Spiking-Transformer-based Model for Multimodal Medical Image Processing

Ma Xiaochen[1]

[1]Shandong Academy of Sciences, Key Laboratory of Computing Power Network and Information Security, No. 28666, Jingshi East Road, Licheng District, China, 18654513605

Spiking Neural Networks (SNNs), the third-generation neural networks, exhibit spatiotemporal encoding and energy-efficient characteristics, with applications in brain-computer interface (BCI), machine perception, and natural language processing (NLP). However, SNNs exhibit limited performance compared to conventional deep neural networks (DNNs), mainly due to unstable training cycles [1] and accumulation of multi-timestep latency. This study develops a Spiking-Transformer hybrid model that combines spatiotemporal encoding with a multi-head self-attention mechanism, enabling effective modeling of long-range spatial dependencies in medical images, aiming to improve model classification accuracy and generalization ability for various image processing tasks. The proposed architecture introduces a time-efficient training (TET) mechanism and image enhancement methods to reduce the computational cost while retaining the spatiotemporal information. Unlike existing Spike-driven Transformer [2] models, the proposed framework supports both 2D and 3D medical imaging modalities and is compatible with existing deep learning pipelines. We trained the model on MedMNIST [3] benchmark datasets with experiments conducted on an NVIDIA T4 GPU, using cross-entropy loss ($LOSS_{CE}$) for evaluation. The model's average loss is around 2 by epoch 3 (i.e., the total number of training rounds), and the average top-5 accuracy on the test set is approximately 0.74. These results show a decreasing trend in model loss within the first three epochs, which is consistent with the SNNs' slow convergence characteristics, indicating that the model is effectively learning, almost without signs of overfitting. Further work will increase the training epoch and extend the architecture to segmentation tasks, aiming to mitigate vanishing gradients or learning rate decay, ultimately making the model suitable for real-world multimodal tasks and AI-driven clinical decision-assisted diagnosis.

References.

1. K. Yamazaki, V.-K. Vo-Ho, D. Bulsara, and N. Le, “Spiking Neural Networks and Their Applications: A Review,” *Brain Sci.*, vol. 12, no. 7, p. 863, Jun. 2022, doi: 10.3390/brainsci12070863.
2. M. Yao *et al.*, “Spike-driven Transformer V2: Meta Spiking Neural Network Architecture Inspiring the Design of Next-generation Neuromorphic Chips,” Feb. 2024, doi: 2404.03663v1.
3. J. Yang *et al.*, “MedMNIST v2 - A large-scale lightweight benchmark for 2D and 3D biomedical image classification,” *Sci. Data*, vol. 10, no. 1, p. 41, Jan. 2023, doi: 10.1038/s41597-022-01721-8.