Spike-Event-Driven Deep Spiking Neural Network With Temporal Encoding

Zhixuan Zhang and Qi Liu, *Member, IEEE*

***Abstract*—Feature extraction plays an important role before pattern recognition takes place. The existing artificial neural networks (ANNs), however, ignore to learn and represent temporal information, instead of only utilizing spatial information for recognition. Moreover, the substantial computational and energy costs resulted from the conventional ANN-based classifiers, limit their uses in mobile and embedded applications. In this work, we develop a sparse temporal encoding method which exploits both spatial and temporal information. On the basis of spike-timing-dependent plasticity and multi-scale structure, the resulting temporal feature representation integrates with a temporal spiking neural network (SNN) classifier to achieve high efficiency of parallel computing for feature extraction. Experimental evaluation on four benchmark datasets from image classification and speech recognition tasks show the proposed SNN model yielding state-of-the-art accuracy.**

***Index Terms*—Image classification, multi-scale, speech recognition, spike-event-driven, spiking neural network.**

I. Introduction

THE multi-scale theory, has received much attention for its plethora of applications in computer vision, resulting in better pattern recognition performance [1], [2], [3], [4]. In [2], a multi-task convolutional neural network (m-CNN) model is proposed via the use of the multiple resolutions of inputs, leading to a promising accuracy. By leveraging the tradeoff between smaller model parameters and higher classification result, an inception network employs the parallel and multi-branch structure, and features thus, are extracted from different reception fields. However, when the multi-scale theory comes to deep artificial neural networks (ANNs), there is a major task to be done because ANNs are both computationally demanding and memory intensive, rendering them challenging to deploy on embedded systems of mobile and wearable devices with limited hardware resources [5], [6]. This prompts us to look into energy-efficient solutions, namely, spiking neural networks (SNNs).

The brain-inspired SNNs, as the third generation of neural networks models with asynchronous event-driven information representation and communication paradigm, offer attractive energy-saving owing to their biological neural systems via event-driven computation strategy [7]–[9]. The energy is mostly consumed only when spikes generation and communication take place, where the information is carried by these discrete spike trains (i.e., all-or-nothing impulses). Therefore, integrating the algorithmic power of deep SNN models and energy-efficiency of emerging neuromorphic computing architectures represents an intriguing solution for low-power big data applications. Currently, from the viewpoint of learning method, the SNN-based classifiers are mainly twofold: BP-based or its approximate methods [10], [11], [12], [13], [14] and STDP-based methods [15], [16], [17], [18], [19], [20], [21]. The training process of the former necessitates them non-amicable to the hardware implementation, such as on-chip learning. The bio-inspired STDP-based methods, albeit friendly and efficient neuromorphic hardware implementation, show mediocre to poor performance.

To address that, we take full advantage of the multi-scale structure to design a fully spike-event-driven SNN model for parallel computing for further feature extraction. Our work takes notable contributions summarized as follows:

1) Different with the existing ANN-based systems, the proposed system, termed as multi-scale deep spiking neural network (MS-DSNN), is a fully spike-event-driven model. It is more friendly for hardware implementation due to its computational efficiency on feature extraction, as well as its nature of energy-efficient SNN.
2) A new temporal coding method, named as SinSpike, is developed, where the input image contrasts or spectrograms are encoded in the timings of the output spikes on the basis of difference of Gaussians (DoG) filter.
3) To evaluate the effectiveness of the proposed MS-DSNN, we conduct experiments on different datasets for image classification and speech recognition tasks. The proposed system achieves superior performance compared with the existing schemes.

Manuscript received December 14, 2020; revised February 1, 2021; accepted February 8, 2021. Date of publication February 15, 2021; date of current version March 12, 2021. This work was supported in part by the Agency for Science, Technology, and Research (A*STAR) through its RIE2020 Advanced Manufacturing and Engineering (AME) Programmatic under Grant A1687b0033, and through its RIE2020 Industry Alignment Fund-Industry Collaboration Projects (IAF-ICP) under Grant 12001E0053; and in part by the National Research Foundation, Singapore through its National Robotics Programme (NRP)-Robotics Enabling Capabilities, and Technologies (RECT) under Grant 192 25 00054. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Heidi Christensen. *(Corresponding author: Qi Liu.)*

Zhixuan Zhang is with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: zhangzhixuan77@gmail.com).

