NeuroFlow: Manifold-Aware Topological Flow Matching for EEG Decoding

Lei Wang[1], Yueqin Diao[1], Hanyi Yu[1], Yuetong Yang[1], and Yanwu Xu[1,2](✉)

[1] South China University of Technology, Guangzhou, China
[2] Pazhou Lab, Guangzhou, China
ywxu@ieee.org

Abstract. Decoding visual content from non-invasive EEG is fundamentally challenged by representation collapse, semantic ambiguity, and stochastic generation drift. In this work, we propose NeuroFlow, a structured EEG–vision alignment framework that addresses these issues from both representation learning and generation perspectives. To capture transient functional connectivity without inducing dimensional collapse, we introduce a Manifold-Aware Dynamic Graph Encoder regularized by an orthogonality constraint. To further resolve the inherent semantic ambiguity of these encoded features, a Prototype-Guided Semantic Rectification layer is employed to anchor loose EEG representations to the compact support of the visual manifold. Furthermore, we reformulate the generation process using Deterministic Flow Matching, constructing an optimal transport path that ensures stable semantic transport by mitigating the stochastic drift of diffusion models. Experiments on the THINGS-EEG dataset demonstrate that NeuroFlow outperforms state-of-the-art methods in zero-shot retrieval (Top-1 Acc: 32.4%, +4.6% gain) and achieves superior generation fidelity with 5× fewer inference steps. Codes are available at https://github.com/happyeverymoment/NeuroFlow

Keywords: Brain–Computer Interface · Multimodal Learning · Electroencephalography (EEG).

1 Introduction

Visual reconstruction from electroencephalography (EEG) aims to decode external visual content from neural activity, offering significant potential for high-performance Brain-Computer Interfaces (BCIs) and prognostic evaluation tools for cognitive or visual impairments. In contrast to fMRI [20], EEG offers superior temporal resolution and portability. However, extracting semantically structured 2D images from EEG remains an ill-posed problem due to the inherent semantic ambiguity and complex non-stationary dynamics of neural signals.

Despite deep learning advancements, existing methods face three unresolved bottlenecks: **1) Conflict between Topology Modeling and Representation Collapse.** Visual perception relies on rapidly changing functional connectivity [18]. However, most methods model EEG as independent channels or static sequences [9, 12], failing to capture these dynamics. While Graph Neural Networks (GNNs) [23, 22] explicitly model topology, direct aggregation on correlated channels induces over-smoothing, rendering node features indistinguishable. This phenomenon further induces dimensional collapse in the high-dimensional latent space, where information is encoded in only a few dimensions, severely limiting representation capacity. **2) Manifold Misalignment due to Semantic Ambiguity.** Prevalent multimodal frameworks utilize contrastive learning to map EEG into a joint vision-language space [10, 15]. However, due to the low information density and high noise of EEG signals, unconstrained embeddings tend to be loosely distributed in the latent space. Such sparse latent distribution hinders effective cross-modal mapping, resulting in semantic instability, where embeddings fail to form compact clusters for similar visual concepts. **3) Stochastic Semantic Drift in Generative Priors.** SOTA reconstruction methods rely on Latent Diffusion Models (LDMs) [21, 19, 4], which generate images via stochastic reverse sampling. Under weak EEG conditioning, the probabilistic trajectory of LDMs lacks deterministic guidance, often leading to semantic drift [1, 27]. Furthermore, the extensive iterative sampling (50 steps) incurs high inference latency, hindering real-time BCI applications.

To address these challenges, we propose NeuroFlow. Our main contributions are: **(i)** Decorrelated Dynamic Graph Encoding: We propose DynamicResGNN(DRG) to explicitly model instantaneous functional connectivity. By synergizing it with a Barlow Twins constraint, we effectively mitigate over-smoothing and dimensional collapse. **(ii)** Prototype-Guided Manifold Rectification: We design a Prototype Refinement Layer(PRL) that utilizes visual priors as semantic anchors to rectify the EEG manifold structure, enhancing semantic discriminability. **(iii)** Deterministic Optimal Transport Generation: We introduce Flow Matching to EEG visual reconstruction for the first time. By constructing a deterministic Optimal Transport path, we ensure trajectory stability and reduce inference steps by 3–6×.

