image processing. It consists of a Lightness coordinate L^* and two chromatic coordinates u^* and v^*, with distances designed to approximate perceived differences (DmenuACl & HS, 1976), thereby enabling separate manipulation of Lightness structure and Chromaticity variation.

3 Method

We propose Differentially Supervised Lightness-Chromaticity Complementary Alignment (DS-LCCA) as shown in Fig. 2. Two component-specific transformations construct lightness and chromaticity supervision in CIELUV space. Four independent neural–visual alignment branches learn from these views and the original image. Three RPE–ResNet-50 branches support retrieval and classification, while an ATM–ViT-H-14 branch supplies the primary generation condition. The lightness and chromaticity RPE branches provide complementary generation conditions through linear projections.

3.1 Problem Definition

Let $\mathcal{D} = \{(\mathbf{X}_i, I_i, y_i)\}_{i=1}^{N}$ contain neural recordings, paired RGB images, and category labels. Here, $\mathbf{X}_i \in \mathbb{R}^{C_s \times T}$ contains T samples from C_s selected EEG channels or MEG sensors, and $I_i \in [0,1]^{H \times W \times 3}$. EEG and MEG are evaluated separately. We define

$$I^R = I^G = I, \qquad I^L = \Phi_L(I), \qquad I^C = \Phi_C(I). \tag{1}$$

Indices R, L, C denote the original-image, lightness, and chromaticity RPE branches. G denotes the original-image ATM branch dedicated to generation. Retrieval ranks candidate images, classification ranks categories, and generation reconstructs an image from $\mathbf{X}$.

3.2 Differential Lightness–Chromaticity Supervision Priors

As shown in Fig. 2(a), the two transformations independently process lightness and chromatic coordinates while preserving the complementary components.

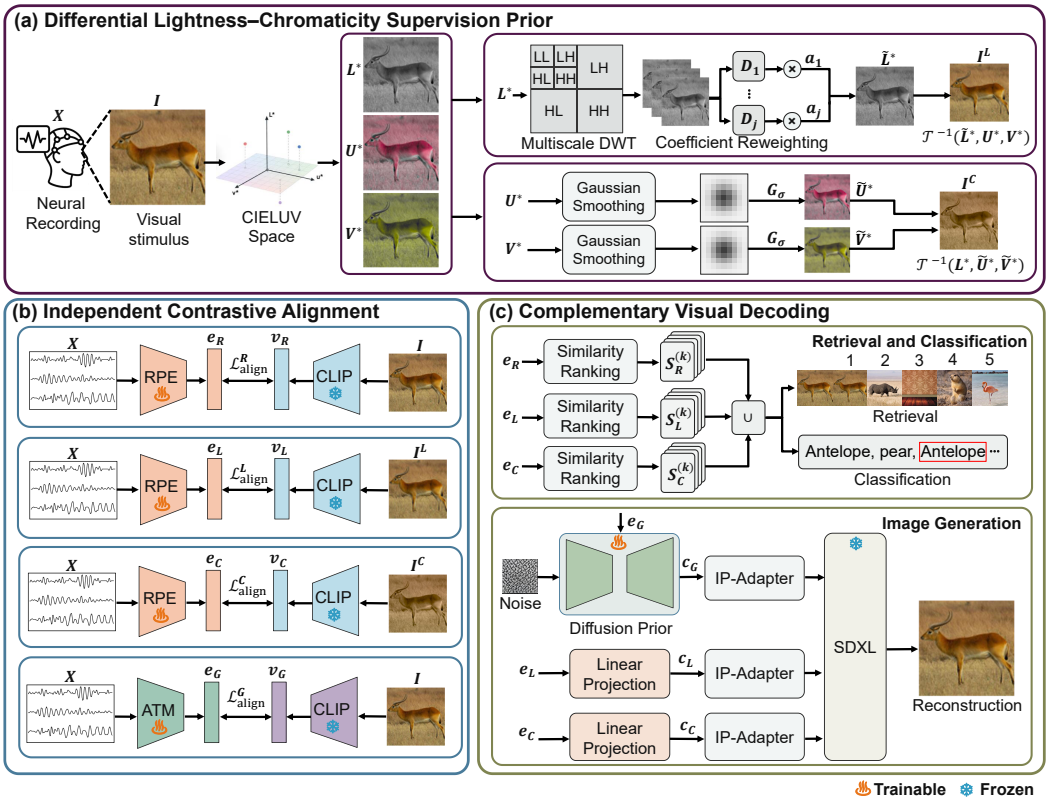

Figure 2: Overview of DS-LCCA.

3.2.1 Colorimetric Transformation

Write $(L^*, u^*, v^*) = \mathcal{T}(I)$ for RGB-to-CIELUV conversion (DmenuACl & HS, 1976). After decoding the RGB transfer function, linear-light RGB is mapped to XYZ using

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R_{\text{lin}} \\ G_{\text{lin}} \\ B_{\text{lin}} \end{bmatrix}. \tag{2}$$

With $\delta = 6/29$ and $f(t) = t^{1/3}$ for $t > \delta^3$, otherwise $f(t) = t/(3\delta^2) + 4/29$, the coordinates are

$$\begin{aligned} L^* &= 116 f(Y/Y_n) - 16, \\ u^* &= 13L^*(u' - u'_n), \quad v^* = 13L^*(v' - v'_n), \\ u' &= 4X/(X + 15Y + 3Z), \quad v' = 9Y/(X + 15Y + 3Z). \end{aligned} \tag{3}$$

We use $Y_n = 1$, $u'_n = 0.197939$, and $v'_n = 0.468311$, assigning zero CIELUV coordinates to black pixels. The inverse $\mathcal{T}^{-1}$ converts processed CIELUV coordinates through XYZ and linear-light RGB back to the input RGB encoding.

3.2.2 Multiscale Lightness Reweighting

An n-level discrete wavelet transform decomposes L^* into approximation A_n and directional detail coefficients D_j^o, where $o \in \{H, V, D\}$ and $j = 1$ denotes the finest scale. Prescribed weights $a_0, \ldots, a_n$ regulate these coefficients:

$$\widetilde{L}^* = \mathcal{W}_n^{-1}\big(a_0 A_n, \{a_j D_j^H, a_j D_j^V, a_j D_j^D\}_{j=1}^n\big), \quad I^L = \mathcal{T}^{-1}(\widetilde{L}^*, u^*, v^*). \tag{4}$$

Weights below one attenuate their coefficient groups. The original chromatic coordinates are retained before conversion back to RGB.

3.2.3 GAUSSIAN CHROMATICITY SMOOTHING

For odd kernel size $k = 2r + 1$ and support $\Omega_k = \{-r, \dots, r\}^2$, define

$$G_\sigma(p, q) = \frac{\exp[-(p^2 + q^2)/(2\sigma^2)]}{\sum_{(a,b)\in\Omega_k} \exp[-(a^2 + b^2)/(2\sigma^2)]}. \tag{5}$$

The Chromaticity branch suppresses fine-scale variations by convolution while retaining L^*:

$$\widetilde{u}^* = G_\sigma * u^*, \qquad \widetilde{v}^* = G_\sigma * v^*, \qquad I^C = \mathcal{T}^{-1}(L^*, \widetilde{u}^*, \widetilde{v}^*). \tag{6}$$

Both supervision targets are RGB images. Their component-specific interventions occur in CIELUV coordinates.

3.3 INDEPENDENT CONTRASTIVE ALIGNMENT

Fig. 2(b) shows the four alignment branches. All four branches receive the same recording $\mathbf{X}$ and have independently parameterized neural encoders. Their input–output mappings are

Branch	Neural embedding	Visual embedding
R	$\mathbf{e}_R = f_R^{\mathrm{RPE}}(\mathbf{X})$	$\mathbf{v}_R = F_{\mathrm{RN50}}(I)$
L	$\mathbf{e}_L = f_L^{\mathrm{RPE}}(\mathbf{X})$	$\mathbf{v}_L = F_{\mathrm{RN50}}(I^L)$
C	$\mathbf{e}_C = f_C^{\mathrm{RPE}}(\mathbf{X})$	$\mathbf{v}_C = F_{\mathrm{RN50}}(I^C)$
G	$\mathbf{e}_G = f_G^{\mathrm{ATM}}(\mathbf{X})$	$\mathbf{v}_G = F_{\mathrm{H14}}(I)$

(7)

For $b \in \{R, L, C\}$, $\mathbf{e}_b, \mathbf{v}_b \in \mathbb{R}^d$. For G, $\mathbf{e}_G, \mathbf{v}_G \in \mathbb{R}^{d_g}$. The RPE branches use the same frozen pretrained CLIP ResNet-50 image encoder (Radford et al., 2021). The generation branch separately trains an ATM neural encoder (Li et al., 2024) and a ViT-H-14 image encoder on original-image pairs.

Residual projection encoder. For each $b \in \{R, L, C\}$, flatten $\mathbf{X}$ into $\mathbf{h}_{b,0} \in \mathbb{R}^{C_s T}$. The RPE computes

$$\begin{aligned}
\mathbf{h}_{b,1} &= \mathrm{LN}(\mathrm{Dropout}(\mathrm{GELU}(\mathbf{W}_{b,1}\mathbf{h}_{b,0} + \mathbf{q}_{b,1}))),\\
\mathbf{h}_{b,\mathrm{res}} &= \mathbf{h}_{b,1} + \mathrm{LN}(\mathrm{Dropout}(\mathrm{GELU}(\mathbf{W}_{b,\mathrm{res}}\mathbf{h}_{b,1} + \mathbf{q}_{b,\mathrm{res}}))),\\
\mathbf{e}_b &= \mathrm{LN}(\mathbf{W}_{b,\mathrm{out}}\mathbf{h}_{b,\mathrm{res}} + \mathbf{q}_{b,\mathrm{out}}).
\end{aligned} \tag{8}$$

The weight dimensions are $1024 \times C_s T$, 1024×1024, and $d \times 1024$, respectively. $\mathbf{q}$ denotes biases. Residual refinement precedes the final projection.

