MOTIF-TOPOLOGY AND REWARD-LEARNING IMPROVED SPIKING NEURAL NETWORK FOR EFFICIENT MULTI-SENSORY INTEGRATION

Shuncheng Jia [1,2,#], *Ruichen Zuo* [1,4,#], *Tielin Zhang* [1,2,#,*], *Hongxing Liu* [1], *Bo Xu* [1,2,3,*]

[1]Institute of Automation, Chinese Academy of Sciences, China
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, China
[3]Center for Excellence in Brain Science and Intelligence Technology, CAS, China.
[4]School of Information and Electronics, Beijing Institute of Technology, China

ABSTRACT

Network architectures and learning principles are key in forming complex functions in artificial neural networks (ANNs) and spiking neural networks (SNNs). SNNs are considered the new-generation artificial networks by incorporating more biological features than ANNs, including dynamic spiking neurons, functionally specified architectures, and efficient learning paradigms. In this paper, we propose a Motif-topology and Reward-learning improved SNN (MR-SNN) for efficient multi-sensory integration. MR-SNN contains 13 types of 3-node Motif topologies which are first extracted from independent single-sensory learning paradigms and then integrated for multi-sensory classification. The experimental results showed higher accuracy and stronger robustness of the proposed MR-SNN than other conventional SNNs without using Motifs. Furthermore, the proposed reward learning paradigm was biologically plausible and can better explain the cognitive McGurk effect caused by incongruent visual and auditory sensory signals.

Index Terms— Spiking Neural Network, Multi-sensory Integration, Motif Topology, Reward Learning

1. INTRODUCTION

Spiking neural networks (SNNs) are considered as the third generation of artificial neural network (ANNs) [1], which are biologically plausible at both network architectures and learning paradigms. The neurons, synapses, networks, and learning principles in SNNs are far more complex and powerful than those used in ANNs [2].

This paper highlights two important features of SNNs, which are also the most differences between SNNs and ANNs, including specific network architectures and efficient learning principles. For the architectures, specific cognitive topologies learned from evolution are highly sparse and efficient in SNNs [3], instead of pure densely-recurrent ones in counterpart ANNs. For the learning principles, SNNs are more tuned by biologically-plausible plasticity principles, e.g., the spike timing-dependent plasticity (STDP) [4], short-term plasticity (STP) [5] (which further includes facilitation and depression), lateral inhibition, Long-Term Potentiation (LTP), Long-Term Depression (LTD), Hebbian learning, synaptic scaling, synaptic redistribution and reward-based plasticity [6], instead of by the pure multi-step backpropagation (BP) of errors in ANNs. The SNNs encode spatial information by fire rate and temporal information by spike timing, giving us hints and inspiration that SNNs are also powerful in integrating visual and auditory sensory signals.

In this paper, we focus more on the key feature of SNNs at information representation, integration, and classification. Hence, a Motif-network and Reward-learning improved SNN (MR-SNN) is proposed and then will be verified efficient on multi-sensory integration. The MR-SNN contains at least three key advantages. First, specific Motif circuits can improve accuracy and robustness at single-sensory and multi-sensory classification tasks. Second, MR-SNN can reach a relatively little higher and lower computation cost than other state-of-the-art SNNs without Motifs. Third, the reward learning paradigm can better describe the McGurk effect [7], which describes an interesting psychological phenomenon that a new but reasonable audio concept might generate as a consequence of giving incongruent visual and auditory inputs. It exhibits a biologically-like behavior by using biologically plausible learning principles.

