MambaMER: Adaptive EEG-Guided Multimodal Emotion Recognition with Mamba

Xiangle Ping, Wenhui Huang* and Yuanjie Zheng

School of Information Science and Engineering, Shandong Normal University, Jinan, China
whhuang.sdu@gmail.com
* Corresponding author

Abstract. In recent years, multimodal emotion recognition has gradually become a research hotspot. Although existing methods have achieved significant results by integrating information from different modalities, irrelevant or conflicting emotional information across modalities often limits its performance improvement. Inspired by Mamba's ability to effectively filter irrelevant information and model long-range dependencies with linear complexity, we propose a new paradigm for EEG-guided adaptive multimodal emotion recognition with Mamba. This paradigm effectively addresses the interference caused by cross-modal information conflicts, enhancing the performance of multimodal emotion recognition. Firstly, to alleviate the interference caused by conflicts between different modalities, we design a multi-scale EEG-guided conflict suppression module. Guided by multi-scale EEG features, this module uses a selective cross state space model to suppress irrelevant information and conflicts in eye movement features, thereby obtaining enhanced eye movement features. Secondly, to deeply integrate the complementary features between the EEG modality and the enhanced eye movement modality, we propose a novel cross-modal fusion mechanism, consisting of Mutual-Cross-Mamba and Merge-Mamba, which effectively captures long-range dependencies in the fused features, thereby enhancing the integration and utilization of cross-modal information. Experimental results on the SEED, SEED-IV, and SEED-V datasets demonstrate that our method significantly surpasses current state-of-the-art methods.

Keywords: EEG · Eye Movement · Multimodal Emotion Recognition · Mamba.

1 Introduction

Multimodal emotion recognition (MER) has gained attention for its ability to decode complex human emotions [9]. By integrating physiological signals and behavioral data, multimodal methods improve emotion recognition accuracy. Electroencephalography (EEG) captures dynamic changes linked to emotional states, making it a key modality [12, 29]. Eye movement data offers valuable behavioral and cognitive cues, especially in response to external stimuli [4, 19,

28]. Combining EEG with eye movement data provides both neurological and behavioral insights, leading to a more comprehensive understanding of emotional states [5, 6, 22].

In recent years, with the continuous advancement of multimodal emotion recognition research, many methods have been proposed that integrate information from different modalities and explore their complementarity and individual characteristics to achieve more accurate emotion recognition [1,25,26]. For example, Fu et al. [5] designed a multimodal feature fusion neural network to capture complementary eye movement and EEG information, while Wang et al. [23] applied an attention mechanism to integrate EEG and eye movement signals for emotion recognition. However, these methods often fail to fully account for irrelevant or conflicting emotional information between modalities. Additionally, using attention mechanisms for cross-modal fusion significantly increases computational complexity.

To address the aforementioned issues, this paper proposes a Mamba-based adaptive EEG-guided multimodal emotion recognition model that effectively suppresses emotion-irrelevant interference and conflicting information in eye movement signals, thereby enhancing the performance of multimodal emotion recognition. The main contributions of this work are as follows:

1. We propose an adaptive EEG-guided multimodal emotion recognition model based on Mamba, which effectively mitigates interference from cross-modal conflicts and enhances recognition accuracy. To the best of our knowledge, this is the first study to apply the Mamba model to resolve cross-modal conflicts in emotion recognition.
2. We designed a multi-scale EEG-guided conflict suppression module, which learns to suppress irrelevant information and conflicts in the eye movement features under the guidance of multi-scale EEG features, thereby enhancing the eye movement features.
3. We propose a novel cross-modal fusion mechanism consisting of Mutual-Cross-Mamba and Merge-Mamba, which aims to realize the deep interaction of multimodal features and effectively capture the remote dependencies in the fused features.

