ISSN 2616-7433 Vol. 7, Issue 4: 58-66, DOI: 10.25236/FSST.2025.070408

RSVP Target Detection Based on MCAMEEGNet

Qianqian Yang[1], Wei Wang[2,*]

[1]School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai, 200093, China
[2]Navy Medical Center, Navy Military Medical University, 200093, Shanghai, China
**Corresponding author*

Abstract: *Target detection plays a critical role in daily applications. While traditional computer vision techniques have demonstrated certain effectiveness in remote sensing image target detection, their performance is constrained by dependence on prior knowledge and limitations in target feature representation. Notably, their detection performance deteriorates significantly under factors such as varying illumination, target occlusion, and limited sample sizes. In contrast, the human brain exhibits remarkable capabilities for target identification in complex environments.This study investigates an electroencephalogram(EEG)-based rapid serial visual presentation(RSVP) method for target detection in remote sensing imagery. By integrating brain-computer interface(BCI) technology, the proposed approach achieves efficient and accurate target detection. Furthermore, we present a novel neural network architecture(MCAMEEGNet)—which combines multi-scale convolution networks, channel attention mechanisms, and a compact Electroencephalography network(EEGNet). Experimental results demonstrate that this model achieves significant improvements in event-related potential(ERP)-based RSVP target recognition, attaining an accuracy of 94.77% on the test set. The research establishes a novel technical framework for EEG-based remote sensing image target detection, offering both theoretical insights and practical applications.*

Keywords: *Rapid Serial Visual Presentation, Event-Related Potentials, Brain-Computer Interface, Convolutional Neural Network*

1. Introduction

Computer vision has been effective in the field of target detection, but it is still ineffective under light intensity, target occlusion, small sample size and real-time requirements. In contrast, the human brain, with its memory, attention and abstraction abilities, is more cognitively aware of unstructured and other complex information and can understand the high-level semantics of images. For humans, social experience constitutes a priori knowledge, and the human brain can directly perform tasks without the support of large training datasets. The human brain only needs a small number of short-term samples to show better performance in the field of target detection. And the human brain's ability to transfer knowledge enables it to quickly adapt to complex environments and overcome sensitive information[1]. Therefore combining the capabilities of the human brain with computers to achieve better results is currently a hot topic in the field of target detection research.

ERP signal is the electrophysiological response of the brain to external stimuli, and two important properties of ERP lie in latency stability and waveform stability[2]. Among the various components of the ERP signals, the P300 potential is closely related to the brain's decision-making process, reflecting the brain's evaluation and classification process of external stimuli.The RSVP paradigm consists of a series of images presented sequentially to the subject at a fast speed, including target and non-target images, and usually the presentation time of each image is only a few tens to a few hundreds of milliseconds, so that the subject can efficiently recognize a large number of images in a short time[3]. When a subject notices a target image, the brain generates ERP signals for target detection.

RSVP based on ERP signals has been used in the field of target detection and many applications benefit from the Benefits of this model. Examples include area surveillance, face recognition and clinical diagnosis[4]. In the field of area surveillance, RSVP target detection has been successfully applied in applications such as birds'-eye view and side-scan sonar images[5]. Barngrover et al. research focuses on the accurate identification of mine-like objects in the undersea environment through sonar imaging[6]. Ana Matran-Fernandez et al. apply the N200 composition for the detection of targets that may appear at any location in aerial images[7]. In the field of face recognition, Touryan

ISSN 2616-7433 Vol. 7, Issue 4: 58-66, DOI: 10.25236/FSST.2025.070408

et al. explored the effect of face familiarity on neural response using ERP signals. The experiment required subjects to recognize three facial images with different degrees of familiarity, and the results showed that the amplitude change of the ERP component of the brain was significantly positively correlated with the degree of face familiarity[8]. WANG et al. designed a single-trial face detection neural decoding framework based on RSVP, which innovatively constructed a Convolutional Neural Network (CNN) to achieve end-to-end EEG classification by analysing the ERP spatio-temporal features generated when subjects viewed the target face and non-target face stimuli. Experiments show that compared with the common Support Vector Machine (SVM), the deep learning model demonstrates a significant advantage in the classification accuracy index[9]. Shenoy et al. successfully detected facial images among numerous pictures by utilizing independent spatial patterns of single-trial N170 components[10]. RSVP target recognition also has great potential for medical applications. It can help clinical diagnosis and disease prevention by quickly identifying targets in images, such as cancer cell identification, epilepsy early warning, etc. Hope et al. used mammography images in order to develop more effective breast cancer screening methods[11]. Future RSVP studies will benefit from iterative changes in design parameters. This will allow comparative studies to be conducted on different design parameters and enable the identification of the parameters that most affect the experimental paradigm. This is a rapidly growing area with a promising future[4].

