Leveraging Visual Blur Perception Characteristics for EEG Decoding

Wenchao Liu[1], Hongwei Li[1], Zhouyang Xu[1], Lin Ma[1], Haifeng Li[1]*

[1] Harbin Institute of Technology, Harbin, China
23b903096@stu.hit.edu.cn, lihongwei@hit.edu.cn, 23b903017@stu.hit.edu.cn,
malin_li@hit.edu.cn, lihaifeng@hit.edu.cn

Abstract

In recent years, electroencephalography (EEG)-based visual decoding research has become a key direction for revealing brain processing mechanisms and realizing brain-computer interfaces. This emerging field has attracted extensive attention in the fields of brain science, cognitive neuroscience, and artificial intelligence. Among various approaches, contrastive learning has demonstrated strong performance in aligning multi-modal data, effectively enabling unified representations across modalities. However, during human visual perception, images are often subject to varying degrees of blurring due to the uneven distribution of retinal photoreceptor cells and the limited speed of lens accommodation. To address the mismatch between EEG and visual representations, we propose a novel visual decoding framework inspired by human perceptual blurring. Specifically, multi-level Gaussian blurring is applied to the image to simulate human visual characteristics, followed by a feature selection module to construct robust visual representations. For EEG decoding, we design a lightweight and efficient network employing positively constrained spatial convolutions to identify channels associated with visual processing. The EEG and visual features are then aligned using contrastive learning. We evaluate the proposed framework on the Things-EEG dataset. Experimental results show significant improvements in the zero-shot brain-to-image retrieval task, achieving a top-1 accuracy of 80% and a top-5 accuracy of 96.9 %, surpassing previous state-of-the-art methods by margins of 29.1% and 17.2%, respectively. These findings highlight the potential of incorporating perceptual properties into EEG-based visual decoding.

Code —
https://github.com/makeitperfect/VisualEEGDecoding

Introduction

How the human brain encodes and decodes visual information has been one of the core issues in the fields of neuroscience, cognitive neuroscience, and artificial intelligence (Tootell et al. 1983; Jin et al. 2008; Tanigawa, Lu, and Roe 2010; Chang, Bao, and Tsao 2017; Roe et al. 2012; Khaligh-Razavi and Kriegeskorte 2014; Güçlü and Van Gerven 2015). By decoding neural activity related to visual stimuli in brain physiological signals, this not only helps to understand the brain's visual information processing mechanisms but also provides critical technical support for applications such as brain-computer interfaces (BCI) (Zhu et al. 2010; Salvaris and Sepulveda 2009) and neural rehabilitation (Chen et al. 2020; Fernández et al. 2021; Beauchamp et al. 2020). In recent years, visual decoding research based on neural signals such as fMRI and EEG has garnered significant attention. Among these, functional Magnetic Resonance Imaging (fMRI) has gained significant attention due to its high spatial resolution, enabling the identification of fine-grained spatial patterns activated in the brain during visual tasks. It is currently the most commonly used visual decoding technique (Du et al. 2023; Horikawa and Kamitani 2017; Allen et al. 2022; Liu et al. 2023; Scotti et al. 2023; Fang, Zheng, and Pan 2023). However, fMRI has limitations such as high cost and low temporal resolution, which restrict its application in BCI and neural rehabilitation scenarios with high requirements for real-time performance and scalability (Smith 2004; Lin, Sprague, and Singh 2022).

In contrast, electroencephalogram (EEG) has demonstrated broad application prospects in the field of visual decoding due to its high temporal resolution and low cost (Khadir et al. 2023; Grootswagers et al. 2022; Di Russo et al. 2002; Hartmann, Schirrmeister, and Ball 2018; Singh et al. 2023; Spampinato et al. 2017; Chen et al. 2023; Zhang et al. 2025). Currently, mainstream EEG visual decoding methods mostly adopt contrastive learning strategies, aligning the features extracted from EEG with image features from pre-trained visual models (such as CLIP (Radford et al. 2021), DINO (Oquab et al. 2023), ViT (Dosovitskiy et al. 2020), and ResNet (He et al. 2016)). These visual models are trained on large-scale image-text datasets, endowing them with discriminative and generalizable capabilities, thereby significantly enhancing the performance and robustness of decoding models even with limited EEG training samples. Song et al. used contrastive learning to align EEG signal representations and CLIP representations, demonstrating excellent performance in zero-shot classification, achieving a top-1 classification accuracy of 15.6% and a top-5 classification accuracy of 42.8% on the Things-EEG dataset. Wei et al. proposed a bidirectional cyclic consistency (MB2C) framework, which not only uses contrastive learning to align image representations extracted by CLIP with features

*Corresponding author.

extracted from EEG, MB2C also employs two generative adversarial networks to further reduce the differences between the two representations. This method achieved a top-1 retrieval accuracy of 28.45% and a retrieval accuracy of 60.37% on the Things-EEG dataset. Li et al. further trained the representations aligned with the CLIP model using diffusion models, achieving visual reconstruction of EEG signals. These methods have improved performance in visual EEG decoding and reconstruction; however, they directly use the representations of raw images for training, ignoring the differences between human visual perception and the presentation of images. This discrepancy is referred to as system gap. Due to the highly uneven distribution of photoreceptor cells on the retina, the fovea region has a higher density of cone cells, which provide excellent spatial resolution and detail perception capabilities. In contrast, the peripheral regions are dominated by rod cells, which are primarily responsible for light sensitivity and are less sensitive to spatial details, resulting in a generally blurred overall perception (Kolb et al. 1995; Curcio et al. 1990; Osterberg 1935). This results in most image regions perceived by humans appearing blurry, far from the completely clear image input (Bharadwaj and Schor 2006; Burge and Geisler 2011; Sprague et al. 2016).

