Sensors and Materials, Vol. 37, No. 11 (2025) 5177–5188
MYU Tokyo

S & M 4239

Multiscale Residual Attention and Multitask Learning for Steady-state Visual Evoked Potential Brain–Computer Interfaces

Yeou-Jiunn Chen,[1] Gwo-Jiun Horng,[2] Qian-Bei Hong,[1] and Kun-Yi Huang[2*]

[1]Department of Electrical Engineering, Southern Taiwan University of Science and Technology, No. 1, Nantai St., Yungkang Dist., Tainan City 710301, Taiwan
[2]Department of Computer Science and Information Technology, Southern Taiwan University of Science and Technology, No. 1, Nantai St., Yungkang Dist., Tainan City 710301, Taiwan

(Received October 13, 2025; accepted November 17, 2025)

Keywords: brain–computer interface, steady-state visual evoked potential, residual attention network, multitask learning, assistive technologies

Steady-state visual evoked potential (SSVEP)-based brain–computer interfaces (BCIs) offer a promising solution for facilitating communication in individuals with severe motor disabilities. Improving recognition accuracy through neural-network-based techniques is essential for developing more reliable and responsive SSVEP-based BCIs. In this study, HybridNet integrating a multiscale block (MSB), a residual attention network (ResAttNet), and a multitask learning (MTL) block is developed to improve the SSVEP-based BCI performance. In this study, we enhance the sensing performance of EEG-based BCIs by improving feature extraction and classification accuracy using EEG sensor data. MSB is used to capture diverse spectral patterns for robust feature extraction. ResAttNet is designed for adaptive channel-spatial feature enhancement. The MTL module further promotes shared feature learning across tasks to improve the classification accuracy. Experiments on two benchmark SSVEP datasets show that HybridNet achieves accuracies of up to 87.0 and 89.1% using 1 s EEG segments, outperforming other approaches. In future work, we will explore real-time and subject-independent implementations through integration with advanced EEG sensing technologies, enhancing signal robustness and practical applicability.

1. Introduction

Brain computer interfaces (BCIs) have emerged as promising assistive technologies that allow individuals with severe motor impairments to communicate or control external devices through neural activity, bypassing muscular pathways. This capability is particularly crucial for patients with amyotrophic lateral sclerosis or spinal cord injuries, enabling them to regain a degree of independence through wheelchair operation, robotic limb control, or communication systems.[1–4]

Compared with motor imagery-based BCIs, which suffer from extensive training demands and intersubject variability, and event-related potential-based BCIs, which are prone to user

*Corresponding author: e-mail: iamkyh77@stust.edu.tw
https://doi.org/10.18494/SAM5976

ISSN 0914-4935
https://myukk.org/

fatigue and temporal sparsity, steady-state visual evoked potential (SSVEP)-BCIs provide a more robust platform for real-time applications.[1] SSVEP-based BCIs are particularly attractive because of their high signal reproducibility, rapid response, high information transfer rate, and minimal training requirements.[2–4] These characteristics make SSVEP-based BCIs especially suitable for individuals with severe motor impairments, enabling them to interact with assistive communication systems and regain a degree of autonomy in daily life.

In SSVEP-based BCIs, the effectiveness of electroencephalogram (EEG) sensors as primary components for detecting neural responses to visual stimuli is paramount. The performance of these systems is strongly affected by the quality, stability, and spatial sensitivity of the EEG electrodes employed. By incorporating artificial intelligent mechanisms into EEG-based frameworks, the overall system performance can be substantially improved, particularly under conditions where sensor signals are noisy or suboptimal. Such integration enhances the robustness of signal interpretation and provides a degree of tolerance to fluctuations in sensor quality or electrode placement. Consequently, this approach contributes to the development of more resilient and dependable assistive technologies, ensuring reliable operation in real-world environments where external factors often degrade signal clarity.

Machine learning and deep learning techniques have been extensively investigated. Deep learning methods, including convolutional neural networks (CNNs), recurrent neural networks, and transformers, automatically extract hierarchical spatial–temporal features from raw EEG signals, thereby achieving superior accuracy.[5–10] Attention mechanisms further enhance decoding by selectively emphasizing informative channels and frequency bands while suppressing noise.[11,12] Integrating attention mechanisms into deep neural networks not only improves feature representation but also significantly boosts the performance and robustness of SSVEP-based BCIs.