2 Methodology

2.1 Overview

As illustrated in Fig. 1, **NeuroFlow** establishes a high-fidelity mapping from raw neural dynamics to visual semantics via two cascaded stages. First, raw EEG signals are encoded into EEG embeddings by NeuroFlow Encoder and aligned with CLIP image and text features[17] to obtain semantically grounded representations. Second, these embeddings are transformed into image embeddings through a flow matching generator to produce the final reconstructed images.

Stage 1: Physio-Semantic Alignment Encoding. This stage projects raw EEG signals into a compact semantic manifold aligned with image embeddings [3, 2]. Given an EEG segment $\mathbf{E}$, we first model the instantaneous functional topology via DynamicResGNN (Sec. 2.2) to capture transient neural dynamics. These topological features are processed by a spatio-temporal backbone [12, 25, 8] and projected into the latent space, yielding the initial embedding

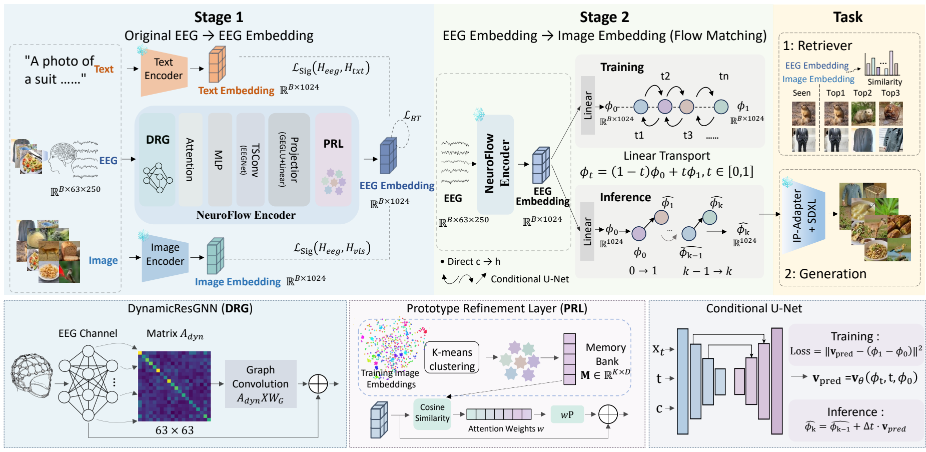

Fig. 1. NeuroFlow framework. It consists of (Stage 1) Physio-Semantic Alignment Encoding using DynamicResGNN (DRG) for topology and Prototype Refinement (PRL) for semantic rectification, and (Stage 2) Conditional Optimal Transport for high-fidelity reconstruction.

$\mathbf{H}_{\text{enc}}$. To resolve inherent semantic ambiguity, we introduce a Prototype Refinement Layer (PRL) (Sec. 2.3). This module acts as a manifold rectification mechanism, anchoring noisy embeddings to stable visual priors to produce the robust representation $\mathbf{H}_{\text{eeg}}$ used for retrieval task. NeuroFlow features a tightly coupled causal design: DRG and orthogonality constraints prevent over-smoothing, PRL rectifies noisy geometries, and Flow Matching provides deterministic transport to eliminate stochastic drift.

Stage 2: Conditional Optimal Transport Generation. Stage 2 generates visual content by transporting the aligned EEG representation $\mathbf{H}_{\text{eeg}}$ to the visual manifold. Departing from stochastic diffusion paradigms [14], we employ Conditional Flow Matching (Sec. 2.4) to construct a deterministic transport path from the EEG manifold to the visual manifold[13]. Specifically, we regress a vector field conditioned on $\mathbf{H}_{\text{eeg}}$, which guides the optimal transport trajectory to predict the target image embedding $\hat{\mathbf{H}}_{\text{vis}}$. Finally, $\hat{\mathbf{H}}_{\text{vis}}$ is decoded into the reconstructed image $\hat{\mathbf{I}}$ via a frozen SDXL backbone with an IP-Adapter[16].