To reduce the discrepancy between EEG representations and pre-trained visual representations, Wu et al. proposed a method called Uncertainty-Aware Blur Prior (UBP), which calculates the uncertainty between sample pairs and dynamically blurs the high-frequency details of the original image using a foveation-like approach, thereby reducing the impact of mismatch and improving alignment. This method achieved a top-1 accuracy of 50.9% and a top-5 accuracy of 79.7% in zero-shot retrieval tasks, significantly enhancing the performance of EEG visual decoding. However, the human visual blur perception mechanism in reality is more complex. Due to variations in ocular structure and visual acuity among individuals, blur perception exhibits highly personalized characteristics (Vera-Diaz, Woods, and Peli 2010). Additionally, constrained by the speed of lens accommodation, humans experience varying degrees of dynamic blur during rapid gaze shifts (Schaeffel, Wilhelm, and Zrenner 1993; Bharadwaj and Schor 2006). The single blur strategy adopted by UBP struggles to accurately model these diverse and complex perceptual characteristics.

To address this issue, we propose an EEG decoding framework based on visual blur perception characteristics. The framework employs multi-level Gaussian blurring to simulate the human visual blur perception mechanism. A pre-trained vision model then extracts features from these blurred images. To enable adaptive adjustment to varying blur degrees, we further design a feature selection module that dynamically selects features from different blur levels. For EEG decoding, we propose a simple yet efficient network architecture employing a positivity-constrained spatial convolution module for vision-relevant neural channel selection. The main contributions of this paper are as follows:

- A simple and efficient visual decoding framework for EEG is proposed. Inspired by the characteristics of human visual blur perception, the framework firstly processes visual stimuli with different blurs, and then uses feature selection and feature adaptation to construct pre-trained picture representations, which effectively reduces the effect of system gap.
- A simple, highly interpretable EEG coding network was constructed. Analysis of the model's interpretability revealed that electrodes in the occipital and temporal lobes play a dominant role in visual EEG decoding, aligning with the ventral pathway of visual processing. Furthermore, region-specific selectivity for different blur levels was observed, further validating the method's reasonableness and effectiveness.
- Our method achieved a top-1 accuracy of 80% and a top-5 accuracy of 96.9% in zero-shot brain-to-image retrieval on the Things-EEG dataset. These results demonstrate the significant potential of EEG visual decoding and provide new insights into neural mechanisms of visual information processing.

Methods

The EEG decoding framework based on visual blur perception characteristics proposed in this paper is shown in Figure 1. In this framework, we use contrast learning to align EEG representations with pre-trained visual representations and learn how to decode image information from EEG signals. To bridge the gap between visual stimuli and human perception, a blurring pipeline is introduced to allow for different blurring for visual stimuli during visual representation construction. Following representation extraction using a pre-trained model, feature selection and adaptation were performed to construct personalized visual representations.For EEG representation extraction, visual stimulus-evoked EEG signals are processed directly by the EEG encoder. During training, the EEG representations were aligned with the pre-trained visual representations using contrast learning. The testing phase employs a zero-shot retrieval approach, where EEG-derived features are compared with candidate image features through similarity computation, with the highest-scoring match selected as the corresponding image. The following sections provide a detailed introduction of this method.

Blurring pipeline

Human visual perception constitutes a complex physiological and cognitive process. The uneven photoreceptor distribution (cones and rods) across the retina results in foveal vision exhibiting maximal clarity and detail resolution, while peripheral vision progressively blurs with decreasing cone density. Furthermore, visual perception is modulated by multiple interdependent factors: ocular accommodation, dynamic visual acuity changes, neural attention mechanisms, cortical information integration, and ambient lighting conditions. To make the extraction of visual representations in the framework more aligned with human visual processing, images are first processed with multi-level blurring to simulate how the human eye perceives objects. Specifically, Gaussian blurring is applied to the images, with the results shown in

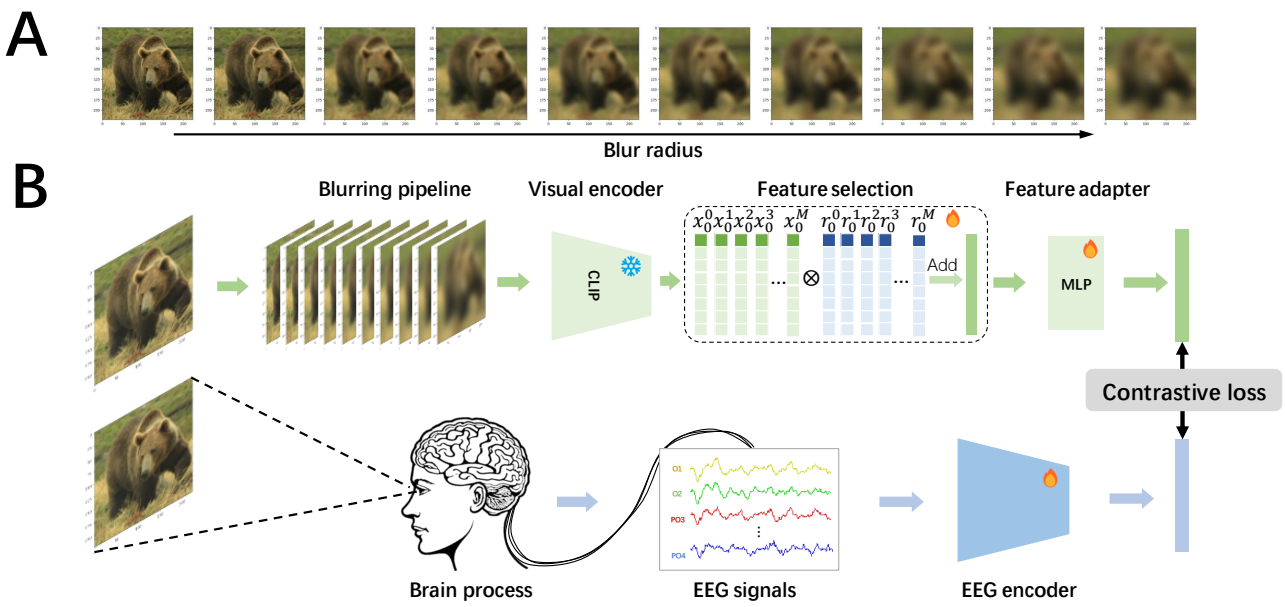

Figure 1: Figure (A) shows different blurred images, and Figure (B) shows the overall framework of EEG decoding based on visual blur perception characteristics. During training, the visual branch uses a pre-trained visual encoder to extract representations of different blurred images and performs feature selection and feature adaptation. The EEG branch uses an EEG encoder to directly obtain representations, and the training objective is to align visual representations and EEG representations using contrastive learning.

