MANCHESTER
1824
The University of Manchester

Image Processing with brain like neural networks

A dissertation submitted to The University of Manchester for the degree of
Master of Science in Advanced Computer Science
in the Faculty of Science and Engineering

Year of submission
2023

Student ID
11144277

School of Engineering

Contents

Word count: 1000

1 Introduction

The human brain still remains a realm away from being stimulated entirely in the real world scenario with the technological advancements of today[1] but nonetheless, there is a steady rise in progress to achieve this goal. Our goal in this project is to stimulate a brain-like image-processing model. There are many neurobiological research findings that show that human visual processing can perform a recognition task in less than 10 ms processing time for each neuron[2].

Due to this many Neurocomputing methods have been developed each reigning supreme in a certain field, but none have managed to rise above the rest like deep learning techniques in machine learning. Due to this, our study focuses on the branch of Spiking Neural Networks (SNN), which can be attributed to being modelled after Deep Learning Neural Networks.

In comparison to other kinds of artificial neural networks, SNNs are modelled more closely after how the human brain functions. SNNs use binary spike signals instead of conventional artificial neural networks' continuous-valued activation functions to transmit information between neurons. As spikes are the basic unit of communication in the brain, this increases their biological plausibility.

Because of their ability to encode and process information via sparse and event-driven spikes, SNNs can conduct complex computations with high efficiency. Furthermore, SNNs can learn via various biologically plausible methods, including spike-timing-dependent plasticity (STDP)[3], which allows the network to modify the strength of its connections based on the precise timing of spikes. Homeostasis plasticity is another method which can help in achieving the same results as its role is to maintain the overall stability and balance of the neural networks. It involves adjusting the strength of the synapses at a level which is maintainable[4].

SNNs were motivated by the discovery that neurons in the brain interact with one another through discrete spikes, which are short, high-amplitude electrical events. Each neuron in SNNs integrates input spikes over time and generates output spikes when its membrane potential crosses a threshold. The accurate timing of spikes in neuronal activity is critical in SNNs because the brain processes information depending on the precise timing of spikes in neuronal activity.

While SNNs are still in the early stages of development and have not yet reached the same level of success as deep learning approaches in some domains, they show promise for tasks like event detection, sensory processing, and neuromorphic computing.

A human brain learns much more from its visual cortex than its hearing[5] and to stimulate that in a real-world environment we have gone forward with the project to showcase that[6]. SNNs are a paradigm which has much more capability in modelling complex information processing in the brain.

In retrospect to classical ANN, SNN's model needs to have their inputs encoded in a certain manner to keep up with the perspective that the brain needs to be stimulated at the same time the spatial and temporal information of the data needs to be conserved which makes it a tricky and a debatable topic. In many studies spike rates[7] have been used to encode the data but there are many findings that the high-speed processing done by our brain cannot be achieved by a rate coding scheme alone

The working of a spiking neural network is based on many of the established criteria in machine learning and deep learning. There are many facets which still remain unexplored in the working of the brain or how it has to be implemented in real word stimulation.

2 Background

Synaptic plasticity refers to the ability of the synapses for learning and memory by changing its strength over time. These changes in strength are in response to various stimuli ,including changes in the frequency or timing of neuronal activity. Though the exact relationships in the overall scenario remains controverisal[8].

The Rate encoding [9] is a neural encoding scheme where the information is carried in the number of spikes. The number of spikes is either averaged over a given duration or over several repeated experiments with the same input. Although this encoding scheme may be biologically plausible, sometimes there is just not enough time to wait for an average number of spikes to generate an appropriate response. An example described in [10] is that a fly can react to new stimuli from the environment and change the direction of flight in just 30 to 40 milliseconds. This reaction is often a response to only a few spikes. There is not enough time to average this input over several runs of the same stimulation or waiting a period of time to average the number of spikes.

The Temporal encoding (also called Latency encoding)[10] is a scheme where the information is carried in the timing of spikes. Fewer spikes are used than with the rate encoding. The information can be carried in the timing of a single spike . In the case of a periodic input stimulation, the spike can be fired when a phase change happens.

Both the encoding schemes can be applied either to a single neuron or a group of neurons [9], [10]. For example, in the case of rate encoding, the average number of spikes can be calculated from a group of neurons. Similarly, when the temporal encoding is used, some patterns in firing times of a group of neurons can be observed based on the stimulation.