2 Method

2.1 Preliminaries

The state-space model (SSM) [8] consists of a state equation describing the system's internal dynamics and an observation equation linking the system state to the observed values. For an input $x(t) \in \mathbb{R}$ and a hidden state $y(t) \in \mathbb{R}$, the system is represented by a linear ordinary differential equation (ODE) as follows:

$$h'(t) = \mathbf{A}h(t) + \mathbf{B}x(t), \quad y(t) = \mathbf{C}h(t). \tag{1}$$

Here, $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the state matrix, $\mathbf{B} \in \mathbb{R}^{N\times 1}$ is the input matrix, and $\mathbf{C} \in \mathbb{R}^{1\times N}$ is the output matrix. Mamba integrates continuous systems with

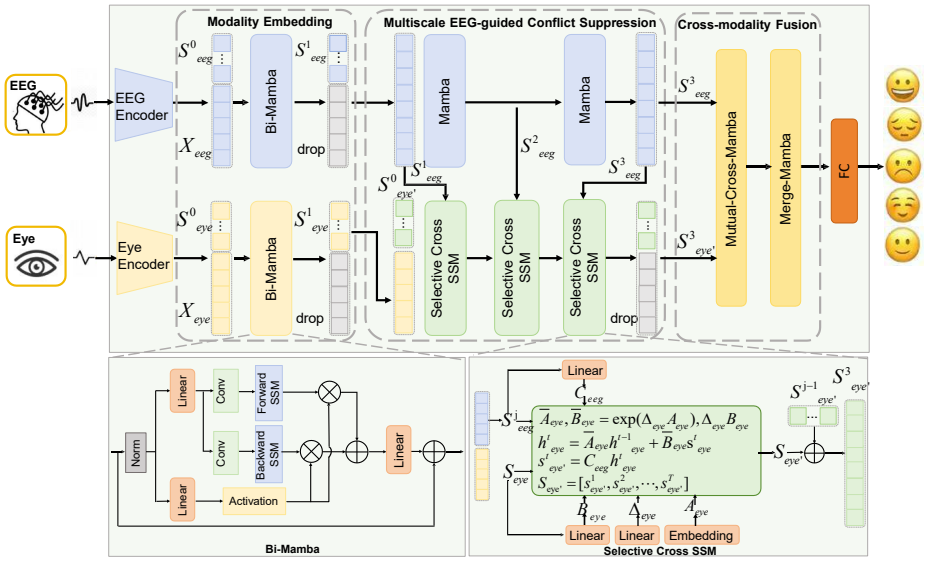

Fig. 1. Our framework consists of three modules: modality embedding, multi-scale EEG-guided conflict suppression, and cross-modal fusion. The modality embedding module utilizes Bi-Mamba to extract unified modality features from EEG and eye movement data. The multi-scale EEG-guided conflict suppression module extracts multi-scale EEG features using two Mamba blocks and uses these features to guide the Selective Cross-SSM in generating enhanced eye movement feature representations. The cross-modal fusion module integrates multimodal information through Mutual-Cross-Mamba and Merge-Mamba, and outputs the final emotion classification result via a fully connected layer.

deep learning algorithms by applying zero-order hold (ZOH) [20] to transform the continuous parameters $\mathbf{A}$ and $\mathbf{B}$ into their discrete counterparts $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$, incorporating the time scaling parameter Δ. The conversion formula is as follows:

$$\overline{\mathbf{A}} = \exp(\Delta\mathbf{A}), \quad \overline{\mathbf{B}} = (\Delta\mathbf{A})^{-1}(\exp(\Delta\mathbf{A}) - \mathbf{I}) \cdot \Delta\mathbf{B}. \tag{2}$$

In practice, following the approach of [7], we approximate $\overline{\mathbf{B}}$ using the first-order Taylor series as follows:

$$\overline{\mathbf{B}} = (e^{\Delta\mathbf{A}} - \mathbf{I})\mathbf{A}^{-1}\mathbf{B} \approx (\Delta\mathbf{A})(\Delta\mathbf{A})^{-1}\Delta\mathbf{B} = \Delta\mathbf{B}. \tag{3}$$

This approach converts the continuous ODE into a discrete form. The expression is as follows:

$$h_t = \overline{\mathbf{A}}h_{t-1} + \overline{\mathbf{B}}x_t, \quad y_t = \mathbf{C}h_t. \tag{4}$$

Building on the aforementioned discrete state-space equations, mamba introduces data dependencies into the model parameters, enabling it to selectively

propagate or forget information based on sequential inputs. Additionally, a parallel scanning algorithm is employed to accelerate the equation solving process [7].

