Noisy Softplus: A Biology Inspired Activation Function

Qian Liu and Steve Furber

Advanced Processor Technologies group, School of Computer Science,
University of Manchester, M13 9PL, Manchester, United Kingdom
qian.liu-3@manchester.ac.uk, steve.furber@manchester.ac.uk

Abstract. The Spiking Neural Network (SNN) has not achieved the recognition/classification performance of its non-spiking competitor, the Artificial Neural Network(ANN), particularly when used in deep neural networks. The mapping of a well-trained ANN to an SNN is a hot topic in this field, especially using spiking neurons with biological characteristics. This paper proposes a new biologically-inspired activation function, Noisy Softplus, which is well-matched to the response function of LIF (Leaky Integrate-and-Fire) neurons. A convolutional network (ConvNet) was trained on the MNIST database with Noisy Softplus units and converted to an SNN while maintaining a close classification accuracy. This result demonstrates the equivalent recognition capability of the more biologically-realistic SNNs and bring biological features to the activation units in ANNs.

Keywords: Noisy Softplus, biologically-inspired, Spiking Neural Network, activation function, LIF neurons

1 Introduction

Deep Neural Networks (DNNs) are the most promising research field in computer vision, even exceeding human-level performance on image classification tasks [8]. To investigate whether brains might work similarly on vision tasks, these powerful DNN models have been converted to spiking neural networks (SNNs). In addition, the spiking DNN offers the prospect of neuromorphic systems that combine remarkable performance with energy-efficient training and operation.

Theoretical studies have shown that biologically-plausible learning, e.g. Spike-Timing-Dependent Plasticity (STDP), could approximate a stochastic version of powerful machine learning algorithms such as Contrastive Divergence [13], Markov Chain Monte Carlo [1] and Gradient Descent [14]. Stochasticity, in contrast with the continuously differentiable functions used by ANNs, is intrinsic to the event-based spiking process, making network training difficult. In practice, ANNs use neuron and synapse models very different from biological neurons, and it remains an unsolved problem to develop SNNs with equivalent performance.

On the other hand, the offline training of an ANN, which is then mapped to an SNN, has shown near loss-less conversion and state-of-the-art classification

accuracy. This research aims to prove that SNNs are equally capable as their non-spiking rivals of pattern recognition, and at the same time are more biologically realistic and energy-efficient. Jug et. al. [10] first proposed the use of the Siegert function to replace the sigmoid activation function in Restricted Boltzmann Machine (RBM) training. The Siegert units map incoming currents driven by Poisson spike trains to the response firing rate of a Leaky Integrate-and-Fire (LIF) neuron. The ratio of the spiking rate to its maximum is equivalent to the output of a sigmoid neuron. A spiking Deep Belief Network (DBN) [15] was implemented on neuromorphic hardware, SpiNNaker [6], to recognise hand written digits in real time. However, cortical neurons seldom saturate their firing rate. Thus Rectified Linear Units (ReLU) were proposed and surpassed the performance of other popular activation units thanks to their advantage of sparsity [7]. Recent work [9] proposed the Soft LIF response function, which is equivalent to Softplus activation.

Even better performance [2, 4] has been demonstrated in Spiking Convolutional Networks (ConvNets), but this employed simple integrate and fire neurons. The training used only ReLUs and zero bias to avoid negative outputs, and applied a deep learning technique, dropout, to increase the classification accuracy. Normalising the trained weights for use on an SNN was relatively straightforward and maintained the classification accuracy. This work was extended to a Recursive Neural Network (RNN) [5] and run on the TrueNorth [12] neuromorphic hardware platform.

The Noisy Softplus activation function proposed here is based on LIF neurons with biological characteristics, and is the first attempt to map a spiking neural response accurately to the activation unit of an ANN. The resulting classification accuracy was tested on a spiking ConvNet; the performance was close to that of the original ConvNet, and was better than using Softplus. This study brings a significant biological feature, noise, to the activation units of an ANN, in the hope of promoting research into noise-based computation.

