Mind the Parameters: Lightweight and Efficient Brain Visual Decoding with Shared Tensor Cores

Haodong Jing[1], **Panqi Yang**[1], **Rongchao Zhang**[2], **Junhao Jia**[3], **Hoi Leong Lee**[4], **Zhipeng Liu**[5], **Yongqiang Ma**[1*], **Nanning Zheng**[1*]

[1]State Key Laboratory of Human-Machine Hybrid Augmented Intelligence, Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University

[2]Peking University [3]Hangzhou Dianzi University

[4]Universiti Malaysia Perlis [5]Northeastern University

jinghd@stu.xjtu.edu.cn, nnzheng@xjtu.edu.cn

Abstract

Linking brain activity to computational representations of visual perception is a central goal at the intersection of neuroscience and machine learning. However, cross-subject decoding from fMRI remains challenging because neural responses vary substantially across individuals, while existing methods often rely on subject-specific fine-tuning or parameter-heavy decoders. This limits generalization to unseen subjects and hinders efficient deployment. We propose BrainTC–Brain Tensor Cores, a lightweight framework for cross-subject brain visual decoding based on shared tensor decomposition. BrainTC separates shared functional structure from subject-specific variation through a shared tensor-core alignment module that maps ROI-wise responses into a common functional space with shared tensor bases and residual subject factors. This supports both zero-shot decoding for unseen subjects and few-shot adaptation by updating only residual parameters. We then develop a Brain Tensor-Transformer with tensorized attention to compactly model inter-ROI dependencies, together with a hierarchical neural-to-visual mapping module for multi-level visual prediction. Experiments show that BrainTC achieves competitive cross-subject decoding performance with substantially improved parameter efficiency, particularly in zero-shot generalization to unseen subjects. Our code and model weights are publicly available at this link.

1 Introduction

Visual perception facilitates authentic understanding and interaction with the physical world, functioning simultaneously as an internal mapping of external reality and an external manifestation of neural cognitive processes [1, 2]. Situated at the intersection of artificial intelligence and cognitive neuroscience, visual decoding serves as a pivotal framework for elucidating complex cognitive mechanisms [3, 4, 5]. Among various neuroimaging modalities, functional Magnetic Resonance Imaging (fMRI) [6] offers a comprehensive observational window, distinguished by its high spatial resolution and non-invasive nature. Consequently, the analysis of fMRI signals not only bridges neural activity with perceptual experiences but also establishes a theoretical foundation for developing brain-inspired computing models and scalable next-generation BCIs [7, 8].

In this domain, reconstructing visual experiences from fMRI signals has become a key testbed for neural representation learning [9, 10]. Existing research primarily centers on two core sub-problems:

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

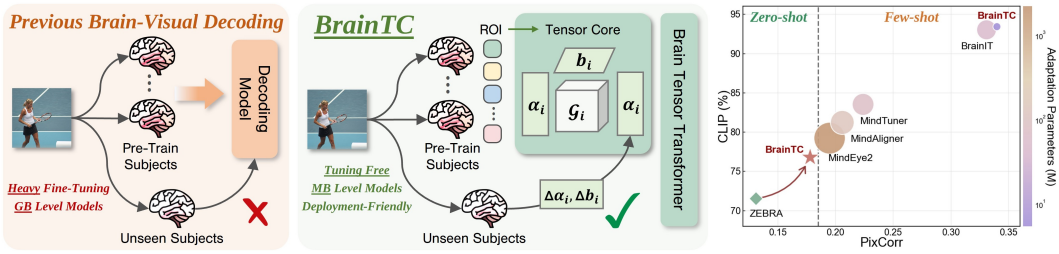


Figure 1: Existing brain decoding models achieve strong reconstruction and generalization performance, but their large fine-tuning parameter costs hinder deployment in resource-constrained settings. BrainTC achieves parameter-efficient decoding by adopting the ROI → Tensor perspective.

first, learning the non-linear mapping from noisy brain signals to fine-grained visual-semantic representations; second, injecting these representations as conditions into generative models to achieve high-fidelity reconstruction. With the construction of large-scale high-quality fMRI datasets and advancements in large-scale pre-trained models, tremendous success has been achieved over the past decade [11, 12, 13, 14, 15, 16, 17, 18, 19] in dimensions such as semantic consistency, spatial distribution, and reconstruction details.

However, despite the exciting flourishing landscape of high-fidelity reconstruction, the "experiment-reality" gap hindering practical application is becoming increasingly prominent. First, although cross-subject decoding has garnered attention, many existing methods still rely heavily on costly subject-specific fine-tuning. While some recent studies attempt zero-shot adaptation without fine-tuning [18, 15], the implicit embedding pre-alignment during decoding inevitably requires massive data to cover the "subject space" [18, 16], thereby limiting their generalization capabilities in large-scale screening or heterogeneous data scenarios [20, 21]. Second, existing models predominantly rely on parameter-heavy backbone networks; traditional attention mechanisms often neglect the inherent "spatial-subject" tensor structure [22, 23] when processing fMRI, compelling models to consume substantial computational resources to relearn these prior relationships. This prohibitive computational burden hinders their practical deployment on resource-constrained mobile systems.

In reviewing the deployment and generalization of LLMs, Parameter-Efficient Fine-Tuning (PEFT) [24, 25, 26] has become an effective solution. Among these, LoRA and its variants [27, 28] are widely adopted due to their simplicity and stability. However, LoRA is inherently confined to low-rank approximation in two-dimensional space; while this mathematical simplification is convenient, it fails to fully capture the impact of weight variations on high-dimensional neural data. As a generalization of matrix decomposition, tensor decomposition methods have emerged as a powerful technique for approximating high-dimensional spaces [29, 30] and have been widely applied in neural network compression and generalization enhancement [31, 32, 33, 34]. Recent studies indicate that introducing dynamic tensor decomposition into attention mechanisms [35, 36] can effectively improve model representation efficiency and significantly reduce memory overhead, providing a new perspective for the lightweight deployment of brain decoding models. Given the theoretical advantages of tensor decomposition in capturing high-dimensional structures, this paper aims to explore a core question: *Can **tensorized shared parameterization** support competitive zero-/few-shot cross-subject brain decoding without full subject-specific decoder fine-tuning, while remaining lightweight?*

To answer this question, we propose **BrainTC** (**Brain** **T**ensor **C**ores), a lightweight and efficient cross-subject brain visual decoding framework built on shared tensor decomposition. As shown in Fig. 1, the main idea of BrainTC is to explicitly model shared neural structure and subject-specific variation through a unified tensor-core parameterization, rather than relying on heavy subject-specific tuning or dense over-parameterized backbones. Specifically, BrainTC first aligns ROI-wise neural responses across subjects into a common functional space using subject-dependent projections generated by **Shared Tensor Cores** and lightweight residual factors. This design naturally supports both zero-shot decoding and few-shot target-subject calibration by updating only a small number of residual parameters. Based on the aligned representations, we further develop a **Brain Tensor-Transformer** to model structured inter-ROI dependencies with tensorized attention and feed-forward layers, substantially improving parameter efficiency. Finally, we introduce a *hierarchical neural-to-visual mapping* module that predicts multi-level visual representations through shared coupling components,

enabling the decoder to capture neural correspondences from perceptual details to semantic concepts. Through these designs, BrainTC provides a unified solution for explicit cross-subject alignment, efficient neural interaction modeling, and lightweight subject adaptation. Extensive experiments demonstrate that BrainTC achieves strong and competitive cross-subject decoding performance while using a substantially more parameter-efficient formulation. Notably, BrainTC shows clear advantages in zero-shot cross-subject decoding, where it generalizes effectively to unseen subjects without requiring full subject-specific retraining.

In summary, our main contributions are as follows:

- We propose **BrainTC**, a lightweight and efficient tensorized framework for cross-subject brain visual decoding, which provides a unified solution to the long-standing trade-off between cross-subject generalization and subject-specific adaptation.
- We introduce a shared tensor-core functional alignment scheme with weight-decomposed subject factors, which disentangles transferable subject-invariant functional structure from individual variability. This design supports parameter-efficient zero-shot/few-shot adaptation to unseen subjects.
- We show that a tensorized architecture, instantiated as a Brain Tensor-Transformer with hierarchical neural-to-visual mapping, can model inter-ROI interactions and neural-to-visual correspondences in a significantly more compact manner, leading to robust performance in cross-subject decoding.

2 Related Works

2.1 fMRI-Based Brain Decoding

Reconstructing visual information from fMRI signals serves as a bridge between neuroscience and AI [3, 4], over the past decade has undergone a transformative shift from linear mapping to generative modeling. Early decoding methods were constrained by linear regression models and small-scale data [37], making it difficult to capture complex visual features. With the emergence of large-scale, high-resolution fMRI datasets [38, 39, 40], the realism of reconstructions has significantly improved [41]. In particular, the introduction of diffusion models has brought about a paradigm shift. Works such as MinD-Vis [42, 9, 10, 43] have achieved high-fidelity reconstruction by aligning fMRI signals to the CLIP [44] space and leveraging the powerful generative prior of LDM [45]. Subsequently, MindEye series [46, 11], UMBRAE [12], NeuroPictor [13], and other works [47, 15, 48, 49, 50] further introduced multimodal contrastive learning constraints, significantly enhancing the semantic consistency and spatial structural validity of reconstructions. Additionally, to overcome the limitations of single-subject models, methods based on latent space alignment, physiological functional registration, and semantic invariant representation decoupling have effectively reduced the reliance on large amounts of fine-tuning data from target subjects [14, 16, 17, 51, 52, 53, 18, 54, 55, 56, 57, 58]. However, generalizability is only one of the challenges brain decoding faces in practical application, greater challenges lie in the high computational costs and inference latency associated with complex models. Some researchers have already begun attempting to construct lightweight brain decoding architectures [59, 60].

2.2 Low-rank and Tensor Decomposition

Low-rank approximation and tensor decomposition form the cornerstone of machine learning [29, 30, 61] for processing high-dimensional data and compressing over-parameterized models. As a mathematical generalization of matrix singular value decomposition (SVD) to multidimensional space [62, 63], tensor decomposition decouples high-order tensors into low-rank factor matrices or compact products of core tensors through paradigms such as CP [64], Tucker [65], and Tensor Train (TT) [29] and Tensor Ring (TR) [66]. This "tensorization" strategy capitalizes on the inherent low-rank redundancy of neural network parameters, often maintaining or even enhancing a model's representational capacity [67] while significantly reducing storage and computational overhead. In the field of LLM compression, training-free SVD schemes based on activation-aware truncation strategies ($e.g.$, SVD-LLM [68]) have become powerful tools for efficiently compressing Transformer weights. Throughout the evolution of deep learning, the effectiveness of tensor decomposition has been repeatedly validated in CNNs [31, 32], RNNs [33, 69, 34, 70], Transformer attention

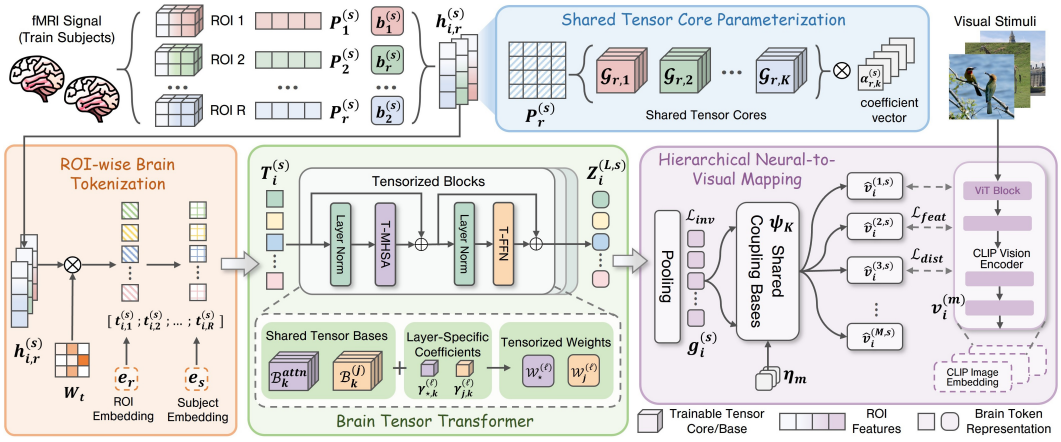


Figure 2: **BrainTC** consists of three modules: Functional Alignment with Shared Tensor Cores for ROI-wise cross-subject alignment, Brain Tensor-Transformer for structured neural decoding, and Hierarchical Neural-to-Visual Mapping for multi-level visual prediction.

mechanisms [35, 36], and MoE architecture [71]. In neuroscience, neural imaging data naturally exhibit higher-order tensor structures [72], and the decoding process is universally challenged by inter-subject variability and data heterogeneity. Tensor decomposition, backed by the guarantee of "essential uniqueness"—a feature lacking in matrix methods such as PCA [73], makes them promising for extracting embeddings from complex brain activity that are clearly interpretable and generalizable.

3 Method

The overall framework of **BrainTC** is shown in Fig. 2. Our goal is to develop a unified cross-subject neural decoding framework that improves both generalization to unseen subjects and adaptation efficiency for visual reconstruction. To this end, we propose **BrainTC**, a tensorized architecture that integrates *Functional Alignment with Shared Tensor Cores*, *Brain Tensor-Transformer*, and *Hierarchical Neural-to-Visual Mapping* within a shared structured parameterization.

