Poster Abstract: TCT: Zero-training two staged Contrastive Transformer network for SSVEP classification

Chenlong Wang[1], Yan Zhuo[1], Han Li[1], Xinlei Chen[1,2,3†]
[1]Shenzhen International Graduate School, Tsinghua University
[2] Pengcheng Laboratory; [3] RISC-V International Open Source Laboratory; [†] Corresponding author.

ABSTRACT

Steady-State Evoked Potential (SSVEP) is a brain response to specific frequency visual stimuli, used in brain-computer interfaces due to its robust and easily detectable signals. Researchers have long applied methods like Canonical Correlation Analysis and deep learning for SSVEP signal decomposition and classification. However, those methods struggle to classify SSVEP signals without new subject's data, and calibration is time-consuming. In this paper, we propose a two-stage, two-Transformer streams network to address the challenge of classifying SSVEP signals from new subjects. We utilize hierarchical contrastive learning to project features into a more discriminable feature space before classification. The comparative experiment demonstrates that our approach exhibits superior performance relative to alternative methods in processing SSVEP signals from new subjects.

KEYWORDS

Steady-state Visual Evoked Potential (SSVEP) , Brain-computer interface, Zero-training

1 INTRODUCTION

The Brain-computer interfaces (BCIs) are systems that enable interaction between the human brain and external devices. Non-invasive BCIs are popular due to their safety, ease of use, and wide applicability.

Steady State Visual Evoked Potential (SSVEP) is one of the most commonly utilized paradigms of non invasive BCIs. The SSVEP principle states that focusing on a flickering visual stimulus causes the brain to produce electrical activity at the same frequency. SSVEP-based BCIs offer advantages such as stronger signals for reliable detection, minimal training for better accessibility, and faster transfer rates for efficient real-time interaction[1–3].

Up to the present, numerous training-free SSVEP signal decoding models based on decomposition have been proposed. These models are capable of decoding SSVEP signals and performing classification tasks without training[4, 5]. However, they lack sufficient accuracy in classification. Similarly, there have been many training-based models based on deep learning[6, 7]. But due to the high variability of SSVEP signals, these models perform poorly when dealing with new subjects, and their calibration processes can be extremely time-consuming. Moreover, these models require that SSVEP signals be collected within the same day and from the same subject, which limit the application of BCIs. These methods are referred to as within-subject methods, and methods that do not require calibration when encountering new subjects are known as zero-training methods[8].

In this study, we propose a zero-training, two stage, two Transformer streams network with self-supervised contrastive pre-training followed by supervised fine-tuning, TCT. Experimental results demonstrate that our model exhibits strong generalization capabilities in handling the SSVEP signal classification task for new subjects, achieving the highest accuracy relative to other models.

2 METHOD

2.1 Hierarchical Contrastive learning

We employed hierarchical contrastive learning for self-supervised pre-training, allowing the model to catch data consistency across latent levels. This approach minimized contrastive loss, distinguishing SSVEP signals for different stimuli in the feature space and aligning those from the same stimuli.

2.1.1 Observation Level. For a Steady-State Visual Evoked Potential (SSVEP) signal sample x_i , we apply a random mask to obscure a portion of its features, resulting in $\tilde{x_i}$. Define t as a timestamp in a SSVEP signal sample, and t^- as another timestamp. Subsequently, an encoder structured based on two transformer streams is employed to extract features of $x_{i,t}$ and $\tilde{x_{i,t}}$, obtain $h_{i,t}$ and $\widetilde{h}_{i,t}$. The observation-level contrastive loss is defined as:

$$\mathcal{L}_{\mathrm{O}}^{i,t} = -\log \frac{\exp\left(\boldsymbol{h}_{i,t} \cdot \widetilde{\boldsymbol{h}}_{i,t}\right)}{\sum_{t^- \in \mathcal{T}} \left(\exp\left(\boldsymbol{h}_{i,t} \cdot \widetilde{\boldsymbol{h}}_{i,t^-}\right) + \exp\left(\boldsymbol{h}_{i,t} \cdot \boldsymbol{h}_{i,t^-}\right)\right)} \tag{1}$$

where $\mathcal{T}$ is the set of all timestamps. The overall loss function $\mathcal{L}_{\mathrm{O}}$ can be obtained by calculating the expected value of all $\mathcal{L}_{\mathrm{O}}^{i,t}$

