EEG-Face: A Facial-Image Stimulated EEG Data-Set for Analysis of Brain Perceived Multimedia

Wuxia Zhang
Shenzhen University
Shenzhen, China
zhangwuxia2020@email.szu.edu.cn

Yang Xin
Shenzhen University
Shenzhen, China
2205278003@mails.szu.edu.cn

Shibo Lv
Shenzhen University
Shenzhen, China
2400101082@mails.szu.edu.cn

Xin Zhang
Shenzhen University
Shenzhen, China
2250271011@email.szu.edu.cn

Xiang Zhong
Shenzhen University
Shenzhen, China
zhongxiang2023@email.szu.edu.cn

Jianmin Jiang*
Shenzhen University
Shenzhen, China
jianmin.jiang@szu.edu.cn

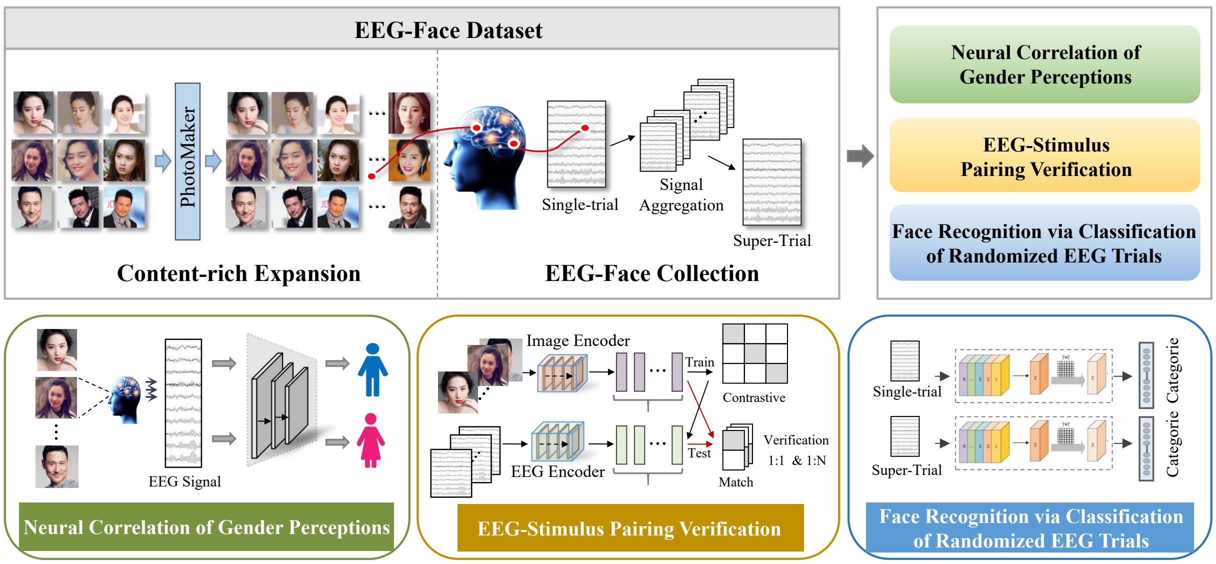


Figure 1: Illustration of architectural construction of EEG-Face and its usability tests.

Abstract

Over recent years, EEG-based brain decoding of perceived multimedia is emerging to be an important multidisciplinary research area. Lack of data sets with multimedia stimuli, however, presents a significant challenge for its further advancement. In this paper, we establish a facial-image stimulated EEG dataset, named as EEG-Face, to address the challenge and provide a crucial support for relevant research, such as brain-computer interface (BCI), face recognition via brain-perceived EEGs, and multimedia content analysis via brain perception activities. As facial images not only distinguish between genders but also dive deeper into individual differences, our proposed EEG-Face provides larger scope, more focus, and greater potential for dedicated research on brain perception of human faces. As shown in Figure 1, the proposed EEG-Face essentially consists of 20,000 brain responded EEG trials stimulated with 40 individual faces, all of whom are Chinese film stars. Following the establishment of the dataset, a range of experiments over EEG-Face is carried out to demonstrate its usability and feasibility, which include: (i) neural correlation of gender perceptions; (ii) EEG-Stimulus pairing verification; and (iii) face recognition via classification of randomized EEG trials. The dataset and the codes for all reported experiments are available from: https://github.com/eeg-wx2024/EEG-Face.

*Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
MM '25, Dublin, Ireland
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3758223

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI).**

Keywords

EEG, Face Perception, EEG-Face, EEG classification

ACM Reference Format:
Wuxia Zhang, Yang Xin, Shibo Lv, Xin Zhang, Xiang Zhong, and Jianmin Jiang. 2025. EEG-Face: A Facial-Image Stimulated EEG Data-Set for Analysis of Brain Perceived Multimedia. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25).* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3746027.3758223

1 Introduction

Over recent years, decoding perceptual information directly from brain activities has increasingly attracted interests across multi-disciplinary research fields, primarily including neuroscience and computer science [5, 6, 23, 25, 26, 28, 31]. While most research reported in the literature is primarily carried out via fMRI-based approaches [24, 32, 34], its inherent high cost, low portability and operational complexity limit its applications in many real-world scenarios. In contrast, electroencephalography (EEG) has shown great potential in brain state analysis, brain-computer interface (BCI) and other fields due to its advantages of high temporal resolution, portable equipment and low cost [3, 10, 12, 27]. Despite the great promise of EEGs, accurate brain decoding directly from EEG signals, especially decoding complex perceptual multimedia content, still faces severe challenges [1, 17].

