Benchmarking Spike-Based Visual Recognition

Qian Liu

I. Introduction

To gain a better understanding of the brain and build biologically-inspired computers, increasing attention is being paid to research into spike-based neural computation. Within the field, the visual pathway and its hierarchical organisation have been extensively studied within the primate brain. Spiking Neural Networks (SNNs) inspired by the understanding of observed biological structure and function have been successfully applied to visual recognition/classification tasks. In addition, implementations on neuromorhpic hardware have made large-scale networks run in (or even faster than) real time, and accessible on mobile robots. Neuromorphic sensors, e.g. silicon retinas, are able to feed such a mobile system with real-time visual stimuli. A new series of vision benchmarks for spike-based neural processing are now needed to quantitatively measure progress within this rapidly advancing field. We propose that a large dataset of spike-based visual stimuli is needed to provide a baseline for comparisons on SNN models and algorithms, and some benchmarking network models are also required to validate the accuracy and cost of these neuromorphic hardware platforms.

First of all, an initial NE (Neuromorphic Engineering) dataset of input stimuli based on standard computer vision benchmarks consisting of digits (from the MNIST database) is presented according to the current research on spike-based image recognition. Within this dataset, all images are centre aligned and having similar scale. We describe how we intend to expand this dataset to fulfil the needs of upcoming research problems. For instance, the data should provide cases to measure position-, scale-, and viewing-angle invariance. The data are in Address-Event Representation (AER) format which is widely used in the neuromorphic engineering field unlike conventional images. These spike trains are produced by various techniques: rate-based Poisson spike generation, rank order encoding and recorded output from a silicon retina with both flashing and oscillating input stimuli. Furthermore a complementary evaluation methodology is also presented to assess both model-level and hardware-level performance.

With this dataset we hope to (1) promote meaningful comparison between algorithms in the field of neural computation, (2) allow comparison with conventional image recognition methods, (3) provide an assessment of the state of the art in spike-based visual recognition, and (4) help researchers identify future directions and advance the field.

Finally, we provide two SNN models to validate their classification capabilities and to assess the performances of their hardware implementations as tentative benchmarks. Future work will focus on proposing equivalent ANN deep learning models on SNN, thus to supplement the comparison with the state-of-the-art of image recognition. One of the main research is to apply biologically-inspired learning algorithms to the Deep Belief Networks (DBN). It requires fully understanding of the probabilistic representation of the Restricted Boltzmann Machine (RBM) and DBN, as well as the pre-training and fine-training process. How to transfer the learning algorithm to spike-based learning will be the major problem to solve. The learning will benefit from exploiting neuromorphic hardware platforms to train a large-scale DBN model in real time. The comparison on time and energy cost of DBN learning between conventional ANN and event-based SNN will be of interest in the neuromorphic community. The other benchmarking model is the Convolutional Network (ConvNet). Some initial experiment [1] has validated the feasibility of real-time hardware simulation of a large-scale convolutional model. The optional research may continue on benchmarking ConvNets using the database we proposed.

II. The Dataset: NE15-MNIST

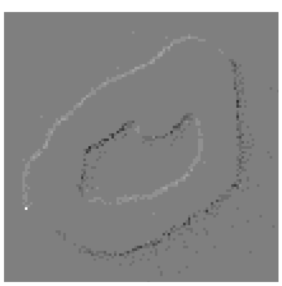
The name of the first proposed dataset in the benchmarking system is NE15-MNIST which stands for Neuromorphic Engineering 2015 on MNIST. The original MNIST dataset is downloaded from the website[1] of THE MNIST DATABASE of handwritten digits [2]. The NE15-MNIST is converted into a spiking version of the original dataset consisting of four subsets which were generated for different purposes:

- *Poissonian* to benchmarking existing methods of rate-based spiking models.
- *FoCal* to promote the study of spatio-temporal algorithms applied to recognition tasks using few input spikes.

[1] http://yann.lecun.com/exdb/mnist/

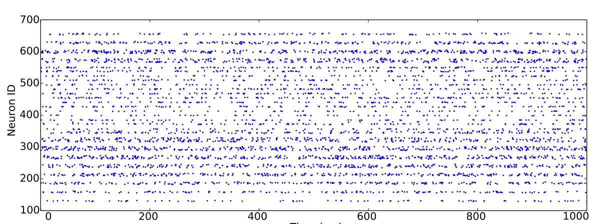
- *DVS recorded flashing input* to encourage research on fast recognition methods which are found in the primate visual pathway.
- *DVS recorded moving input* to trigger the study of algorithms targeting on continuous input from real-world sensors and to implement them on mobile neuromorphic robots.