2.1.2 Sample Level. Similar to the prerequisites at the observation level, given two different SSVEP signal samples x_i and x_j, we can obtain h_i and h_j by the encoder. The sample level loss can be defined as:

$$\mathcal{L}_{\mathrm{S}}^{i} = -\log \frac{\exp\left(\boldsymbol{h}_i \cdot \widetilde{\boldsymbol{h}}_i\right)}{\sum_{j=1}^{|\mathcal{D}|} \left(\exp\left(\boldsymbol{h}_i \cdot \widetilde{\boldsymbol{h}}_j\right) + \exp\left(\boldsymbol{h}_i \cdot \boldsymbol{h}_j\right)\right)} \tag{2}$$

where the $\widetilde{h}_i$ is the feature obtained by the masked x_i, $|\mathcal{D}|$ is the total number of samples. $\mathcal{L}_{\mathrm{S}}$ is the expected value of all $\mathcal{L}_{\mathrm{S}}^{i}$

Therefore, the total contrastive learning loss can be formulated as:

$$\mathcal{L} = \mathcal{L}_{\mathrm{O}} + \mathcal{L}_{\mathrm{S}} \tag{3}$$

2.2 Two Transformer stream encoder

As shown in Fig.1, the SSVEP signal encoder is based on two Transformer streams structure. At the TCT input, we feed both time-domain and frequency-domain SSVEP signal information, mapping them to a common feature space using an embedding layer and a parameter-sharing encoder. This method reduces parameters and facilitates view interaction, optimizing common feature extraction.

Then, cross-attention integrates frequency and time domain information from mixed SSVEP signals, improving the model's ability

979-8-3503-6201-5/24/$31.00 ©2024 IEEE
DOI 10.1109/IPSN61024.2024.00048

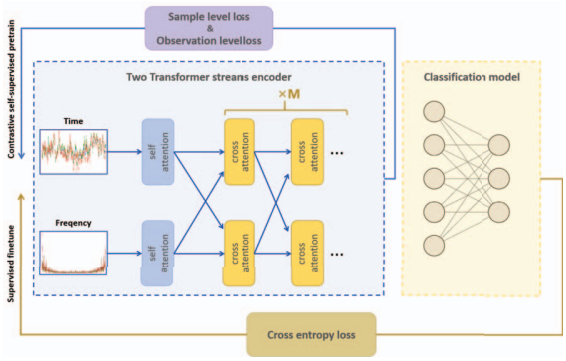

Figure 1: The structure of the proposed TCT model.

to identify complementary features for a more detailed SSVEP signal representation.

In our training pipeline, we first encode original time and frequency domain data, then pretrain this encoder using self-supervised learning to minimize contrastive loss. Subsequently, we attach the classification model to the pretrained encoder and fine-tune it for classification results.

3 EXPERIMENT

3.1 Dataset

The Benchmark[9] dataset includes SSVEP data from 35 healthy subjects using a 40-target BCI speller, with characters identified by unique frequency (8 to 15.8 Hz in 0.2 Hz steps) and phase (0 to 1.5π in 0.5π steps) modulations. The data were collected using 6 channels from occipital electrodes.

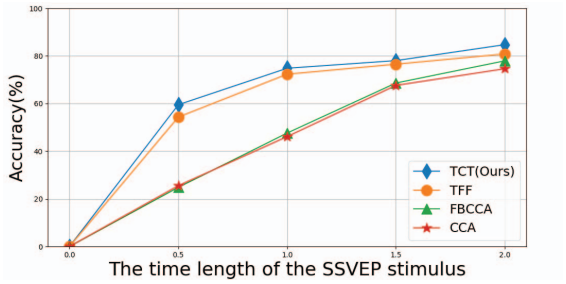

Figure 2: The line chart of the comparative experiment for the SSVEP classification task.

