Efficient Channel Attention-Enhanced ResNet Variants for EOG-Based Reading Classification

Youngmi Song, Mooseop Kim, SuGil Choi, Sungjun Wang, Chi Yoon Jeong
Human Sensory Augmentation Research Section
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
Email: ymsong@etri.re.kr, gomskim@etri.re.kr, sooguri@etri.re.kr, sungjoonkk@etri.re.kr, iamready@etri.re.kr

***Abstract*—This study proposes a deep learning-based framework for recognizing reading activity using electrooculography (EOG) signals recorded in real-world environments. The one-dimensional EOG signals, corresponding to the left and right eyes, are transformed into five-channel two-dimensional representations to capture richer features. These representations are then processed using a ResNet architecture augmented with an efficient channel attention module to classify the reading activity. To address data scarcity and class imbalance, Mixup augmentation technique is employed. The proposed method is evaluated using a leave-one-subject-out cross-validation protocol on a publicly available dataset and achieves superior accuracy, outperforming existing EOG-based reading activity recognition approaches.**

***Index Terms*—electrooculography, reading-activity classification, convolutional neural networks, sensor data analytics**

I. INTRODUCTION

Electrooculography (EOG) signals measure the potential difference between the cornea and retina, corresponding to the anterior and posterior parts of the human eye. EOG is an effective and widely used technique for understanding, characterizing, and classifying eye movements. Research on analyzing human visual behavior and attention using EOG signals is a promising field with applications in various domains, including human-computer interaction (HCI), medical diagnosis, attention research, driver monitoring, industrial assistance robotics, and patient rehabilitation.

Several field studies have been conducted in controlled laboratory environments to detect reading activities using EOG signals [1], [2]. However, detecting reading activities from data collected in real-world environments poses a complex challenge due to significant noise in the collected signals and the limited availability of training data.

Recent advances in EOG acquisition technology have made it feasible to capture EOG signals in real environments using commercial EOG glasses [3]. An EOG dataset comprising 220 hours of recordings was collected from 10 participants using JINS MEME EOG glasses [3]. The collected signals were annotated with four distinct labels: English reading (EN), horizontal Japanese reading (JH), vertical Japanese reading (JV), and no reading (NR). Further research using this publicly available EOG dataset has been actively conducted [4]–[6].

Islam et al. [4] introduced a self-supervised learning (SSL) approach that outperformed conventional supervised deep learning (DL) and support vector machine (SVM) methods in limited data scenarios. SSL enables the model to generate pseudo-labels from unlabeled data, which facilitates effective learning of gaze patterns and eye movements. Despite the limited number of labeled samples, the proposed method excelled in classifying reading activities. Baray et al. [5] effectively addressed class imbalance by employing a nested classification approach for the detailed identification of reading activities. This method converted 1D EOG signals into 2D spectrograms and extracted features using pre-trained models such as ResNet, VGG, and ViT. The combined features were subsequently used to train a machine learning classifier, achieving an accuracy of 66.56% in reading activity classification. However, the accuracy achieved remains relatively low, suggesting that enhancements to the classification pipeline are necessary to improve performance in real-world settings.

This study aims to enhance the performance of EOG signal analysis by systematically optimizing the classification pipeline using a publicly available dataset designed to detect reading activities in real-world environments. Accordingly, the architecture of the proposed method incorporates three main components: (1) two-dimensional signal images that capture enriched features derived from various EOG signal variants, (2) a ResNet backbone enhanced with an attention mechanism, and (3) the Mixup augmentation strategy to improve generalization performance under limited data conditions.

II. PROPOSED METHOD

This study improves the analysis of EOG signals by utilizing real-world data and optimizing the classification pipeline. The approach involves data preprocessing, signal image generation, and network modeling, as depicted in Fig. 1. Specifically, 1D EOG signals are converted to 2D images for input into a feature extraction network. ResNet with an integrated efficient channel attention module (ECA) [7] is employed to construct a lightweight and effective classification model.

EOG signals from two channels, corresponding to the left and right eyes, were segmented into fixed-length sequences to meet the minimum input requirements of the classification network. Following previous research [5], [6], the continuous signals were divided into non-overlapping 30-second segments. The Savitzky–Golay (SG) filter [8] was applied to enhance signal quality while preserving intrinsic waveform