2.2 Decorrelated Dynamic Graph Encoding

We propose a Decorrelated Dynamic Graph Encoding scheme, implemented via DynamicResGNN with a Barlow Twins orthogonality constraint [29], to capture transient neural connectivity while preventing feature collapse associated with graph aggregation. Instead of relying on a predefined static topology, our module infers sample-specific functional connectivity directly from the input. Specifically, we map the EEG trial to a temporal feature subspace $\mathbf{E}_\mu \in \mathbb{R}^{C\times d}$ and derive the dynamic adjacency matrix $\mathbf{A}_{\text{dyn}}$ via inner product:

$$\mathbf{A}_{\text{dyn}} = \text{Softmax}\left(\frac{\mathbf{E}_\mu \mathbf{E}_\mu^\top}{\sqrt{d}}\right), \tag{1}$$

where C denotes the number of channels. To preserve local temporal details against global topological smoothing, we employ a residual fusion mechanism:

$$\mathbf{F}_{\text{topo}} = \mathbf{A}_{\text{dyn}}\mathbf{E}\mathbf{W}_G, \qquad \mathbf{E}_{\text{out}} = \alpha\mathbf{E} + (1-\alpha)\mathbf{F}_{\text{topo}}, \tag{2}$$

where $\mathbf{W}_G$ is the graph convolution weight and α is a learnable coefficient controlling the residual contribution.

To combat dimensional collapse in the latent space, we impose Barlow Twins Loss, an orthogonality constraint as a regularizer. We compute the auto-correlation matrix $\mathcal{C}$ of the batch-normalized EEG embeddings $\mathbf{H}_{\text{eeg}}$(Sec. 2.3) and minimize:

$$\mathcal{L}_{\text{BT}} = \sum_i (1 - \mathcal{C}_{ii})^2 + \lambda \sum_{i \neq j} \mathcal{C}_{ij}^2. \tag{3}$$

2.3 Prototype-Guided Semantic Rectification

To resolve the semantic ambiguity of noisy EEG signals, we introduce the Prototype Refinement Layer (PRL). This module rectifies neural representations by anchoring them to a visual memory bank. We first construct a learnable memory bank $\mathbf{M} \in \mathbb{R}^{K \times D}$ to store representative visual prototypes. We initialize $\mathbf{M}$ via K-means clustering on the training visual features $\mathcal{V} = \{\mathbf{v}_i\}_{i=1}^N$. Specifically, we solve for the optimal centroids $\boldsymbol{\mu}$ that minimize the intra-cluster variance:

$$\min_{\mathbf{M}} \sum_{i=1}^{N} \sum_{k=1}^{K} \mathbb{I}(c_i = k) \|\mathbf{v}_i - \boldsymbol{\mu}_k\|_2^2, \tag{4}$$

where $\mathbb{I}(\cdot)$ is the indicator function and c_i is the cluster assignment. The resulting centroids $\{\boldsymbol{\mu}_k\}$ serve as the initial values for $\mathbf{M}$. Given this semantic memory, we then refine the encoded EEG features by retrieving relevant prototypes. For the coarse query $\mathbf{H}_{\text{enc}}$, we compute its attention weights over the memory bank:

$$\mathbf{w} = \text{Softmax}\left(\frac{\mathbf{H}_{\text{enc}}\mathbf{M}^\top}{\tau}\right), \tag{5}$$

where τ controls the attention sharpness. Using the retrieved prototypes, we obtain the rectified representation $\mathbf{H}_{\text{eeg}}$ via adaptive gating:

$$\mathbf{H}_{\text{anchor}} = \mathbf{w}\mathbf{M}, \quad \mathbf{H}_{\text{eeg}} = (1-\beta)\mathbf{H}_{\text{enc}} + \beta\mathbf{H}_{\text{anchor}}, \tag{6}$$

where β is a learnable parameter balancing the input features and the retrieved prototypes. This process effectively pulls outlier embeddings back towards the compact semantic centers. Finally, to enforce cross-modal alignment, we optimize a multi-task objective:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{Sig}}(\mathbf{H}_{\text{eeg}}, \mathbf{H}_{\text{vis}}) + \lambda_{\text{txt}}\mathcal{L}_{\text{Sig}}(\mathbf{H}_{\text{eeg}}, \mathbf{H}_{\text{txt}}) + \lambda_{\text{BT}}\mathcal{L}_{\text{BT}}. \tag{7}$$

The targets $\mathbf{H}_{\text{vis}}, \mathbf{H}_{\text{txt}} \in \mathbb{R}^D$ are extracted from a frozen CLIP encoder, sharing the same dimension as $\mathbf{H}_{\text{eeg}}$ to ensure a stable semantic space. We employ SigLIP loss $\mathcal{L}_{\text{Sig}}$ [28] for its superior training stability over standard CLIP. Text embeddings $\mathbf{H}_{\text{txt}}$ serve as auxiliary constraints to regularize the joint embedding space.

