





Sample-Based Data Augmentation Based on Electroencephalogram Intrinsic Characteristics

Ruilin Li , Lipo Wang , *Senior Member, IEEE*, P. N. Suganthan , *Fellow, IEEE*, and Olga Sourina , *Senior Member, IEEE*

Abstract—Deep learning for electroencephalogram-based classification is confronted with data scarcity, due to the time-consuming and expensive data collection procedure. Data augmentation has been shown as an effective way to improve data efficiency. In addition, contrastive learning has recently been shown to hold great promise in learning effective representations without human supervision, which has the potential to improve the electroencephalogram-based recognition performance with limited labeled data. However, heavy data augmentation is a key ingredient of contrastive learning. In view of the limited number of sample-based data augmentation in electroencephalogram processing, three methods, performance-measure-based time warp, frequency noise addition and frequency masking, are proposed based on the characteristics of electroencephalogram signal. These methods are parameter learning free, easy to implement, and can be applied to individual samples. In the experiment, the proposed data augmentation methods are evaluated on three electroencephalogram-based classification tasks, including situation awareness recognition, motor imagery classification and brain-computer interface steady-state visually evoked potentials speller system. Results demonstrated that the convolutional models trained with the proposed data augmentation methods yielded significantly improved performance over baselines. In overall, this work provides more potential methods to cope with the problem of limited data and boost the classification performance in electroencephalogram processing.

Index Terms—Data augmentation, electroencephalogram, situation awareness, motor imagery, SSVEP.

Manuscript received 30 December 2021; revised 11 June 2022; accepted 17 June 2022. Date of publication 23 June 2022; date of current version 5 October 2022. (Corresponding author: Ruilin Li.)

Ruilin Li and Olga Sourina are with Cognitive Human-Machine Interaction Lab, Fraunhofer, Singapore 639798, and also with Nanyang Technological University, Singapore 639798 (e-mail: ruilin001@e.ntu.edu.sg; EOSourina@ntu.edu.sg).

Lipo Wang is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: elp-wang@ntu.edu.sg).

P. N. Suganthan is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, and also with the KINDI Center for Computing Research, College of Engineering, Qatar University, Doha 2713, Qatar (e-mail: epsugan@ntu.edu.sg).

This article has supplementary downloadable material available at <https://doi.org/10.1109/JBHI.2022.3185587>, provided by the authors.

Digital Object Identifier 10.1109/JBHI.2022.3185587

I. INTRODUCTION

ELECTROENCEPHALOGRAM (EEG) signals are related to human brain activities and have been used to recognize different human states [1]. Various machine learning models have been applied for EEG processing such as universum support vector machine [2]. Furthermore, convolutional neural network (CNN) has been shown as a promising method to enhance and simplify EEG-based classification pipelines [3]. Due to the difficulty to get rich EEG data, improving the classification accuracy of CNN is a research focus in EEG processing. Specifically, when a limited number of data is used to train an exceedingly complex model, such as having too many parameters compared to the number of training samples, over-fitting might happen and weaken the classification ability of models. Other fields such as image processing have shown that using data augmentation to improve data efficiency can help obtain better classification accuracy of the model [4]. Therefore, well-designed data augmentation methods are required for EEG processing.

Previous works mainly employed two modes of data augmentation in EEG processing. One was to generate synthetic samples typically by combining different samples. The swaps between segments in frequency domain [5] and time domain [6] were typical methods to combine samples. Furthermore, attempts have been made to mix the decomposed components such as performing crossover of independent components obtained by independent components analysis (ICA) [7], and intermixing the intrinsic mode function obtained by empirical mode decomposition [8]. Apart from the mentioned combination methods, using generative adversarial networks (GANs) [9] to design data augmentation can be regarded as an interpolation among training samples. For instance, Fan *et al.* [10] exploited GAN to perform data augmentation in EEG processing, leading to the significantly improved classification performance. Similarly, Xu *et al.* [11] proposed a balanced Wasserstein GAN with gradient penalty to generate minority class data in the rapid serial visual presentation classification task. Luo *et al.* [12] proposed a selective Wasserstein GAN to augment differential entropy (DE) features and improve the performance of emotion recognition. Fu *et al.* [13] exploited GAN to generate samples and proposed a hybrid model of broad-deep networks to increase the similarity between generated features and real features.

Alternatively, sample-based data augmentation was another direction, which was to generate synthetic samples by performing a transformation on each sample. A classical method, noise

addition was an example of sample-based data augmentation, which has shown its effectiveness for machine learning [14]. Other examples of sample-based data augmentation include scaling [15] and flipping [16]. In addition, the over-sampling methods were also exploited to perform data augmentation in EEG processing such as using the synthetic minority over-sampling technique (SMOTE) [17] for class imbalance problem [18]. In EEG processing, besides improving classification performance, the sample-based methods can also be used for learning invariant features from real samples and synthetic samples, *e.g.*, contrastive learning [19]. However, the number of validated sample-based data augmentation methods is limited in EEG processing.

Furthermore, previous works mainly performed data augmentation on extracted hand-engineering features for EEG-based classification tasks. For example, Wang *et al.* [20] added Gaussian noise to the training data which was extracted DE features for emotion recognition. The recognition performance showed significant improvements. Yin *et al.* [21] applied Gaussian noise on the feature vector to increase the recognition accuracy. With the development of end-to-end deep learning for EEG-based classification tasks, data augmentation methods on raw EEG signals are also required. However, less attention has been paid to that.