2. RELATED WORKS

For biological connections, the visual-haptic integration pathway [8], visual-vestibular integration pathway [9], and visual-auditory integration pathway [10] have been identified and played important roles in opening the black box of cognitive
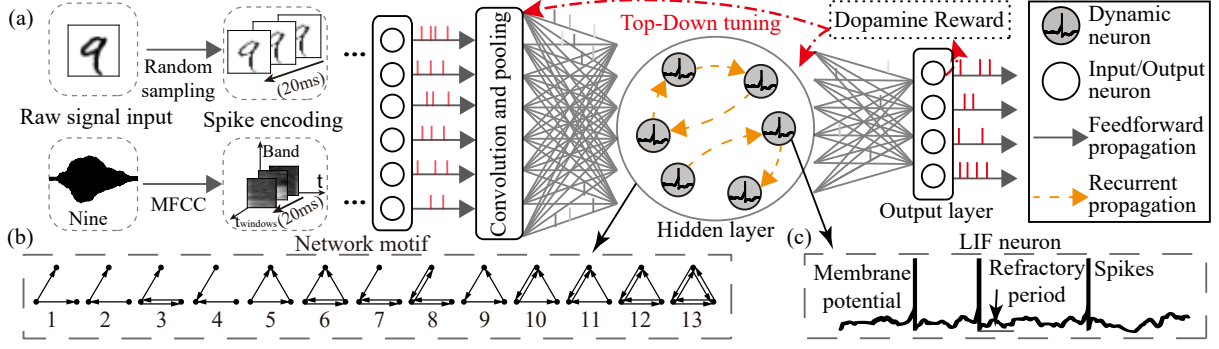
Fig. 1. The architecture of MR-SNN. (**a**) The architecture of MR-SNN on the multi-sensory integration task. (**b**) The example of artificial 3-node Motifs. (**c**) The spiking neuron with dynamic membrane potential.

functions in the biological brain [11].

For learning paradigms, besides biologically plausible principles (e.g., STDP, STP), some efficient algorithms have been proposed, such as ANN-to-SNN conversion (i.e., directly train ANNs with BP first and then equivalently convert to SNNs) [12], proxy gradient learning (i.e., replacing the non-differential membrane potential at firing threshold by an infinite gradient value) [13], and temporal BP learning (e.g., SpikeProp) [14].

For application tasks, SNNs have shown powerful on motion planning [15], visual pattern recognition [16, 5] and probabilistic inference [17].

3. METHODS

3.1. The Motif topology

All the 13 types of 3-node Motifs in Fig. 1(b) have been used to analyze functions in various types of biological networks [11]. Here, for simplicity, we transform the synaptic weights in SNNs into the range of 0 to 1 by using the Sigmoid function first and then count the Motif distribution to generate the Motif mask, named as $M_t^{r,l}$ at the recurrent layer l.

3.2. The spiking neurons and recurrent architectures

The leaky integrated-and-fire (LIF) neurons are the most simple and basic neurons for simulating biologically plausible spiking dynamics, including non-differential membrane potential and refractory period, as shown in Fig. 1(c).

The recurrent connections between LIF neurons are usually used to simulate the network-scale dynamics in SNNs. As shown in Fig. 1(a), we designed a four-layer SNN architecture containing visual and auditory input encoding, multi-sensory integration in a recurrent hidden layer, and readout layer. The synaptic weights between neurons in hidden layers are adaptive, while predefined Motif masks decide the connections between neurons. The membrane potentials in hidden layers are the integration of feedforward potential and recurrent potential, shown as follows:

$$\begin{cases} S(t) = S^f(t) + S^r(t) \\ V_i(t) = V_i^f(t) + V_i^r(t) \\ C\frac{dV_i^f(t)}{dt} = g(V_i(t) - V_{rest})(1 - S(t)) + \sum_{j=1}^{N} W_{i,j}^f X_j \\ C\frac{dV_i^r(t)}{dt} = \sum_{j=1}^{N} W_{i,j}^r S(t) \cdot M_t^{r,l} \end{cases},$$

(1)

where C is the capacitance, $S(t)$ is the firing flag at timing t, $V_i(t)$ is the membrane potential of neuron i that incorporates feed-forward $V_i^f(t)$ and recurrent $V_i^r(t)$, V_{rest} is the resting potential, $W_{i,j}^f$ is the feed-forward synaptic weight from the neuron i to the neuron j, $W_{i,j}^r$ is the recurrent synaptic weight from the neuron i to the neuron j. $M_t^{r,l}$ is the mask that incorporates Motif topology to influence the feed-forward propagation further. The historical information is stored in the forms of recurrent membrane potential $V_i^r(t)$, where spikes are generated after potential reaching a firing threshold, shown as follows:

$$\begin{cases} V_i^f(t) = V_{reset}, S^f(t) = 1 & if(V_i^f(t) = V_{th}) \\ V_i^r(t) = V_{reset}, S^r(t) = 1 & if(V_i^r(t) = V_{th}) \\ S^f(t) = 1 & if(t - t_{sf} < \tau_{ref}, t \in (1, T_1)) \\ S^r(t) = 1 & if(t - t_{sr} < \tau_{ref}, t \in (1, T_2)) \end{cases},$$