2.4 Conditional Optimal Transport via Flow Matching

Stochastic diffusion paths often diverge given weak EEG signals. We reframe generation as a Conditional Optimal Transport (OT) problem, regressing a vector field $\mathbf{u}_t$ that maps the EEG source to the visual target via a deterministic linear trajectory. We define the flow from source $\phi_0 = \mathbf{H}_{\text{eeg}}$ to target $\phi_1 = \mathbf{H}_{\text{vis}}$. The optimal transport path is defined as a linear interpolation:

$$\phi_t = (1-t)\phi_0 + t\phi_1, \quad t \in [0,1]. \tag{8}$$

This linear path corresponds to a constant conditional vector field $\mathbf{u}_t^*$ that minimizes the transport cost:

$$\mathbf{u}_t^*(\phi_t \mid \phi_0, \phi_1) = \frac{\mathrm{d}\phi_t}{\mathrm{d}t} = \phi_1 - \phi_0. \tag{9}$$

Here, the velocity represents the constant semantic shift required to transform the EEG prior into the visual target. We train a neural velocity estimator $\mathbf{v}_\theta$ conditioned on ϕ_0. We construct a standard conditional U-Net [7]. This network minimizes the Conditional Flow Matching (CFM) loss:

$$\mathcal{L}_{\text{CFM}} = \mathbb{E}_{t,\phi_0,\phi_1} \left\| \mathbf{v}_\theta(\phi_t, t, \phi_0) - (\phi_1 - \phi_0) \right\|_2^2, \tag{10}$$

where $t \sim \mathcal{U}(0,1)$ is sampled uniformly. During inference, we integrate the ODE $\mathrm{d}\hat{\phi}_t/\mathrm{d}t = \mathbf{v}_\theta(\hat{\phi}_t, t, \phi_0)$ from $\phi_0 = \mathbf{H}_{\text{eeg}}$. Using a first-order Euler solver with step size $\varDelta t = 1/N$, the deterministic update rule is:

$$\hat{\phi_{k+1}} = \hat{\phi_k} + \varDelta t \cdot \mathbf{v}_\theta(\hat{\phi_k}, \tau_k, \phi_0). \tag{11}$$

With a small step count (e.g., $N = 20$), this formulation enables rapid, deterministic transport to the image manifold, mitigating the semantic drift typical of stochastic sampling.

3 Experiments

3.1 Experimental Setup

Dataset and Metrics. We use the THINGS-EEG dataset[6], containing 10 subjects and 1,854 concepts (16,540 training/200 test images).We follow a strict

Table 1. Performance comparison of different models used as EEG encoders for retrieval task. Each row shows the average accuracy and standard deviation across all subjects for each method. Best results are in **bold**, and second-best are underlined.

