Structural-Semantic Aware Information Reduction for Asymmetric EEG-Visual Alignment

Hongan Chen[1], Yan Kong[1], Caifeng Shan[1*], and Yuqi Fang[1,2*]

[1] School of Intelligence Science and Technology, Nanjing University, Suzhou, Jiangsu, 215163, China

[2] National Institute of Healthcare Data Science at Nanjing University, Nanjing University, Nanjing, Jiangsu, 210093, China

[*] Co-corresponding authors

yqfang@nju.edu.cn

Abstract. Understanding visual processing is a foundational pursuit in computational neuroscience. However, effective EEG-visual alignment is hindered by a severe information asymmetry. This disparity manifests as a massive capacity gap between dense image pixels and sparse EEG signals, compounded by a semantic mismatch where image backgrounds overshadow the primary subject, and physiological noise contaminates the neural data. Consequently, existing data-driven alignment methods—lacking explicit structural constraints—often overfit to superficial visual details, thereby limiting zero-shot generalization. To overcome these challenges, we propose the Structural-Semantic Aware Information Reduction (SAIR) framework. SAIR explicitly bridges this modality gap through three core components: (1) utilizing neuroscience-inspired strategies (edge extraction, depth estimation, and foveal blurring) to reduce visual redundancy and emphasize key structures; (2) employing parallel encoding architectures to systematically process multi-granular EEG features; and (3) applying a hierarchical contrastive objective for robust feature alignment. Extensive experiments on the Things-EEG dataset demonstrate that SAIR achieves state-of-the-art performance, with Top-1 accuracies of 66.9% and 17.5% in intra-subject and inter-subject settings, respectively, surpassing previous benchmarks by 8.8% and 3.8% absolute percentage points. Code is available at https://github.com/ThomasHC5/SAIR.

Keywords: Electroencephalography · Asymmetric Alignment · Visual Decoding · Information Reduction · Zero-shot Retrieval.

1 Introduction

Decoding visual perception from brain activity is a longstanding goal in neuroscience. Establishing a principled mapping between neural responses and visual stimuli is crucial for advancing our understanding of human cognition and developing practical brain-computer interfaces (BCIs) [9,15], allowing direct interaction with external devices via mental representation. Among available recording

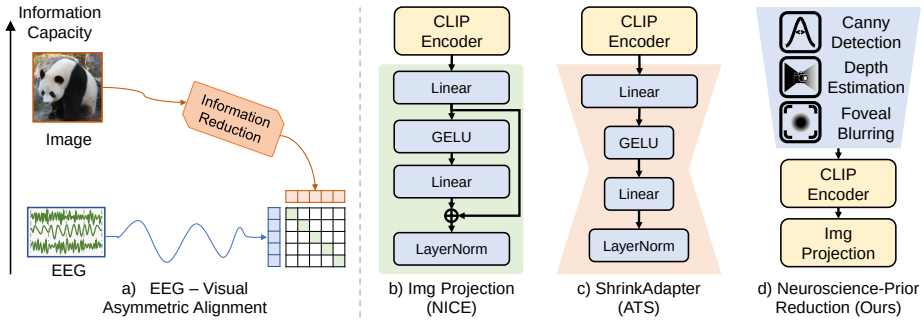

Fig. 1. Motivation. a) Common EEG-visual asymmetric alignment framework of EEG-visual decoding. b) Img Projection without any reduction. c) Bottleneck style deep-learning-based adapter. d) Neuroscience-Prior Reduction proposed in this paper.

modalities, electroencephalography (EEG) is particularly attractive due to its non-invasive nature, high temporal resolution, and ease of deployment.

Despite its advantages, effective EEG-visual alignment faces a severe information asymmetry [7,19]. This disparity manifests in two primary ways. First, there is a massive hardware capacity disparity: advanced EEG devices record sparse data matrices (e.g., 250×63 for a 1-second trial at 250Hz across 63 channels), which are physically overwhelmed by dense, high-resolution images (e.g., $500 \times 500 \times 3$). Second, semantic alignment is inherently complex; visual targets typically occupy only a fraction of an image cluttered with distracting backgrounds, while EEG signals concurrently encode mixed, non-visual noise such as emotional and physiological responses. Therefore, it is essential to explicitly reduce visual information capacity and highlight key subjects before feature alignment.

