

Data Augmentation for EEG Visual Classification with Deep Neural Networks

Shuning Xue, Jie Jiang, Longteng Guo and Jing Liu, *Member, IEEE*

Abstract—Electroencephalography (EEG) is a non-invasive technique used to record electrical activity in the brain. It is commonly applied in research related to brain-computer interface (BCI) systems and visual decoding. Recently, researchers have developed deep neural networks aimed at EEG-based visual classification. While these deep learning models improve the performance of visual decoding in EEG-based BCIs, they often face challenges such as overfitting due to limited data availability and the low signal-to-noise ratio of EEG signals. To address the problem of insufficient training samples, this study introduces data augmentation methods for EEG visual classification. Experimental results show that the Noise Adding and Time Masking methods, chosen from six augmentation methods, lead to significant performance improvements across three classification tasks involving two EEG datasets. Furthermore, to overcome the limitations associated with the scarcity of high signal-to-noise ratio EEG signals, this study averages the EEG signals evoked by the same visual stimulus. Subsequently, data augmentation methods are applied to these averaged signals to generate high-quality augmented samples. By combining the augmented and averaged signals with the original signals, a comprehensive training set is created, enabling the EEG model to better distinguish features relevant to visual concept classification. The extensive experiments also investigated the impact of augmented sample sizes on different training data volumes, as well as the effects of different parameter settings for augmentation and training. Overall, this work introduces data augmentation for EEG visual classification for the first time, significantly improving the performance of EEG deep learning models.

Index Terms—Brain-computer interface (BCI), electroencephalography (EEG), EEG-based visual recognition, deep learning, data augmentation

I. INTRODUCTION

Brain-computer interfaces (BCIs) enable humans to interact with the external environment through neural signals [1]. Electroencephalography (EEG) is a widely studied non-invasive

This research is supported by Artificial Intelligence-National Science and Technology Major Project (2023ZD0121200) and the National Natural Science Foundation of China (U21B2043, 6243000159, 62102416), and the Key Research and Development Program of Jiangsu Province under Grant BE2023016-3. (Corresponding author: Jing Liu.)

Shuning Xue and Jing Liu are with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China, and also with the Zidongtaichu Foundation Model Research Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: xueshuning2021@ia.ac.cn; jliu@nlpr.ia.ac.cn).

Jie Jiang, and Longteng Guo are with the Zidongtaichu Foundation Model Research Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: jie.jiang@nlpr.ia.ac.cn; longteng.guo@nlpr.ia.ac.cn).

method for recording electrophysiological signals to develop BCI systems [2], [3]. Recently, there has been growing interest among researchers in the visual decoding of EEG signals [4]–[8]. Researchers have conducted tasks such as semantic classification [9] visual stimulus retrieval [10], and natural image generation [11], [12] using visual stimulus-evoked EEG datasets. Among these tasks, enhancing the accuracy of EEG visual classification is crucial for creating a reliable BCI system based on visual decoding.

With the rise of artificial intelligence and deep learning technology, researchers have successfully developed high-performance deep neural network models and algorithms [13]–[15]. These methods replaced traditional EEG feature extraction methods that rely on manual design and improved the accuracy of EEG visual classification tasks. Despite the impressive learning capabilities of deep neural networks and their potential contributions to building BCIs based on EEG visual decoding, these models often encounter the problem of overfitting during training. This issue arises because deep learning is a data-driven approach that optimizes model parameters through gradient backpropagation over multiple training iterations. Such methods typically require large amounts of high-quality training data to avoid deep neural networks fitting to noisy signals. However, acquiring EEG signals is time-consuming and labour-intensive. While researchers have made substantial efforts in collecting large-scale datasets [16]–[19] for EEG visual decoding, the amount of EEG signals remains insufficient compared to the extensive datasets available in computer vision and natural language processing.

Data augmentation is a widely used technique that increases the number of training samples and enhances the diversity of datasets. This approach has shown significant benefits by helping models focus on relevant features specific to particular tasks while minimizing the impact of irrelevant input features. In the field of EEG signal decoding, data augmentation methods have been applied to various tasks, including steady-state visual evoked potential (SSVEP) [20], emotion recognition [21], motor imagery [22], disease prediction [23], and sleep stage classification [24]. Researchers have explored a variety of signal transformation methods to augment EEG signals for these tasks. Li et al. [25] achieved promising results in the motor imagery task with deep neural networks by applying amplitude perturbation to the input signals. Rommel et al. [26] explored three data augmentation methods for sleep stage classification, including Time Reversal, Sign Flip, and Frequency Shift. Ding et al. [27] introduced Time Masking to augment EEG signals and enhance the classification

performance of the asynchronous SSVEP-BCI system. In addition to transformations for time series data, researchers have also explored data augmentation methods based on the spatial characteristics of EEG. Wang et al. [28] introduced a straightforward knowledge-driven approach called Channel Reflection. This method involves reflecting the electrodes from the left and right sides of the brain in EEG signals, swapping the labels for the motor imagery task, while retaining the original labels for SSVEP, P300, and seizure classification tasks.

Data augmentation has been utilized in various EEG tasks; however, there have been relatively few studies focused specifically on EEG visual decoding. Most research in this area has concentrated on developing deep neural network architectures [29], [30] and leveraging knowledge from pre-trained visual models to enhance feature learning during EEG model training [31], [32]. Despite this, there has been limited investigation into training techniques aimed at addressing overfitting issues in deep neural networks. To tackle the challenge of suboptimal classification performance stemming from insufficient data, we introduced data augmentation methods for EEG visual classification. We conducted experiments using two public EEG datasets that employ different visual stimulus-evoked paradigms. In a six-class classification task using the EEG72 dataset [5], a hybrid EEG model [33] demonstrated a performance improvement of 0.45% with the implementation of data augmentation methods. For the 72-class classification task using the same EEG72 dataset, the model achieved a performance enhancement of 2.37%. Similarly, in a 200-class classification task with the THINGS-EEG-5Hz dataset [18], a performance increase of 2.29% was observed.

When selecting signal transformation methods for data augmentation, previous studies indicate that the same method can have varying effects across different tasks [34], [35]. The inherent complexity of neural signals has led to inconsistencies in choosing augmentation techniques for various EEG decoding tasks. In some cases, an augmentation method that is effective for one task may actually degrade performance in another. To explore the impact of different data augmentation methods on EEG visual classification tasks, our study investigated six distinct transformations: Noise Addition, Time Masking, Time Reversal, Sign Flip, Channel Reflection, and Time Shift. We adjusted the parameters associated with these methods and analyzed performance variations across different settings. The results showed that Noise Addition and Time Masking are effective strategies that enhance the performance of five EEG models across two EEG datasets. Conversely, Time Reversal, Sign Flip, Channel Reflection, and Time Shift were found to be less effective and may not be suitable for EEG visual classification tasks.

In addition to insufficient training data, a low signal-to-noise ratio in EEG signals is a key reason why models tend to overfit. EEG signals are non-invasive electrophysiological measurements that often contain noise and artifacts from muscle activity, circuit interference, and environmental factors [36]. When using data augmentation, introducing additional variables can negatively affect the training of EEG models [34]. To enhance the signal-to-noise ratio of the augmented samples, we utilized

a signal-averaging method. This technique improves the clarity of event-related potential (ERP) waveforms by averaging multiple trials triggered by the same stimulus [37]. Both EEG datasets we examined for visual decoding consist of signals collected from subjects repeatedly exposed to the same image. To minimize the impact of irrelevant components in visual stimulus-evoked EEG signals, we combined data augmentation with the signal-averaging method. Specifically, we first averaged the samples corresponding to the same visual stimulus in the training set and then applied the augmentation method to these averaged samples. This combined approach resulted in augmented samples with a higher signal-to-noise ratio. Experimental results showed that incorporating the signal-averaging method significantly improved model performance. In the six-class classification task on the EEG72 dataset, performance increased by 0.88%. Similarly, in the 72-class classification task on the same dataset, performance rose by 1.06%, while the 200-class classification task on the THINGS-EEG-5Hz dataset improved by 4.01%.