Symmetric contrastive objective. For each branch $b \in \{R, L, C, G\}$ and minibatch of B pairs, define $s_{ij}^b = \overline{\mathbf{e}}_{b,i}^\top \overline{\mathbf{v}}_{b,j}/\tau_b$, where $\overline{\mathbf{x}} = \mathbf{x}/\|\mathbf{x}\|_2$. Each branch independently minimizes the symmetric InfoNCE loss

$$\mathcal{L}_{\mathrm{align}}^b = -\frac{1}{2B}\sum_{i=1}^{B}\left[\log\frac{\exp(s_{ii}^b)}{\sum_j \exp(s_{ij}^b)} + \log\frac{\exp(s_{ii}^b)}{\sum_j \exp(s_{ji}^b)}\right]. \tag{9}$$

This objective updates each RPE against frozen RN50 targets and updates the ATM–ViT-H-14 encoders in the separately trained generation branch.

3.4 COMPLEMENTARY VISUAL DECODING

3.4.1 RETRIEVAL AND CLASSIFICATION

The upper part of Fig. 2(c) uses branches $\mathcal{B}_{\mathrm{disc}} = \{R, L, C\}$ for retrieval and classification. With $\Phi_R(I) = I$, their retrieval and classification scores are

$$\begin{aligned}
s_b^{\mathrm{ret}}(\mathbf{X}, I_j) &= \overline{\mathbf{e}}_b^\top \overline{F_{\mathrm{RN50}}(\Phi_b(I_j))},\\
s_b^{\mathrm{cls}}(\mathbf{X}, y) &= \overline{\mathbf{e}}_b^\top \overline{F_T(p(y))},
\end{aligned} \tag{10}$$

where F_T is the frozen CLIP text encoder paired with RN50 and $p(y)$ is the category prompt. For branch-specific top-k sets $\mathcal{S}_{b,i}^{(k)}$, we report

$$\mathcal{U}_i^{(k)} = \bigcup_{b\in\{R,L,C\}} \mathcal{S}_{b,i}^{(k)}, \qquad \mathrm{Hit}_\cup @k = \frac{1}{N_{\mathrm{test}}}\sum_i \mathbf{1}[a_i \in \mathcal{U}_i^{(k)}], \tag{11}$$

where a_i is the target image identity or category. This three-branch union counts a hit if any branch succeeds. Its set contains at most $3k$ candidates.

3.4.2 IMAGE RECONSTRUCTION

The lower part of Fig. 2(c) uses G as the primary generation branch and L, C as auxiliary branches.

Primary diffusion prior. The prior targets $\mathbf{z}_0 = \mathbf{v}_G$ and is conditioned on $\mathbf{e}_G$. With $\boldsymbol{\epsilon}_p \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and cumulative noise schedule $\bar{\alpha}_t$,

$$\mathbf{z}_t = \sqrt{\bar{\alpha}_t}\mathbf{v}_G + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_p, \qquad \mathcal{L}_{\text{prior}} = \mathbb{E}\|\boldsymbol{\epsilon}_p - \boldsymbol{\epsilon}_\theta(\mathbf{z}_t, t, \mathbf{e}_G)\|_2^2. \tag{12}$$

Conditional reverse diffusion from Gaussian noise produces $\mathbf{c}_G = \widehat{\mathbf{z}}_0 \sim p_\theta(\mathbf{z}_0 \mid \mathbf{e}_G)$.

Auxiliary projections and joint conditioning. Two affine projections produce conditions in the ViT-H-14 embedding space:

$$\mathbf{c}_L = \mathbf{A}_L\mathbf{e}_L + \mathbf{r}_L, \qquad \mathbf{c}_C = \mathbf{A}_C\mathbf{e}_C + \mathbf{r}_C, \qquad \mathbf{A}_b \in \mathbb{R}^{d_g \times d}, \quad \mathbf{r}_b \in \mathbb{R}^{d_g}. \tag{13}$$

Separate IP-Adapters (Ye et al., 2023) inject $\mathbf{c}_G, \mathbf{c}_L, \mathbf{c}_C$ into a pretrained SDXL generator (Podell et al., 2024). Their attention contributions combine in parallel:

$$\mathbf{H}_{\text{out}} = \mathbf{H}_{\text{base}} + \sum_{b\in\{G,L,C\}} \lambda_b \,\text{softmax}(\mathbf{Q}\mathbf{K}_b^\top/\sqrt{d_h})\mathbf{V}_b, \tag{14}$$

where $\mathbf{K}_b, \mathbf{V}_b$ are derived from the corresponding adapter tokens and λ_b controls its contribution.

Generation objective. Let $\mathbf{u}_0 = \mathcal{E}_{\text{VAE}}(I)$ be the SDXL image latent and $\mathbf{u}_s = \sqrt{\bar{\gamma}_s}\mathbf{u}_0 + \sqrt{1-\bar{\gamma}_s}\boldsymbol{\epsilon}_g$, with independent Gaussian noise $\boldsymbol{\epsilon}_g$. The generation loss is

$$\mathcal{L}_{\text{gen}} = \mathbb{E}\,\|\boldsymbol{\epsilon}_g - \boldsymbol{\epsilon}_\psi(\mathbf{u}_s, s; \mathbf{c}_G, \mathbf{A}_L\mathbf{e}_L + \mathbf{r}_L, \mathbf{A}_C\mathbf{e}_C + \mathbf{r}_C)\|_2^2\,. \tag{15}$$

Gradients propagate through the adapter pathways to both auxiliary projections. Alignment, prior, and generation losses supervise their respective training stages. At inference, $\widehat{I} = \mathcal{G}_{\text{SDXL}}(\boldsymbol{\xi}; \mathbf{c}_G, \mathbf{c}_L, \mathbf{c}_C)$ starts from Gaussian latent noise $\boldsymbol{\xi}$.

4 EXPERIMENTS AND RESULTS

4.1 COMPARISON WITH SOTA METHODS

To evaluate DS-LCCA across recording modalities and decoding tasks, we compare it with state-of-the-art methods on THINGS-EEG and THINGS-MEG. Due to space limitations, Appendix A.1 provides dataset descriptions, preprocessing procedures, evaluation metrics, baselines, and implementation details. DS-LCCA achieves superior performance over all evaluated state-of-the-art methods on every task across both datasets, including retrieval, classification, and reconstruction, as summarized in Fig. 3. Detailed results are reported in Appendix A.2. Tables 5 and 6 provide subject-wise intra- and inter-subject retrieval results for EEG and MEG, respectively, while Table 7 evaluates zero-shot classification under 50-, 100-, and 200-way settings. Table 8 further compares low-level structural and high-level semantic reconstruction metrics across both modalities. Fig. 6 presents ranked EEG retrieval examples for qualitative analysis. Together, these results demonstrate that Lightness–Chromaticity supervision consistently benefits neural visual decoding across modalities and tasks.

4.2 ABLATION STUDY

Lightness–Chromaticity supervision.

To assess the contribution of each supervision component, we compare hromaticity, Lightness, and their combination on both datasets. Table 1 shows that either Chromaticity or Lightness supervision alone consistently underperforms their combination across retrieval, classification, and reconstruction. Joint Lightness–Chromaticity supervision yields stronger performance across all three tasks on both

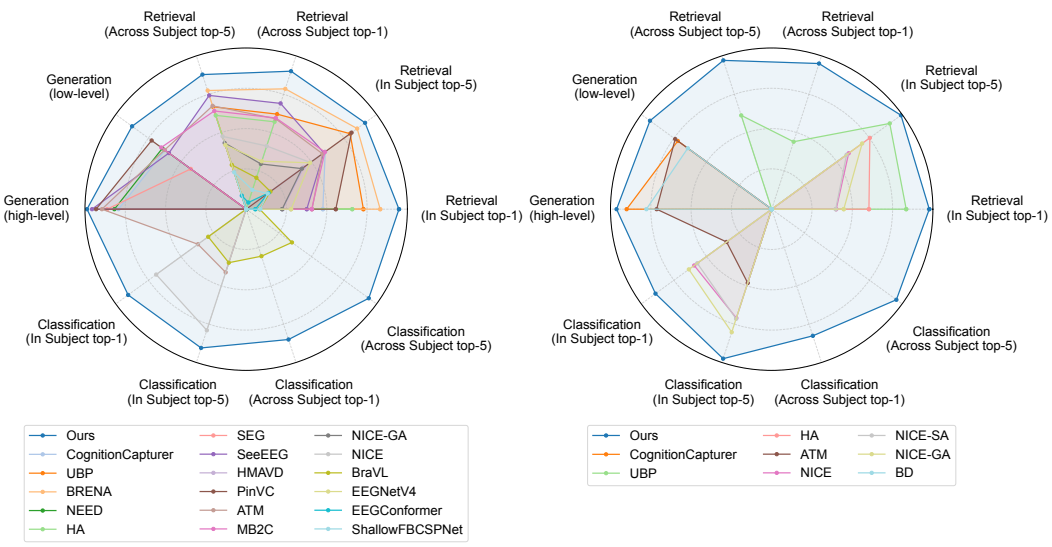

Figure 3: Retrieval, classification, and reconstruction on THINGS-EEG (left) and THINGS-MEG (right).

Table 1: Lightness–Chromaticity ablation on THINGS-EEG and THINGS-MEG.

Method	Retrieval				Classification				Reconstruction						
	Intra-subject		Inter-subject		Intra-subject		Inter-subject		Low-level		High-level				
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	PixCorr↑	SSIM↑	AlexNet(2)↑	AlexNet(5)↑	Inception↑	CLIP↑	SwAV↓
THINGS-EEG															
Chromaticity	49.60	81.30	10.10	31.90	12.80	35.90	3.55	13.10	0.1594	0.4656	0.8258	0.8836	0.7836	0.8021	0.5349
Lightness	52.50	82.90	12.30	33.60	12.95	37.45	3.60	13.50	0.1668	0.4719	0.8156	0.8007	0.7666	0.8169	0.5441
DS-LCCA	**66.35**	**91.15**	**18.00**	**43.90**	**18.10**	**45.25**	**5.10**	**18.80**	**0.1914**	**0.5174**	**0.8488**	**0.9098**	**0.8220**	**0.8460**	**0.5294**
THINGS-MEG															
Chromaticity	21.56	48.81	2.25	10.50	9.10	27.60	2.38	9.75	0.1083	0.3885	0.8120	0.8635	0.7830	0.7860	0.5520
Lightness	23.16	51.35	2.66	9.60	8.10	28.80	2.13	10.13	0.1130	0.3920	0.8025	0.8630	0.7660	0.8000	0.5615
DS-LCCA	**31.28**	**60.46**	**4.75**	**16.50**	**15.15**	**39.00**	**4.13**	**16.25**	**0.1300**	**0.4313**	**0.8178**	**0.8895**	**0.7847**	**0.8095**	**0.5480**

THINGS-EEG and THINGS-MEG. Qualitative comparisons in Appendix A.3, Fig. 7, further show that combined Lightness–Chromaticity supervision produces reconstructions that more faithfully preserve the semantic identity, characteristic shape, orientation, and overall spatial configuration of the reference objects.