Aiming to mitigate this modality gap, current state-of-the-art approaches employ deep neural networks to implicitly compress visual features into a latent space aligned with EEG (Fig. 1(a)). For example, NICE [17] employs an ImgProjection technique to align feature spaces (Fig. 1(b)), while ATS [19] introduces a bottleneck architecture called ShrinkAdapter to force the reduction of visual information (Fig. 1(c)). However, lacking explicit structural guidance, these unguided data-driven methods tend to overfit to superficial image details (e.g., background textures) rather than invariant topologies. This shortcut learning severely compromises their zero-shot generalization on unseen subjects or in cross-domain scenarios.

To address these limitations, we propose the Structural-Semantic Aware Information Reduction (SAIR) framework (Fig. 1(d)). SAIR explicitly bridges the modality gap by reducing visual dimensionality and focusing on the core semantic subject. Specifically, we employ three neuroscience-prior strategies to filter redundancies: Canny edge detection for shape topology [8], depth estimation for geometric layout [6,11], and foveal blurring for the visual system's non-uniform spatial attention [3]. To align these filtered visual representations with neural

signals, SAIR leverages parallel encoding architectures to systematically process multi-granular EEG features. The resulting representations are then fused via a structure-guided semantic modulation mechanism and optimized using a hierarchical contrastive objective. By aligning the information granularity of visual inputs with the neural substrate, our method achieves state-of-the-art accuracy for 200-way zero-shot retrieval in both inter-subject and intra-subject settings.

Our main contributions are summarized as follows:

- Motivated by the misalignment between the uncontrollable latent spaces of existing end-to-end EEG decoding methods and the sparse semantic encoding of the human brain, we introduce explicit constraints on information compression.
- We propose a novel Structural-Semantic Aware Information Reduction framework (SAIR), which incorporates neuroscience-prior filters to disentangle information, effectively reduce visual dimensionality and highlight the core subject.
- We design a parallel encoding and structure-guided semantic modulation mechanism to fuse multi-granular features, coupled with a hierarchical contrastive objective for comprehensive EEG-visual alignment.
- Extensive experiments on the Things-EEG dataset demonstrate that our method significantly outperforms state-of-the-art approaches.

2 Methodology

2.1 Problem Formulation

The zero-shot retrieval task serves as the standard paradigm for EEG-visual decoding. Given a training split $D_{train} = \{(x_i, y_i, h_i)\}_{i=1}^{N}$, where h_i is an image and $x_i \in \mathbb{R}^{C\times T}$ is the EEG response, C and T denote the numbers of channels and time steps, respectively, and $y_i \in Y_{train}$ is the class label, the model is trained only on D_{train}. It is then evaluated on a zero-shot test split $D_{test} = \{(x_i^t, y_i^t, h_i^t)\}$, where $h_i^t \in H_{test}$, $y_i^t \in Y_{test}$, and $Y_{train} \cap Y_{test} = \emptyset$. Given an EEG trial $\hat{x}$, the goal is to identify the corresponding image $\hat{i}$ from the test image set H_{test}. We present our proposed SAIR framework through three key components: image information reduction, modality encoding and feature alignment, as demonstrated in Fig. 2.

2.2 Image Information Reduction

As highlighted in Sec. 1, the inherent modality gap between dense visual stimuli and sparse EEG signals necessitates an explicit visual information reduction. To align visual representations with EEG's coarse-grained semantic, we employ three complementary strategies to extract structural invariants while filtering redundancies: **edge extraction**, **depth estimation**, and **foveal blurring**.