To gain a comprehensive understanding of how data augmentation affects EEG visual classification, we evaluated the impact of varying augmented sample sizes on different volumes of training data for the EEG72 and THINGS-EEG-5Hz datasets. For all three tasks using the full training data, we consistently observed an improvement in accuracy as the size of the augmented samples increased. However, data augmentation did not generally enhance performance when the volume of training data was relatively small. We also examined how different degrees of averaging influenced the algorithm's performance. In the six-class task using the EEG72 dataset, the model's performance declined as the number of overlapping signals increased. Conversely, in the 72-class task with the same EEG72 dataset, performance initially improved with a higher number of overlapping signals before eventually declining. In the 200-class task using the THINGS-EEG-5Hz dataset, the model's performance consistently improved as the number of overlapping signals increased. Furthermore, our sensitivity experiments revealed that among the various parameters affecting training, dropout rates had a significant impact on the performance of artificial neural networks.

In summary, the main contributions of this study are as follows:

- (1) We introduce data augmentation methods for EEG visual classification with deep neural networks, which led to performance improvements on two public EEG datasets.

- (2) We evaluate the effects of six transformation-based data augmentation methods on EEG visual classification tasks and discuss potential reasons for the distinct outcomes.

- (3) We apply data augmentation on averaged signals to enhance both the quantity and quality of the augmented samples, further improving the accuracy of EEG visual classification.

- (4) We perform extensive experiments to examine the effects of data augmentation parameters and different degrees of signal averaging. Additionally, we evaluate the effect of augmented sample sizes on various training data volumes and assess the sensitivity of our method to training parameters.

II. MATERIALS AND METHODS

TABLE I
DESCRIPTION OF THE TWO PUBLIC EEG DATASETS.

Dataset	Exp. paradigm	# of channels	# of class	# of subjects	# of trials per class	Epoch length (ms)
EEG72 [5]	low-speed	124	6 or 72	10	864 or 72	500
THINGS-EEG-5Hz [18]	RSVP	64	200	10	80	500

*Exp.: Experimental, low-speed: low-speed serial visual presentation, RSVP: rapid serial visual presentation

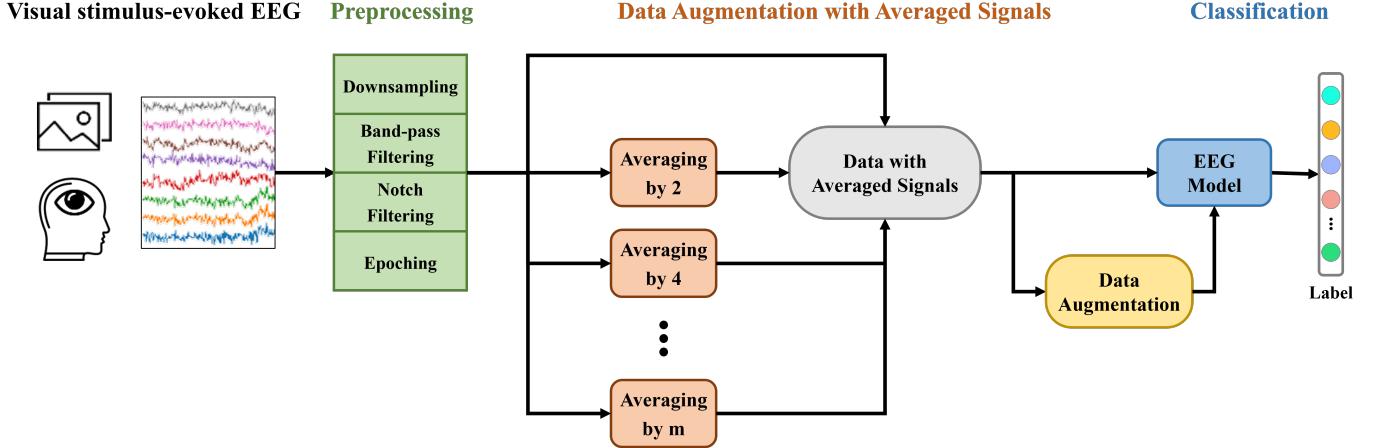


Fig. 1. Our overall framework with data augmentation for EEG visual classification.

A. Datasets

Table I details the two public EEG datasets for EEG visual classification.

1) **EEG72 [5]**: The first EEG dataset includes EEG data collected from 10 participants. This study involved presenting color images from six semantic categories. In total, 72 images were shown on a mid-grey background, each displayed for 500 ms. The EEG data was recorded using the 128-channel EGI HCGSN 110 nets.

2) **THINGS-EEG-5Hz [18]**: The second EEG dataset also comprises EEG data collected from 10 participants. The data was gathered using the Rapid Serial Visual Presentation (RSVP) paradigm, which involved presenting natural images cropped to a square shape. These images were chosen from 1,854 object concepts in the THINGS database [38]. While collecting EEG signals, each image was shown for 100 milliseconds, followed by a 100 milliseconds blank screen. The signals were collected using a BrainVision actiCHamp amplifier with 64-channel EASYCAP. This dataset is divided into two parts: 1654 concepts with 10 images and 4 repetitions each, and 200 concepts with 1 image and 80 repetitions each. In our study, we used only the 200 concepts that had a high number of repetitions, which is more suitable for deep-learning and signal-averaging experiments.

B. Data Preprocessing

The EEG signals from EEG72 has already been preprocessed when it was released. The samples were filtered within the frequency range of 1 to 25 Hz. The signals were subsequently downsampled to 62.5 Hz and segmented into trials consisting of 32 time samples.

For the THINGS-EEG-5Hz datasets, the MNE package [39] was utilized for preprocessing. The sampling rate of the data was resampled to 250 Hz. Then the signals were filtered to include frequencies from 0.1 to 100 Hz, and a notch filter was applied to eliminate 50 Hz power line interference. The epochs were created from 0 to 500 ms relative to stimulus onset. Additionally, we applied z-score normalization and clamped large voltage values. No further preprocessing or artifact correction methods were applied.

C. Framework Overview

This section provides a brief overview of our proposed framework. The overall framework is depicted in Fig. 1. First, we preprocess the EEG signals evoked by visual stimuli. Next, we epoch the time series into samples, assigning each sample a corresponding label based on the concept of visual stimulus. For the samples stimulated by the same visual stimulus, we calculate the averaged samples. Along with the original samples, we augment these samples to increase the size of the training set. Finally, we utilized the augmented data with averaged signals to train the EEG model, which iteratively learns to perform the EEG visual classification task and outputs the predicted probabilities of multiple visual concepts.

D. Data Augmentation

This section provides a description of six data augmentation methods. The legends of temporal transformation methods are depicted in Fig. 2 and the legend of the Channel Reflection method is depicted in Fig. 3.