3.1 Preliminaries

The decoder is trained on source subjects $\mathcal{S}_{\text{tr}}$ and evaluated on an unseen target subject $s^* \notin \mathcal{S}_{\text{tr}}$. For subject s, the dataset is $\mathcal{D}^{(s)} = \{(x_i^{(s)}, y_i)\}_{i=1}^{N_s}$, each fMRI response is organized into R predefined ROIs, $x_i^{(s)} = \{x_{i,r}^{(s)}\}_{r=1}^{R}$, $x_{i,r}^{(s)} \in \mathbb{R}^{V_r}$, V_r denotes the fixed ROI input dimensionality after padding or truncation. The decoder predicts hierarchical visual representations $f_\theta(x_i^{(s)}) = \{\hat{v}_i^{(m,s)}\}_{m=1}^{M}$, supervised by multi-level targets $\{v_i^{(m)}\}_{m=1}^{M}$ extracted from a pretrained visual encoder.

3.2 Functional Alignment with Shared Tensor Cores

Subject-Dependent Functional Projection. Neural responses across subjects exhibit substantial inter-subject variability in voxel geometry and functional organization, making it difficult to apply a shared decoder across subjects directly. To reduce this discrepancy, we first map ROI-wise responses into a shared functional space using subject-dependent alignment parameters. Given the ROI-wise neural response $x_i^{(s)}$ from subject s, we map it into a shared functional space and obtain an aligned representation $h_i^{(s)} \in \mathbb{R}^{R \times D}$, where D is the latent dimensionality of the shared functional space.

Concretely, for each ROI r, the neural response $x_{i,r}^{(s)} \in \mathbb{R}^{V_r}$ is mapped to:

$$h_{i,r}^{(s)} = P_r^{(s)} x_{i,r}^{(s)} + b_r^{(s)}, \tag{1}$$

where $P_r^{(s)} \in \mathbb{R}^{D \times V_r}$ and $b_r^{(s)} \in \mathbb{R}^D$ denote the projection matrix and bias for ROI r and subject s. Stacking all ROI representations yields $h_i^{(s)} \in \mathbb{R}^{R \times D}$. These aligned features are not yet the final Transformer tokens; instead, they form a subject-harmonized intermediate representation, which is subsequently converted into ROI-wise brain tokens for Transformer-based decoding.

Shared Tensor Core Parameterization. A direct implementation of the ROI-wise projection matrices $\{P_r^{(s)}\}_{r=1}^R$ would require learning an independent transformation for each ROI and each subject. Such a formulation is parameter-intensive and does not explicitly exploit the common functional structure shared across subjects. To address this issue, we parameterize the subject-dependent projection weights by a set of shared tensor cores together with lightweight subject-specific factors. Specifically, $P_r^{(s)}$ is reshaped into a tensor $\mathcal{P}_r^{(s)}$ and reconstructed as:

$$\mathcal{P}_r^{(s)} \in \mathbb{R}^{d_1 \times \cdots \times d_P \times v_1 \times \cdots \times v_Q}, \tag{2}$$

where $\prod_{p=1}^P d_p = D$ and $\prod_{q=1}^Q v_q = V_r$. Instead of learning $\mathcal{P}_r^{(s)}$ independently for every subject, we express it through a shared tensor core set $\mathcal{G}_r = \{\mathcal{G}_{r,1}, \mathcal{G}_{r,2}, \ldots, \mathcal{G}_{r,K}\}$, which captures ROI-level transformation patterns common to all subjects. Subject-specific variability is introduced by a coefficient vector $\alpha_r^{(s)} \in \mathbb{R}^K$, which is parameterized as $\alpha_r^{(s)} = \bar{\alpha}_r + \Delta\alpha_r^{(s)}$.

Using these coefficients, the projection tensor is constructed as $\mathcal{P}_r^{(s)} = \sum_{k=1}^K \alpha_{r,k}^{(s)} \mathcal{G}_{r,k}$, and the corresponding matrix form $P_r^{(s)}$ is obtained by folding $\mathcal{P}_r^{(s)}$ back to $\mathbb{R}^{D \times V_r}$.

This parameterization provides an explicit decomposition for cross-subject functional alignment. The shared tensor cores encode transferable functional geometry that is reused across subjects, while the coefficient vector $\alpha_r^{(s)}$ captures how each subject deviates from the shared template. Compared with learning fully independent projection matrices, this design greatly reduces the number of subject-specific parameters and improves statistical efficiency in low-data cross-subject settings. More discussion of this design is provided in *Appendix.* **??**.

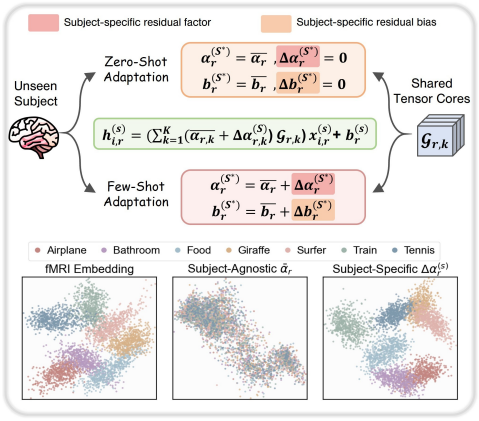


Figure 3: Zero/few-shot adaptation of new subjects, and t-SNE of fMRI, agnostic coefficients $\bar{\alpha}_r$, and specific factor $\Delta\alpha_r^{(s)}$.

Weight-Decomposed Tensor Adaptation. The decomposed tensor parameterization naturally enables efficient adaptation to an unseen target subject s^*. Rather than optimizing the full ROI-wise projection weights, we adapt only a small set of subject-specific factors while keeping the shared tensor cores fixed. For each ROI r, the target-subject projection tensor is:

$$\mathcal{P}_r^{(s)} = \sum_{k=1}^K \left(\bar{\alpha}_{r,k} + \Delta\alpha_{r,k}^{(s)}\right) \mathcal{G}_{r,k}, \qquad b_r^{(s)} = \bar{b}_r + \Delta b_r^{(s)}, \tag{3}$$

where $\mathcal{G}_{r,k}$ are the shared tensor cores, $\bar{\alpha}_r \in \mathbb{R}^K$ denotes the subject-agnostic base coefficients, $\Delta\alpha_r^{(s)} \in \mathbb{R}^K$ is a subject-specific residual factor, $\bar{b}_r \in \mathbb{R}^D$ is a shared base bias, and $\Delta b_r^{(s)} \in \mathbb{R}^D$ is a subject-specific residual bias.

As shown in Fig. 3, during source-subject training, both shared parameters and source residuals are learned. For a target subject, zero-shot decoding sets $\Delta\alpha_r^{(s^*)} = \Delta b_r^{(s^*)} = 0$, while few-shot adaptation updates only these residual factors with the shared tensor cores frozen. This weight-decomposed adaptation preserves the shared functional geometry learned from source subjects while allowing flexible personalization to the target subject.

3.3 Brain Tensor-Transformer

ROI-wise Brain Tokenization. Given the aligned ROI-wise representation $h_i^{(s)}$, we next convert each ROI feature into a brain token for Transformer-based decoding. Each aligned ROI feature

$h_{i,r}^{(s)} \in \mathbb{R}^D$ is projected into a token embedding space of dimension d_t:

$$t_{i,r}^{(s)} = W_t h_{i,r}^{(s)} + e_r + e_s, \tag{4}$$

where $W_t \in \mathbb{R}^{d_t \times D}$ is a learnable token projection matrix, and $e_r, e_s \in \mathbb{R}^{d_t}$ are ROI and subject embeddings. For unseen target subjects, the subject embedding is initialized as the shared mean embedding and kept frozen in zero-shot decoding. The token sequence is denoted as $T_i^{(s)} = [\, t_{i,1}^{(s)}; t_{i,2}^{(s)}; \ldots; t_{i,R}^{(s)} \,] \in \mathbb{R}^{R \times d_t}$.

This ROI-wise tokenization assigns one token to each ROI, thereby enabling the subsequent Transformer to capture long-range interactions across distributed cortical areas. In addition, the token sequence remains naturally compatible with the ROI-based alignment module, since each token is constructed directly from an aligned functional feature rather than from raw voxel measurements.

Tensorized Transformer Blocks. We model inter-ROI dependencies using a stack of tensorized Transformer blocks, our design is motivated by the observation that interactions among brain tokens are highly structured, involving ROI identity, aligned functional content, and context-dependent dependencies across cortical regions. A standard Transformer parameterizes these interactions with large dense projection matrices, which are prone to overfitting in cross-subject neural decoding. To obtain a more compact and transferable architecture, we generate the major projection weights from shared tensor bases with lightweight layer-specific coefficients.

Let $Z_i^{(0,s)} = T_i^{(s)}$, and let $Z_i^{(\ell,s)} \in \mathbb{R}^{R \times d_t}$ denote the hidden token representation at layer ℓ. Each block follows the standard pre-norm residual form:

$$\tilde{Z}_i^{(\ell,s)} = Z_i^{(\ell-1,s)} + \text{T-MHSA}\big(\text{LN}(Z_i^{(\ell-1,s)})\big), \;\; Z_i^{(\ell,s)} = \tilde{Z}_i^{(\ell,s)} + \text{T-FFN}\big(\text{LN}(\tilde{Z}_i^{(\ell,s)})\big), \tag{5}$$

where T-MHSA is tensorized multi-head self-attention, and T-FFN is a tensorized FFN.

For each layer ℓ, we reconstruct the corresponding weight tensors as:

$$\mathcal{W}_\star^{(\ell)} = \sum_{k=1}^{K_a} \gamma_{\star,k}^{(\ell)} \mathcal{B}_k^{\text{attn}}, \; \star \in \{Q, K, V, O\}; \qquad \mathcal{W}_j^{(\ell)} = \sum_{k=1}^{K_j} \gamma_{j,k}^{(\ell)} \mathcal{B}_k^{(j)}, \; j \in \{1, 2\}. \tag{6}$$

Here $\{\mathcal{B}_k^{\text{attn}}\}$ are shared attention bases, $\{\mathcal{B}_k^{(1)}\}$ and $\{\mathcal{B}_k^{(2)}\}$ are feed-forward bases, and γ denotes layer-specific coefficients. The reconstructed tensors are folded into matrices and used in standard self-attention and feed-forward operations.

First, this shared-factor tensorization substantially reduces the number of free parameters by replacing full projection matrices with a small set of shared tensor bases and lightweight coefficients. Second, because the shared tensor bases capture reusable interaction patterns among ROI-wise brain tokens, the resulting model is encouraged to learn stable cross-region structure.

Hierarchical Neural-to-Visual Mapping. Since visual perception is inherently hierarchical, ranging from low-level structures to high-level semantics, a single flat decoder is often insufficient to align neural representations with visual targets at different abstraction levels. We therefore adopt a hierarchical neural-to-visual mapping strategy, in which the decoded brain representation is progressively aligned to multi-level visual features.

Let $\{v_i^{(m)}\}_{m=1}^M$ denote multi-level visual representations extracted from stimulus y_i by a pretrained visual encoder. From the final Transformer output $Z_i^{(L,s)}$, we can obtain a global neural summary $g_i^{(s)} = \text{Pool}\big(Z_i^{(L,s)}\big)$, where $g_i^{(s)} \in \mathbb{R}^{d_t}$ is a global neural representation obtained by token pooling.

We predict each visual level using shared neural-to-visual coupling bases:

$$\hat{v}_i^{(m,s)} = \Phi_m(g_i^{(s)}) = \sum_{k=1}^{K_v} \eta_{m,k} \, \psi_k(g_i^{(s)}), \qquad m = 1, \ldots, M, \tag{7}$$

where $\{\psi_k(\cdot)\}_{k=1}^{K_v}$ are shared basis mappings that capture reusable neural-to-visual coupling patterns, and $\eta_m \in \mathbb{R}^{K_v}$ is a level-specific coefficient vector. This hierarchical formulation allows different visual levels to reuse common neural-to-visual coupling patterns, by combining hierarchical supervision with it, the proposed mapping module improves parameter efficiency while preserving the flexibility needed to predict visual representations at different semantic levels.

3.4 Training and Inference

We train the model on source subjects $\mathcal{S}_{\text{tr}}$ with hierarchical visual supervision, distribution-aware feature matching, and subject-invariant neural regularization. The overall objective is:

$$\mathcal{L} = \mathcal{L}_{\text{feat}} + \beta\mathcal{L}_{\text{dist}} + \mu\mathcal{L}_{\text{inv}}, \tag{8}$$

where hierarchical feature loss is $\mathcal{L}_{\text{feat}} = \sum_{s\in\mathcal{S}_{\text{tr}}} \sum_{i=1}^{N_s} \sum_{m=1}^{M} \lambda_m \, \ell\big(\hat{v}_i^{(m,s)}, v_i^{(m)}\big)$, λ_m is the weight for the m-th visual level, and $\ell(\cdot,\cdot)$ combines squared error and cosine distance to preserve both feature fidelity and semantic consistency. The distribution-aware alignment term is $\mathcal{L}_{\text{dist}} = \sum_{s\in\mathcal{S}_{\text{tr}}} \sum_{i=1}^{N_s} \sum_{m=1}^{M} \lambda_m \left\| \big(\hat{v}_i^{(m,s)} - v_i^{(m)}\big)\Sigma_m^{-1/2} \right\|_2^2$, Σ_m is the empirical covariance of the m-th level visual features from the training set. The invariant loss is defined as $\mathcal{L}_{\text{inv}} = \sum_i \sum_{(s_a,s_b)\in\mathcal{P}_i} \left\| g_i^{(s_a)} - g_i^{(s_b)} \right\|_2^2$, $\mathcal{P}_i$ denotes the set of subject pairs viewing stimulus y_i.