EEG-based brain decoding of object images was pioneered by Standford researchers in 2015 [13], where an EEG dataset stimulated by object images with 6 classes was published, and a classification rate of 28.87% was also reported. This work represents the first attempt on EEG-based decoding of brain-perceived multimedia.

In ACM-MM'2017, Florida researchers published another dataset, EEG-ImageNet, via block-based stimulations, where all images from the same class are presented to the test subject in a deterministic order [14]. With such deterministic EEG trials, which are generated from 2000 natural images directly selected from ImageNet with 40 classes, Spampinato et al. further reported a classification rate of 82.9% via a RNN-based deep model [30]. This work is seen as an important milestone towards EEG-based brain decoding of perceived natural images [7, 23, 35]. Following the questioning that EEG-ImageNet may suffer from problems such as signal drift and filtering bias, however, the dataset is further refined and the corresponding classification rate is reduced to 21.8%, yet a new deep model, EEGChannelNet, is introduced and reported to achieve a classification rate at 48.1% [23].

In CVPR2021, Purdue researchers argued that the block-based deterministic trials in EEG-ImageNet may suffer from the confound of blocking effect, and thus recreated EEGImageNet, where natural images are presented to the test subject in randomized orders. Over such randomized EEG trials, classification of the same 40 classes by a number of existing classifiers is significantly degraded, and the maximum top-1 classification rate is dropped to 7% [1].

Recently, some studies have begun to explore multimodal decoding methods via exploitation of stimuli images to assist brain decoding from EEG signals [16, 28]. Based on the dataset established by Gifford et al. [8], a comparative learning framework is proposed to capture the correlation between the features of stimuli images and the representation of EEGs, leading to a new paradigm of brain decoding via classifying stimuli-EEG pairings, where the top-1 classification rate is reported to reach 15.6%.

Given the fact that extensive research has been reported across the areas of multimedia, computer vision, and computer graphics over topics of face recognition, generation of talking heads, and cognitive analysis of facial emotions etc [20], understanding the neural mechanisms of their perceptions and representations is paramount for advancing computational intelligence technologies, such as integration of artificial intelligence with human brain intelligence. To this end, introduction of dedicated facial stimuli for EEGs will evoke more stable and distinguishable neural response patterns, and thus enable new research initiatives over EEG-based decoding from brain perception of facial images. Yet existing EEG datasets are limited to visual stimuli with general concepts, leaving a significant gap to forward the designated research field.

To address the aforementioned gap, we propose to construct a new EEG dataset, EEG-Face, in this paper to open up a range of research possibilities around the essential topic: computational interpretation of how human brains perceive faces, a critical category of multimedia information. We highlight our contributions as follows:

- EEG-Face comprises 20,000 high-density (128 electrodes/ channels) and randomized EEG trials, which are recorded from a subject while viewing a large and diverse collection of 20,000 face images across 40 distinct and gender-balanced identities.
- To sustain content richness and maximize the potential and scope for further research, we applied a diffusion model (i.e. PhotoMaker [19]) to generate a range of new facial images with different styles and appearances for each individual, significantly expanding its varieties and diversities without incurring any changes over its identity.
- Over EEG-Face extensive experiments are carried out to demonstrate its great potential for serving a range of multi-disciplinary research objectives, which can be highlighted as: (i) to expand the existing research on multimedia facial processing into EEG-based brain decoding of human face perceptions; (ii) to foster the development and rigorous evaluation of novel EEG decoding algorithms for recognition of facial identities and other perceived attributes; (iii) to pioneer new research initiatives for analyzing and understanding multimedia content perception, such as visualization of brain perceived facial images etc. (iv) To inspire new generation of multimedia learning models via bridging the gap between artificial intelligence (AI) and brain intelligence (BI).

2 Related Work

Over the emerging research area on EEG-based brain decoding of perceived multimedia, existing datasets have experienced two stages of visual stimuli constructions, i.e. from single object images to natural images with multi-objects and semantics. While the former has typical single object set in clattered background, the latter tends to be closer to the practical brain perception of the real world. Specific details are summarized as follows.

The first EEG dataset to pioneer the work on EEG-based decoding of single object perception is established by Stanford researchers in 2015 [13]. The dataset contains 6 class labels, each of which has 12 images, totaling 72 stimulation images. In this work, Kaneshiro et al.