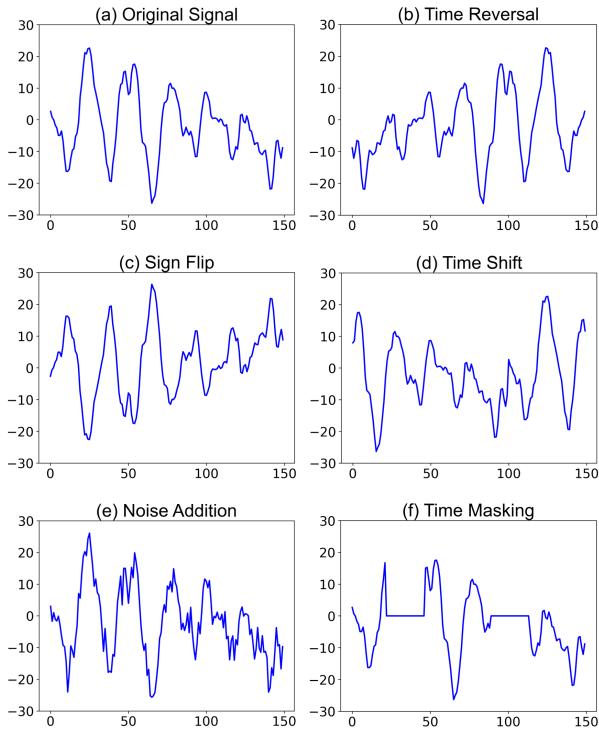


Fig. 2. Data augmentation methods based on temporal transformation for EEG signals. (a) Original Signal. (b) Time Reversal. (c) Sign Flip. (d) Time Shift. (e) Noise Addition. (f) Time Masking.

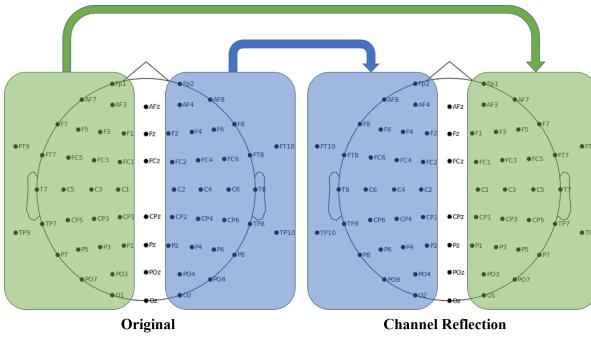


Fig. 3. The Channel Reflection augmentation method for THINGS-EEG-5Hz.

1) Time Reversal: Time Reversal refers to flipping the time axis of a signal. The legend of Time Reversal method is depicted in Fig. 2(b). This transformation method is particularly suitable for tasks sensitive to frequency domain information, as altering the timing does not affect the relative proportions of different frequencies.

2) Sign Flip: Sign Flip refers to the process of flipping the electrical potentials captured by EEG sensors by multiplying the outputs of all channels by -1. The legend of Sign Flip method is depicted in Fig. 2(c). This transformation can be viewed as inverting the direction of charge flow in neuron dendrites.

3) Time Shift: Time Shift refers to advancing or postponing a change in the signal. In our implementation, we perform a cyclic shift on a single trial that has been epoched. This means that any signal that moves out of the epoching range

due to the shift is filled into the position that was vacated. The legend of Time Shift method is depicted in Fig. 2(d). To investigate the influence of different response delays on the visual classification task, we performed experiments with time shifts of 16 ms, 32 ms, 48 ms, 464 ms, 480 ms, and 496 ms for EEG72 and 8 ms, 16 ms, 24 ms, 476 ms, 484 ms, and 492 ms THINGS-EEG-5Hz. The best results among different shifts were presented in tables.

4) Noise Addition : Noise Addition refers to the process of adding Gaussian noise to raw time series data. The legend of Noise Addition method is depicted in Fig. 2(e). This transformation method helps prevent deep neural networks from fitting the noise components, encouraging the model to resist the influence of noise. To investigate the impact of varying noise levels on the visual classification task, we performed experiments using Noise Addition with noise amplitude proportions of 0.0025, 0.005, 0.01, 0.02, and 0.03 for EEG72 and THINGS-EEG-5Hz.

5) Time Masking : Time Masking involves applying a mask window to a sample and setting the values within that window to zero. In our implementation, we randomly apply a mask in the time domain for each electrode of a single trial, ensuring that the mask position varies for different trials. This approach prevents the model from becoming overly reliant on specific time segments, allowing it to learn how to handle data across those segments more effectively. The legend of Time Masking method is depicted in Fig. 2(f). This transformation technique encourages deep neural networks to learn global features rather than focusing solely on particular time segments. To investigate the effects of different mask ratios and the continuity of masks on visual classification tasks, we performed experiments with various combinations of mask numbers and mask lengths, including (1,1), (1,2), (2,1), (1,4), (2,2), and (4,1) for EEG72 and THINGS-EEG-5Hz.

6) Channel Reflection : Channel Reflection involves reflecting the left and right brain electrodes (or channels) of EEG signals while simultaneously swapping the labels for motor imagery tasks [28]. The legend of Channel Reflection method is depicted in Fig. 3. For EEG visual classification, we reflect the left and right electrodes while retaining the original labels. This transformation technique encourages deep neural networks to consider the bilateral symmetry of the brain.

E. Data Augmentation with Averaged Signals

This section describes the process of data augmentation with averaged signals, which involves taking samples that share the same visual stimulus and calculating their average. The concept of averaging EEG signals with the same stimulus is illustrated in Fig. 5. The averaged data can also be subjected to signal transformation methods to further enhance the sample size. Note that only the training set data applied signal-averaging methods, while the test set data remains unaveraged to align with single-trial classification in practical applications.

In the following sections, we introduce the two methods used for averaging EEG signals in this study, along with the specific signal-averaging designs applied to the two datasets.

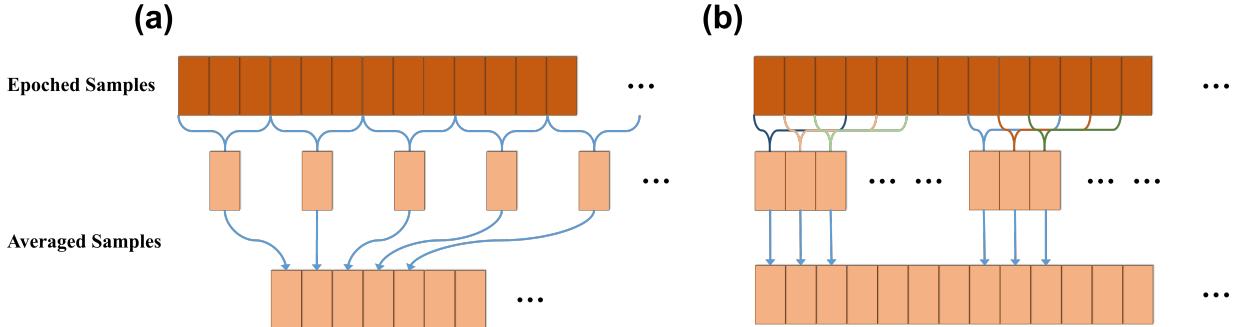


Fig. 4. The two averaging methods for EEG signals. (a) Non-overlapping Averaging. (b) Overlapping Averaging.

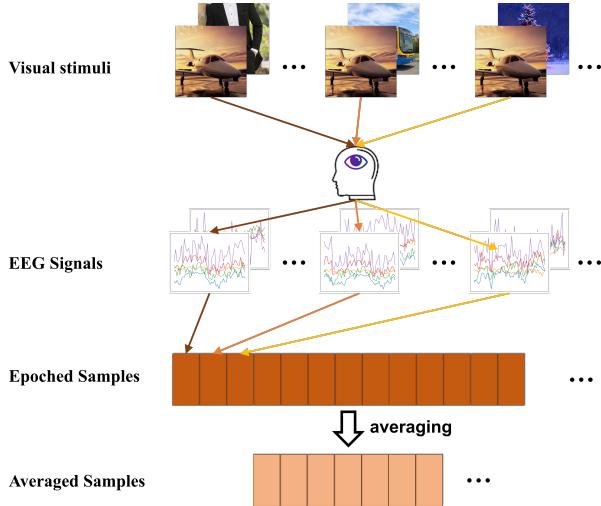


Fig. 5. The legend of averaging EEG signals with the same stimulus.