2.2 Modality Embedding

Given the raw EEG signal U_{eeg} and eye movement U_{eye}, we extract EEG features $X_{\text{eeg}} \in \mathbb{R}^{n \times d_1}$ and eye movement features $X_{\text{eye}} \in \mathbb{R}^{n \times d_2}$ using the method from [11], where n is the number of samples, and d_1, d_2 are the feature dimensions. To unify the feature representations for each modality, we introduce two Bi-Mamba blocks and initialize low-dimensional token sequences $S^0_{\text{eeg}}, S^0_{\text{eye}} \in \mathbb{R}^{T \times d}$, where T is the sequence length and d is the token feature dimension. The EEG and eye movement features are then embedded into their respective tokens via the Bi-Mamba blocks:

$$S^1_{eeg} = \mathrm{E}^0_{eeg}\left(\text{concat}\left(S^0_{eeg}, X_{eeg}\right), \theta_{E^0_{eeg}}\right) \in \mathbb{R}^{T \times d}, \tag{5}$$

$$S^1_{eye} = \mathrm{E}^0_{\text{eye}}\left(\text{concat}\left(S^0_{eye}, X_{eye}\right), \theta_{E^0_{eye}}\right) \in \mathbb{R}^{T \times d}. \tag{6}$$

Here, $\mathrm{E}^0_{\text{eeg}}$ and $\mathrm{E}^0_{\text{eye}}$ denote the Bi-Mamba layers with parameters $\theta_{\mathrm{E}^0_{\text{eeg}}}$ and $\theta_{\mathrm{E}^0_{\text{eye}}}$, respectively. The operation concat$(\cdot)$ represents concatenation. Bi-Mamba embeds the modality features into tokens, generating the unified feature representations S^1_{eeg} and S^1_{eye}.

2.3 Multi-Scale EEG-Guided Conflict Suppression

After achieving a unified representation transformation for the features of each modality, our goal is to filter out irrelevant or conflicting interference information between modalities. Specifically, we define the S^1_{eeg} as low-scale EEG features and further extract medium-scale and high-scale EEG features (i.e., S^2_{eeg} and S^3_{eeg}) by introducing two Mamba layers:

$$S^i_{eeg} = \mathrm{E}^i_{eeg}\left(S^{i-1}_{eeg}, \theta_{E^i_{eeg}}\right) \in \mathbb{R}^{T \times d}, \tag{7}$$

where $i \in \{2, 3\}$ represents EEG features at different scales, with $\mathrm{E}^i_{\text{eeg}}$ and $\theta_{E^i_{\text{eeg}}}$ denoting the i-th Mamba layer and its parameters.

Next, we initialize the enhanced eye movement feature $S^0_{eye'} \in \mathbb{R}^{T \times d}$ and update it by calculating the relationship between EEG and eye movement features using selective cross-state-space computation. The process is as follows:

$$\overline{\mathbf{A}}_{\text{eye}}, \overline{\mathbf{B}}_{eye} = \exp(\Delta_{\text{eye}}\mathbf{A}_{\text{eye}}), \Delta_{\text{eye}}\mathbf{B}_{\text{eye}}, \tag{8}$$

$$h^t{}_{eye} = \overline{\mathbf{A}}_{eye} h^{t-1}{}_{eye} + \overline{\mathbf{B}}_{eye} s^t{}_{eye}, \tag{9}$$

$$s^t{}_{eye'} = \mathbf{C}^i_{eeg} h^t{}_{eye}, \tag{10}$$

$$S_{eye'} = \left[s^1{}_{eye'}, s^2{}_{eye'}, ..., s^T{}_{eye'}\right]. \tag{11}$$