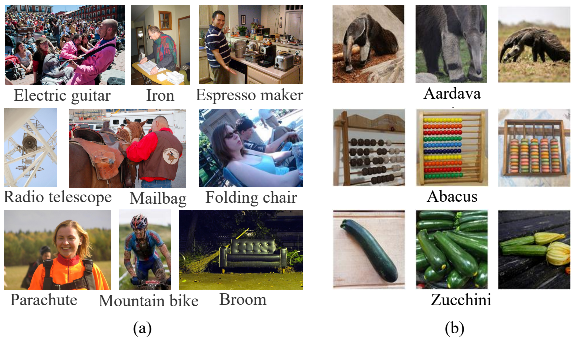


Figure 2: Illustration of representative stimulation images: (a) samples for EEG-ImageNet; (b) samples for EEG data set by Gifford et al. [8]

[13] integrated single-trial EEG classification with Representational Similarity Analysis (RSA) to validate that both object categories (6-way) and specific instances (72-way) could be successfully decoded solely from single-trial EEG signals, with accuracy significantly exceeding chance levels.

To explore the perception of more images and make it closer to the real-world scenarios of brain perception, Spampinato et al. [30] published the second dataset, EEG-ImageNet, with a total number of 2000 images that are organized in 40 classes, and each class contains 50 natural images directly selected from ImageNet. Figure 2 (a) illustrates some representative samples of such natural images together with their class labels. Using an EEG system with 128 channels, the EEG collection protocol is designed as such that all images are divided into 40 blocks, each of which consists of 50 images from the same class. Brain stimulation is arranged in a sequence of 40 blocks, and within each block, images are presented to the subject in a deterministic order to leverage the complexity of natural image perception. A rapid event design is employed, in which the subject views each image for 0.5 seconds and continues until the end of viewing each block. To increase the diversity, 6 subjects are arranged to view all stimuli images and thus 6 sets of 2000 deterministic EEG trials are made available. Associated with the dataset, Spampinato et al. also reported two types of exploratory research, including: (i) multi-class (40-way) classification of EEG trials via deep learning models [30] to recognize the brain perceived class concepts (labels); (ii) visualizing the decoded class concepts (labels) via generative learning and GAN-based image generations, pioneering exploratory research on mind-reading via computational intelligence [30].

By addressing the potential issue that EEG-ImageNet may suffer from the confounds of blocking effect, Ahmed et al. [1] recreated EEG-ImageNet, which features in the following changes: (i) the total number of images is increased to 40,000, and thus each class has 1000 sample images, 20 times more than the original EEG-ImageNet; (ii) For EEG recording, all 40,000 images are divided into 100 sets, and each set contains 400 images equally spread over all 40 classes; (iii) A single male adult is arranged to view all the 100 sets of images and EEGs are collected in 10 sessions in the run of {10 sets, 8 sets, 10 sets, 11 sets, 11 sets, 10 sets, 10 sets, 10 sets, 10 sets, 10 sets}; (iv) with a 96-channel BioSemi EEG system, the rapid-event design is employed and the selection of all images to be viewed by the subject is randomized. For the convenience of description, we refer to this updated dataset as EEG-ImageNet with randomized trials, and the original dataset as EEG-ImageNet with deterministic trials.

In 2022, Gifford et al. published the third EEG dataset to support data-driven computational neuroscience in modeling human object recognition [8]. The dataset is collected with 10 subjects, to whom both the training images and test images are presented in a pseudo-randomized order. In total, the training set contains 1654 concepts, each of which includes 10 object images, totaling 16,540 images, and the test set contains 200 concepts, each of which is represented by 1 object image. Figure 2 (b) illustrates some representative samples of the object images. Using a rapid serial visual presentation (RSVP), all images are presented to the subject for 100 ms, followed by 100 ms blanking, and EEGs are collected with 63 channels at a 1000 Hz sampling rate. While the collection of EEGs are replicated 4 times over the training set, collection of EEGs over the test set is replicated 8 times to provide a richer EEG dataset for research purposes. Compared with EEG-ImageNet, this dataset is not suitable for direct classification of EEG trials to recognize the relevant semantics (i.e. class labels). This is due to the fact that: (i) lack of training EEG signals since each concept only contains 10 images to stimulate brains; (ii) there are too many concepts, 1654 in total, making it difficult to classify even over the stimulating images. As a result, relevant research reported in the literature is limited to classifying brain perception via stimuli-EEG pairings rather than EEGs alone [28].

