

HMSNN: Hippocampus inspired Memory Spiking Neural Network

Tielin Zhang*, Yi Zeng*[†], Dongcheng Zhao*, Liwei Wang[‡], Yuxuan Zhao* and Bo Xu*[†]

*Institute of Automation, Chinese Academy of Sciences, Beijing, China

[†]CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai, China

[‡]School of Software and Microelectronics, Peking University, Beijing, China

Email: {tielin.zhang, yi.zeng}@ia.ac.cn

Abstract—Human beings receive stimulations in primary sensory cortex and transfer them to higher brain regions automatically. What happened in this procedure? In this paper, we will focus on one of these regions (hippocampus) and try to simulate its working procedure by building an HMSNN (Hippocampus inspired Memory Spiking Neural Network) model. Dentate Gyrus (DG) and Cornu Ammonis area 3 (CA3) are the main regions of hippocampus and will be simulated by feed forward Spiking Neural Network (SNN) and recurrent Hopfield-like network respectively. From the structural perspective, the computational unit and the connectivity between neurons in HMSNN are all consistent with the anatomical-experimental results in hippocampus. From the functional perspective, the multi-scale memory formation, memory abstraction and memory retention will be shown in HMSNN model. In addition, the HMSNN is tested on MNIST handwritten digit dataset (with static images) and robot walking dataset (with dynamical images). The experimental result shows that: biological neural circuit inspired HMSNN shows comparable classification performance on both datasets compared to the state-of-art convolutional neural networks (CNNs), and shows significantly better performance compared to CNN when noises are introduced to the original images.

I. INTRODUCTION

HUMAN brain, in which most of its internal operation mechanisms are still in black box, is attracting more and more scientists from various fields to work together for a step forward to get better understanding of it. With the deeper understanding of the brain, more brain-inspired neural network models are proposed and try to reveal the nature of biological brain intelligence through precise bottom-up simulation on both structural and functional perspectives. Since the brain is a complicated network with hierarchical and modular structures, most of the work about brain focus on specific brain functions and on limited target brain regions.

Learning and memory are essential for all mammals. There are various kinds of brain regions which are related with different kinds of memories. Hippocampus is an essential one, which is fundamentally important for the converting procedure from short term memory to long term memory.

Many models about the regions in hippocampal formation (HF) have been proposed, but to the best of our knowledge, an integrated one with both solid biological basis and good behavior level application results has not been built. Wills

proposed the attractor network structure which could make a detailed description about the mechanism of the memory (e.g. memory retention) [1]. O'reilly found the special connection styles of Cornu Ammonis area three (CA3) in anatomical experiments of hippocampus and proposed a recurrent network model to realize the functions of storage and retrieval of sensory data [2]. Paul Frankland found memory could be erased by new born neurons in Dentate Gyrus (DG) region of hippocampus [3]. Fred H. Gage gave the hypothesis that DG naturally has the ability of pattern abstraction which could get the invariant features from the input knowledge [4]. When the environment changes, the neurons and synapses of DG will change accordingly [4]. Fukushima applied invariant recognition of the network on realistic applications and got good performances on specific classification tasks [5].

In this paper, we try to build a hippocampus inspired model to realize the procedure of abstraction of invariant concepts in DG, the memory retention and storage in CA3, and the judgement of the competition in CA1. The model is applied on robot learning, memory and decision making.

In the procedure of cognitive robot memory formation, concept generation is usually the first step, and then is the association of the concepts (e.g. recall the previous experiences to make robot adapt to the current environment). Traditional robotics focus more on specific robot tasks [6], for example, chess-playing robots, robots on manufacturing industry, and robots for painting or guiding. However, these robots are mostly based on the precise pre-programming and have no ability to make the connections between neurons change automatically. They can neither effectively store, accumulate and integrate the acquired knowledge, nor recall and utilize the previous experience in the new tasks. The improvement of robot behavior performance relies on tuning parameters. In this paper we will try to solve the above problems by modeling the procedure of information storage and recall for a better robotic cognitive intelligence.

How does the brain understand hierarchical concepts? What is the mechanism for invariant representation in hippocampus? In this paper, we try to build hippocampus inspired model to answer these questions. The model will combine the spiking DG network with Hopfield-like CA3 network.