Residual networks (ResNets) effectively address vanishing gradient problems in deep models and have been effectively applied to EEG decoding tasks.[13] Integrating ResNet with attention mechanisms improves robustness and classification performance by emphasizing critical spatial and spectral EEG components. This integration significantly improves the decoding accuracy and generalization capability of SSVEP-based BCIs, making them more reliable for assistive technologies.

Multi-domain features (MDFs), which integrate time, frequency, and spatial information, can now capture richer signal characteristics and improve generalization.[6–12] Hybrid CNN–RNN models with MDFs have demonstrated significant gains in physiological monitoring and movement intention recognition.[10] Another complementary approach is multiscale convolution, which utilizes filters of varying kernel sizes to capture both global and local features, thereby enhancing the representation of hierarchical and multifrequency patterns inherent in SSVEP signals.[12] The integration of spatial- and frequency-domain features enables more accurate EEG decoding, which in turn enhances the overall performance and reliability of SSVEP-based BCI systems.

We present HybridNet, a novel neural network architecture designed for SSVEP-based BCIs. HybridNet integrates a multiscale convolution block (MSB), a residual attention network (ResAttNet), and a multitask learning (MTL) block. To capture diverse temporal-spectral

patterns, MSB is used for robust feature extraction to find a meaningful representation of EEG signals. To find robust embedding features, ResAttNet is proposed to enhance generalization by selectively amplifying discriminative features. To improve classification performance, MTL is developed by leveraging shared representation across related tasks. The proposed framework is aimed at advancing SSVEP-based BCIs and developing practical and effective assistive technologies.

The novelty of this study lies in three key aspects: (1) the integration of multiscale convolution and residual attention to capture diverse temporal–spectral patterns in EEG signals, (2) the application of an MTL strategy to promote shared feature learning across multiple visual stimuli targets, and (3) the enhancement of EEG sensor data interpretation through channel-spatial attention mechanisms, improving robustness under short-time recordings. Unlike previous CNN-, RNN-, or Transformer-based approaches, HybridNet achieves higher accuracy with reduced model complexity, making it suitable for real-time sensor-driven BCI systems.

2. SSVEP-based BCI

The proposed architecture of HybridNet for SSVEP-based BCI is shown in Fig. 1. The system takes preprocessed EEG signals in both time and frequency domains, where frequency-domain representations are obtained via Fourier transform to complement temporal information. MSB is first applied to extract domain-dependent features, followed by a channel-combination block that integrates spatial information across EEG channels. Feature learning is further enhanced through a dilated convolutional block, which enlarges the receptive field, and ResAttNet that adaptively emphasizes salient components. The feature fusion module aggregates time- and frequency-domain features through element-wise summation, after which, a MTL block performs classification. To mitigate overfitting, dropout layers are inserted after each major block. The detailed design of each module is presented in the following subsections.

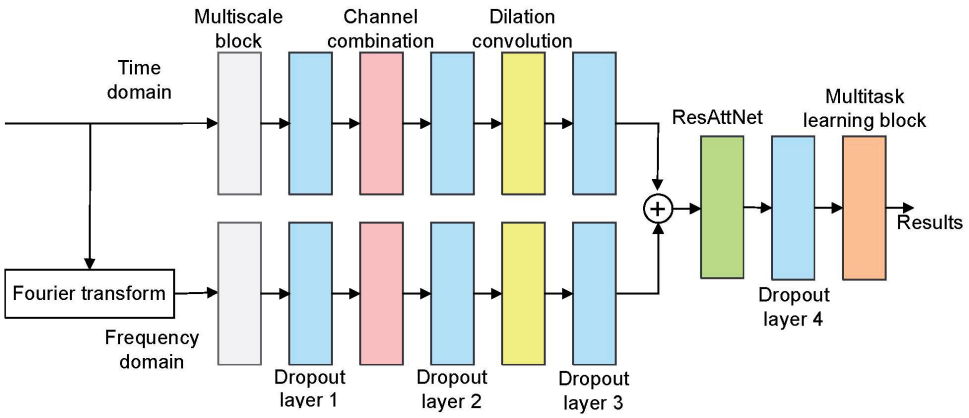

Fig. 1. (Color online) Proposed architecture of HybridNet for SSVEP-based BCI.