Qi Liu is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583 (e-mail: elelqi@nus.edu.sg).

Digital Object Identifier 10.1109/LSP.2021.3059172

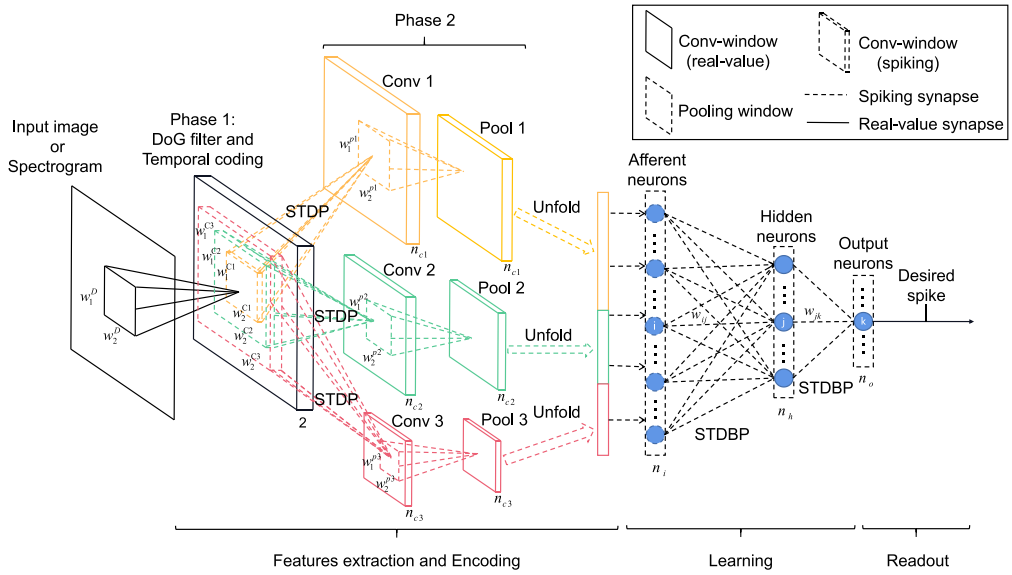

Fig. 1. Schematic of the feedforward computational model for pattern recognition.

II. Multi-Scale Deep Spiking Neural Network

A. Spiking Neural Network Model

The SNNs also mimic the biological process of the human brain that evolved by nature, and share the same network structure with conventional ANNs in terms of either feedforward or recurrent. The difference is that neuronal communication is mainly executed by stereotypical action potentials or spikes in the SNNs. Both the firing rate and temporal structure of the spike train are considered as the important information carriers in the biological neural systems. In this work, a non-leaky intergrate-and-fire (LIF) model is utilized for all the spiking neurons, where neurons accumulate input spikes from the former layer and generate output spikes from the soma whenever their membrane potentials surpass the firing threshold. Herein, the subthreshold membrane dynamics of the ith LIF neuron at time step t is modeled as follows:

$$V_j(t) = V_j(t-1) + \sum_i w_{i,j} s_i(t-1) \tag{1}$$

where $w_{i,j}$ is the synaptic weight of feedforward connection between the jth presynaptic and ith postsynaptic neurons. $s_i(t-1)$ denotes the incoming spike train from the jth presynaptic neuron, and $s_i(t-1) \in \{0, 1\}$. The spike generation process can be described by

$$V_j(t) = 0 \text{ and } s_i(t) = 1, \text{ if } V_j(t) \geq V_{th}. \tag{2}$$

Without loss of generality, we assume the firing threshold $V_{th} = 1$ and set the resting potential to 0 in this work.