Edge Extraction. Mimicking the orientation-selective neurons in the primary visual cortex (V1) [8], we utilize the Canny operator [2] to extract the structural topology of the image. This transformation preserves the object's

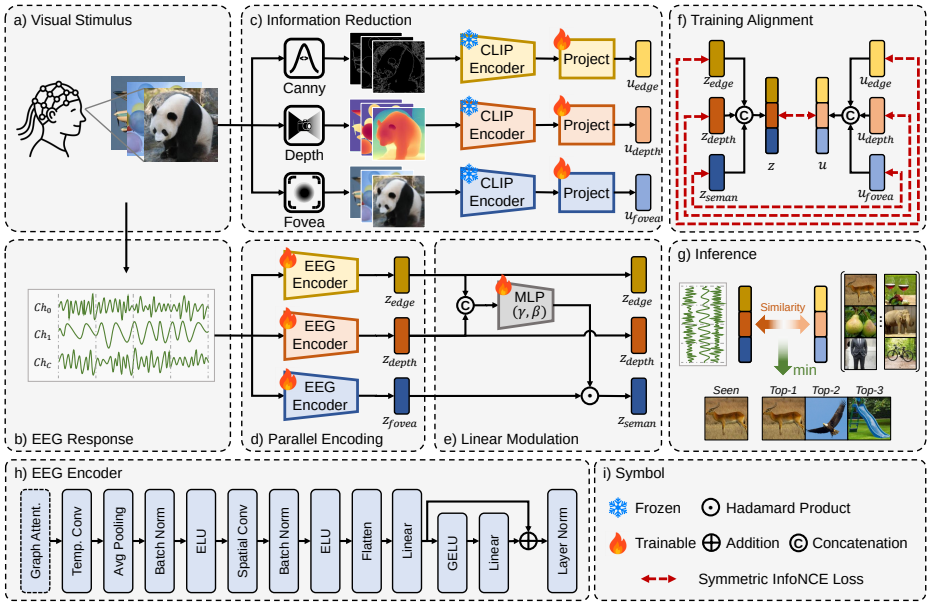

Fig. 2. Overview of the proposed Structural-Semantic Aware Information Reduction (SAIR) framework. a) Visual stimulus is c) reduced by filters and encoded to image embedding; paired b) EEG response is d) parallel encoded to three features and fused by e) linear modulation. Image and EEG embeddings are aligned by f) InfoNCE loss.

shape—a primary feature encoded in the ventral visual stream—while discarding dataset-specific pixel statistics, stripping away texture and color biases that often lead to spurious correlations. Formally, given an image h, we obtain the edge map $h_{edge} = \mathcal{T}_{Canny}(h)$, where $\mathcal{T}_{Canny}$ denotes the Canny edge detection function.

Depth Estimation. While edge maps capture 2D contours, they lack 3D spatial context. We employ monocular depth estimation to provide a geometric abstraction of the scene. Depth maps emphasize the spatial layout and distinct object boundaries, offering a geometry-centric representation that aligns with the dorsal stream's processing mechanisms [11,6]. We utilize DepthAnything2 [20], a pre-trained depth estimator, to generate the depth map $h_{depth} = \mathcal{T}_{depth}(h)$.

Foveal Blurring. Inspired by the spatially non-uniform distribution of photoreceptors in the human retina, which allocates high resolution primarily to the foveal region [3], we apply a foveation transform. This operation simulates the non-uniform sampling of the retina by maintaining resolution at the fixation point (center) while progressively blurring the periphery. This effectively suppresses background clutter (visual noise) that is irrelevant to the core semantic concept. The foveated image is defined as $h_{fovea} = \mathcal{T}_{fovea}(h, \sigma)$, where σ controls the blur radius.

2.3 Encoding and Fusion

To align neural signals with the multi-granular visual information defined, we adopt a parallel encoding architecture, followed by a structure-guided fusion.

Parallel EEG and Image Encoding. Given the processed EEG signal $x \in \mathbb{R}^{C \times T}$, we employ a Temporal-Spatio Convolutional Network (TSConv) [17] with Graph Attention (GAT) [1] as the backbone encoder. To capture distinct neurological signatures corresponding to shape, geometry, and semantic attention, we instantiate three parallel TSConv encoders $\{E_{edge}, E_{depth}, E_{fovea}\}$ with identical architectures. Each encoder maps the raw EEG into a latent feature vector $z_k = E_k(x)$, where $k \in \{edge, depth, fovea\}$.