2.1 EEG signal acquisition and sensor configuration

In this study, a 64-channel SynAmps2 EEG acquisition system was employed, operating at a sampling rate of 1000 Hz and subsequently downsampled to 250 Hz to balance temporal resolution and computational efficiency. Nine electrodes positioned over the occipital and parietal regions (O1, Oz, O2, PO3, POz, PO4, Pz, PO5, and PO6) were selected owing to their high sensitivity to visual stimuli and their critical role in capturing SSVEPs. These electrodes form the core sensing interface responsible for detecting frequency-specific neural responses induced by visual flickers.

To maximize the utility of the acquired signals, advanced signal decoding and feature optimization techniques were incorporated into the HybridNet framework. This architecture enhances both the accuracy and robustness of the sensing-to-decoding pipeline, allowing for effective operation even under conditions of sensor noise or variability in electrode placement. The integration of sophisticated neural network methodologies with optimized EEG sensing configurations demonstrates how sensor-level signal acquisition and intelligent decision systems can jointly advance the performance and reliability of assistive technologies in real-world environments.

2.2 Multiscale convolutional block

The MSB illustrated in Fig. 2 is adopted to extract the embedding features of the EEG signals. Given an input, x, the block applies four parallel convolutions with kernel sizes k_1, k_2, k_3, and k_4. The ith convolution F_i is defined as

$$F_i(x) = ELU\left(BN\left(Conv_{1\times k_i}(x, w_i)\right)\right), \tag{1}$$

where $ELU(\cdot)$, $BN(\cdot)$, and $Conv_{1\times k_i}$ denote the exponential linear unit, batch normalization, and a convolution with kernel size $1\times k_i$, respectively. w_i represents the corresponding kernel weights. The output x_c values of the parallel convolutions are concatenated along the channel dimension to integrate multiscale contextual information:

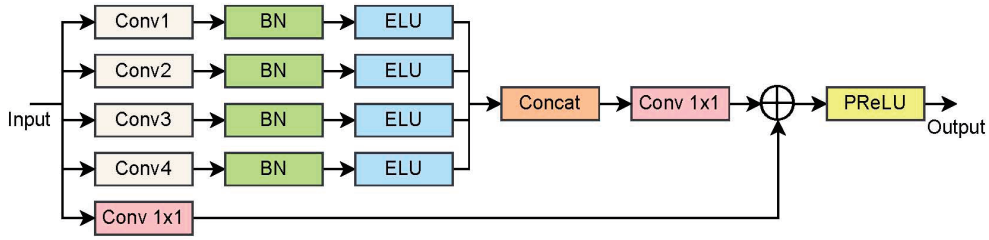

Fig. 2. (Color online) MSB with parallel convolutions of different kernel sizes and a 1×1 convolution for feature fusion.

$$x_c = Concat\left(F_1(x), F_2(x), F_3(x), F_4(x)\right). \tag{2}$$

To maintain a compact representation, a subsequent 1×1 convolution is applied to fuse the concatenated features for reducing the number of filters, thereby lowering computational complexity. Moreover, a residual connection is incorporated to stabilize training and alleviate the vanishing gradient problem. Therefore, an element-wise addition, ⊕, is adopted to fuse these embedding features. Finally, the output x_{ms} is activated by the parametric rectified linear unit (PReLU) as

$$x_{ms} = PReLU\left(Conv_{1\times1}(x_c, w_c) \oplus Conv_{1\times1}(x, w_r)\right), \tag{3}$$

where w_c and w_r are the kernel weights of a 1×1 convolution for x_c and the input x. Through this design, MSB enhances the network's ability to extract discriminative features across different temporal or spatial resolutions while preserving computational efficiency.