B. Architecture of the Proposed MS-DSNN

As shown in Fig. 1, the proposed MS-DSNN model consists of three functional parts: feature extraction and encoding, learning, and readout. The information stream flows in a feedforward way and we then describe individual parts in detail.

1) Data Preprocessing: Feature extraction layer gains the spatiotemporal information of spike patterns, which serves as a key factor in pattern recognition. Given raw speech signals, to achieve fully event-driven frontend, they are decomposed into multiple independent frequency bands based on universal cochlear filterbanks using the constant-Q transform (CQT) [22] to replace the traditional mel-frequency cepstral coefficients (MFCC) [23], leading to some spectrograms with visual features. While it is different for processing the gray-scale images, the data is directly normalized and then moves forward to Phase 1, namely, the DoG filter and temporal coding module.

2) Feature Extraction and SinSpike Temporal Encoding: The first layer applies ON- and OFF- center DoG filters of size $w_1^D \times w_2^D$ on the input images or spectrograms. The DoG filter is a feature enhancement algorithm to increase the visibility of edges and preserve other spatial information. After that, the resulting spectrograms or image contrasts are encoded in the timings of the output spikes via the encoding method. As is well-known, the traditional linear latency coding approach has been used intensively in temporal encoding scheme. Given the normalization data X_i, the corresponding spike time for each real value p is shown as:

$$\text{Linear}: \ t_i = T_{\max} - X_i * (T_{\max} - T_{\min}) \tag{3}$$

where $T_{\max}$ and $T_{\min}$ represent the boundary of the encoding window. From (3), we can see that the larger the real value is, the shorter the delay time is. It is hoped that those features with smaller real values, such as background noise and edge information, can be encoded more later, while ones with larger real values (i.e, carried more important information) should go into the subsequent neural network earlier and the rest keeps with appropriate delay time. The reason behind that is because we expect the proposed encoding method enjoys the background noise suppression as well as preserves the edge information for further performance improvement. This motivates us to devise

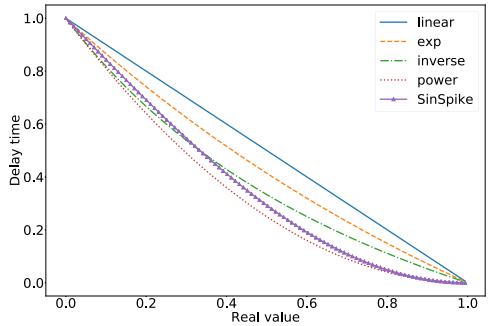

Fig. 2. Illustration of the curves of the proposed SinSpike and other schemes.

the non-linear SinSpike encoding approach:

$$\text{Non-linear}: \ t_i = (1 - \sin(X_i * \pi/2)) * T_{\max} \tag{4}$$

which is inspired by the curve shape of sine function in $[0, \pi/2]$ and its gradient decreases between 0 to $\pi/2$, going from fast to slow, as plotted in Fig. 2. Therefore, intensities of spectrograms (or images) are transformed into spikes by taking reciprocal of the value after Phase 1, which means higher intensities corresponds to earlier spikes.

To speed up the training calculation time for feature extraction, a new parallel computing structure composed of different kernel sizes of convolution maps and max pooling blocks in Phase 2 is proposed to further detect the local features, which is motivated by the multi-scale theory. Herein, STDP-based unsupervised learning method [24] is introduced to achieve invariance representation of visual inputs. That is:

$$\begin{cases} \Delta w_{ij} = a^+ * w_{ij} * (1 - w_{ij}), \text{if } t_j - t_i \leq 0 \\ \Delta w_{ij} = a^- * w_{ij} * (1 - w_{ij}), \text{if } t_j - t_i \geq 0 \end{cases} \tag{5}$$

where $a^+ = 0.004$ and $a^- = -0.003$.

Finally, temporal information are unfolded together for afferent neurons on spike-timing-dependent back-propagation (STDBP)-based SNN classifier [25].