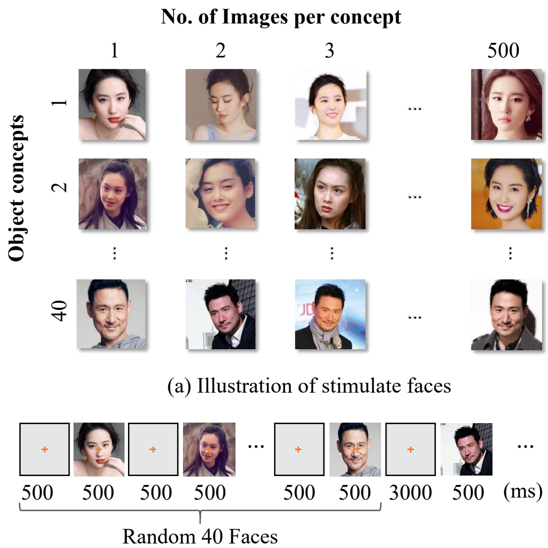
3 The Proposed EEG-Face

3.1 Establishment of the Core Stimuli Set

Given the limited availability of original photographs from public sources, we propose to establish a core stimuli set first by collecting 2000 facial images for 40 individuals, each of which contains 50 faces with different emotion and appearances. To facilitate the best possible flexibility, a number of selection criteria is also set up, which includes high resolution, clear viewing angles, no occlusion, neutral or no-strong emotional expressions. Correspondingly, we prioritize those publicly familiar faces, such as celebrities or well-known film stars, to ensure straightforward identification by the subject. To maximize the scope of research and achieve the best possible usability, we maintain a gender balance of 20/20 between male and female faces to facilitate a broad range of research activities, such as face recognition and gender identification etc. All EEG trials were collected from a single healthy adult male subject.

3.2 Content-rich Expansion

To address the possible issue that the core stimuli set with 50 photographs per individual maybe insufficient in both quantity and diversity for larger scale training of deep models, we propose to employ a diffusion model based technique (PhotoMaker [19]) to implement a scheme of content-rich expansion for the core stimuli set. Such expansion scheme increases the number of original images per identity from 50 to 500, bringing the total number of stimuli to 20,000, where each identity is verified by manual review to guarantee its authenticity. The essential advantage of this method lies



(a) Illustration of stimulate faces

(b) Illustration of stimulate presentations

Figure 3: Structural overview of stimulation paradigm

in the fact that the richness of the brain stimulation via faces can be significantly enhanced, including introduction of diverse viewpoints, lighting conditions [33], and even subtle natural expression variations while strictly preserving the core identity features (i.e. identity preservation). As a result, not only the validity and identity consistency of the generated images are well sustained, but also the dimensionality of the entire dataset is greatly enriched, leading to significant improvement of the perceptual quality and robustness.

3.3 Stimulation Paradigm

Figure 3 illustrates an structural overview of our designed stimulation paradigm for the proposed EEG-Face. As seen, its essential objective is to precisely control the presentation of 20,000 visual stimuli, ensuring a high degree of standardization and consistency across all trials. Throughout the stimulation period, a fixation cross ("+") is continuously displayed at the center of the screen against a constant mid-gray background, in order to sustain a continuous focus of attention. In each trial, the target stimulus (a color face image) is presented centrally on this background without any surrounding border. To ensure a consistent retinal image size across all stimuli, the display size is also carefully calibrated to maintain a visual angle of 7.0 degrees (width) × 6.5 degrees (height) at the preset viewing distance. These unified visual presentation settings are implemented to guide the subject's gaze, providing a stable visual baseline, minimizing potential interference from edge information, and standardizing perceptual input. This level of control is particularly crucial for reducing eye movement artifacts, especially when using eye-tracking sensitive recording techniques such as EEG and MEG [4, 21].

As shown in Figure 3, each face is presented for 500 milliseconds, which is designed to provide sufficient time for perceptual identification of the stimulating face while limiting the intra-trial cognitive interferences. Following the offset of the stimulus, an Inter-Stimulus Interval (ISI) of 500 milliseconds is inserted, during which only the mid-gray background and the central fixation cross are displayed. The primary purpose of this ISI is to allow the participant's visual system to return to a relatively stable baseline level, thereby effectively preventing residual effects from corrupting the independence of individual trials. In addition, the entire sequence of presentations is organized in 500 blocks, each of which contains 40 faces of all individuals, as shown in the column of faces in Figure 3 (a). Specific presentation of each block, together with their durations and blanks, is demonstrated in Figure 3 (b), where the order of their appearances is fully randomized to prevent the test subject from predicting local stimulus patterns. Upon completion of each block, a 3-second Inter-Block-Interval (IBI) is introduced to provide the subject with a brief rest to mitigate the cognitive fatigue.

3.4 EEG recording and preprocessing

EEG sequences are recorded by using a 128-channel GREENTEK system, where electrodes are arranged according to the standard 10-10 system [22], and signals are amplified by a BrainVision actiCHamp Plus amplifier. Following the raw data acquisition, a number of preprocessing is also applied, which can be summarized as: (i) down sampling from the original 5000 Hz to 2000 Hz; (ii) re-referencing the data to electrode A1 and A2, which are clipped to the two earlobes; (iii) sequence of filtering, including zero-phase shift Butterworth filter, low-pass filter with cutoff frequency at 0.3 Hz, and high-pass filter with cutoff frequency at 40 Hz, 50 Hz notch filter to remove power line interference; (iv) application of Independent Component Analysis (ICA) to remove ocular artifacts caused by eye movements such as blinks etc; (v) semi-automatic artifact detection and rejection; (vi) segmenting the continuous EEG waves into event-related epochs, with a time window from 200 milliseconds pre-stimulus to 500 milliseconds post-stimulus; (vii) baseline correction by subtracting the mean value within 200 ms to 0 ms pre-stimulus time window for each trial; and (viii) multivariate noise normalization [9].