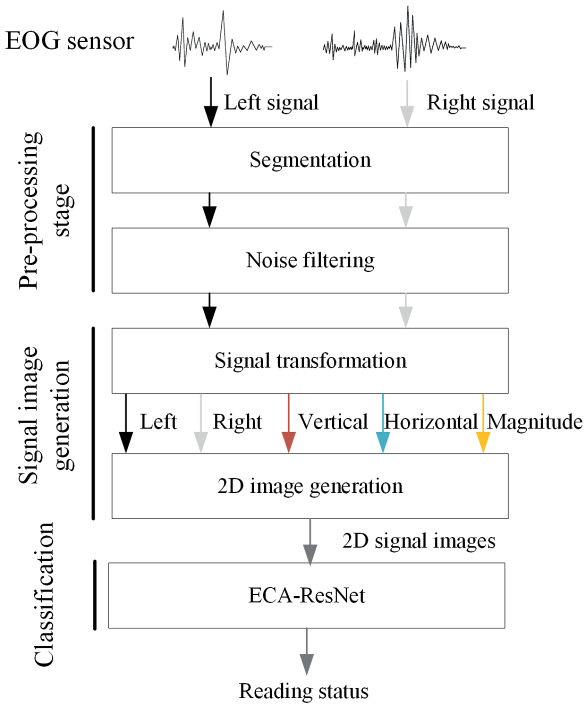

Fig. 1. The proposed framework for reading classification includes pre-processing, transforming signals into images, and classification via a neural network enhanced with the ECA module

characteristics. In addition, horizontal and vertical EOG components were computed from the bilateral EOG signals to extract features related to eye movement. These components were calculated using the equations proposed by Kanoh et al. [9], and their magnitudes were also obtained to enhance the feature representation. The signals were calculated as follows:

$$EOG_h = EOG_l - EOG_r \ , \tag{1}$$

$$EOG_v = -EOG_l + \frac{EOG_r}{2} \ , \tag{2}$$

$$EOG_m = \sqrt{EOG_h^2 + EOG_v^2} \ . \tag{3}$$

Recent advances in image-based signal analysis have increasingly focused on using sophisticated classification networks that leverage 2D signal representations as input. In this study, the original 1D EOG signals, recorded at a frequency of 100 Hz, were divided into 30-second segments, resulting in 3000-sample time series for both horizontal and vertical components. These 1D sequences (100Hz, 30s) were reshaped into square 2D matrices and concatenated along the channel axis to form five-channel composite inputs for a neural network classifier.

This study adopts a channel attention approach that captures lightweight cross-channel interactions while preserving the original channel dimensionality. An ECA module [7] is one of the attention mechanisms used to improve deep convolutional neural networks (CNNs). The ECA module aggregates convolutional features using global average pooling without dimensionality reduction. Subsequently, it adaptively determines the kernel size k, applies a 1D convolution, and uses a sigmoid function to learn channel attention. This module exhibits flexible structural characteristics that facilitate adaptive selection of the kernel size in the 1D convolution, which in turn determines the range of local channel interactions. The ECA module improves the learning of discriminative feature representation and is easily integrated into the CNN backbone. In related studies [7], [10], the effectiveness of the ECA module was demonstrated through performance improvements in various deep CNN architectures, including ResNet and the lightweight MobileNetV2. Thus, the proposed method employed the ResNet-18 network [11] with an integrated ECA module to construct a lightweight and effective classification model.

Data augmentation techniques were employed to enhance model performance and to address the challenges associated with collecting large-scale EOG data, as well as the limited size of public datasets. Specifically, Mixup [12] was adopted considering the characteristics of the EOG-based classification and the structure of the 2D signal images.

III. EXPERIMENTAL RESULTS

The performance of the proposed method was evaluated by applying the leave-one-subject-out cross-validation (LOSO) protocol, which processes the data of each participant in a single fold. By applying the LOSO protocol, the data from one participant served as a test set, while 80% of the data from the remaining participants was used as a training set, and the remaining 20% was used as a validation set. This approach mitigates the risk of specific data that could influence the results, such as individual patterns of participants, leaking into the training and test sets. To address the class imbalance problem of the EOG dataset, the weighted F1 score was used to demonstrate consistent performance, as observed in previous studies [5]. This protocol effectively mitigates data leakage between the training and test sets, ensuring that individual-specific patterns do not bias the evaluation results.

The proposed method was evaluated on a publicly available dataset [3] by integrating the ECA module into a ResNet architecture and employing Mixup augmentation. The network was trained for 25 epochs with a batch size of 16, using the ADAM optimizer and a learning rate of 0.001. The publicly available dataset consists of EOG signals collected from ten participants over two days using commercial EOG glasses. The total number of samples, segmented into 30-second intervals, was 26,421. Participant 7 had the lowest number of samples (2,062), while Participant 4 had the highest (2,873). The average number of samples per participant was 2,642. The LOSO protocol was used to assess the variability in performance between participants.

The reported results represent the average classification performance for the reading activity, obtained from 10 independent experimental runs across 10 participants. The results are summarized in Table I. As shown in Table I, the proposed method achieved an accuracy of 71.42% and an F1 score of

67.21%. Participant 8 achieved the highest performance, with an accuracy of 82.91% and an F1 score of 81.89%, whereas Participant 7 exhibited the lowest results, likely due to subject-specific signal variability or noise. The relatively consistent results across most participants indicate the robustness of the proposed approach, while the drop in performance for Participant 7 suggests potential for subject-specific adaptation in future work.