The dataset can be found in the GitHub repository at: https://github.com/qian-liu/benchmarking. Two file formats are supported in the dataset: jAER format [3] (.dat or .aedat), and binary file in NumPy .npy format. The address event representation (AER) interface has been widely used in neuromorphic systems, especially for vision sensors. The spikes are encoded as time events with corresponding addresses to convey information. The spikes in jAER format, both recorded from a DVS retina and artificially generated, can be displayed in jAER software. Figure 1a is a snapshot of the software displaying an .aedat file which is recorded by a DVS retina [4]. The resolution of the DVS recorded data is 128×128. As an example of the data of the dataset, a Poissonian representation of the same image is shown in Figure 1b and its raster plot in Figure 1c. The other format of spikes used is a list of spike source arrays in PyNN [5], a description language for building spiking neuronal network models. Python code for converting one file format to and from the other is also provided.



(a) A snapshot of jAER playing the DVS recorded spikes.

(b) A snapshot of jAER playing Poissonian spike trains.



(c) The raster plot of the Poissonian spike trains.

Figure 1: Snapshots of jAER software playing spike presented videos. The same image of digit “0” is transformed to spikes by DVS recording and the Poissonian generation respectively. A raster plot of the Poissonian spike trains is also provided.

III. Evaluations on Spike-Based Recognition

The complementary evaluation methodology is essential to assess both the model-level and hardware-level performances. For a network model, its topology, neuron and synapse models, and training methods are major descriptions for any kind of neural networks, including SNNs. While the recognition accuracy, network latency and also the biological time taken for both training and testing are specific performance measurements of a spike-based model. To build any SNN model on a hardware platform, its network size will be constrained by the scalability of the hardware. Neural and synaptic models are limited to the ones that are physically implemented, unless the hardware platform supports programmability. The accuracy of the results (e.g. CA) are naturally affected by the precision of the variable representing the membrane potential and synaptic weights. Any attempt to implement an on-line learning algorithm on neuromorphic hardware must be backed by synaptic plasticity support. Running an identical SNN model on different neuromorphic hardware platforms can not only expose if any of the previously mentioned capacities are supported, but also benchmark their performance on simulation time and energy usage.

The test results of the case studies as examples using the dataset are listed in Table I and evaluated on the proposed metrics. While the hardware evaluation metrics is listed in Talbe II

Table I: Hardware independent comparison

	Pre	Network	Training	Recognition
Exp 1	None	FC Network, LIF neurons	K-means clusters, Supervised STDP 18,000 s of training	92.98% 1 s per test 10.70 ms latency
Exp 2	None	4-layer RBM, LIF neurons	Off-line trained, unsupervised	94.94% 16 ms latency

IV. Case Studies

We present two recognition SNN models working on the Poissonian subset of the NE15-MNIST dataset. Their network components, training and testing methods are described in the paper [11] (under corrections of the reviews) according to the evaluation methodology stated above. The specific spike-based evaluations on input event rates and/or responding latency are also provided. Meanwhile, as tentative benchmarks the models are implemented on SpiNNaker to assess the performance against software simulators. Presenting proper

Table II: Hardware dependent comparison

	System	Neuron Model	Synaptic Plasticity	Precision	Simulation Time	Energy/Power Usage
SpiNNaker [6]	Digital, Scalable	Programmable Neuron/Synapse, Axonal delay	Programmable learning rule	11- to 14-bit synapses	Real-time Flexible time resolution	8 nJ/SE 54.27 MSops/W
TrueNorth [7]	Digital, Scalable	Fixed models, Config params, Axonal delay	No plasticity	122 bits params & states, 4-bit synapse (4 signed int + on/off state)	Real-time	46 GSops/W
Neurogrid [8]	Mixed-mode, Scalable	Fixed models, Config params	Fixed rule	13-bit shared synapses	Real-time	941 pJ/SE
HI-CANN [9]	Mixed-mode, Scalable	Fixed models, Config params	Fixed rule	4-bit synapses	Faster than real-time	198 pJ/SE 13.5 MSops/W (network only)
iAER-IFAT [10]	Mixed-mode, Scalable	Fixed models, Config params	No plasticity	Analogue neuron/synapse	Real-time	20GSops/W