Lightness wavelet subbands. To assess the contribution of Lightness approximation and detail components, we retain one wavelet subband at a time and set the remaining coefficients to zero. Table 2 shows that the LL approximation outperforms each directional detail subband across the reported tasks. Qualitative comparisons are provided in Appendix A.3 (Fig. 8). In these examples, LL-based supervision better preserves overall object shape and spatial structure. Reconstructions from individual detail subbands are less coherent. Together, these results indicate that the coarse Lightness structure retained by LL provides more effective supervision than isolated directional details under the evaluated decoding protocol. This finding is consistent with the contribution of coarse lightness structure to visual perception, suggesting that preserving such information can facilitate neural–visual alignment.

Chromaticity processing. To evaluate the spatial operator used for Chromaticity supervision, we compare several image-processing operations. Gaussian smoothing achieves the best reported results in Table 3. Corresponding reconstructions appear in Appendix A.3 (Fig. 9). In the displayed examples, Gaussian-smoothed supervision better preserves object identity and overall structure, suggesting that

Table 2: Lightness wavelet subbands on THINGS-EEG.

Subbands	Retrieval				Classification				Reconstruction						
	Intra-subject		Inter-subject		Intra-subject		Inter-subject		Low-level		High-level				
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	PixCorr↑	SSIM↑	AlexNet(2)↑	AlexNet(5)↑	Inception↑	CLIP↑	SwAV↓
LL	**66.35**	**91.15**	**18.00**	**43.90**	**18.10**	**45.25**	**5.10**	**18.80**	**0.1914**	**0.5174**	**0.8488**	**0.9098**	**0.8220**	**0.8460**	**0.5249**
LH1	54.85	86.25	15.05	41.50	14.95	42.80	4.20	17.75	0.1637	0.5077	0.8131	0.8864	0.7761	0.8176	0.5305
HL1	54.30	85.55	14.75	41.20	14.80	42.50	4.15	17.65	0.1612	0.5046	0.8218	0.8944	0.7735	0.8212	0.5326
HH1	54.45	86.65	14.75	41.70	14.85	43.00	4.20	17.90	0.1650	0.5080	0.8246	0.8857	0.7933	0.8307	0.5368
LH2	55.10	86.65	14.95	41.75	15.00	43.15	4.25	17.85	0.1825	0.5047	0.8270	0.8973	0.7795	0.8329	0.5294
HL2	55.10	86.90	14.90	41.85	14.95	43.10	4.20	17.90	0.1902	0.5061	0.8269	0.9037	0.7747	0.8183	0.5315
HH2	52.85	85.90	14.30	41.35	14.40	42.65	4.05	17.70	0.1823	0.5058	0.8371	0.9032	0.7882	0.8217	0.5265

Table 3: Chromaticity processing on THINGS-EEG.

Method	Retrieval				Classification				Reconstruction						
	Intra-subject		Inter-subject		Intra-subject		Inter-subject		Low-level		High-level				
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	PixCorr↑	SSIM↑	AlexNet(2)↑	AlexNet(5)↑	Inception↑	CLIP↑	SwAV↓
K-means	64.80	90.80	17.55	43.25	17.55	43.85	4.20	17.10	0.1773	0.5105	0.8007	0.8866	0.7790	0.8263	0.5309
Mean Shift	64.95	88.85	17.65	43.10	17.75	43.95	4.30	17.50	0.1794	0.5097	0.8315	0.8976	0.7847	0.8305	0.5303
DBSCAN	64.30	86.40	17.35	42.90	17.20	43.50	4.15	16.85	0.1865	0.5114	0.8074	0.8907	0.7847	0.8345	0.5303
Fuzzy C-means	64.10	86.20	17.10	42.55	17.20	43.25	4.10	16.55	0.1902	0.5156	0.8177	0.8883	0.7865	0.8173	0.5327
Median Cut	65.10	90.75	17.80	43.30	17.85	44.10	4.85	18.10	0.1684	0.5064	0.8152	0.8816	0.7860	0.8192	0.5344
Low Resolution	63.90	85.80	16.85	42.05	16.55	42.95	3.95	16.15	0.1700	0.5053	0.8198	0.8894	0.7871	0.8239	0.5382
Uniform Blur	64.00	86.45	17.10	42.50	17.10	43.10	4.15	16.75	0.1696	0.5111	0.8169	0.8973	0.7830	0.8242	0.5325
Foveated Blur	64.85	87.55	17.35	42.55	17.55	43.40	4.25	17.10	0.1825	0.5103	0.8179	0.8968	0.7809	0.8294	0.5296
Gaussian Smoothing	**66.35**	**91.15**	**18.00**	**43.90**	**18.10**	**45.25**	**5.10**	**18.80**	**0.1914**	**0.5174**	**0.8488**	**0.9098**	**0.8220**	**0.8460**	**0.5294**

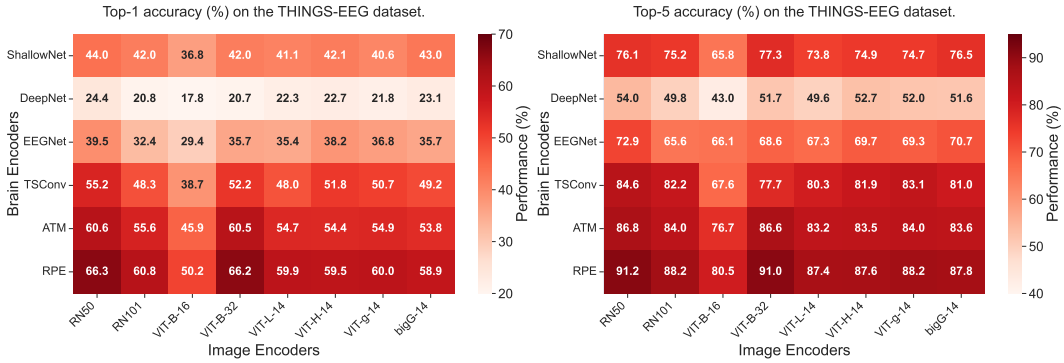

Figure 4: Retrieval with different neural and image encoders on THINGS-EEG. Left and right show Top-1 and Top-5 results.

suppressing fine-scale chromatic variations provides effective supervision for reconstruction while retaining useful visual information.

Neural and image encoder assignments. To determine whether retrieval and reconstruction benefit from the same encoder pair, we compare neural–image encoder combinations and primary–auxiliary assignments. Figure 4 shows that RPE performs best across the tested visual backbones, with RN50 outperforming ViT-H-14 for retrieval. Additional retrieval and classification results appear in Appendix A.4, Table 9. For reconstruction, Table 4 compares all four assignments of RPE–RN50 and ATM–ViT-H-14. Using ATM–ViT-H-14 for the primary branch and RPE–RN50 for the auxiliary branches achieves the best result on every reported metric. This task-dependent pattern supports a dedicated original-image representation for primary conditioning and representation-specific RPE features for auxiliary conditioning.

Table 4: Primary and auxiliary encoder assignments for reconstruction on THINGS-EEG. Each pair lists the neural encoder followed by the image encoder.

Primary	Auxiliary	Low-level		High-level				
		PixCorr↑	SSIM↑	AlexNet(2)↑	AlexNet(5)↑	Inception↑	CLIP↑	SwAV↓
RPE–RN50	RPE–RN50	0.1487	0.4027	0.7731	0.8514	0.7574	0.7846	0.5734
ATM–ViT-H-14	ATM–ViT-H-14	0.1792	0.4197	0.7946	0.8438	0.7025	0.7522	0.5921
RPE–RN50	ATM–ViT-H-14	0.1528	0.3890	0.7395	0.7677	0.6536	0.6945	0.6304
ATM–ViT-H-14	**RPE–RN50**	**0.1914**	**0.5174**	**0.8488**	**0.9098**	**0.8220**	**0.8460**	**0.5294**

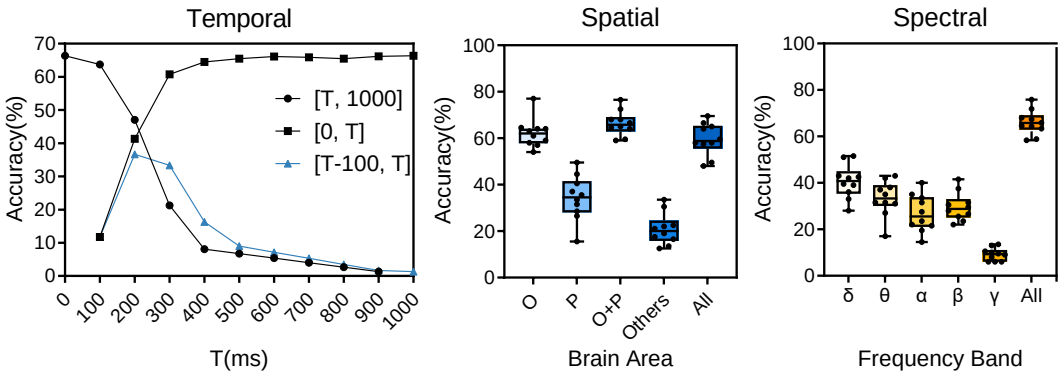

Figure 5: EEG decoding across temporal windows (left), channel subsets (middle), and frequency bands (right). Results are averaged across subjects. O and P denote occipital and parietal EEG channels, respectively.

4.3 Temporal, Spatial, and Spectral Analysis

To locate informative neural measurements, we vary the post-stimulus window, EEG channel subset, and frequency band. Figure 5 shows that the 100–200 ms window is the strongest isolated interval, while cumulative windows gain most of their performance within the first 400 ms. Occipital–Parietal channels perform best for retrieval and classification. Full-passband recordings outperform individual frequency bands, with delta performing best among the isolated bands. These findings support task-dependent channel selection and the use of temporally integrated responses. Appendix A.5 provides complete results in Tables 10–15 and reconstruction examples in Figs. 10 and 11.