1) Data Augmentation with Non-overlapping Averaged Signals: In the EEG72 and THINGS-EEG-5Hz datasets, the training sets consist of 64 and 72 samples for each class, respectively. Consequently, when averaging the training set data, the maximum number of samples that can be averaged is 64 and 72, respectively.

It is important to clarify that no duplicate samples were used when conducting the non-overlapping averaging. Specifically, when averaging original n samples by a number m , we first average m samples to create a new sample. These m samples are then excluded from the subsequent signal averaging process. In the end, n/m new samples are generated. In this study, we refer to this method as Non-overlapping Averaging. The legend of Non-overlapping Averaging is depicted in Fig. 4(a).

In the non-overlapping experiment, we combined different averaging sample sizes: 64, 32, 16, 8, 4, and 2 samples for the EEG72 dataset, and 72, 36, 18, 8, 4, and 2 samples for the THINGS-EEG-5Hz dataset. This approach ensures that the size of the merged averaged signals is approximately equal to the original sample size.

2) Data Augmentation with Overlapping Averaged Signals: The second signal-averaging method reuses the samples. In this approach, the original n samples are also averaged by a number m . After randomly rearranging the 64 samples, we sum the first m samples and divide the sum by m to create the

first new sample. For the second new sample, we subtract the first sample from the sum of the m samples, add the $(m+1)^{th}$ sample, and then divide this new sum by m . This process continues for all subsequent samples: we subtract the sample with the smallest index from the new sum, add a new sample, and divide by m to generate a new sample. In this study, we refer to this method as Overlapping Averaging. The legend of Overlapping Averaging is depicted in Fig. 4(b).

In the overlapping experiment, we explored the effects of averaging different sample sizes, specifically 2, 3, 4, 5, 6, 7, 8, 9, and 10 samples, for both the EEG72 and THINGS-EEG-5Hz datasets.

III. EXPERIMENTS AND RESULTS

A. EEG Models and Evaluation Metrics

1) EEG Models: To evaluate data augmentation methods, we primarily assessed classification accuracies using a hybrid model [33]. Below are brief descriptions of the model: In our study, we reproduced four modules from the hybrid model. The Electrode Reweight module adaptively reweights the input electrodes, focusing on vision-related electrodes. The Temporal CNN module extracts local temporal features from the EEG signals of each electrode, while the Spatial CNN module integrates these temporal features across all electrodes. Finally, the Transformer module derives global EEG embeddings from the collected features.

To investigate the effect of data augmentation on various models, we conducted experiments using Time Masking and Noise Addition across four additional models. The results of these experiments are detailed in Section IV-B. The four models included are: the linear model, ShallowConvNet [40], Temporal CNN [30], and ATM [11]. The implementation codes for ShallowConvNet were obtained from LMDA [41]. The Temporal CNN model was reproduced based on the temporal stream structure of TSCNN [30]. And the implementation codes for ATM were sourced from ATM [11].

For the classification task, we used fully connected (FC) networks with a single hidden layer. The sizes of the hidden layers are set to 128 for the EEG72 dataset, and 512 for the THINGS-EEG-5Hz dataset. We implemented the EEG models along with data augmentation methods using the publicly available PyTorch 1.6.0 framework. To determine optimal hyperparameters for the hybrid model, we used a grid search

TABLE II
ACCURACY (%) ON EEG72 AND THINGS-EEG-5Hz DATASETS FOR SIX AUGMENTATION METHODS.

Method	EEG72			THINGS-EEG-5Hz
	6-class	72-class	200-class	
baseline	53.20 ± 0.43	25.01 ± 0.82	37.19 ± 2.79	
+ Channel Reflection	—	—	18.31 ± 1.80	
+ Sign Flip	32.25 ± 0.34	7.06 ± 0.21	26.88 ± 2.16	
+ Time Reversal	42.27 ± 0.36	15.42 ± 0.37	24.51 ± 1.90	
+ Time Shift	51.82 ± 0.42	24.32 ± 0.50	32.23 ± 2.00	
+ Time Masking	53.39 ± 0.39	<u>27.26</u> ± 0.47**	38.45 ± 3.08**	
+ Noise Addition	53.66 ± 0.50*	26.73 ± 0.55**	39.73 ± 3.13**	
+ Noise + Mask	53.65 ± 0.50*	27.38 ± 0.49**	39.48 ± 2.99**	

where **bold** fonts indicate the best results, and underlined fonts are the second best results. * denotes the classification accuracy of the model trained with this specific augmentation are significantly better than the baseline (paired t-test $p < 0.05$). ** denotes $p < 5e^{-4}$.

strategy to optimize the learning rate, batch size, and other parameters. For optimization, we utilized the Adam algorithm [42].

2) *Evaluation Metrics*: In our experiments, we utilized either 9-fold or 10-fold cross-validation to evaluate our methods.

For the EEG72 dataset, each image was presented to participants 72 times. We conducted a 9-fold cross-validation, using 64 of the 72 samples for each image as the training set, while the remaining eight samples were used as the test set.

For the THINGS-EEG-5Hz dataset, which includes 200 images as visual stimuli, each image was presented to participants 80 times. We divided the 80 EEG signals induced by each image into 10 folds, and employed 10-fold cross-validation. During each training process, nine folds of data were used as the training set, consisting of 72 samples for each class, while the remaining fold served as the test set.

We averaged the accuracies obtained from each test fold and averaged the subject-wise accuracies across all subjects. The results are reported in the following sections.

B. Overall Performance of Data Augmentation

In this section, we evaluate the classification accuracies of six data augmentation methods using the hybrid model on two EEG datasets. Table II demonstrates that models employing specific data augmentation methods achieve higher classification accuracy, while other methods show poorer performance.

Generally, Time Masking and Noise Addition can significantly enhance model performance. In contrast, methods such as Channel Reflection, Time Reversal, Sign Flip, and Time Shift do not help and result in decreased accuracies. Specifically, Time Masking yielded a 0.19% improvement in accuracy on 6-class tasks for EEG72 ($p = 0.355$), and a 2.25% improvement on 72-class tasks for EEG72 ($p = 2.22e^{-4}$). Moreover, it led to a 1.26% improvement on 200-class tasks for THINGS-EEG-5Hz ($p = 7.68e^{-7}$). Noise Addition produced a 0.46% improvement in accuracy on 6-class tasks for EEG72 ($p = 2.09e^{-2}$), and a 1.72% improvement on 72-class tasks for EEG72 ($p = 6.67e^{-6}$). It also contributed a 2.54% improvement on 200-class tasks for THINGS-EEG-5Hz ($p = 1.63e^{-19}$). Additionally, in most cases, the combination of Time Masking and Noise Addition did not produce further accuracy improvements. This indicates that the effect of data

augmentation methods has its limits and that mixed augmentation is not always better. For the effect of different augmented sample sizes, we did relevant experiments in Section IV-A.

Overall, the improvements in classification accuracy suggest that data augmentation is an effective approach for EEG visual classification. It enhances the classification performance across three tasks of two datasets, demonstrating its ability to improve the generalization of the deep-learning model on the test set.

C. Ablation Studies with Averaged Signals

In this section, we further evaluated the impact of data augmentation using averaged signals on two EEG datasets with the hybrid model. Table III presents the classification performance across five conditions for the EEG72 and THINGS-EEG-5Hz datasets.

For both the EEG72 and THINGS-EEG-5Hz datasets, training with averaged signals significantly improved the performance of the hybrid models. In the 6-class task of EEG72, accuracy increased by 0.89% ($p = 2.17e^{-4}$). For the 72-class task of EEG72, accuracy improved by 3.54% ($p = 1.49e^{-12}$). In the 200-class task of THINGS-EEG-5Hz, accuracy rose by 8.49% ($p = 5.11e^{-56}$).