In recent years CNN play an important role in EEG signal processing and classification. Cecotti used CNN to detect P300 signals and achieved good detection performance, which shows that CNN can efficiently and accurately extract key spatio-temporal information from ERP signals[12].Pedram et al. 2022 developed the TGT-MHOG-CNN framework for detecting P300 components, which employs a triangular function-based tuned Gabor transform(GT) to transform the raw EEG signal into a spectrogram with time-frequency features. On this basis, an improved directional gradient histogram algorithm is introduced for feature extraction, and finally the signals fused with spatio-temporal features are identified by deep convolutional neural networks[13], and the organic combination of traditional image processing features and deep learning models effectively improves the accuracy of ERP signal analysis. In order to cope with the challenge of the distribution variability of ERP signals in BCI-RSVP in different datasets, Zhao proposed a multilayer frequency spatio-temporal feature classification framework (STAEE) in 2024[14], which is composed of four core modules, such as the filter bank module, time window decomposition module, spatial-temporal domain filtering module, and region-of-interest selection module, and effectively improves the system's classification performance. Aiming at the technical difficulties of strong coupling of spatio-temporal features and significant individual differences of P300 EEG signals, Yuan et al. proposed PSAEEGNet in 2024, which is composed of three module: the standard convolutional layer, the Pyramid Squeeze Attention (PSA) module, and the depthwise convolutional layer[15], which effectively extracts spatio-temporal features of P300 and significantly improves the classification performance. In 2025, Chen et al. proposed a new framework, Target Stack Convolutional Self-Encoder (CS-IRE-TSCAE)[16], based on reconstructing the invariant representation in order to overcome the difficulties in the cross-subject scenario in the dual-target BCI-RSVP. It effectively solves the problem of data distribution differences in cross-subject scenarios in the dual-target BCI-RSVP task and mitigates the subject dependency effect. And the model was validated on the ERP dataset of the 2022 BCI Controlled Robotics Competition.

In this paper, a convolutional deep learning network is used to classify ERP signals, improving on the structure of EEGNet, a compact convolutional neural network for EEG proposed by Lawhern in 2017, which uses a combination of deep convolution and separable convolution for more efficient ERP feature extraction[17]. The model plays an important role in EEG signal processing and classification. In this paper, the original single convolutional model is improved to multi-scale convolution networks, which captures features at different scales to make the feature representation richer and more comprehensive. And the attention module is added to enhance the feature selection ability of the model, which enhances important features and suppresses irrelevant features by adaptively adjusting the channel weights.

2. Experimental Procedure

2.1 Participants

Ten students were selected as experimental subjects(male / female = 5 / 5, mean age of 23.5±6.7 years). Prior to the experiment, it was ensured that all subjects met the following criteria: normal or corrected-to-normal visual function, no history of neurological disease or psychiatric disorder, and no

ISSN 2616-7433 Vol. 7, Issue 4: 58-66, DOI: 10.25236/FSST.2025.070408

continuous use of any medication for three months. Prior to the experiment, participants were required to get enough sleep for two consecutive nights of at least seven hours, to refrain from consuming caffeinated or alcoholic beverages on the day of the test, and to avoid high-intensity exercise for two hours prior to the start of the experiment. Subjects were required to complete an informed consent form.

The experiment was conducted in a standardized, soundproof and noise-reduced laboratory, with subjects sitting in a comfortable chair in a standard posture, with their eyes looking at a monitor 60 cm in front of them. The experimental procedure consisted of two phases: firstly, a practice section was conducted to help the subjects Familiarize with the operation interface and response requirements. Afterwards, a formal experiment was conducted, in which all the instructions were shown to the subjects through the display. Quietness and environmental stability were maintained throughout the experiment to ensure reliable data collection.

2.2 Experiment Setup

The original remote sensing image was segmented into 8*8 subgraphs as the experimental material, some of which contained randomly positioned target aircraft. A total of 8 groups were set up for this experiment, and each group contained 80 stimulus units. Each stimulus unit consists of 8 consecutive rapidly presented remote sensing images, one of which may contain an aircraft target (if it exists, the aircraft randomly appears within one of the images), and the target location is randomly distributed in the images. The experiment used a staged rate design: the first four groups (Group1-Group4) presented the images at 400 ms to provide subjects with a more adequate observation time; the last four groups (Group5-Group8) switched to a high-speed mode of 200 ms to significantly compress the visual processing window. Immediately after each stimulus unit, subjects were asked to provide feedback on whether the target was detected or not by pressing a key ("1" means detected, "0" means not detected). After the completion of each set of tasks, subjects were allowed to take a short rest with their eyes closed to Significantly Compress the Visual Processing Window. The target detection efficacy at both presentation rates was good. Practice is required before the formal experiment to ensure that subjects are familiar with the experimental rules. Figure 1 shows the experimental stimulus presentation protocol.