At inference, we evaluate zero/few-shot cross-subject decoding and follow the diffusion reconstruction protocol of MindEye2 [11], with additional details in *Appendix.* **??**.

4 Experiments

4.1 Experimental Setup

Dataset. We evaluate on the Natural Scenes Dataset (NSD) [74], a large-scale benchmark containing high-resolution 7-T fMRI recordings. The dataset consists of whole-brain responses from subjects viewing 73,000 natural scenes from MS-COCO [75]. Following established protocols [46], we utilize data from the four subjects (subj01, 02, 05, 07) who completed the full experimental sessions, viewing 10,000 images each with three repetitions. The training set for each subject comprises 8,859 image stimuli and 24,980 fMRI trials, whereas the test set comprises 982 image stimuli and 2,770 fMRI trials. The test set stimuli are shared across all subjects, whereas the training set stimuli are subject-specific and non-overlapping. For preprocessing, we use the provided 1.8-mm native volume space data and apply the `Nsdgeneral` mask to isolate voxels in the visual cortex ($\sim$ 40k voxels).

Implementation. All experiments are conducted on a single NVIDIA A800 GPU with bf16 mixed-precision training. We train the model with the AdamW optimizer for 60 epochs using a batch size of 64. The initial learning rate is set to 1×10^{-4} with a weight decay of 5×10^{-2}, and is scheduled with cosine annealing. Gradient clipping with a maximum norm of 1.0 is applied to stabilize optimization. During source-subject training, the shared tensor cores/bases and source-subject residual factors are jointly optimized. For few-shot calibration on a new subject, we use a smaller batch size of 16 and train for 20 epochs with a learning rate of 5×10^{-5}. To improve numerical stability in the distribution-aware loss, the covariance statistics are estimated on the training set and small eigenvalues are clipped when computing the inverse square root.

Metrics. Following standard protocols, we assess reconstruction quality using eight metrics. For low-level, we use Pixel-wise Correlation (`PixCorr`) and Structural Similarity (`SSIM`) [77]. For semantic evaluation, we compare features extracted from the 2/5-th layers of AlexNet (`Alex(2)`, `Alex(5)`) [78], Inception-v3 (`Incep`) [79], and `CLIP` [44]. We also report the average correlation distance using EfficientNet-B1 (`Eff`) [80] and SwAV-ResNet50 (`SwAV`) [81]. We further evaluate model complexity using the parameters (`Para.`) and `FLOPs` in new subject adaptation.

4.2 Comparative Experimental Results

Quantitative Results. We report results averaged over the evaluated subjects. As shown in Tab. 1, BrainTC achieves the best performance in the few-shot setting using only 1 hour of target-subject data. Compared with the strongest few-shot baseline, BrainIT, BrainTC improves low-level reconstruction quality, raising PixCorr from 0.331 to 0.340 and SSIM from 0.473 to 0.481. High-level perceptual and semantic metrics also show consistent gains: AlexNet(2) increases from 97.1% to 97.9%, CLIP similarity improves from 93.0% to 93.4%, and both Eff and SwAV distances are reduced. These results indicate that BrainTC better preserves visual structure while producing semantically faithful reconstructions under extremely limited calibration data. A key advantage of BrainTC is efficiency. Despite achieving the best reconstruction quality, it requires only 5.44M adaptation parameters and 0.03G FLOPs in the few-shot setting, substantially lower than all competing methods. In particular,

Table 1: Quantitative Performance. We compared the performance of BrainTC across different approaches (average results of subjects 1, 2, 5, and 7 in NSD). Fully fine-tuned refers to using 40 hours data from test subject, few-shot refers to using only 1 hour data, and zero-shot refers to not using any test data. The Adaptation computational cost refers to the fine-tuning computational cost.

Method		Low-Level				High-Level				Adapt. Cost	
		PixCorr↑	SSIM↑	Alex(2)↑	Alex(5)↑	Incep↑	CLIP↑	Eff↓	SwAV↓	Para.↓	FLOPs↓
▼*Fully fine-tuned*											
Takagi. [9]	[CVPR'23]	0.246	0.410	78.9%	85.6%	83.8%	82.1%	0.811	0.504	1.7B	18.6G
MindEye [46]	[NeurIPS'23]	0.319	0.360	92.8%	96.9%	94.6%	93.3%	0.648	0.377	2.9B	32.4G
UMBRAE [12]	[ECCV'24]	0.283	0.341	95.5%	97.0%	91.7%	93.5%	0.700	0.393	2.4B	26.8G
NeuroPictor [13]	[ECCV'24]	0.229	0.375	96.5%	98.4%	94.5%	93.3%	0.639	0.350	2.0B	12G
NeuroVLA [76]	[NeurIPS'24]	0.265	0.357	93.1%	97.1%	96.8%	97.5%	0.633	0.321	674.60M	7.5G
MindBridge [17]	[CVPR'24]	0.151	0.263	87.7%	95.5%	92.4%	94.7%	0.712	0.418	118.55M	1.3G
MindEye2 [11]	[ICML'24]	0.322	0.431	96.1%	98.6%	95.4%	93.0%	0.619	0.344	2.2B	528G
MindTuner [14]	[AAAI'25]	0.322	0.421	95.8%	98.8%	95.6%	93.8%	0.612	0.340	76.71M	18.4G
▼ *Few-shot*											
MindEye2 [11] (1 h)		0.195	0.419	84.2%	90.6%	81.2%	79.2%	0.810	0.468	2.2B	13.2G
MindTuner [14] (1 h)		0.224	0.420	87.8%	93.6%	84.8%	83.5%	0.780	0.440	76.71M	0.46G
MindAligner [16] (1 h)	[ICML'25]	0.206	0.414	85.6%	91.6%	83.0%	81.2%	0.802	0.463	139.23M	0.84G
BrainIT [15] (1 h)	[ICLR'26]	0.331	0.473	97.1%	98.6%	94.4%	93.0%	0.648	0.370	44.82M	0.26G
BrainTC		**0.340**	**0.481**	**97.9%**	**98.8%**	**94.6%**	**93.4%**	**0.625**	**0.362**	**5.44M**	**0.03G**
▼*Zero-shot*											
ZEBRA [18]	[NeurIPS'25]	0.131	0.375	74.6%	81.2%	72.2%	71.5%	0.837	0.506	N/A	N/A
BrainTC		**0.178**	**0.392**	**78.8%**	**83.6%**	**77.4%**	**76.8%**	**0.813**	**0.489**	N/A	N/A

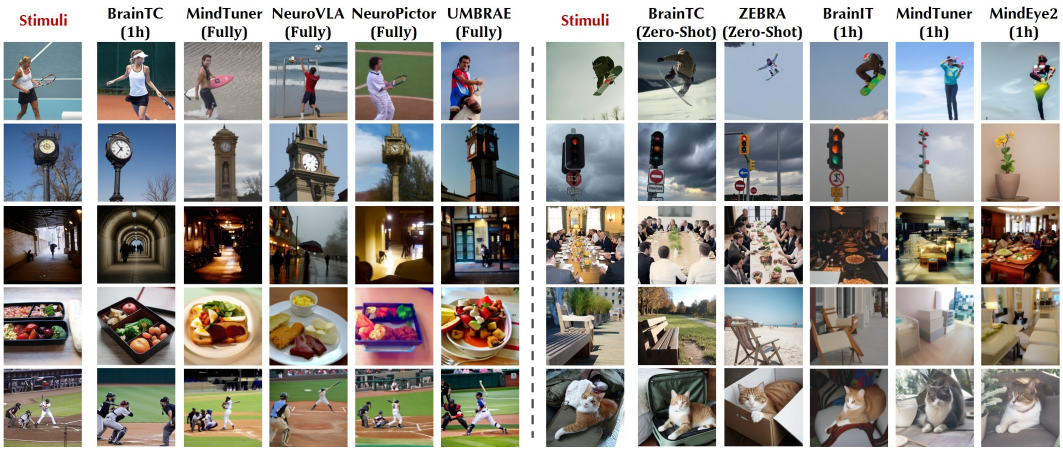


Figure 4: Qualitative Comparison. We present BrainTC's 1-hour / zero-shot reconstructions alongside methods that support fully fine-tuned / zero-shot / 1-hour reconstructions.

compared with BrainIT, BrainTC reduces adaptation parameters by about $8\times$ and computation by nearly $9\times$, yet still achieves better performance. This demonstrates that the proposed weight-decomposed tensor adaptation provides a superior performance-efficiency trade-off over conventional adaptation schemes. In the more challenging zero-shot setting, BrainTC also surpasses the recent cross-subject method ZEBRA on every metric, improving PixCorr from 0.131 to 0.178 and CLIP from 71.5% to 76.8%. These gains verify that the shared tensor cores learn transferable functional structure across subjects, enabling robust decoding without target-subject calibration.

Qualitative Results. Fig. 4 shows representative reconstructions. In few-shot, BrainTC produces images that are visually closer to the stimuli than fully fine-tuned baselines, with more accurate object shape, scene layout, and color composition. This is particularly evident in examples such as the *tennis player, clock tower, and baseball scenes*. In zero-shot, BrainTC also yields more coherent and semantically faithful reconstructions, while remaining competitive with few-shot methods. It more reliably captures the global scene structure in challenging samples. Overall, the qualitative results are

Table 2: Results of ablation experiments on the components of BrainTC on Subject 1. The first eight metrics are reported under the zero-shot setting, while the last two show few-shot performance.

Setting	Low-Level				High-Level				Few-shot	
	PixCorr↑	SSIM↑	Alex(2)↑	Alex(5)↑	Incep↑	CLIP↑	Eff↓	SwAV↓	PixCorr↑	CLIP↑
▼ *Ablation on Functional Alignment Design*										
w/o Alignment	0.129	0.372	73.8%	79.5%	72.1%	71.5%	0.833	0.503	0.264	83.0%
w/o Tensor decomposition	0.145	0.381	76.0%	81.8%	74.4%	73.8%	0.823	0.495	0.283	86.6%
w/o Shared tensor cores	0.140	0.378	75.2%	81.0%	73.6%	73.0%	0.827	0.498	0.271	84.7%
w/o Subject residual factors	0.173	0.393	78.4%	84.0%	77.0%	76.6%	0.808	0.482	0.279	86.1%
▼ *Ablation on Tensorized Transformer Blocks*										
Standard Transformer	0.164	0.389	77.6%	83.1%	76.1%	75.6%	0.814	0.488	0.364	93.4%
LoRA-style attention	0.171	0.392	78.3%	83.8%	76.9%	76.4%	0.810	0.484	0.381	93.0%
$only$ Tensorized attention	0.178	0.395	79.0%	84.5%	77.7%	77.3%	0.805	0.480	0.391	94.5%
$only$ Tensorized FFN	0.177	0.394	78.9%	84.4%	77.5%	77.1%	0.806	0.481	0.389	94.4%
Full Model	**0.187**	**0.399**	**79.6%**	**85.1%**	**78.5%**	**78.2%**	**0.799**	**0.475**	**0.405**	**95.1%**

consistent with the quantitative gains, showing that BrainTC improves both structural fidelity and semantic consistency under limited or even no target-subject calibration.

4.3 Ablation & Discussion

Components Ablation Analysis. We conduct ablations on Subject 1, with results summarized in Tab. 2. Functional alignment is the most critical component. Removing it causes the largest drop, decreasing zero-shot PixCorr from 0.187 to 0.129 and CLIP from 78.2% to 71.5%, while substantially degrading few-shot performance. This confirms that explicit ROI-wise alignment is essential for reducing inter-subject mismatch before decoding. Replacing tensorization with a non-decomposed version lowers zero-shot and few-shot performance, showing that tensor decomposition provides an effective, data-efficient representation of subject-specific projections. Removing shared tensor cores or subject residual factors also hurts performance. These results indicate shared cores capture transferable functional structure, whereas lightweight residual factors enable subject-specific adaptation. We analyze tensorized Transformer blocks. A standard Transformer performs worse than all tensorized variants, and LoRA-style attention partially closes the gap. Tensorizing attention or FFN already improves results, while tensorizing both achieves the best performance. The visual comparisons in Fig. 5 (A) are consistent with these trends: the full model produces sharper objects, more coherent layouts, and fewer semantic confusions than ablated variants.

Pareto Frontier Analysis. We further analyze the trade-off between adaptation efficiency and reconstruction quality by plotting Pareto frontiers of few-shot CLIP versus adaptation FLOPs in Fig. 5 (B). BrainTC clearly dominates prior adaptation methods, including BrainIT, MindTuner, and MindAligner, as its frontier lies consistently in the upper-left region. This means that BrainTC achieves better semantic fidelity with substantially lower adaptation cost. The same advantage holds when comparing different architectural choices within our framework: BrainTC surpasses both the LoRA-style variant and the standard Transformer over the entire compute range. These results demonstrate that the proposed tensorized adaptation is not only more effective, but also significantly more efficient, yielding the best practical trade-off for few-shot cross-subject decoding.