(2)

where $V_i^f(t)$ is the feed-forward membrane potential, $V_i^r(t)$ is the recurrent membrane potential, $S^f(t)$ and $S^r(t)$ are spike flags of feed-forward and recurrent membrane potentials, respectively, V_{reset} is reset membrane potential.

3.3. The local principle of gradient approximation

The membrane potential at the firing time is a non-differential spike, so local gradient approximation (pseudo-BP) [18] is usually used to make the membrane potential differentiable by replacing the non-differential part with a predefined number, shown as follows:

$$Grad_{local} = \frac{\partial S_i(t)}{\partial V_i(t)} = \begin{cases} 1 & if\ (|V_i(t) - V_{th}| < V_{win}) \\ 0 & else \end{cases}, \tag{3}$$

where $Grad_{local}$ is the local gradient of membrane potential at the hidden layer, $S_i(t)$ is the spike flag at neuron i, $V_i(t)$ is the membrane potential of neuron i, V_{th} is the firing threshold. This approximation makes the membrane potential $V_i(t)$ differentiable at the spiking time between an upper bound of $V_{th} + V_{win}$ and a lower bound of $V_{th} - V_{win}$.

3.4. The global principle of reward learning

The reward propagation has been proposed in our previous work [18], where the reward signal is directly given to all hidden neurons without layer-to-layer backpropagation, shown as follows:

$$\begin{cases} Grad_R = B_{rand}^{f,l} \cdot R_t - h^{f,l} \\ \Delta W_t^{f,l} = -\eta^f (Grad_R) \\ \Delta W_t^{r,l} = -\eta^r (Grad_{t+1} + Grad_R) \cdot M_t^{r,l} \end{cases}, \tag{4}$$

where $h^{f,l}$ is the current state of layer l, R_t is the predefined reward for current input signal. A predefined random matrix $B_{rand}^{f,l}$ is designed to generate the reward gradient $Grad_R$. $W_t^{f,l}$ is the synaptic weight at layer l in feed-forward phase, $\Delta W_t^{r,l}$ is the recurrent-type synaptic modification at layer l which is defined by both $Grad_R$ by reward learning and $Grad_{t+1}$ by iterative membrane-potential learning [19]. The $M_t^{r,l}$ is the mask that incorporates Motif topology to further influence the propagated gradients.

4. EXPERIMENTS

4.1. Visual and auditory Datasets

The MNIST dataset [20] was selected as the visual sensory dataset, containing 70,000 28×28 one-channel gray images of handwritten digits from zero to nine. Among them, 60,000 images are selected for training, while the remaining 10,000 ones are left for testing. The TIDigits dataset [21] was selected as the auditory sensory dataset, containing 4,144 spoken digit recordings from zero to nine, corresponding to those in the MNIST dataset. Each recording was sampled as 20KHz for around 1 second. Some examples are shown in Fig. 1(a).

4.2. Experimental configurations

We built the SNN in Pytorch and trained on TITAN Xp GPU. The network architectures for MNIST and TIDigits are the same, containing one convolutional layer (with a kernel size of 5×5), one full-connection or integrated layer (with 200 LIF neurons), and one output layer (with ten output neurons). The

capacitance C is 1μ F/cm^2, conductivity g is 0.2 nS, time constant τ_{ref} is 1 ms, resting potential V_{rest} is equal to reset potential V_{reset} with 0 mV. The learning rate is $1e$-4, the firing threshold V_{th} is 0.5 mV, the simulation time T is set as 28 ms, the gradient approximation range V_{win} is 0.5 mV.