2 Methods

2.1 Neural Science Background

This paper proposes a new activation function, Noisy Softplus, which is inspired by neuroscience observations of LIF neurons. The LIF neuron model follows the following membrane potential dynamics:

$$\tau_m \frac{\mathrm{d}V}{\mathrm{d}t} = V - V_{rest} + R_m I(t) \ . \tag{1}$$

The membrane potential V changes in response to the input current I, starting at the resting membrane potential V_{rest}, where the membrane time constant is $\tau_m = R_m C_m$, R_m is the membrane resistance and C_m is the membrane capacitance. The central idea in converting spiking neurons to activation units lies in the
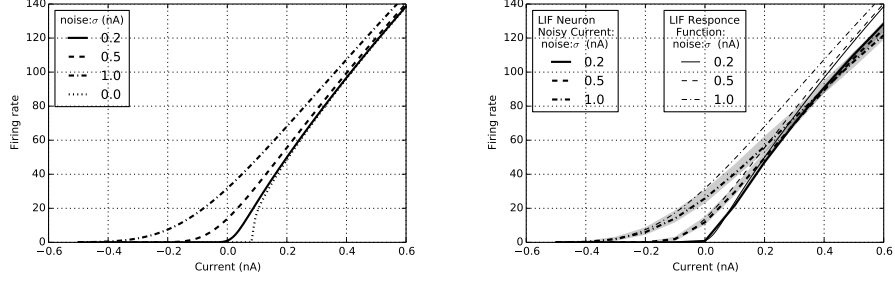
(a) Response function with noisy currents.

(b) Recorded response firing rate.

Fig. 1: (a) Response function of the LIF neuron with noisy input currents with different standard deviations. (b) Comparing the recorded firing rates of the LIF neuron simulation driven by noisy currents to the response function shown in (a).

Table 1: Parameter setting for the current-based LIF neurons using PyNN.

Parameters	cm	tau_m	tau_refrac	tau_syn_E	tau_syn_I	v_rest	v_thresh	i_offset
Values	0.25	20.0	1.0	5.0	5.0	-65.0	-50.0	0.1
Units	nF	ms	ms	ms	ms	mV	mV	nA

response function of a neuron model. Given a constant current injection I, the response function, i.e. firing rate, of the LIF neuron is:

$$\lambda_{out} = \left[t_{ref} - \tau_m \log \left(1 - \frac{V_{th} - V_{rest}}{IR_m} \right) \right]^{-1}, \text{ when } IR_m > V_{th} - V_{rest}, \quad (2)$$

otherwise the membrane potential cannot reach the threshold V_{th} and the output firing rate is zero. The absolute refractory period t_{ref} is included, where all input during this period is ignored. The dotted (zero noise) line in Fig.1(a) illustrates the response function of an LIF neuron, which inspired the proposal of ReLUs. The parameters of the LIF neuron are all biologically valid (see the listed values in Table 1), and the same parameters are used throughout this paper. In practice, a noisy current generated by the arrival of spike trains, rather than a constant current, flows into the neurons. The response function of the LIF neuron to a noisy current is as follows, where μ and σ are the mean and variance of the current:

$$\lambda_{out} = \left[t_{ref} + \tau_m \int_{\frac{V_{rest} - \mu\tau_m}{\sigma\sqrt{\tau_m}}}^{\frac{V_{th} - \mu\tau_m}{\sigma\sqrt{\tau_m}}} \sqrt{\pi} \exp(u^2)(1 + erf(u)) \mathrm{d}u \right]^{-1}. \quad (3)$$

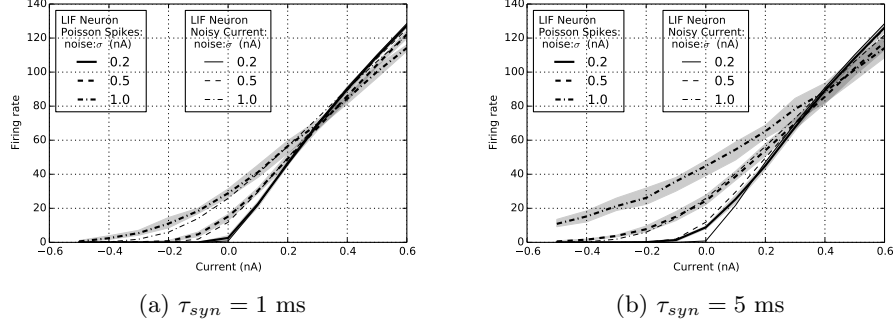(a) $\tau_{syn} = 1$ ms   (b) $\tau_{syn} = 5$ ms

Fig. 2: Recorded response firing rate of two LIF neurons with different synaptic constants. The driving noisy current is simulated with Poission spike trains and the results are compared to the noisy current source.