4.4 Hyperparameter Analysis

To assess sensitivity to the supervision settings, we examine four hyperparameters on THINGS-EEG. For the Chromaticity branch, Table 16 compares Gaussian kernel sizes, with the best results obtained using a 35×35 kernel (Appendix A.6.1). Table 17 evaluates smoothing strength and favors $\sigma = 10$ (Appendix A.6.2). For the Lightness branch, Table 18 supports two decomposition levels (Appendix A.6.3), while Table 19 favors the db2 basis (Appendix A.6.4).

5 Conclusion

DS-LCCA constructs complementary visual supervision through Lightness wavelet reweighting and Chromaticity smoothing. Independent alignment branches support discriminative decoding, while a dedicated original-image branch and two auxiliary projections jointly condition reconstruction. Experiments on THINGS-EEG and THINGS-MEG demonstrate the utility of component-specific supervision across retrieval, classification, and reconstruction. One limitation is that the current formulation focuses on Lightness–Chromaticity decomposition and does not explicitly model broader perceptual factors. Future work will extend the supervision design to richer perceptual mechanisms, and assess robustness across acquisition settings.

6 AI USE STATEMENT

AI tools assisted with language editing and manuscript organization. The authors reviewed the resulting text and take responsibility for its accuracy and scientific content.

7 ETHICS STATEMENT

This study analyzes publicly available EEG and MEG datasets and involves no new participant recruitment or data collection. Neural recordings can contain sensitive information, and decoding outputs should be treated as model estimates. Applications should respect participant consent and neural-data privacy.

8 REPRODUCIBILITY STATEMENT

We use publicly available THINGS-EEG and THINGS-MEG datasets. Appendix A.1 describes preprocessing, evaluation, and training settings, and Appendices A.2–A.6 provide detailed results and sensitivity analyses. The anonymous code repository is linked in the abstract.

REFERENCES

Yohann Benchetrit, Hubert Banville, and Jean-Rémi King. Brain decoding: toward real-time reconstruction of visual perception. In *International Conference on Learning Representations*, volume 2024, pp. 7846–7858, 2024a.

Yohann Benchetrit, Hubert Banville, and Jean-Rémi King. Brain decoding: Toward real-time reconstruction of visual perception, 2024b. URL https://arxiv.org/abs/2310.19812. Version 3.

F. W. Campbell and J. G. Robson. Application of fourier analysis to the visibility of gratings. *The Journal of Physiology*, 197(3):551–566, 1968. doi: 10.1113/jphysiol.1968.sp008574. URL https://doi.org/10.1113/jphysiol.1968.sp008574.

A. M. Derrington, J. Krauskopf, and P. Lennie. Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of Physiology*, 357:241–265, 1984. doi: 10.1113/jphysiol.1984.sp015499. URL https://doi.org/10.1113/jphysiol.1984.sp015499.

Ernst DmenuACl and R HS. Commission internationale de l'eclairage c. 1976.

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. doi: 10.1109/TPAMI.2023.3263181.

Shuai Huang, Huan Luo, Haodong Jing, Qixian Zhang, Litao Chang, Yating Feng, Xiao Lin, Chendong Qin, Han Chen, Shuwen Jia, et al. Need: Cross-subject and cross-task generalization for video and image reconstruction from eeg signals. *Advances in Neural Information Processing Systems*, 38:173134–173173, 2026.

Haodong Jing, Yongqiang Ma, Panqi Yang, Haibo Hua, and Nanning Zheng. Pinpointing visual content: Disentangled features in multimodal model for eeg representation learning and decoding. *Knowledge-Based Systems*, pp. 114212, 2025.

Kendrick N. Kay, Thomas Naselaris, Ryan J. Prenger, and Jack L. Gallant. Identifying natural images from human brain activity. *Nature*, 452:352–355, 2008. doi: 10.1038/nature06713.

Jun-Mo Kim, Woohyeok Choi, Sang-Jun Park, Keun-Soo Heo, Young-Han Son, Ji-Hye Oh, Dong-Hee Shin, and Tae-Eui Kam. Seeeeg: Semantic-aware eeg-based multi-modal retrieval-augmented generation for high-fidelity visual brain decoding. In *2025 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 4883–4892. IEEE, 2025.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: A compact convolutional neural 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, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/ba5f1233efa77787ff9ec015877dbd1f-Abstract.html.

Lu Meng and Chuanhao Yang. Semantics-guided hierarchical feature encoding generative adversarial network for visual image reconstruction from brain activity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:1267–1283, 2024.

Xingyu Miao, Bing Zhai, Haoran Duan, Yang Long, et al. Decoding visual neural representations by multimodal with dynamic balancing. *arXiv preprint arXiv:2509.03433*, 2025.

Yoichi Miyawaki, Hajime Uchida, Okito Yamashita, Masa-aki Sato, Yusuke Morito, Hiroki C. Tanabe, Norihiro Sadato, and Yukiyasu Kamitani. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, 60(5):915–929, 2008. doi: 10.1016/j.neuron.2008.11.004.

K. T. Mullen. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology*, 359:381–400, 1985. doi: 10.1113/jphysiol.1985.sp015591. URL https://physoc.onlinelibrary.wiley.com/doi/10.1113/jphysiol.1985.sp015591.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving latent diffusion models for high-resolution image synthesis. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=di52zR8xgf.

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, pp. 8748–8763, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Nona Rajabi, Antonio H. Ribeiro, Miguel Vasco, Farzaneh Taleb, Mårten Björkman, and Danica Kragic. Human-aligned image models improve visual decoding from the brain. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 51009–51038, 2025. URL https://proceedings.mlr.press/v267/rajabi25a.html.

Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human Brain Mapping*, 38(11):5391–5420, 2017.

Yonghao Song, Qingqing Zheng, Bing-Chuan 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. doi: 10.1109/TNSRE.2022.3230250.

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. URL https://openreview.net/forum?id=dhLIno8FmH.

Yu Takagi and Shinji Nishimoto. High-resolution image reconstruction with latent diffusion models from human brain activity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14453–14463, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Takagi_High-Resolution_Image_Reconstruction_With_Latent_Diffusion_Models_From_Human_Brain_CVPR_2023_paper.html.

Yayun Wei, Lei Cao, Hao Li, and Yilin Dong. Mb2c: Multimodal bidirectional cycle consistency for learning robust visual neural representations. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 8992–9000, 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*, pp. 2246–2257, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Wu_Bridging_the_Vision-Brain_Gap_with_an_Uncertainty-Aware_Blur_Prior_CVPR_2025_paper.html.

Lukun Wu, Jie Li, Ziqi Ren, Kaifan Zhang, and Xinbo Gao. Shrinking the teacher: An adaptive teaching paradigm for asymmetric EEG-vision alignment. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(21):17859–17867, 2026. doi: 10.1609/aaai.v40i21.38844. URL https://ojs.aaai.org/index.php/AAAI/article/view/38844.

Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. IP-Adapter: Text compatible image prompt adapter for text-to-image diffusion models, 2023. URL https://arxiv.org/abs/2308.06721.

Kaifan Zhang, Lihuo He, Xin Jiang, Wen Lu, Di Wang, and Xinbo Gao. CognitionCapturer: Decoding visual stimuli from human EEG signal with multimodal information. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(13):14486–14493, 2025. doi: 10.1609/aaai.v39i13.33587. URL https://ojs.aaai.org/index.php/AAAI/article/view/33587.

Yanan Zhu, Ziwei Xiang, Jiamin Wu, Jinyang Guo, Hongyuan Zhang, Chunfeng Song, Hongjian Fang, Qihao Zheng, Yufei Guo, Xuelong Li, and Xianglong Liu. BRENA: Brain-inspired hierarchical neural alignment framework for visual decoding from EEG signals, 2025. URL https://openreview.net/forum?id=ybKEYQvM0L.

A Appendix

A.1 Experimental settings

Datasets and preprocessing. THINGS-EEG contains recordings from 10 participants viewing natural images. For each participant, training includes 1,654 concepts with 10 images per concept and four repetitions per image, yielding 66,160 trials. Testing includes 200 unseen concepts with one image per concept and 80 repetitions, yielding 16,000 trials. We bandpass-filter recordings from 0.1 to 100 Hz, downsample from 1,000 to 250 Hz, and subtract the mean of the 200-ms pre-stimulus baseline. We use 17 posterior electrodes and extract 0–1000 ms post-stimulus epochs. The 80 test repetitions of each image are averaged. THINGS-MEG contains recordings from four participants. Images are presented for 500 ms, followed by a blank interval of 1000 ± 200 ms. The training set contains 1,854 concepts with 12 images per concept, each presented once. The test set contains 200 concepts with one image per concept, repeated across 12 sessions. We bandpass-filter recordings from 0.1 to 40 Hz, downsample to 200 Hz, and apply baseline correction using the 100-ms pre-stimulus interval. We extract 0–1000 ms epochs and average the 12 test repetitions for each image.

Evaluation metrics. We report retrieval and classification results at $k = 1$ and $k = 5$. Intra-subject evaluation trains and tests on the same participant, while inter-subject evaluation holds out one participant for testing. Classification additionally considers 50, 100, and 200 candidate categories. Reconstruction uses PixCorr and SSIM for low-level similarity and AlexNet(2), AlexNet(5), Inception, CLIP, and SwAV for feature-based similarity. Higher values are better except for SwAV. Reported values are averaged across subjects.

Baselines. The compared methods include CognitionCapturer (Zhang et al., 2025), UBP (Wu et al., 2025), BRENA (Zhu et al., 2025), NEED (Huang et al., 2026), human-aligned encoders (Rajabi et al., 2025), SEG (Meng & Yang, 2024), SeeEEG (Kim et al., 2025), HMAVD (Miao et al., 2025), PinVC (Jing et al., 2025), ATM (Li et al., 2024), MB2C (Wei et al., 2024), NICE variants (Song et al., 2024), BraVL (Du et al., 2023), EEGNet (Lawhern et al., 2018), EEGConformer (Song et al., 2023), ShallowFBCSPNet (Schirrmeister et al., 2017), and MEG decoding (Benchetrit et al., 2024a).

Implementation Details. Experiments use one NVIDIA RTX 4090 GPU and Adam with a learning rate of 10^{-4}. Retrieval and classification models are trained for 50 epochs, and reconstruction models for 100 epochs. The contrastive temperature is initialized to 0.07. RPE uses 1024-dimensional hidden layers with GELU, dropout, LayerNorm, and a residual connection, as described in Section 3.3. RPE–RN50 supports discriminative decoding and auxiliary reconstruction conditions, while ATM–ViT-H-14 supplies the primary reconstruction condition. The Lightness branch uses a two-level db2 decomposition, retaining approximation coefficients and setting detail coefficients to zero. The Chromaticity branch smooths u^* and v^* separately with a 35×35 Gaussian kernel and $\sigma = 10$.