Tensor Design Analysis. We study tensor capacity by varying K and K_a, which determine basis size in functional alignment and tensorized attention. As shown in Tab. 3, increasing capacity from small settings improves semantic metrics, but gains are not monotonic. The balanced configuration $(4, 4)$ gives the best results, achieving the highest Inception and CLIP scores and the lowest Eff and SwAV distances with only 5.44M parameters and 0.03G FLOPs. By contrast, a larger design such as $(8, 8)$ increases cost without benefit. This confirms that BrainTC benefits from structured tensorization rather than simply scaling model capacity.

Table 3: Design analysis of K and K_a tensor capacity allocation in BrainTC.

K	K_a	Incep ↑	CLIP ↑	Eff ↓	SwAV ↓	Para.	FLOPs
2	2	92.9%	91.5%	0.657	0.381	3.21M	0.02G
8	2	93.8%	92.4%	0.641	0.372	6.48M	0.03G
2	8	93.4%	92.0%	0.648	0.376	4.87M	0.04G
4	4	**94.6%**	**93.4%**	**0.625**	**0.362**	**5.44M**	**0.03G**
8	8	94.4%	93.1%	0.628	0.364	8.96M	0.05G

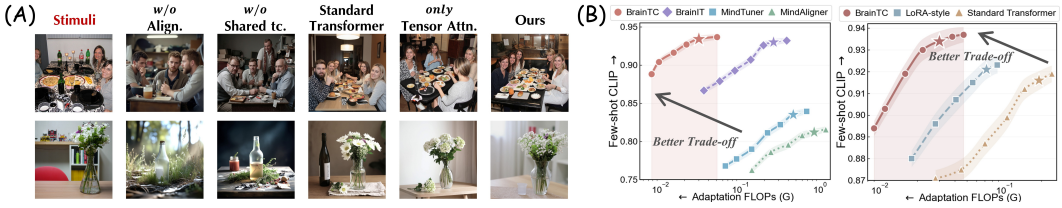


Figure 5: (A) Visualization results of ablation of different components in BrainTC. (B) Pareto frontier dominance of the BrainTC's model and tensor block architecture.

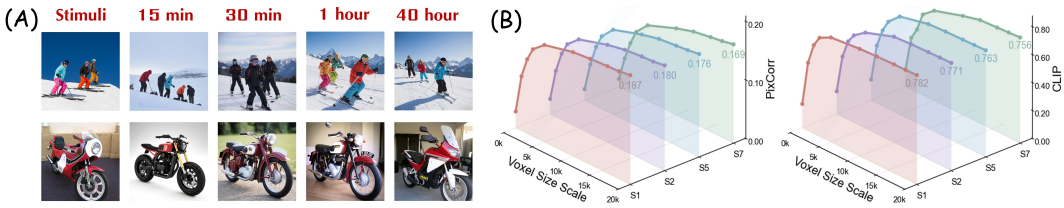


Figure 6: (A) Qualitative reconstruction results fine-tuned for Subject 1 with limited data. (B) Number of voxels scaling on Subject 1/2/5/7 in BrainTC training.

Voxel Scale. We analyze how voxel scale affects BrainTC. As shown in Fig. 6 (A): reconstructions become progressively sharper and more semantically faithful from 15 min to 1 hour, with clearer object boundaries, more accurate shapes, and better color composition. Notably, the 1-hour setting already approaches the visual quality of the 40-hour model in several examples, indicating that BrainTC can exploit larger voxel scales efficiently and achieve strong reconstruction quality even under limited calibration data. The quantitative results in Fig. 6 (B) further support this observation, performance consistently improves as more voxels are included, with PixCorr and CLIP rising across all subjects before saturating at larger scales. This trend suggests that increasing voxel coverage provides richer visual information, while BrainTC remains stable without obvious overfitting.

5 Conclusion

In this work, we presented BrainTC, an efficient framework for cross-subject brain visual decoding built on shared tensor decomposition. By decomposing ROI-wise alignment into shared tensor cores and lightweight residual subject factors, BrainTC provides a principled way to capture both transferable functional structure and individual variability. Combined with a tensorized Transformer and hierarchical neural-to-visual mapping, it supports both zero-shot generalization to new subjects and efficient few-shot adaptation. Through extensive experimentation and discussion, BrainTC achieved superior performance, showing that competitive cross-subject decoding does not require heavy retraining or over-parameterized decoders. Our findings demonstrate that structured tensor parameterization offers a practical route to scalable, deployment-friendly BCI systems.

Acknowledgments

This work was supported by he Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China under Grant JYB2025XDXM504, the Guangdong Major Project of Basic and Applied Basic Research under Grant No. 2023B0303000009, and the National Natural Science Foundation of China NO. 32541016, 62436005.

References

[1] J. K. O'regan, "Solving the" real" mysteries of visual perception: the world as an outside memory.," *Canadian Journal of Psychology/Revue canadienne de psychologie*, vol. 46, no. 3, p. 461, 1992. 1

[2] N. Wade and M. Swanston, *Visual perception: An introduction*. Psychology Press, 2013. 1

[3] Y. Wang, S. Wang, Y. Zhang, C. Du, C. Fan, D. Li, H. Zhou, H. Zhang, J. Li, Q. Liu, *et al.*, "Representation, alignment, and generation: A comprehensive survey of foundation models for non-invasive brain decoding," *bioRxiv*, pp. 2025–11, 2025. 1, 3

[4] P. Liu, G. Dong, D. Guo, K. Li, F. Li, X. Yang, M. Wang, and X. Ying, "A survey on fmri-based brain decoding for reconstructing multimodal stimuli," *arXiv preprint arXiv:2503.15978*, 2025. 1, 3

[5] S. R. Oota, Z. Chen, M. Gupta, R. S. Bapi, G. Jobard, F. Alexandre, and X. Hinaut, "Deep neural networks and brain alignment: Brain encoding and decoding (survey)," *arXiv preprint arXiv:2307.10246*, 2023. 1

[6] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, "Neurophysiological investigation of the basis of the fmri signal," *nature*, vol. 412, no. 6843, pp. 150–157, 2001. 1

[7] M. Bamdad, H. Zarshenas, and M. A. Auais, "Application of bci systems in neurorehabilitation: a scoping review," *Disability and Rehabilitation: Assistive Technology*, vol. 10, no. 5, pp. 355–364, 2015. 1

[8] M. Fleury, P. Figueiredo, A. Vourvopoulos, and A. Lécuyer, "Two is better? combining eeg and fmri for bci and neurofeedback: a systematic review," *Journal of neural engineering*, vol. 20, no. 5, p. 051003, 2023. 1

[9] Y. Takagi and S. 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. 1, 3, 8

[10] F. Ozcelik and R. VanRullen, "Natural scene reconstruction from fmri signals using generative latent diffusion," *Scientific Reports*, vol. 13, no. 1, p. 15666, 2023. 1, 3

[11] P. S. Scotti, M. Tripathy, C. K. T. Villanueva, R. Kneeland, T. Chen, A. Narang, C. Santhirasegaran, J. Xu, T. Naselaris, K. A. Norman, and T. M. Abraham, "Mindeye2: shared-subject models enable fmri-to-image with 1 hour of data," in *Proceedings of the 41st International Conference on Machine Learning*, ICML'24, JMLR.org, 2024. 2, 3, 7, 8, 17

[12] W. Xia, R. de Charette, C. Öztireli, and J.-H. Xue, "UMBRAE: Unified Multimodal Brain Decoding," July 2024. 2, 3, 8

[13] J. Huo, Y. Wang, Y. Wang, X. Qian, C. Li, Y. Fu, and J. Feng, "Neuropictor: Refining fmri-to-image reconstruction via multi-individual pretraining and multi-level modulation," in *European Conference on Computer Vision*, pp. 56–73, Springer, 2024. 2, 3, 8

[14] Z. Gong, Q. Zhang, G. Bao, L. Zhu, R. Xu, K. Liu, L. Hu, and D. Miao, "Mindtuner: Cross-subject visual decoding with visual fingerprint and semantic correction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, pp. 14247–14255, 2025. 2, 3, 8, 17

[15] R. Beliy, A. Zalcher, J. Kogman, N. Wasserman, and M. Irani, "Brain-it: Image reconstruction from fmri via brain-interaction transformer," *arXiv preprint arXiv:2510.25976*, 2025. 2, 3, 8, 17

[16] Y. Dai, Z. Yao, C. Song, Q. Zheng, W. Mai, K. Peng, S. Lu, W. Ouyang, J. Yang, and J. Wu, "Mindaligner: Explicit brain functional alignment for cross-subject visual decoding from limited fmri data," in *International Conference on Machine Learning*, pp. 12214–12228, PMLR, 2025. 2, 3, 8, 17

[17] S. Wang, S. Liu, Z. Tan, and X. Wang, "MindBridge: A Cross-Subject Brain Decoding Framework," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Seattle, WA, USA), pp. 11333–11342, IEEE, June 2024. 2, 3, 8

[18] H. Wang, J. Lu, H. Li, and X. Li, "Zebra: Towards zero-shot cross-subject generalization for universal brain visual decoding," in *The Thirty-ninth Annual Conference on Neural Information Processing Systems*. 2, 3, 8

[19] H. Jing, P. Yang, D. Jiang, Z. Liu, N. Zheng, and Y. Ma, "Evoke: Efficient and high-fidelity eeg-to-video reconstruction via decoupling implicit neural representation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, 2026. 2

[20] R. Sitaram, A. Caria, R. Veit, T. Gaber, G. Rota, A. Kuebler, and N. Birbaumer, "Fmri brain-computer interface: A tool for neuroscientific research and treatment," *Computational intelligence and neuroscience*, vol. 2007, no. 1, p. 025487, 2007. 2

[21] S. Saha, K. A. Mamun, K. Ahmed, R. Mostafa, G. R. Naik, S. Darvishi, A. H. Khandoker, and M. Baumert, "Progress in brain computer interface: Challenges and opportunities," *Frontiers in systems neuroscience*, vol. 15, p. 578875, 2021. 2

[22] X. Song, L. Meng, Q. Shi, and H. Lu, "Learning tensor-based features for whole-brain fmri classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 613–620, Springer, 2015. 2

[23] G. Ma, L. He, C.-T. Lu, P. S. Yu, L. Shen, and A. B. Ragin, "Spatio-temporal tensor analysis for whole-brain fmri classification," in *Proceedings of the 2016 siam international conference on data mining*, pp. 819–827, SIAM, 2016. 2

[24] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, M. Attariyan, and S. Gelly, "Parameter-efficient transfer learning for nlp," in *International conference on machine learning*, pp. 2790–2799, PMLR, 2019. 2

[25] B. Lester, R. Al-Rfou, and N. Constant, "The power of scale for parameter-efficient prompt tuning," in *Proceedings of the 2021 conference on empirical methods in natural language processing*, pp. 3045–3059, 2021. 2

[26] X. L. Li and P. Liang, "Prefix-tuning: Optimizing continuous prompts for generation," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 4582–4597, 2021. 2

[27] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, *et al.*, "Lora: Low-rank adaptation of large language models.," *Iclr*, vol. 1, no. 2, p. 3, 2022. 2

[28] S.-Y. Liu, C.-Y. Wang, H. Yin, P. Molchanov, Y.-C. F. Wang, K.-T. Cheng, and M.-H. Chen, "Dora: Weight-decomposed low-rank adaptation," in *Forty-first International Conference on Machine Learning*, 2024. 2

[29] I. V. Oseledets, "Tensor-train decomposition," *SIAM Journal on Scientific Computing*, vol. 33, no. 5, pp. 2295–2317, 2011. 2, 3

[30] M. Wang, Y. Pan, Z. Xu, G. Li, X. Yang, D. Mandic, and A. Cichocki, "Tensor networks meet neural networks: A survey and future perspectives," *arXiv preprint arXiv:2302.09019*, 2023. 2, 3

[31] T. Garipov, D. Podoprikhin, A. Novikov, and D. Vetrov, "Ultimate tensorization: compressing convolutional and fc layers alike," *arXiv preprint arXiv:1611.03214*, 2016. 2, 3

[32] W. Wang, Y. Sun, B. Eriksson, W. Wang, and V. Aggarwal, "Wide compression: Tensor ring nets," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 9329–9338, 2018. 2, 3

[33] J. Su, W. Byeon, J. Kossaifi, F. Huang, J. Kautz, and A. Anandkumar, "Convolutional tensor-train lstm for spatio-temporal learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 13714–13726, 2020. 2, 3

[34] Y. Pan, J. Xu, M. Wang, J. Ye, F. Wang, K. Bai, and Z. Xu, "Compressing recurrent neural networks with tensor ring for action recognition," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 4683–4690, 2019. 2, 3

[35] X. Ma, P. Zhang, S. Zhang, N. Duan, Y. Hou, M. Zhou, and D. Song, "A tensorized transformer for language modeling," *Advances in neural information processing systems*, vol. 32, 2019. 2, 4

[36] Y. Zhang, Y. Liu, H. Yuan, Z. Qin, Y. Yuan, Q. Gu, and A. C.-C. Yao, "Tensor product attention is all you need," *arXiv preprint arXiv:2501.06425*, 2025. 2, 4