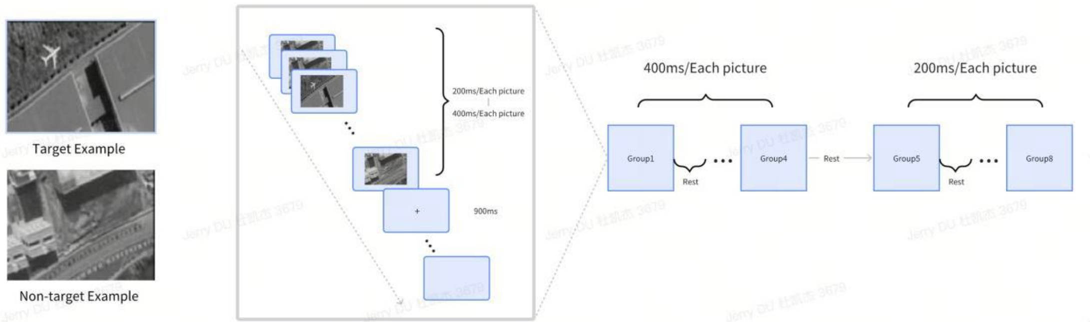

Figure 1: Schematic diagram of RSVP paradigm.

2.3 Data Recordings

The RSVP experimental procedure used a 64-channel wireless EEG acquisition system (NeuSen W 64, Neuracle) to record the participants' EEG activity. Its 64 electrode positions follow the 10-20 international system[18]. It is a standardized EEG electrode placement system where electrodes are placed in specific positions according to the shape and size of the head to enable systematic recording of electrical activity in different areas of the brain.The 64 electrodes are Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC3, FC1, FCz, FC2, FC4, FC6, C5, C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CPz, CP2, CP4, CP6, P7, P3, Pz, P4, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, T7, T8, TP7, TP8, FT7, FT8, AF7, AF3, AFz, AF4, AF8, F1, F2, F5, F6, FC1, FC2, C1, C2, A1, A2, GND, REF. All electrodes' impedance were kept below 10 kΩ during EEG signal acquisition. The sampling rate was 1000 Hz.

2.4 Data Preprocessing and Epoch Extraction

After recording the EEG signals of the subjects, the EEG data need to be preprocessed first, firstly the electrodes will be positioned, and the M1 and M2 bilateral mastoids will be selected as the

ISSN 2616-7433 Vol. 7, Issue 4: 58-66, DOI: 10.25236/FSST.2025.070408

reference electrodes. And remove the useless electrodes (ECG, HEOR, HEOL, VEOU, VEOL), after removing the useless electrodes, the raw data were band-pass filtered using Butterworth filter (0.1-45 Hz). The EEG signals were then downsampled to 256 Hz to reduce the burden of subsequent calculations by lowering the sampling rate. Since the different components of the original EEG signal are linearly superimposed, Independent Component Analysis (ICA) can separate the signal into different EEG components, which are then identified in order to facilitate the removal of artifacts from the original signal, such as eye movements, muscle movements, heart movements, etc. and referenced to the mean value. (EEG data segments starting 200 ms before image presentation and ending 800 ms after target presentation, and use the data from 200 ms before the stimulus to calculate the baseline in order to eliminate the influence of poor electrode contact on the signal, retaining only those segments that were judged correctly by the subject). The sample size of the processed data was 59*256. Due to the imbalance between the target and non-target sample sizes, the number of non-target samples was scaled down to match the target samples to improve the accuracy and reduce the training intensity. all the preprocessing procedures of the EEG signals were carried out using the EEGLAB toolbox in MATLAB R2020a (The MathWorks Inc, USA)[20]. The EEG signals preprocessed as described above remove a large amount of noise and obtain relatively clean data conducive to improving the subsequent classification training.The flow of EEG data processing is shown in Figure 2.

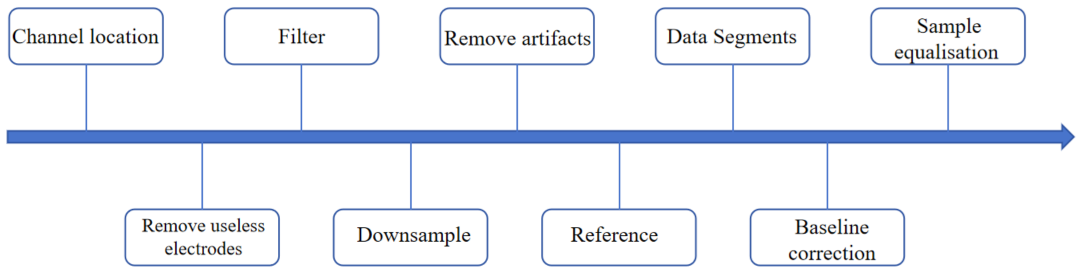

Figure 2: Diagram of the data processing process.