2.2   LIF Neuron Simulation

To verify the response function, a simulation was carried out using PyNN [3] to compare with the analytical results. A noisy current with a particular μ and σ was injected into an LIF neuron for 10 s. The firing rate was the average among 10 trials, see Fig.1(b). The slight difference compared to the analytical results (dashed lines) comes from the time resolution of the simulated noisy current. A more realistic simulation of a noisy current is generated by a Poisson spike train, where the mean and variance are given by:

$$\mu = \tau_{syn} \sum_i w_i \lambda_i \ , \ \sigma^2 = \frac{1}{2}\tau_{syn} \sum_i w_i^2 \lambda_i \ , \tag{4}$$

where τ_{syn} is the synaptic time constant, and each Poisson spike train connects to the neuron with a strength of w_i and a firing rate of λ_i. Two populations of Poisson spike sources, for excitatory and inhibitory synapses respectively, were connected to a single LIF neuron to mimic the noisy currents. The firing rates of the Poisson spike generators were determined by the given μ and σ. Fig. 2 illustrates the recorded firing rates responding to the spike trains compared to the result driven by noisy currents. The use of noisy currents assumes that the post-synaptic potential (PSP) is a delta function, e.g. τ_{syn} tends to the limits of 0. However, in practice the release of neurotransmitter takes time and the noise added to the mean current is not pure white noise. Thus the experiments show that a longer τ_{syn} increases the level of noise and widens the variance of the output firing rate.

(a) Noisy Softplus

(b) Noisy Softplus in 3D

(c) $k = 0.17$ and $\tau_{syn} = 1$ ms

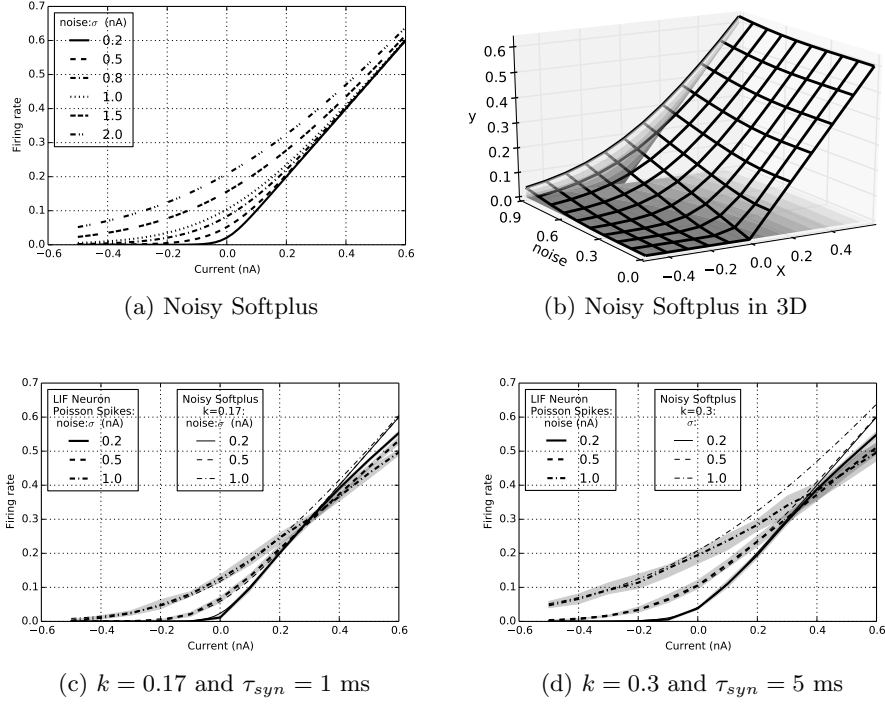(d) $k = 0.3$ and $\tau_{syn} = 5$ ms

Fig. 3: Noisy Softplus fits to the response function of the LIF neuron. Noisy Softplus in (a) curve sets and (b) 3D. (c) and (d) show how Noisy Softplus fits to the response firing rates of LIF neurons with different synaptic constants.

2.3   Noisy Softplus

Inspired by the set of response functions triggered by different levels of noise, we propose the Noisy Softplus activation function:

$$y = f_{ns}(x, \sigma) = k\sigma \log[1 + \exp(\frac{x}{k\sigma})], \tag{5}$$

where x refers to the mean current, y is the normalised output firing rate, σ plays an important role to define the noise level, and k, which is determined by the neuron parameters, controls the curve scaling. Note that the novel activation function we propose contains two parameters, the current and its noise; both are naturally obtained in spiking neurons. Fig.3(a) and 3(b) show the activation function in curve sets and in a 3D plot. The Noisy Softplus fits well to the recorded response firing rate of the LIF neuron with suitable calibration of k, see Fig.3(c) and 3(d). The derivative is the logistic function scaled by $k\sigma$:

$$\frac{\partial f_{ns}(x, \sigma)}{\partial x} = \frac{1}{1 + exp(-\frac{x}{k\sigma})}. \tag{6}$$