A.2 Overall Performance

Retrieval. Tables 5 and 6 report subject-wise retrieval results on THINGS-EEG and THINGS-MEG, respectively. On THINGS-EEG, the combined configuration achieves intra-subject hits of 66.4% and 91.2% at $k = 1$ and $k = 5$, and inter-subject hits of 18.0% and 43.9%, with consistent advantages in the corresponding average results despite variability in cross-participant transfer. On THINGS-MEG, the combined configuration also improves the aggregate results over UBP in both settings, reaching inter-subject hits of 4.8% and 16.5%. These results demonstrate the effectiveness of the supervision design across both recording modalities. Qualitative examples in Fig. 6 further illustrate the ranked retrieval results on EEG, where the displayed candidates often preserve object category and broad appearance even when the exact stimulus is not ranked first. The figure shows ranked candidates for visualization and does not define an ordering for the multi-branch candidate union.

Classification. To test semantic discrimination as the candidate set grows, Table 7 compares 50-way, 100-way, and 200-way zero-shot classification. DS-LCCA consistently outperforms the listed EEG and MEG baselines across all reported candidate settings, with the advantage maintained as the classification space becomes more challenging. On THINGS-EEG, DS-LCCA achieves 31.20%/67.20%, 28.70%/60.10%, and 18.10%/45.25% Top-1/Top-5 accuracy for 50-, 100-, and 200-way classification, respectively. On THINGS-MEG, the corresponding results are 25.75%/57.70%, 22.70%/50.70%, and 15.15%/39.00%. Although performance decreases as the number of candidate

Table 5: Subject-wise retrieval on THINGS-EEG. Values are percentages.

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																						
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	12.3	36.6	10.4	33.9	13.1	39.0	16.4	47.0	8.0	26.9	14.1	40.6	15.2	42.1	20.0	49.9	13.3	37.1	14.9	41.9	13.8	39.5
NICE-SA	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	41.2	70.5	51.2	80.9	51.2	82.0	51.1	76.9	42.2	72.8	57.5	83.5	49.0	79.9	58.6	85.8	45.1	76.2	61.5	88.2	50.9	79.7
MB2C	23.7	56.3	22.7	50.5	26.3	60.2	34.8	67.0	21.3	53.0	31.0	62.3	25.0	54.8	39.0	69.3	27.5	59.3	33.2	70.8	28.5	60.4
Neural-MCRL	24.2	52.5	17.3	45.9	22.0	50.8	27.7	56.1	16.6	40.8	23.6	48.9	25.3	49.3	33.1	62.8	21.5	46.7	24.1	56.8	23.5	51.1
CognitionCapturer	31.4	79.7	31.4	77.8	38.2	85.7	40.4	85.8	24.4	66.3	34.8	78.8	34.7	81.0	48.1	88.6	37.4	79.4	35.6	79.3	35.6	80.2
PinVC	32.6	77.9	32.5	78.9	35.0	85.4	39.8	85.1	26.5	73.4	34.7	78.2	36.5	80.4	48.3	83.7	37.8	84.0	36.6	80.7	38.9	80.8
DS-LCCA	**59.5**	**86.0**	**68.0**	**91.5**	**65.0**	**93.5**	**64.0**	**91.5**	**59.0**	**86.0**	**68.0**	**92.0**	**64.5**	**91.5**	**72.5**	**95.0**	**66.5**	**88.0**	**76.5**	**96.5**	**66.4**	**91.2**
Inter-subject																						
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	11.5	29.7	15.5	40.0	9.8	27.0	13.0	32.3	8.8	33.8	11.7	31.0	10.2	23.8	12.2	32.2	15.5	40.5	16.0	43.5	12.4	33.4
MB2C	10.5	28.2	11.3	32.8	8.8	27.7	13.7	33.5	10.7	27.5	12.2	33.2	11.5	31.8	12.0	32.2	12.2	31.3	16.2	42.2	11.9	32.0
Neural-MCRL	10.1	26.1	9.0	22.1	5.8	18.2	8.6	21.0	5.1	15.5	7.9	22.0	7.3	19.9	10.6	26.6	10.6	24.7	12.1	28.2	8.7	22.4
DS-LCCA	**19.5**	**43.5**	**26.0**	**57.5**	11.0	33.0	**22.0**	**43.0**	**15.5**	**40.0**	**15.5**	**40.0**	14.5	38.5	**17.5**	**45.5**	**18.5**	**45.5**	20.0	**52.0**	**18.0**	**43.9**

Table 6: Subject-wise retrieval on THINGS-MEG. Values are percentages.

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
Intra-subject										
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	28.1	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	15.0	38.0	46.0	80.5	27.3	59.0	18.5	43.5	26.7	55.2
DS-LCCA	**22.5**	**53.0**	**66.0**	**92.0**	**50.5**	**80.0**	**25.5**	**61.0**	**34.3**	**63.9**
Inter-subject										
UBP	2.0	5.7	1.5	17.2	2.7	10.5	2.5	8.0	2.2	10.4
DS-LCCA	**2.0**	**9.5**	**5.0**	**20.0**	**7.5**	**18.0**	**4.5**	**18.5**	**4.8**	**16.5**

categories increases, DS-LCCA retains clear gains over the compared methods, indicating that the proposed supervision remains effective under increasingly difficult semantic discrimination.

Reconstruction. To assess both structural and semantic fidelity, Table 8 compares seven reconstruction metrics. On THINGS-EEG, DS-LCCA achieves the best performance on six of the seven metrics, including PixCorr of 0.1914 and SSIM of 0.5174, indicating improved preservation of pixel-level structure. It also attains the highest AlexNet(2), AlexNet(5), Inception, and SwAV scores, showing consistent gains in high-level feature similarity. On THINGS-MEG, DS-LCCA surpasses ATM on all seven metrics, with particularly large improvements in AlexNet(2), AlexNet(5), Inception, and CLIP similarity. These results indicate that the proposed reconstruction design improves both low-level image structure and high-level semantic consistency, and that the benefit transfers consistently across EEG and MEG.

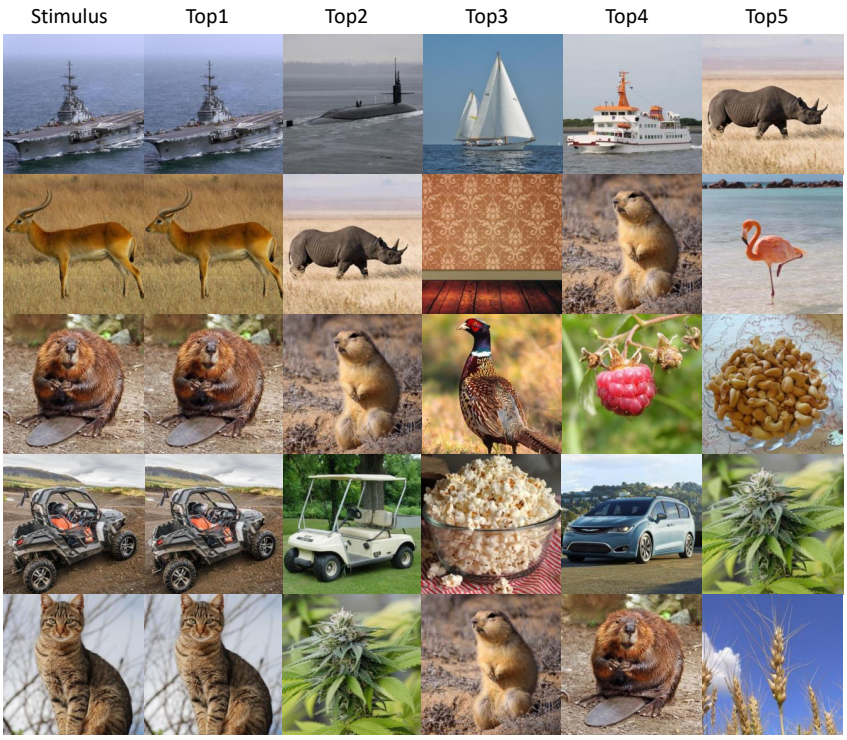

Figure 6: EEG retrieval examples. Each row shows the stimulus followed by five ranked candidates.

Table 7: Zero-shot classification with 50, 100, and 200 candidate categories. Values are percentages. The fMRI block provides results on a separate dataset.

Dataset	Model	50-way		100-way		200-way	
		Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
GOD-Wiki (fMRI)	CADA-VAE (V&T)	10.02	40.37	-	-	-	-
	MVAE (V&T)	10.04	39.60	-	-	-	-
	MMVAE (V&T)	11.68	43.29	-	-	-	-
	MoPoE-VAE (V&T)	12.90	51.78	-	-	-	-
	BraVL (V&T)	13.99	53.13	-	-	-	-
THINGS-MEG	NICE	-	-	-	-	10.10	28.40
	NICE-SA	-	-	-	-	9.70	28.60
	NICE-GA	-	-	-	-	10.80	32.10
	ATM	15.63	41.38	11.75	29.25	5.88	19.25
	DS-LCCA	**25.75**	**57.70**	**22.70**	**50.70**	**15.15**	**39.00**
THINGS-EEG	BraVL	14.33	40.28	-	-	5.82	17.45
	ATM	17.40	39.40	11.50	28.50	7.40	20.60
	DS-LCCA	**31.20**	**67.20**	**28.70**	**60.10**	**18.10**	**45.25**

Table 8: Reconstruction on THINGS-EEG and THINGS-MEG.

Methods	Datasets	Low-level		High-level				
		PixCorr↑	SSIM↑	AlexNet(2)↑	AlexNet(5)↑	Inception↑	CLIP↑	SwAV↓
ATM	THINGS-EEG	0.1600	0.3450	0.7760	0.8660	0.7340	0.7860	0.5820
PinVC	THINGS-EEG	0.1890	0.3900	0.7940	0.8810	0.8060	0.8090	0.5630
CognitionCapturer	THINGS-EEG	0.1500	0.3470	0.7540	0.6230	0.6690	0.7150	0.5900
SeeEEG	THINGS-EEG	0.1300	0.3440	0.8090	0.9040	0.7840	**0.8690**	0.5390
DS-LCCA	**THINGS-EEG**	**0.1914**	**0.5174**	**0.8488**	**0.9098**	**0.8220**	0.8460	**0.5294**
ATM	THINGS-MEG	0.1040	0.3400	0.6130	0.6720	0.6190	0.6030	0.6510
DS-LCCA	**THINGS-MEG**	**0.1300**	**0.4313**	**0.8178**	**0.8895**	**0.7847**	**0.8095**	**0.5480**



Figure 7: Reconstructions with original-image, Chromaticity, Lightness, and combined supervision.