To design the sample-based data augmentation methods that can be applied to raw EEG signals, the intrinsic properties of EEG signals should be considered. Firstly, subjects may have different “mind speeds” (*i.e.*, EEG signals with similar envelopes may have a different period). However, limited number of subjects and data restricts the variance of the periods. Therefore, time warp is a potential data augmentation method to solve this problem. Um *et al.* [22] proposed a time warp method to be used on the data augmentation of wearable sensor data for Parkinson’s disease monitoring, which stretched and compressed the signals to the different extents in different regions. To apply time warp to EEG processing, the intrinsic properties should be considered. One example of such application was that Bassi *et al.* [23] used time warp on the spectrograms of steady-state visually evoked potentials (SSVEP) signals. However, the transformation was only studied in the time-frequency domain. The design of time warp in the time domain and the use of the method for different types of EEG signals remain to be further investigated.

Secondly, the frequency domain information also plays an important role in classification. Previous works mainly explored data augmentation methods in the time domain. For example, Kang *et al.* [7] proposed an ICA - evolution method to generate synthetic samples in the time domain. Moreover, spatial characteristics were exploited to design data augmentation. Krell *et al.* [24] proposed to rotate electrodes along three main axes of the head and then perform interpolation to generate synthetic data. However, limited studies investigated the data augmentation methods in the frequency domain. One related example was that Li *et al.* [25] tried noise addition on the magnitude after short-time Fourier transformation (STFT). Since frequency information can represent different human brain activities [26], more data augmentation methods in the frequency domain such as noise addition and occlusion methods require to be further investigated.

To address the aforementioned problems, three data augmentation methods are proposed based on the intrinsic properties of EEG signals. Specifically, a performance-measure-based (PMB) time warp is proposed to simulate similar signals but with period change, aiming to reduce the non-stationary problem in EEG datasets. Moreover, phase information is considered in the design of frequency noise addition. The performance of using noise addition on the magnitude, phase and their combination are compared in this work. To the author’s best knowledge, this is the first time to compare the use of noise addition on different components in the frequency domain in EEG processing. In addition, this work proposes to investigate frequency masking in different EEG-based classification tasks. The proposed methods are evaluated in three different EEG-based classification tasks, including situation awareness (SA) recognition, motor imagery (MI) classification and brain-computer interface (BCI) SSVEP speller system. Results demonstrated that the proposed methods could help to boost classification performance. In overall, this work provides more potential methods to cope with the problem of limited data in EEG processing.

II. METHODOLOGY

This section presents three proposed data augmentation methods for EEG processing, namely PMB time warp, frequency noise addition, and frequency masking.

A. Performance-Measure-Based Time Warp

Time warp stretches or compresses the signal along time axes. Previous time warp method [22] performed the combination of stretching and compressing with varying extent on different parts of the signals in a single data extraction window. However, in the experiment of inducing different human states, it is highly likely that the condition of the brain’s functional state stays consistent within the short data extraction window. Hence, time warp in one mode, *i.e.*, either stretching or compressing the extracted signals is more suitable for EEG signals. Furthermore, the beginning part of the extracted signals is usually less related to the induced states. For instance, SA for drivers can be labeled based on the reaction time [27]. The signal that is close to the performance measure point can better reflect the induced state. Similarly, in MI tasks, the beginning part of the extracted signals contains the cue-response action which is not related to the desired states. Therefore, this work proposes a PMB time warp to stretch or compress the signals by fixing the start point and pulling/pushing the end point of the data extraction window. The difference between the previous time warp method [22] and the proposed PMB time warp is shown in Supplementary Fig. S1.

A $m \times n$ EEG signal is denoted as $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l, \dots, \mathbf{x}_m]^T$, where $\mathbf{x}_l = [x_l(1), x_l(2), \dots, x_l(q), \dots, x_l(n)]$. The process of time warp can be divided into two steps. Firstly, the time axis is warped. For instance, the original time points $\{0, 1, 2, \dots, n\}$ are warped as $\{0, 1.1, 2.3, \dots, n + 30\}$. Denoting the warping function on time axis as $\tau(\cdot)$, the channel l of the transformed signal \mathbf{x}_{tw} with warped time axis is given by

$$\tilde{\mathbf{x}}_l = [x_l(\tau(1)), x_l(\tau(2)), \dots, x_l(\tau(q)), \dots, x_l(\tau(n))]. \quad (1)$$

A uniformly distributed random number w generated from $[-W, W]$ is exploited to define the extent of warping. When $w < 0$, the signal is compressed. When $w > 0$, the signal is stretched. The last time point of the extracted signal is warped as $n + w$, while the start point remains unchanged after the time warp. Then, a thin-plate spline is applied to compute the rest of the warped time points. Subsequently, warped signal x_{tw} is generated by applying interpolation to compute the signal value on warped time points.

B. Frequency Noise Addition

Although noise addition has been widely used as a data augmentation method for EEG processing [28], there are limited studies on noise addition in the frequency domain. Some attempts have been made to add noise to the magnitude of the transformed time-frequency domain images [25]. However, the noises in EEG signals may not only affect the magnitude but also the phase. Therefore, this work proposes to investigate the effect of adding noise on the phase, as well as the effect of adding noise on the combination of the magnitude and the phase in EEG processing. The example of frequency noise addition is shown in Supplementary Fig. S1.