As a result, the above recording and the pipeline of preprocessing ultimately yields 20,000 EEG matrices, the size of which can be represented as: 126 (EEG channels) × 500 (EEG digital points).

4 Usability of EEG-Face: Empirical Analysis of Brain-Perceived Multimedia

In this section, we demonstrate the usability of our proposed EEG-Face via a number of experiments carried out towards analysis of brain-perceived multimedia content, i.e. facial images, directly out of their responded EEG trials. Following the spirit of the existing work on EEG-based brain decoding and classification, we selected a wide range of benchmarking classifiers, not only including those advanced deep models, such as EEGNet, EEGChannelNet, SyncNet and Transformers etc. but also those representative traditional unsupervised classifiers such as SVM etc. Details of all these experiment design and result analysis are described as follows.

Neural Correlation of Gender Perceptions: The balanced gender composition (i.e. equal numbers of male and female faces) inside the proposed EEG-Face makes it an ideal platform for investigating how the brain processes and represents gender information from human faces. Relevant analysis can be designed to identify

Table 1: Gender identification results via classifying randomized EEG trials

Model	Accuracy	Model	Accuracy
SVM[17]	52.3%	EEGChannelNet[23]	54.1%
KNN[17]	51.2%	EEGConformer[29]	54.0%
LSTM[11]	54.3%	NICE-EEG[28]	54.1%
MLP[17]	53.8%	EEGNet[15]	54.4%
CNN[17]	54.1%	EEGNet[15] (10 males & 20 females)	68.8%
SyncNet[18]	54.0%		

specific EEG patterns responded to gender information out of variable proportions of female and male faces, leading to dedicated interpretation of key cortical activities elicited by viewing the male versus female faces. Table 1 lists the gender identification results achieved by all the selected models based on the balanced female and male faces. As seen, the proposed EEG-Face indeed contains discernible gender-related representational information that models can learn to differentiate. For instance, among all the models tested, EEGNet achieves the highest accuracy at 54.4%, and several other models including LSTM, EEGChannelNet and NICE-EEG etc. also demonstrated comparable performance levels. To demonstrate the flexibility of EEG-Face, we also selected EEGNet as a representative to test its performance on a non-balanced gender set with 10 male faces and 20 female faces. Over this gender set, EEGNet delivered a classification rate at 68.8%.

While the margins by which these achieved accuracies exceed the chance level (50% for the balanced male-female case) are not exceptionally large, the results achieved by all models consistently and statistically surpass this threshold, and thus providing foundational experimental support for future development of more advanced gender decoding models.

EEG-Stimulus Pairing Verification: By applying the EEG-based brain decoding framework NICE-EEG [28], this experiment is designed to verify the fine-grained and instance-level correspondence between a specific EEG trial and its associated visual stimulus. Essentially, NICE-EEG relies on an image encoder and an EEG encoder to extract modality specific features. These features are aligned through a contrastive learning framework, which pulls the representations of matched EEG-image pairs closer in the embedding space while pushing further apart those of mismatched pairs. This self-supervised approach enables the model to learn cross-modal associations without relying on explicit class labels.

We adopt two test protocols to assess the quality of the learned cross-modal representations:

- 1:1 Verification: Given an EEG-image pair, the model determines whether the two modalities correspond to the same stimulus or not.
- 1:N Verification: Given a query EEG trial, the model selects the most likely matching image from a set of 40 face classes.

Table 2 lists the classification rates achieved by NICE-EEG [28], from which it can be seen that NICE-EEG achieves a top-1 pairing accuracy of approximately 58%, and 3.9%/17.5% for top-1/top-5 accuracy in the 1:1 and 1:N matching scenarios, respectively (with N = 40). To improve the quality (SNR) of the EEG-Face data, we formulate super-trials or meta-trials by randomly averaging the

Table 2: Results of EEG-Image Pairings

Trial Setting	1:1 Verification		1:N Verification	
	Top-1	–	Top-1	Top-5
Single-trial	58.0%	–	3.9%	17.5%
Super-trial	67.0%	–	4.4%	20.5%

Table 3: Results of brain-perceived face recognition

Model	Top-1	Top-5	Params
SVM[17]	3.2%	13.3%	1.9M
KNN[17]	2.7%	11.8%	-
LSTM[11]	3.3%	14.1%	0.1M
MLP[17]	3.1%	13.5%	6.3M
CNN[17]	3.4%	14.4%	0.004M
EEGNet[15]	3.4%	14.7%	0.01M
EEGChannelNet[23]	3.2%	14.2%	4.5M
SyncNet[18]	3.4%	14.8%	0.04M
EEGConformer[29]	3.3%	14.1%	0.6M
NICE-EEG[28]	3.2%	14.1%	0.3M

signals of 10 EEG trials within each class [2]. In this way, EEG-Face reduces its size to 50 super-trials per face and 2000 super-trials in total. As a result, when evaluated on these super-trials, NICE-EEG demonstrates improved performances, achieving 67% top-1 accuracy in the 1:1 setting and 4.4%/20.5% top-1/top-5 accuracy in the 1:N setting.