As shown in Fig. 1(a), before being given to the input layer, the raw input signals were encoded to spike trains first by random sampling (for that in spatial image data) or by the temporal encoding of MFCC [22] (for that in temporal auditory data). The encoding aimed to convert the input to spike trains by comparing each number with a random number generated from Bernoulli sampling at each time slot of time window T.

Algorithm 1 The MR-SNN algorithm.

1. Initialize the network by resetting weights and all related parameters.
2. Encode to spike trains from the raw numbers in datasets.
3. Train the proposed MR-SNN on single-sensory datasets.
3.1 Learn synaptic weights w_{ij} and Motif masks $M_t^{r,l}(s)$, $M_t^{r,l}(t)$ of networks with pseudo-BP and reward-learning in two datasets, respectively.
3.2 Save the Motif masks during single-sensory classification, in which the spatial $M_t^{r,l}(s)$ and temporal $M_t^{r,l}(t)$ were generated from visual and auditory tasks, respectively.
4. Synthesize the integrated masks $M_t^{r,l}$ from spatial and temporal masks, where $M_t^{r,l} = 1/2 \times \left(M_t^{r,l}(s) + M_t^{r,l}(t) \right)$. And retrain the $w_{i,j}$ of the network with frozen Motif mask $M_t^{r,l}$.
5. Test the performance of SNNs using these new masks in the multi-sensory classification tasks, and make comparison.


4.3. Learning Motif-topology from single-sensory tasks

The Motif distributions in visual and auditory categorizations were shown in Fig. 2(a,c), learned from different single-sensory classification tasks. We further set "credible frequency" by multiplying the occurrence frequency and $1 - P$, where the P is the P-value of a selected Motif after comparing it to 2,000 matrices that each element is uniformly and randomly distributed, indicating more credibility given by a lower P-value.

The 3rd Motif in MNIST and the 13th Motif in TIDigits are the most credibly frequent topologies. The corresponding visualization of these two types of Motif distributions is shown in Fig. 2(b,d), where the distribution on visual task is sparser than that on auditory task, which is also corresponding with the biological findings [23, 24]. The following sections will verify these key Motif connections playing important roles in improving accuracy and robustness during spatial and temporal classifications.

4.4. Motif-topology for stronger robust network

As shown in Table 1, the higher accuracies than SVM demonstrated the capacity of SNNs for dealing with sensory information. Furthermore, the experimental results showed that the SNNs using feedforward and Motif-topology (FF-Motif)
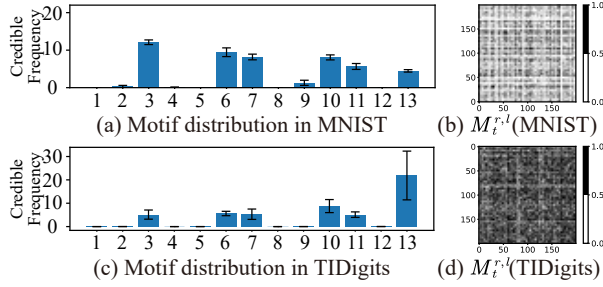
Fig. 2. Key Motif distributions and examples of visualization learned from the MNIST (**a**, **b**) and TIDigits (**c**, **d**) datasets. All figures are averaged over 5 repeating experiments with different random seeds.

were more robust than those using the feedforward connection (FF). The standard SNNs-Motif with dopamine learning [18] reached a higher accuracy than other algorithms, with an improvement of around 0.03%, 0.43%, and 0.03% for MNIST, TIDigits, and integrated sensory, respectively. After giving uniformly distributed random noise to raw input, the accuracy improvement of SNN-Motif was higher than not giving noise, reaching around 7.81%, 7.36%, and 0.76% for MNIST, TIDigits, and integrated sensory, respectively. It showed that these special Motif-topologies (the third one in MNIST and the 13th one in TIDigits) also contributed to the robust computation of SNNs during classification.

