Proceedings of CONF-SEML 2026 Symposium: Multimodal Data Acquisition: Applications in Physiological and Behavioral Research
DOI: 10.54254/2755-2721/2026.AD34066

Deep Learning for Steady-State Visual Evoked Potential-Based Brain-Computer Interfaces

Chenyi Zhang

College of Communication Engineering, Jilin University, Changchun, China
cyzhang2024@mails.jlu.edu.cn

Abstract. With the rapid development of neuroscience and computer technology, Brain-Computer Interface (BCI) has become an important research field. Among various BCI paradigms, Steady-State Visual Evoked Potential (SSVEP) has attracted wide attention because of its high information transfer rate, strong anti-interference ability, and relatively simple implementation. In recent years, the release of high-quality electroencephalography (EEG) datasets and the continuous progress of deep learning have greatly promoted the development of SSVEP recognition. This paper reviews the recent progress of deep learning methods for SSVEP-based BCI systems. It first introduces the basic principles of SSVEP recognition and traditional signal processing methods, including Canonical Correlation Analysis (CCA) and its derivative methods. It then summarizes the applications of major deep learning models, such as Artificial neural networks (ANN), Convolutional neural networks (CNN), Recurrent neural networks (RNN), and Transformers, and analyzes their advantages and limitations in feature extraction, temporal modeling, cross-subject adaptation, and classification performance. Finally, the paper discusses the main challenges in current research, including model generalization, computational efficiency, interpretability, and practical deployment, and points out possible future directions for building more accurate, robust, and user-friendly SSVEP-based BCI systems.

Keywords: SSVEP, brain-computer interface, deep learning, EEG decoding

1. Introduction

Steady-state visual evoked potential (SSVEP)-based electroencephalography (EEG) systems typically consist of three components: a stimulation device, EEG signal acquisition equipment, and a signal processing module. Due to the high requirements for stability and effectiveness in BCI systems, the capability of feature extraction and recognition, which directly determines system performance, is particularly critical and remains a key research focus in the BCI field. However, traditional signal processing methods still suffer from many issues, such as low signal-to-noise ratio, poor model generalization, and unsatisfactory user experience. To overcome these limitations, applying deep learning techniques, with their powerful ability for automatic feature extraction and modeling of high-dimensional and nonlinear SSVEP signals, has become a core strategy. This paper reviews common SSVEP processing methods and the current state of deep learning applications, summarizes existing challenges in this field, and provides future perspectives, aiming to offer

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.AD34066

references for related researchers. Steady-State Visual Evoked Potential (SSVEP) is an important branch of EEG signal processing technology. A typical SSVEP-based Brain-Computer Interface (BCI) system requires users to focus on targets flickering in a specific manner, while recording an EEG, and determine the target the user is focusing on by analyzing the EEG signal. CCA is one of the earliest proposed calibration-free methods [1]. After this, more advanced algorithms such as Filter Bank Canonical Correlation Analysis (FBCCA) were also developed, demonstrating improved performance [2]. Subsequent research has mainly focused on calibration-based methods, among which Task-Related Component Analysis (TRCA) is a representative algorithm [3]. Building upon these, researchers have further developed more effective methods. With the advancement of deep learning, a series of calibration-free methods and hybrid approaches combining calibration-based methods with deep learning have been proposed, becoming a current research hotspot. In addition, commonly used EEG datasets include the Tsinghua Benchmark dataset, the Tsinghua BETA dataset, and the UCSD dataset [4].

2. Applications of deep learning in SSVEP

Deep learning is a multi-layer computational structure inspired by the human brain. By constructing complex neural networks with nonlinear processing units, it can learn hierarchical representations from massive data and perform feature extraction, enabling it to directly learn complex signals. Compared with traditional machine learning methods, it demonstrates higher performance and has been widely applied in EEG signal processing. Currently, ANN, CNN, RNN, and Transformers are common deep learning models used for SSVEP recognition.