Figure 1(A). For the input image x, the processing steps are as follows:

$$x_{\text{blur}}(i,j) = \sum_{m=-r}^{r} \sum_{n=-r}^{r} x(i+m, j+n) \cdot G(m,n), \quad (1)$$

where $x(i,j)$ denotes the pixel corresponding to the ith row and jth column in the image, r is the kernel radius, the corresponding kernel size is $k = 2r+1$, and $G(m,n)$ is the weight corresponding to the point $x(i+m, j+n)$. The Gaussian kernel used is:

$$G(m,n) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{m^2+n^2}{2\sigma^2}\right) \quad (2)$$

The degree of image blurring can be controlled by adjusting two key parameters: the kernel radius r and the standard deviation σ of the Gaussian kernel.

Visual representation construction

By processing M distinct blurred images through the pre-trained visual representation model, we extracted their corresponding feature representations $F_{blur} = \{f_{blur_0}, f_{blur_1}, f_{blur_2}, \cdots, f_{blur_M}\}$, where $f_{blur_*} \in \mathbb{R}^d$, and d is the dimension corresponding to each representation. Considering the differences in blur perception characteristics among individuals, a feature selection method that performs element-wise weighted selection was adopted to construct subject-dependent pre-trained visual representations. When inputting different blurred pre-trained representations F_{blur}, the feature selection calculation is as follows:

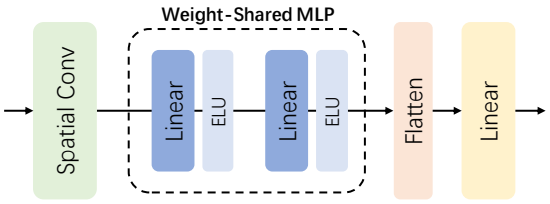

Figure 2: The structure of the EEG encoder.

$$x_{selected}(i) = \frac{\sum_{k=0}^{M} e^{R(k,i)} \cdot F_{blur}(k,i)}{\sum_{k=0}^{M} e^{R(k,i)}} \quad (3)$$

Where $x_{selected} \in \mathbb{R}^d$, $x_{selected}(i)$ is the i-th feature of the feature vector; $F_{blur}(k,i)$ is the i-th feature of the k-th blurred visual representation, and $R(k,i)$ is the learnable weight corresponding to this feature. After feature selection, a feature adaptive layer is used to map the visual representations to a semantic space similar to the EEG representations.

EEG representation construction

To investigate neural mechanisms underlying visual processing, we developed an interpretable EEG decoding architecture (Figure 2). The network consists of a spatial convolution layer, a shared-parameter MLP, a flatten layer, and a linear layer.

Motivated by functional specialization in visual cortex, we implemented constrained spatial convolution with (ch,1)

kernels to automatically select task-relevant electrodes. Kernel weights were rectified (absolute value operation) to enforce positive constraints, effectively filtering non-visual channels. Then, a MLP with shared weights is used to extract EEG information related to visual decoding from each channel. Finally, the features from all channels are flattened and mapped to a feature space similar to the visual representation of CLIP using a single linear layer. Compared to visual decoding models, EEG decoding models have a simpler structure, primarily because EEG signals represent neural electrical signals generated during visual processing in the brain, which already contain processed information.

Training and testing

We adopted a contrastive learning approach to construct a visual EEG decoding framework. The visual encoding model (CLIP) used in this paper was trained on a large amount of image-text data sets and has superior zero-shot classification performance and generalization ability. Therefore, the optimization goal of the model is to align the representations extracted from EEG with those extracted from images. Let (z_i^I, z_j^E) denote a sample pair composed of visual representation z_i^I and EEG representation z_j^E. When $i = j$, it represents a positive sample pair, and when $i \neq j$, it represents a negative sample pair. The InfoNCE loss is adopted as the optimization objective:

$$L_{I2E} = -\frac{1}{B}\sum_{i=0}^{B} log\frac{\exp(z_i^I z_i^E/\tau)}{\sum_{j=0}^{B}\exp(z_i^I z_j^E/\tau)} \tag{4}$$

$$L_{E2I} = -\frac{1}{B}\sum_{i=0}^{B} log\frac{\exp(z_i^I z_i^E/\tau)}{\sum_{j=0}^{B}\exp(z_j^I z_i^E/\tau)} \tag{5}$$

$$L_{InfoNCE} = \frac{1}{2}\left(L_{I2E} + L_{E2I}\right) \tag{6}$$

Where B represents the batch size during training, and τ is a hyperparameter that controls the temperature.

During evaluation, we employed an N-way zero-shot retrieval paradigm. Each test EEG sample was first encoded into a feature representation using the EEG encoder. Subsequently, we computed the similarity scores between this EEG representation and N candidate image representations. The final classification was determined by selecting the top-K most similar images

Experiment

Dataset and preprocessing

We used two datasets, Things-EEG (Gifford et al. 2022) and Things-MEG (Hebart et al. 2023), for testing.