2.3 Residual attention network

The proposed ResAttNet illustrated in Fig. 3 is designed to enhance feature representation by jointly exploiting channel and spatial attention within a residual learning framework. Given an input embedding feature, a 1×1 convolution is first applied to reduce the channel dimensionality, producing the compressed intermediate feature. This intermediate feature is then processed by three parallel 1×k convolutions, followed by PReLU activation. Among the three branches, two are followed by distinct attention mechanisms: the channel attention module and the spatial attention module. Their outputs are combined through element-wise multiplication (denoted as ⊗) to capture interdependences between channel-wise and spatial representations. Then, this combined attention weighting is subsequently multiplied element-wise with the output of the third 1×k convolution branch, which does not incorporate attention, thereby refining the features maps and emphasizing informative representations. Finally, a residual connection integrates the refined feature with the input embedding feature through element-wise addition. Each convolution is followed by the PReLU activation function.

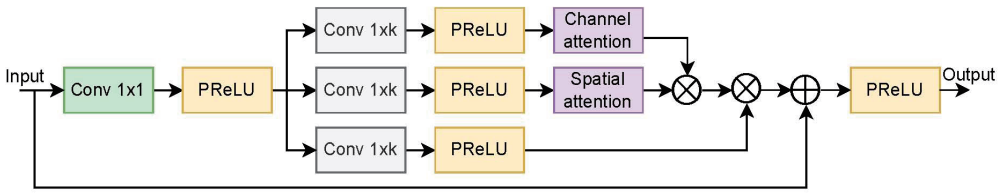

Fig. 3. (Color online) Architecture of the proposed ResAttNet, consisting of parallel convolutional branches with channel and spatial attention.

2.4 Multitask learning block

To enhance the discriminative capability of HybridNet, a MTL block is incorporated as the final convolutional module. The MTL block performs classification by treating each target recognition as an independent task.[4] Traditional MTL architectures typically allocate separate task-specific layers for each output branch, resulting in high computational redundancy and limited scalability. In this study, the MTL block adopts a group convolution strategy, allowing separate convolution operations to be performed within a single layer for each target. This design effectively enables parallel computation without duplicating all convolutional layers, leading to more efficient GPU utilization and reduced memory overhead.

For the input embedding feature, $x' \in R^{c \times l}$, where c is the number of channels and l is the length of the input embedding feature. To enable task-parallel processing without fully duplicating convolution, x' is replicated and concatenated m times along the channel dimension, where m is the number of targets. Therefore, a duplicated embedding feature, $x'' \in R^{(m \cdot c) \times l}$, can be obtained. A single grouped 1-d convolution is then applied to x'', so that each group performs an independent convolution on its corresponding channel slice. Finally, the sigmoid activation is performed to output the predicted likelihood of the corresponding targets.

3. Experimental Results and Discussion

In this study, two publicly available SSVEP datasets were employed: Datasets 1 and 2.[14,15] During data collection, participants were seated approximately 60 cm from a display presenting multiple flickering visual targets encoded with specific frequencies and phases. EEG signals were recorded using wet electrodes positioned in accordance with the international 10-20 system, primarily over the occipital and parietal regions, which are the most responsive to visual stimulation.

Dataset 1 consists of 40 visual targets encoded using joint frequency and phase modulation (JFPM) with stimulation frequencies ranging from 8 to 15.8 Hz (0.2 Hz interval) and phases spaced by 0.5π0.5\pi0.5π. EEG signals were recorded from 35 healthy participants (18 males, 17 females; mean age: 22 years) using a 64-channel SynAmps2 system at 1000 Hz and later downsampled to 250 Hz. Nine occipital and parietal channels (O1, Oz, O2, PO3, POz, PO4, Pz, PO5, and PO6) were selected.

Dataset 2 comprises 12 JFPM-coded targets with stimulation frequencies between 9.25 and 14.75 Hz (0.5 Hz interval). EEG data were acquired from 10 healthy subjects (9 males, 1 female; mean age: 28 years) using an 8-channel BioSemi ActiveTwo system at 2048 Hz and downsampled to 256 Hz. Eight occipital and parietal channels (O1, Oz, O2, PO3, POz, PO4, PO7, and PO8) were used.

The recorded signals were bandpass-filtered between 1 and 40 Hz, downsampled, and segmented into 1 to 4 s windows before being fed into the proposed HybridNet architecture. Model evaluation was conducted using a leave-one-subject-out cross-validation scheme, where the data of one subject served as the test set and the remaining data were used for training, with one-fifth of the training set further held back for validation. The network hyperparameters are

summarized in Table 1, where NS, KS, FS, DL, and R denote the number of samples, the kernel size, the number of filters, the dilation length, and the dropout rate, respectively, and T and F indicate time-domain and frequency-domain features, respectively. Model training employed the Adam optimizer for stable and efficient optimization across diverse data and architectures. Other training parameters were set as follows: learning rate = 0.001, epochs = 100, batch size = 64, and early stopping patience = 10.