Method	v200 Top1	v200 Top5	v4 Top1	v50 Top5	v100 Top5
LSTM	1.4±0.5	2.8±0.8	25.4±2.0	10.7±1.6	5.1±1.0
CNN	5.8±2.0	16.5±6.3	55.0±9.9	35.4±10.3	24.1±7.1
ShallowFBCSPNet	1.1±0.2	2.9±0.4	25.4±2.4	10.5±1.2	5.6±0.8
EEGITNet	1.0±0.1	2.7±0.2	28.1±3.1	10.4±0.7	5.5±0.4
MLP	10.5±1.9	26.3±5.1	71.4±4.5	56.3±7.1	39.8±6.5
MetaEEG [5]	13.8±3.2	36.1±5.9	74.9±6.3	63.6±6.7	49.4±7.2
NICE [23]	23.4±6.1	50.5±7.4	84.7±4.5	79.6±6.1	66.1±6.3
EEGNet [11]	19.3±4.3	44.3±8.3	81.5±3.7	72.9±7.2	58.6±5.9
SeeEEG [9]	22.4±5.7	47.6±8.8	80.9±5.5	74.1±6.0	62.0±7.1
DreamDiffusion [1]	21.7±4.5	45.9±8.2	84.0±4.0	73.4±5.9	59.1±6.5
BrainDreamer [27]	20.3±6.7	43.4±10.5	78.2±5.9	68.8±8.5	56.5±8.4
WPFModel [24]	25.5±5.4	53.80±7.7	84.6±5.1	<u>80.9±6.6</u>	67.5±8.2
CBraMod [26]	5.6±1.9	16.8±6.4	56.4±9.2	39.5±10.4	26.1±8.8
CoCapturer [30]	<u>27.8±5.2</u>	<u>56.2±8.3</u>	85.2±5.2	80.2±6.2	68.2±7.5
ATM_E [12]	27.2±5.5	53.7±7.4	<u>86.5±3.3</u>	80.6±5.9	<u>68.9±5.7</u>
NeuroFlow(Ours)	**32.4±4.9**	**60.7±7.8**	**88.7±3.9**	**83.2±6.0**	**73.6±7.9**

zero-shot protocol with disjoint training and test categories. Using a single RTX 4090,we train Stage-1 via AdamW ($lr = 3 \times 10^{-4}$, $bs = 1024$) for 60 epochs. Stage-2 adopts a 1D conditional U-Net backbone trained for 30 epochs. For inference, we employ Euler integration with $N = 20$ steps. For retrieval, we report Top-1/5 accuracy. For generation, we assess semantic (CLIP Score), structural (SSIM), and pixel (PixCorr) fidelity, alongside efficiency.

3.2 Retrieval Performance

Main Results. As shown in Table 1, NeuroFlow achieves a state-of-the-art Top-1 accuracy of **32.4%**, significantly outperforming the best baseline (CoCapturer) by **+4.6%**. Top-5 accuracy also improves by **+4.5%** over the strongest baseline, indicating more reliable semantic ranking. This substantial gain validates the effectiveness of our manifold-aware encoding in resolving semantic ambiguity while preserving the intrinsic topology of EEG representations.

3.3 Generation Performance

Quality vs. Speed. As shown in Table 2, NeuroFlow surpasses SOTA methods (ATM, CoCapturer) in both semantic (CLIP, Inception) and structural (SSIM) fidelity. The marginal drop in PixCorr reflects an inherent property of Flow Matching: it prioritizes semantic manifold alignment over fitting high-frequency stochastic noise. Leveraging the deterministic Optimal Transport trajectory, NeuroFlow converges in just 20 steps (7.7 s), achieving a 6.5× speedup over

Table 2. Comparison of image generation performance and inference efficiency across different metrics. ↑ indicates higher is better, ↓ indicates lower is better.

Method	SSIM ↑	Inception ↑	CLIP ↑	PixCorr ↑	Inference Time ↓ (s)
EEG Embedding Only	0.360	0.661	0.769	0.133	–
ATM (50 steps)	0.374	0.731	0.828	**0.164**	50.1
CoCapturer (50 steps)	0.347	0.669	0.715	<u>0.150</u>	–
NeuroFlow (10 steps)	<u>0.413</u>	0.728	0.820	0.138	**3.8**
NeuroFlow (20 steps)	**0.414**	<u>0.734</u>	**0.841**	0.142	<u>7.7</u>
NeuroFlow (50 steps)	0.412	**0.758**	<u>0.833</u>	0.134	18.4

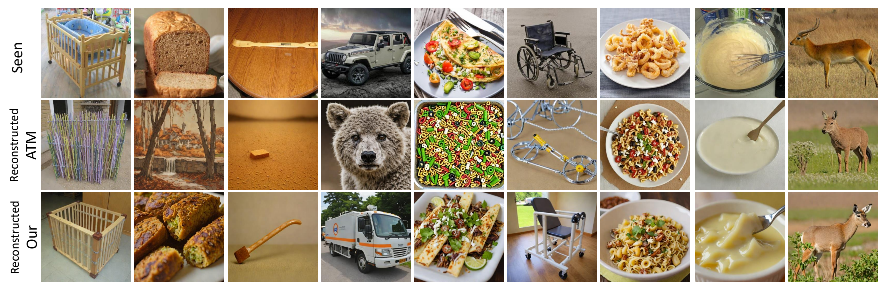

Fig. 2. Qualitative comparison. Our method significantly improves semantic consistency and reduces structural artifacts.

stochastic diffusion baselines (50.1 s). Qualitative reconstruction examples are shown in Fig. 2, which shows that NeuroFlow has stronger semantic consistency.