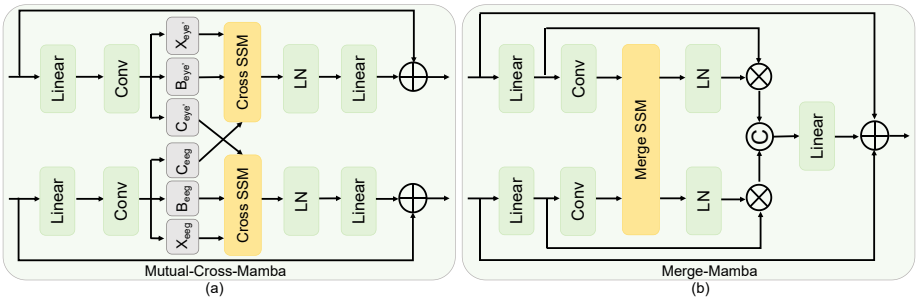

Fig. 2. Cross-modality fusion. (a) Mutual-Cross-Mamba and (b) Merge-Mamba

Here, $\mathbf{C}_{\text{eeg}}$ is computed from the EEG features at different scales, and the enhanced eye movement modality feature $S^{j}_{eye'}$ can be updated as follows:

$$S^{j}_{eye'} = S^{j-1}_{\text{eye}'} + S_{\text{eye}'}, \tag{12}$$

where $j \in \{1, 2, 3\}$ and $S^{j}_{eye'} \in \mathbb{R}^{T \times d}$ represent the j-th selective cross-state-space model and its corresponding output.

2.4 Cross-Modal Fusion

The proposed cross-modal fusion mechanism consists of the Mutual-Cross-Mamba (MutCroMB) block and the Merge-Mamba (MerMB) block. MutCroMB enhances inter-modal interactions by applying the SSM in both directions across modalities. MerMB uses a selective scanning mechanism to fuse interacting features, producing the final fusion result. The fusion process is described below:

$$\begin{aligned} \widehat{S}_{\text{eeg}}, \widehat{S}_{\text{eye}'} &= \text{MutCroMB}\big(S^{3}_{\text{eeg}}, S^{3}_{\text{eye}'}\big), \\ S_{\text{fusion}} &= \text{MerMB}\big(\widehat{S}_{\text{eeg}}, \widehat{S}_{\text{eye}'}\big). \end{aligned} \tag{13}$$

Specifically, MutCroMB enhances modal feature interaction via a cross-multiplication mechanism. As shown in Fig. 2 (a), the two input features are processed by convolution and then passed into the cross SSM. According to Eq. (4), the matrix $\mathbf{C}$ decodes information from the hidden state h_t to generate the output y_t.

In MutCroMB, features from both modalities interact to generate cross-modal enhanced features. To capture long-range dependencies, the MerMB module integrates MutCroMB's output. As shown in Fig. 2 (b), $\widehat{S}_{\text{eeg}}$ and $\widehat{S}_{\text{eye}'}$ are first processed through linear layers and convolutional layers, and then enter the Merge SSM block. To ensure comprehensive information capture, we perform a reverse scan on the merged sequence S_{Merge}, producing S_{Inverse}, which undergoes further processing to yield $\hat{S}_{\text{Merge}}$ and $\hat{S}_{\text{Inverse}}$. Finally, the inverted output is reversed and subsequently added to the merged sequence. This process can be represented as:

$$\widetilde{S}_{\text{eeg}}, \widetilde{S}_{\text{eye}'} = \text{Conv}(\text{Linear}(\hat{S}_{\text{eeg}})), \text{Conv}(\text{Linear}(\hat{S}_{\text{eye}'})), \tag{14}$$