The transformation that noises are added on the combination of magnitude and phase components can be written in the form:

$$\mathbf{x}_{Fnoise} = \mathcal{F}^{-1}\{X(j\omega)e^{j\omega t} + \zeta(\omega)\}, \quad (2)$$

where $X(j\omega)e^{j\omega t} = \mathcal{F}\{x(t)\}$ and $\zeta \sim \mathcal{N}(0, \eta^2)$. \mathcal{F} and \mathcal{F}^{-1} represent the Fourier transform and the inverse Fourier transform, respectively.

The frequency Gaussian noise addition on magnitude and phase can be separately expressed as (3) and (4):

$$\mathbf{x}_{Fnoise_m} = \mathcal{F}^{-1}\{[X(j\omega) + \zeta(\omega)]e^{j\omega t}\}, \quad (3)$$

$$\mathbf{x}_{Fnoise_ph} = \mathcal{F}^{-1}\{X(j\omega)e^{j[\omega t + \zeta(\omega)]}\}, \quad (4)$$

C. Frequency Masking

Data augmentation methods of masking have been used in other fields [4]. However, frequency masking has not been comprehensively studied in different EEG paradigms. This work proposes investigating frequency masking in EEG-based classification tasks. The impact of different masking values and different ranges that the masking is operated on is further studied. The specific method is described as follows.

Based on the Fast Fourier Transform (FFT) algorithm, $(n + 1)/2$ frequency points are obtained for each channel of the EEG signal. There are two hyper-parameters, namely the number of masking points I_{fm} , and the number of masking area a . The start point of the selected frequency band is randomly initialized between 0 and $n - I_{fm}$. Then, each frequency point in I_{fm} is assigned to a random value. The magnitude and the phase of the same frequency band are masked simultaneously for all channels. Finally, an inverse FFT is performed to transform the obtained frequency domain data to the time domain and these synthetic data are then used for training. The example of frequency masking is shown in Supplementary Fig. S1.

III. EXPERIMENTS

In this section, the proposed data augmentation methods are evaluated on three EEG-based classification tasks, including SA recognition, MI classification and target identification in the BCI SSVEP speller system. The specific experiment setting and the results are described as follows. For convenience, in the following tables and figures, frequency masking, frequency noise addition and PMB time warp are denoted as FreqMasking, FreqNoise and PMB TW, respectively. The subject-dependent setting and the subject-independent setting are denoted as S-dependent and S-independent, respectively.

A. Datasets Introduction and Preprocessing

Three datasets were employed in the experiments. For each dataset, raw data were exploited without further feature extraction. The specific dataset splits are described as follows:

1) *Taiwan Driving Dataset*: The pre-processed dataset was used in this study [29]. In the experiment, lane-departure events were randomly induced to make the car drift from the original cruising lane towards the left or right sides (deviation onset). Each participant was instructed to quickly compensate for this perturbation by steering the wheel (response onset) to cause the car to move back to the original cruising lane (response offset). A complete trial included events with deviation onset, response onset, and response offset.

In this study, fatigue-related SA was analyzed. Three seconds of the EEG data prior to the deviation onset were extracted. By following [27], the local reaction time (RT) and the global RT were used to label data. An alert-RT was set. When both the local and global RT were shorter than 1.5 times the alert-RT, the extracted data were labeled as “high SA”. When both the local and global RT were longer than 2.5 times the alert-RT, the data were labeled as “low SA”. Transitional states with moderate performance were not considered in this work. Then, the EEG data were down-sampled to 128 Hz. Finally, a balanced SA dataset was obtained which included 2022 samples of 11 subjects.

2) *Brain-Computer Interface Competition IV 2b Dataset*: There were two classes of EEG data, the MI of the left hand and right hand, from nine subjects (B01–B09) in the BCI Competition IV 2b dataset [31]. There were 5 sessions in the MI tasks, in which the subject received feedback only in the last 3 sessions. Each subject conducted 120 trials in the first 2 sessions, and 160 trials in the remaining sessions. The subjects were required to imagine the corresponding hand movement over a period of 4 seconds after a visual cue was shown. The 3-channel EEG signals were sampled at 250 Hz. In this work, the data from the 4th second to the 7th second of each trial were intercepted as a sample. Finally, 750 time steps were obtained in each sample.

3) *The Benchmark Dataset*: The benchmark dataset [32] was recorded in a BCI SSVEP speller experiment with 35 subjects. There were 6 blocks in the experiment. During each block, the subjects were required to watch the screen where a matrix (5×8) of 40 target characters was flickering at various frequencies (in the range 8–15.8 Hz with 0.2 Hz increments) with at least 0.5π phase difference between adjacent frequencies. The