3) SNN-Based Temporal Classifier: As shown in Fig. 1, encoding neurons are fully connected to output neurons and the carried information is propagated via precise spike timing. In this work, we mainly apply the STDBP learning rule as the SNN-based temporal classifier, given by

$$\begin{aligned} \frac{\partial t_j}{\partial w_{ij}} &= \frac{\partial t_j}{\partial V_j(t_j)} \frac{\partial V_j(t_j)}{\partial w_{ij}} = \frac{t_i - t_j}{\sum_i w_{ij}} \\ \frac{\partial t_j}{\partial t_i} &= \frac{\partial t_j}{\partial V_j(t_j)} \frac{\partial V_j(t_j)}{\partial t_i} = \frac{w_{ij}}{\sum_i w_{ij}} \end{aligned} \tag{6}$$

corresponding to the derivatives of the first spike time t_j with respect to synaptic weights w_{ij} and input spike times t_i, respectively, and $t_j < t_i$.

4) Readout: The readout part is to extract output spikes in the last SNN-based temporal classifier layer, where each learning neuron corresponds to one category for a classification task. The category of an input pattern will be determined by one of the neurons that generates the lowest spike distance. Here, we utilize the softmax function on the negative values of the spike times in the output layer, to minimize spike times of the desired neurons as well as to simultaneously maximize ones of the undesired neurons. The resulting distance is measured by the cross-entropy loss function:

$$\mathcal{L}(g, \mathbf{t}^o) = -\ln \frac{\exp(-\mathbf{t}^o[g])}{\sum_i \exp(-\mathbf{t}^o[i])} \tag{7}$$

where $\mathbf{t}^o$ represents the vector of the spike times in the output layer and g denotes the desired class index.

TABLE I
COMPARISON WITH THE EXISTING LATENCY CODING APPROACHES IN TERMS OF ACCURACY (%)

	MNIST	Caltech101	TIDIGITS	RWCP
Linear	97.51	97.83	94.20	99.50
Exponential	98.78	98.07	94.16	100
Inverse	98.84	98.43	94.60	100
Power	98.72	98.67	93.92	100
SinSpike	**98.90**	**98.79**	**95.09**	**100**

III. EXPERIMENTAL EVALUATION

In this section, we investigate the proposed MS-DSNN system on image classification and speech recognition tasks, based on the MNIST [26], Caltech101 (face/motor bike) [27], and RWCP [28], TIDIGITS [29] datasets, respectively.

A. Datasets

The MNIST dataset consists of 60K 28×28 grayscale images (i.e., handwritten digits 0 - 9) for training and 10K for testing. The Caltech101 dataset contains 101 categories and each category has 40 to 800 300×200 images with complex background noise. We only evaluate all compared models on face and motor bike categories, where 200 randomly selected images per category are used for training and the rest for testing. For speech recognition task, RWCP and TIDIGITS datasets are utilized. RWCP dataset consists of high-fidelity natural sound samples recorded in the real acoustic environment at a sampling rate of 16 kHz. We use the same 10 environmental sound classes, including 'cymbals,' 'horn,' 'phone4,' 'bells5,' 'kara,' 'bottle1,' 'buzzer,' 'metal15,' 'whistle1' and 'ring'. We randomly selected 40 samples from each class, wherein 20 samples are used to train an ANN-based sound classifier and the rest are used for evaluation. The TIDIGITS dataset comprises of reading digit sequences of variable lengths from 21 dialectical regions of the United States. We use the subset of isolated spoken digits from 11 classes (i.e., 'zero' to 'nine' and 'oh'), which consists of 2464 train and 2486 test utterances. The utterances are spoken by 111 male and 114 female speakers at a sampling rate of 20 kHz.