[37] S. Nishimoto, A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu, and J. L. Gallant, "Reconstructing visual experiences from brain activity evoked by natural movies," *Current biology*, vol. 21, no. 19, pp. 1641–1646, 2011. 3

[38] E. J. Allen, G. St-Yves, Y. Wu, J. L. Breedlove, J. S. Prince, L. T. Dowdle, M. Nau, B. Caron, F. Pestilli, I. Charest, *et al.*, "A massive 7t fmri dataset to bridge cognitive neuroscience and artificial intelligence," *Nature neuroscience*, vol. 25, no. 1, pp. 116–126, 2022. 3

[39] N. Chang, J. A. Pyles, A. Marcus, A. Gupta, M. J. Tarr, and E. M. Aminoff, "BOLD5000, a public fMRI dataset while viewing 5000 visual images," *Scientific Data*, vol. 6, p. 49, May 2019. 3

[40] T. Horikawa and Y. Kamitani, "Generic decoding of seen and imagined objects using hierarchical visual features," *Nature Communications*, vol. 8, no. 1, p. 15037, 2017. 3

[41] R. Beliy, G. Gaziv, A. Hoogi, F. Strappini, T. Golan, and M. Irani, "From voxels to pixels and back: Self-supervision in natural-image reconstruction from fmri," *Advances in Neural Information Processing Systems*, vol. 32, 2019. 3

[42] Z. Chen, J. Qing, T. Xiang, W. L. Yue, and J. H. Zhou, "Seeing beyond the brain: Conditional diffusion model with sparse masked modeling for vision decoding," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22710–22720, 2023. 3, 15, 23

[43] J. Chen, Y. Qi, and G. Pan, "Rethinking visual reconstruction: experience-based content completion guided by visual cues," in *Proceedings of the 40th International Conference on Machine Learning*, pp. 4856–4866, 2023. 3, 23

[44] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*, pp. 8748–8763, PMLR, 2021. 3, 7, 15

[45] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022. 3

[46] P. Scotti, A. Banerjee, J. Goode, S. Shabalin, A. Nguyen, A. Dempster, N. Verlinde, E. Yundler, D. Weisberg, K. Norman, *et al.*, "Reconstructing the mind's eye: fmri-to-image with contrastive learning and diffusion priors," *Advances in Neural Information Processing Systems*, vol. 36, pp. 24705–24728, 2023. 3, 7, 8, 15

[47] B. Zeng, S. Li, X. Liu, S. Gao, X. Jiang, X. Tang, Y. Hu, J. Liu, and B. Zhang, "Controllable Mind Visual Diffusion Model," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 6935–6943, Mar. 2024. 3

[48] H. Jing, P. Yang, R. Zhang, Z. Liu, Y. Liu, X. Bai, Y. Ma, and N. Zheng, "Follow-your-mind: Towards inversion-free brain-driven visual context synthesis and editing," in *European Conference on Computer Vision*, pp. 552–572, Springer, 2026. 3

[49] H. Jing, Y. Ma, W. Gao, D. Jiang, S. Huang, and N. Zheng, "Mind-vad: Brain-inspired fmri-to-video precise reconstruction via cross-modal autoregressive diffusion," *IEEE Transactions on Circuits and Systems for Video Technology*, 2026. 3

[50] H. Jing, D. Jiang, J. Wang, J. Jia, Y. Li, Y. Ma, and N. Zheng, "More natural, more real: Object-aware gaussian splatting for 3d visual decoding from human brain," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19033–19044, 2026. 3

[51] H. Jing, Y. Ma, P. Yang, H. Hua, and N. Zheng, "Pinpointing visual content: Disentangled features in multimodal model for eeg representation learning and decoding," *Knowledge-Based Systems*, p. 114212, 2025. 3

[52] Y. Liu, Y. Ma, G. Zhu, H. Jing, and N. Zheng, "See through their minds: Learning transferable brain decoding models from cross-subject fmri," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, pp. 5730–5738, 2025. 3, 23

[53] H. Jing, D. Jiang, Y. Ma, H. Hua, B. Huang, and N. Zheng, "Beyond brain decoding: Visual-semantic reconstructions to mental creation extension based on fmri," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 19258–19268, 2025. 3

[54] Y. Wei, Y. Zhang, X. Xiao, T. Wang, X. Wang, and V. D. Calhoun, "More-brain: Routed mixture of experts for interpretable and generalizable cross-subject fmri visual decoding," *arXiv preprint arXiv:2505.15946*, 2025. 3

[55] M. Yu, M. Nan, H. Adeli, J. S. Prince, J. A. Pyles, L. Wehbe, M. M. Henderson, M. J. Tarr, and A. F. Luo, "Meta-learning an in-context transformer model of human higher visual cortex," *arXiv preprint arXiv:2505.15813*, 2025. 3

[56] W. Mai, J. Wu, Y. Zhu, Z. Yao, D. Zhou, A. F. Luo, Q. Zheng, W. Ouyang, and C. Song, "Synbrain: Enhancing visual-to-fmri synthesis via probabilistic representation learning," *arXiv preprint arXiv:2508.10298*, 2025. 3

[57] H. Jing, Y. Ma, P. Yang, H. Li, S. Huang, B. Chen, and N. Zheng, "Damind: Zero-shot visual cross-domain alignment and representation for eeg decoding," *IEEE Transactions on Image Processing*, 2026. 3

[58] S. Huang, H. Luo, H. Jing, Q. Zhang, L. Chang, Y. Feng, X. Lin, C. Qin, H. Chen, S. Jia, *et al.*, "Need: Cross-subject and cross-task generalization for video and image reconstruction from eeg signals," in *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. 3

[59] Z. Gong, Q. Zhang, G. Bao, L. Zhu, Y. Zhang, K. Liu, L. Hu, and D. Miao, "Lite-mind: Towards efficient and robust brain representation learning," in *ACM Multimedia 2024*, 2024. 3

[60] S. Zhang, D. Liang, and Y. Zhou, "Neuroswift: A lightweight cross-subject framework for fmri visual reconstruction of complex scenes," in *Proceedings of the 7th ACM International Conference on Multimedia in Asia*, pp. 1–8, 2025. 3

[61] Y. Cao, X. Liu, F. Zhang, S. Deng, Z. Ren, and J. Wang, "Beyond static allocation: Dynamic sensitivity-aware fine-tuning for vision transformers," in *Forty-third International Conference on Machine Learning*. 3

[62] Z. Yuan, Y. Shang, Y. Song, D. Yang, Q. Wu, Y. Yan, and G. Sun, "Asvd: Activation-aware singular value decomposition for compressing large language models," *arXiv preprint arXiv:2312.05821*, 2023. 3

[63] X. Huang, Y.-L. Huang, and Z. Wen, "Sola: Leveraging soft activation sparsity and low-rank decomposition for large language model compression," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, pp. 17494–17502, 2025. 3

[64] V. Lebedev, Y. Ganin, M. Rakhuba, I. Oseledets, and V. Lempitsky, "Speeding-up convolutional neural networks using fine-tuned cp-decomposition," *arXiv preprint arXiv:1412.6553*, 2014. 3

[65] Y.-D. Kim, E. Park, S. Yoo, T. Choi, L. Yang, and D. Shin, "Compression of deep convolutional neural networks for fast and low power mobile applications," *arXiv preprint arXiv:1511.06530*, 2015. 3

[66] Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki, "Tensor ring decomposition," *arXiv preprint arXiv:1606.05535*, 2016. 3

[67] X. Liu and K. K. Parhi, "Tensor decomposition for model reduction in neural networks: A review [feature]," *IEEE Circuits and Systems Magazine*, vol. 23, no. 2, pp. 8–28, 2023. 3

[68] X. Wang, Y. Zheng, Z. Wan, and M. Zhang, "Svd-llm: Truncation-aware singular value decomposition for large language model compression," *arXiv preprint arXiv:2403.07378*, 2024. 3

[69] Y. Yang, D. Krompass, and V. Tresp, "Tensor-train recurrent neural networks for video classification," in *International conference on machine learning*, pp. 3891–3900, PMLR, 2017. 3

[70] R. Mehta, R. Chakraborty, Y. Xiong, and V. Singh, "Scaling recurrent models via orthogonal approximations in tensor trains," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10571–10579, 2019. 3

[71] X. Yuebin, Y. WANG, X. Peng, H. Zang, C. Minghao, P. Xia, and Z. Wen, "Td-moe: Tensor decomposition for moe models," in *The Fourteenth International Conference on Learning Representations*. 4

[72] A. Cichocki, "Tensor decompositions: new concepts in brain data analysis?," *Journal of the Society of Instrument and Control Engineers*, vol. 50, no. 7, pp. 507–516, 2011. 4

[73] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010. 4

[74] E. J. Allen, G. St-Yves, Y. Wu, J. L. Breedlove, J. S. Prince, L. T. Dowdle, M. Nau, B. Caron, F. Pestilli, I. Charest, *et al.*, "A massive 7t fmri dataset to bridge cognitive neuroscience and artificial intelligence," *Nature neuroscience*, vol. 25, no. 1, pp. 116–126, 2022. 7, 15

[75] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pp. 740–755, Springer, 2014. 7, 15

[76] G. Shen, D. Zhao, X. He, L. Feng, Y. Dong, J. Wang, Q. Zhang, and Y. Zeng, "Neuro-vision to language: Enhancing brain recording-based visual reconstruction and language interaction," *Advances in Neural Information Processing Systems*, vol. 37, pp. 98083–98110, 2024. 8

[77] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004. 7, 15

[78] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012. 7, 15

[79] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016. 7, 15

[80] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, pp. 6105–6114, PMLR, 2019. 7, 15

[81] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, "Unsupervised learning of visual features by contrasting cluster assignments," *Advances in neural information processing systems*, vol. 33, pp. 9912–9924, 2020. 7, 15

[82] X. Zhang, R. Quan, W. Wang, and Y. Yang, "Moving beyond diffusion: Hierarchy-to-hierarchy autoregression for fmri-to-image reconstruction," *arXiv preprint arXiv:2510.22335*, 2025. 17, 23

[83] T. Horikawa and Y. Kamitani, "Generic decoding of seen and imagined objects using hierarchical visual features," *Nature communications*, vol. 8, no. 1, p. 15037, 2017. 22, 23

[84] M. N. Hebart, O. Contier, L. Teichmann, A. H. Rockter, C. Y. Zheng, A. Kidder, A. Corriveau, M. Vaziri-Pashkam, and C. I. Baker, "Things-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior," *Elife*, vol. 12, p. e82580, 2023. 23

[85] B. Zeng, S. Li, X. Liu, S. Gao, X. Jiang, X. Tang, Y. Hu, J. Liu, and B. Zhang, "Controllable mind visual diffusion model," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 38, pp. 6935–6943, 2024. 23

[86] D. Li, H. Qin, M. Wu, C. Wei, and Q. Liu, "Brainflora: Uncovering brain concept representation via multimodal neural embeddings," in *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 5577–5586, 2025. 23

A More Implementation Details

A.1 Dataset & Preprocessing

Dataset. We use the Natural Scenes Dataset (NSD) [74], a large-scale public fMRI-image dataset for studying visual representations in the human brain. NSD contains high-resolution 7T fMRI recordings from eight participants collected during a natural image viewing experiment. During data acquisition, subjects viewed images drawn from the MS-COCO dataset [75]. Each image was presented for 3 seconds, followed by a 1-second inter-stimulus interval. Functional images were acquired using whole-brain gradient-echo EPI at 1.8 mm isotropic resolution with a repetition time (TR) of 1.6 s.

Following common practice in fMRI-to-image reconstruction studies [46, 42], we focus on the four subjects who completed the full set of sessions, namely Subj-01, Subj-02, Subj-05, and Subj-07. Each subject viewed 10,000 unique images in total, including 9,000 training images and 1,000 test images. Since each image was repeated up to three times, the training set contains 24,980 single-trial fMRI samples, while the test set contains 2,770 trials before averaging. The test images are shared across subjects, whereas the training images are subject-specific. The visual stimuli were square-cropped natural images displayed at approximately $8.4^\circ \times 8.4^\circ$ of visual angle. For functional analysis, we use the preprocessed single-trial beta estimates released by NSD, which are obtained from a general linear model fit to the fMRI time series.

Preprocess. To focus on visual cortical activity relevant to image perception, we adopt the official `Nsdgeneral` region-of-interest mask provided by NSD. This mask spans a broad set of visual regions, ranging from early visual cortex to higher-order ventral visual areas. For the four subjects used in our experiments, the number of selected voxels is 15,724 for Subj-01, 14,278 for Subj-02, 13,039 for Subj-05, and 12,682 for Subj-07. These voxels are extracted from the 1.8 mm functional space and used as neural input to the model.

We use the NSD preprocessed functional data. The original preprocessing pipeline includes standard corrections such as temporal interpolation for slice-timing offsets and spatial interpolation for motion correction. Based on the released single-trial beta estimates, we further apply several preprocessing steps before model training. The voxels within the `Nsdgeneral` mask are further grouped according to the NSD visual ROI annotations. For each subject, we preserve the ROI identity and construct ROI-wise voxel vectors. Since the number of voxels within the same ROI may differ across subjects, each ROI vector is converted to a fixed length by padding or truncation, and a binary mask is used to exclude padded entries from the projection computation. After ROI extraction, the selected voxel responses are organized as ROI-wise vectors and used as the input to the functional alignment module. For notational compatibility with standard fMRI decoding formulations, the concatenation of all ROI responses can also be viewed as a flattened vector, and used as the fMRI input to the neural encoder, denoted as $x_i^{(s)} \in \mathbb{R}^V$.