$$S_{\text{Merge}} = \text{Concat}(\widetilde{S}_{\text{eeg}}, \widetilde{S}_{\text{eye}'}), \tag{15}$$

$$S_{\text{Inverse}} = \text{Inverse}(S_{\text{Merge}}), \tag{16}$$

$$\hat{S}_{\text{Merge}}, \hat{S}_{\text{Inverse}} = \text{SSM}(S_{\text{Merge}}), \text{SSM}(S_{\text{Inverse}}), \tag{17}$$

$$S_{\text{fusion}} = \hat{S}_{\text{Merge}} + \text{Inverse}(\hat{S}_{\text{Inverse}}). \tag{18}$$

Through this process, we obtain the deeply fused feature S_{fusion}, which is then passed through a fully connected layer and a softmax function for classification. The model is optimized by minimizing the cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N} \sum_{i} \sum_{c=1}^{M} y_{ic} \log \left(p_{ic} \right), \tag{19}$$

where M is the total number of categories, y_{ic} is a binary indicator (0 or 1) for sample i belonging to category c, and p_{ic} is the predicted probability of sample i belonging to category c.

3 Experiment

3.1 Datasets

We conducted experiments on three datasets: SEED [16], SEED-IV [27], and SEED-V [14]. The SEED dataset includes data from 12 Chinese subjects watching 15 clips covering happy, neutral, and sad emotions. The SEED-IV contains data from 15 subjects, involving four emotions (happy, sad, fear, and neutral). The SEED-V includes data from 16 subjects, covering five emotions (happy, sad, fear, neutral, and disgust).

3.2 Implementation Details

We followed the training/testing protocols from the original papers for each dataset, using the same data splits as previous studies [14, 16, 27]. The model was trained for 100 epochs with a batch size of 16, using the Adam optimizer and a learning rate of $1e^{-4}$. A dropout layer with a rate of 0.5 was added to prevent overfitting. All experiments were performed on an NVIDIA RTX 3090 with CUDA 11.8, using Python 3.10.13 and PyTorch 2.1.1.

3.3 Experiment Results

We evaluated the proposed method on the SEED, SEED-IV, and SEED-V datasets and conducted a comparative analysis with other multimodal methods.

Table 1. Comparison of the average accuracy and standard deviation (%) of various multimodal methods across different datasets. (Bold indicates the best accuracy.)

Methods	Modality	SEED	SEED-IV	SEED-V
Fuzzy Intergral [15, 18]	EEG and EM	87.60±19.9	73.60±16.7	73.20±8.70
BDAE [17, 27]	EEG and EM	91.01±8.91	85.11±11.79	79.70±4.76
DCCA [15]	EEG and EM	94.60±6.12	87.50±9.20	85.30±5.6
DCCA-FCP [24]	EEG and EM	95.08±6.42	-	84.51±5.11
CAN [21]	EEG and EM	94.03±6.62	87.71±9.74	-
ECO-FET [10]	EEG and EM	93.69±8.22	87.76±9.19	77.13±4.16
ATAM [13]	EEG and EM	94.80±7.50	91.60±10.0	-
Ours	EEG and EM	**96.82±5.20**	**94.93±6.12**	**86.95±5.89**

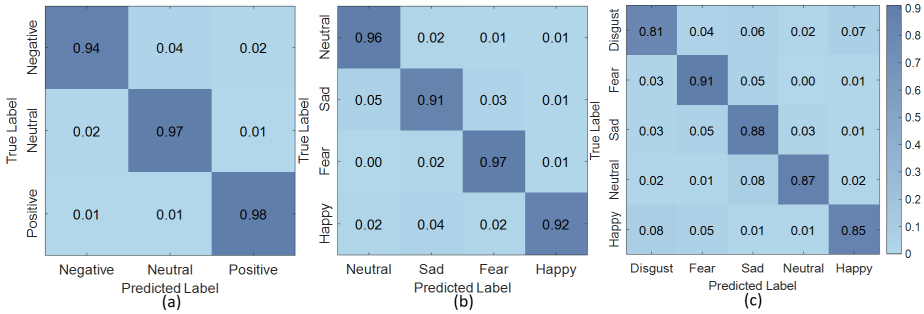

Fig. 3. Confusion matrix (a) SEED, (b) SEED-IV and (c) SEED-V.