2.5 Network Model

EEGNet is a compact CNN architecture that can be applied to motion imagery classification tasks and ERP and steady-state visual evoked potentials.The advantage of EEGNet is that it can be trained with a limited number of datasets and can produce separable features. In addition, the EEGNet model has good generalization ability. Based on these advantages, in this paper, an improved EEGNet——MCAMEEGNet to detect ERP signals based on the RSVP paradigm target detection task. Figure 3 shows the overall visualization of the MCAMEEGNet structure.

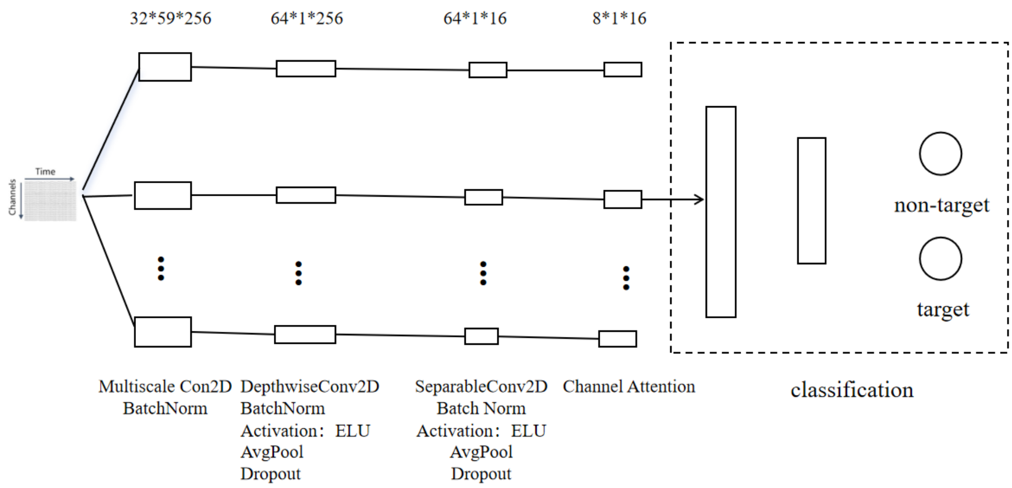

Figure 3: Overall visualization of the MCAMEEGNet structure.

The original image (C, T), C is the number of channels and T is the sampling point. Firstly, the data end is upscaled (1, C, T) as the input of the network, firstly, different time window features are captured by multi-scale convolution kernel, the size of the convolution kernel is (1, 64), (1, 32), (1, 16),

ISSN 2616-7433 Vol. 7, Issue 4: 58-66, DOI: 10.25236/FSST.2025.070408

after multi-scale convolution networks, the corresponding feature maps are respectively outputted, as the convolution is easy to lose the original signals, so it fills in with the same mode, and after convolution, the three different kinds of convolutional The superposition of the features can capture features at different scales, and this structure effectively improves the model's ability to perceive the EEG signal features by concurrently connecting different sized convolution kernels, making the features more rich and comprehensive. The feature map of the output by multi-scale fusion is (F1, C, T), and the mathematical expression of this process is as follows:

$$\mathrm{F}_{11} = \mathrm{ReLU}(\mathrm{BN}(\mathrm{W}_{1*64} * X)) \tag{1}$$

$$\mathrm{F}_{12} = \mathrm{ReLU}(\mathrm{BN}(\mathrm{W}_{1*32} * X)) \tag{2}$$

$$\mathrm{F}_{13} = \mathrm{ReLU}(\mathrm{BN}(\mathrm{W}_{1*16} * X)) \tag{3}$$

$$\mathrm{F}_{1} = \mathrm{X} + \mathrm{BN}(\mathrm{W}_{\mathrm{compress}} * concat([F_{11}, F_{12}, F_{13}])) \tag{4}$$

The lightweight convolutional architecture is subsequently implemented as follows: First, spatial feature extraction across individual channels is achieved through DepthwiseConv2D, where depthwise convolution operates on a single channel at a time. Distinct from traditional convolution, separate kernels are applied to different channels, enabling the model to independently learn features for each channel. A depth multiplier of 2 is configured to control the number of output channels generated, acting on each output channel. A kernel size of (C, 1) is adopted to effectively capture complex inter-channel relationships, enhancing decoding accuracy while substantially reducing the number of trainable parameters. Dropout regularization is incorporated with a probability of 0.5 to Prevent overfitting during sample training. Finally, dimensionality reduction is performed via average pooling with a kernel size of (1, 4).