These results suggest the presence of a learnable correspondence between EEG responses and their associated face-stimuli correlation within the dataset. While performance remains modest in absolute terms, the model demonstrates a consistent ability to distinguish the matched EEG-Image pairs from those mismatched ones. These findings support the potential of the dataset for exploring cross-modal representation learning, particularly when strategies are employed to improve the quality of data collections.

Face Recognition via Classification of Randomized EEG Trials: While face recognition has been widely researched across areas of multimedia and computer vision for the past decades, face recognition out of brain-perceived EEG trials has never been attempted to the best of our knowledge. Considering the fact that classification of randomized EEG trials with 40 general class concepts are already proved to be extremely difficult [1, 17], we expect that this task would be more challenging. Nonetheless, our essential objective is to demonstrate the substantial potential of our proposed EEG-Face for launching new research initiatives to bridge the gap between the macroscopic neural activities and multimedia content, apart from providing a valuable resource for multi-disciplinary researchers.

Table 3 lists the face recognition results achieved by all selected benchmarking baselines, including SVM[17], KNN[17], LSTM[11], MLP[17], CNN[17], EEGNet[15], EEGChannelNet[23], SyncNet[18], EEGConformer[29], and NICE-EEG [28]. In addition to the recognition rates in top-1 and top-5, these results also include the count of parameters used by every model, in order to test and compare the efficiency of all models, i.e. the resources required for achieving the

Table 4: Test results for the varying number of faces

#Classes	EEGNet		#Classes	EEGNet	
	Top-1(%)	Top-5(%)		Top-1(%)	Top-5(%)
1	-	-	21	6.5+0.5	27.3+0.8
2	60.5+2.6	-	22	5.7+0.2	25.0+0.4
3	39.0+2.0	-	23	5.7+0.3	23.9+0.2
4	32.3+1.9	-	24	5.5+0.3	23.6+0.8
5	25.6+1.1	-	25	5.2+0.2	22.1+0.4
6	20.7+1.0	87.3+0.5	26	5.1+0.2	21.6+0.5
7	18.3+1.3	75.7+1.2	27	4.8+0.1	21.2+0.7
8	14.9+0.4	66.5+0.7	28	4.8+0.3	20.3+0.5
9	15.3+1.2	59.7+0.6	29	4.6+0.2	19.2+0.2
10	11.5+0.6	53.1+1.0	30	4.4+0.1	18.8+0.4
11	10.7+0.5	48.5+0.7	31	4.3+0.3	18.1+0.4
12	10.4+0.4	43.6+0.3	32	4.2+0.3	17.4+0.6
13	9.9+0.3	44.5+1.6	33	4.1+0.2	17.3+0.5
14	9.4+0.6	38.6+0.2	34	3.7+0.2	16.7+0.3
15	8.6+0.3	36.3+0.6	35	4.0+0.3	16.1+0.3
16	8.7+0.6	34.5+0.6	36	3.9+0.4	16.2+0.7
17	7.4+0.3	31.7+0.3	37	3.6+0.2	15.8+0.5
18	7.3+0.6	31.3+0.8	38	3.7+0.3	15.3+0.7
19	6.9+0.2	29.1+0.4	39	3.5+0.2	14.9+0.4
20	6.5+0.4	28.7+1.0	40	3.4+0.6	14.7+0.7

recognition rates. As seen, among all the tested models, CNN[17], EEGNet[15], and SyncNet[18] jointly achieved the highest top-1 accuracy of 3.4%. SyncNet[18] obtained the highest top-5 accuracy at 14.8%, which is closely followed by EEGNet[15] and CNN[17]. In addition, it is also noted that models such as CNN and EEGNet only require 0.004 million parameters and 0.01 million parameters, respectively, for their achieved recognition rates. Other models, including LSTM[11] and EEGConformer[29], also demonstrated competitive performances. While the absolute values of the highest top-1 (3.4%) and top-5 (14.8%) rates are modest, they are still above the corresponding chance levels (2.5% for top-1 and 12.5% for top-5), providing a positive validation for the usability of our proposed EEG-Face.

To enable an in-depth evaluation of EEG-Face, we select the well-performed EEGNet as a representative and test its face recognition performances by varying the number of identities from 2 to 40, the results of which are listed in Table 4. As seen, when the number of identities increases, the difficulty of such brain perceived face recognition is substantially intensified, resulting in a progressive decline of both top-1 and top-5 recognition rates. Yet all recognition rates still remain significantly above their respective chance levels. Consequently, this systematic evaluation serves not only to benchmark the EEGNet model's performance across a spectrum of complexities but also to consistently demonstrate the positive existence of classifiable neural signatures within EEG-Face across all the varying number of identities. These findings underscore the dataset's usability for investigating the frontiers of EEG-based recognition of brain-perceived faces.

