Gesture Recognition with Brain-inspired Spiking Neural Networks

Tang Ching Wa 21035605D

***Abstract* — The goal of the project is to create a gesture recognition system using two brain-inspired technologies: Dynamic Vision Sensors (DVS) and Spiking Neural Networks (SNN). With the help of these two technologies, an efficient and effective gesture recognition system can be created.**

Keywords* — *Gesture Recognition, Deep learning, Dynamic Vision Sensors, Spiking Neural Networks

I. Introduction

DVS camera, also known as an event camera, captures human gestures with high dynamic range, low redundancy, and low power consumption. It detects brightness changes in pixels and generates events when the brightness exceeds a threshold. This event-driven approach is suitable for high-speed motion recording and reduces energy consumption. [1] The captured data from the DVS will be processed using SNNs, which are inspired by biological neurons in the human brain. SNNs follow an event-driven computational paradigm and can handle discrete, time-varying spikes of activity. They are well-suited for processing neuromorphic datasets and have potential in sensory processing, time-series analysis, and pattern recognition tasks. [2]

II. design/methodology/implementation

A. *Leaky Integrate-and-Fire Neuron Models*

The Leaky Integrate-and-Fire (LIF) neuron model is a simplified mathematical description of the behavior of a neuron in a neural network, particularly in the context of spiking neural networks (SNNs). It is a commonly used model due to its simplicity and ability to capture essential aspects of neuronal dynamics. [2]

B. *Spatio-temporal backpropagation (STBP) algorithm*

Spatio-temporal backpropagation refers to the process of propagating gradients or errors through a spatio-temporal domain in a neural network during the training phase. Since SNNs are non-differentiable, the STBP algorithm solves these problems by introducing approximate derivatives of spike activity for gradient descent training. It combines spatial and temporal domains in a layer-by-layer manner without requiring additional complex techniques. [3]

C. *Surrogate Gradient (SG)*

The surrogate gradient method addresses the non-differentiable problem by introducing a continuous relaxation of the non-smooth spiking nonlinearity. It approximates the true gradients, allowing for gradient computation and enabling backpropagation. [4][5]

D. *Residual Network (ResNet)*

Residual Networks (ResNets) are deep convolutional neural network architectures designed to solve the problem of vanishing gradients in deep networks. They introduce shortcut connections that allow layers to learn residual mappings, making optimization easier and improving accuracy compared to plain networks. [6]

III. Evaluauation and Results

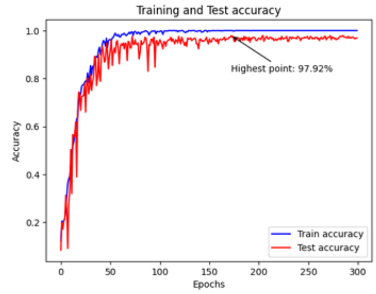

Figure 1. Model training and test accuracy.

Figure 1 demonstrates that the model achieves a high-test accuracy of approximately 97% for the last 100 epochs. The highest test accuracy recorded during training is 97.92%. These results indicate the strong generalization ability of the model, as it consistently provides accurate predictions on both the training and new, unseen data.

IV. conclusion

The project "Gesture Recognition with Brain-inspired Spiking Neural Networks" aims to develop an accurate and efficient Spiking Neural Network (SNN) model for gesture recognition using the DVS128 gesture dataset. Initially, a plain sequential model was created as a baseline, but as the project progressed, more complex architectures were incorporated, starting with the VGG architecture. The VGG architecture improved performance, achieving 93% accuracy. Later, the model was upgraded to use the ResNet architecture, which further improved accuracy. Finally, The SNN model achieved a rank of second place with 97.92% accuracy in the DVS128 gesture benchmark.

References

[1] Gallego, Guillermo, et al. "Event-based vision: A survey." IEEE transactions on pattern analysis and machine intelligence 44.1 (2020): 154-180.

[2] K. Roy, A. Jaiswal, and P. Panda, "Towards spike-based machine intelligence with

[3] Wu, Y., Deng, L., Li, G., Zhu, J., & Shi, L. (2018). Spatio-temporal backpropagation for training high-performance spiking neural networks. Frontiers in neuroscience, 12, 331.

[4] Jason K. Eshraghian, Max Ward, Emre Neftci, Xinxin Wang, Gregor Lenz, Girish Dwivedi, Mohammed Bennamoun, Doo Seok Jeong, and Wei D. Lu. "Training Spiking Neural Networks Using Lessons From Deep Learning". arXiv preprint arXiv:2109.12894, September 2021.

[5] E. O. Neftci, H. Mostafa, and F. Zenke, "Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks,"

[6] He, K., Zhang, X., Ren, S., & Sun, J. (2015, December 10). Deep residual learning for image recognition. arXiv.org. https://arxiv.org/abs/1512.03385