A.2 Evaluation Metrics

We assess reconstruction quality using eight metrics spanning low-level fidelity and high-level semantic consistency. For low-level fidelity, we use Pixel-wise Correlation (`PixCorr`) and Structural Similarity (`SSIM`) [77]. `PixCorr` measures the correlation between the reconstructed image and the corresponding ground-truth image at the pixel level, while `SSIM` evaluates their structural similarity in terms of luminance, contrast, and local image structure.

For semantic evaluation, we extract features from several pretrained vision models. Specifically, `Alex(2)` and `Alex(5)` are based on the second and fifth layers of AlexNet [78], `Incep` is based on Inception-v3 [79], and `CLIP` is based on CLIP [44]. Following standard practice, these metrics are computed using a two-way identification protocol, where the reconstruction is considered correct if its feature representation is more similar to that of the matched ground-truth image than to that of a mismatched image. The chance level for this evaluation is 50%.

We additionally report the average correlation distance between reconstructed and ground-truth features using EfficientNet-B1 (`Eff`) [80] and SwAV-ResNet50 (`SwAV`) [81]. These two metrics complement the identification-based measures by directly quantifying the closeness of the feature embeddings.

Table 4: Training hyperparameter settings of BrainTC.

Setting	Source-subject Training	Target-subject Calibration
Optimizer	AdamW	
AdamW Momentum	$\beta_1 = 0.9, \beta_2 = 0.999$	
Weight Decay	5×10^{-2}	
Learning Rate Schedule	Linear Warmup + Cosine Annealing	
Gradient Clipping	1.0	
Numerical Precision	bf16	
Progressive Dropout	$0 \rightarrow 0.3$ (linear)	Disabled
Warmup Steps	200	50
Training Epochs	60	20
Batch Size	64	16
Initial Learning Rate	1×10^{-4}	5×10^{-5}
Trainable Parameters	Shared cores/bases + source residuals	Target residuals

To demonstrate computational efficiency, we further report the total number of learnable parameters (`Para.`) and floating-point operations (`FLOPs`) required during target-subject calibration, excluding frozen image generation backbones unless otherwise specified.

A.3 Cross-subject Evaluation Protocol

For cross-subject evaluation on NSD, we adopt a leave-one-subject-out protocol. For each target subject, BrainTC is trained on the remaining source subjects and evaluated on the held-out target subject. Specifically, when one subject is selected as the target subject, all response-stimulus pairs from this subject are excluded from source-subject training. The reported results are averaged over the four full-session NSD subjects, namely Subj-01, Subj-02, Subj-05, and Subj-07.

A.4 Training Details

BrainTC is trained in two stages: source-subject training and target-subject few-shot calibration. We summarize the training hyperparameters in Tab. 4. Unless otherwise specified, the same training and inference pipeline is used across the main experiments, ablation studies, and cross-dataset evaluations.

In the first stage, the shared tensor cores/bases and source-subject residual factors are jointly optimized to learn transferable functional structures and subject-specific variations. Source-subject training is conducted for 60 epochs with a batch size of 64 and an initial learning rate of 1×10^{-4}. We use 200 warmup steps, followed by cosine annealing. During source-subject training, progressive dropout is linearly increased from 0 to 0.3 to improve training robustness.

In the second stage, BrainTC adapts to a new target subject with limited calibration data. The shared tensor cores/bases are frozen, and only the target-subject residual factors are updated, including the residual coefficients $\Delta\alpha_r^{(s^*)}$ and residual biases $\Delta b_r^{(s^*)}$. This lightweight calibration preserves the pretrained cross-subject structure while enabling subject-specific correction. Target-subject calibration is performed for 20 epochs with a batch size of 16 and a learning rate of 5×10^{-5}. We use 50 warmup steps followed by cosine annealing. Progressive dropout are disabled during calibration to avoid over-regularizing the limited target-subject adaptation data. The full decoder is not fine-tuned, since unrestricted adaptation may disturb the shared structure learned from source subjects.

For optimization, we use AdamW with momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and a weight decay of 5×10^{-2}. Gradient clipping with a maximum norm of 1.0 is applied throughout training. All experiments are conducted on a single NVIDIA A800 GPU with bf16 mixed precision. For model configuration, we set the latent functional dimension and token embedding dimension to $D = 256$ and $d_t = 512$, respectively. The tensor capacity parameters are set to $K = 4$ and $K_a = 4$, which provide the best trade-off between reconstruction quality and computational cost in the tensor design analysis. The loss weights are set to $\beta = 0.1$ and $\mu = 0.05$.

To improve numerical stability in the distribution-aware loss, covariance statistics are estimated on the training set, and small eigenvalues are clipped when computing the inverse square root. This stabilizes covariance-based feature matching, especially under limited-data calibration settings.

Table 5: Quantitative results of visual reconstruction across subjects of NSD.

Method	Low-Level				High-Level			
	PixCorr↑	SSIM↑	Alex(2)↑	Alex(5)↑	Incep↑	CLIP↑	Eff↓	SwAV↓
▼ *Subject 1*								
MindHier	0.273	0.394	97.0%	99.3%	96.7%	97.2%	0.598	0.321
MindEye2 (1 h)	0.235	0.428	88.0%	93.3%	83.6%	80.8%	0.798	0.459
MindAligner (1 h)	0.226	0.415	88.2%	93.3%	83.5%	81.8%	0.800	0.459
MindTuner (1 h)	0.262	0.422	90.6%	94.9%	85.8%	84.6%	0.774	0.433
BrainIT (1 h)	0.396	0.484	99.4%	99.6%	95.9%	94.7%	0.619	0.348
ZEBRA (zero-shot)	0.153	0.384	76.1%	81.8%	73.4%	72.3%	0.814	0.490
BrainTC (1 h)	0.405	0.491	99.5%	99.7%	96.2%	95.1%	0.608	0.340
BrainTC (zero-shot)	0.187	0.399	79.6%	85.1%	78.5%	78.2%	0.799	0.475
▼ *Subject 2*								
MindHier	0.237	0.383	95.1%	98.8%	96.1%	96.0%	0.611	0.328
MindEye2 (1 h)	0.200	0.433	85.0%	92.1%	81.9%	79.4%	0.807	0.467
MindAligner (1 h)	0.218	0.426	88.1%	93.3%	84.1%	82.5%	0.791	0.452
MindTuner (1 h)	0.225	0.425	89.1%	95.1%	84.8%	83.7%	0.781	0.440
BrainIT (1 h)	0.320	0.472	97.0%	98.5%	94.7%	93.4%	0.642	0.365
ZEBRA (zero-shot)	0.135	0.382	76.0%	81.6%	72.0%	71.3%	0.839	0.496
BrainTC (1 h)	0.327	0.480	97.6%	98.7%	94.6%	93.3%	0.623	0.358
BrainTC (zero-shot)	0.180	0.395	79.1%	84.0%	77.8%	77.1%	0.808	0.483
▼ *Subject 5*								
MindHier	0.218	0.378	91.4%	97.8%	97.0%	97.4%	0.576	0.316
MindEye2 (1 h)	0.175	0.405	83.1%	91.0%	84.3%	82.5%	0.781	0.444
MindAligner (1 h)	0.197	0.409	84.7%	91.6%	84.6%	82.8%	0.784	0.454
MindTuner (1 h)	0.208	0.415	86.8%	93.7%	87.7%	85.9%	0.750	0.422
BrainIT (1 h)	0.310	0.469	96.4%	98.3%	94.0%	92.7%	0.650	0.374
ZEBRA (zero-shot)	0.119	0.365	72.6%	80.9%	72.4%	72.1%	0.825	0.503
BrainTC (1 h)	0.318	0.477	97.3%	98.5%	94.1%	92.9%	0.628	0.366
BrainTC (zero-shot)	0.176	0.390	78.6%	83.2%	77.4%	76.3%	0.813	0.490
▼ *Subject 7*								
MindHier	0.213	0.372	92.3%	97.6%	94.0%	95.0%	0.640	0.352
MindEye2 (1 h)	0.170	0.408	80.7%	85.9%	74.9%	74.3%	0.854	0.504
MindAligner (1 h)	0.183	0.407	81.5%	88.3%	79.9%	77.8%	0.834	0.487
MindTuner (1 h)	0.202	0.417	84.5%	90.8%	80.7%	79.6%	0.817	0.465
BrainIT (1 h)	0.298	0.467	95.6%	98.0%	93.0%	91.2%	0.681	0.393
ZEBRA (zero-shot)	0.117	0.370	73.8%	80.4%	71.0%	70.4%	0.869	0.534
BrainTC (1 h)	0.310	0.476	97.2%	98.3%	93.5%	92.3%	0.641	0.384
BrainTC (zero-shot)	0.169	0.384	77.9%	82.1%	75.9%	75.6%	0.832	0.508
▼ *Average*								
MindHier [82]	0.235	0.381	94.0%	98.4%	95.9%	96.4%	0.606	0.329
MindEye2 [11] (1 h)	0.195	0.419	84.2%	90.6%	81.2%	79.2%	0.810	0.468
MindTuner [14] (1 h)	0.224	0.420	87.8%	93.6%	84.8%	83.5%	0.780	0.440
MindAligner [16] (1 h)	0.206	0.414	85.6%	91.6%	83.0%	81.2%	0.802	0.463
BrainIT [15] (1 h)	0.331	0.473	97.1%	98.6%	94.4%	93.0%	0.648	0.370
ZEBRA (zero-shot)	0.131	0.375	74.6%	81.2%	72.2%	71.5%	0.837	0.506
BrainTC (1 h)	0.340	0.481	97.9%	98.8%	94.6%	93.4%	0.625	0.362
BrainTC (zero-shot)	0.178	0.392	78.8%	83.6%	77.4%	76.8%	0.813	0.489

B More Results

B.1 More Visual Reconstruction Results

Tab. 5 reports detailed reconstruction results for each NSD subject, which are consistent with the averaged performance in the main paper. With only 1 hour of target-subject calibration data, BrainTC achieves strong and stable performance across subjects 1, 2, 5, and 7, especially on low-level metrics such as PixCorr and SSIM. This indicates that the proposed weight-decomposed tensor adaptation effectively recovers subject-specific visual details while updating only compact residual parameters. BrainTC also maintains competitive high-level perceptual and semantic performance, preserving both visual fidelity and semantic consistency across individuals.

Table 6: Quantitative results of few-shot transfer learning on Subj-01 using different amounts of target-subject calibration data.

Method	Low-Level				High-Level			
	PixCorr ↑	SSIM ↑	Alex(2) ↑	Alex(5) ↑	Incep ↑	CLIP ↑	Eff ↓	SwAV ↓
▼ *15 min*								
BrainIT	0.336	0.476	97.6%	98.4%	90.7%	91.3%	0.706	0.404
BrainTC	**0.348**	**0.482**	**98.1%**	**98.7%**	**92.0%**	**92.2%**	**0.688**	**0.392**
▼ *30 min*								
BrainIT	0.378	0.480	99.1%	99.3%	94.3%	93.3%	0.652	0.372
BrainTC	**0.389**	**0.486**	**99.2%**	**99.4%**	**95.0%**	**94.1%**	**0.635**	**0.363**
▼ *1 hour*								
BrainIT	0.396	0.484	99.4%	99.6%	95.9%	94.7%	0.619	0.348
BrainTC	**0.405**	**0.491**	**99.5%**	**99.7%**	**96.2%**	**95.1%**	**0.608**	**0.340**

Despite expected inter-subject variations in fMRI responses and signal quality, BrainTC remains robust under both few-shot and zero-shot settings. In the few-shot setting, it achieves the best or among the best reconstruction quality across subjects compared with existing baselines. In the zero-shot setting, BrainTC outperforms ZEBRA for every subject, demonstrating that the shared tensor cores learn transferable functional structure without target-specific calibration. These results further support the cross-subject generalization and adaptation efficiency of BrainTC.

Fig. 7 provides qualitative comparisons between BrainTC and representative reconstruction methods. Compared with the zero-shot baseline ZEBRA, BrainTC produces more faithful global layouts, clearer object shapes, and more consistent scene semantics. Although fully fine-tuned methods may recover finer textures, BrainTC narrows the gap between zero-shot and calibration-based reconstruction, often yielding semantically coherent results with competitive scene structure. These visual comparisons are consistent with the quantitative results and confirm the effectiveness of shared tensorized functional alignment for unseen subjects.

B.2 Results of Few-Shot Learning with More Limited Data

Tab. 6 reports transfer learning results on Subject 1 using different amounts of target-subject calibration data, including 15 minutes, 30 minutes, and 1 hour. BrainTC consistently outperforms BrainIT across all evaluated data regimes and metrics, demonstrating its strong adaptation ability under extremely limited calibration. Even with only 15 minutes of subject-specific data, BrainTC achieves competitive reconstruction quality, indicating that the shared tensor cores provide effective transferable structure while the lightweight residual factors can rapidly capture subject-specific variations.