On the visual side, to obtain supervision signals, we utilize a pre-trained CLIP encoder (ResNet50). The three types of reduced images ($h_{edge}, h_{depth}, h_{fovea}$) are passed through the frozen CLIP encoder to obtain their embeddings u_k.

Topology-Agnostic Adaptation for Cross-Subject Decoding. While the GAT module effectively captures subject-specific electrode topologies for intra-subject decoding, anatomical variation and electrode-placement shifts can cause the learned adjacency matrix to misalign across subjects, leading to negative transfer. To ensure robust cross-subject generalization, we adopt a Topology-Agnostic Adaptation strategy: bypassing the GAT module during inter-subject evaluation. By relying solely on TSConv backbones, our framework naturally decouples subject-specific priors and avoids overfitting to individual anatomical biases.

Structure-Guided Fusion via FiLM. A key insight of our approach is that *structure should modulate semantics*. The edge and depth features represent the *skeleton* of the visual stimulus, while the foveal features represent the *content*. To implement this, we propose a Structure-Guided Fusion module based on Feature-wise Linear Modulation (FiLM) [13].

First, we construct a structural context vector z_{struct} by concatenating the edge and depth embeddings as $z_{struct} = \text{Concat}(z_{edge}, z_{depth})$. We then use z_{struct} to generate channel-wise modulation parameters γ (scale) and β (shift) via a lightweight multi-layer perceptron (MLP). These parameters modulate the semantic (foveal) feature z_{fovea} with a residual connection:

$$z_{seman} = \gamma(z_{struct}) \odot z_{fovea} + \beta(z_{struct}) + z_{fovea}, \tag{1}$$

where $\odot$ denotes the Hadamard product. By conditioning the foveal features on structural context, this mechanism effectively suppresses geometrically inconsistent semantics, producing a refined representation z_{seman} for final alignment.

2.4 Feature Alignment and Zero-Shot Retrieval

Hierarchical Contrastive Objective. We introduce a hierarchical alignment strategy that enforces EEG-visual consistency at both the holistic level and the local component level. First, to construct a comprehensive representation that encompasses both the structural topology (edge and depth) and the

modulated semantic content (fovea), we concatenate the explicit embeddings to form the final EEG representation $z = \text{Concat}(z_{edge}, z_{depth}, z_{seman})$. Similarly, the corresponding visual representation is $u = \text{Concat}(u_{edge}, u_{depth}, u_{fovea})$.

To align the EEG manifold with the visual semantic space, we employ the symmetric InfoNCE loss [12,14]. For a batch of B paired embeddings $\{(x_k, y_k)\}_{k=1}^{B}$, the symmetric InfoNCE loss function $\mathcal{L}(x, y)$ is defined as:

$$\mathcal{L}(x,y) = -\frac{1}{2B}\sum_{k=1}^{B}\left[\log\frac{\exp(x_k^\top y_k/\tau)}{\sum_{j=1}^{B}\exp(x_k^\top y_j/\tau)} + \log\frac{\exp(y_k^\top x_k/\tau)}{\sum_{j=1}^{B}\exp(y_k^\top x_j/\tau)}\right], \quad (2)$$

where τ is a learnable temperature parameter. To establish this hierarchical structure, we apply the symmetric InfoNCE loss at both the global concatenated representations and the local individual components to encourage more consistent feature alignment, as defined:

$$\mathcal{L}_{SAIR} = \mathcal{L}(z,u) + \lambda(\mathcal{L}(z_{edge}, u_{edge}) + \mathcal{L}(z_{depth}, u_{depth}) + \mathcal{L}(z_{seman}, u_{fovea})), \quad (3)$$

where λ is a weighting coefficient to balance the terms.

Zero-Shot Retrieval. During the inference phase, we perform zero-shot retrieval on a disjoint test set. For a query EEG x, we compute its holistic embedding z. We then calculate the cosine similarity between z and the embeddings of all candidate images in the test gallery H_{test}. The image $\hat{h}$ with the maximum similarity is retrieved as the predicted stimulus: $\hat{h} = \arg\max_{h \in H_{test}} \text{Sim}(z, u_h)$.