4.5. Motif-topology for better multi-sensory integration

For the experiments of multi-sensory integration, the visual and auditory signals were simultaneously given to an SNN using integrated Motif distributions learned from single-sensory tasks. As shown in Table 1, the learning accuracies for the SNNs using single visual (98.50%) or auditory (98.20%) sensories were lower than that using multi sensories (99.55%). In addition, the means of accuracies for single-sensory tasks were improved up to 0.43% for SNNs using Motif distributions than those without using them.

Table 1. The comparison of accuracy (%) after giving Motif topology with (∗) or without (−) 80% additional noise to the input layers.

Topology	MNIST	TIDigits	Integration
SVM[25]	97.92	94.53	−
FF(−)[18]	98.50±0.02	98.20±0.14	99.55±0.06
FF-Motif(−)	98.53±0.09	98.63±0.10	99.58±0.09
FF(∗)	90.55±1.17	83.71±2.23	98.37±0.09
FF-Motif(∗)	98.36±0.11	91.07±1.42	99.13±0.11

We found a notable accuracy increase for the paradigms with input noises, reaching 98.36%, 91.07%, and 99.13% for visual, auditory, and integrated sensory, respectively. In conclusion, the Motif-topology would improve the accuracy of all

three tasks under circumstances of whether giving additional noise or not.

4.6. Motif-topology for the explainable McGurk effect

The McGurk effect [7] describes an interesting psychological phenomenon where incongruent voice saying and face articulating will result in a new concept. For example, we believe we hear as [d] but actually voice [b] and face [g] are given, as shown in Fig. 3(a).
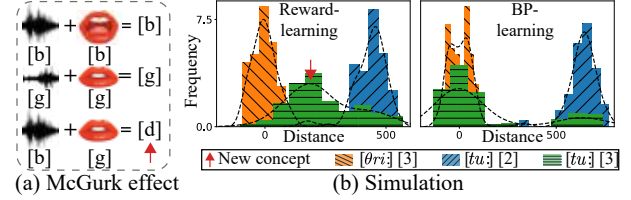
Fig. 3. McGurk effect and probability density distribution of feature distances at integrated layers. [tu:] and [θri:] are the pronunciations of spoken two and three, [2] and [3] are the hand-digit images of two and three.

We simulated this phenomenon with our proposed MR-SNN model. For simplicity, we used handwritten digits (images of [2] and [3]) and spoken digits (pronunciation of [tu:] and [θri:]) to represent face articulating and voice saying, respectively.

The two histogram figures in Fig. 3(b) were calculated to approximate the probability density of feature distance in the integrated layer (see experimental configurations for more details). We could only see two distinguished normally distributed distributions representing congruent sensory signals to SNNs using pseudo-BP, including [θri:, 3] (orange bar) and [tu:, 2] (blue bar). However, besides two old concepts, a new concept (green) would generate for SNNs using reward learning and integrated Motif circuits, especially after receiving incongruent sensory signals [tu:, 3] (green bar). This is a small step in understanding the McGurk effect from simple algorithmic comparison. We will further discuss it in our further works.

5. CONCLUSION

We proposed a new Motif-topology and Reward-learning improved SNN (MR-SNN), exhibiting two important features. First, the Motif topology learned from spatial or temporal data could improve accuracy and robustness than standard SNNs without using Motifs. Second, with the biologically plausible reward learning, the proposed MR-SNN could simulate the McGurk effect found in cognitive integration of human brains, where the traditional SNNs using pseudo-BP would be failed. The source code of the models and experiments can be found at https://github.com/thomasaimondy/Motif-SNN.

6. REFERENCES

[1] Wolfgang Maass, "Networks of spiking neurons: the third generation of neural network models," *Neural Networks*, vol. 10, no. 9, pp. 1659–1671, 1997.

[2] Demis Hassabis, Dharshan Kumaran, Christopher Summerfield, and Matthew Botvinick, "Neuroscience-inspired artificial intelligence," *Neuron*, vol. 95, no. 2, pp. 245–258, 2017.

[3] Liqun Luo, "Architectures of neuronal circuits," *Science*, vol. 373, no. 6559, pp. eabg7285, 2021.