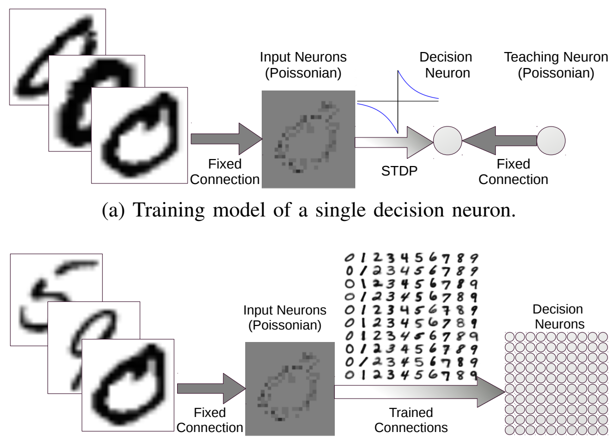
benchmarks for vision recognition systems is still under investigation, the case studies only make first attempt.

A. 2-Layer STDP

The case study is a simple two-layered network where the input neurons receive Poissonian presented spike trains from the dataset and form an FC network with the decision neurons. The model utilises LIF neurons, and the parameters are all with biological means.

In order to fully assess the performance, different settings have been configured on the network, such as network size, input rate and testing images duration. There are two layers in the model: 28×28 input neurons fully connect to 100 decision neurons. Each decision neuron responds to a certain template of a digit. In the standard configuration, there are 10 decision neurons answering to the same digit with slightly different templates, e.g. altogether 10×10 decision neurons. Those templates are embedded in the connection weights between the two layers. Fig. 2a shows how the connections to a single decision neuron are tuned.

The training set of 60,000 hand written digits are firstly classified into 100 classes, 10 subclasses per digit, using K-means clusters. So the images in a certain subclass are used to train one corresponding decision neuron. The firing rates of the input neurons are assigned linearly according to their intensities and normalised with a total firing rate of 2,000 Hz. All the images together are presented for 18,000 s (about 300 ms per image) during training and at the same time a teaching signal of 50 Hz is conveyed to the decision neuron to trigger STDP learning. The trained weights are plotted in accordance with the positions of the decision neurons in Fig. 2b.



(a) Training model of a single decision neuron.

(b) Testing model of the spiking neural network.

Figure 2: The training and testing model of the two-layered spiking neural network.

After training the weights of the plastic synapses are set to static, keeping the state of the weights at the last moment of training. The feed-forward testing network is shown in Fig. 2b where Poissonion spike trains are generated the same way as in the training with a total firing rate of 2,000 Hz per image. The input neurons convey the same spike trains to every decision neuron through its responding trained synaptic weights. Every testing image (10,000 images in total) is presented once and lasts 1 s with a silence of 200 ms between them. The output neuron with the highest firing rate decides what digit was recognized.

Among different network configurations, the network of 500 decision neurons (50 clusters per digit) reaches the highest recognition rate. The network achieved a CA of 92.98% and average latency of 10.70 ms, and the simulation costs SpiNNaker 0.41 W on power use and 4,920 J on energy use 3.

Table III: Comparisons of NEST (N) on a PC and SpiNNaker (S) performances.

Clusters per digit	Accuracy (%) N	Accuracy (%) S	Simulation (s) N	Simulation (s) S	Power Use (W) N	Power Use (W) S
1	79.62	79.50	554.77			0.38
10	91.29	91.43	621.74			0.38
50	92.98	92.92	1,125.12	12,000	21.0	0.41
100	87.27	86.83	1,406.01			0.44
1000	89.65	89.74	30,316.88			1.50

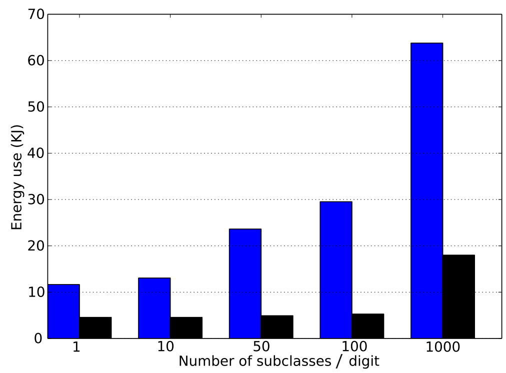


Figure 3: Energy usages of different network size both using NEST (blue) on a PC and SpiNNaker (black).