The third layer is a separable convolutional layer that combines depth convolution with point-by-point convolution(1, 1) using SeparableConv2D. Firstly, spatial convolution in the depth direction is performed and then the resulting channels are doped together for point-by-point convolution. It is just deepwiseconv2d and point-by-point convolution. The size of the input feature map of the network is (D×F1)×1×(T/4) and the size of the output feature map needs to be (F2, 1, T/4)[19].

Depthwise separable convolution aids the model in extracting more representative features through the combination of various attributes, enabling it to learn rich features with fewer parameters and reducing computational complexity. Subsequently, average pooling with a kernel size of (1, 8) is applied for dimensionality reduction[17].

Subsequently the attention mechanism is carried out and the mechanism employs dynamically learn channel weights to enhance feature representation based on the channel attention module in the CBAM model. Two spatial compression methods, Global Average Pooling (GAP) and Global Maximum Pooling (GMP), are also used to form complementary channel description information. A two-layer MLP with shared weights is used to process two-way features. The two-way results are summed element-by-element and then passed through the sigmoid function to generate the channel attention weight M_c, which indicates the importance of each channel. The whole process can be expressed by the following equation:

$$\mathrm{M}_{\mathrm{C}} = \sigma(\mathrm{W1} \bullet \delta(W0 \bullet AvgPool(X) + W1 \bullet \delta(W0 \bullet MaxPool(X))) \tag{5}$$

$$\mathrm{X}' = \mathrm{Mc} \otimes \mathrm{X} \tag{6}$$

Following each convolutional layer, EEGNet employs Batch Normalization to regularize the input data, stabilizing the distribution of input data during the deep learning optimization process. After batch normalization, the ELU (Exponential Linear Unit) activation function is applied to add more nonlinear elements to the network model, thereby enhancing the network's ability to fit complex nonlinear relationships.

This is followed by a fully connected layer and finally in softmax classification, softmax classifies the transmitted features in N units of neurons, where N is 2, to accurately classify the data into categories.Figure 4 shows ERP signals for targets trials and non-targets trials.The softmax formula is given as follows:

ISSN 2616-7433 Vol. 7, Issue 4: 58-66, DOI: 10.25236/FSST.2025.070408

$$\mathrm{p}(i) = \frac{\exp(\theta_i{}^T x)}{\sum_k^K \exp(\theta_k^T x)} \tag{7}$$

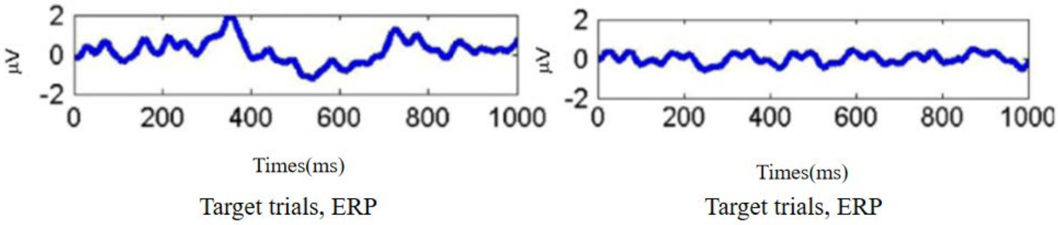

Figure 4: ERP signals for targets trials and non-targets trials.

3. Result

3.1 Classification Performance Analysis

In this study, we adopt the MCAMEEGNet model to classify the ERP signals based on the RSVP target recognition task, and judge the stability of the model by the training process and results of MCAMEEGNet on the dataset. Because of the individual variability of EEG signals among different subjects, this paper focuses on the training of the model and the analysis of the results for each subject, in which a total of about 4,000 samples were collected and processed for each subject for the target detection task under the RSVP paradigm, and in which 80% of the samples were randomly selected as the training set, 10% as the validation set, and 10% as the test set. The MCAMEEGNet model was trained using the Adam optimiser, cross-entropy loss was used for gradient updating and the model was trained for 50 epochs.

The training and validation process evaluates the performance of the model through changes in the accuracy (acc) and loss function, which provide a combined measure of the model's performance in terms of accuracy and completeness. the formula for acc is as follows:

$$\mathrm{acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{8}$$

The formula TP denotes the instances where the model correctly predicts the positive category as positive. That is, when the actual category of an instance is a positive category and the model predicts it to be a positive category. FN denotes an instance where the model incorrectly predicts a positive category as a negative category. That is, when the actual category of an instance is positive, but the model predicts it to be negative. FP indicates that the model incorrectly predicts the negative category to be the positive category real column. That is, when an instance's actual category is negative, but the model predicts it to be positive. TN indicates that the model correctly predicts the negative category as an instance of the negative category. That is, when an instance's actual category is a negative category and the model predicts it to be a negative category[21].