Things-EEG. The Things-EEG dataset contains EEG data collected from 10 participants during a visual rapid presentation task. Each participant's data was divided into a training set and a test set. The training set includes EEG data from participants viewing 16,540 images, with each image repeated 4 times. The test set includes EEG data from participants viewing 200 images, with each image repeated 80 times. The image categories in the training set and test set do not overlap. We employed a standard preprocessing method, averaging the repeated EEG signals to reduce noise. We performed 250 Hz resampling, with sample lengths set to 1 second after the presentation of each image stimulus. After preprocessing, each participant's dataset comprised 16,540 training samples and 200 test samples.

Things-MEG. The Things-MEG dataset contains MEG data collected from 4 participants while viewing images. Each participant's dataset is divided into a training set and a test set. The training set contains MEG data corresponding to the participants viewing 22,248 images, with each image repeated once. The test set contains MEG data from the participants viewing 200 images, with each image repeated 12 times. The categories of images in the training set and test set do not overlap. We employed a standard preprocessing method to average the repeated MEG signals.

Experimental setup

All experiments were conducted using PyTorch on a GeForce 5060ti GPU, with AdamW used for parameter optimization, a learning rate of 0.0001, and a batch size of 1024. We applied 10 different levels of blurring to the images in the visual branch. The visual model utilized the RN50 model weights provided by OpenCLIP. We conducted 200-way zero-shot retrieval tests with both intra-subject and inter-subject designs. In the intra-subject tests, each subject was tested individually, with both training and testing data sourced from the same subject. In the inter-subject tests, we employed a leave-one-out method, using the training set from other subjects for model training and the testing set from the target subject for evaluation. During training, we adhered to a strict training-validation-test split. We randomly divided 0.05 of the training set as the validation set for model selection. Each test was repeated five times to ensure the stability of the results.

Results

Zero-shot retrieval. To validate the proposed method, we conducted 200-way zero-shot retrieval tests on the Things-EEG dataset and compared it with some current benchmark methods: BraVL (Du et al. 2023), NICE (Song et al. 2024), ATM-S (Li et al. 2024), MB2C (Wei et al. 2024),VE-SDN (Chen et al. 2024), CognitionCapturer (Zhang et al. 2025) and UBP (Wu et al. 2025). The intra-subject results are shown in Table 1. As shown in the table, our method significantly improves zero-shot retrieval accuracy, achieving an average accuracy of 80.0% (Top-1) and 96.9% (Top-5) across all subjects. Compared to the state-of-the-art UBP method, the Top-1 and Top-5 accuracy increased by 29.1% and 17.2%, respectively. The inter-subject results are presented in Table 2. Our method still outperformed existing approaches, with improvements of 7.6% (Top-1) and 14.6% (Top-5) over UBP. The smaller margin compared to intra-subject tests is attributed to individual differences among subjects, which challenge model generalization to unseen data.

To further verify the generality of the method proposed in this paper, we conducted another test on the Things-

Method	BraVL		NICE		MB2C		ATM-S		VE-SDN		CognitionCapturer		UBP		Ours	
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
Sub 1	6.1	17.9	13.3	40.2	23.7	56.3	25.6	60.4	32.6	63.7	27.2	59.5	41.2	70.5	**81.9**	**96.8**
Sub 2	4.9	14.9	12.1	36.1	22.7	50.5	22.0	54.5	34.4	69.9	28.7	57.0	51.2	80.9	**81.7**	**96.8**
Sub 3	5.6	17.4	15.3	39.6	26.3	60.2	25.0	62.4	38.7	73.5	37.2	66.1	51.2	82.0	**78.3**	**96.0**
Sub 4	5.0	15.1	15.9	49.0	34.8	67.0	31.4	60.9	39.8	72.0	37.7	63.2	51.1	76.9	**76.9**	**97.2**
Sub 5	4.0	13.4	9.8	34.4	21.3	53.0	12.9	43.0	29.4	58.6	21.8	47.8	42.2	72.8	**71.4**	**93.7**
Sub 6	6.0	18.2	14.2	42.4	31.0	62.3	21.3	51.1	34.5	68.8	31.6	58.1	57.5	83.5	**84.3**	**98.7**
Sub 7	6.5	20.4	17.9	43.6	25.0	54.8	30.5	61.5	34.5	68.3	32.8	59.6	49.0	79.9	**78.2**	**96.7**
Sub 8	8.8	23.7	18.2	50.2	39.0	69.3	38.8	72.0	49.3	79.8	47.6	73.5	58.6	85.8	**84.5**	**98.1**
Sub 9	4.3	14.0	14.4	38.7	27.5	59.3	34.4	51.5	39.0	69.6	33.4	57.6	45.1	76.2	**78.7**	**96.8**
Sub 10	7.0	19.7	16.0	42.8	33.2	70.8	29.1	63.5	39.8	75.3	35.1	63.6	61.5	88.2	**84.1**	**98.1**
Avg	5.8	17.5	14.7	41.7	28.4	60.4	28.5	60.4	37.2	69.9	33.3	60.6	50.9	79.7	**80.0**	**96.9**

Table 1: Intra-subject test on Things-EEG dataset (train and test on one subject).