3.1 Experimental results of frequency domain information

In this study, the contributions of temporal (T), frequency-domain (Fcomp), and combined (T+Fcomp) inputs for HybridNet are evaluated and the results are shown in Table 2. As the data length increases, classification performance consistently improves, demonstrating the positive correlation between signal duration and SSVEP response stability.

For Dataset 1, Fcomp achieved a higher accuracy than the time-domain input across all durations, indicating that frequency components provide more discriminative features for steady-state response identification. The combination T+Fcomp further enhanced performance, reaching 87.0% at 1 s. A similar trend is observed for Dataset 2, where Fcomp and T+Fcomp inputs outperform the pure temporal features, achieving an accuracy of 89.1% at 1 s.

These results confirm that incorporating spectral information effectively complements temporal patterns, leading to improved robustness and generalization. The integration of multi-domain features enables the proposed model to exploit both phase synchronization and temporal dynamics, enhancing decoding efficiency even under shorter observation windows.

Table 1
Network hyperparameter settings of HybridNet for Datasets 1 and 2.

Layer	Hyperparameter	Dataset 1	Dataset 2
Input layer	NS	250 (1 s), 500 (2 s), 750 (3 s), 1000 (4 s)	256 (1 s), 512 (2 s), 768 (3 s), 1024 (4 s)
MSB block	KS, FS, DL	T:25~55/F:23~53, 4, 1	T:23~53/F:23~53, 8, 1
Dropout layer 1,2,3,4	R	0.5	0.5
Convolutional block 2	KS, FS, DL	9, 4, 1	8, 2, 1
Convolutional block 3	KS, FS, DL	T:35/F:19, 8, 4	T:51/F:87, 4, 4
ResAttNet	KS, FS, DL	35, 8, 1	11, 8, 1
MTL block	KS	250 (1 s), 500 (2 s), 750 (3 s), 1000 (4 s)	256 (1 s), 512 (2 s), 768 (3 s), 1024 (4 s)
Output layer	NS	40	12

Table 2
Classification accuracy (%) of the proposed framework using time-, frequency-, and multi-domain inputs under different EEG signal durations.

Dataset	Input domain	Data length			
		1 s	2 s	3 s	4 s
Dataset 1	T	83.5 ± 16.8	88.4 ± 15.2	91.3 ± 14.0	93.5 ± 11.8
	Fcomp	86.2 ± 14.4	94.1 ± 10.0	95.5 ± 11.6	96.0 ± 8.9
	T+Fcomp	87.0 ± 13.8	94.3 ± 9.7	96.0 ± 8.9	96.4 ± 8.6
Dataset 2	T	90.4 ± 13.2	94.6 ± 7.4	95.0 ± 7.4	95.9 ± 5.7
	Fcomp	91.0 ± 11.5	94.3 ± 9.7	94.6 ± 11.0	95.8 ± 8.4
	T+Fcomp	89.1 ± 13.7	93.8 ± 10.0	95.1 ± 7.8	95.7 ± 7.6

3.2 Experimental results of attention mechanisms

To further investigate the contribution of the attention modules, ablation experiments were conducted by selectively removing the channel attention (CA) and spatial attention (SA) mechanisms from the HybridNet architecture. Specifically, ResAttNet-CA excludes the CA module, ResAttNet-SA excludes the SA module, and ResNet with both removed serves as a baseline. The results for both datasets under different signal durations are summarized in Table 3.

For the 1 s data length, the proposed HybridNet achieved the highest classification accuracies of 87.0 and 89.1% on Datasets 1 and 2, respectively. In contrast, removing the CA or SA module caused accuracy drops of approximately 2 to 3%, indicating that both mechanisms play essential yet complementary roles. Compared with the baseline ResNet, HybridNet consistently achieved superior performance, confirming that the joint integration of CA and SA enables the model to effectively capture critical embedding features. These results indicate that embedding both CA and SA simultaneously within the residual structure yields the strongest improvements, particularly for smaller data lengths. The consistent advantage across both datasets further demonstrates the robustness and generalizability of the proposed attention-enhanced framework.