For each subject, we trained an independent model using their EEG data and evaluated its classification performance on the test set for that subject only. Figure 5 shows the variation of the loss function for different number of iterations for the subjects. The loss rate tends to stabilize after 50 training rounds. And the loss rate stays low throughout the training process, showing better convergence.

ISSN 2616-7433 Vol. 7, Issue 4: 58-66, DOI: 10.25236/FSST.2025.070408

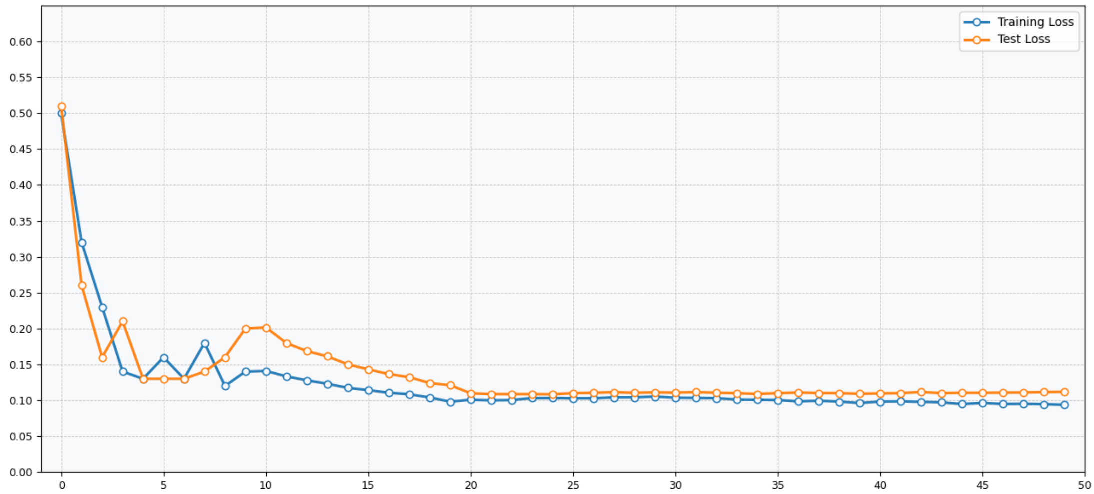

Figure 5: Training and validation loss variation of MCAMEEGNet.

In addition, this paper compares the MCAMEEGNet model with current deep learning models applied to EEG signals on a test set of 10 subjects. The results indicate that the MCAMEEGNet outperforms other deep learning algorithms in ERP signal detection, and the average accuracy of the subjects can be as high as 94.77%. The accuracy of all 10 subjects on the test set is shown in Figure 6.

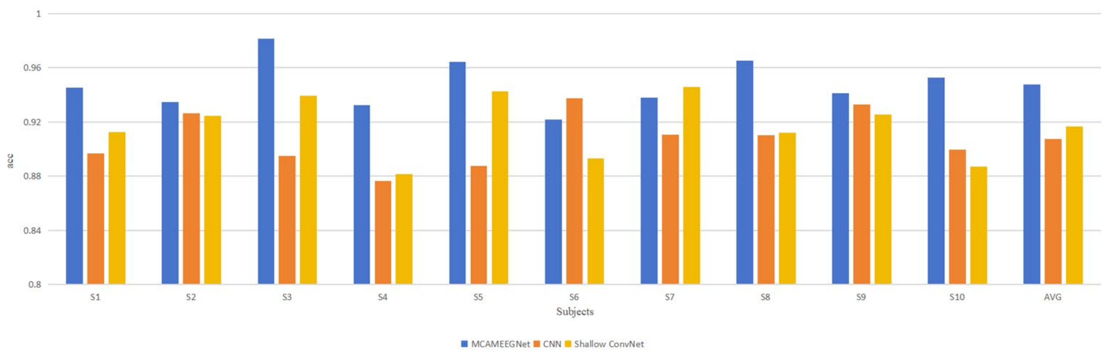

Figure 6: Comparison of Accuracy for Different Algorithms.

3.2 Cross-Subject Performance

All of the above results are for single subject, to improve the generalization of the algorithms or models across subjects, we also conducted a Subject-independent study which used leave-one-subject cross-validation. The classification results are shown in Table 1, which indicates that the accuracy of Subject-dependent models is greater than that of the Subject-independent models. Moreover, the classification results of MCAMEEGNet under both stimuli are better than other CNN classification models, further confirming the effectiveness of MCAMEEGNet. This study promotes the generalization and practicality of ERP-based RSVP research.