B. *Case Study II*

Paper [12] proposed a method to map off-line trained DBNs into a spiking neural networks and take advantage of the real-time performance and energy efficiency of neuromorphic platforms. For this work we used an off-line trained spiking DBN with a 784-500-500-10 network topology. Simulations take place on a software spiking neural network simulator named Brian [13] and results are verified on the SpiNNaker platform.

A similar experiment to the one presented for the Case Study I was performed; its purpose was to establish the relation that input spike rates hold with latency and classification accuracy. The input rates were varied from 500 Hz to 2,000 Hz and the results are summarised in Figure 4. Simulations ran in Brian for all 10,000 MNIST digits of the testing set and for 4 trials. The mean classification latency for the particular spiking DBN on SpiNNaker is 16 ms which is identical to the Brian simulation seen in Figure 4.

V. Preliminary and Future Work

One of the main research in the future is to apply biologically-inspired learning algorithms to the SDBN

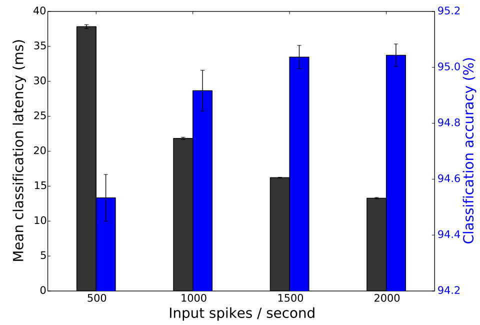


Figure 4: Mean classification latency (black) and classification accuracy (blue) as a function of the input spikes per second for the spiking DBN. Results are averaged over 4 trials, error bars show standard deviations.

thus to provide the comparison with the state-of-the-art of the deep learned image recognition. The event-based learning will benefit from exploiting neuromorphic hardware platforms to train a large-scale DBN model in real time. The comparison on time and energy cost of DBN learning between conventional ANN and event-based SNN will be of interest in the neuromorphic community. The current work starts from layer-by-layer pre-training of RBMs and the fine-training of a DBN, and future work will focus on applying spiking learning rules to DBN training, see Gantt chart Figure 5. This work is accumulatively written in my personal report [14].

Works done: (* indicates potential research)

1) Contrastive Divergence (CD)
   a) Product of Experts Problem (PoE)
   b) Markov Chain Monte Carlo Sampling
   c) Gibbs Sampling
   d) CD Instead of Kullback-Leibler (KL) *
2) RBM
   a) Objective Function
   b) CD with 1-Step Gibbs Sampling
   c) CD Validation Methods *
   d) Comparison of CD_1 and CD_k *
3) DBN
   a) The Probabilistic Model *
   b) Greedy Algorithm
   c) Fine Training
   d) Practical Training Configurations *

The future work will firstly reproduce the recursive network Neftci and etc. proposed [15] for training SRBM using event-driven contrastive divergence (eCD). It will followed by improving and formalisation of their methods and proposing new algorithms. The fine-training on pre-trained layered SRBMs in an SDBN will be studied afterwords, and SpiNNaker on-line trained models will be provided as benchmarks.

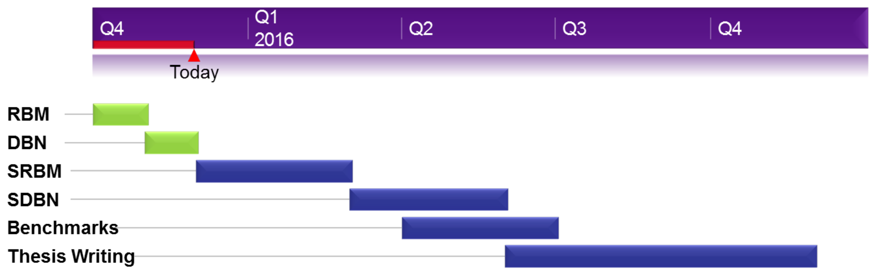


Figure 5: Gantt Chart of the future work.

References

[1] L. Camunas-Mesa, C. Zamarreño-Ramos, A. Linares-Barranco, A. J. Acosta-Jiménez, T. Serrano-Gotarredona, and B. Linares-Barranco, "An event-driven multi-kernel convolution processor module for event-driven vision sensors," *Solid-State Circuits, IEEE Journal of*, vol. 47, no. 2, pp. 504–517, 2012.