Method	BraVL		NICE		NICE-SA		NICE-GA		ATM-S		MB2C		UBP		Ours	
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
Sub 1	2.3	8.0	7.6	22.8	7.0	22.6	5.9	21.4	10.5	26.8	10.5	28.2	11.5	29.7	**27.3**	**54.9**
Sub 2	1.5	6.3	5.9	20.5	6.6	23.2	6.4	22.7	7.1	24.8	11.3	32.8	15.5	40.0	**32.0**	**63.0**
Sub 3	1.4	5.9	6.0	22.3	7.5	23.7	5.5	20.1	11.9	33.8	8.8	27.7	9.8	27.0	**11.1**	**34.5**
Sub 4	1.7	6.7	6.3	20.7	5.4	21.4	6.1	21.0	14.7	39.4	13.7	33.5	13.0	32.3	**18.6**	**46.0**
Sub 5	1.5	5.6	4.4	18.3	6.4	22.2	4.7	19.5	7.0	23.9	10.7	27.5	8.8	33.8	**16.9**	**40.7**
Sub 6	1.8	7.2	5.6	22.2	7.5	22.5	6.2	22.5	11.1	35.8	12.2	33.2	11.7	31.0	**16.1**	**43.2**
Sub 7	2.1	8.1	5.6	19.7	3.8	19.1	5.9	19.1	16.1	43.5	11.5	31.8	10.2	23.8	**16.0**	**45.4**
Sub 8	2.2	7.6	6.3	22.0	8.5	24.4	7.3	25.3	15.0	40.3	12.0	32.2	12.2	32.2	**18.3**	**49.0**
Sub 9	1.6	6.4	5.7	17.6	7.4	22.3	4.8	18.3	4.9	22.7	12.2	31.3	15.5	40.5	**13.6**	**41.2**
Sub 10	2.3	8.5	8.4	28.3	9.8	29.6	6.2	26.3	20.5	46.5	16.2	42.2	16.0	43.5	**30.1**	**62.1**
Avg	1.8	7.0	6.2	21.4	7.0	23.1	5.9	21.6	11.8	33.7	11.9	32.0	12.4	33.4	**20.0**	**48.0**

Table 2: Inter-subject test on Things-EEG dataset (leave one subject out for test).

Method	Sub 1 Top1	Sub 2 Top1	Sub 3 Top1	Sub 4 Top1	Avg Top1	Avg Top5
Intra-subject: train and test on one subject						
NICE	9.6	18.5	14.2	9	12.8	36.0
NICE-SA	9.8	18.6	10.5	11.7	12.7	35.0
NICE-GA	8.7	21.8	16.5	10.3	14.3	42.3
UBP	15.0	46	27.3	18.5	26.7	55.2
Ours	**25.4**	**78.1**	**37.3**	**35.3**	**44.0**	**72.0**
Inter-subject: leave one subject out for test						
UBP	2.0	1.5	2.7	2.5	2.2	10.4
Ours	**2.9**	**7.7**	**5.8**	**4.7**	**5.3**	**15.9**

Table 3: Intra-subject and inter-subject test on Things-MEG dataset.

MEG dataset. Similarly, we conducted intra-subject and inter-subject 200-way retrieval tests. The results are shown in Table 3. In the intra-subject test, the method proposed in this paper achieved an average classification accuracy of 44.03% and 72.05% for Top-1 and Top-5, respectively, which is an improvement of 17.3% and 16.8% over the current state-of-the-art UBP method. In inter-subject tests, the proposed method achieved average classification accuracy rates of 5.28% and 15.85% for Top-1 and Top-5, respectively, representing improvements of 3.08% and 5.45% over the current state-of-the-art UBP method. This further proves the effectiveness and universality of the method proposed in this paper.

Method	Top - 1	Top - 3	Top - 5
w/o Blurring pipeline	57.27±5.09	79.91±4.83	86.95±4.18
w/o Feature adapter	56.21±6.86	77.87±5.77	85.68±4.43
w/o Spatial Conv	67.04±6.11	85.36±4.65	91.45±3.27
w/ All	**80.00±4.19**	**93.92±2.30**	**96.89±1.43**

Table 4: Comparison of zero-shot retrieval accuracy (%) for different variants on the Things-EEG dataset.

Ablation study. To validate the effectiveness of each module in our method, we conducted ablation experiments on the Things-EEG dataset evaluating three variants: without the blurring pipeline, without spatial convolution, and without the feature adapter. As shown in Table 4, all variants exhibited performance degradation, with the spatial convolution removal reducing Top-1/3/5 accuracy by 12.96%/8.56%/5.44%, while the feature adapter and blurring pipeline ablations caused more substantial declines of 23.79%/16.05%/11.21% and 22.73%/14.01%/9.94% respectively. These results demonstrate that both the blurring

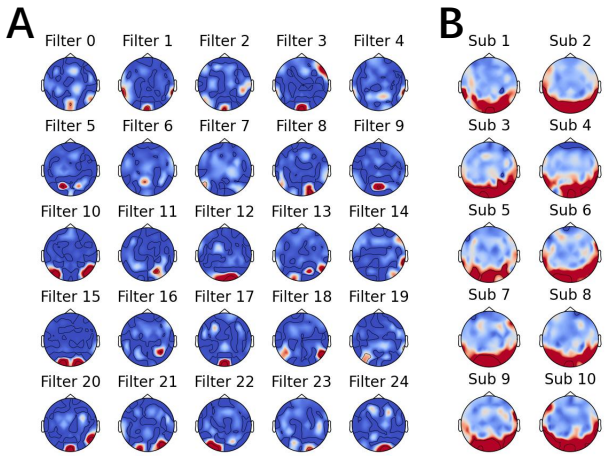

Figure 3: Visualization of spatial filters. Figure (A) shows the distribution of all spatial filters for subject 1 on the brain topography map. Figure (B) shows the distribution of all spatial filters for each subject after integration on the brain topography map, where red represents high weight and blue represents low weight.

pipeline and feature adaptation layer play critical roles in EEG visual decoding by effectively reducing representational discrepancies between images and EEG signals, while the spatial convolution's importance suggests that electrode selection is crucial as not all channels contribute equally to decoding performance, with redundant information potentially impairing model accuracy.

Visualization of spatial filter. The kernels in the spatial convolution of the EEG encoding model were visualized, as presented in Figure 3. Each convolution kernel corresponds to a distinct spatial filter. Non-negative constraints were imposed on the spatial filters, thereby restricting each filter's function to selecting electrodes engaged in visual decoding. In Figure 3(A), we visualized each spatial filter using a brain topography map. It can be seen that each spatial filter focuses only on EEG information from specific brain regions, enabling effective channel selection. In Figure 3(B), to identify the brain regions used in visual decoding, we combined all spatial filters using the maximum operation. It can be seen that the regions primarily involved in EEG visual decoding are concentrated in the occipital and temporal lobes of the brain, which correspond to the ventral pathway in visual processing.