In Section 2, the detailed model of DG and CA3 will be introduced and integrated together based on the experi-

mental biological connectivity atlas from Allen Mouse Brain Atlas [7]. In Section 3, MNIST handwritten digit dataset (with static images) and robot walking dataset (with dynamical images) will be introduced, and the experiments about the invariant feature abstraction and memory formation will be tested and compared with traditional convolutional neural networks (CNNs) on these two datasets.

II. BIOLOGICAL BACKGROUND OF HIPPOCAMPUS

As Fig. 1 shows, the hippocampus contains DG, Cornu Ammonis area one (CA1), CA3 and Subiculum (Sub). The proportion of excitatory neurons in hippocampus is around 80%-90%, and they could be the pyramidal neurons in CA1 and CA3, or granule cells in DG [8]. Information from Entorhinal Cortex (EC) to hippocampus has two main processing maps: one is from EC to CA3 by perforant pathway, and then from CA3 to CA1 by Schaeffer pathway; the other is from EC to DG, from DG to CA3 by mossy fiber, and then from CA3 to CA1 by Schaeffer pathway.

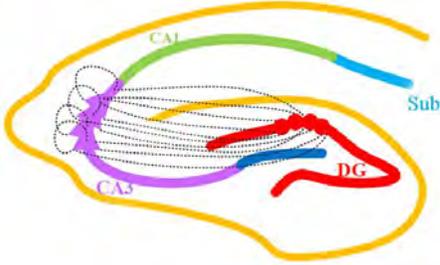


Fig. 1. The connectivity and structure of hippocampus.

The main sub regions in hippocampus are CA1-CA3, the Sub and the DG. Perforant pathway and trisynaptic pathway are the two main projections from EC to hippocampus. The output regions from HF to EC are CA1 and Sub. A sketch on the mesoscale connection among different sub regions of HF and EC is shown in Fig. 2.

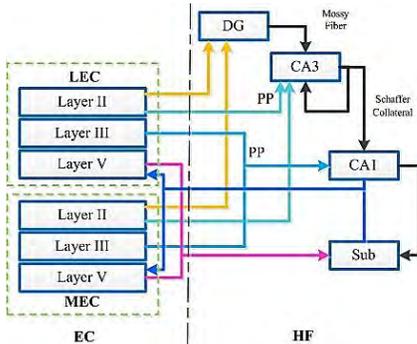


Fig. 2. The connectivity diagram of the EC-HF network. This image is integrated and refined from [9].

The number of CA3 neurons in mouse and human are 0.33×10^6 and 2.3×10^6 respectively, and the number of DG in mouse and human are 0.30×10^6 and 8.8×10^6 respectively. It is obvious that the number of neurons in human and mouse is

similar for CA3, but not for DG. This intuitively means that the DG region is more likely to be a key region for the feature abstraction and representation, and the CA3 is more like to be only a middle procedure of the memory retention.

III. THE FEEDFORWARD SNN FOR DG

The DG is an important sub region of hippocampus and the main functions of DG are the pattern abstraction and the memory retention or elimination [3], [4]. Inspired by the connectivity of DG, we design a feed forward spiking neural network to form the DG-like function of pattern abstraction.

A. The multi-layer SNN architecture of DG

The function of DG is to form the stable representation of the input information from EC, which is called the pattern abstraction procedure [4]. As shown in Fig. 3, the multi-layer SNN is with full connection. After training, the network could get the proper stable representation of information even when different proportions of distortions or noises are introduced.

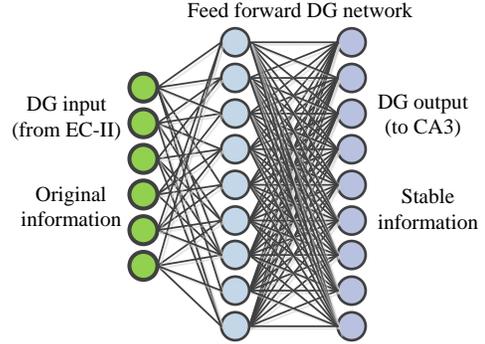


Fig. 3. The two layers of SNN architecture inspired by DG

The general multi-layer SNN architecture can be divided into three parts: the spike generation part, the hidden layers part and the output part. The spike generation part converts different kinds of signals into the standard spikes signals. In the hidden layers part, some network learning rules such as synaptic dynamics based on triphasic Spike Timing Dependent Plasticity (STDP) rules [10], the Possion type noise [11] are coordinated together for a better training performance to implement the feature abstraction function. The output part will realize the classification function by winner-take-all rules [12]. In the DG network, we only use the first and second parts to form the function of feature abstraction and realize the stable features with a static vector size. The stable feature will be the source of the CA3 network for memory storage and retrieval.