3 Experiment and Discussion

Dataset. We evaluate our framework on Things-EEG [5], a large-scale public dataset for EEG-visual decoding. It contains recordings from 10 participants collected under a Rapid Serial Visual Presentation (RSVP) paradigm. During recording, a central bull's eye fixation target was displayed to guide subjects' attention. The training split includes 16,540 images (1,654 concepts, 4 repetitions/image), while the disjoint test split holds 200 images (200 novel concepts, 80 repetitions/image).

Following the standard pipeline in [17], 63-channel EEG signals are band-pass filtered (0.1-100 Hz), downsampled to 250 Hz, and trial-averaged to enhance the signal-to-noise ratio (SNR).

Implementation Details. Implemented in PyTorch (NVIDIA RTX 4090 GPU), the model is optimized via Adam with a batch size of 1000 and a learning rate of 1×10^{-4} for 200 epochs. The hyperparameter λ is set to 1/3. Following [18], intra-subject validation utilizes 740 random training trials. For the inter-subject leave-one-subject-out (LOSO) setting, the validation set aggregates test data from the 9 source subjects. We repeat each experiment five times and report the average results.

Overall Performance. We compare our method with several latest and effective baselines, including NICE [17], ATM-S [10], CogCap [21], UBP [18],

Table 1. Top-1 (upper) and Top-5 (lower) accuracy (%) for 200-way zero-shot retrieval. **Bold** and underline denote the best and second-best.

Method	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Avg
Intra-subject (Subject dependent) - train and test on one subject											
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
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
CogCap	31.4 79.6	31.4 77.8	38.1 85.6	40.3 85.8	24.4 66.3	34.8 78.7	34.6 80.9	48.1 88.6	37.4 79.3	35.5 79.2	35.6 80.2
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
ATS	48.7 78.4	59.5 87.9	60.8 89.2	59.4 85.3	**54.5** 85.3	65.1 90.4	49.8 82.5	65.8 91.3	53.5 87.1	64.3 91.4	58.1 86.7
SAIR	**68.9** **93.0**	**66.0** **93.7**	**64.8** **92.7**	**64.7** **93.7**	52.0 **86.8**	**76.4** **93.8**	**65.9** **93.0**	**76.0** **96.0**	**63.3** **89.9**	**71.3** **94.5**	**66.9** **92.7**
Inter-subject (Subject independent) - leave one subject out for test											
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
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
CogCap	16.3 42.3	16.2 37.9	8.8 26.8	15.4 37.6	10.1 31.7	14.0 35.4	10.7 26.9	13.9 34.2	9.0 32.4	15.3 38.6	13.0 34.4
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
ATS	14.6 35.9	**19.6** **48.3**	8.0 20.3	15.1 40.3	8.2 26.1	14.6 36.8	9.7 30.7	11.3 32.6	11.2 28.3	24.3 53.4	13.7 35.3
SAIR	**22.2** **51.2**	17.2 43.2	**13.2** **36.8**	**17.0** **45.1**	**13.1** **36.3**	**19.9** **47.4**	15.1 43.4	**17.8** **46.8**	14.1 34.1	**25.4** **56.7**	**17.5** **44.1**

and the state-of-the-art ATS [19]. As shown in Table 1, our SAIR significantly outperforms baselines in both inter-subject and intra-subject settings. Notably, SAIR achieves Top-1 and Top-5 accuracies of 66.9% and 92.7%, respectively, in the intra-subject setting, surpassing ATS by 8.8% and 6.0%, respectively.

SAIR also demonstrates strong cross-subject generalization under the inter-subject LOSO protocol, reaching 17.5% Top-1 accuracy and 44.1% Top-5 accuracy, which corresponds to 27.7% and 24.9% relative improvements over ATS, a deep-learning-based projection paradigm. These results confirm that structural-semantic aware information alignment is an effective strategy for visual decoding.

We further report comparisons with more baselines in Fig. 3(a), including BrainAlign [16] and BraVL [4]. Our SAIR framework not only achieves strong average performance across subjects, but also markedly improves the worst-subject accuracy (from 48.7% to 52.0%), indicating better generalization and robustness.