(a) Pixel firing rates   (b) 5x5 kernel   (c) Output firing rates

(d) 10 input digits as a raster plot   (e) Observed 'activation functions'
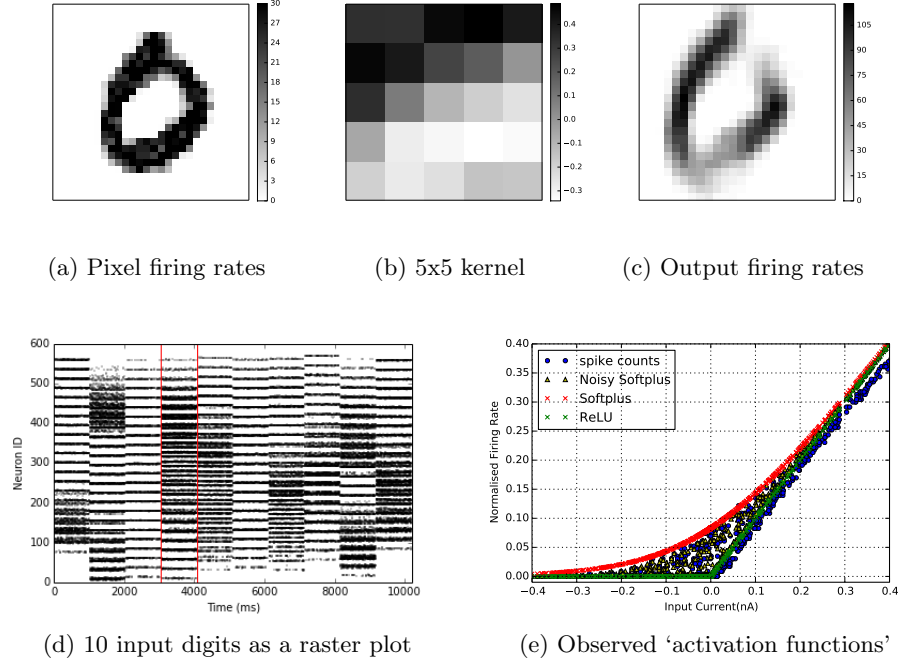
Fig. 4: Noisy Softplus fits to the neural response firing rate in an SNN simulation. The 28x28 Poisson spike trains in (a) firing rate, and (d) raster plots, are convolved with a 5x5 kernel (b). (c) the convolved map with the firing rates of each neuron. (e) the normalised firing rate compared with Noisy Softplus, Softplus and ReLU activation functions.

3   Results

A ConvNet model was trained on MNIST, a popular database in neuromorphic vision, using Noisy Softplus neurons. The architecture contains 28x28 input units, followed by two convolutional layers c5-2s-12c5-2s, and the 10 output neurons represents the classified digit. All the convolution and average sampling neurons use Noisy Softplus units with no bias, while the output neurons are softmax units converting a vector of values into the range (0, 1) that add up to 1. The weights were updated using a fixed learning rate, 50 images per batch and 10 epochs. Before testing on spiking LIF neurons, the weights of each layer were scaled to ensure that input synaptic currents stay within a valid range.

To validate how well the Noisy Softplus activation fits to the response firing rate of LIF neurons in a real application, we simulated the model on Nest using the Poisson MNIST dataset [11] and the neurons of a convolutional map were observed. Fig. 4 shows the convolution of a 5x5 kernel with an input digit '0'

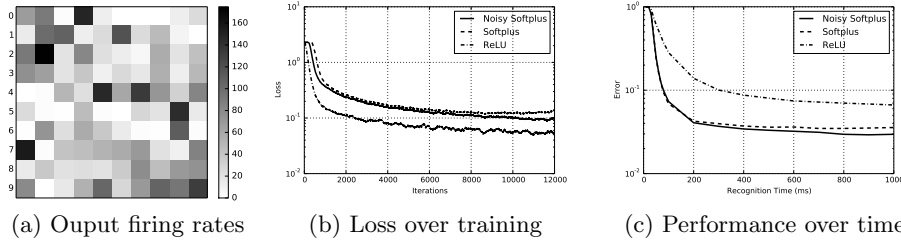(a) Ouput firing rates   (b) Loss over training   (c) Performance over time

Fig. 5: Classification performance is calculated by the firing rate of output neurons (a). (b) shows how the loss varies over training. (c) illustrates the accuracy over short response times.

represented by spike trains. The estimated spike counts using Noisy Softplus fit to the real recorded firing rate much more accurately than the Softplus and ReLU activation functions. In Fig. 4(e), we manually selected a suitable scaling factor for Softplus which located on the top slope of the response activity. However, the scale factor remains static for all the neurons, thus resulting in a mismatch at different level of noise. Noisy Softplus adapts to noise automatically.