TABLE I
CLASSIFICATION PERFORMANCE (%) OF USING DATA AUGMENTATION METHODS ON TAIWAN DRIVING DATASET

Method	S-dependent			S-independent		
	Avg. Acc.	Std.	McNemar test	Avg. Acc.	Std.	McNemar test
InterpretableCNN [30]	85.89	1.65		75.06	6.41	
Jittering [14]	88.84	1.83	$p < 0.01$	77.16	3.37	$p < 0.01$
Scaling [31]	88.74	1.53	$p < 0.01$	78.82	3.02	$p < 0.01$
Permutation [7]	88.69	2.23	$p < 0.01$	77.40	4.72	$p < 0.01$
FreqNoise	89.03	2.76	$p < 0.01$	80.55	4.22	$p < 0.01$
PMB TW	89.22	2.28	$p < 0.01$	79.93	2.84	$p < 0.01$
FreqMasking	90.19	2.23	$p < 0.01$	80.90	3.47	$p < 0.01$

EEG data were recorded through 64 channels. Each trial started with a visual cue that was displayed for 0.5 seconds on the screen to guide the subject's gaze to the desired target and then conducted the stimulation for 5 seconds that was followed by an offset of 0.5 seconds. The EEG was down-sampled to 250 Hz. The average visual latency of the subjects was approximately estimated as 140 ms in this dataset. In this work, the target identification in BCI SSVEP speller systems was regarded as a multi-class classification problem.

B. Situation Awareness Recognition

The Taiwan driving dataset was used for the SA recognition task. The performance of the model trained in the subject-dependent and subject-independent settings was evaluated. The subject-dependent setting involved training and test sets from the same subjects. In the subject-independent setting, the training set and the test set were from different subjects. The dataset split of both settings is introduced here. In the subject-dependent setting, the data of each subject were divided into training, validation and test sets in the ratio of 6 : 2 : 2. In the subject-independent setting, the data of 11 subjects were randomly divided into the training set from 7 subjects' data, the validation set from 2 subjects' data, and the test set from 2 subjects' data. Data augmentation was only performed on the training set.

The InterpretableCNN model [30] which achieved the state-of-the-art (SOTA) cross-subject SA recognition performance was adopted in the experiment. Adam optimizer with momentum $\beta_1 = 0.9$ and $\beta_2 = 0.99$ was used. The mini-batch was 32 and the learning rate was 0.001. The model was trained for 50 epochs. Cross-entropy was employed as the cost function for training. The statistical significance of the improvement was evaluated by the McNemar test [33]. The hyper-parameters of the models and the data augmentation methods were optimized by using grid search. The specific search space can be found in Supplementary Section I. This hyper-parameter optimization was also employed in all of the following EEG-based classification tasks.

1) *Recognition Evaluation*: The recognition results were shown in Table I. In the subject-dependent setting, the hyper-parameters adopted for PMB time warp, frequency noise addition and frequency masking were $W = 20$, $\eta = 1$, $I_{fm} = 10$ and $a = 1$, respectively. The frequency noise addition on the combination of the magnitude and the phase was used in this experiment. In the subject-independent setting, the hyper-parameters were $W = 10$, $\eta = 0.5$, $I_{fm} = 5$ and $a = 2$. Results

indicated that the average accuracy of the model trained with the proposed data augmentation methods showed significant improvement as compared to the baseline, demonstrating the effectiveness of the proposed data augmentation methods for SA recognition.

The effect of hyper-parameters of three proposed data augmentation methods was evaluated. Results were shown in Supplementary Fig. S2. Consistent improvement was observed under all parameter settings, which demonstrated that three methods were robust to the corresponding hyper-parameters.

Then, the proposed methods were compared with three commonly used data augmentation methods, which were jittering [20], scaling [15] and permutation [7]. Specifically, jittering added Gaussian noise in the time domain. Scaling scaled the signal of channels with different factors. Permutation divided the signal into several parts, then randomly permuted the segments, and finally concatenated these segments. From Table I, the model trained with the proposed data augmentation methods all showed consistent improvement compared with the three commonly used methods.

Subsequently, the effect of using the combination of the proposed data augmentation methods, which means applying more than one method to each sample to generate synthetic samples, was investigated. Also, the performance of using a mixture of the augmented training sets was investigated. The results were shown in Supplementary Table S5 and Table S6. It was observed in Supplementary Table S5 that the performance of using the combination of different transformations slightly outperformed the performance of using a single method. From Supplementary Table S6, the use of the mixture of augmented training sets consistently improved the performance over the baseline in both settings.

2) *Comparison Between Time Warp Methods*: The classification performance of using Time warp 1 [22] and the proposed PMB time warp was compared in Supplementary Table S7. It was observed that the average accuracy of PMB time warp was better than that of Time warp 1 in the subject-independent setting. The possible reason was that the PMB time warp which only stretched or compressed the signal rather than using a combination of warping modes like in Time warp 1 helped the model to recognize the similar signals but with different periods from different subjects. In contrast, the period difference of an individual subject may not vary much during one experiment. Thus, the two warping methods showed similar results in the subject-dependent setting.

3) *Frequency Noise Addition on Different Components*: To analyze the contribution of noise addition on the phase, the classification performance of using noise addition on different frequency components was then evaluated. Denoting the noise addition on the combination of the magnitude and the phase as FreqNoise_Combine, noise addition on the magnitude as FreqNoise_Mag, and noise addition on the phase as FreqNoise_Phase, the comparison results were shown in Supplementary Table S8. It was observed that the phase-included frequency noise addition showed better average accuracy in both settings. This may have attributed to the improved classification ability of the model by training with noise addition on phase.