A.3 Qualitative Ablation Results

To complement the quantitative comparisons in Table 1, Fig. 7 presents visual reconstructions obtained with original-image, Chromaticity, Lightness, and combined Lightness–Chromaticity supervision. Across the examples shown, combined supervision produces reconstructions that more closely preserve the semantic identity of the reference objects, together with their characteristic shapes, orientations, and overall spatial configurations. The individual branches exhibit complementary strengths: Chromaticity supervision captures characteristic color information, whereas Lightness supervision better preserves object contours and structural organization. Combining these cues yields more coherent reconstructions, with improved agreement between object appearance and geometry. Compared with original-image supervision, the combined formulation also better maintains recognizable object features and their arrangement, supporting visual correspondence with the reference images beyond broad category similarity. These qualitative observations complement the quantitative gains and illustrate the benefits of jointly incorporating Lightness and Chromaticity cues into neural visual reconstruction.

To examine how Lightness subbands affect reconstruction, Fig. 8 presents qualitative comparisons corresponding to Table 2. In the displayed examples, LL-based supervision better preserves overall object shape and spatial structure than individual detail subbands. These observations support the value of coarse Lightness structure for reconstruction.

To examine how Chromaticity processing affects reconstruction, Fig. 9 presents qualitative comparisons corresponding to Table 3. Gaussian-smoothed supervision produces reconstructions that better preserve object identity and overall structure in the displayed examples. This qualitative advantage complements the quantitative results, suggesting that suppressing fine-scale chromatic variations can provide effective supervision for image reconstruction while retaining useful visual information.

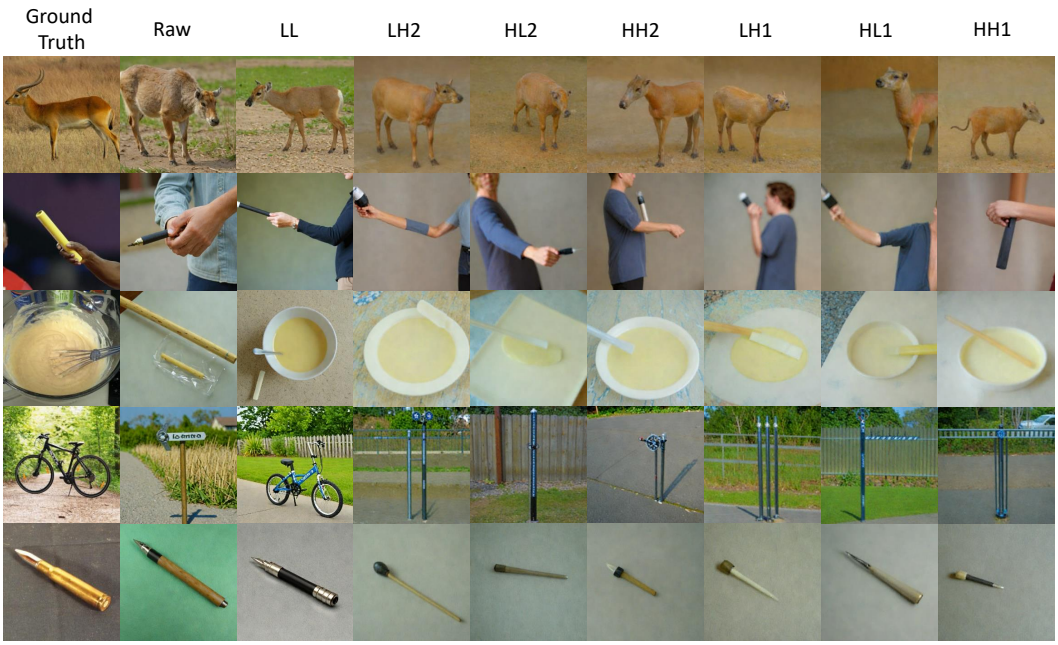

Figure 8: Reconstructions with individual Lightness wavelet subbands.

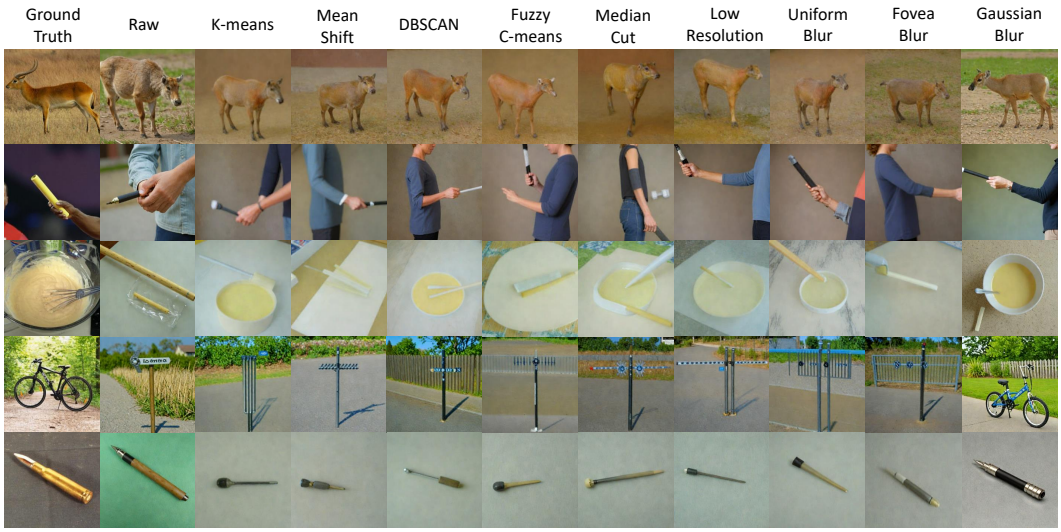

Figure 9: Reconstructions with different Chromaticity processing methods.

A.4 Additional Encoder Comparisons

To complement the retrieval heatmap in Fig. 4, Table 9 compares RPE–RN50 with ATM–ViT-H-14 across retrieval and classification settings. RPE–RN50 performs better on all listed measures. Together with the primary–auxiliary assignment experiment in Table 4, this supports different encoder roles for discriminative decoding and reconstruction.

Table 9: Neural–image encoder pairs for retrieval and classification on THINGS-EEG.

Method	Retrieval				Classification			
	Intra-subject		Inter-subject		Intra-subject		Inter-subject	
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
ATM–ViT-H-14	54.40	83.55	12.35	38.15	10.85	33.25	2.70	11.90
RPE–RN50	**66.35**	**91.15**	**18.00**	**43.90**	**12.80**	**35.90**	**3.55**	**13.10**

Table 10: Decoding with non-overlapping 100-ms EEG windows.

t(ms)	Retrieval				Classification				Reconstruction						
	Intra-subject		Inter-subject		Intra-subject		Inter-subject		Low-level		High-level				
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	PixCorr↑	SSIM↑	AlexNet(2)↑	AlexNet(5)↑	Inception↑	CLIP↑	SwAV↓
0-100	11.75	33.40	4.15	16.00	6.60	15.62	1.86	6.49	0.0630	0.1886	0.2794	0.3316	0.2706	0.3083	0.8550
100-200	**36.65**	**69.50**	**12.94**	**31.29**	**12.73**	**32.72**	**3.67**	**11.24**	**0.1365**	**0.3724**	**0.5714**	**0.6600**	**0.5439**	**0.6016**	**0.6669**
200-300	33.35	41.60	11.78	18.73	7.62	29.77	2.20	10.23	0.1242	0.2229	0.5200	0.3950	0.4950	0.3601	0.6868
300-400	16.30	41.60	5.76	19.93	8.22	21.67	2.32	9.00	0.0874	0.2349	0.3876	0.4130	0.3753	0.3840	0.7665
400-500	9.00	25.40	3.28	11.25	4.93	11.96	1.49	4.97	0.0583	0.1308	0.2140	0.2524	0.2072	0.2475	0.9049
500-600	7.16	18.27	2.61	8.09	3.54	9.52	1.07	3.95	0.0463	0.0941	0.1702	0.1816	0.1649	0.1781	0.9199
600-700	5.32	15.14	1.94	6.71	2.94	7.07	0.89	2.94	0.0344	0.0780	0.1265	0.1505	0.1225	0.1476	0.9235
700-800	3.48	12.01	1.27	5.32	2.33	4.63	0.70	1.92	0.0225	0.0619	0.0827	0.1194	0.0801	0.1170	0.9441
800-900	1.64	8.88	0.60	3.93	1.72	2.18	0.52	0.91	0.0106	0.0457	0.0390	0.0883	0.0378	0.0865	0.9525
900-1000	1.30	5.75	0.51	2.77	1.14	1.69	0.32	0.69	0.0086	0.0326	0.0327	0.0574	0.0322	0.0534	0.9530

Table 11: Decoding with cumulative EEG windows from stimulus onset.

t(ms)	Retrieval				Classification				Reconstruction						
	Intra-subject		Inter-subject		Intra-subject		Inter-subject		Low-level		High-level				
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	PixCorr↑	SSIM↑	AlexNet(2)↑	AlexNet(5)↑	Inception↑	CLIP↑	SwAV↓
0-100	11.75	33.40	4.15	16.00	6.60	15.62	1.86	6.49	0.0630	0.1886	0.2794	0.3316	0.2706	0.3083	0.8550
0-200	41.35	75.50	13.20	32.45	13.38	33.50	3.77	13.92	0.1404	0.3824	0.6224	0.6725	0.6028	0.6253	0.6235
0-300	60.75	86.15	16.48	41.49	17.11	41.43	4.82	17.21	0.1752	0.4890	0.7771	0.8599	0.7526	0.7996	0.5879
0-400	64.50	88.45	17.50	42.60	17.56	43.99	4.95	18.28	0.1861	0.5021	0.8251	0.8828	0.7991	0.8209	0.5374
0-500	65.50	90.75	17.77	43.71	18.02	44.67	5.08	18.56	0.1889	0.5151	0.8379	0.9058	0.8115	0.8422	0.5304
0-600	66.10	91.10	17.93	43.88	18.09	45.08	5.10	18.73	0.1907	0.5171	0.8456	0.9093	0.8189	0.8455	0.5300
0-700	65.90	90.10	17.88	43.39	17.89	45.89	5.04	19.07	0.1901	0.5114	0.8430	0.8993	0.8164	0.8362	0.5369
0-800	65.45	89.95	17.76	43.32	17.86	44.64	5.03	18.54	0.1888	0.5106	0.8373	0.8978	0.8109	0.8348	0.5312
0-900	66.15	91.10	17.95	43.88	18.09	45.11	5.10	18.74	0.1908	0.5171	0.8462	0.9093	0.8195	0.8455	0.5296
0-1000	**66.35**	**91.15**	**18.00**	**43.90**	**18.10**	**45.25**	**5.10**	**18.80**	**0.1914**	**0.5174**	**0.8488**	**0.9098**	**0.8220**	**0.8460**	**0.5294**

A.5 Temporal, Spatial, and Spectral Results

A.5.1 Temporal Windows

To isolate informative response intervals, Table 10 compares non-overlapping 100-ms windows. The 100–200 ms interval gives the strongest results across the reported tasks. Later isolated intervals show lower performance, indicating that early responses are particularly informative under this protocol.