B. Neuron and synaptic models in DG

In the DG network, the excitatory neurons are granule cells [13]. Considering the computational efficiency, we choose Leaky Integrate and Fire (LIF) model as the neuron model

for network learning [14]. Standard LIF models are shown in Equation (1) and Equation (2).

$$C_m \frac{dV}{dt} = -g_L (V - V_L) + I_{syn} \quad (1)$$

$$\tau_m \frac{dV}{dt} = -(V - V_L) + \frac{I_{syn}}{g_L} \quad (2)$$

C_m is the membrane capacitance of the neuron, V is the membrane potential of the neuron, g_L is the conductance of the membrane, V_L is the steady-state leaky potential, I_{syn} is the input current of presynaptic neuron. $\tau_m = \frac{C}{g_L}$ represents the voltage delay time, and different types of neurons have different values of τ_m .

For neurons in the network, all received excitatory electrical conductivity is denoted as g^E , Excitatory reversal potential is represented as V^E . The electrical activity of each neuron in the DG network is shown in Equation (3), (4) and (5).

$$I_i^{syn} = -g^E (V - V^E) \quad (3)$$

$$\tau_m \frac{dV}{dt} = -(V - V_L) - \frac{g^E}{g_L} (V - V^E) \quad (4)$$

$$V \rightarrow V_{reset} \quad \text{if } (V > V_{th}) \quad (5)$$

When $V > V_{th}$, the potential value of the neuron is reset to the reference value as the neuron fires. The initial values are set accordingly, namely $V_L = -70\text{mv}$, $V_{th} = -50\text{mv}$, $V_{reset} = -55\text{mv}$, $\tau_m = 20\text{ms}$, $V^E = -70\text{mv}$, $V^I = -70\text{mv}$ and $g_L = 20\text{ns}$.

The unique dynamic plasticity of synapse is the key to the realization of learning and memory in biological neural networks [15]. We use a common model of synaptic connectivity, the Conductance-based Synapse Model [16]. The model assumes that the presynaptic spike affects on the electric potential of the postsynaptic membrane potential through opening the specific ion channel. The excitatory synapses bind AMPA receptors by releasing glutamate [17]. The ion channel in the receptors will be temporarily open after binding the neurotransmitter, allowing the ion to pass through the membrane of the receptors, which leads to an increase in the membrane conductance. Therefore, we set the ion concentrations related conductance value to excitatory synaptic receptors. Without the presynaptic input spikes, the electrical conductivity of excitatory and inhibitory neurons will gradually attenuate to zero with time constant τ_E , as shown in Equation (6).

$$\frac{dg_i^E}{dt} = -\frac{g_i^E}{\tau_E} \quad (6)$$

Whenever a presynaptic neuron j spikes, it will increase the synaptic conductance between neurons $g_{i,j}$ by an amount of $w_{i,j}$. t_k^j is the spiking time of neuron j . At this time, the membrane potential of the neuron j is increased to V_{peak} , and the membrane potential is slowly being reset to the reference potential when the neuron no longer fire for a long time. Finally, we can get Equation (7) through the upper four equations.

$$\Delta g_i^E(t) = -\frac{g_i^E(t)}{\tau_E} \Delta t + \sum_{j \in C_E} w_{i,j} \sum_k \delta_{t,t_k^j} \quad (7)$$

Here $\tau_E = 2\text{ms}$, $\tau_I = 5\text{ms}$. In the training procedure, the connection weights between granule cells will follow the STDP learning mechanism to update their values.

IV. THE RECURRENT NETWORK FOR CA3

The CA3 network connection architecture is shown in Fig. 4. The average convergence and divergence number of synapses from EC is 3750 and 4600, and from itself is 5500 and 6000. The network could be considered as a discrete Hopfield network [18]. Information received by CA3 has three sources: from DG, from Layer 2 of EC, or from itself.