TABLE I
PERFORMANCE COMPARISON OF FIVE-CLASS CLASSIFICATION USING LOSO PROTOCOL WITH ECA.

Test subject	**Accuracy (%)**	**F1 score (%)**
P1	76.61	72.84
P2	74.98	71.76
P3	74.29	68.11
P4	68.92	69.74
P5	73.81	70.31
P6	71.29	65.31
P7	52.23	43.80
P8	82.91	81.89
P9	70.32	66.90
P10	68.81	61.41
Average	71.42	67.21

Table II summarizes the performance comparison between the proposed method and state-of-the-art approaches. Experimental results indicate that the proposed method achieved an accuracy of 65.85% without data augmentation, and 71.42% when Mixup augmentation was applied, outperforming all prior methods. These findings clearly demonstrate the effectiveness of the proposed architecture and highlight the critical role of Mixup in improving classification performance under limited data conditions.

TABLE II
PERFORMANCE COMPARISON WITH STATE-OF-THE-ART (SOTA) METHODS.

Methods	Accuracy (%)
Statistical features with SVM [3]	32.00
Self-Supervised Learning [4]	57.36
SimCLR [13]	60.38
Supervised DL [13]	56.42
ML, DL combined approach [5]	66.56
Ours (No aug.)	65.85
Ours (w/ Mixup)	71.42

IV. CONCLUSION

In this study, we proposed a deep learning-based framework for recognizing reading activity using EOG signals recorded in real-world environments. By transforming one-dimensional EOG signals into five-channel two-dimensional representations and processing them with a ResNet architecture enhanced by an ECA module, the proposed method effectively captured informative features relevant to eye movement. To address the challenges posed by data scarcity and class imbalance, the Mixup augmentation technique was employed. The experimental results demonstrate that the proposed method is effective and suitable for classifying EOG signals. Although the results demonstrate improved accuracy, alternative architectures, such as Temporal Convolutional Networks (TCNs) and Transformers with efficient self-attention mechanisms, may further enhance model optimization for sequential data. Future research will investigate applying EOG signal analysis across various architectures to identify more efficient models.

ACKNOWLEDGMENT

This work was supported by the Electronics and Telecommunications Research Institute grant funded by Korean government [Fundamental Technology Research for Human-Centric Autonomous Intelligent Systems under Grant 25ZB1200].

REFERENCES

[1] F. Latifoğlu, R. İleri, E. Demirci, and G. Altıntop, "Detection of reading movement from eog signals," in *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2020, pp. 1–5.

[2] F. Latifoğlu, R. İleri, and E. Demirci, "Assessment of dyslexic children with eog signals: Determining retrieving words/re-reading and skipping lines using convolutional neural networks," *Chaos, Solitons Fractals*, vol. 145, p. 110721, 2021.

[3] S. Ishimaru, T. Maruichi, M. Landsmann, K. Kise, and A. Dengel, "Electrooculography dataset for reading detection in the wild," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, ser. UbiComp/ISWC '19 Adjunct. New York, NY, USA: Association for Computing Machinery, 2019, p. 85–88.

[4] M. R. Islam, S. Sakamoto, Y. Yamada, A. W. Vargo, M. Iwata, M. Iwamura, and K. Kise, "Self-supervised learning for reading activity classification," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 5, no. 3, Sep. 2021.

[5] S. B. Baray, M. U. Ahmed, M. E. H. Chowdhury, and K. Kise, "Eog-based reading detection in the wild using spectrograms and nested classification approach," *IEEE Access*, vol. 11, pp. 105 619–105 632, 2023.

[6] C. Y. Jeong, Y. Song, S. Wang, M. Kim, and S. Choi, "End-to-end reading classification in the wild using 2d eog signal images and resnet variants," *ETRI Journal*, pp. 1–12, 2025.

[7] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," *CoRR*, vol. abs/1910.03151, 2019. [Online]. Available: http://arxiv.org/abs/1910.03151

[8] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, pp. 1627–1639, Jan. 1964.

[9] S. Kanoh, S. Ichi-nohe, S. Shioya, K. Inoue, and R. Kawashima, "Development of an eyewear to measure eye and body movements," *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2267–2270, 2015.

[10] C. Y. Jeong, Y. Song, S. Shin, and M. Kim, "Efficient pitch-estimation network for edge devices," *ETRI Journal*, vol. 47, no. 1, pp. 112–122, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.4218/etrij.2023-0430

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[12] H. Zhang, M. Cissé, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018.

[13] M. R. Islam, A. W. Vargo, M. Iwata, M. Iwamura, and K. Kise, "Evaluating contrastive learning for fine-grained reading detection," in *2022 12th International Congress on Advanced Applied Informatics (IIAI-AAI)*, 2022, pp. 430–435.