2.1. Artificial Neural Network

The neurons in the neural network are organized in a hierarchical structure and are generally divided into three parts: the input layer, the output layer, and the hidden layer which is used to extract key features. Through forward propagation, loss computation, and backpropagation, the model learns the mapping between inputs and outputs. Compared with traditional methods, ANN can optimize all parameters end-to-end to achieve global optimality, while integrating multi-domain information and capturing multidimensional features of SSVEP signals, thus being widely used. However, due to the large number of trainable parameters, ANN usually requires a large amount of calibration data, which often leads to high power consumption for data transmission and high costs for signal acquisition. Therefore, addressing these issues has become a major research direction. To tackle the problem of excessive calibration data requirements, Wang et al. designed a more compact model. This model integrates existing knowledge with the attentional neural network and fully utilizes the high interpretability of the attention layer, eliminating the traditional spatial filtering algorithm [5]. Moreover, this model uses the shared weights between the SSVEP signal model and the stimulus as constraints, reducing inter-layer connections and compressing training parameters. Compared with existing correlation analysis-based algorithms, this method reduces more than 80% of trainable parameters and improves recognition performance by 7%. The performance comparison is shown in Figure 1. In addition, Zhang et al. proposed a personalized signal compression and reconstruction method along with a non-uniform quantization approach and applied it to SSVEP to reduce data transmission power [6]. This method first trains a baseline ANN using a dataset, then employs incremental learning to train subject-specific ANN models, thereby compressing individual data. During compression, a newly proposed non-uniform quantization method is used to reduce

DOI: 10.54254/2755-2721/2026.AD34066

compression-induced errors. This method achieves only a 3.79% decrease in recognition accuracy under a compression ratio of 12.7% on the BETA dataset.

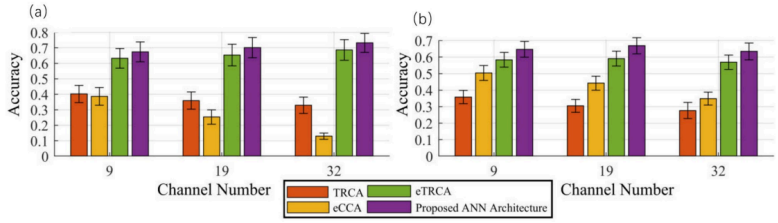

Figure 1. Comparison of the performance of the proposed method in this article with that of various CA-based methods [5] (a) The recognition result of the 0.25-second signal using the benchmark dataset. (b) The recognition result of the 0.4-second signal using the BETA dataset

2.2. Convolutional Neural Network

EEG signals are typically of low quality because of the interference of other noises and the loss incurred in the transmission process of the signal. This is also why the traditional machine learning techniques cannot be used in this area. Nevertheless, CNN is able to learn internal data relationships automatically without need to extract features manually. To solve problems like the giant individual differences, a CNN network architecture was created by Yang et al. using a parallel architecture and incorporating multi-frequency information [7]. The input of this model is multi-band signals whose role is to extract the information that the signals carry in time domain and frequency domain and therefore leverage more effectively on the signals. The model is further partitioned into three modules to extract spatial, temporal and frequency-domain features. Figure 2 shows the structure. The CNN module later fused these multi-dimensional features to attain correct classification results. The model had accuracies of 99.37 and 93.96 when tested on two datasets, respectively, and has a high degree of robustness. Moreover, Shao et al. suggested a new filter bank-based multi-branch CNN to detect SSVEP [8]. The initial component of the model is a multi-branch network architecture to extract time-domain and frequency-domain features, and it consists of two special subnetworks. The second section adds a feature fusion coefficient α to enable the model to be personalized and optimized to fit different users. The model recorded an accuracy of 85.51% in the offline testing of a signal length of 1.0s and online testing has also recorded a high performance of 96.70 which has indicated a high level of robustness and generalization.

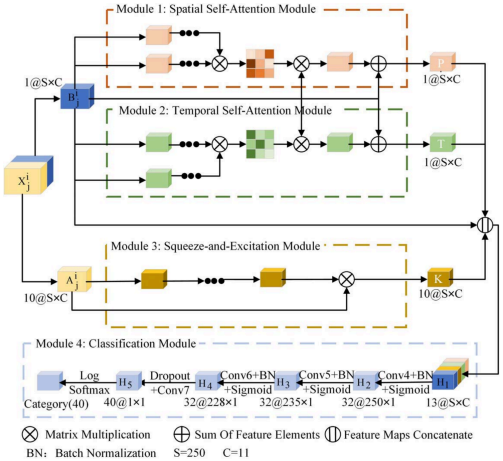

Figure 2. Overall architecture of the parallel multi-band fusion CNN [7]