As the amount of calibration data increases, BrainTC further improves both low-level fidelity and high-level semantic alignment, showing a stable scaling trend from 15 minutes to 1 hour. These results suggest that the proposed weight-decomposed tensor adaptation is not only parameter-efficient, but also highly data-efficient, making BrainTC suitable for practical cross-subject visual reconstruction scenarios where collecting extensive fMRI data for each new subject is costly.

Fig. 8 provides qualitative comparisons under different amounts of target-subject calibration data. As the calibration data increase from 15 minutes to 30 minutes and 1 hour, BrainTC further improves fine-grained details, object shapes, and background consistency, showing a clear and stable adaptation trend. Compared with BrainIT under the same data budgets, BrainTC generates more faithful and semantically aligned reconstructions in many challenging cases, such as animals, sports scenes, vehicles, and human activities. These qualitative results are consistent with the quantitative results in Tab. 6, demonstrating the strong data efficiency of the proposed weight-decomposed tensor adaptation.

B.3 More Ablation Results

The upper part of Tab. 7 studies which target-subject parameters should be updated during adaptation. Updating only the residual bias yields the weakest performance, since it can mainly compensate for subject-specific activation offsets but cannot modify the ROI-wise functional projection itself. In contrast, updating only the residual coefficients leads to clear improvements, suggesting that

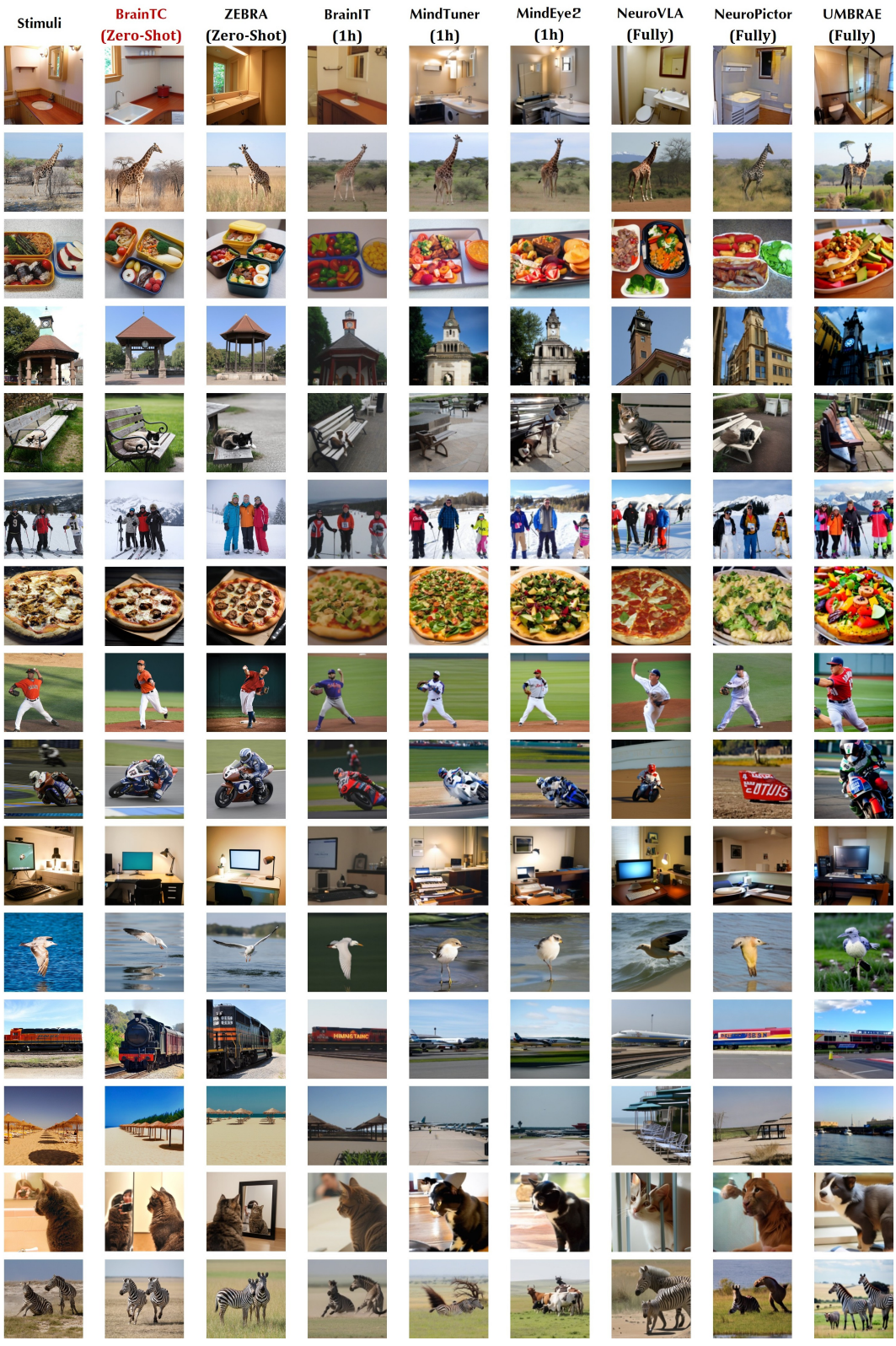


Figure 7: More qualitative comparisons of reconstructions performed for Subj-01.

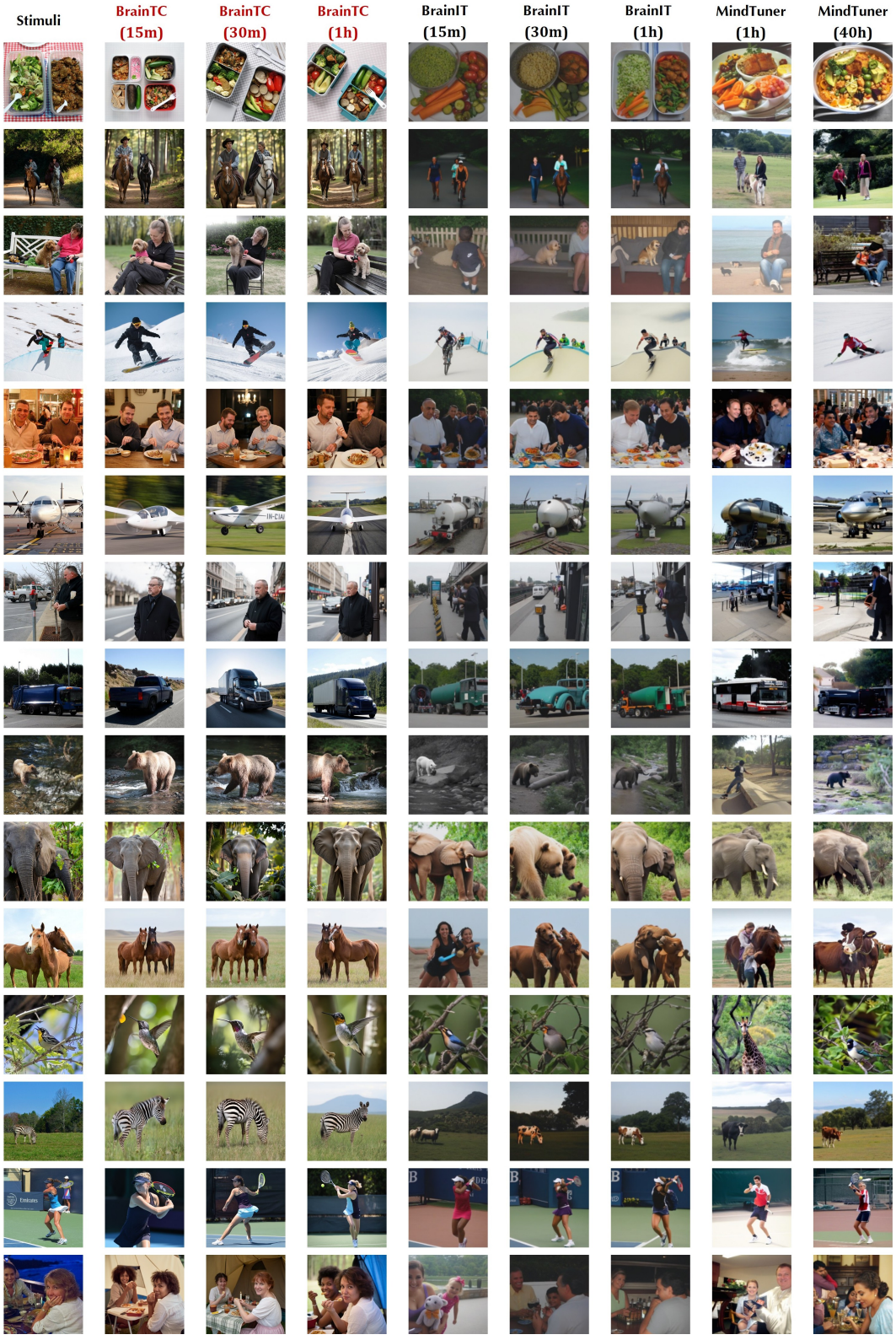


Figure 8: Qualitative comparisons of transfer learning for Subj-01.

Table 7: Ablation on updated parameter subsets and sensitivity analysis of loss weights.

Setting		Low-Level				High-Level				Few-shot	
		PixCorr↑	SSIM↑	Alex(2)↑	Alex(5)↑	Incep↑	CLIP↑	Eff↓	SwAV↓	PixCorr↑	CLIP↑
▼ *Ablation on Updated Parameter Subset*											
Residual coefficient only	$\Delta\alpha_r^{(s^*)}$	0.181	0.392	78.7%	84.4%	77.8%	77.4%	0.811	0.483	0.390	94.0%
Residual bias only	$\Delta b_r^{(s^*)}$	0.176	0.388	78.1%	83.8%	77.0%	76.6%	0.823	0.491	0.375	92.9%
Ours	$\Delta\alpha_r^{(s^*)}$+$\Delta b_r^{(s^*)}$	**0.187**	**0.399**	**79.6%**	**85.1%**	**78.5%**	**78.2%**	**0.799**	**0.475**	**0.405**	**95.1%**
+Subject embedding	$\Delta\alpha_r^{(s^*)}$+$\Delta b_r^{(s^*)}$+e_s	0.185	0.397	79.3%	84.8%	78.2%	77.9%	0.804	0.478	0.401	94.6%
Full decoder fine-tuning		0.183	0.394	79.0%	84.6%	78.0%	77.7%	0.808	0.481	0.397	94.2%
▼ *Ablation on Loss Functions*											
Feature only	$+\mathcal{L}_{\text{feat}}$	0.176	0.394	78.6%	84.3%	77.2%	76.9%	0.807	0.481	0.394	94.4%
Feature + Distribution	$+\mathcal{L}_{\text{feat}}+\mathcal{L}_{\text{dist}}$	0.181	0.396	79.1%	84.7%	77.9%	77.6%	0.803	0.478	0.399	94.8%
Ours	$+\mathcal{L}_{\text{feat}}+\mathcal{L}_{\text{dist}}+\mathcal{L}_{\text{inv}}$	**0.187**	**0.399**	**79.6%**	**85.1%**	**78.5%**	**78.2%**	**0.799**	**0.475**	**0.405**	**95.1%**
▼ *Sensitivity Analysis of Loss Weights*											
$\beta = 0.01$	$\mu = 0.1$	0.178	0.390	78.2%	83.9%	77.1%	76.8%	0.824	0.492	0.383	93.1%
$\beta = 0.01$	$\mu = 0.05$	0.181	0.392	78.5%	84.2%	77.4%	77.1%	0.818	0.488	0.388	93.6%
$\beta = 0.01$	$\mu = 0.01$	0.179	0.391	78.3%	84.0%	77.2%	76.9%	0.822	0.490	0.381	93.2%
$\beta = 0.05$	$\mu = 0.1$	0.183	0.395	79.0%	84.5%	77.8%	77.5%	0.811	0.483	0.394	94.1%
$\beta = 0.05$	$\mu = 0.05$	0.185	0.397	79.3%	84.8%	78.1%	77.8%	0.805	0.479	0.399	94.6%
$\beta = 0.05$	$\mu = 0.01$	0.182	0.394	78.8%	84.4%	77.7%	77.3%	0.813	0.485	0.392	94.0%
$\beta = 0.1$	$\mu = 0.01$	0.184	0.396	79.1%	84.7%	78.0%	77.7%	0.807	0.480	0.396	94.4%
$\beta = 0.1$	$\mu = 0.05$	**0.187**	**0.399**	**79.6%**	**85.1%**	**78.5%**	**78.2%**	**0.799**	**0.475**	**0.405**	**95.1%**

recombining the shared tensor cores is more effective for capturing subject-specific functional variations. Our default strategy, which updates both the residual coefficients and residual biases, achieves the best performance across low-level and high-level metrics. This confirms that the two components are complementary: the residual coefficients adjust the functional projection, while the residual biases provide additional subject-specific calibration. We also evaluate two alternative adaptation strategies. Adding a target-subject embedding slightly degrades performance, likely because the extra free parameters introduce mild overfitting under limited calibration data. Fully fine-tuning the decoder also performs worse than our default adaptation, despite using more trainable parameters. This indicates that unrestricted fine-tuning can disturb the shared cross-subject structure learned during pretraining, whereas the proposed residual-only adaptation preserves transferable tensor cores while allowing efficient personalization. These results validate the design choice of updating only $\Delta\alpha_r^{(s^*)}$ and $\Delta b_r^{(s^*)}$ for few-shot target-subject adaptation.