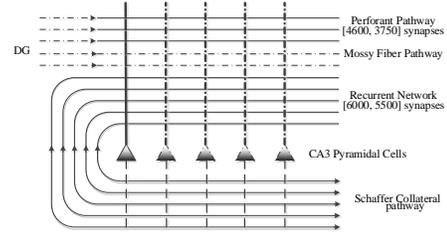


Fig. 4. Connectivity and number of synapses in CA3.

Based on the characteristics of the CA3 network, we use a discrete Hopfield network to mimic CA3. This network is with the functions of memory association, memory storage and memory retrieval. The Hopfield network could simulate the memory mechanisms of biological neural network with a very similar way with human memory.

A. Architecture of the CA3 network

Here we presents an illustrative example of the CA3 network, as shown in Fig. 5. The network has the following characteristics: (1) symmetric weights ($W_{ij} = W_{ji}$); (2) no neuron has a connection with itself. Symmetric weights are typically used to guarantee the energy function and decrease monotonically while following the activation rules. The network may exhibit some periodic or chaotic behavior if non-symmetric weights are used. Each neuron may receive information both from outside and from the network itself, thus forming the mutual coupling structure between neurons in the network. The state of each neuron is not only affected by the input information from outside, it is also related to the states of other neurons. Its relationship with the external input is no longer a simple mapping, and can be regarded as the result of joint action with the outside and inside units.

B. Neural computation model in CA3 network

Neurons in CA3 networks are binary units, which only use two different values to describe their states and the values are determined by the rules that whether or not the voltages of neurons exceed thresholds. Actually, the network is bipolar and nonlinear system, namely, using -1 instead of 0 to represent the negative state. We use McCulloch-Pitts neuron model. Due

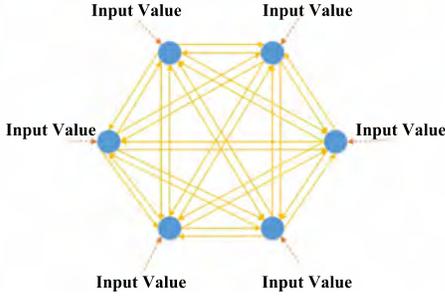


Fig. 5. An illustrative example of a discrete Hopfield network with 6 neurons.

to the connection pattern, the state of each neuron restricts others, while the corresponding status of other neurons in turn to restrict the stability of the specific state of the neuron, which makes the whole system balanced and stable in fixed points. Without the influence of external signals, the state will remain stable. But once the strong enough external stimulus has been added into the network, original state of equilibrium conditions will no longer exist. The system has different probabilities to move to the original or a new state.

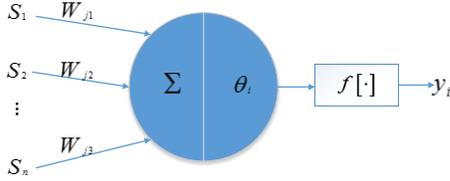


Fig. 6. Computing model of the neuron in CA3.

The computational neuron model in the network is shown in Fig. 6. Here, S_i is the state of neuron i , W_{ij} is the synaptic strength (connectivity weight) between neuron j and neuron i , θ_i is the threshold of neuron i . The update rules of a neuron is represented as Equation (8).

$$S_i = \begin{cases} +1, & \text{if } \sum_j W_{ij} S_j \geq \theta_i; \\ -1, & \text{otherwise.} \end{cases} \quad (8)$$

C. The principle of associative memory

The stable property of CA3 network inspires us to use associative memory function to model hippocampus network. In order to achieve this goal, we should try to store patterns as steady states of the network, and the network can transit from the original state to another steady state under the excitation of the external signal. we set the threshold of each neuron to be zero, namely, $\theta_i = 0$, and let the maximum number of storage patterns as N , which has effects on each neuron i and is predefined as the number of classes of images, as shown in Equation (9):

$$\text{sgn}\left(\sum_j W_{ij} S_j\right) = S_j \quad (9)$$

Obviously, if set

$$W_{ij} = \frac{1}{N} S_i S_j \quad (10)$$

then, the Equation (9) is established. It also needs to test the stability of the network to resist external disturbance. Suppose there are several neurons whose status is inconsistent with the storage pattern (denoted by A_j). The general input for neuron i is denoted as the following:

$$h_i = \sum_{j=0}^{n-1} S_i S_j S_j + \sum_{j=n}^{N-1} S_i S_j A_j \quad (11)$$

Obviously, if $n < \frac{N}{2}$, only the first one will determine the sign of the function (equation (11)), suggesting that the presence of some error information does not influence the neurons deviate from the correct status. The Equation (9) to (11) can explain why CA3 network can have associative memory function. Given parts of the information of the stored pattern, the network could provide complete information after a few calculation iterations. The state of each neuron is determined by weighted input information and a symbol function. Each neuron receives all the information provided by other neurons, and the joint mode mainly plays the role of mutual support. If the state of any neuron does not match the storage pattern, it will react against to the opposite. As long as the counterproductive neuron quantity is limited, not up to the change of the polarity of h_i between which the neuron can still maintain or turn into the right state. This is the fundamental reason of the network why it can realize associative memory.