3.3 Experimental results of ablation study

To further assess the contribution of individual components in the proposed HybridNet, an ablation study was conducted by selectively replacing the MSB, ResAttNet, and MTL block with their simplified counterparts. Specifically, MSB was substituted with conventional convolutional layers (1×55 and 1×53 for Dataset 1; 1×31 and 1×53 for Dataset 2), ResAttNet was replaced by standard convolutional layers (1×35 for Dataset 1 and 1×51 for Dataset 2), and the MTL block was replaced by a fully connected layer with 128 nodes. The experimental results are summarized in Table 4.

Experimental results show that replacing the MTL layer with a fully connected layer led to the most pronounced degradation at smaller data lengths, with accuracy drops of up to 22.0% on Dataset 1 and 9.4% on Dataset 2 for 1 s trials. This indicates that MTL plays a crucial role in extracting discriminative features when limited temporal information is available. In contrast,

Table 3
Classification accuracy (%) of the proposed framework without CA or SA.

Dataset	Model	Data length			
		1 s	2 s	3 s	4 s
Dataset 1	HybridNet	87.0 ± 13.8	94.3 ± 9.7	96.0 ± 8.9	96.4 ± 8.6
	ResAttNet-CA	84.8 ± 15.9	91.8 ± 13.1	92.7 ± 13.6	93.3 ± 13.4
	ResAttNet-SA	84.3 ± 15.2	91.4 ± 14.3	92.6 ± 13.5	93.3 ± 13.3
	ResNet	84.6 ± 15.6	91.5 ± 12.3	92.0 ± 15.7	94.0 ± 13.2
Dataset 2	HybridNet	89.1 ± 13.7	93.8 ± 10.1	95.1 ± 7.8	95.7 ± 7.6
	ResAttNet-CA	88.3 ± 14.2	93.2 ± 9.6	93.7 ± 10.5	94.3 ± 9.8
	ResAttNet-SA	88.7 ± 13.1	93.0 ± 9.4	93.1 ± 11.4	94.3 ± 9.9
	ResNet	86.8 ± 16.6	93.0 ± 10.6	94.7 ± 8.7	94.8 ± 8.5

Table 4
Ablation study results of HybridNet on Datasets 1 and 2 in terms of average accuracy (%).

Testing dataset	Model	Data length			
		1 s	2 s	3 s	4 s
Dataset 1	HybridNet	87.0	94.3	96.0	96.4
	- MSB	85.5	93.0	93.7	94.1
	- ResAttNet	83.1	91.8	92.7	93.0
	- MTL	65.0	84.3	90.4	93.1
Dataset 2	HybridNet	89.1	93.8	95.1	95.7
	- MSB	88.2	92.5	93.6	94.2
	- ResAttNet	86.3	91.7	92.6	93.0
	- MTL	79.7	90.3	92.9	94.6

replacing MSB with standard convolutional layers resulted in relatively smaller drops for small data lengths (≤2 s), but larger performance decreases at longer durations (3 to 4 s). This suggests that MSB is particularly effective when more complete temporal–frequency structures are present, leveraging multiscale representations for improved performance. Finally, removing ResAttNet yielded consistent accuracy reductions across both datasets, with the largest effect observed at small data lengths (up to 3.9 and 2.8% for Datasets 1 and 2, respectively). This highlights the importance of attention mechanisms in enhancing the network's ability to emphasize informative channels and spatial regions. Therefore, the combined effect provides a significant improvement in overall accuracy, validating the design choices of the proposed architecture.

3.4 Experimental results compared with those of other approaches

To further evaluate the effectiveness of HybridNet, its performance was compared with those of several other approaches including EEGNet,[6] GDNet-EEG,[12] Compact CNN,[6] SSVEPformer,[9] FB-SSVEPformer,[9] FBCNN,[8] and MTL.[4] The results on Datasets 1 and 2 are summarized in Table 5.