To measure the benefit of accumulating responses, Table 11 extends the endpoint while fixing onset at 0 ms. Intra-subject retrieval reaches 64.50% at 400 ms and 66.35% at 1000 ms. Most of the gain occurs within the first 400 ms, with smaller subsequent changes.

To assess the contribution of early responses, Table 12 delays the starting point while retaining the 1000-ms endpoint. Performance declines as early measurements are removed. Fig 10 provides corresponding reconstruction examples, supporting the temporal trends summarized in Fig. 5.

Table 12: Decoding with delayed EEG windows ending at 1000 ms.

t(ms)	Retrieval				Classification				Reconstruction						
	Intra-subject		Inter-subject		Intra-subject		Inter-subject		Low-level		High-level				
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	PixCorr↑	SSIM↑	AlexNet(2)↑	AlexNet(5)↑	Inception↑	CLIP↑	SwAV↓
0-1000	**66.35**	**91.15**	**18.00**	**43.90**	**18.10**	**45.25**	**5.10**	**18.80**	**0.1914**	**0.5174**	**0.8488**	**0.9098**	**0.8220**	**0.8460**	**0.5294**
100-1000	63.75	90.10	17.29	43.39	17.89	43.48	5.04	18.06	0.1839	0.5114	0.8155	0.8993	0.7898	0.8362	0.5486
200-1000	47.05	78.95	12.76	38.02	15.68	32.09	4.42	13.33	0.1357	0.4481	0.6019	0.7880	0.5829	0.7327	0.6754
300-1000	21.25	50.40	7.76	20.27	10.01	15.49	2.52	9.02	0.0913	0.2261	0.3718	0.4030	0.3633	0.3678	0.7695
400-1000	8.10	24.35	3.20	11.73	4.84	10.52	1.36	4.30	0.0534	0.1382	0.2036	0.2430	0.2003	0.2260	0.8646
500-1000	6.74	20.63	2.66	9.94	4.10	8.76	1.15	3.57	0.0444	0.1171	0.1694	0.2059	0.1667	0.1915	0.9195
600-1000	5.38	16.91	2.12	8.14	3.36	6.99	0.95	2.85	0.0354	0.0960	0.1352	0.1688	0.1331	0.1569	0.9143
700-1000	4.02	13.19	1.59	6.35	2.62	5.22	0.74	2.13	0.0265	0.0749	0.1011	0.1317	0.0994	0.1224	0.9332
800-1000	2.66	9.47	1.05	4.56	1.88	3.46	0.53	1.41	0.0175	0.0538	0.0669	0.0945	0.0658	0.0879	0.9475
900-1000	1.30	5.75	0.51	2.77	1.14	1.69	0.32	0.69	0.0086	0.0326	0.0327	0.0574	0.0322	0.0534	0.9530



Figure 10: Reconstructions from different post-stimulus EEG windows.

A.5.2 EEG Channel Subsets

To examine spatial coverage, Table 13 compares occipital, parietal, combined occipital–parietal, remaining, and all electrodes. Occipital–parietal inputs perform best for retrieval and classification, exceeding either subset alone and the all-channel configuration. This supports posterior channel selection for discriminative decoding without establishing anatomical sources from scalp measurements.

To test whether reconstruction requires broader spatial coverage, Table 14 compares regional and all-channel inputs. All channels give the strongest reconstruction scores. Figure 11 illustrates the corresponding outputs. This difference from discriminative decoding supports task-dependent channel selection.

A.5.3 EEG Frequency Bands

To assess spectral contributions, Table 15 compares delta, theta, alpha, beta, gamma, and full-passband inputs. Delta gives the best isolated-band results, while the full passband performs best overall. The advantage of the full passband supports retaining information across frequencies under the evaluated filtering protocol.

Table 13: Retrieval and classification with EEG channel subsets. O and P denote occipital and parietal EEG channels, respectively.

Brain Area	Retrieval				Classification			
	Intra-subject		Inter-subject		Intra-subject		Inter-subject	
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
O	62.15	88.85	16.86	42.79	16.86	44.11	4.75	18.33
P	34.25	66.30	9.29	31.93	10.34	32.91	2.83	13.75
O+P	**66.35**	**91.15**	**18.00**	**43.90**	**18.10**	**45.25**	**5.10**	**18.80**
Others	20.85	50.95	6.65	24.54	6.78	25.29	1.82	10.51
All	59.55	87.50	16.16	42.14	16.24	43.44	4.58	18.05

Table 14: Reconstruction with EEG channel subsets.

Brain Area	Low-level		High-level				
	PixCorr↑	SSIM↑	AlexNet(2)↑	AlexNet(5)↑	Inception↑	CLIP↑	SwAV↓
Frontal	0.0844	0.3572	0.4869	0.4972	0.5098	0.5029	0.6896
Central	0.0831	0.3568	0.5079	0.5206	0.4883	0.4697	0.6831
Parietal	0.0932	0.3611	0.5193	0.5128	0.5257	0.5211	0.6903
Temporal	0.1064	0.3942	0.5209	0.5191	0.4914	0.5128	0.6777
Occipital	0.1003	0.4166	0.4933	0.5074	0.5147	0.5086	0.6771
All	**0.1914**	**0.5174**	**0.8488**	**0.9098**	**0.8220**	**0.8460**	**0.5249**

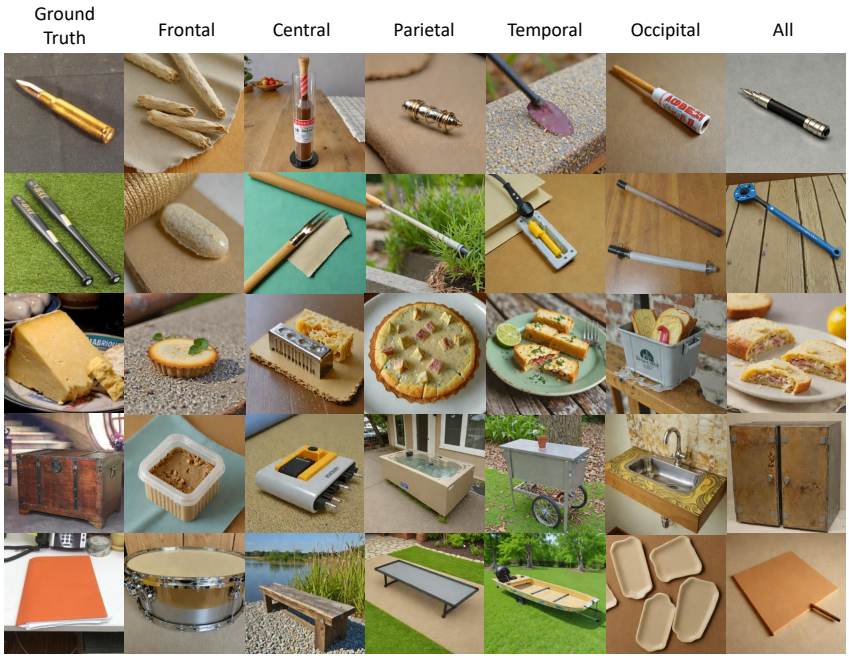

Figure 11: Reconstructions from different EEG channel subsets.

A.6 Hyperparameter Sensitivity

A.6.1 Gaussian Kernel Size

To assess the spatial support required for Chromaticity smoothing, we compare seven Gaussian kernel sizes on THINGS-EEG. Table 16 shows that performance improves as the kernel size increases from 5 to 35, with the 35×35 kernel achieving the best results across retrieval, classification, and reconstruction. Further increases generally reduce performance. This pattern suggests that a small neighborhood provides insufficient spatial averaging, whereas broader support can suppress useful

Table 15: Decoding with individual EEG frequency bands and the full passband.

Frequency Band	Retrieval				Classification				Reconstruction						
	Intra-subject		Inter-subject		Intra-subject		Inter-subject		Low-level		High-level				
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	PixCorr↑	SSIM↑	AlexNet(2)↑	AlexNet(5)↑	Inception↑	CLIP↑	SwAV↓
delta	40.55	73.00	13.00	30.83	13.06	31.65	3.72	13.49	0.1370	0.3562	0.6187	0.6560	0.6024	0.6170	0.6089
theta	33.30	65.75	11.03	24.03	10.53	21.83	2.85	11.40	0.1157	0.2787	0.4644	0.5049	0.4710	0.4630	0.6515
alpha	26.85	57.50	9.89	20.38	9.49	18.60	2.40	10.19	0.1033	0.2347	0.3844	0.4171	0.3898	0.3833	0.7073
beta	29.60	60.25	10.90	22.47	10.04	20.46	2.64	10.99	0.1101	0.2587	0.4238	0.4598	0.4288	0.4226	0.6750
gamma	9.15	27.15	3.93	13.08	5.39	13.48	1.52	5.60	0.0570	0.1541	0.2528	0.2710	0.2448	0.2520	0.8319
all	**66.35**	**91.15**	**18.00**	**43.90**	**18.10**	**45.25**	**5.10**	**18.80**	**0.1914**	**0.5174**	**0.8488**	**0.9098**	**0.8220**	**0.8460**	**0.5249**

Table 16: Effect of Gaussian kernel size on THINGS-EEG.