3.4 Ablation & Interpretability Analysis

Table 3 highlights a trade-off: DynamicResGNN alone degrades performance (27.2% → 26.8%) due to topological over-smoothing. However, coupling it with the Barlow Twins constraint yields a substantial boost to **30.2%** (+3.0%). This demonstrates that feature decorrelation is a prerequisite for effective dynamic topology modeling, preventing representation collapse. The full model further reaches 32.4% with prototype refinement. These analyses mathematically corroborate the resolution of semantic ambiguity and dimensional collapse. **Universality.** We further evaluate the generality of our joint objective by applying the SigLIP+BT Loss to multiple EEG Encoder backbone architectures. Replacing the standard CLIP loss with our objective improves performance (Table 3), demonstrating that the orthogonality constraint serves as a robust, architecture-agnostic regularizer.

We validate NeuroFlow through multiple interpretability analyses (Fig. 3). **Topological Importance.** Channel ablation (Fig. 3a) identifies the occipital region as the dominant information source, aligning with neurophysiological priors on visual processing. **Manifold Rectification.** t-SNE visualization (Fig. 3b) confirms that PRL pulls loose EEG embeddings into compact clusters around

Table 3. Unified ablation study showing component configurations and corresponding retrieval performance.

Method	$\mathcal{L}_{\mathrm{BT}}$	PRL	DRG	v200 Top1	v200 Top5	v100 Top5	v50 Top5	Δ Top1
MetaEEG	✓	–	–	15.2±4.3	36.9±7.8	50.8±6.8	64.6±7.5	+1.4
NICE	✓	–	–	27.8±5.9	59.7±8.0	75.0±5.4	86.6±5.0	+4.3
EEGNet	✓	–	–	23.7±6.8	52.2±9.3	65.4±8.4	78.6±5.9	+4.4
SeeEEG	✓	–	–	23.4±6.2	47.2±7.7	60.6±9.1	75.3±8.5	+1.1
NeuroFlow	–	–	✓	26.8±6.4	53.6±8.5	66.2±7.4	78.4±6.3	-0.5
NeuroFlow	✓	–	–	28.3±4.7	56.6±6.5	71.0±5.5	83.7±4.5	+1.1
NeuroFlow	✓	–	✓	30.2±6.1	58.6±5.9	72.8±4.5	83.7±4.4	+3.0
NeuroFlow	✓	✓	–	29.6±6.4	58.5±6.7	72.5±6.6	83.9±5.2	+2.4
NeuroFlow	✓	✓	✓	**32.4±4.9**	**60.7±7.8**	**73.6±7.9**	**83.2±6.0**	**+5.2**

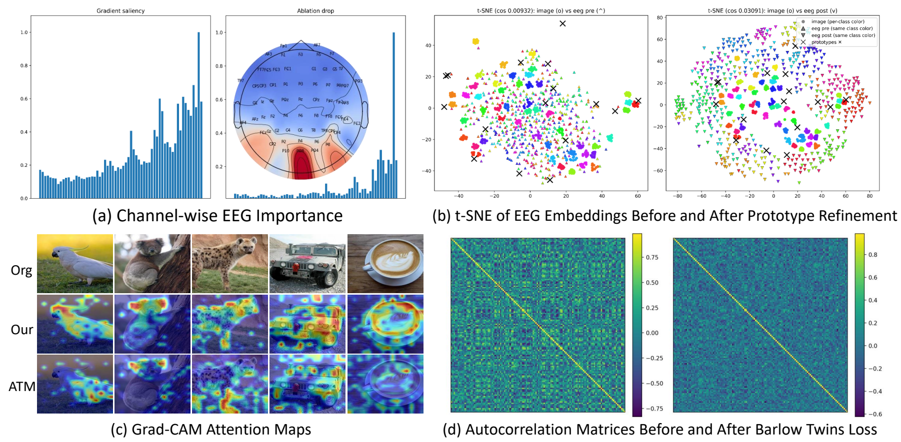

Fig. 3. Mechanism analysis. (a) Saliency aligns with visual cortex. (b) PRL improves clustering (t-SNE). (c) Attention focuses on subjects. (d) Barlow Twins ($\mathcal{L}_{\mathrm{BT}}$) decorrelates features.