The experimental results in Table 1 show that, on the SEED dataset, the average recognition accuracy of our model reaches 96.82%, an improvement of 1.74% over the best-performing model; on the SEED-IV dataset, the average recognition accuracy is 94.93%, an increase of 3.33%; and on the SEED-V dataset, the average recognition accuracy is 86.95%, surpassing the best model by 1.65%. These results thoroughly demonstrate the model's outstanding performance in emotion recognition tasks.

Fig. 3 presents the confusion matrices of our proposed method on the SEED, SEED-IV, and SEED-V datasets. It is clear that on the SEED dataset neutral and positive emotions are easy to distinguish. On the SEED-IV, neutral and fear emotions are easier to recognise than sadness and happiness. On the SEED-V, fear is the easiest emotion to recognise, while disgust is the most difficult.

3.4 Ablation Study and Analysis

Effects of different modalities and dominant modalities. We conducted ablation experiments to assess the impact of different modalities. Table 2 presents results on the SEED, SEED-IV, and SEED-V datasets, comparing EEG, eye movement, and both modalities, with each signal as the dominant modality. On SEED, the model's accuracy with only eye movement signals is 83.18%, while using eye movement as the dominant modality improves accuracy to 87.34%.

Table 2. Effects of different modalities and dominant modalities (mean ± std. dev (%)).

Modality	SEED	SEED-IV	SEED-V
EM	83.18±7.05	80.35±7.23	71.68±8.11
EEG	94.21±5.66	91.29±6.03	83.62±6.41
EEG+EM (EM Dominant)	87.34±6.89	83.68±6.47	75.39±7.15
EEG+EM (EEG Dominant)	96.82±5.20	94.93±6.12	86.95±5.89

Table 3. Effects of different components (mean ± std. dev (%).)

Methods	SEED	SEED-IV	SEED-V
w/o MECS	86.76±7.38	85.45±7.96	78.69±5.63
w/o CMF	91.52±6.59	89.03±6.28	81.72±7.04
w/o ME	92.18±6.93	90.52±7.21	83.26±7.16
MambaMER	96.82±5.20	94.93±6.12	86.95±5.89

Table 4. Performance comparison of mamba and transformer.

Methods	TransformerMER	MambaMER
Accuracy	96.21%	96.82%
FLOPs	1.58B	0.86B
Parameters	86.72M	28.56M

However, eye movement as the dominant modality performs worse than EEG alone (94.21%) or EEG as the dominant modality (96.82%). This indicates that EEG, with its stronger emotional expression, more effectively guides eye movement signals, a conclusion consistent across the other datasets.

Effects of Different Components. To evaluate the contributions of each component in the proposed model, we conducted ablation experiments on the SEED, SEED-IV, and SEED-V datasets, with the results presented in Table 3. Removing the EEG-guided conflict suppression (MECS) module led to a significant drop in performance, highlighting its crucial role in capturing cross-modal relationships. Similarly, removing the cross-modal fusion (CMF) module also resulted in a performance decline, further emphasizing its importance in integrating complementary information and enhancing inter-modal interactions. Additionally, removing the modality embedding (ME) module also reduced performance. These results thoroughly demonstrate the key role of the MECS, CMF, and ME modules in enhancing the model's expressive power and robustness.

Performance Comparison of Mamba and Transformer. By replacing Mamba with Transformer [2, 3] in the model, the TransformerMER model was constructed. We compared the accuracy, FLOPs, and parameters of MambaMER and the deformation model on the SEED dataset to validate the effectiveness and superiority of the proposed MambaMER model. Table 4 summarizes the experimental results of the two methods. The results indicate that MambaMER slightly outperforms the deformation model in terms of accuracy, while demonstrating a significant advantage in FLOPs and parameters. Therefore, the experimental results provide strong evidence for the feasibility of the MambaMER model and its advantages over TransformerMER.