Ablation Study. We conduct an ablation study to evaluate the contribution of each component. As summarized in Fig. 3(b), SAIR achieves a peak Top-1

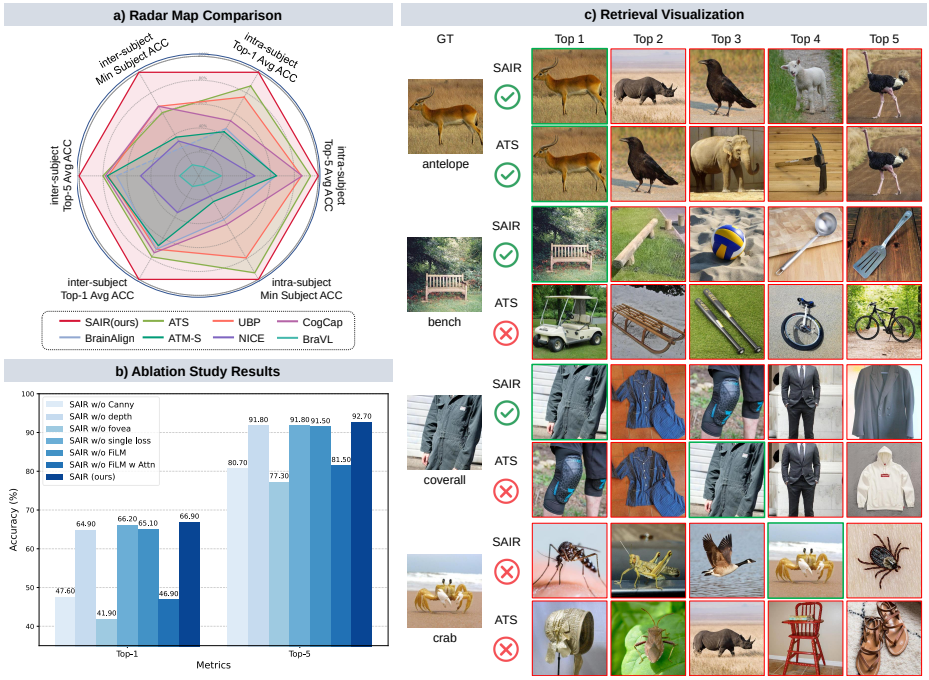

Fig. 3. (a) Radar-map visualization comparing SAIR with competing methods. (b) Ablation study results in intra-subject setting. (c) Visualization of retrieval examples comparing SAIR and ATS.

accuracy of 66.9% when all components are enabled; however, removing any image reduction strategy leads to a performance drop, demonstrating its necessity and effectiveness. Replacing FiLM with an attention block results in a substantial percentage-point decrease in accuracy, validating the importance of one-way modulation from structure to semantics.

Retrieval Visualization. Fig. 3(c) visualizes intra-subject retrieval results for subject 8. While both methods retrieve the "antelope" (Row 1), ATS's subsequent candidates (e.g., "pickax") reveal a heavy reliance on shared background colors, whereas SAIR consistently retrieves structurally related animals. For challenging queries like "bench" and "coverall" (Rows 2 and 3), SAIR achieves accurate Top-1 retrieval, while ATS yields geometrically inconsistent objects (e.g., "golf cart") due to texture distraction. Notably, even in failure cases like the "crab" (Row 4), SAIR retrieves structurally similar multi-legged candidates (e.g., "mosquito"), unlike ATS which outputs disjoint items like a "chair". These results qualitatively confirm that SAIR learns robust, shape-biased representations instead of overfitting to superficial textures.

4 Conclusion

In this paper, we introduce SAIR, a Structural-Semantic Aware Information Reduction framework for EEG-visual decoding. SAIR adopts three neuroscience-prior image reduction strategies, rather than relying on deep-learning-based adaptive compression, and designs a structure-guided semantic modulation mechanism to substantially improve the generalization of EEG retrieval. Through extensive experiments, we validate the effectiveness of our framework and its key components. SAIR not only achieves new state-of-the-art performance with significant accuracy gains, but also offers insights into image-information reduction for future brain-computer interface applications.