Leaky Integrate and Fire (LIF) neuron model[11] is based neuron models which can reproduce electrophysical signals to a minute degree of accuracy but it becomes expensive and computationally complex. Due to this simple spiking neuron models with low computational cost are highly popular for studies in regards to the components of this matter. The LIF is a one dimensional spiking neuron model with low costs and it is defined by the following equation See (1) for info.

$$_M(\frac{dv_m(t)}{dt}) = -(v_m(t) - E_r) + R_m I(t) \quad (1)$$

where $v_m(t)$ is the membrane potential , $_M$ is the membrane time constant, E_r is the membrane rest potential which is a constant , R_m is the membrane resistance and I(t) is the sum of the current supplied by the following equation(2) for info.

$$I_t = W.S_t \quad (2)$$

When $v_m(t)$ reaches the threshold level V_{th} , an output spike is generated and the membrane voltage is reset to the rest potential for a period of time which is called refractory period

ANN is an abstraction and simulation of the structure and function of structure and function [12] of biological systems which is widely used in pattern recognition. However, neurocomputing science explores the mathematical and physical basis of biological systems, which reveals the nonprogrammed and adaptive information processing capabilities of neural systems. Compared with biological systems traditional ANN has its inherent limitation,which are mainly shown in two aspects (1) The ANN neuron model is too simple to stimulate the changes of biological neuron potential and its spiking firing process. (2) The ANN model does not use the time information of a single spike ubternally but process analog quantities. To solve these problems, SNN with more biological authencity is proposed to solve intelligent application problems in engineering. SNN has more realistic modelling and analysis of biological systsms based on its unique information coding and processing methods, further deepning the understanding of the basic function of the nervous systems and synaptic plasticity[13].

At present, ANN-SNN is the most effective way to realize the deep SNN. The SNN obtained by conversion also shows excellent performance on complex image processing tasks [14]. In terms of the conversion method, [15] customized an ANN architecture based on ReLu without bias. However, the disadvantage of their algorithm is that it is necessary to set the spiking neuron threshold for each layer of the network. Based on this, [16] realized the automatic determination of the threshold, which is a promotion of [16]. The automatic determination of the threshold depends on the model-based normalization and data-based normalization methods. It can be considered that Diehl's work [16] is the cornerstone of the ANN-SNN method. To facilitate the conversion, most of the previous work [16] has made certain constraints on the original ANN, such as using the ReLu activation function for neurons, bias ,0, and using average pooling instead of maximum pooling. Even so, the performance of SNN suffers a loss of precision compared with the original model. The methods of scaling synapses [17], adding noise and constraining synaptic discharge rate [18] are used to reduce the precision loss of model conversion, but these technologies also complicate the conversion process.

Most of the existing SNN algorithm models [15], [17]–[19] use the hard reset mechanism. Under this mechanism, no matter how much the membrane potential of the neuron exceeds the threshold, the membrane potential will be reset to 0 after it emits a spike. The hard reset mechanism causes the loss of membrane potential, which is also a major source of deep model conversion errors. Therefore,[20] proposed residual membrane potential neurons, which solved the problem of membrane potential loss and improved the spike firing rate. However, we found that the soft reset mechanism has the problem of spiking neuron over-activation. When the input of the sample is large, the membrane potential of the neuron will be many times as the threshold. After the neuron performs the soft reset, its membrane potential still exceeds the threshold. This problem could seriously affect the network's ability to distinguish the input.

3 Research Methodology

To facilitate the milestones set for our studies we have taken into consideration the case study of both basic ANN and SNN models and more complex ANN models ie enhanced and established CNN model and combination of both ANN-SNN models in a way that complements each other and embraces the rules and procedures of both thus enhancing the result produced by such a model. In the end, we have also decided to study the claim of how fast recognition and classification an SNN model can perform and does it hold a candle to the level of a human brain.

For our case study between ANN and SNN models, we have set some basic parameters where the results are to be compared ie Accuracy, model complexity, parameters, energy consumption and time taken for the results to be achieved. Another parameter here is the database, but this one is a bit flexible as image datasets can achieve good results with both models with many variations[12], [19], [21], [22]. For SNN models Event driven datasets known as Dynamic vision Sensor (DVS) are more inclined towards its nature and performs better with it than with normal image datasets like N-MNIST, CIFAR10-DVS [23]. For now, we are only utilizing MNIST dataset for all those parameters we have set.

In the current SNN model, we have utilized Dog-Filter as the first layer followed by three Convoltion2d layer altenating with MaxPool Layer. The activation layers used with those conv2d layer is relu. After that the inputs are passed from the MaxPool layer to a Rate coding layer where the inputs are encoded as deemed necessary according to spike generation methodology and then passed to Leaky Integrate Fire(LIF) node, the inputs form those are passed to a linear node followed by a LIF node and then passed to another Linear node and finally the output is predicted by a softmax layer[19], [24] . Another variation tested in this environment is a Support vector Machine used for the output. The CNN model used is almost the same but without the rate coding layer and LIF layer and both the models have produced results which have negligible difference and the only difference lies in the parameters the neural network had taken in for training.