To further explore the potential of our proposed EEG-Face, we also carried out additional experiments via their super-trials, and the face recognition results are demonstrated in Figure 4. As seen,

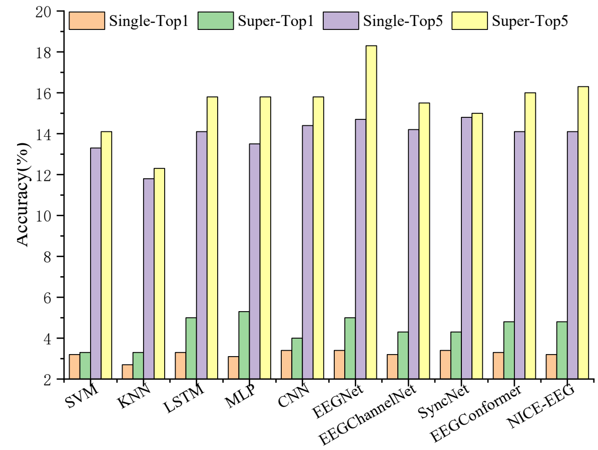


Figure 4: Comparative demonstration of face recognition between single trials and super-trials

compared with the results upon single trials, the recognition rates upon super-trials are generally enhanced across all the tested models, aligning with the expectation for the improved SNRs. In terms of top-1 recognition rates, MLP achieved the best at 5.3%, and in terms of top-5 recognition rates, EEGNet achieved the best rate at 18.3%, a notable improvement from 14.7% on single trials. In summary, while the magnitude of improvement varies across models, most of them, especially deep learning ones, have exhibited clear performance gains when the scheme of super-trials is applied.

5 Conclusions

In this paper, we introduce EEG-Face, a large-scale and high density EEG dataset, which is specifically designed to facilitate research initiatives on brain's perception of facial images. The essential objective is to address the limitations of the existing datasets that all categories constructed are general concepts. Through extensive experiments, it is sufficiently demonstrated that the proposed EEG-Face has great potential for a range of research and applications, which can be highlighted as: (i) further exploitation of generative expansion of face samples to enrich the variation and diversity of facial expressions, thereby facilitating in-depth studies on the neural correlations of emotion perceptions, which is a crucial aspect of social cognition highly relevant to multimedia understanding; (ii) brain decoding of finer-grained facial attributes, such as age or gaze directions etc.; (iii) out of EEG-Face, new research initiatives can be launched to visualize brain-perceived human identities via exploiting generative learning models; and (iv) providing a unique and valuable resource for understanding how human brains perceive multimedia content.

6 Acknowledgments

The authors wish to acknowledge the financial support from Natural Science Foundation China (NSFC) under the Grant No. 62032015 and W2412099.

References

[1] Hamad Ahmed, Ronnie B Wilbur, Hari M Bharadwaj, and Jeffrey Mark Siskind. 2021. Object classification from randomized EEG trials. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3845–3854.

[2] Hari M Bharadwaj, Ronnie B Wilbur, and Jeffrey Mark Siskind. 2023. Still an Ineffective Method With Supertrials/ERPs—Comments on "Decoding Brain Representations by Multimodal Learning of Neural Activity and Visual Features". *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).

[3] Fernando Lopes Da Silva. 2023. EEG: origin and measurement. In *EEG-fMRI: physiological basis, technique, and applications*. Springer, 23–48.

[4] Carlos De La Torre-Ortiz and Tuukka Ruotsalo. 2024. Perceptual Visual Similarity from EEG: Prediction and Image Generation. In *Proceedings of the 32nd ACM International Conference on Multimedia*. 11146–11155.

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

[6] Yiqun Duan, Jinzhao Zhou, Zhen Wang, Yu-Kai Wang, and Chin-Teng Lin. 2023. Dewave: Discrete eeg waves encoding for brain dynamics to text translation. *arXiv preprint arXiv:2309.14030* (2023).

[7] Ahmed Fares, Sheng-hua Zhong, and Jianmin Jiang. 2020. Brain-media: A dual conditioned and lateralization supported GAN (DCLS-GAN) towards visualization of image-evoked brain activities. In *Proceedings of the 28th ACM International Conference on Multimedia*. 1764–1772.

[8] Alessandro T Gifford, Kshitij Dwivedi, Gemma Roig, and Radoslaw M Cichy. 2022. A large and rich EEG dataset for modeling human visual object recognition. *NeuroImage* 264 (2022), 119754.

[9] Matthias Guggenmos, Philipp Sterzer, and Radoslaw Martin Cichy. 2018. Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *Neuroimage* 173 (2018), 434–447.

[10] Ahmed G Habashi, Ahmed M Azab, Seif Eldawlatly, and Gamal M Aly. 2023. Generative adversarial networks in EEG analysis: an overview. *Journal of NeuroEngineering and Rehabilitation* 20, 1 (2023), 40.

[11] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[12] Mahboobeh Jafari, Afshin Shoeibi, Marjane Khodatars, Sara Bagherzadeh, Ahmad Shalbaf, David López García, Juan M Gorriz, and U Rajendra Acharya. 2023. Emotion recognition in EEG signals using deep learning methods: A review. *Computers in Biology and Medicine* (2023), 107450.