Table 1: Accuracy for Subject-dependent model and Subject-independent model

acc(%)	MCAMEEGNet	DeepConvNet	Shallow ConvNet
Subject-dependent model	94.77±3.8	90.12±5.4	92.78±4.8
Subject-independent model	87.99±7.3	82.84±8.7	85.62±6.3

3.3 Ablation Experiment

Furthermore, in order to validate the effectiveness of the MCAMEEGNet model proposed in this paper, that is the role of multi-scale convolution networks and attention mechanisms in the model, we conducted specific ablation experiments. We compared the classification accuracy of the benchmark models EEGNet, Attention+EEGNet, Multi-scale+EEGNet and MCAMEEGNet. Table 2 further demonstrates the classification results of 10 subjects on each model, and the results show that MCAMEEGNet outperforms all other models in terms of the average classification accuracy of the

ISSN 2616-7433 Vol. 7, Issue 4: 58-66, DOI: 10.25236/FSST.2025.070408

subjects.The results of the three models, Attention+EEGNet, Muti-cale+EEGNet, and MCAMEEGNet, are closer for both S3 and S8. This may be due to the more significant ERP signal and higher stability of this subject. This result further confirms that multi-scale convolution networks and hybrid attention mechanisms play an important role in MCAMEEGNet for classification based on ERP target detection signals, resulting in a greater improvement in the performance of MCAMEEGNet.Comparison of Accuracy for Ablation Study is shown in Figure 7.

Table 2: Ablation Study of MCAMEEGNet.

Subjects	MCAMEEGNet	Multiscale+EEGNet	Attention+EEGNet	EEGNet
S1	0.9453	0.9237	0.9287	0.8932
S2	0.9348	0.9178	0.9149	0.9276
S3	0.9816	0.9864	0.9807	0.9554
S4	0.9324	0.9025	0.9102	0.9063
S5	0.9642	0.9419	0.9478	0.9496
S6	0.9217	0.8918	0.9137	0.9156
S7	0.9378	0.9226	0.9273	0.9029
S8	0.9652	0.9658	0.9629	0.9227
S9	0.9413	0.9045	0.9327	0.9054
S10	0.9529	0.9195	0.9224	0.9176
Avg	0.9477	0.9369	0.9383	0.9281

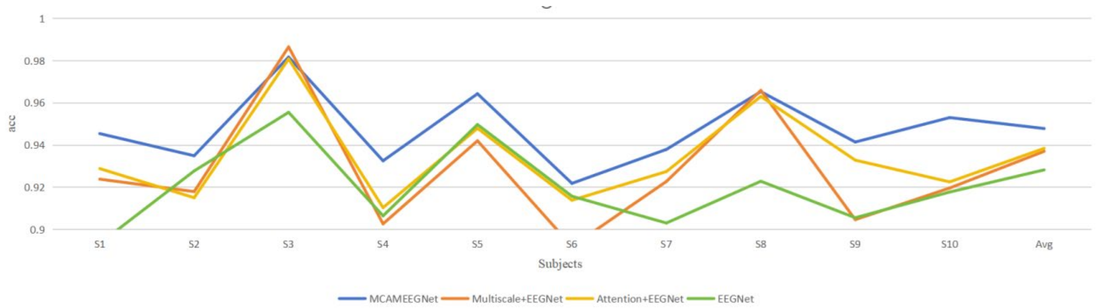

Figure 7: Comparison of Accuracy for Ablation Study.

4. Conclusion

In this paper, we investigate target recognition based on the RSVP paradigm, using EEG signals to distinguish between targets and non-targets. In this paper, we use convolutional neural networks for classification and propose a novel network model——MCAMEEGNet, which incorporates multi-scale convolution networks and attention mechanisms in the benchmark model——EEGNet.MCAMEEGNet firstly adds parallel convolution kernels to the convolutional layer to obtain richer features, which demonstrates excellent performance in RSVP target recognition based on ERP signals, with accuracy up to 94.77%, higher than other EEG classification models mentioned in this paper. Additionally we conducted a cross-subject study, and the classification performance across subjects was lower than that of a single subject, which is expected because the CNN parameters are tuned for each subject's data, indicating that there is individual variability among EEG signals. Finally, specific ablation experiments were conducted to confirm that multi-scale convolution networks and attention mechanisms play an important role in the MCAMEEGNet architecture. Overall, the MCAMEEGNet model has good performance in the target detection and classification task, and future exploration is needed to face more complex scenarios to verify the generalization ability and adaptability of deep learning in different scenarios. In addition, future research will address the challenge of online decoding, emphasising the importance of short-time decoding strategies. By focusing on real-time processing, it aims to improve the responsiveness and accuracy of the model in real-world applications, paving the way for more effective brain-computer interfaces.