DOI: 10.54254/2755-2721/2026.AD34066

2.3. Recurrent Neural Network

As feedforward neural networks do not store data, they cannot effectively handle sequential data because they can only work with the existing inputs. RNNs overcome this limitation by adding hidden layers which give internal memory. RNNs are also used in conjunction with Long Short-Term Memory (LSTM) units or Gated Recurrent Unit (GRU) units, which also increase the ability to remember over the long term and decrease the complexity of computation, respectively. As EEG signals are time-varying sequential data with temporal dependencies, it is quite appropriate to apply RNNs to model long-range dependencies. The single-channel dry electrode SSVEP-based biometric approach suggested by Gorur et al. uses three RNN architectures: LSTM, Bidirectional Long Short-Term Memory (BiLSTM) and GRU to identify individuals individually (Figure 3 represents the BiLSTM architecture [9]). They also proposed two data augmentation techniques, i.e. time-domain augmentation and spectrogram augmentation in order to solve the problems of high-noise, dependence on multi-channel signals, and challenging training in the traditional techniques. High robustness was shown using this method with a 100 percent accuracy and an ITR of 3.459 on 11 subjects. Moreover, Kim et al. suggested a two-phase asynchronous BCI system, which consists of an asynchronous model with autocorrelation and LSTM and an EEGNet-based classifier to effectively identify user intent [10]. This system was experimented on an SSVEP dataset of 40 subjects and attained a sensitivity of 98.24 +2.21% and a recognition accuracy rate of 77.05 +14.95% in a one-second time period which showed great potential.

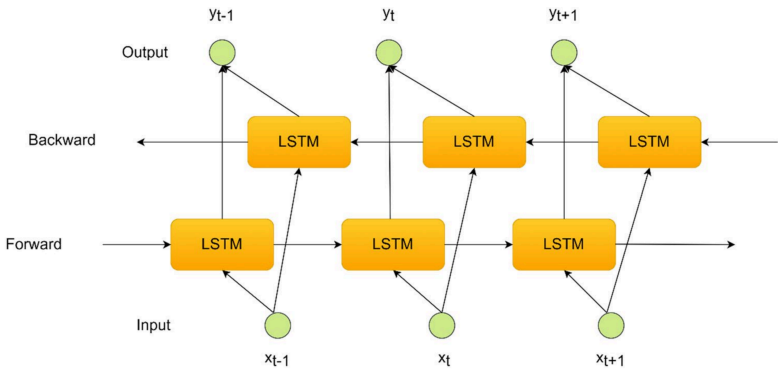

Figure 3. Schematic diagram of the BiLSTM architecture [9]

2.4. Transformer and attention mechanism

EEG signals are long time-series data, so the data at various time points could be correlated with each other, so long-range dependencies and global features could be important to capture. Transformers self-attention mechanism can capture relationships in the whole sequence, and successfully represent long-range dependencies, so they are useful in modeling global contextual information in EEG signals and in the overall brain activity patterns. Zhu et al. introduced the ID-Protoformer which is a model that improves the ability to recognize SSVEP [11]. This model is designed on Transformer and combines dynamic identity prototypes. It initially provides raw signals to information-rich, spatiotemporal embeddings, followed by accentuating discriminative features via an adaptive token reweighting procedure. The recognition benefit of this model is that it introduces subject specific perceptual signals into the token space, thus enhancing system robustness. Secondly, Liu et al. introduced the DGConformer model, a Transformer-based cross-subject model with the ability to quickly calibrate it on a person [12]. The model is a hybrid of the

DOI: 10.54254/2755-2721/2026.AD34066

attention mechanism of Transformer and parallel local CNNs to learn global temporal attributes and local information, and adds a Patch Embedding module to improve the fitting of features. An explicit domain generalization module, StableNet, is also part of the model, removing spurious correlations through reweighting of training examples, thus enhancing performance of generalization. It was tested on both the Benchmark and BETA datasets in calibration-free, calibration-based and incomplete calibration settings with greatly improved performance compared to the traditional methods FBCCA and eTRCA.

3. Future perspectives

Despite the fact that a lot of progress has been achieved towards combining SSVEP with deep learning, a number of challenges are still present. To begin with, existing EEG data are of small scale, experiment conditions are few and signal-to-noise ratios are minimal, which leads to poor generalization. Second, deep learning models with high precision tend to be computationally expensive and expensive to deploy, making the deployment of EEG-based systems less portable and practical. EEG data preprocessing could be further enhanced in future to get more clean signals with better signal-to-noise ratios. Moreover, the principles of deep learning models should be examined more thoroughly, such as quantization and knowledge distillation to create more effective lightweight models and provide a compromise between the accuracy and efficiency.