We compared the training using ReLU, Softplus, and Noisy Softplus by their loss during training averaged over 6 trials, see Fig.5. The trained networks were scaled to SNNs and compared on recognition rates, 93.34%, 96.43% and 97.03% with a conversion loss of 4.76%, 0.91% and 0.74%. As it is a major concern in neuromorphic vision, the recognition performance over short response times is also estimated in Fig.5(c).

4   Discussion

The biologically-inspired activation function, Noisy Softplus, adapts to the noise level of input currents automatically, and is the first attempt to map activation units accurately to the firing response of LIF neurons. Noisy Softplus not only brings more biological features to the activation function, but also proves capable of performing well in a spiking ConvNet recognition task. The spiking version of Noisy Softplus wins on accuracy over the sigmoid neuron, compared to the result [15] of using Siegert units. As a result of its more accurate mapping, Noisy Softplus outperforms Softplus.

Future work on SNNs will include constraints during training to limit the function within the active range, which is equivalent to constraining the maximum firing rate of an LIF neuron. As a result there should be no need for the scaling process after training. For more accurate mapping, the scale factor k should be (numerically) derived to avoid calibration. In ANNs, it could be useful to study noise as extra information to be gathered by Softplus activation to enhance classification.

Acknowledgments

References

1. Buesing, L., Bill, J., Nessler, B., Maass, W.: Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons. PLoS Comput Biol 7(11), e1002211 (2011)
2. Cao, Y., Chen, Y., Khosla, D.: Spiking deep convolutional neural networks for energy-efficient object recognition. International Journal of Computer Vision 113(1), 54–66 (2015)
3. Davison, A.P., Brüderle, D., Eppler, J., Kremkow, J., Muller, E., Pecevski, D., Perrinet, L., Yger, P.: PyNN: a common interface for neuronal network simulators. Frontiers in neuroinformatics 2 (2008)
4. Diehl, P.U., Neil, D., Binas, J., Cook, M., Liu, S.C., Pfeiffer, M.: Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. In: Neural Networks (IJCNN), International Joint Conference on. IEEE (2015)
5. Diehl, P.U., Zarrella, G., Cassidy, A., Pedroni, B.U., Neftci, E.: Conversion of artificial recurrent neural networks to spiking neural networks for low-power neuromorphic hardware. arXiv preprint (2016)
6. Furber, S.B., Galluppi, F., Temple, S., Plana, L., et al.: The SpiNNaker Project. Proceedings of the IEEE 102(5), 652–665 (2014)
7. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: International Conference on Artificial Intelligence and Statistics. pp. 315–323 (2011)
8. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1026–1034 (2015)
9. Hunsberger, E., Eliasmith, C.: Spiking Deep Networks with LIF Neurons. arXiv preprint (2015)
10. Jug, F., Lengler, J., Krautz, C., Steger, A.: Spiking networks and their rate-based equivalents: does it make sense to use Siegert neurons? In: Swiss Soc. for Neuroscience (2012)
11. Liu, Q., Pineda-García Garibaldia nd Stromatias, E., Serrano-Gotarredona, T., Furber, S.: Benchmarking spike-based visual recognition: a dataset and evaluation. Frontiers in Neuroscience (under review) (2016)
12. Merolla, P.A., Arthur, J.V., Alvarez-Icaza, R., Cassidy, A.S., Sawada, J., Akopyan, F., Jackson, B.L., Imam, N., Guo, C., Nakamura, Y., et al.: A million spiking-neuron integrated circuit with a scalable communication network and interface. Science 345(6197), 668–673 (2014)
13. Neftci, E., Das, S., Pedroni, B., Kreutz-Delgado, K., Cauwenberghs, G.: Event-driven contrastive divergence for spiking neuromorphic systems. Frontiers in neuroscience 7 (2013)
14. O'Connor, P., Welling, M.: Deep spiking networks. arXiv preprint (2016)
15. Stromatias, E., Neil, D., Galluppi, F., Pfeiffer, M., Liu, S.C., Furber, S.: Scalable energy-efficient, low-latency implementations of trained spiking deep belief networks on SpiNNaker. In: Neural Networks (IJCNN), International Joint Conference on. IEEE (2015)