We plan on using the best SNN model that we form from the above case study to use in our gaze detection model where we would be using perpetual motion of the eyes to point out direction in which the person in question is looking towards. Through this we are also trying to establish how good a SNN model can perform in retrospect to a human as we are trying to imitate a human brain in real-world stimulation.

The datasets to be used for gaze detection purposes are the NVIDIA Gaze dataset[25] , Labeled Pupil in the wild dataset[26], , Dr(eye)ve Dataset [27]. A common feature among all these datasets is the orientation of the pupils in the eye and the co-ordinates of the pupil in the eye. The Dr(eye)ve dataset differs from the other two as it's data is for the purpose of detecting where the drivers eyes are looking towards while driving and it also gives us the general idea of the orientation of drivers head while performing various actions during driving. This dataset would serve the purpose of showcasing the models learning from the other two datasets and help us in gaining a deeper understanding of how the models different parameters differ from the original results gained by earlier results through deep learning and other techniques and also in finding out the gaps if any in our model.

4 Ethical and Professional Consideration

The datasets and models availed and studied in our research are public while some of the datasets belong to organizations, which can be availed to the public given proper reason to use those datasets and don't harm the organization in any way. The purpose of the datasets used in our study is to analyze how a gaze detection model can be built with SNN and how the architecture's performance and other criteria differ from the more famous deep learning models. At the same time, we have made sure to properly use the concepts, observations, and libraries at an appropriate level.

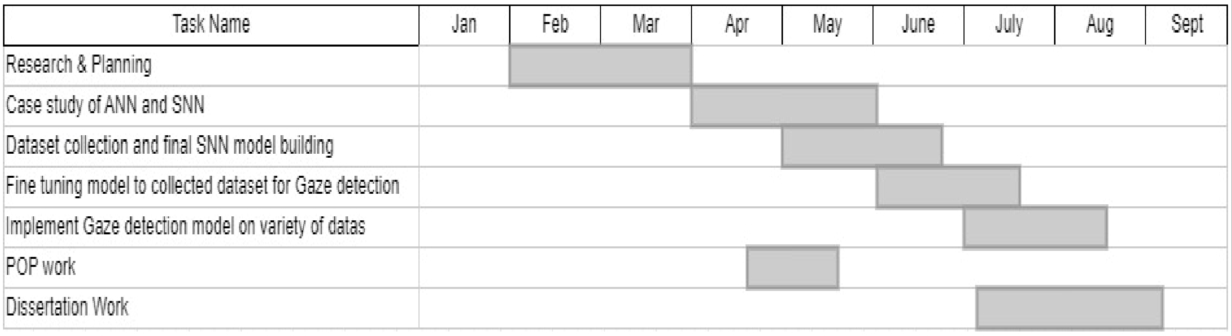
5 Risk Consideration

Risks come in various levels and aspects and most of the risk to our study and purpose lies in the technical aspect. There are risks in regard to the dataset, the architecture of the model, the various components used in the architecture which do not conform to the base of our study, the computing resources required to compute the parameters obtained through our architecture and mitigating the circumstances that would push our study towards the boundary of unachievable.

6 Project Evaluation

For project evaluation, our main aim is towards the parameters we have described before so that a comprehensive case study can be formed between ANN and SNN models. For our Gaze detection model, there are many more parameters like the computational costs which may go beyond the ideal situation, the success criteria for the same would be how accurate would our model's sensitivity to detect the precise direction in which the eyes are looking and also how the variation in co-ordinates from each source of data would affect the model.

7 Planning



References

[1] F. Dehais, W. Karwowski, and H. Ayaz, *Brain at work and in everyday life as the next frontier: Grand field challenges for neuroergonomics*, 2020 (cited on p. 3).

[2] R. Van Rullen and S. J. Thorpe, "Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex," *Neural computation*, vol. 13, no. 6, pp. 1255–1283, 2001 (cited on p. 3).

[3] J. Sjöström and W. Gerstner, "Spike-timing dependent plasticity," *Scholarpedia*, vol. 5, no. 2, p. 1362, 2010, revision #184913. DOI: 10.4249/scholarpedia.
1362 (cited on p. 3).

[4] G. G. Turrigiano and S. B. Nelson, "Homeostatic plasticity in the developing nervous system," *Nature reviews neuroscience*, vol. 5, no. 2, pp. 97–107, 2004 (cited on p. 3).

[5] N. Kanwisher and E. Wojciulik, "Visual attention: Insights from brain imaging," *Nature reviews neuroscience*, vol. 1, no. 2, pp. 91–100, 2000 (cited on p. 3).

[6] A. Taherkhani, A. Belatreche, Y. Li, G. Cosma, L. P. Maguire, and T. M. McGinnity, "A review of learning in biologically plausible spiking neural networks," *Neural Networks*, vol. 122, pp. 253–272, 2020 (cited on p. 3).