4. Conclusion

Deep learning-based SSVEP recognition is of great significance. Such methods can automatically extract complex, high-dimensional spatiotemporal features from SSVEP signals, demonstrating strong performance in scenarios with few channels and short time windows. ANN has a simple structure and is easy to implement, making it suitable for extracting high-quality frequency-domain features; CNN excels at extracting spatial features across channels and is currently the most widely used; RNN is effective at capturing temporal features and is suitable for modeling long-term dependencies; Transformer models global temporal relationships through self-attention but requires large amounts of data for training. In practice, CNN is often combined with RNN or Transformer to leverage complementary advantages and achieve better performance. With the integration of lightweight network designs, current deep learning systems are expected to become more user-friendly and robust, which helps bring SSVEP-BCI technology into practical use, with typical scenarios covering healthcare and intelligent interaction. In summary, deep learning has improved SSVEP recognition by enhancing feature extraction, reducing dependence on manual signal processing, and supporting more flexible decoding frameworks. CNN-based models are effective for spatial-spectral feature extraction, RNN-based models are useful for temporal dependency modeling, and Transformer-based models show potential in personalized and calibration-efficient recognition. Future studies should further address limited training data, cross-subject generalization, model interpretability, and real-time deployment in practical BCI systems.

References

[1] Lin, Z., Zhang, C., Wu, W., & Gao, X. (2006). Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. IEEE Transactions on Biomedical Engineering, 53(12), 2610–2614. https://doi.org/10.1109/TBME.2006.886577

[2] Chen, X., Wang, Y., Gao, S., Jung, T.-P., & Gao, X. (2015). Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface. Journal of Neural Engineering, 12(4), 046008.

DOI: 10.54254/2755-2721/2026.AD34066

https: //doi.org/10.1088/1741-2560/12/4/046008

[3] Nakanishi, M., Wang, Y., Chen, X., Wang, Y.-T., Gao, X., & Jung, T.-P. (2018). Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. IEEE Transactions on Biomedical Engineering, 65(1), 104–112. https: //doi.org/10.1109/TBME.2017.2694818

[4] Liang, L., Zhang, Q., Zhou, J., Li, W., & Gao, X. (2023). Dataset evaluation method and application for performance testing of SSVEP-BCI decoding algorithm. Sensors, 23(14), Article 6310. https: //doi.org/10.3390/s23146310

[5] Wang, Z., Wong, C.-M., Wang, B., Feng, Z., Cong, F., & Wan, F. (2023). Compact artificial neural network based on task attention for individual SSVEP recognition with less calibration. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31, 2525–2534. https: //doi.org/10.1109/TNSRE.2023.3280821

[6] Zhang, S., Ma, K., Yin, Y., Ren, B., & Liu, M. (2022). A personalized compression method for steady-state visual evoked potential EEG signals. Information, 13(4), Article 186. https: //doi.org/10.3390/info13040186

[7] Yang, J., Zhao, S., Fu, Z., & Liu, X. (2024). PMF-CNN: Parallel multi-band fusion convolutional neural network for SSVEP-EEG decoding. Biomedical Physics & Engineering Express, 10(3), 035002. https: //doi.org/10.1088/2057-1976/ad2e36

[8] Shao, Y., Zheng, H., & Chen, X. (2025). A multibranch CNN framework based on time–frequency feature fusion for SSVEP detection. IEEE Sensors Journal, 25(20), 38416–38428. https: //doi.org/10.1109/JSEN.2025.3605622

[9] Gorur, K., Eraslan, H., & Kuruoglu, E. E. (2022). The single-channel dry electrode SSVEP-based biometric approach: Data augmentation techniques against overfitting for RNN-based deep models. Brain Sciences, 12(7), Article 943. https: //doi.org/10.3390/brainsci12070943

[10] Kim, H., Ahn, M., & Jun, S. C. (2025). A brain switch for SSVEP-based BCI speller using an RNN-based detection approach. In 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp.1–4). IEEE. https: //doi.org/10.1109/EMBC58623.2025.11252734

[11] Zhu, J., Jin, X., & Kong, W. (2025). ID-ProtoFormer: A dynamic identity prototype-infused transformer for SSVEP-based biometric recognition. IEEE Signal Processing Letters, 32, 4104–4108. https: //doi.org/10.1109/LSP.2025.3622525

[12] Liu, J., Wang, R., Yang, Y., Zong, Y., Leng, Y., Zheng, W., & Ge, S. (2024). Convolutional transformer-based cross subject model for SSVEP-based BCI classification. IEEE Journal of Biomedical and Health Informatics, 28(11), 6581–6593. https: //doi.org/10.1109/JBHI.2024.3454158