HybridNet consistently outperformed conventional deep learning and transformer-based models across both datasets. In the time domain, HybridNet achieved accuracies of 83.5 and 90.4% on Datasets 1 and 2, respectively. It exceeds the accuracies of EEGNet and Compact CNN by more than 10% on average. Furthermore, HybridNet further improved its performance to 86.2 and 91.0% by utilizing Fcomp, demonstrating that spectral information provides more discriminative features than time-domain signals alone.

The best performance was obtained with the multi-domain features (T+Fcomp), achieving 87.0% on Dataset 1 and 89.1% on Dataset 2. This demonstrates that integrating both temporal and spectral features allows the model to capture complementary patterns of EEG signals, thereby improving classification robustness. Compared with transformer-based methods such as SSVEPformer and FB-SSVEPformer, HybridNet maintains competitive or superior accuracy with a more compact and computationally efficient design, making it well suited to real-time BCI applications.

Table 5
Accuracies (%) of proposed HybridNet and other approaches on Datasets 1 and 2.

Method	Input	Dataset 1		Dataset 2	
		No. of channels	Accuracy	No. of channels	Accuracy
EEGNet	Time	9	72.00	8	80.00
GDNet-EEG			75.28		—
Compact CNN			—		80.00
SSVEPformer	Fcomp	9	80.40	8	84.16
FB-SSVEPformer			83.19		88.37
FBCNN			73.52		83.46
HybridNet	Time	9	83.50	8	90.40
	Fcomp		86.20		91.00
	T+Fcomp		87.00		89.10
MTL	Time	11	84.00	—	—
HybridNet	Time	11	85.60	—	—

Compared with previous SSVEP decoding networks such as EEGNet, GDNet-EEG, and SSVEPformer, the proposed HybridNet achieves superior decoding accuracy and computational efficiency owing to its novel hybrid architecture. The multiscale and attention-based design enables adaptive feature extraction directly from sensor-level EEG data, enhancing the system's ability to identify subtle spatial–spectral correlations even under variable sensing conditions. Furthermore, the integration of the MTL block promotes task-specific generalization by sharing discriminative representations across multiple visual stimuli. This joint optimization of sensing and decoding processes represents a significant advancement over existing methods and will contribute to more reliable and efficient EEG sensor-based BCI systems.

5. Conclusions

In this study, HybridNet, a novel neural network architecture that integrates MSB, ResAttNet, and a MTL module, was proposed for the development of SSVEP-based BCIs. MSB effectively captured the diverse spectral patterns and generated robust embedding features for EEG signals. ResAttNet enhanced feature representation by adaptively emphasizing salient channel-spatial components. The MTL module significantly improved classification performance by promoting shared feature learning across tasks. Experimental results demonstrated that HybridNet outperforms existing approaches including CNN, Transformer, EEGNet, SSVEPformer, and FBCNN in terms of both accuracy and computational efficiency. Furthermore, the results of ablation studies confirmed that each component contributed complementary benefits to the overall performance. The proposed HybridNet enhanced the decoding capability of EEG sensor data by improving spatial and spectral feature discrimination. This contribution advances the application of EEG sensors in BCIs, making them more effective for assistive communication and control systems. In future work, we will extend HybridNet to real-time and subject-independent applications by integrating it with advanced EEG sensing systems. The aim of this integration is to enhance the signal stability, adaptive calibration, and material efficiency of EEG sensors, thereby strengthening the overall sensing-decoding synergy and promoting reliable assistive applications in practical environments.

Acknowledgments

The authors would like to thank the National Science and Technology Council, Taiwan for the financial support (NSTC 114-2221-E-218-018) and the Higher Education Sprout Project of the Ministry of Education, Taiwan.