B. Experimental Results

1) Comparison With the Existing Latency Coding Approaches: To evaluate the validity of the proposed SinSpike temporal coding method, we compare with the existing latency coding approaches on our MS-DSNN system, as shown in Table I. The proposed method performs the best in terms of accuracy among all competitors, namely, Linear latency coding

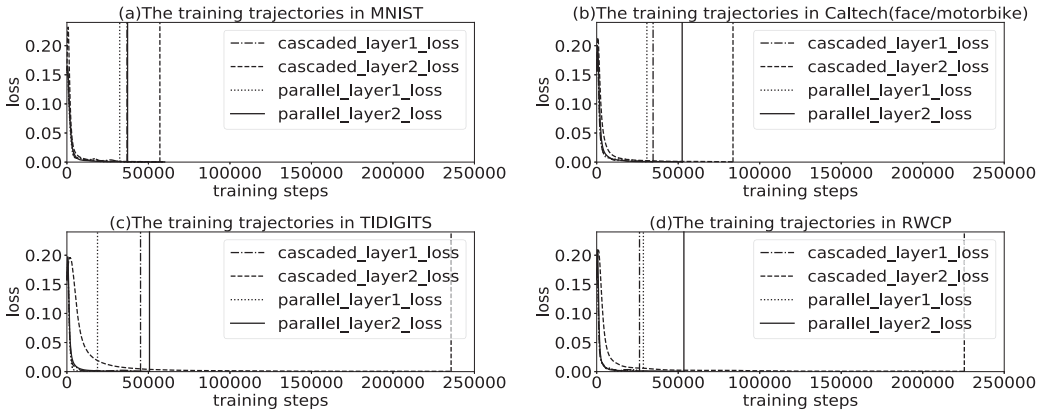

Fig. 3. Training time comparison between the cascaded and parallel structures with different datasets.

TABLE II
TRAINING TIME COMPARISON BETWEEN THE CASCADED AND PARALLEL STRUCTURES WITH DIFFERENT DATASETS

	MNIST	Caltech101	TIDIGITS	RWCP
Cascaded	5134s	6438s	10076s	9880s
Parallel	**2848s**	**2542s**	**1723s**	**1988s**

method and its variants [30]. The reason behind this result can be explained by Fig. 2, we observe that the proposed method costs more time to generate spikes compared with other non-linear ones when the intensities of real values are less than 0.4, while it performs faster than others when the intensities of real values are larger than 0.8. It is consistent with our above analysis.

2) Training Time Comparison Between the Cascaded and Parallel Structures With Different Datasets: As shown in Table II, the parallel computing scheme of the proposed MS-DSNN system takes the advantage of speeding up the training time, as compared to its cascaded structure. This is also verified by Fig. 3, where the number of training steps of each layer[1] in parallel structure is less than that in cascaded structure. The learning convergence is measured by the loss function as the same with that in [15]. Without loss of generality, the loss threshold is set at 0.0005 for both cascaded and parallel structures. From Fig. 3, we can see that our system with both cascaded and parallel structures can converge.

3) Comparison With the Existing Systems: To further test the performance of the proposed MS-DSNN system, we compare it with other systems among different datasets for both image classification and speech recognition tasks. We conduct experiment on the MNIST dataset, and the results are provided at Table III. Ours outperforms other rate-based or spike-based models for image classification, at the accuracy up to 98.9%. Similar with the results on Caltech101 (face/motor bike) dataset, the proposed model achieves higher accuracy at 98.79%, compared to 98.2% in [20]. On the other hand, Table IV presents the comparison results for speech recognition task on the TIDIGITS dataset, and ours performs the best among all compared models. Moreover, the proposed technique achieves superior recognition result on the RWCP dataset with an accuracy 100%, as comparison with 98.5%, 97.5% and 99.1% for [31], [32], [14], respectively.

[1]"Each layer i" means that the cascaded combination of Conv i and Pool i. Herein, to show the figure more clearly, only two layers are involved.