The combination of data augmentation techniques with averaged signals produced mixed results. In the 6-class task of EEG72, combining Time Masking with averaged signals improved accuracy by 0.15% compared to training with averaged signals alone ($p = 0.254$). The addition of Noise Addition with averaged signals increased accuracy by 0.33% ($p = 1.82e^{-2}$). Furthermore, combining both Time Masking and Noise Addition with averaged signals resulted in a further accuracy improvement of 0.39% ($p = 3.04e^{-3}$). In the 72-class task of EEG72, however, the combination of data augmentation techniques with averaged signals yielded similar accuracy to the baseline with averaged signals. For the 200-class task of THINGS-EEG-5Hz, the combination of data augmentation with averaged signals significantly decreased accuracy.

The experimental results suggest that both data augmentation and signal averaging contribute positively to improving classification accuracies. However, using averaged signals during training may reduce the advantages of data augmentation. Data augmentation techniques are designed to mitigate the overfitting of deep learning models to noise present in the

TABLE III
ABLATION EXPERIMENTS ON EEG72 AND THINGS-EEG-5Hz DATASETS FOR THE AUGMENTATION WITH AVERAGED SIGNALS.

Method	EEG72			THINGS-EEG-5Hz
	6-class	72-class	200-class	
baseline	53.20 ± 0.43	25.01 ± 0.82	37.19 ± 2.79	
+ Average	$54.09 \pm 0.34^*$	$\underline{28.55} \pm 0.34^*$	$\underline{45.68} \pm 3.51^*$	
+ Mask + Average	54.24 ± 0.32	28.25 ± 0.45	43.59 ± 3.34	
+ Noise + Average	$54.42 \pm 0.18_{\dagger}$	$\underline{\underline{28.68}} \pm 0.41$	44.65 ± 3.40	
+ Noise + Mask + Average	$\underline{\underline{54.53}} \pm 0.47_{\dagger}$	28.44 ± 0.54	43.49 ± 3.24	

where **bold** fonts indicate the best results, and underlined fonts are the second best results. $*$ denotes the classification accuracy of the model trained with averaged signals are significantly better than the baseline (paired t-test $p < 5e^{-4}$). \dagger denotes the specific augmentation with averaged signals are significantly better than the baseline with averaged signals (paired t-test $p < 0.05$).

training data. Since averaged signals have already eliminated a substantial amount of irrelevant noise compared to single-trial signals.

In summary, when data augmentation is applied to averaged signals to increase the number of training samples, it presents a double-edged sword. On one hand, this approach exposes the model to more diverse artificial data, enhancing generalization. On the other hand, it can dilute the proportion of averaged signals in the training set. This dilution risks reintroducing some noise into the training data, leading to a situation where the model fits more noisy data rather than clean averaged signals. This issue was evident in poorer performance on the 72-class task of EEG72 and the 200-class task of THINGS-EEG-5Hz when combining data augmentation with averaged signals. Nonetheless, applying data augmentation to averaged signals did yield a slight improvement in the 6-class task of EEG72, indicating that there can still be potential benefits from data augmentation methods.

IV. DISCUSSION

A. Data Augmentation on Different Training Data Volumes and the Effect of Augmented Sample Size

In this section, we evaluated the effect of augmented sample sizes on various training data volumes of the EEG72 and THINGS-EEG-5Hz datasets. The classification performance across different augmented sample sizes is presented in Fig. 6.

From Fig. 6 (a) and (b), it can be observed that using 100% of the training data with the augmentation method significantly improves performance on the 72-class tasks, while the effect is less pronounced for the 6-class tasks. This difference may be due to the differing number of samples per class; while the training sample sizes are consistent between the 6-class and 72-class tasks, the tasks with a higher number of classifications have fewer samples available for each class. As a result, these classes are more vulnerable to noise, making data augmentation particularly effective in mitigating interference from noise.

Furthermore, for all three tasks using 100% of the training data, a continuous improvement in accuracy can be observed with increased augmented sample size. This indicates the effectiveness of augmentation with more data could yield further improvement in enhancing the generalization capabilities of deep learning models.

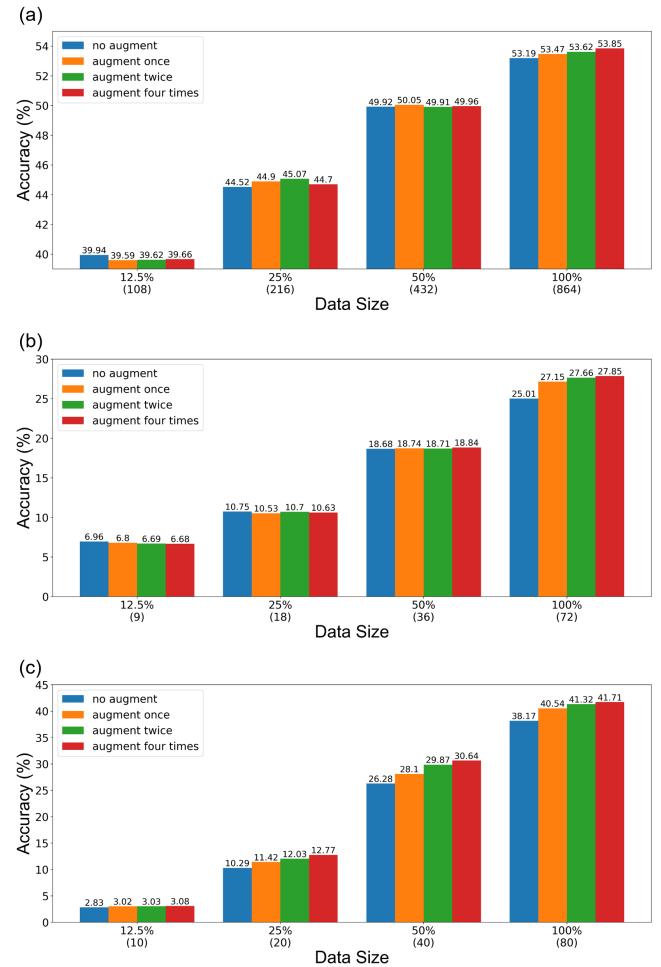


Fig. 6. Experiments on EEG72 and THINGS-EEG-5Hz for Different Training Data Volumes and Augmented Sample Sizes Using Noise Addition Method. (a) 6-class task of EEG72. (b) 72-class task of EEG72. (c) 200-class task of THINGS-EEG-5Hz.

It's important to highlight that, in all three tasks, data augmentation generally does not enhance performance when the training data volume is small. This may occur because the diversity introduced by data augmentation may not adequately address the temporal variations present in EEG signals. The augmented signals are derived from the original training data and primarily aim to increase signal diversity to counteract

TABLE IV
EXPERIMENTS OF FIVE EEG MODELS FOR EEG72 AND THINGS-EEG-5Hz DATASETS.