4) *Analysis of Frequency Masking*: The frequency masking was further analyzed in two aspects, namely the range that the masking is operated on and the masking value. Firstly, four types of masking value were evaluated, which were 1) zero values, 2) random values generated from a Gaussian distribution with the mean and standard deviation of the masked frequency band, 3) the maximum magnitude of the masked band with unchanged phases, and 4) random Gaussian noise with the mean of 0 and standard deviation of 0.5. Moreover, the performance of using the range of 0–30 Hz and using the full range of 0–64 Hz was compared. The comparison results were presented in Supplementary Table S9.

It is observed that frequency masking with different types of masking values and different ranges consistently improved the recognition accuracy compared with the baseline of Interpretable CNN (as shown in Table I). However, masking in the range from 0 to 30 Hz did not show as good performance as the counterpart of 0–64 Hz. The possible reason was that the higher frequency band may still contain useful information to some extent. The masking across the full range (0–64 Hz) could generate samples that consider more general partial information loss cases. Regarding the four types of masking value, in the subject-independent setting, the masking value of noise obtained the best result. In the subject-dependent setting, the zero masking value showed the best performance.

5) *Metrics Analysis for Proposed Methods*: For further analysis, average precision, recall, specificity, and F1-score were computed and shown in Supplementary Table S10. Better metrics results were obtained for the model trained with the three proposed transformations in the subject-dependent setting. Generally, the proposed methods showed better performance in comparison with the baseline and the commonly used data augmentation methods. Higher specificity was also obtained by using frequency masking, which is desired when measuring the operators' SA states in dynamic systems to avoid the false detection of the low SA state. Therefore, frequency masking can be the better choice.

C. Motor Imagery Classification

The proposed data augmentation methods were further evaluated on different CNN models on BCI Competition IV 2b dataset in the subject-dependent setting. Specifically, DeterministicCNN [35] which achieved SOTA on this dataset, as well as the EEGNet-8, 2 [34] and InterpretableCNN [30] which

succeeded on EEG classification tasks were employed. For training, Adam optimizer with momentum $\beta_1 = 0.9$ and $\beta_2 = 0.99$ was used. The mini-batch was 32 and the learning rate was 0.001. The model was trained for 50 epochs and cross-entropy was employed as the cost function. For the dataset split of subject-dependent setting, the first 3 sessions were utilized as the training set, and the remaining 2 sessions were taken as the test set. The hyper-parameters of the proposed data augmentation methods were optimized by using 5-fold cross-validation. The selected hyper-parameters used for EEGNet were $W = 30$, $\eta = 0.01$, $I_{fm} = 30$ and $a = 1$. The selected hyper-parameters used for DeterministicCNN were $W = 30$, $\eta = 0.05$, $I_{fm} = 20$ and $a = 1$. The selected hyper-parameters used for InterpretableCNN were $W = 20$, $\eta = 0.01$, $I_{fm} = 30$ and $a = 2$. The statistical significance of the improvement was evaluated by the paired-sample t -test. The comparison results were shown in Table II. Results demonstrated that the proposed data augmentation methods significantly improved the classification performance of different baseline models.

An experiment on a large-scale dataset [36] that included 109 subjects was conducted as well. ShallowCNN [37] and InterpretableCNN [30] were adopted as the baseline models. The results demonstrated the effectiveness of the proposed methods on large-scale dataset. The specific results were presented in Supplementary Section II.

For MI classification, the noises from the measuring device or from other brain activities may affect the phase information of the signal. Therefore, the proposed phase-included noise addition can help to increase the classification capability of the models. Furthermore, frequency masking allows a lower reliance of CNN models on specific frequency band information for classification. As such, the model can learn from the frequency bands with partial information loss, which can help the model to be more general for recognizing different types of signal patterns. In addition, PMB time warp was applicable for EEG signals of MI as well. In MI tasks, the beginning part of the collected signals in each trial corresponded to the response to the cue and was not much related to the label. Moreover, the period of MI signals could be different across samples. Therefore, PMB time warp that fixes the start point of the data extraction window and warps the signal in one mode by pulling or pushing the end point of the data extraction window could be used to increase the diversity of the MI datasets and further improve the classification accuracy.

D. Target Identification in Brain-Computer Interface Steady-State Visually Evoked Potentials Speller System

In this section, the effect of data augmentation on BCI SSVEP signals is investigated. The SOTA deep CNN [38] was employed as the baseline method. In this experiment, the hyper-parameters of the three proposed data augmentation methods were optimized to $W = 10$, $\eta = 0.5$, $I_{fm} = 10$ and $a = 1$. The hyper-parameters of the SOTA deep CNN model and the training process were the same as that of the original work. To perform a fair comparison, this work adopted the same EEG data extraction as [38], including filtering the EEG signals into 3 sub-bands