Visualization of activation maps with different levels of blurriness. To further investigate the impact of varying blurring levels on visual decoding, the activation regions of EEG representations in images were visualized using the method proposed by (Li et al. 2025), previously applied in the CLIP study to examine language representation activations. In this study, language representations were replaced with EEG representations to examine the information prioritized by the brain during visual processing. The results are presented in Figure 4, where 'Fusion' denotes the activation map following feature selection integration, while the remaining maps correspond to individual blurred representations. The 'Fusion' results demonstrate that EEG-derived representations effectively emphasize the image's main subject, facilitating human perception of key visual information. Furthermore, the activation maps reveal that in low-blur images, EEG representations primarily localize to regions associated with the main subject. With increasing blurriness, the activated regions expand progressively and shift toward background areas. This observation is consistent with the brain's tendency to discriminate foreground from background based on blurriness, where higher blurriness typically corresponds to background regions.

Image retrieval visualization. To verify whether the representations extracted by the EEG encoding model contain vision-related information, we visualized the similarity between the EEG representations and the visual representations. as shown in Figure 5. In Figure5 (A), we first roughly divided the test data into six categories: animals, food, vehicles, tools, clothes, and household items, and constructed a similarity matrix between the EEG representations and visual representations for all 200 categories. The figure shows the average similarity matrix of 10 participants. As can be seen from the figure, EEG representations and their corresponding visual representations generally have the highest similarity, and EEG representations and visual representations within the same category have high similarity, proving that the method proposed in this paper can effectively extract visual-related information from EEG. In Figure5 (B), the images corresponding to the five visual representations with the highest similarity to the EEG representations are visualized. It can be observed that the images corresponding to high-similarity visual representations also exhibit certain similarities in color, background, and semantic content with the images viewed by the participants.

Impact of repetition. To reduce the impact of low signal-to-noise ratio in EEG, the current EEG visual decoding work generally uses the method of averaging the EEG signals of repeated stimuli. To study the impact of different repetition times in averaging, we conducted experiments with different repetition times on the test set. The results are shown in Figure 6. From the results, we can see that as the repetition times increase, the classification performance gradually improves, which is consistent with other papers. However, unlike other methods, our proposed method achieves zero-shot retrieval accuracy of Top-1 13.8%, Top-3 27.95%, and Top-5 35.95% even without averaging. When the repetition time is 10, the proposed method achieves Top-1 59.55%, Top-3 79.9%, and Top-5 87.25%, surpassing the classification performance of the current best method, UBP, at 80 repetitions.

Image generation visualization. We also used the features extracted from EEG to guide the diffusion model for image generation, attempting to reconstruct the images seen from EEG. We adopted the method proposed in the paper (Li et al. 2024), using the features extracted by the EEG encoder to train a diffusion model, and then using the features output by the diffusion model as prompts for SDXL+IP-Adapter for image generation. The results are shown in Fig-

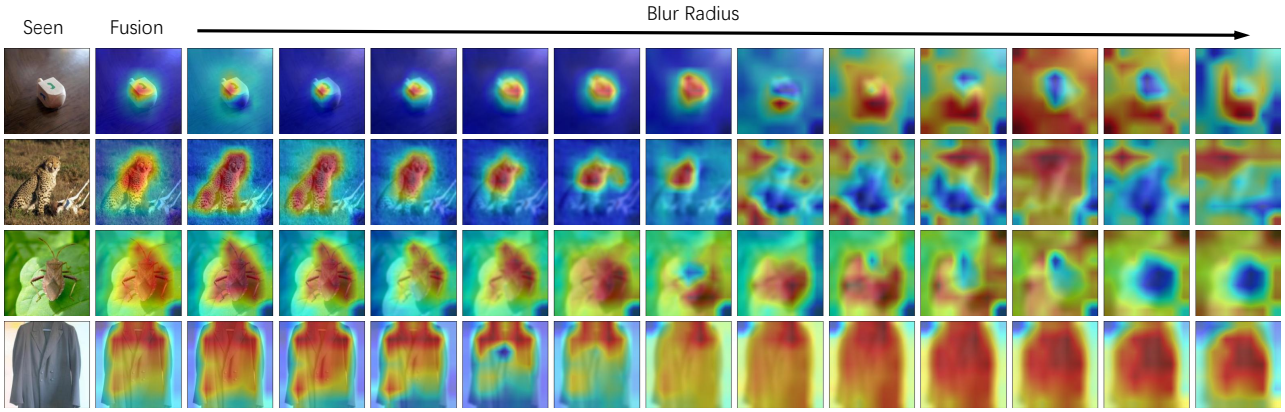

Figure 4: EEG activation maps for different blurred images, with red representing the image regions corresponding to the EEG representations. As can be seen from the figure, in low-blur images, the regions corresponding to the EEG representations are typically local regions related to the main subject of the image. As blur increases, the activated regions gradually expand and shift to background regions.

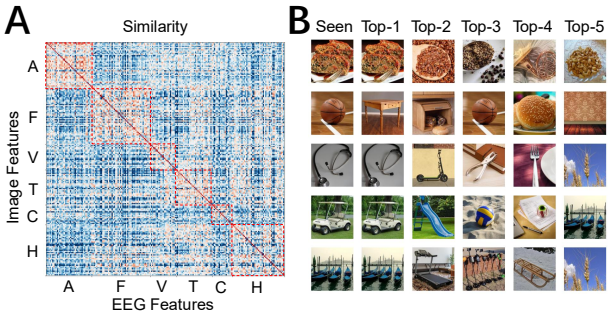

Figure 5: Figure (A) shows the similarity visualization between EEG features and image features. The test samples are roughly divided into six categories: Animals, Food, Vehicle, Tools, Clothes, and Household items, represented by A, F, V, T, C, and H respectively. Figure (B) shows the visualization of the 5-shot image retrieval results.