4 Conclusion and Discussion

In this paper, we propose an adaptive EEG-guided multimodal emotion recognition paradigm based on Mamba. To mitigate interference caused by conflicts between different modalities, we design a multi-scale EEG-guided conflict suppression module. Guided by EEG features at different scales, a selective cross-state space model extracts conflict-independent representations from eye movement features, effectively filtering out irrelevant or conflicting information. Additionally, we introduce a novel cross-modal fusion mechanism to enhance feature interaction and capture long-range dependencies. Experiments on SEED, SEED-IV, and SEED-V datasets show that our method outperforms state-of-the-art approaches. In the future, we will focus on addressing the issue of missing modalities in cross-modal fusion and explore more adaptive mechanisms to effectively compensate for the impact of information loss on emotion recognition performance.

Acknowledgements. This work is supported by funds from the National Natural Science Foundation of China (62476156, 62003196, 62076249, 62072289, and 62073201), the Young Talent of Lifting Engineering for Science and Technology in Shandong (SDAST2024QTA091), the Youth Innovation Technology Project of Higher Education in Shandong Province (2023KJ193) and the Provincial Natural Science Foundation of Shandong Province of China (ZR2020QF032).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Che, T., Zheng, Y., Yang, Y., Hou, S., Jia, W., Yang, J., Gong, C.: Sdof-gan: Symmetric dense optical flow estimation with generative adversarial networks. IEEE Transactions on Image Processing **30**, 6036–6049 (2021)
2. Chen, Z., Zheng, Y., Gee, J.C.: Transmatch: a transformer-based multilevel dual-stream feature matching network for unsupervised deformable image registration. IEEE transactions on medical imaging **43**(1), 15–27 (2023)
3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
4. Fan, X., Xu, P., Zhao, Q., Hao, C., Zhao, Z., Wang, Z.: A domain adaption approach for eeg-based automated seizure classification with temporal-spatial-spectral attention. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 14–24. Springer (2024)
5. Fu, B., Gu, C., Fu, M., Xia, Y., Liu, Y.: A novel feature fusion network for multimodal emotion recognition from eeg and eye movement signals. Frontiers in Neuroscience **17**, 1234162 (2023)