3.2 Comparative results

In our comparative experiments, we apply TFF[8], FBCCA[5], and CCA[4] as the baselines. Among these, TFF is a SSVEP classification model based on symmetric transformers, which previously achieved SOTA results in zero-training SSVEP signal classification tasks. FBCCA and CCA represent traditional training-free classification methods. The comparative experiment results in Fig.2 show that TCT outperforms other models in SSVEP signal processing accuracy for new subjects, even with short time lengths. This indicates TCT's strong generalization across subjects and potential for real-time applications.

4 CONCLUSION

In this paper, we propose a zero-training, two-stage, two Transformer streams network for SSVEP classification for new subjects without calibration, using hierarchical contrastive learning to enhance feature discrimination. Our method demonstrates state-of-the-art performance in processing SSVEP signals from new subjects. Future work will focus on enhancing BCI reliability by processing SSVEP signals from complex environments.

5 ACKNOWLEDGMENTS

This paper was supported by the National Key R&D program of China No. 2022YFC3300700, the Natural Science Foundation of China under Grant No. 62371269. Guangdong Innovative and Entrepreneurial Research Team Program No. 2021ZT09L197, Shenzhen 2022 Stabilization Support Program No. WDZC20220811103500001, and Tsinghua Shenzhen International Graduate School Cross-disciplinary Research and Innovation Fund Research Plan No. JC20220011.

REFERENCES

[1] Rui Na, Chun Hu, Ying Sun, Shuai Wang, Shuailei Zhang, Mingzhe Han, Wenhan Yin, Jun Zhang, Xinlei Chen, and Dezhi Zheng. An embedded lightweight ssvep-bci electric wheelchair with hybrid stimulator. *Digital Signal Processing*, 116:103101, 2021.

[2] Ying Sun, Wenzheng Ding, Xiaolin Liu, Dezhi Zheng, Xinlei Chen, Qianxin Hui, Rui Na, Shuai Wang, and Shangchun Fan. Cross-subject fusion based on time-weighting canonical correlation analysis in ssvep-bcis. *Measurement*, 199:111524, 2022.

[3] Rui Na, Dezhi Zheng, Ying Sun, Mingzhe Han, Shuai Wang, Shuailei Zhang, Qianxin Hui, Xinlei Chen, Jun Zhang, and Chun Hu. A wearable low-power collaborative sensing system for high-quality ssvep-bci signal acquisition. *IEEE Internet of Things Journal*, 9(10):7273–7285, 2021.

[4] Zhonglin Lin, Changshui Zhang, Wei Wu, and Xiaorong Gao. Frequency recognition based on canonical correlation analysis for ssvep-based bcis. *IEEE transactions on biomedical engineering*, 53(12):2610–2614, 2006.

[5] Xiaogang Chen, Yijun Wang, Shangkai Gao, Tzyy-Ping Jung, and Xiaorong Gao. Filter bank canonical correlation analysis for implementing a high-speed ssvep-based brain–computer interface. *Journal of neural engineering*, 12(4):046008, 2015.

[6] Qianxin Hui, Xiaolin Liu, Yang Li, Susu Xu, Shuailei Zhang, Ying Sun, Shuai Wang, Xinlei Chen, and Dezhi Zheng. Riemannian geometric instance filtering for transfer learning in brain-computer interfaces. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 1162–1167, 2022.

[7] Xiaolin Liu, Rongye Shi, Qianxin Hui, Susu Xu, Shuai Wang, Rui Na, Ying Sun, Wenbo Ding, Dezhi Zheng, and Xinlei Chen. Tcacnet: Temporal and channel attention convolutional network for motor imagery classification of eeg-based bci. *Information Processing & Management*, 59(5):103001, 2022.

[8] Xujin Li, Wei Wei, Shuang Qiu, and Huiguang He. Tff-former: Temporal-frequency fusion transformer for zero-training decoding of two bci tasks. In *Proceedings of the 30th ACM international conference on multimedia*, pages 51–59, 2022.

[9] Yijun Wang, Xiaogang Chen, Xiaorong Gao, and Shangkai Gao. A benchmark dataset for ssvep-based brain–computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1746–1752, 2016.