Method	EEG72		THINGS-EEG-5Hz
	6-class	72-class	200-class
linear	47.48 \pm 0.41	17.84 \pm 0.44	22.47 \pm 2.81
+ Time Masking	47.48 \pm 0.44	18.50 \pm 0.62*	23.50 \pm 3.15**
+ Noise Addition	47.57 \pm 0.56	18.57 \pm 0.41*	23.69 \pm 2.97**
ShallowConvNet [40]	36.95 \pm 0.34	9.38 \pm 0.42	14.74 \pm 1.34
+ Time Masking	39.48 \pm 0.39**	10.61 \pm 0.51*	15.66 \pm 1.73**
+ Noise Addition	39.46 \pm 0.50**	10.77 \pm 0.59*	15.56 \pm 1.81*
Temporal CNN [30]	51.33 \pm 0.23	22.16 \pm 0.46	37.14 \pm 3.75
+ Time Masking	51.26 \pm 0.35	23.67 \pm 0.61**	38.43 \pm 3.95**
+ Noise Addition	51.21 \pm 0.33	23.89 \pm 0.44**	38.47 \pm 3.69**
ATM [11]	50.44 \pm 0.55	22.86 \pm 0.44	36.65 \pm 3.32
+ Time Masking	50.44 \pm 0.43	24.47 \pm 0.33**	36.40 \pm 3.61
+ Noise Addition	50.97 \pm 0.32*	25.11 \pm 0.62**	37.05 \pm 3.46*
Hybrid [33]	53.20 \pm 0.43	25.01 \pm 0.82	37.19 \pm 2.79
+ Time Masking	53.39 \pm 0.39	27.26 \pm 0.47**	38.45 \pm 3.08**
+ Noise Addition	53.66 \pm 0.50*	26.73 \pm 0.55**	39.73 \pm 3.13**

where * denotes the classification accuracy of the model trained with this specific augmentation are significantly better than the baseline (paired t-test $p < 0.05$). ** denotes $p < 5e^{-4}$.

TABLE V
EXPERIMENTS ON EEG72 AND THINGS-EEG-5Hz DATASETS FOR DIFFERENT NOISE AMPLITUDE.

Noise Amplitude	EEG72		THINGS-EEG-5Hz
	6-class	72-class	200-class
0	53.20 \pm 0.43	25.01 \pm 0.82	37.19 \pm 2.79
0.0025	53.60 \pm 0.23*	27.09 \pm 0.37**	39.95 \pm 3.29**
0.005	53.72 \pm 0.39*	27.26 \pm 0.52**	39.67 \pm 3.00**
0.01	<u>53.66</u> \pm 0.50*	27.26 \pm 0.47**	39.73 \pm 3.13**
0.02	53.54 \pm 0.27	26.99 \pm 0.63**	39.77 \pm 3.41**
0.03	53.46 \pm 0.42	26.70 \pm 0.79**	39.46 \pm 3.10**

where **bold** fonts indicate the best results, and underlined fonts are the second best results. * denotes the classification accuracy of the model trained with averaged signals are significantly better than the baseline (paired t-test $p < 0.05$). ** denotes $p < 5e^{-4}$.

noise interference. This can assist the model in better capturing semantic features and reducing the effects of noise. However, since EEG signals are non-stationary, their characteristics often vary significantly over time. Thus, merely alleviating noise is insufficient; exposing the model to a broader range of diverse data across different sessions is crucial.

B. Data Augmentation on Different EEG Models

To investigate the effect of data augmentation on different EEG models, we conducted experiments on two EEG datasets with five models. Table IV presents the experimental results using two data augmentation methods: Time Masking and Noise Addition. The experimental results indicate that for the 72-class task using the EEG72 dataset and the 200-class task using the THINGS-EEG-5Hz dataset, classification accuracy consistently improved with the application of data augmentation methods across all five models. In contrast, for the 6-class task using the EEG72 dataset, the effect of data augmentation methods was weak in most cases. This suggests that data augmentation is more effective for classification tasks with a greater number of categories, as there tends to be a smaller sample size for each class. Conversely, in classification tasks with fewer categories, the sample size for each class is

often sufficient. This means original samples likely containing similar features to the augmented samples. Consequently, the effect of data augmentation is less significant in classification tasks with fewer categories, where there are ample samples available [43].

C. Impact of Noise Addition

In this section, we evaluated the impact of noise amplitude on the Noise Addition approach across two EEG datasets. The classification performance at different noise amplitudes is presented in Table V. The experimental results indicate that noise amplitude influences the effectiveness of data augmentation. For the EEG72 dataset, optimal noise amplitudes are set to 0.005 or 0.01 to achieve favourable results. In the case of the THINGS-EEG-5Hz dataset, a noise amplitude of 0.0025 is effective. Furthermore, it is important to note that excessive noise amplitude can lead to a decline in performance.

D. Impact of Time Masking

In this section, we evaluated the impact of the number of masks and the length of masks on the Time Masking approach across two EEG datasets. The "Mask Number" refers to the quantity of continuous masks applied, while "Mask Length"

TABLE VI
EXPERIMENTS ON EEG72 AND THINGS-EEG-5Hz DATASETS FOR DIFFERENT MASK RATIO.

Mask Number	Mask Length	EEG72			THINGS-EEG-5Hz
		6-class	72-class	200-class	
0	0	53.20 ± 0.43	25.01 ± 0.82	37.19 ± 2.79	
1	1	53.39 ± 0.39	<u>26.73</u> ± 0.55**	<u>38.45</u> ± 3.08**	
1	2	53.11 ± 0.37	25.94 ± 0.47*	37.12 ± 3.06	
2	1	53.30 ± 0.32	<u>26.63</u> ± 0.41**	37.78 ± 2.83*	
1	4	52.42 ± 0.42	24.31 ± 0.30	34.25 ± 2.94	
2	2	52.53 ± 0.52	24.86 ± 0.43	35.00 ± 2.78	
4	1	52.91 ± 0.47	25.91 ± 0.37*	36.29 ± 2.99	

where **bold** fonts indicate the best results, and underlined fonts are the second best results. * denotes the classification accuracy of the model trained with averaged signals are significantly better than the baseline (paired t-test $p<0.05$). ** denotes $p<5e^{-4}$.

TABLE VII
EXPERIMENTS ON EEG72 AND THINGS-EEG-5Hz FOR DIFFERENT NUMBER OF AVERAGED SIGNALS.

Number of Averaged Signals	EEG72			THINGS-EEG-5Hz
	6-class	72-class	200-class	
1	53.20 ± 0.43	25.01 ± 0.82	37.19 ± 2.79	
2	<u>54.14</u> ± 0.39**	28.10 ± 0.47**	43.62 ± 3.26**	
3	54.07 ± 0.34*	28.54 ± 0.37**	44.54 ± 3.35**	
4	54.09 ± 0.34**	28.55 ± 0.34**	45.69 ± 3.24**	
5	53.91 ± 0.58*	28.52 ± 0.56**	46.78 ± 3.30**	
6	53.93 ± 0.64*	28.68 ± 0.48**	47.53 ± 3.39**	
7	53.67 ± 0.40	28.55 ± 0.33**	<u>47.85</u> ± 3.49**	
8	53.70 ± 0.39	28.42 ± 0.31**	47.77 ± 3.50**	
9	53.70 ± 0.40	28.24 ± 0.70**	47.92 ± 3.57**	
10	53.64 ± 0.53	27.94 ± 0.40**	47.77 ± 3.59**	
Non-overlap	54.33 ± 0.24**	<u>28.67</u> ± 0.54**	45.68 ± 3.51**	

where **bold** fonts indicate the best results, and underlined fonts are the second best results. * denotes the classification accuracy of the model trained with averaged signals are significantly better than the baseline (paired t-test $p<0.05$). ** denotes $p<5e^{-4}$.

refers to the duration of each mask. It is important to note that when implementing the Time Masking approach, we apply random masking to the signal of each electrode individually. Afterwards, we merge the masked signals from all electrodes to create an augmented sample. The classification performance under different mask settings is presented in Table VI. For the two datasets, the accuracy was highest when the mask rate was the lowest. We observed a decrease in accuracy as the mask rate increased. These experimental results suggest the mask ratio should be kept at a lower value for EEG visual classification.

E. Impact of Averaged Signals

In this section, we evaluated the impact of different signal-averaging methods on the EEG72 and THINGS-EEG-5Hz datasets. The classification performance under different signal-averaging methods is presented in Table VII. For the 6-class task on the EEG72 dataset, the highest accuracy was achieved using non-overlapping averaged signals. In the case of the 72-class task on the EEG72 dataset, the best performance was obtained with overlapping averaged signals, using an average of 6 samples. The highest accuracy was also obtained with overlapping averaged signals for the 200-class task on the THINGS-EEG-5Hz dataset, using an average of 9 samples.