TABLE III
THE CLASSIFICATION ACCURACIES OF DIFFERENT MODELS ON THE MNIST DATASET

Model	Neural coding	Accuracy (%)
Hussain et al. [33]	Rate-based	90.3
Zhao et al. [13]	Spike-based	91.3
Querlioz et al. [19]	Spike-based	93.5
O'Connor et al. [34]	Rate-based	94.1
Diehl et al. [18]	Spike-based	95.0
Kheradpisheh et al. [15]	Spike-based	98.4
Ours	Spike-based	**98.9**

TABLE IV
THE RECOGNITION ACCURACIES OF DIFFERENT MODELS ON THE TIDIGITS DATASET

Model	Accuracy (%)
Zhang et al. [35]	92.3
Tavanaei et al. [21]	91.0
Abdollahi et al. [36]	78.7
Ours	**95.09**

IV. CONCLUSION

This work proposes an efficient computational feedforward architecture based on multi-scale structure and fully event-driven SNN. Features are represented as temporal information using the proposed SinSpike coding method throughout the architecture. Experimental results demonstrate the effectiveness of the proposed MS-DSNN system with the advantages of parallel computing on feature extraction and superior performance compared with counterparts for image classification and speech recognition tasks.

REFERENCES

[1] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[2] X. Yin and X. Liu, "Multi-task convolutional neural network for pose-invariant face recognition," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 964–975, Feb. 2018.