visual anchors, effectively resolving semantic ambiguity. **Visual Saliency.** Unlike the uniform attention of baselines, NeuroFlow (Fig. 3c) explicitly focuses on semantically salient image regions, demonstrating precise cross-modal alignment. **Feature Orthogonality.** Correlation matrices (Fig. 3d) reveal a sparse off-diagonal structure compared to baselines. This confirms that our orthogonality constraint effectively mitigates dimensional collapse.

4 Conclusion

We present NeuroFlow, a unified framework resolving semantic ambiguity and dimensional collapse in neural decoding. By synergizing manifold-aware dynamic topology with prototype rectification, we ensure precise physio-semantic alignment. Furthermore, our Conditional Optimal Transport paradigm enables rapid, deterministic generation, reducing stochastic drift. Empirical analyses show high parameter robustness with no statistically significant performance fluctuations. This work establishes structured manifold learning and deterministic transport as critical foundations for high-fidelity BCIs. While promising, our evaluation is currently limited to a single benchmark dataset, and broader validation across diverse recording conditions is an important direction for future work. In addition, exploring more flexible transport formulations may further improve fine-grained reconstruction fidelity.

Acknowledgments. This work was supported in part by the Major Science and Technology Special Projects for Cancer, Cardiovascular, Respiratory and Metabolic Diseases (No.2025ZD0547202), and the Guangzhou Science and Technology Department (No.2024D03J0013).

Disclosure of Interests. The authors declare no competing interests.

References

1. Bai, Y., Wang, X., pei Cao, Y., Ge, Y., Yuan, C., Shan, Y.: Dreamdiffusion: Generating high-quality images from brain eeg signals (2023), https://arxiv.org/abs/2306.16934
2. Cao, X., Gong, P., Zhang, L., Zhang, D.: Eeg-clip: A transformer-based framework for eeg-guided image generation. Neural Networks **194**, 108167 (2026). https://doi.org/https://doi.org/10.1016/j.neunet.2025.108167
3. Chen, H., He, L., Liu, Y., Yang, L.: Visual neural decoding via improved visual-eeg semantic consistency. CoRR **abs/2408.06788** (2024). https://doi.org/10.48550/ARXIV.2408.06788
4. Guo, Z., Wu, J., Song, Y., Bu, J., Mai, W., Zheng, Q., Ouyang, W., Song, C.: Neuro-3d: Towards 3d visual decoding from eeg signals (2025), https://arxiv.org/abs/2411.12248
5. Han, J.W., Bak, S., Kim, J.M., Choi, W., Shin, D.H., Son, Y.H., Kam, T.E.: Meta-eeg: Meta-learning-based class-relevant eeg representation learning for zero-calibration brain–computer interfaces. Expert Systems with Applications **238**, 121986 (2024)
6. Hebart, M.N., Contier, O., Teichmann, L., Rockter, A.H., Zheng, C.Y., Kidder, A., Corriveau, A., Vaziri-Pashkam, M., Baker, C.I.: Things-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. Elife **12**, e82580 (2023)
7. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
8. Kim, H.J., Park, D., Jang, J.W., Ju, Y.G., Won, D.O.: SSPNet: Towards Feasible Spatio-Spectral Portraits-Based Deep Learning Framework for Neurodegenerative Disease Multi-Classification . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. vol. LNCS 15963. Springer Nature Switzerland (September 2025)
9. Kim, J.M., Choi, W., Park, S.J., Heo, K.S., Son, Y.H., Oh, J.H., Shin, D.H., Kam, T.E.: Seeeeg: Semantic-aware eeg-based multi-modal retrieval-augmented generation for high-fidelity visual brain decoding. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops. pp. 4824–4833 (October 2025)