It is worth noting that in the 6-class task using the EEG72 dataset, the model's performance declined as the average number of overlapping signals increased. Conversely, in the 72-

class task with the same EEG72 dataset, performance initially improved with a rising average number of overlapping signals before eventually decreasing. In contrast, in the 200-class task with the THINGS-EEG-5Hz dataset, the model's performance consistently improved as the average number of overlapping signals increased. This suggests that as the difficulty of the classification task increases, the requirement for signal quality rises. Consequently, the effect of signal-averaging at higher average numbers tends to be more beneficial for the 200-class task with the THINGS-EEG-5Hz dataset. However, for the EEG72 dataset, it is essential to note that the average number should not be excessively large. A very high average number may lead to signal over-smoothing, resulting in a significant difference between the training data distribution and the single-trial signals in the test set.

Additionally, for the EEG72 dataset, non-overlapping averaging generally outperformed most of the overlapping averaging methods. Conversely, for the THINGS-EEG-5Hz dataset, non-overlapping averaging exhibited weaker performance compared to overlapping averaging with a higher average number. This discrepancy arises because different tasks have varying preferences for average numbers. In the non-overlapping averaging method, a higher average number results in fewer samples, which diminishes its effectiveness on the THINGS-EEG-5Hz dataset, where a higher average number is preferable.

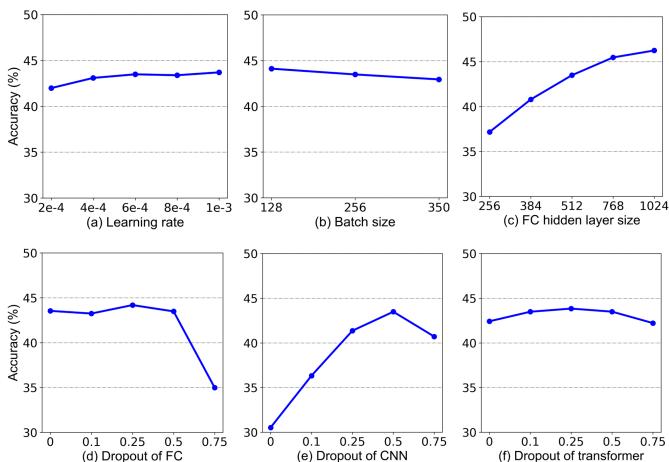


Fig. 7. Performance comparison with different hyperparameters on THINGS-EEG-5Hz dataset. (a) Learning rate. (b) Batch size. (c) FC hidden layer size. (d) Dropout of FC. (e) Dropout of CNN. (f) Dropout of transformer.

F. Parameter Sensitivity

Deep learning models are sensitive to hyperparameter settings, making it essential to optimize these parameters to achieve high performance. In this section, we conducted experiments on the THINGS-EEG-5Hz dataset and made adjustments to the hyperparameters related to both training and model structure.

The learning rate and batch size are fundamental parameters that impact the training of deep learning algorithms. From Fig. 7 (a), it is evident that setting the learning rate around 6e-4 resulted in desirable accuracies. A learning rate of 2e-4 may cause the model to fall into a local optimum, leading to lower accuracy. On the other hand, a learning rate of 1e-3 also demonstrated promising accuracy, indicating that further adjustments to the learning rate could enhance performance.

Fig. 7 (b) shows that a batch size of 128 produces the highest accuracy. A smaller batch size allows for more frequent model updates, increasing the chances of escaping local optima, which can result in higher accuracy. The experimental results suggest that a lower batch size may further enhance performance.

As illustrated in Fig. 7 (c), using an increasing size of the fully connected (FC) hidden layer enhances the model's performance. The results suggest that even better performance could be achieved by further increasing the parameters or optimizing the structure.

Fig. 7 (d), (e), and (f) demonstrate that the dropout rate has a significant impact on the model's performance. Specifically, the model performs significantly better when the dropout rate for the FC model is lower than 0.5, while performance drops sharply at a rate of 0.75. The model's accuracy is highly sensitive to the dropout rate of the convolutional neural network (CNN), with differences of nearly 15% between the highest and lowest accuracy rates. Conversely, the dropout rate of the transformer model appears to have little influence on performance.

G. Limitations and Future Work

For data augmentation methods, we focused exclusively on temporal and spatial transformations. Future work could explore the effects of frequency domain transformations and generative models for data augmentation in the EEG visual classification task. In our study, we applied data augmentation methods to single-subject tasks and did not combine data from different subjects. For future research involving cross-subject tasks, researchers might consider exchanging features between same-class samples from different subjects as a form of data augmentation. This approach may help models overlook differences in subject-dependent features and focus on task-related features. In the future, we plan to investigate self-supervised pre-training methods that integrate data augmentation for contrastive learning. By training a pre-trained model for EEG visual classification, we aim to enhance classification accuracy and explore cross-subject training algorithms that improve the model's generalization performance across subjects.

V. CONCLUSION

In this study, we introduced data augmentation methods to address the issue of insufficient data in deep learning training for the EEG visual classification task. We conducted experiments on two EEG datasets, and the results demonstrated that strategies such as Noise Addition and Time Masking effectively improved the classification accuracy of five EEG models. Additionally, we combined the augmented samples with averaged signals, leading to further improvements in the 6-class task of the EEG72 dataset. For the 72-class task of EEG72 and the 200-class task of THINGS-EEG-5Hz, training with averaged signals alone showed better performance than training with augmented samples. We compared the effectiveness of various data augmentation strategies and found that techniques such as Time Reverse, Sign Flip, Channel Reflection, and Time Shift were less effective for EEG visual classification tasks. Our experiments involved varying training data volumes and augmented sample sizes. The results indicated that using the entire training set consistently led to improvements in accuracy as the size of the augmented samples increased. However, data augmentation did not typically enhance performance when the training data volume was small. We also examined the impact of data augmentation across different parameter settings and investigated how various levels of averaged data influenced classification accuracy. The results suggested that as the difficulty of the classification task increased, there was a greater need for high-quality signals averaged from a larger number of raw signals. Additionally, parameter sensitivity experiments demonstrated that the dropout rate significantly affects the performance of artificial neural networks. Overall, this work is expected to advance research in EEG visual classification using artificial neural networks.

REFERENCES

- [1] 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.