TABLE II
CLASSIFICATION PERFORMANCE (%) OF USING THE PROPOSED DATA AUGMENTATION METHODS ON BCI COMPETITION IV 2B DATASET

Method		Subject									Avg. Acc.	Improvement in Acc.
		B01	B02	B03	B04	B05	B06	B07	B08	B09		
EEGNet [35]	w/o DA ¹	73.75	54.64	56.56	95.94	78.13	78.75	73.75	83.13	86.88	76.02	
	w FreqNoise	72.81	56.43	65.63	95.94	88.44	82.5	78.75	86.25	88.12	79.75*	3.73
	w FreqMasking	76.56	56.43	64.69	94.06	83.44	82.81	80.31	83.44	86.56	79.01*	2.99
	w PMB TW	70.63	60.36	61.56	94.69	87.19	79.38	80.63	85.94	87.81	78.94*	2.92
DeterministicCNN [36]	w/o DA	67.25	56.14	57.81	93.63	84.69	72.75	77.56	84.88	87.69	76.10	
	w FreqNoise	67.88	57.71	58.50	93.13	87.81	77.56	78.75	84.88	87.38	77.34*	1.24
	w FreqMasking	66.94	54.93	58.63	92.81	87.50	78.88	79.63	86.94	86.75	77.31*	1.21
	w PMB TW	69.88	56.64	59.38	92.81	88.13	78.50	78.25	87.25	86.00	77.72*	1.62
InterpretableCNN [30]	w/o DA	74.69	60.36	63.31	96.31	79.56	81.75	74.31	91.25	86.50	78.44	
	w FreqNoise	79.19	57.64	65.06	94.25	88.06	80.88	76.31	88.69	88.94	79.98*	1.54
	w FreqMasking	80.00	57.29	63.50	94.13	87.81	82.31	77.19	89.38	89.31	80.17*	1.73
	w PMB TW	79.25	56.36	62.31	94.69	86.94	82.69	76.38	89.69	88.56	79.78*	1.34

¹DA: data augmentation.

*Significantly different from baselines with $p < 0.05$.

TABLE III
AVERAGE ACCURACY (%) AND AVERAGE ITR (BITS/MIN) OF USING THE PROPOSED DATA AUGMENTATION METHODS ON THE BENCHMARK DATASET

	Avg. Acc.	ITR
Deep CNN [39]	74.07	217.10
FreqMasking	74.57	218.74
FreqNoise	75.50*	223.50*
PMB TW	75.57*	223.60*

*Significantly different from baselines with $p < 0.05$.

by the band-pass filter, and using the stimulation time of 0.4 seconds and 9 channels (Pz, PO3, PO5, PO4, PO6, POz, O1, Oz, O2) signals. The subject-dependent setting was employed in this experiment. The dataset split was performed as follows. One block data were randomly selected from 6 blocks data as the test set. Another block data were then randomly selected as the validation set. The remaining 4 blocks data were used as the training set.

The comparison results were shown in Table III. It was observed that the deep CNN with three data augmentation methods all presented better average accuracy than the baseline. Moreover, the paired-sample t -test was then conducted which showed that frequency noise addition and PMB time warp achieved significant improvement in average accuracy and ITRs compared with the baseline, while frequency masking only showed slight improvement. The possible reason was that the number of masked points spanned a large frequency band (10 masking points corresponded to around 25 Hz), resulting in a large information loss. Therefore, the number of masking points for signals with fewer time steps and higher sampling frequency requires to be carefully designed.

IV. DISCUSSION

Based on the results of comparative studies, the three proposed data augmentation methods can increase the classification ability of the CNN models. Furthermore, the proposed data augmentation methods can be used to cope with the limited training data problem and increase the diversity of the datasets such that the

classification performance can be increased. Specifically, time warp can be used to stretch or compress the signal, simulating different period changes of similar signals. Frequency noise addition can help the model to better learn the samples with phase noises. Frequency masking can force the model to learn more general features and have a lower reliance on specific frequency band information. In overall, the proposed data augmentation methods can be applied in various EEG-based tasks to boost the classification performance.

The practical application is discussed here. Due to the time-consuming data collection procedure, the amount of data collected for training is usually limited. One example is the limited size of MI EEG datasets that could be collected for rehabilitation engineering. Furthermore, while experiencing higher time cost, the issue of limited data collected may be even more prominent in human mental states recognition tasks because the desired mental states (*e.g.* fatigue) of some subjects may not be successfully induced in experiments, leading to fewer datasets obtained for training. In practice, taking monitoring SA for air traffic controllers as an example, to reduce the impact of the abovementioned limitations, the proposed data augmentation methods can be applied to increase data efficiency and thus, allowing significantly higher classification accuracy.

To gain a deeper understanding on the effect of the proposed data augmentation methods, three aspects of the proposed three methods in the SA recognition task were further discussed in this section.

1) *Chirality of Proposed Methods*: Data augmentation has a cost of approximation error wherever the same distribution as the original data does not hold. Lin *et al.* [39] investigated the visual chirality in image processing where the augmented image might have different distribution compared with the original image. They proposed to use self-supervised learning to evaluate the approximation error and analyze the chirality of the augmented images. In EEG processing, the measure of this approximation error has not been investigated. In this work, this self-supervised learning method was adopted to evaluate the approximation error of the proposed data augmentation methods in EEG processing. Specifically, the original data was labeled as class "0," while the augmented data was labeled as class "1". The InterpretableCNN [30] was trained to recognize the original

and transformed data. Interesting results were obtained. For PMB time warp and frequency masking, the model can achieve around 90% accuracy to classify the test set. The model trained with frequency noise addition presented random guess results on the test set with 50% accuracy. The results illustrated that chirality also exists in the proposed data augmentation methods. Similar characteristics have also been found in some well-known data augmentation methods such as flipping and random cropping [39]. In EEG processing, such properties should be taken into account when exploiting augmented data for analysis, *e.g.*, developing the models for learning invariant features from real samples and synthetic samples such as contrastive learning. The chiral features of EEG datasets will be investigated in future works.