10. Lan, Y.T., Ren, K., Wang, Y., Zheng, W.L., Li, D., Lu, B.L., Qiu, L.: Seeing through the brain: Image reconstruction of visual perception from human brain signals (2023), https://arxiv.org/abs/2308.02510
11. Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J.: Eegnet: A compact convolutional network for eeg-based brain-computer interfaces. Journal of Neural Engineering **15**(5), 056013.1–056013.17 (2018)
12. Li, D., Wei, C., Li, S., Zou, J., Qin, H., Liu, Q.: Visual decoding and reconstruction via eeg embeddings with guided diffusion (2024), https://arxiv.org/abs/2403.07721
13. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. arXiv preprint arXiv:2210.02747 (2022)
14. Liu, X.H., Lu, B.L., Zheng, W.L.: Eegmirror: Leveraging eeg data in the wild via montage-agnostic self-supervision for eeg to video decoding. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 18273–18283 (October 2025)
15. Lopez, E., Sigillo, L., Colonnese, F., Panella, M., Comminiello, D.: Guess what i think: Streamlined eeg-to-image generation with latent diffusion models (2025), https://arxiv.org/abs/2410.02780
16. Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., Rombach, R.: Sdxl: Improving latent diffusion models for high-resolution image synthesis. arXiv preprint arXiv:2307.01952 (2023)
17. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PmLR (2021)
18. Sadaghiani, S., Poline, J.B., Kleinschmidt, A., D'Esposito, M.: Ongoing dynamics in large-scale functional connectivity predict perception. Proceedings of the National Academy of Sciences **112**(27), 8463–8468 (2015)
19. Scotti, P.S., Banerjee, A., Goode, J., Shabalin, S., Nguyen, A., Cohen, E., Dempster, A.J., Verlinde, N., Yundler, E., Weisberg, D., Norman, K.A., Abraham, T.M.: Reconstructing the mind's eye: fmri-to-image with contrastive learning and diffusion priors (2023), https://arxiv.org/abs/2305.18274
20. Scotti, P.S., Tripathy, M., Villanueva, C.K.T., Kneeland, R., Chen, T., Narang, A., Santhirasegaran, C., Xu, J., Naselaris, T., Norman, K.A., Abraham, T.M.: Mindeye2: Shared-subject models enable fmri-to-image with 1 hour of data (2024), https://arxiv.org/abs/2403.11207
21. 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)
22. Shi, S., Zheng, X., Wu, T., Hao, X.: Various Attention Mechanism Graph Convolutional Network with Multi-Source Domain Adaptation for Cross-Subject EEG Emotion Recognition . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. vol. LNCS 15971. Springer Nature Switzerland (September 2025)
23. Song, Y., Liu, B., Li, X., Shi, N., Wang, Y., Gao, X.: Decoding natural images from eeg for object recognition (2024), https://arxiv.org/abs/2308.13234
24. Su, C.: Dlinear makes efficient long-term predictions. In: Proceedings of ACM Conference (Baidu KDD Cup) (2022)
25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. arXiv (2017)

26. Wang, J., Zhao, S., Luo, Z., Zhou, Y., Jiang, H., Li, S., Li, T., Pan, G.: Cbramod: A criss-cross brain foundation model for eeg decoding. arXiv preprint arXiv:2412.07236 (2024)
27. Wang, L., Wu, C., Wang, L.: Braindreamer: Reasoning-coherent and controllable image generation from eeg brain signals via language guidance (2024), https://arxiv.org/abs/2409.14021
28. Wu, H., Li, Q., Zhang, C., He, Z., Ying, X.: Bridging the vision-brain gap with an uncertainty-aware blur prior (2025), https://arxiv.org/abs/2503.04207
29. Zbontar, J., Jing, L., Misra, I., LeCun, Y., Deny, S.: Barlow twins: Self-supervised learning via redundancy reduction. In: International conference on machine learning. pp. 12310–12320. PMLR (2021)
30. Zhang, K., He, L., Jiang, X., Lu, W., Wang, D., Gao, X.: Cognitioncapturer: Decoding visual stimuli from human eeg signal with multimodal information (2024), https://arxiv.org/abs/2412.10489