[2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[3] T. Delbruck, "Frame-free dynamic digital vision," in *Proceedings of Intl. Symp. on Secure-Life Electronics, Advanced Electronics for Quality Life and Society*, pp. 21–26, 2008.

[4] T. Serrano-Gotarredona and B. Linares-Barranco, "A 128×128 1.5% contrast sensitivity 0.9% FPN 3μs latency 4 mW asynchronous frame-free dynamic vision sensor using transimpedance preamplifiers," *Solid-State Circuits, IEEE Journal of*, vol. 48, no. 3, pp. 827–838, 2013.

[5] A. P. Davison, D. Brüderle, J. Eppler, J. Kremkow, E. Muller, D. Pecevski, L. Perrinet, and P. Yger, "PyNN: a common interface for neuronal network simulators," *Frontiers in neuroinformatics*, vol. 2, 2008.

[6] E. Stromatias, F. Galluppi, C. Patterson, and S. Furber, "Power analysis of large-scale, real-time neural networks on SpiNNaker," in *Neural Networks (IJCNN), The 2013 International Joint Conference on*, pp. 1–8, 2013.

[7] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, *et al.*, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, no. 6197, pp. 668–673, 2014.

[8] B. V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran, J.-M. Bussat, R. Alvarez-Icaza, J. V. Arthur, P. Merolla, K. Boahen, *et al.*, "Neurogrid: a mixed-analog-digital multichip system for large-scale neural simulations," *Proceedings of the IEEE*, vol. 102, no. 5, pp. 699–716, 2014.

[9] J. Schemmel, D. Bruderle, A. Grubl, M. Hock, K. Meier, and S. Millner, "A wafer-scale neuromorphic hardware system for large-scale neural modeling," in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, pp. 1947–1950, 2010.

[10] T. Yu, J. Park, S. Joshi, C. Maier, and G. Cauwenberghs, "65k-neuron Integrate-and-Fire array transceiver with address-event reconfigurable synaptic routing," in *Biomedical Circuits and Systems Conference (BioCAS), 2012 IEEE*, pp. 21–24, 2012.

[11] Q. Liu, G. Pineda-Garcia, E. Stromatias, T. Serrano-Gotarredona, and S. Furber, "Benchmarking Spike-Based Visual Recognition: A Dataset and Evaluation," *Frontiers in Neuromorphic Enginerring*, under review.

[12] P. O'Connor, D. Neil, S.-C. Liu, T. Delbruck, and M. Pfeiffer, "Real-time classification and sensor fusion with a spiking deep belief network," *Frontiers in neuroscience*, vol. 7, 2013.

[13] D. Goodman and R. Brette, "Brian: a simulator for spiking neural networks in Python," *Frontiers in neuroinformatics*, vol. 2, 2008.

[14] Q. Liu, "On-line training of spiking deep belief network," 2015.

[15] E. Neftci, S. Das, B. Pedroni, K. Kreutz-Delgado, and G. Cauwenberghs, "Event-driven contrastive divergence for spiking neuromorphic systems," *Frontiers in neuroscience*, vol. 7, 2013.

Appendix A
Thesis Outline

The following section-level outline gives the planned thesis structure for this project. Sections which are reliant on upcoming work are indicated with a star (*);

1) Introduction
2) Background
    a) Neural Network on Image Recognition
    b) Neuron Models and Spiking Neural Network
    c) Spiking Neural Network Simulation
    d) Neuromorphic Simulators
3) Related Works
    a) Vision Databases and Benchmarks
    b) Deep Neural Networks
    c) Spike-Based Image Recognition
    d) Real-Time Neuromorphic Vision System
4) Benchmarking Spike-Based Visual Recognition
    a) Database
    b) Evaluation Methodology
5) Spiking Deep Belief Network
    a) Restricted Boltzmann Machine
    b) Deep Belief Network
    c) Spiking RBM and DBN *
    d) Formalisation of SDBN *
6) Benchmarks
    a) STDP Learned 2-Layer Network
    b) Spiking DBN *
    c) Off-line Trained Convolutional Network *
7) Discussions *
    a) Benefits of Spike-based Processing *
    b) Scalability of Hardware SDBN *
    c) Future NE Database Development *
8) Future Work *
    a) SDBN Toolbox on SpiNNaker *
    b) Video-Based Recognition and Benchmarks *
    c) Natural Language Processing and Benchmarks *