- [2] S. Vaid, P. Singh, and C. Kaur, "Eeg signal analysis for bci interface: A review," in *2015 fifth international conference on advanced computing & communication technologies*. IEEE, 2015, pp. 143–147.
- [3] K. Värbü, N. Muhammad, and Y. Muhammad, "Past, present, and future of eeg-based bci applications," *Sensors*, vol. 22, no. 9, p. 3331, 2022.
- [4] I. Simanova, M. Van Gerven, R. Oostenveld, and P. Hagoort, "Identifying object categories from event-related eeg: toward decoding of conceptual representations," *PloS one*, vol. 5, no. 12, p. e14465, 2010.
- [5] B. Kaneshiro, M. Perreau Guimaraes, H.-S. Kim, A. M. Norcia, and P. Suppes, "A representational similarity analysis of the dynamics of object processing using single-trial eeg classification," *Plos one*, vol. 10, no. 8, p. e0135697, 2015.
- [6] C. Spampinato, S. Palazzo, I. Kavasidis, D. Giordano, N. Souly, and M. Shah, "Deep learning human mind for automated visual classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6809–6817.
- [7] T. Grootswagers, A. K. Robinson, and T. A. Carlson, "The representational dynamics of visual objects in rapid serial visual processing streams," *NeuroImage*, vol. 188, pp. 668–679, 2019.
- [8] W. Li, P. Zhao, C. Xu, Y. Hou, W. Jiang, and A. Song, "Deep learning for eeg-based visual classification and reconstruction: Panorama, trends, challenges and opportunities," *IEEE Transactions on Biomedical Engineering*, 2025.
- [9] S. Bagchi and D. R. Bathula, "Eeg-convtransformer for single-trial eeg-based visual stimulus classification," *Pattern Recognition*, vol. 129, p. 108757, 2022.
- [10] Z. Ye, L. Yao, Y. Zhang, and S. Gustin, "Self-supervised cross-modal visual retrieval from brain activities," *Pattern Recognition*, vol. 145, p. 109915, 2024.
- [11] D. Li, C. Wei, S. Li, J. Zou, and Q. Liu, "Visual decoding and reconstruction via eeg embeddings with guided diffusion," *arXiv preprint arXiv:2403.07721*, 2024.
- [12] C.-S. Chen and C.-S. Wei, "Mind's eye: Image recognition by eeg via multimodal similarity-keeping contrastive learning," *arXiv preprint arXiv:2406.16910*, 2024.
- [13] J. Luo, W. Cui, S. Xu, L. Wang, X. Li, X. Liao, and Y. Li, "A dual-branch spatio-temporal-spectral transformer feature fusion network for eeg-based visual recognition," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 2, pp. 1721–1731, 2023.
- [14] C. Du, K. Fu, J. Li, and H. He, "Decoding visual neural representations by multimodal learning of brain-visual-linguistic features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- [15] Y. Song, B. Liu, X. Li, N. Shi, Y. Wang, and X. Gao, "Decoding natural images from eeg for object recognition," *arXiv preprint arXiv:2308.13234*, 2023.
- [16] H. Ahmed, R. B. Wilbur, H. M. Bharadwaj, and J. M. Siskind, "Object classification from randomized eeg trials," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3845–3854.
- [17] T. Grootswagers, I. Zhou, A. K. Robinson, M. N. Hebart, and T. A. Carlson, "Human eeg recordings for 1,854 concepts presented in rapid serial visual presentation streams," *Scientific Data*, vol. 9, no. 1, p. 3, 2022.
- [18] A. T. Gifford, K. Dwivedi, G. Roig, and R. M. Cichy, "A large and rich eeg dataset for modeling human visual object recognition," *NeuroImage*, vol. 264, p. 119754, 2022.
- [19] S. Xue, B. Jin, J. Jiang, L. Guo, J. Zhou, C. Wang, and J. Liu, "A multi-subject and multi-session eeg dataset for modelling human visual object recognition," *Scientific Data*, vol. 12, no. 1, p. 663, 2025.
- [20] J. Huang, H. Wen, P. Yang, K. Yang, and B. Xiong, "Shift aliasing signal data augmentation method for ssvep-based bcis," in *2024 30th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*. IEEE, 2024, pp. 1–6.
- [21] Z. Zhang, Y. Liu, and S.-h. Zhong, "Ganser: A self-supervised data augmentation framework for eeg-based emotion recognition," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 2048–2063, 2022.
- [22] Y. Pei, Z. Luo, Y. Yan, H. Yan, J. Jiang, W. Li, L. Xie, and E. Yin, "Data augmentation: Using channel-level recombination to improve classification performance for motor imagery eeg," *Frontiers in Human Neuroscience*, vol. 15, p. 645952, 2021.
- [23] H. Jia, Z. Huang, C. F. Caiafa, F. Duan, Y. Zhang, Z. Sun, and J. Solé-Casals, "Assessing the potential of data augmentation in eeg functional connectivity for early detection of alzheimer's disease," *Cognitive Computation*, vol. 16, no. 1, pp. 229–242, 2024.
- [24] E. Khalili and B. M. Asl, "Automatic sleep stage classification using temporal convolutional neural network and new data augmentation technique from raw single-channel eeg," *Computer Methods and Programs in Biomedicine*, vol. 204, p. 106063, 2021.
- [25] Y. Li, X.-R. Zhang, B. Zhang, M.-Y. Lei, W.-G. Cui, and Y.-Z. Guo, "A channel-projection mixed-scale convolutional neural network for motor imagery eeg decoding," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 6, pp. 1170–1180, 2019.
- [26] C. Rommel, T. Moreau, J. Paillard, and A. Gramfort, "Cadda: Class-wise automatic differentiable data augmentation for eeg signals," *arXiv preprint arXiv:2106.13695*, 2021.
- [27] W. Ding, A. Liu, L. Guan, and X. Chen, "A novel data augmentation approach using mask encoding for deep learning-based asynchronous ssvep-bci," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 875–886, 2024.
- [28] Z. Wang, S. Li, J. Luo, J. Liu, and D. Wu, "Channel reflection: Knowledge-driven data augmentation for eeg-based brain-computer interfaces," *Neural Networks*, vol. 176, p. 106351, 2024.
- [29] S. Bagchi and D. R. Bathula, "Adequately wide 1d cnn facilitates improved eeg based visual object recognition," in *2021 29th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1276–1280.
- [30] J. Kalafatovich, M. Lee, and S.-W. Lee, "Learning spatiotemporal graph representations for visual perception using eeg signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 97–108, 2022.
- [31] Z. Jiao, H. You, F. Yang, X. Li, H. Zhang, and D. Shen, "Decoding eeg by visual-guided deep neural networks," in *IJCAI*, vol. 28. Macao, 2019, pp. 1387–1393.
- [32] M. Ferrante, T. Boccato, S. Bargione, and N. Toschi, "Decoding visual brain representations from electroencephalography through knowledge distillation and latent diffusion models," *Computers in Biology and Medicine*, p. 108701, 2024.
- [33] S. Xue, B. Jin, J. Jiang, L. Guo, and J. Liu, "A hybrid local-global neural network for visual classification using raw eeg signals," *Scientific Reports*, vol. 14, no. 1, p. 27170, 2024.
- [34] C. He, J. Liu, Y. Zhu, and W. Du, "Data augmentation for deep neural networks model in eeg classification task: a review," *Frontiers in Human Neuroscience*, vol. 15, p. 765525, 2021.
- [35] C. Rommel, J. Paillard, T. Moreau, and A. Gramfort, "Data augmentation for learning predictive models on eeg: a systematic comparison," *Journal of Neural Engineering*, vol. 19, no. 6, p. 066020, 2022.
- [36] C. Q. Lai, H. Ibrahim, M. Z. Abdullah, J. M. Abdullah, S. A. Suandi, and A. Azman, "Artifacts and noise removal for electroencephalogram (eeg): A literature review," in *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*. IEEE, 2018, pp. 326–332.
- [37] M. A. Boudevyn, S. J. Luck, J. L. Farrell, and E. S. Kappenstein, "How many trials does it take to get a significant erp effect? it depends," *Psychophysiology*, vol. 55, no. 6, p. e13049, 2018.
- [38] M. N. Hebart, A. H. Dickter, A. Kidder, W. Y. Kwok, A. Corriveau, C. Van Wicklin, and C. I. Baker, "Things: A database of 1,854 object concepts and more than 26,000 naturalistic object images," *PloS one*, vol. 14, no. 10, p. e0223792, 2019.
- [39] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen, "Mne software for processing meg and eeg data," *neuroimage*, vol. 86, pp. 446–460, 2014.
- [40] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [41] Z. Miao, M. Zhao, X. Zhang, and D. Ming, "Lmda-net: A lightweight multi-dimensional attention network for general eeg-based brain-computer interfaces and interpretability," *NeuroImage*, p. 120209, 2023.
- [42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [43] F. Fahimi, S. Dosen, K. K. Ang, N. Mrachacz-Kersting, and C. Guan, "Generative adversarial networks-based data augmentation for brain-computer interface," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 9, pp. 4039–4051, 2020.