Kernel size k	Retrieval				Classification				Reconstruction						
	Intra-subject		Inter-subject		Intra-subject		Inter-subject		Low-level		High-level				
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	PixCorr↑	SSIM↑	AlexNet(2)↑	AlexNet(5)↑	Inception↑	CLIP↑	SwAV↓
5	61.87	88.40	15.32	38.96	15.43	39.54	4.33	16.34	0.1671	0.4517	0.7412	0.7942	0.7476	0.7587	0.5568
15	63.81	88.52	16.18	40.85	16.26	41.67	4.57	17.24	0.1764	0.4763	0.7824	0.8386	0.7575	0.7798	0.5553
25	65.48	90.31	17.26	42.94	17.11	43.86	4.82	18.17	0.1857	0.5010	0.8229	0.8819	0.7967	0.8202	0.5468
35	**66.35**	**91.15**	**18.00**	**43.90**	**18.10**	**45.25**	**5.10**	**18.80**	**0.1914**	**0.5174**	**0.8488**	**0.9098**	**0.8220**	**0.8460**	**0.5294**
45	66.08	89.58	17.08	42.11	16.84	43.14	4.74	17.87	0.1845	0.4978	0.8172	0.8757	0.7912	0.8145	0.5515
55	65.08	89.58	16.87	41.89	16.66	42.68	4.69	17.68	0.1829	0.4935	0.8102	0.8682	0.7845	0.8075	0.5570
65	65.05	88.52	16.42	40.67	16.18	41.46	4.55	17.15	0.1788	0.4825	0.7930	0.8498	0.7678	0.7905	0.5681

Table 17: Effect of Gaussian standard deviation on THINGS-EEG.

σ	Retrieval				Classification				Reconstruction						
	Intra-subject		Inter-subject		Intra-subject		Inter-subject		Low-level		High-level				
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	PixCorr↑	SSIM↑	AlexNet(2)↑	AlexNet(5)↑	Inception↑	CLIP↑	SwAV↓
6	63.88	88.80	16.12	40.77	16.23	41.59	4.58	17.28	0.1735	0.4791	0.7892	0.8543	0.7748	0.7865	0.5638
7	64.75	89.85	16.58	41.89	16.68	42.74	4.70	17.74	0.1782	0.4817	0.7904	0.8671	0.7754	0.7878	0.5633
8	65.18	90.21	16.90	42.53	16.97	43.48	4.78	18.04	0.1811	0.4896	0.8232	0.8707	0.7877	0.8103	0.5544
9	65.89	90.68	17.34	43.21	17.45	44.71	4.92	18.55	0.1856	0.5017	0.8332	0.8822	0.7972	0.8204	0.5413
10	**66.35**	**91.15**	**18.00**	**43.90**	**18.10**	**45.25**	**5.10**	**18.80**	**0.1914**	**0.5174**	**0.8488**	**0.9098**	**0.8220**	**0.8460**	**0.5294**
11	65.67	90.43	17.09	42.82	17.17	43.99	4.84	18.27	0.1835	0.4961	0.8338	0.8721	0.7981	0.8211	0.5302
12	65.23	90.07	16.81	42.29	16.89	43.28	4.76	17.97	0.1806	0.4882	0.8207	0.8681	0.7853	0.7980	0.5471
13	64.98	89.92	16.62	41.93	16.71	42.81	4.71	17.78	0.1787	0.4831	0.8127	0.8595	0.7776	0.7902	0.5519
14	64.52	89.60	16.34	41.24	16.40	42.03	4.62	17.45	0.1761	0.4762	0.8013	0.8472	0.7665	0.7887	0.5635
15	64.34	89.36	16.21	40.90	16.28	41.72	4.59	17.32	0.1747	0.4724	0.7949	0.8404	0.7604	0.7823	0.5678

chromatic information. These results favor an intermediate spatial support for constructing effective Chromaticity supervision.

A.6.2 Gaussian Standard Deviation

To evaluate the strength of Chromaticity smoothing, we compare ten standard deviations from $\sigma = 6$ to $\sigma = 15$. As shown in Table 17, performance improves up to $\sigma = 10$, which achieves the best results across all reported metrics. Larger values generally reduce performance, although individual reconstruction metrics exhibit small fluctuations. Smaller standard deviations concentrate the averaging weights near each pixel, while larger values distribute them more broadly within the kernel. The observed optimum suggests that moderate smoothing suppresses fine-scale chromatic variation while retaining information useful for neural–visual alignment and reconstruction.

A.6.3 Wavelet Decomposition Depth

To assess the spatial scale of Lightness supervision, we compare one to five wavelet decomposition levels. Table 18 shows that two levels outperform the other settings across all reported metrics. Three

Table 18: Effect of wavelet decomposition depth on THINGS-EEG.

Levels	Retrieval				Classification				Reconstruction						
	Intra-subject		Inter-subject		Intra-subject		Inter-subject		Low-level		High-level				
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	PixCorr↑	SSIM↑	AlexNet(2)↑	AlexNet(5)↑	Inception↑	CLIP↑	SwAV↓
1	59.86	87.43	15.23	38.97	15.32	39.28	4.32	16.31	0.1642	0.4441	0.7283	0.7805	0.7051	0.7257	0.6172
2	**66.35**	**91.15**	**18.00**	**43.90**	**18.10**	**45.25**	**5.10**	**18.80**	**0.1914**	**0.5174**	**0.8488**	**0.9098**	**0.8220**	**0.8460**	**0.5294**
3	64.90	89.04	17.12	42.11	17.19	44.05	4.84	18.28	0.1826	0.4937	0.8098	0.8679	0.7842	0.8072	0.5521
4	49.20	75.15	11.25	29.87	11.30	29.01	3.19	12.05	0.1423	0.3306	0.6424	0.6812	0.6251	0.6405	0.6923
5	40.13	69.49	9.43	26.55	9.48	25.61	2.68	10.64	0.1226	0.2974	0.5552	0.5879	0.5408	0.5538	0.7529

Table 19: Effect of wavelet basis on THINGS-EEG.

Wavelet basis	Retrieval				Classification				Reconstruction						
	Intra-subject		Inter-subject		Intra-subject		Inter-subject		Low-level		High-level				
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	PixCorr↑	SSIM↑	AlexNet(2)↑	AlexNet(5)↑	Inception↑	CLIP↑	SwAV↓
haar	62.87	86.53	15.56	38.87	15.64	40.09	4.41	16.63	0.1676	0.4532	0.7436	0.7969	0.7200	0.7410	0.6047
bior2.2	63.32	90.22	16.25	42.35	16.32	41.82	4.60	17.34	0.1752	0.4736	0.7771	0.8329	0.7524	0.7744	0.5789
bior3.3	64.77	90.44	16.85	42.75	16.92	43.36	4.77	17.98	0.1802	0.4871	0.7992	0.8565	0.7738	0.7965	0.5599
bior4.4	64.01	90.77	16.55	42.98	16.61	42.56	4.68	17.66	0.1780	0.4812	0.7895	0.8461	0.7645	0.7870	0.5694
bior5.5	66.22	90.66	17.65	43.13	17.73	**45.44**	5.00	18.58	0.1886	0.5100	0.8368	0.8968	0.8103	0.8340	0.5378
dmey	62.87	89.72	16.14	41.79	16.20	41.52	4.57	17.23	0.1734	0.4689	0.7694	0.8245	0.7449	0.7667	0.5869
db2	**66.35**	**91.15**	**18.00**	**43.90**	**18.10**	45.25	**5.10**	**18.80**	**0.1914**	**0.5174**	**0.8488**	**0.9098**	**0.8220**	**0.8460**	**0.5294**
db4	64.01	89.83	16.42	42.21	16.49	42.27	4.65	17.54	0.1768	0.4780	0.7843	0.8405	0.7594	0.7817	0.5740
db8	65.15	90.00	16.98	42.58	17.06	43.72	4.81	18.14	0.1825	0.4934	0.8095	0.8676	0.7840	0.8070	0.5559
db16	64.64	90.33	16.74	42.64	16.79	43.03	4.73	17.85	0.1799	0.4864	0.7982	0.8554	0.7730	0.7957	0.5633
db32	65.02	90.27	16.92	42.70	17.00	43.56	4.79	18.07	0.1817	0.4913	0.8061	0.8639	0.7806	0.8035	0.5587
sym4	64.20	90.15	16.55	42.51	16.62	42.61	4.68	17.67	0.1782	0.4818	0.7905	0.8472	0.7655	0.7880	0.5679
sym8	65.65	90.22	17.16	42.70	17.23	44.16	4.86	18.33	0.1842	0.4981	0.8174	0.8760	0.7915	0.8148	0.5509
sym16	65.09	90.60	17.06	42.98	17.11	43.86	4.82	18.20	0.1830	0.4948	0.8120	0.8703	0.7864	0.8094	0.5542
coif2	64.39	90.16	16.62	42.50	16.68	42.76	4.70	17.74	0.1788	0.4835	0.7934	0.8503	0.7684	0.7911	0.5665
coif4	66.22	89.83	17.59	42.48	17.69	45.34	4.99	18.42	0.1881	0.5085	0.8345	0.8943	0.8082	0.8318	0.5399
coif8	65.84	90.00	17.21	42.65	17.28	44.28	4.87	18.38	0.1847	0.4994	0.8195	0.8782	0.7936	0.8170	0.5490
coif16	65.28	89.72	16.93	42.10	17.00	43.56	4.79	18.07	0.1822	0.4927	0.8084	0.8664	0.7828	0.8057	0.5571

levels remain competitive, but performance declines substantially at levels four and five. Under the evaluated coefficient-processing scheme, this pattern suggests that deeper decomposition removes useful structure along with fine detail. Conversely, a single level provides less effective supervision than the two-level configuration. These results support two levels as a suitable balance between reducing fine-scale detail and preserving informative Lightness structure.

A.6.4 Wavelet Basis

To examine whether Lightness supervision depends on the wavelet basis, we compare Haar, biorthogonal, discrete Meyer, Daubechies, Symlet, and Coiflet configurations. Table 19 shows that db2 achieves the best results on all reported metrics except intra-subject classification Top-5, where bior5.5 performs best. The bior5.5 and coif4 configurations also provide competitive results across the remaining metrics, indicating that effective supervision is not restricted to a single wavelet family. Within the Daubechies, Symlet, and Coiflet families, increasing the order does not consistently improve performance. Overall, these comparisons support db2 as the default basis for the evaluated Lightness transformation.