[4] Tielin Zhang, Yi Zeng, Dongcheng Zhao, and Mengting Shi, "A plasticity-centric approach to train the non-differential spiking neural networks," in *The 32th AAAI Conference on Artificial Intelligence (AAAI-2018)*, 2018.

[5] Tielin Zhang, Yi Zeng, Dongcheng Zhao, and Bo Xu, "Brain-inspired balanced tuning for spiking neural networks.," in *IJCAI*, 2018, pp. 1653–1659.

[6] Wickliffe C Abraham and Mark F Bear, "Metaplasticity: the plasticity of synaptic plasticity," *Trends in neurosciences*, vol. 19, no. 4, pp. 126–130, 1996.

[7] K. Tiippana, "What is the mcgurk effect?," *Frontiers in Psychology*, vol. 5, 2014.

[8] Marc O Ernst and Martin S Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, no. 6870, pp. 429–433, 2002.

[9] Zhixian Cheng and Yong Gu, "Distributed representation of curvilinear self-motion in the macaque parietal cortex," *Cell Reports*, vol. 15, no. 5, pp. 1013–1023, 2016.

[10] Barry E Stein, M Alex Meredith, W Scott Huneycutt, and Lawrence McDade, "Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli," *Journal of Cognitive Neuroscience*, vol. 1, no. 1, pp. 12–24, 1989.

[11] K. Shen, G. Bezgin, R. M. Hutchison, J. S. Gati, and A. R. Mcintosh, "Information processing architecture of functionally defined clusters in the macaque cortex," *Journal of Neuroscience the Official Journal of the Society for Neuroscience*, vol. 32, no. 48, pp. 17465–76, 2012.

[12] Peter U Diehl, Daniel Neil, Jonathan Binas, Matthew Cook, Shih-Chii Liu, and Michael Pfeiffer, "Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing," in *The 2015 International Joint Conference on Neural Networks (IJCNN-2015)*. IEEE, 2015, pp. 1–8.

[13] Jun Haeng Lee, Tobi Delbruck, and Michael Pfeiffer, "Training deep spiking neural networks using backpropagation," *Frontiers in Neuroscience*, vol. 10, 2016.

[14] Sander M Bohte, Joost N Kok, and Han La Poutre, "Error-backpropagation in temporally encoded networks of spiking neurons," *Neurocomputing*, vol. 48, no. 1, pp. 17–37, 2002.

[15] Elmar Rueckert, David Kappel, Daniel Tanneberg, Dejan Pecevski, and Jan Peters, "Recurrent spiking networks solve planning tasks," *Scientific reports*, vol. 6, pp. 21142, 2016.

[16] Tielin Zhang, Yi Zeng, Dongcheng Zhao, and Mengting Shi, "A plasticity-centric approach to train the non-differential spiking neural networks," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[17] Alireza Soltani and Xiao-Jing Wang, "Synaptic computation underlying probabilistic inference," *Nature neuroscience*, vol. 13, no. 1, pp. 112–119, 2010.

[18] Tielin Zhang, Shuncheng Jia, Xiang Cheng, and Bo Xu, "Tuning convolutional spiking neural network with biologically plausible reward propagation," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–11, 2021.

[19] Paul J Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.

[20] Yann LeCun, "The mnist database of handwritten digits," *http://yann. lecun. com/exdb/mnist/*, 1998.

[21] R. Gary Leonard and George Doddington, "Tidigits ldc93s10," *Web Download. Philadelphia: Linguistic Data Consortium*, 1993.

[22] Beth Logan, "Mel frequency cepstral coefficients for music modeling," in *In International Symposium on Music Information Retrieval*. Citeseer, 2000.

[23] William E Vinje and Jack L Gallant, "Sparse coding and decorrelation in primary visual cortex during natural vision," *Science*, vol. 287, no. 5456, pp. 1273–1276, 2000.

[24] Tomáš Hromádka, Michael R DeWeese, and Anthony M Zador, "Sparse representation of sounds in the unanesthetized auditory cortex," *PLoS biology*, vol. 6, no. 1, pp. e16, 2008.

[25] John Platt, "Sequential minimal optimization : A fast algorithm for training support vector machines," *Microsoft Research Technical Report*, 1998.