[3] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.
[4] Y. Zhou, X. Tian, and H. Li, "Multi-task waveRNN with an integrated architecture for cross-lingual voice conversion," *IEEE Signal Process. Lett.*, vol. 27, pp. 1310–1314, 2020.
[5] J. Wu, Y. Chua, M. Zhang, H. Li, and K. C. Tan, "A spiking neural network framework for robust sound classification," *Front. Neurosci.*, vol. 12, pp. 836.1–836.17, 2018.
[6] Z. Pan, Y. Chua, J. Wu, M. Zhang, H. Li, and E. Ambikairajah, "An efficient and perceptually motivated auditory neural encoding and decoding algorithm for spiking neural networks,"*Front. Neurosci.*, vol. 13, pp. 1420.1–1420.17, 2020.
[7] M. Zhang, H. Qu, A. Belatreche, Y. Chen, and Z. Yi, "A highly effective and robust membrane potential-driven supervised learning method for spiking neurons," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 123–137, Jan. 2019.
[8] M. Zhang *et al.*, "MPD-AL: An efficient membrane potential driven aggregate-label learning algorithm for spiking neurons," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, 2019, pp. 1327–1334.
[9] M. Zhang, H. Qu, A. Belatreche, and X. Xie, "EMPD: An efficient membrane potential driven supervised learning algorithm for spiking neurons," *IEEE Trans. Cogn. Develop. Syst.*, vol. 10, no. 2, pp. 151–162, Jun. 2018.
[10] Y. Cao, Y. Chen, and D. Khosla, "Spiking deep convolutional neural networks for energy-efficient object recognition," *Int. J. Comput. Vis.*, vol. 113, pp. 54–66, 2015.
[11] S. B. Shrestha and G. Orchard, "SLAYER: Spike layer error reassignment in time," in part of Advances in Neural Information Processing Systems 31, 2018, *arXiv:1810.08646*.
[12] R. Xiao, Q. Yu, R. Yan, and H. Tang, "Fast and accurate classification with a multi-spike learning algorithm for spiking neurons," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 1445–1451.
[13] B. Zhao, R. Ding, S. Chen, B. Linares-Barranco, and H. Tang, "Feedforward categorization on AER motion events using cortex-like features in a spiking neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 9, pp. 1963–1978, Sep. 2015.
[14] Y. Yao, Q. Yu, L. Wang, and J. Dang, "A spiking neural network with distributed keypoint encoding for robust sound recognition," in *Proc. Int. Joint Conf. Neural Netw.*, 2019, pp. 1–8.
[15] S. R. Kheradpisheh, M. Ganjtabesh, S. J. Thorpe, and T. Masquelier, "STDP-based spiking deep convolutional neural networks for object recognition," *Neural Netw.*, vol. 99, pp. 56–67, 2018.
[16] M. Mozafari, M. Ganjtabesh, A. Nowzari-Dalini, S. J. Thorpe, and T. Masquelier, "Bio-inspired digit recognition using reward-modulated spike-timing-dependent plasticity in deep convolutional networks," *Pattern Recognit.*, vol. 94, pp. 87–95, 2019.
[17] Q. Yu, H. Tang, K. C. Tan, and H. Li, "Rapid feedforward computation by temporal encoding and learning with spiking neurons," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 10, pp. 1539–1552, Oct. 2013.
[18] P. Diehl and M. Cook, "Unsupervised learning of digit recognition using spike-timing-dependent plasticity," *Front. Comput. Neurosci.*, vol. 9, pp. 99.1–99.9, 2015.
[19] D. Querlioz, O. Bichler, P. Dollfus, and C. Gamrat, "Immunity to device variations in a spiking neural network with memristive nanodevices," *IEEE Trans. Nanotechnol.*, vol. 12, no. 3, pp. 288–295, May 2013.
[20] M. Mozafari, S. R. Kheradpisheh, T. Masquelier, A. Nowzari-Dalini, and M. Ganjtabesh, "First-spike-based visual categorization using reward-modulated STDP," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6178–6190, Dec. 2018.
[21] A. Tavanaei and A. Maida, "A spiking network that learns to extract spike signatures from speech signals," *Neurocomputing*, vol. 240, pp. 191–199, Feb. 2017.
[22] H. B. Sailor and H. A. Patil, "Novel unsupervised auditory filterbank learning using convolutional RBM for speech recognition," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 12, pp. 2341–2353, Dec. 2016.
[23] K. A. Darabkh, L. Haddad, S. Z. Sweidan, M. Hawa, R. Saifan, and S. H. Alnabelsi, "An efficient speech recognition system for arm-disabled students based on isolated words," *Comput. Appl. Eng. Educ.*, vol. 26, no. 2, pp. 285–301, 2018.
[24] T. Masquelier and S. J. Thorpe, "Unsupervised learning of visual features through spike timing dependent plasticity," *PLOS Comput. Biol.*, vol. 3, no. 2, pp. 1–11, Feb. 2007.
[25] M. Zhang *et al.*, "Rectified linear postsynaptic potential function for backpropagation in deep spiking neural networks," in arXiv, 2020.
[26] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[27] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop*, 2004, pp. 178–178.
[28] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, "Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition," in *Proc. 2nd Int. Conf. Lang. Resour. Eval.* Athens, Greece: *Eur. Lang. Resour. Assoc.*, May 2000, pp. 965–968.
[29] R.G. Leonard and G. Doddington, "TIDIGITS LDC93S10," Web Download. Philadelphia: Linguistic Data Consortium, DOI: 10.35111/72xz-6x59, 1993.
[30] H. He *et al.*, "Constructing an associative memory system using spiking neural network," *Front. Neurosci.*, vol. 13, no. 650, pp. 1–15, 2019.
[31] J. Dennis, Q. Yu, H. Tang, H. D. Tran, and H. Li, "Temporal coding of local spectrogram features for robust sound recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 803–807.
[32] R. Xiao, R. Yan, H. Tang, and K. C. Tan, "A spiking neural network model for sound recognition," in *Proc. Cogn. Syst. Signal Process.*, F. Sun, H. Liu, and D. Hu, Eds. Singapore: Springer Singapore, 2017, pp. 584–594.
[33] S. Hussain, S. Liu, and A. Basu, "Improved margin multi-class classification using dendritic neurons with morphological learning," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2014, pp. 2640–2643.
[34] 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," *Front. Neurosci.*, vol. 7, no. 178, pp. 1–13, Oct. 2013.
[35] Y. Zhang, P. Li, Y. Jin, and Y. Choe, "A digital liquid state machine with biologically inspired learning and its application to speech recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 11, pp. 2635–2649, Nov. 2015.
[36] M. Abdollahi and S.-C. Liu, "Speaker-independent isolated digit recognition using an AER silicon cochlea," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, 2011, pp. 269–272.