2) Improved Recognized Regions: To visualize the useful regions for classification after using the proposed data augmentation methods, a class activation mapping (CAM) based visualization technique [30] was employed. The visualization results were shown in Supplementary Figs. S4–S6. The recognized regions of the model trained with data augmentation were discussed. Specifically, the regions of the occipital lobe and frontal lobe can be recognized after applying the proposed data augmentation methods for the hard samples that were wrongly classified as low SA state previously (as shown in Fig. S4 and Fig. S6). Previous study [40] has shown that the artifacts in the frontal lobe can be an indicator of wakeful EEG signals which is related to the high SA state. Cui *et al.* [30] have also found that the Electromyography (EMG) activities in the frontal lobe could be a common factor that affects the classification of alert state across different subjects. Alert state can be regarded as the low-level part of SA [41]. Moreover, Catherwood *et al.* [42] have shown that during the perception of SA, the occipital lobe plays an important role. In addition, the regions of the centroparietal EEG channels can be recognized for those hard samples that were wrongly classified as high SA state previously (as shown in Fig. S5). The recognition results were compatible with the previous study [30].

3) Spectrogram Visualization: To better understand the effect of data augmentations, the EEG signals were visualized using a power spectrogram-based topology map. Since the delta [30], theta, and alpha [43] band are correlated with fatigue and sustained attention which are the low-level part of SA [41], the spectrograms on these frequency bands were mainly presented. Specifically, for each sample, STFT was applied to produce the spectrogram. Each sample was divided into segments of 1 s length with an overlap of 875 ms. The final topology map was normalized to [0, 1]. The topology maps of Subject 1 data were presented in Supplementary Fig. S3. For jittering and frequency noise addition, fewer changes in spectrograms were observed compared with the original data. Furthermore, for time warp methods and frequency masking, spectrogram enhancement was observed in theta and alpha frequency bands.

V. CONCLUSION

In this work, three data augmentation approaches are proposed, namely PMB time warp, frequency noise addition, and

frequency masking, for EEG-based classification tasks. The intrinsic properties are considered in the design of the proposed methods. Specifically, 1) PMB time warp is designed to reduce the non-stationary problem in EEG processing; 2) Phase-included frequency noise addition is investigated in three different EEG-based paradigms; and 3) A comprehensive study on using frequency masking in EEG processing is conducted. The proposed methods are evaluated in three EEG-based classification tasks, including SA recognition, MI classification and SSVEP speller system. Results demonstrated that the proposed data augmentation methods can consistently improve the performance of the SOTA models in these tasks. In practice, the proposed methods can be applied to reduce the impact of the limited data problem and boost the classification performance in different EEG-based paradigms. More investigations of the proposed methods on the different types of EEG data will be conducted in the future.

REFERENCES

- [1] B. Kumar and D. Gupta, "Universum based Lagrangian twin bounded support vector machine to classify EEG signals," *Comput. Methods Programs Biomed.*, vol. 208, 2021, Art. no. 106244.
- [2] B. Richhariya and M. Tanveer, "EEG signal classification using universum support vector machine," *Expert Syst. Appl.*, vol. 106, pp. 169–182, 2018.
- [3] S. Liu *et al.*, "3DCANN: A spatio-temporal convolution attention neural network for EEG emotion recognition," *IEEE J. Biomed. Health Informat.*, early access, May 25, 2021, doi: [10.1109/JBHI.2021.3083525](https://doi.org/10.1109/JBHI.2021.3083525).
- [4] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 13001–13008.
- [5] W. Ma *et al.*, "A novel multi-branch hybrid neural network for motor imagery EEG signal classification," *Biomed. Signal Process. Control*, vol. 77, 2022, Art. no. 103718.
- [6] A. M. Roy, "An efficient multi-scale CNN model with intrinsic feature integration for motor imagery EEG subject classification in brain-machine interfaces," *Biomed. Signal Process. Control*, vol. 74, 2022, Art. no. 103496.
- [7] J.-S. Kang, S. Kavuri, and M. Lee, "ICA-evolution based data augmentation with ensemble deep neural networks using time and frequency kernels for emotion recognition from EEG-data," *IEEE Trans. Affect. Comput.*, vol. 13, no. 2, pp. 616–627, Apr.–Jun. 2022.
- [8] K. Takahashi *et al.*, "Data augmentation for convolutional LSTM based brain computer interface system," *Appl. Soft Comput.*, vol. 122, 2022, Art. no. 108811.
- [9] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2014, vol. 27, pp. 2672–2680.
- [10] J. Fan *et al.*, "EEG data augmentation: Towards class imbalance problem in sleep staging tasks," *J. Neural Eng.*, vol. 17, no. 5, Oct. 2020, Art. no. 56017.
- [11] M. Xu, Y. Chen, Y. Wang, D. Wang, Z. Liu, and L. Zhang, "BWGAN-GP: An EEG data generation method for class imbalance problem in RSVP tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 251–263, Jan. 2022. [Online]. Available: <https://ieeexplore.ieee.org/remotexs.ntu.edu.sg/document/9417097>
- [12] Y. Luo, L.-Z. Zhu, Z.-Y. Wan, and B.-L. Lu, "Data augmentation for enhancing EEG-based emotion recognition with deep generative models," *J. Neural Eng.*, vol. 17, no. 5, 2020, Art. no. 0 56021.
- [13] R. Fu, Y. Wang, and C. Jia, "A new data augmentation method for EEG features based on the hybrid model of broad-deep networks," *Expert Syst. Appl.*, vol. 202, 2022, Art. no. 117386.
- [14] C. M. Bishop, "Training with noise is equivalent to Tikhonov regularization," *Neural Comput.*, vol. 7, no. 1, pp. 108–116, 1995.
- [15] D. Freer and G.-Z. Yang, "Data augmentation for self-paced motor imagery classification with c-LSTM," *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016041.
- [16] Y. Pei *et al.*, "Data augmentation: Using channel-level recombination to improve classification performance for motor imagery EEG," *Front. Hum. Neurosci.*, vol. 15, 2021, Art. no. 645952.
- [17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