6. Gong, X., Chen, C.P., Zhang, T.: Cross-cultural emotion recognition with eeg and eye movement signals based on multiple stacked broad learning system. IEEE Transactions on Computational Social Systems (2023)
7. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
8. Gu, A., Goel, K., Gupta, A., Ré, C.: On the parameterization and initialization of diagonal state space models. Advances in Neural Information Processing Systems **35**, 35971–35983 (2022)
9. Hazmoune, S., Bougamouza, F.: Using transformers for multimodal emotion recognition: Taxonomies and state of the art review. Engineering Applications of Artificial Intelligence **133**, 108339 (2024)
10. Jiang, W.B., Li, Z., Zheng, W.L., Lu, B.L.: Functional emotion transformer for eeg-assisted cross-modal emotion recognition. In: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1841–1845. IEEE (2024)
11. Li, L., Deng, W., Liao, S., Qiang, X., Rong, Y., Yang, Y., Liu, S., Zhang, Y.: Stmnet based spatial-temporal multi-modal fusion network for emotion recognition. In: International Conference on Image, Signal Processing, and Pattern Recognition (ISPP 2024). vol. 13180, pp. 1250–1255. SPIE (2024)
12. Liu, H., Lou, T., Zhang, Y., Wu, Y., Xiao, Y., Jensen, C.S., Zhang, D.: Eeg-based multimodal emotion recognition: a machine learning perspective. IEEE Transactions on Instrumentation and Measurement (2024)
13. Liu, W., Luo, Y., Lu, Y., Lu, Y.: A multitask framework for emotion recognition using eeg and eye movement signals with adversarial training and attention mechanism. In: 2023 IEEE International conference on bioinformatics and biomedicine (BIBM). pp. 2551–2558. IEEE (2023)
14. Liu, W., Qiu, J.L., Zheng, W.L., Lu, B.L.: Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition. IEEE Transactions on Cognitive and Developmental Systems **14**(2), 715–729 (2021)
15. Liu, W., Qiu, J.L., Zheng, W.L., Lu, B.L.: Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition. IEEE Transactions on Cognitive and Developmental Systems **14**(2), 715–729 (2021)
16. Liu, W., Zheng, W.L., Li, Z., Wu, S.Y., Gan, L., Lu, B.L.: Identifying similarities and differences in emotion recognition with eeg and eye movements among chinese, german, and french people. Journal of Neural Engineering **19**(2), 026012 (2022)
17. Liu, W., Zheng, W.L., Lu, B.L.: Emotion recognition using multimodal deep learning. In: Neural Information Processing: 23rd International Conference, ICONIP 2016, Kyoto, Japan, October 16–21, 2016, Proceedings, Part II 23. pp. 521–529. Springer (2016)
18. Lu, Y., Zheng, W.L., Li, B., Lu, B.L.: Combining eye movements and eeg to enhance emotion recognition. In: IJCAI. vol. 15, pp. 1170–1176. Buenos Aires (2015)
19. Nordfält, J., Ahlbom, C.P.: Utilising eye-tracking data in retailing field research: A practical guide. Journal of Retailing **100**(1), 148–160 (2024)
20. Pechlivanidou, G., Karampetakis, N.: Zero-order hold discretization of general state space systems with input delay. IMA Journal of Mathematical Control and Information **39**(2), 708–730 (2022)
21. Qiu, J.L., Li, X.Y., Hu, K.: Correlated attention networks for multimodal emotion recognition. In: 2018 IEEE international conference on bioinformatics and biomedicine (BIBM). pp. 2656–2660. IEEE (2018)

22. Wang, J., Zheng, Y., Ma, J., Li, X., Wang, C., Gee, J., Wang, H., Huang, W.: Information bottleneck-based interpretable multitask network for breast cancer classification and segmentation. Medical image analysis **83**, 102687 (2023)
23. Wang, Y., Jiang, W.B., Li, R., Lu, B.L.: Emotion transformer fusion: Complementary representation properties of eeg and eye movements on recognizing anger and surprise. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1575–1578. IEEE (2021)
24. Wu, X., Zheng, W.L., Li, Z., Lu, B.L.: Investigating eeg-based functional connectivity patterns for multimodal emotion recognition. Journal of neural engineering **19**(1), 016012 (2022)
25. Yang, D., Huang, S., Kuang, H., Du, Y., Zhang, L.: Disentangled representation learning for multimodal emotion recognition. In: Proceedings of the 30th ACM international conference on multimedia. pp. 1642–1651 (2022)
26. Zhao, Y., Gu, J.: Feature fusion based on mutual-cross-attention mechanism for eeg emotion recognition. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 276–285. Springer (2024)
27. Zheng, W.L., Liu, W., Lu, Y., Lu, B.L., Cichocki, A.: Emotionmeter: A multimodal framework for recognizing human emotions. IEEE transactions on cybernetics **49**(3), 1110–1122 (2018)
28. Zheng, Y., Sui, X., Jiang, Y., Che, T., Zhang, S., Yang, J., Li, H.: Symreg-gan: symmetric image registration with generative adversarial networks. IEEE transactions on pattern analysis and machine intelligence **44**(9), 5631–5646 (2021)
29. Zheng, Y., Yang, Y., Che, T., Hou, S., Huang, W., Gao, Y., Tan, P.: Image matting with deep gaussian process. IEEE transactions on neural networks and learning systems **34**(11), 8879–8893 (2022)