D. The training procedure in CA3

The training procedure does not require any iteration. It only uses the outer product of the input vector and its transpose (Equation (12)). The diagonal numbers in the matrix are possible states in which only 1 could be stored.

$$W = x \cdot x^T = \begin{bmatrix} x_1^2 & x_1 x_2 & \dots & x_1 x_n \\ x_2 x_1 & x_2^2 & \dots & x_2 x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n x_1 & x_n x_2 & \dots & x_n^2 \end{bmatrix} \quad (12)$$

If a lot of vectors need to be stored in the network, each row of the input vector matrix x could be an input vector (as shown in Equation (13)).

$$W = x^T x - mI \quad (13)$$

V. HMSNN: HIPPOCAMPUS INSPIRED MEMORY SPIKING NEURAL NETWORK

Based on the detailed description of DG model and CA3 model, we propose the HMSNN architecture for an integrated hippocampus model. As shown in Fig. 7, the model mainly contains two parts: the DG part for the information abstraction; the CA3 part for the information retrieval and storage.

VI. EXPERIMENTAL VALIDATION AND RESULTS

The MNIST dataset and robot walking dataset are used for validation. A state-of-the-art method, convolutional neural network, are used for comparative studies.

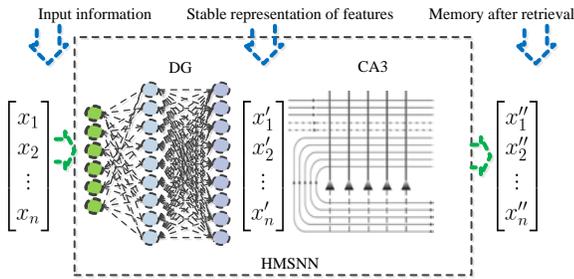


Fig. 7. The architecture of HMSNN.

A. Memory performance of HMSNN on MNIST

MNIST is a large database of handwritten digits, and is often used to train and test various image processing system. It contains 60,000 training images and 10,000 test images.

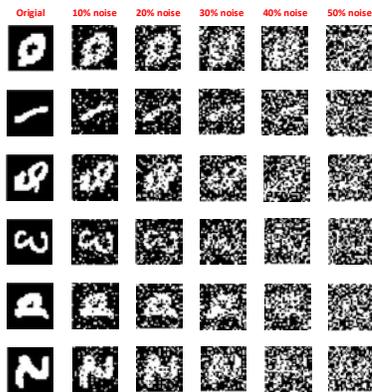


Fig. 8. Different degrees of noise adding on the original digital images (taking 0-5 as an example).

We test the proposed HMSNN model on the MNIST dataset by adding noise on the original images. Firstly, we train both HMSNN and CNN with images without noise. After the training procedure, we test them with images that contains noises. the test images with different proportion of noises are shown in Fig. 8. Consider that HMSNN is not a full classification network which is only a kind of stable memory abstraction and memory network, we go a small step further by making the AND mechanisms of the stable description of the test image with the standard golden rules to get the right class number.

For the stable representations, we have converted the vectors of them into 28×28 matrixs and shown them as images, as shown in Fig. 9. We can easily find stable representation in different distribution of classes even when different noises are introduced into the network.

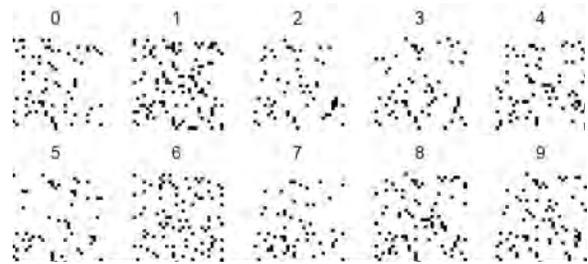


Fig. 9. Transform the stable representation vectors into images with 28×28 pixels.