- [18] F. Mattioli, C. Porcaro, and G. Baldassarre, "A 1D CNN for high accuracy classification and transfer learning in motor imagery EEG-based brain-computer interface," *J. Neural Eng.*, vol. 18, no. 6, Dec. 2021, Art. no. 066053.
- [19] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.
- [20] F. Wang, S.-H. Zhong, J. Peng, J. Jiang, and Y. Liu, "Data augmentation for EEG-based emotion recognition with deep convolutional neural networks," in *Proc. Int. Conf. Multimedia Model.*, 2018, pp. 82–93.
- [21] Z. Yin and J. Zhang, "Cross-subject recognition of operator functional states via EEG and switching deep belief networks with adaptive weights," *Neurocomputing*, vol. 260, pp. 349–366, 2017.
- [22] T. T. Um et al., "Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks," in *Proc. 19th ACM Int. Conf. Multimodal Interact.*, 2017, pp. 216–220.
- [23] P. R. Bassi, W. Rampazzo, and R. Attux, "Transfer learning and specaugment applied to SSVEP based BCI classification," *Biomed. Signal Process. Control*, vol. 67, 2021, Art. no. 102542.
- [24] M. M. Krell and S. K. Kim, "Rotational data augmentation for electroencephalographic data," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2017, pp. 471–474.
- [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 Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1170–1180, Jun. 2019.
- [26] M. Simon et al., "EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions," *Clin. Neurophysiol.*, vol. 122, no. 6, pp. 1168–1178, 2011.
- [27] R. Li, L. Wang, and O. Sourina, "Subject matching for cross-subject EEG-based recognition of driver states related to situation awareness," *Methods*, vol. 202, pp. 136–143, 2021.
- [28] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. W. Shalaby, "EEG-based emotion recognition using 3D convolutional neural networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 329–337, 2018.
- [29] Z. Cao, C.-H. Chuang, J.-K. King, and C.-T. Lin, "Multi-channel EEG recordings during a sustained-attention driving task," *Sci. Data*, vol. 6, no. 1, pp. 1–8, 2019.
- [30] J. Cui, Z. Lan, O. Sourina, and W. Müller-Wittig, "EEG-based cross-subject driver drowsiness recognition with an interpretable convolutional neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Feb. 16, 2022, doi: [10.1109/TNNLS.2022.3147208](https://doi.org/10.1109/TNNLS.2022.3147208).
- [31] R. Leeb, C. Brunner, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, *BCI Competition 2008–Graz Data Set B*. Austria: Graz Univ. Technol., Graz, Austria, pp. 1–6, 2008.
- [32] Y. Wang, X. Chen, X. Gao, and S. Gao, "A benchmark dataset for SSVEP-based brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1746–1752, Oct. 2017.
- [33] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.
- [34] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, Art. no. 056013.
- [35] D. M. Hermosilla et al., "Shallow convolutional network excel for classifying motor imagery EEG in BCI applications," *IEEE Access*, vol. 9, pp. 98275–98286, 2021.
- [36] A. L. Goldberger et al., "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [37] H. Dose, J. S. Møller, H. K. Iversen, and S. Puthusserypady, "An end-to-end deep learning approach to mi-EEG signal classification for BCIS," *Expert Syst. Appl.*, vol. 114, pp. 532–542, 2018.
- [38] O. B. Guney, M. Oblokulov, and H. Ozkan, "A deep neural network for SSVEP-based brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 2, pp. 932–944, Feb. 2022.
- [39] Z. Lin, J. Sun, A. Davis, and N. Snavely, "Visual chirality," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 12295–12303.
- [40] J. W. Britton et al., "Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children, and infants," Chicago, IL, USA: American Epilepsy Society, 2016.
- [41] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Hum. Factors*, vol. 37, no. 1, pp. 32–64, 1995.
- [42] D. Catherwood et al., "Mapping brain activity during loss of situation awareness: An EEG investigation of a basis for top-down influence on perception," *Hum. Factors*, vol. 56, no. 8, pp. 1428–1452, 2014.
- [43] T. Åkerstedt, G. Kecklund, and A. Knutsson, "Manifest sleepiness and the spectral content of the EEG during shift work," *Sleep*, vol. 14, no. 3, pp. 221–225, May 1991.