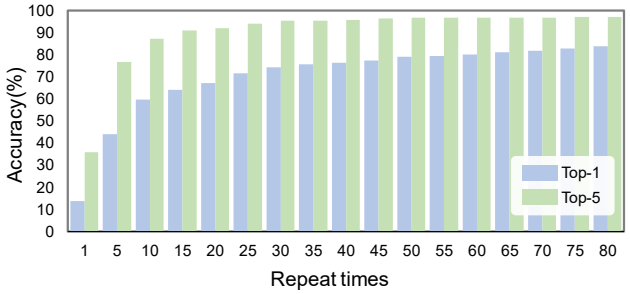

Figure 6: Comparison of the number of repetitions in the test phase.

Figure 7: EEG-guided image generation visualization. Figures (A), (B), and (C) are examples of high-quality, medium-quality, and low-quality images, respectively.

ure 7. In Figure 7(A), the generated images match the seen images in terms of subject category, color, and structure. In Figure 7(B), most of the generated images match in terms of structure and color, while in Figure7(C), the generated images only match in terms of color and structure.

Conclusion

In this work, we propose an EEG-based visual decoding framework that leverages blur perception characteristics. To bridge the gap between external visual stimuli and internal human perception, we introduce a blurring pipeline that simulates neural processing of object recognition. Additionally, we design a personalized feature selection module to construct subject-specific visual representations. For EEG signal processing, we developed an interpretable encoding architecture that incorporates a constrained spatial convolutional layer to perform vision-relevant channel selection. Experiments on the Things-EEG dataset show a significant advantage in zero-shot retrieval tasks, achieving a top-1 accuracy of 80% and a top-5 accuracy of 96.9% in intra-subject tests, which are 29.1% and 17.2% higher than the current state-of-the-art method, further exploring the potential of EEG visual decoding. By analyzing EEG activation maps across varying blur levels, we observed that distinct blur intensities elicit activity in different perceptual regions and spatial extents. These results not only validate our proposed method but also demonstrate the functional role of blur processing in visual perception, and may provide insights into brain visual processing mechanisms.

Acknowledgments

This work was supported in part by special funds of the National Natural Science Foundation of China under Grant 32441112.

References

Allen, E. J.; St-Yves, G.; Wu, Y.; Breedlove, J. L.; Prince, J. S.; Dowdle, L. T.; Nau, M.; Caron, B.; Pestilli, F.; Charest, I.; et al. 2022. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature neuroscience*, 25(1): 116–126.

Beauchamp, M. S.; Oswalt, D.; Sun, P.; Foster, B. L.; Magnotti, J. F.; Niketeghad, S.; Pouratian, N.; Bosking, W. H.; and Yoshor, D. 2020. Dynamic stimulation of visual cortex produces form vision in sighted and blind humans. *Cell*, 181(4): 774–783.

Bharadwaj, S. R.; and Schor, C. M. 2006. Dynamic control of ocular disaccommodation: first and second-order dynamics. *Vision Research*, 46(6-7): 1019–1037.

Burge, J.; and Geisler, W. S. 2011. Optimal defocus estimation in individual natural images. *Proceedings of the National Academy of Sciences*, 108(40): 16849–16854.

Chang, L.; Bao, P.; and Tsao, D. Y. 2017. The Representation of Colored Objects in Macaque Color Patches. *Nature Communications*, 8(1): 2064.

Chen, H.; He, L.; Liu, Y.; and Yang, L. 2024. Visual neural decoding via improved visual-EEG semantic consistency. *arXiv preprint arXiv:2408.06788*.

Chen, X.; Wang, F.; Fernandez, E.; and Roelfsema, P. R. 2020. Shape perception via a high-channel-count neuroprosthesis in monkey visual cortex. *Science*, 370(6521): 1191–1196.

Chen, Z.; Qing, J.; Xiang, T.; Yue, W. L.; and Zhou, J. H. 2023. 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*, 22710–22720.

Curcio, C. A.; Sloan, K. R.; Kalina, R. E.; and Hendrickson, A. E. 1990. Human photoreceptor topography. *Journal of comparative neurology*, 292(4): 497–523.

Di Russo, F.; Martínez, A.; Sereno, M. I.; Pitzalis, S.; and Hillyard, S. A. 2002. Cortical sources of the early components of the visual evoked potential. *Human brain mapping*, 15(2): 95–111.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Du, C.; Fu, K.; Li, J.; and He, H. 2023. Decoding visual neural representations by multimodal learning of brain-visual-linguistic features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9): 10760–10777.

Fang, T.; Zheng, Q.; and Pan, G. 2023. Alleviating the semantic gap for generalized fmri-to-image reconstruction. *Advances in Neural Information Processing Systems*, 36: 15096–15107.

Fernández, E.; Alfaro, A.; Soto-Sánchez, C.; Gonzalez-Lopez, P.; Lozano, A. M.; Peña, S.; Grima, M. D.; Rodil, A.; Gómez, B.; Chen, X.; et al. 2021. Visual percepts evoked with an intracortical 96-channel microelectrode array inserted in human occipital cortex. *The Journal of clinical investigation*, 131(23).

Gifford, A. T.; Dwivedi, K.; Roig, G.; and Cichy, R. M. 2022. A large and rich EEG dataset for modeling human visual object recognition. *NeuroImage*, 264: 119754.

Grootswagers, T.; Zhou, I.; Robinson, A. K.; Hebart, M. N.; and Carlson, T. A. 2022. Human EEG recordings for 1,854 concepts presented in rapid serial visual presentation streams. *Scientific Data*, 9(1): 3.

Güçlü, U.; and Van Gerven, M. A. 2015. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, 35(27): 10005–10014.

Hartmann, K. G.; Schirrmeister, R. T.; and Ball, T. 2018. EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals. *arXiv preprint arXiv:1806.01875*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

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

Horikawa, T.; and Kamitani, Y. 2017. Generic decoding of seen and imagined objects using hierarchical visual features. *Nature communications*, 8(1): 15037.