ISSN 2616-7433 Vol. 7, Issue 4: 58-66, DOI: 10.25236/FSST.2025.070408

References

[1] Xiyu S, Bin Y, Li T, et al. Asynchronous Video Target Detection Based on Single-Trial EEG Signals[J]. IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society, 2020, 28(9): 1931-1943.

[2] RAMADAN R A, VASILAKOS A V, et al. Brain computer interface: control signals review[J]. Neurocomputing, 2017 (223): 26-44.

[3] Zhimin L, Ying Z, Hui G, et al. Multirapid Serial Visual Presentation Framework for EEG-Based Target Detection[J]. BioMed research international, 2017, 20(4): 90-94.

[4] Lees S, Dayan N, Cecotti H, et al. A review of rapid serial visual presentation-based brain–computer interfaces[J]. Journal of Neural Engineering, 2018, 15(2): 021001.

[5] Nima B, Andrey V, et al. Brain activity-based image classification from rapid serial visual presentation[J]. IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society, 2008, 16(5): 432-41.

[6] Barngrover C et al. A brain–computer interface (BCI) for the detection of mine-like objects in sidescan sonar imagery[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 4(1): 124–39.

[7] Matran-Fernandez A, Poli R. Collaborative brain-computer interfaces for target localisation in rapid serial visual presentation[C]. Computer Science and Electronic Engineering Conference, 2014: 127-132.

[8] Jon T, Laurie G, et al. Real-time measurement of face recognition in rapid serial visual presentation[J]. Frontiers in psychology, 2011, 242.

[9] Haofei W, Yiwen W. Convolutional Neural Network for Target Face Detection using Single-trial EEG Signal[J]. IEEE Engineering in Medicine and Biology Society, 2018, 2008-2011.

[10] Pradeep S, Desney S. Human-Aided Computing: Utilizing Implicit Human Processing to Classify Images[C]. The 26th Annual CHI Conference on Human Factors in Computing Systems (CHI 2008), 2008, 11(1):845-854.

[11] Chris H, Annette S, Premkumar E, et al. High throughput screening for mammography using a human-computer interface with rapid serial visual presentation(RSVP)[J].Univ. of Surrey (United Kingdom);The Royal Surrey County Hospital NHS Trust (United Kingdom);Univ. of California, Santa Barbara (United States);Univ. of Pittsburgh (United States), 2013, 86(73): 3-8.

[12] Hubert C, Axel G. Convolutional neural networks for P300 detection with application to brain-computer interfaces[J]. IEEE transactions on pattern analysis and machine intelligence, 2011, 33(3): 433-45.

[13] Pedram H, Maryam Z, Elham M, et al. An efficient deep learning framework for P300 evoked related potential detection in EEG signal[J]. Computer methods and programs in biomedicine, 2022, 229, 107324-107324.

[14] Ziwei Z, Yangfei L, et al. RSVP Target Classification Algorithm for Multi-Layer Frequency Spatio-Temporal Feature Extraction[J]. Journal of Beijing Institute of Technology, 2024, 44(03): 312-320.

[15] Yuan Z, Zhou Q, Wang B, et al. PSAEEGNet: pyramid squeeze attention mechanism-based CNN for single-trial EEG classification in RSVP task[J]. Frontiers in Human Neuroscience, 2024, 18,1385360-1385360.

[16] Chen H, Wang D, Xu M, et al. A multi-source classification framework with invariant representation reconstruction for dual-target RSVP-BCI tasks in cross-subject scenario[J]. Neurocomputing, 2025, 620,129239-129239.

[17] Lawhern J V, Solon J A, Waytowich R N, et al. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces[J]. Journal of Neural Engineering, 2018, 15(5): 056013.

[18] G. H. Klem, et al. The ten-twenty electrode system of the International Federation[J]. Electroencephalogr Clin Neurophysiol, 1999, 52(1): 3-6.

[19] Hongfei Z, Zehui W, Yinhu Y, et al. An improved EEGNet for single-trial EEG classification in rapid serial visual presentation task[J]. Brain Science Advances, 2022, 8(2): 111-126.

[20] Arnaud D, Scott M. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis[J]. Journal of neuroscience methods, 2004, 134(1): 9-21.

[21] Subasi A, Gursoy I M. EEG signal classification using PCA, ICA, LDA and support vector machines[J]. Expert Systems with Applications, 2010, 37(12): 8659-8666.