Convolutional Neural Network (CNN) is very successful in image classification. The comparative study of HMSNN and CNN is shown in in Fig 10. We could conclude that CNN has bad classification performance on recognition task with noises, especially when the percentage of noises goes very high. Its recognition rate decreases very quickly when more noises are introduced into the images. While the memory mechanism of HMSNN makes it have much better performance compared to CNN when the percentage of noises increase. This shows the great advantages of memory storage and association than traditional feed forward networks.

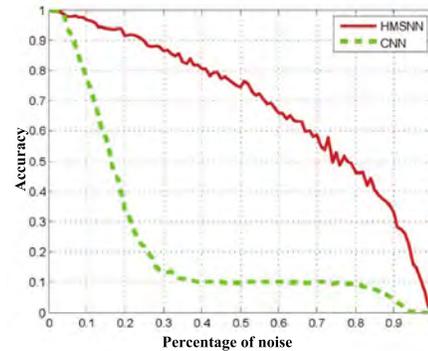


Fig. 10. A comparative study on the recognition performance between HMSNN and CNN.

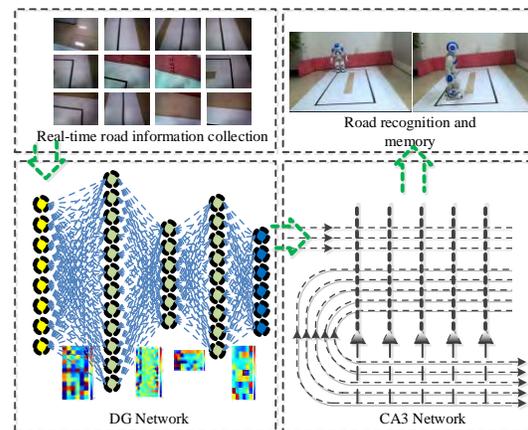


Fig. 11. HMSNN model for robot path recognition.

B. Memory performance of HMSNN in realistic road images.

We choose the dynamical robot path recognition task to test the memory and classification performance of the HMSNN model. As shown in Fig. 11, during the training process, the robot walks along a circle, collects the road tag images and makes the judgement in real time. The data set contains three different types of the road images and each image is with the size of 640×480 .

In the training procedure, the total number of training images is 1200 (60 seconds and 20 frames for each second). The robot extracts stable representations through DG in the HMSNN model and stores them in CA3. During the test procedure, robot recalls the stable representation of the network after the realtime image stimulus.

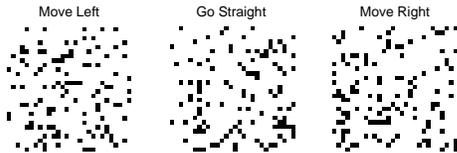


Fig. 12. Transform the stable representation vectors into 28×28 visual images.

HMSNN learns real time road images and forms three types of stable representations for road information in DG and CA3, as shown in the images (28×28 pixels) in Fig. 12.

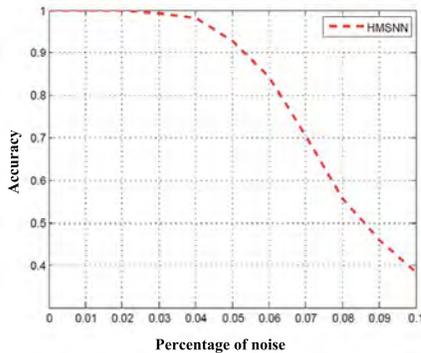


Fig. 13. The relationship between accuracy and the ratio of noise.

Fig. 13 shows the accuracy of classification on the road images with different proportions of noises in the test images. As the percentage of noises increases, the recognition accuracy of HMSNN decreases smoothly. This shows that the robustness and efficiency of HMSNN make it work well on the real road path recognition tasks.

VII. CONCLUSION

Understanding the computational nature of learning and memory is essential both for brain research and Artificial Intelligence. Although we only combine the feed forward network and recurrent network in this paper, it already shows its great potentials in learning tasks. We try to uncover the mysteries of memory by mimicking both the structural and functional mechanisms of the hippocampus. The proposed HMSNN

model is with both biological structural and mechanism basis and intelligent classification behaviors. Experimental results show that HMSNN could get better performance compared with CNN on learning with noise, information storage and retrieval tasks.