The middle part of Tab. 7 analyzes the contribution of different training objectives. Using only the hierarchical feature loss $\mathcal{L}_{\text{feat}}$ provides a reasonable baseline, since it directly supervises the predicted representations with multi-level visual targets. However, the performance is still limited because this objective mainly enforces sample-wise feature regression and does not explicitly account for the distributional structure of the visual representation space. After adding the distribution-aware alignment loss $\mathcal{L}_{\text{dist}}$, the model obtains consistent improvements on both low-level and high-level metrics. This indicates that $\mathcal{L}_{\text{dist}}$ helps calibrate prediction errors according to the empirical covariance of visual features, leading to more balanced and faithful feature-space alignment.

When the subject-invariant regularization $\mathcal{L}_{\text{inv}}$ is further incorporated, the model achieves the best performance across all metrics, including the few-shot evaluation. This demonstrates that enforcing consistency among neural representations induced by the same stimulus across different source subjects encourages the decoder to learn more stimulus-driven and subject-agnostic representations. Such invariance is particularly important for cross-subject decoding, where the target subject is unseen during training. Overall, these results show that the three losses are complementary: $\mathcal{L}_{\text{feat}}$ provides direct hierarchical visual supervision, $\mathcal{L}_{\text{dist}}$ improves distribution-aware feature matching, and $\mathcal{L}_{\text{inv}}$ enhances cross-subject representation consistency.

B.4 Sensitivity Analysis of Hyperparameters

We first analyze the sensitivity of the loss weights in the lower part of Tab. 7. Overall, BrainTC shows stable performance across different combinations of the distribution-aware loss weight and the subject-invariance loss weight, indicating that the proposed training objective is not overly sensitive to a specific hyperparameter choice. A relatively small distribution-aware weight leads to

Table 8: Design analysis of latent functional dimension D and token embedding dimension d_t on few-shot setting (average results of subjects 1, 2, 5, and 7 in NSD).

D	d_t	Incep ↑	CLIP ↑	Eff ↓	SwAV ↓	Para.	FLOPs
64	128	92.7%	91.4%	0.661	0.384	3.62M	0.02G
128	256	93.8%	92.5%	0.640	0.371	4.48M	0.02G
256	256	94.1%	92.8%	0.635	0.368	4.93M	0.03G
128	512	94.0%	92.7%	0.637	0.369	5.08M	0.03G
256	512	**94.6%**	**93.4%**	**0.625**	**0.362**	**5.44M**	**0.03G**
512	512	94.4%	93.1%	0.629	0.365	7.21M	0.04G
256	768	94.3%	93.0%	0.631	0.366	6.86M	0.05G
512	768	94.2%	92.9%	0.634	0.367	9.34M	0.06G

Table 9: To demonstrate the effectiveness of the method, we evaluated Subj-01 using BrainIT's testing standards and additional quantitative metrics.

Method	SSIM-color ↑	1000-way CLIP ↑	LPIPS ↓
NeuroPictor	0.375	0.255	0.628
MindEye2	0.383	0.238	0.624
MindTuner	0.371	0.231	0.638
BrainIT	0.472	0.393	0.596
BrainTC	**0.493**	**0.406**	**0.577**

weaker reconstruction performance, suggesting that visual feature distribution matching is important for preserving perceptual and semantic consistency. Meanwhile, an appropriate subject-invariance weight improves cross-subject generalization by encouraging aligned neural representations for shared stimuli. The best performance is obtained with $\beta = 0.1$ and $\mu = 0.05$, which provides a good balance between distribution-aware visual alignment and subject-invariant neural regularization.

We further study the effects of the latent functional dimension D and token embedding dimension d_t in Tab. 8. Increasing D and d_t generally improves reconstruction quality, as larger dimensions provide stronger capacity for representing ROI-wise functional responses and inter-ROI dependencies. However, the performance gain saturates when the dimensions become too large, while the number of adaptation parameters and FLOPs continue to increase. In particular, setting $D = 256$ and $d_t = 512$ achieves the best overall performance with moderate computational cost. Larger configurations, such as $D = 512$ or $d_t = 768$, do not bring further improvement and instead introduce additional complexity. Therefore, we adopt $D = 256$ and $d_t = 512$ as the default setting, which offers the best trade-off between reconstruction accuracy and adaptation efficiency.

B.5 Additional Metrics Evaluation

Some commonly used metrics for fMRI-based visual reconstruction, especially retrieval-style metrics with limited candidate sets, can become partially saturated and may not fully reflect perceptual differences between methods. In addition, standard grayscale SSIM ignores chromatic information, which is important for evaluating visual reconstruction quality. To provide a more comprehensive assessment, we further evaluate BrainTC on Subj-01 using BrainIT's testing protocol with additional metrics, including SSIM-color, 1000-way CLIP retrieval, and LPIPS. As shown in Tab. 9, BrainTC achieves the best performance across all additional metrics. Compared with BrainIT, BrainTC improves SSIM-color, indicating better preservation of both structural and color information. It also obtains higher 1000-way CLIP retrieval accuracy, suggesting stronger semantic alignment between reconstructed and ground-truth images under a more challenging retrieval setting. Moreover, BrainTC achieves the lowest LPIPS score, demonstrating improved perceptual similarity that is more consistent with human visual judgment. These results further confirm that the improvements of BrainTC are not limited to the standard metrics reported in the main paper, but also generalize to more discriminative evaluations of color fidelity, semantic retrieval, and perceptual quality.

B.6 Generalizability Across Other fMRI Datasets

To further demonstrate the robustness and generalizability of BrainTC beyond the NSD, we additionally evaluate it on the Generic Object Decoding (GOD) dataset [83] and the THINGS-fMRI

Table 10: Quantitative comparison on GOD [83] and THINGS-fMRI [84] datasets.

Method		Low-Level				High-Level			
		PixCorr↑	SSIM↑	Alex(2)↑	Alex(5)↑	Incep↑	CLIP↑	Eff↓	SwAV↓
▼ *GOD Dataset*									
VQ-fMRI [43]	[ICML'23]	0.180	0.423	87.2%	92.7%	78.7%	74.8%	0.838	0.498
MinD-Vis [42]	[CVPR'23]	0.120	0.389	82.8%	93.8%	81.1%	82.6%	0.836	0.491
CMVDM [85]	[AAAI'24]	0.279	0.455	88.3%	94.0%	81.7%	82.1%	0.810	0.485
STTM [52]	[AAAI'25]	0.253	0.367	92.0%	95.6%	81.3%	87.0%	0.870	0.505
BrainTC		**0.288**	**0.469**	**92.7%**	**96.3%**	**84.5%**	**88.3%**	**0.808**	**0.481**
▼ *THINGS-fMRI Dataset*									
BrainFLORA [86]	[ACM MM'25]	0.079	0.348	59.5%	62.2%	60.4%	57.6%	0.812	0.661
MindHier [82]	[ICLR'26]	0.109	0.357	71.0%	81.9%	71.9%	73.1%	0.900	0.589
BrainTC		**0.115**	**0.381**	**78.9%**	**83.0%**	**75.9%**	**78.6%**	**0.778**	**0.565**

dataset [84]. These datasets differ from NSD in subject scale, stimulus organization, and acquisition protocols, thereby providing a broader assessment of cross-dataset generalization.

GOD is a specialized benchmark for fMRI-based object decoding, containing data from 5 subjects collected using a 3T scanner. The dataset includes images from 200 representative object categories derived from ImageNet. Following the standard protocol, the training session consists of 1,200 images, corresponding to 8 images from each of 150 categories, while the test session contains 50 images, one from each of the remaining 50 categories. Notably, the test categories are disjoint from the training categories, making the task particularly suitable for evaluating out-of-category generalization. THINGS-fMRI contains recordings from 3 subjects. Its training set comprises 720 concepts, each associated with 12 images viewed once, whereas the test set includes 100 concepts, with one image selected from each concept and repeated 12 times. Compared with NSD, THINGS-fMRI poses an additional challenge due to its distinct acquisition settings and data structure.

For both datasets, we keep the overall training and inference pipeline consistent with that used in the main experiments, so that the supplementary results primarily reflect the robustness across datasets of the proposed framework rather than changes in the optimization setting.

As shown in Tab. 10, BrainTC achieves the best performance on both GOD and THINGS-fMRI across all evaluated metrics, demonstrating strong generalization beyond the NSD benchmark. On the GOD dataset, BrainTC consistently improves both low-level and high-level reconstruction quality. Compared with the strongest prior results, BrainTC improves PixCorr from 0.279 to 0.288 and SSIM from 0.455 to 0.469, while also achieving higher semantic alignment scores, including AlexNet, Inception, and CLIP metrics. These results indicate that BrainTC can effectively handle the challenging out-of-category setting of GOD, where test categories are disjoint from training categories. On THINGS-fMRI, BrainTC also outperforms existing methods by a clear margin. Overall, these results confirm that the proposed tensorized alignment and decoding framework is not tailored to a single dataset, but can generalize effectively across different stimulus distributions, acquisition protocols, and subject cohorts.

C Impact and Future Work

C.1 Broader Impact

BrainTC contributes to the development of cross-subject brain decoding by showing that structured tensorized parameterization can improve generalization to unseen subjects while substantially reducing adaptation cost. More broadly, the results suggest that explicit factorization of shared functional structure and subject-specific variation may provide a useful design principle for neural decoding models, especially in settings where subject heterogeneity and limited calibration data are major bottlenecks. In this sense, BrainTC may be relevant not only for visual decoding, but also for other brain–computer interface scenarios that require efficient transfer across individuals.

At the same time, improving cross-subject decoding performance also raises important ethical and privacy concerns. Brain signals contain highly sensitive information about perception and internal mental states, and models that generalize across subjects may further increase the risk of large-scale

misuse if appropriate safeguards are absent. If neural data are collected, shared, or decoded without rigorous protection, they may expose private cognitive information in ways that are difficult to control or reverse. We believe that progress in brain decoding must be accompanied by equally strong commitments to informed consent, secure data governance, transparent model usage protocols, and interdisciplinary discussions on the legal and ethical boundaries of neural data technologies.

C.2 Limitations

Although BrainTC achieves strong and competitive performance in cross-subject visual decoding, several challenges still remain.

First, while BrainTC demonstrates clear advantages in zero-shot and few-shot settings, the current study is mainly validated on visual decoding from a limited number of subjects in the NSD, and additional evaluations are provided on GOD and THINGS-fMRI. Although this setting is standard and widely adopted, it does not fully cover the broader variability encountered across datasets, imaging protocols, preprocessing pipelines, and more diverse subject populations. The robustness of shared tensor-core alignment under stronger domain shifts still requires further investigation.

Second, the current formulation adopts ROI-wise neural organization as the basic unit for alignment and decoding. This choice improves efficiency and supports structured modeling, but it may also discard fine-grained voxel-level variation and subject-specific cortical geometry that could be useful for decoding. More generally, the present model emphasizes compactness and transferability, and may not fully exploit all available spatial and temporal structure in neural measurements.

Finally, although BrainTC is substantially more parameter-efficient than many existing approaches, practical deployment in real-world brain-computer interface systems remains challenging. End-to-end latency in realistic settings is influenced not only by model complexity, but also by neural signal acquisition, preprocessing, calibration, and downstream reconstruction pipelines. We believe these limitations reflect broader challenges in cross-subject brain decoding rather than weaknesses unique to our approach.

D Use of Large Language Models (LLMs)

During the preparation of this manuscript, large language models (LLMs), including tools such as ChatGPT, Gemini, and DeepSeek, were used only for editorial and formatting assistance. Their use was limited to non-scientific support tasks, primarily in the following aspects:

1. improving the English expression of the manuscript, such as correcting grammar, refining wording, and enhancing overall readability;
2. assisting with LaTeX editing and typesetting to improve the presentation and consistency of equations, tables, figures, and related formatting elements.

The LLMs were not involved in the development of the main scientific ideas, theoretical derivations, algorithm design, experimental procedures, or interpretation of the results. All substantive academic contributions were completed and verified by the authors. The authors carefully reviewed and revised the manuscript and take full responsibility for its content, originality, and scientific soundness.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims made in the abstract and introduction accurately reflect the paper's contributions and scope.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We have discussed the limitations and future work in the Appendix.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

Justification: Our work does not include the steps of theoretical formula derivation and proof.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We have reported all the information, including module details, parameters and implementation details, in the main text and appendices, and have provided the pseudocode for the algorithm. We have submitted supplementary materials (including parts of the source codes), and the dataset is open-source to facilitate reproduction.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We have submitted a GitHub link, and the dataset is open-source to facilitate reproduction.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We have provided detailed training details and parameter settings.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We have provided statistical verification for the experimental results.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We have provided information about the computer resources of our experiments in the Experimental Setup section.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: The research conducted in this paper conforms to the Code of Ethics of the Conference on Neural Information Processing Systems (NeurIPS).

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: In the Appendix, we explore the broader impact of our work on the field of BCI research.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All the reference materials have been properly cited.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: This paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: Our paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: Our paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: We only use LLMs for formatting purposes, and we also reported the usage details of LLM in the Appendix.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.