Acknowledgments. This study was funded by Young Scientists Fund of the National Natural Science Foundation of China (13001024).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Brody, S., Alon, U., Yahav, E.: How attentive are graph attention networks? In: The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022. OpenReview.net (2022)
2. Canny, J.: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-8**(6), 679–698 (1986). https://doi.org/10.1109/TPAMI.1986.4767851
3. Curcio, C.A., Sloan, K.R., Kalina, R.E., Hendrickson, A.E.: Human photoreceptor topography. Journal of Comparative Neurology **292**(4), 497–523 (1990). https://doi.org/10.1002/cne.902920402
4. Du, C., Fu, K., Li, J., He, H.: 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). https://doi.org/10.1109/TPAMI.2023.3263181
5. Gifford, A.T., Dwivedi, K., Roig, G., Cichy, R.M.: A large and rich eeg dataset for modeling human visual object recognition. NeuroImage **264**, 119754 (2022). https://doi.org/10.1016/j.neuroimage.2022.119754
6. Goodale, M.A., Milner, A.D., Jakobson, L.S., Carey, D.P.: A neurological dissociation between perceiving objects and grasping them. Nature **349**(6305), 154–156 (1991)
7. Grootswagers, T., Robinson, A.K., Carlson, T.A.: The representational dynamics of visual objects in rapid serial visual processing streams. NeuroImage **188**, 668–679 (2019). https://doi.org/10.1016/j.neuroimage.2018.12.046
8. Hubel, D.H., Wiesel, T.N.: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of Physiology **160** (1962)
9. Lettvin, J.Y., Maturana, H.R., McCulloch, W.S., Pitts, W.H.: What the frog's eye tells the frog's brain. Proceedings of the IRE **47**(11), 1940–1951 (1959). https://doi.org/10.1109/JRPROC.1959.287207

10. Li, D., Wei, C., Li, S., Zou, J., Liu, Q.: Visual decoding and reconstruction via eeg embeddings with guided diffusion. In: Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., Zhang, C. (eds.) Advances in Neural Information Processing Systems. vol. 37, pp. 102822–102864. Curran Associates, Inc. (2024)
11. Mishkin, M., Ungerleider, L.G.: Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behavioural Brain Research **6**(1), 57–77 (1982). https://doi.org/10.1016/0166-4328(82)90081-X
12. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
13. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
14. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision. In: Meila, M., Zhang, T. (eds.) Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 139, pp. 8748–8763. PMLR (18–24 Jul 2021)
15. Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L.: A default mode of brain function. Proceedings of the National Academy of Sciences **98**(2), 676–682 (2001). https://doi.org/10.1073/pnas.98.2.676
16. Shi, E., Hu, H., Yuan, Q., Zhao, K., Yu, S., Zhang, S.: Brainalign: Eeg-vision alignment via frequency-aware temporal encoder and differentiable cluster assigner. In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. vol. LNCS 15966. Springer Nature Switzerland (September 2025)
17. Song, Y., Liu, B., Li, X., Shi, N., Wang, Y., Gao, X.: Decoding natural images from eeg for object recognition. In: Kim, B., Yue, Y., Chaudhuri, S., Fragkiadaki, K., Khan, M., Sun, Y. (eds.) International Conference on Learning Representations. vol. 2024, pp. 47648–47665 (2024)
18. Wu, H., Li, Q., Zhang, C., He, Z., Ying, X.: Bridging the vision-brain gap with an uncertainty-aware blur prior. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 2246–2257 (2025)
19. Wu, L., Li, J., Ren, Z., Zhang, K., Gao, X.: Shrinking the teacher: An adaptive teaching paradigm for asymmetric eeg-vision alignment (2025), https://arxiv.org/abs/2511.11422
20. Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H.: Depth anything v2. In: Advances in Neural Information Processing Systems. vol. 37, pp. 21875–21911. Curran Associates, Inc. (2024). https://doi.org/10.52202/079017-0688
21. Zhang, K., He, L., Jiang, X., Lu, W., Wang, D., Gao, X.: Cognitioncapturer: Decoding visual stimuli from human eeg signal with multimodal information. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 14486–14493 (2025)