[7] T. Masquelier and G. Deco, "Learning and coding in neural networks," *Principles of neural coding*, pp. 513–526, 2013 (cited on p. 4).

[8] A. Morrison, M. Diesmann, and W. Gerstner, "Phenomenological models of synaptic plasticity based on spike timing," *Biological cybernetics*, vol. 98, pp. 459–478, 2008 (cited on p. 4).

[9] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014 (cited on p. 4).

[10] J. K. Eshraghian, M. Ward, E. Neftci, *et al.*, "Training spiking neural networks using lessons from deep learning," *arXiv preprint arXiv:2109.12894*, 2021 (cited on p. 4).

[11] W. Gerstner and W. M. Kistler, *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press, 2002 (cited on p. 4).

[12] J. H. Lee, T. Delbruck, and M. Pfeiffer, "Training deep spiking neural networks using backpropagation," *Frontiers in neuroscience*, vol. 10, p. 508, 2016 (cited on pp. 5, 6).

[13] I. Boussaïd, P. Siarry, and M. Ahmed-Nacer, "A survey on search-based modeldriven engineering," *Automated Software Engineering*, vol. 24, pp. 233–294, 2017 (cited on p. 5).

[14] L. Zhang, S. Zhou, T. Zhi, Z. Du, and Y. Chen, "Tdsnn: From deep neural networks to deep spike neural networks with temporal-coding," vol. 33, no. 01, pp. 1319–1326, 2019 (cited on p. 5).

[15] Y. Cao, Y. Chen, and D. Khosla, "Spiking deep convolutional neural networks for energy-efficient object recognition," *International Journal of Computer Vision*, vol. 113, pp. 54–66, 2015 (cited on pp. 5, 6).

[16] P. U. Diehl, D. Neil, J. Binas, M. Cook, S.-C. Liu, and M. Pfeiffer, "Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing," pp. 1–8, 2015 (cited on p. 5).

[17] Y. Kim and P. Panda, "Revisiting batch normalization for training low-latency deep spiking neural networks from scratch," *Frontiers in neuroscience*, p. 1638, 2021 (cited on pp. 5, 6).

[18] A. Sengupta, Y. Ye, R. Wang, C. Liu, and K. Roy, "Going deeper in spiking neural networks: Vgg and residual architectures," *Frontiers in neuroscience*, vol. 13,
p. 95, 2019 (cited on pp. 5, 6).

[19] S. R. Kheradpisheh, M. Ganjtabesh, S. J. Thorpe, and T. Masquelier, “Stdpbased spiking deep convolutional neural networks for object recognition,” *Neural Networks*, vol. 99, pp. 56–67, 2018 (cited on p. 6).

[20] B. Han, G. Srinivasan, and K. Roy, “Rmp-snn: Residual membrane potential neuron for enabling deeper high-accuracy and low-latency spiking neural network,” pp. 13 558–13 567, 2020 (cited on p. 6).

[21] S. Zhou, Y. Chen, X. Li, and A. Sanyal, “Deep scnn-based real-time object detection for self-driving vehicles using lidar temporal data,” *IEEE Access*, vol. 8, pp. 76 903–76 912, 2020 (cited on p. 6).

[22] L.-Y. Niu, Y. Wei, and Y. Liu, “Event-driven spiking neural network based on membrane potential modulation for remote sensing image classification,” *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106 322, 2023 (cited on p. 6).

[23] E. Sadovsky, R. Jarina, and R. Orjesek, “Image recognition using spiking neural networks,” in *2021 31st International Conference Radioelektronika (RADIOELEKTRONIKA)*, IEEE, 2021, pp. 1–5 (cited on p. 6).

[24] Q. Xu, J. Peng, J. Shen, H. Tang, and G. Pan, “Deep covdensesnn: A hierarchical event-driven dynamic framework with spiking neurons in noisy environment,” *Neural Networks*, vol. 121, pp. 512–519, 2020 (cited on p. 6).

[25] J. Kim, M. Stengel, A. Majercik, *et al.*, “Nvgaze: An anatomically-informed dataset for low-latency, near-eye gaze estimation,” in *Proceedings of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1–12 (cited on p. 7).

[26] M. Tonsen, X. Zhang, Y. Sugano, and A. Bulling, “Labelled pupils in the wild:

A dataset for studying pupil detection in unconstrained environments," in *Proceedings of the ninth biennial ACM symposium on eye tracking research & applications*, 2016, pp. 139–142 (cited on p. 7).

[27] A. Palazzi, D. Abati, F. Solera, R. Cucchiara, *et al.*, "Predicting the driver's focus of attention: The dr (eye) ve project," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 7, pp. 1720–1733, 2018 (cited on p. 7).