References

1 X. Wang, V. Liesaputra, Z. Liu, Y. Wang, and Z. Huang: Artif. Intell. Med. **147** (2024) 102738. https://doi.org/10.1016/j.artmed.2023.102738
2 R. Ajmeria, M. Mondal, R. Banerjee, T. Halder, P. K. Deb, D. Mishra, P. Nayak, S. Misra, S. K. Pal, and D. Chakravarty: IEEE Commun. Surv. Tutorials **25** (2023) 184. https://doi.org/10.1109/COMST.2022.3232576
3 W. Ding, A. Liu, L. Guan, and X. Chen: IEEE Trans. Neural Syst. Rehabil. Eng. **32** (2024) 875. https://doi.org/10.1109/TNSRE.2024.3366930
4 H. J. Khok, V. T. C. Koh, and C. Guan: Proc. 2020 IEEE Int. Conf. Syst., Man, Cybern. (SMC) (IEEE, 2020) 1280–1285. https://doi.org/10.1109/SMC42975.2020.9283310
5 X. Li, S. Yang, N. Fei, J. Wang, W. Huang, and Y. Hu: Bioeng. **11** (2024) 613. https://doi.org/10.3390/bioengineering11060613
6 N. Waytowich, V. J. Lawhern, J. O. Garcia, J. Cummings, J. Faller, P. Sajda, and J. M. Vettel: J. Neural Eng. **15** (2018) 066031. https://doi.org/10.1088/1741-2552/aae5d8
7 J. Xie: Appl. Comput. Eng. **103** (2024) 9. https://doi.org/10.54254/2755-2721/103/20241117
8 D. Zhao, T. Wang, Y. Tian, and X. Jiang: IEEE Access **9** (2021) 147129. https://doi.org/10.1109/ACCESS.2021.3124238
9 J. Chen, Y. Zhang, Y. Pan, P. Xu, and C. Guan: Neural Netw. **164** (2023) 521. https://doi.org/10.1016/j.neunet.2023.04.045
10 H. Yao, K. Liu, X. Deng, X. Tang, and H. Yu: J. Neurosci. Methods **379** (2022) 109674. https://doi.org/10.1016/j.jneumeth.2022.109674
11 L. Gong, M. Li, T. Zhang, and W. Chen: Biomed. Signal Process. Control **84** (2023) 104835. https://doi.org/10.1016/j.bspc.2023.104835
12 Z. Wan, W. Cheng, M. Li, R. Zhu, and W. Duan: Front. Neurosci. **17** (2023) 586. https://doi.org/10.3389/fnins.2023.1160040
13 K. He, X. Zhang, S. Ren, and J. Sun: Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (IEEE, Las Vegas, NV, USA, 2016) 770–778. https://doi.org/10.1109/CVPR.2016.90
14 Y. Wang, X. Chen, X. Gao, and S. Gao: IEEE Trans. Neural Syst. Rehabil. Eng. **25** (2017) 1746. https://doi.org/10.1109/TNSRE.2016.2627556
15 M. Nakanishi, Y. Wang, Y. T. Wang, and T. P. Jung: PLoS ONE **10** (2015) e0140703. https://doi.org/10.1371/journal.pone.0140703

About the Authors

Yeou-Jiunn Chen received his Ph.D. degree from the Institute of Information Engineering, National Cheng Kung University, Tainan, Taiwan, in 2000, respectively. He is currently a distinguished professor at the Department of Electrical Engineering, Southern Taiwan University of Science and Technology, Tainan, Taiwan. His research interests include biomedical signal processing, natural language processing, artificial intelligence, and deep learning. (chenyj@stust.edu.tw)

Gwo-Jiun Horng received his Ph.D. degree (2013) in computer science and information engineering from National Cheng Kung University, Taiwan. He is a full distinguished professor in the Department of Computer Science and Information Engineering, Southern Taiwan University of Science and Technology, Tainan, Taiwan. His research interests include mobile services, AIoT, intelligent computing, and cloud networks. (grojium@stust.edu.tw)

Qian-Bei Hong received his Ph.D. degree from the Graduate Program of Multimedia Systems and Intelligent Computing, National Cheng Kung University and Academia Sinica, Tainan, Taiwan, in 2023. He is currently an assistant professor in the Department of Electrical Engineering at Southern Taiwan University of Science and Technology. His research interests include deep learning, voice spoofing detection, speaker recognition, automatic speech recognition, and speech synthesis. (qianbeihong@stust.edu.tw)

Kun-Yi Huang received his Ph.D. degree from the Institute of Computer Science and Information Engineering, National Cheng Kung University (NCKU), Tainan, Taiwan, in 2019. He is currently an assistant professor in the Department of Computer Science and Information Engineering of Southern Taiwan University of Science and Technology. His research interests include artificial intelligence and deep learning, speech and speaker recognition, multimedia emotion recognition, biomedical signal processing and spoken language processing. (iamkyh77@stust.edu.tw)