Jin, J. Z.; Weng, C.; Yeh, C.-I.; Gordon, J. A.; Ruthazer, E. S.; Stryker, M. P.; Swadlow, H. A.; and Alonso, J.-M. 2008. On and off Domains of Geniculate Afferents in Cat Primary Visual Cortex. *Nature Neuroscience*, 11(1): 88–94.

Khadir, A.; Maghareh, M.; Sasani Ghamsari, S.; and Beigzadeh, B. 2023. Brain activity characteristics of RGB stimulus: an EEG study. *Scientific Reports*, 13(1): 18988.

Khaligh-Razavi, S.-M.; and Kriegeskorte, N. 2014. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS computational biology*, 10(11): e1003915.

Kolb, H.; Fernandez, E.; Jones, B.; and Nelson, R. 1995. Webvision: the organization of the retina and visual system [Internet].

Li, D.; Wei, C.; Li, S.; Zou, J.; Qin, H.; and Liu, Q. 2024. Visual Decoding and Reconstruction via EEG Embeddings with Guided Diffusion. arXiv:2403.07721.

Li, Y.; Wang, H.; Duan, Y.; Zhang, J.; and Li, X. 2025. A closer look at the explainability of Contrastive language-image pre-training. *Pattern Recognition*, 162: 111409.

Lin, S.; Sprague, T.; and Singh, A. K. 2022. Mind reader: Reconstructing complex images from brain activities. *Advances in Neural Information Processing Systems*, 35: 29624–29636.

Liu, Y.; Ma, Y.; Zhou, W.; Zhu, G.; and Zheng, N. 2023. Brainclip: Bridging brain and visual-linguistic representation via clip for generic natural visual stimulus decoding. *arXiv preprint arXiv:2302.12971*.

Oquab, M.; Darcet, T.; Moutakanni, T.; Vo, H. V.; Szafraniec, M.; Khalidov, V.; Fernandez, P.; Haziza, D.; Massa, F.; El-Nouby, A.; Howes, R.; Huang, P.-Y.; Xu, H.; Sharma, V.; Li, S.-W.; Galuba, W.; Rabbat, M.; Assran, M.; Ballas, N.; Synnaeve, G.; Misra, I.; Jegou, H.; Mairal, J.; Labatut, P.; Joulin, A.; and Bojanowski, P. 2023. DINOv2: Learning Robust Visual Features without Supervision.

Osterberg, G. A. 1935. Topography of the layer of rods and cones in the human retina. *Acta ophthalmologica*.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PmLR.

Roe, A. W.; Chelazzi, L.; Connor, C. E.; Conway, B. R.; Fujita, I.; Gallant, J. L.; Lu, H.; and Vanduffel, W. 2012. Toward a Unified Theory of Visual Area V4. *Neuron*, 74(1): 12–29.

Salvaris, M.; and Sepulveda, F. 2009. Visual modifications on the P300 speller BCI paradigm. *Journal of neural engineering*, 6(4): 046011.

Schaeffel, F.; Wilhelm, H.; and Zrenner, E. 1993. Interindividual variability in the dynamics of natural accommodation in humans: relation to age and refractive errors. *The Journal of Physiology*, 461(1): 301–320.

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

Singh, P.; Pandey, P.; Miyapuram, K.; and Raman, S. 2023. EEG2IMAGE: image reconstruction from EEG brain signals. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Smith, S. M. 2004. Overview of fMRI analysis. *The British Journal of Radiology*, 77(suppl_2): S167–S175.

Song, Y.; Liu, B.; Li, X.; Shi, N.; Wang, Y.; and Gao, X. 2024. Decoding Natural Images from EEG for Object Recognition. In *International Conference on Learning Representations*.

Song, Y.; Wang, Y.; He, H.; and Gao, X. 2025. Recognizing Natural Images From EEG With Language-Guided Contrastive Learning. *IEEE Transactions on Neural Networks and Learning Systems*.

Spampinato, C.; Palazzo, S.; Kavasidis, I.; Giordano, D.; Souly, N.; and Shah, M. 2017. Deep learning human mind for automated visual classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6809–6817.

Sprague, W. W.; Cooper, E. A.; Reissier, S.; Yellapragada, B.; and Banks, M. S. 2016. The natural statistics of blur. *Journal of Vision*, 16(10): 23–23.

Tanigawa, H.; Lu, H. D.; and Roe, A. W. 2010. Functional Organization for Color and Orientation in Macaque V4. *Nature Neuroscience*, 13(12): 1542–1548.

Tootell, R. B. H.; Silverman, M. S.; De Valois, R. L.; and Jacobs, G. H. 1983. Functional Organization of the Second Cortical Visual Area in Primates. *Science*, 220(4598): 737–739.

Vera-Diaz, F. A.; Woods, R. L.; and Peli, E. 2010. Shape and individual variability of the blur adaptation curve. *Vision Research*, 50(15): 1452–1461.

Wei, Y.; Cao, L.; Li, H.; and Dong, Y. 2024. MB2C: Multimodal Bidirectional Cycle Consistency for Learning Robust Visual Neural Representations. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 8992–9000. Melbourne VIC Australia: ACM. ISBN 979-8-4007-0686-8.

Wu, H.; Li, Q.; Zhang, C.; He, Z.; and Ying, X. 2025. Bridging the Vision-Brain Gap with an Uncertainty-Aware Blur Prior. arXiv:2503.04207.

Zhang, K.; He, L.; Jiang, X.; Lu, W.; Wang, D.; and Gao, X. 2025. CognitionCapturer: Decoding Visual Stimuli From Human EEG Signal With Multimodal Information. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 14486–14493.

Zhu, D.; Bieger, J.; Garcia Molina, G.; and Aarts, R. M. 2010. A survey of stimulation methods used in SSVEP-based BCIs. *Computational intelligence and neuroscience*, 2010(1): 702357.