[13] Blair Kaneshiro, Marcos Perreau Guimaraes, Hyung-Suk Kim, Anthony M Norcia, and Patrick Suppes. 2015. A representational similarity analysis of the dynamics of object processing using single-trial EEG classification. *Plos one* 10, 8 (2015), e0135697.

[14] Isaak Kavasidis, Simone Palazzo, Concetto Spampinato, Daniela Giordano, and Mubarak Shah. 2017. Brain2image: Converting brain signals into images. In *Proceedings of the 25th ACM International Conference on Multimedia*. ACM, 1809–1817.

[15] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. 2018. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of neural engineering* 15, 5 (2018), 056013.

[16] Dongyang Li, Chen Wei, Shiying Li, Jiachen Zou, Haoyang Qin, and Quanying Liu. 2024. Visual decoding and reconstruction via eeg embeddings with guided diffusion. *arXiv preprint arXiv:2403.07721* (2024).

[17] Ren Li, Jared S Johansen, Hamad Ahmed, Thomas V Ilyevsky, Ronnie B Wilbur, Hari M Bharadwaj, and Jeffrey Mark Siskind. 2020. The perils and pitfalls of block design for EEG classification experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 1 (2020), 316–333.

[18] Yitong Li, Kafui Dzirasa, Lawrence Carin, David E Carlson, et al. 2017. Targeting EEG/LFP synchrony with neural nets. *Advances in Neural Information Processing Systems* 30 (2017).

[19] Zhen Li, Mingdeng Cao, Xintao Wang, Zhongang Qi, Ming-Ming Cheng, and Ying Shan. 2024. Photomaker: Customizing realistic human photos via stacked id embedding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8640–8650.

[20] Jinyuan Liu, Xingyuan Li, Zirui Wang, Zhiying Jiang, Wei Zhong, Wei Fan, and Bin Xu. 2024. PromptFusion: Harmonized semantic prompt learning for infrared and visible image fusion. *IEEE/CAA Journal of Automatica Sinica* (2024).

[21] Jun Ma and Tuukka Ruotsalo. 2024. Cognition-Supervised Saliency Detection: Contrasting EEG Signals and Visual Stimuli. In *Proceedings of the 32nd ACM International Conference on Multimedia*. 7744–7753.

[22] Marc R Nuwer, Giancarlo Comi, Ronald Emerson, Anders Fuglsang-Frederiksen, Jean-Michel Guérit, Hermann Hinrichs, Akio Ikeda, Fransisco Jose C Luccas, and Peter Rappelsburger. 1998. IFCN standards for digital recording of clinical EEG. *Electroencephalography and clinical Neurophysiology* 106, 3 (1998), 259–261.

[23] Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, Daniela Giordano, Joseph Schmidt, and Mubarak Shah. 2020. Decoding brain representations by multimodal learning of neural activity and visual features. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 11 (2020), 3833–3849.

[24] Ruijie Quan, Wenguan Wang, Zhibo Tian, Fan Ma, and Yi Yang. 2024. Psychometry: An omnifit model for image reconstruction from human brain activity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 233–243.

[25] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.

[26] Steffen Schneider, Jin Hwa Lee, and Mackenzie Weygandt Mathis. 2023. Learnable latent embeddings for joint behavioural and neural analysis. *Nature* (2023), 1–9.

[27] G Siddhad, A Gupta, D P Dogra, et al. 2024. Efficacy of transformer networks for classification of EEG data. *Biomedical Signal Processing and Control* 87 (2024), 105488.

[28] Yonghao Song, Bingchuan Liu, Xiang Li, Nanlin Shi, Yijun Wang, and Xiaorong Gao. 2024. Decoding Natural Images from EEG for Object Recognition. *arXiv preprint arXiv:2308.13234* (2024).

[29] Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. 2022. EEG conformer: Convolutional transformer for EEG decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 31 (2022), 710–719.

[30] Concetto Spampinato, Simone Palazzo, Isaak Kavasidis, Daniela Giordano, Nasim Souly, and Mubarak Shah. 2017. Deep learning human mind for automated visual classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6809–6817.

[31] Yu Takagi and Shinji Nishimoto. 2023. 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*. 14453–14463.

[32] Shizun Wang, Songhua Liu, Zhenxiong Tan, and Xinchao Wang. 2024. Mindbridge: A cross-subject brain decoding framework. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11333–11342.

[33] Xu Wu, Zhihui Lai, Jie Zhou, Xianxu Hou, Witold Pedrycz, and Linlin Shen. 2024. Light-aware contrastive learning for low-light image enhancement. *ACM Transactions on Multimedia Computing, Communications and Applications* 20, 9 (2024), 1–24.

[34] Huzheng Yang, James Gee, and Jianbo Shi. 2024. Brain Decodes Deep Nets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 23030–23040.

[35] Sheng-hua Zhong, Ahmed Fares, and Jianmin Jiang. 2019. An attentional-LSTM for improved classification of brain activities evoked by images. In *proceedings of the 27th ACM international conference on multimedia*. 1295–1303.