ACKNOWLEDGEMENT

This study was funded by the Strategic Priority Research Program of the Chinese Academy of Sciences (XD-B02060007), and Beijing Municipal Commission of Science and Technology (Z151100000915070, Z161100000216124).

REFERENCES

- [1] P. B. Sederberg, M. J. Kahana, M. W. Howard, E. J. Donner, and J. R. Madsen, "Theta and gamma oscillations during encoding predict subsequent recall," *The Journal of Neuroscience*, vol. 23, no. 34, pp. 10 809–10 814, 2003.
- [2] R. C. O'reilly and J. L. McClelland, "Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off," *Hippocampus*, vol. 4, no. 6, pp. 661–682, 1994.
- [3] J.-H. Han, S. A. Kushner, A. P. Yiu, H.-L. L. Hsiang, T. Buch, A. Waisman, B. Bontempi, R. L. Neve, P. W. Frankland, and S. A. Josselyn, "Selective erasure of a fear memory," *Science*, vol. 323, no. 5920, pp. 1492–1496, 2009.
- [4] W. Deng, M. Mayford, and F. H. Gage, "Selection of distinct populations of dentate granule cells in response to inputs as a mechanism for pattern separation in mice," *Elife*, vol. 2, p. e00312, 2013.
- [5] K. Fukushima, "A hierarchical neural network model for associative memory," *Biological cybernetics*, vol. 50, no. 2, pp. 105–113, 1984.
- [6] K. Kawamura and S. Gordon, "From intelligent control to cognitive control," in *Automation Congress, 2006. WAC'06. World. IEEE*, 2006, pp. 1–8.
- [7] S. W. Oh, J. A. Harris, L. Ng, B. Winslow, N. Cain, S. Mihalas, Q. Wang, C. Lau, L. Kuan, A. M. Henry *et al.*, "A mesoscale connectome of the mouse brain," *Nature*, vol. 508, no. 7495, pp. 207–214, 2014.
- [8] D. Song, R. H. Chan, V. Z. Marmarelis, R. E. Hampson, S. A. Deadwyler, and T. W. Berger, "Estimation and statistical validation of event-invariant nonlinear dynamic models of hippocampal ca3-ca1 population activities," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. IEEE*, 2011, pp. 3330–3333.
- [9] L. Seress, "Interspecies comparison of the hippocampal formation shows increased emphasis on the regio superior in the ammon's horn of the human brain," *Journal fur Hirnforschung*, vol. 29, no. 3, pp. 335–340, 1987.
- [10] A. Waddington, P. A. Appleby, M. De Kamps, and N. Cohen, "Triphasic spike-timing-dependent plasticity organizes networks to produce robust sequences of neural activity," *Frontiers in computational neuroscience*, vol. 6, 2012.
- [11] X. Xie and H. S. Seung, "Learning in neural networks by reinforcement of irregular spiking," *Physical Review E*, vol. 69, no. 4, p. 041909, 2004.
- [12] W. Maass, "On the computational power of winner-take-all," *Neural Computation*, vol. 12, no. 11, p. 2519, 2000.
- [13] J. Tejada, G. M. Arisi, N. Garcia-Cairasco, and A. C. Roque, "Morphological alterations in newly born dentate gyrus granule cells that emerge after status epilepticus contribute to make them less excitable," *PLoS one*, vol. 7, no. 7, p. e40726, 2012.
- [14] N. Burkitt, "A review of the integrate-and-fire neuron model: I. homogeneous synaptic input," *Biological Cybernetics*, vol. 95, no. 1, p. 1, 2006.
- [15] N. E. Sharkey and S. A. Jackson, *An internal report for connectionists*. Springer, 1995.
- [16] A. Destexhe, "Conductance-based integrate-and-fire models," *Neural Computation*, vol. 9, no. 3, pp. 503–514, 1997.
- [17] R. J. O'Brien, L. Lau, and R. L. Haganir, "Molecular mechanisms of glutamate receptor clustering at excitatory synapses," *Current Opinion in Neurobiology*, vol. 8, no. 3, p. 364, 1998.
- [18] Y. S. Abu-Mostafa and J.-M. S. Jacques, "Information capacity of the hopfield model," *Information Theory, IEEE Transactions on*, vol. 31, no. 4, pp. 461–464, 1985.