

# It may be time to perfect the neuron of artificial neural network

**Gang Liu**

A doctoral student at the Institute of Robotics and Intelligent Systems, School of Mechanical Engineering, Xi'an Jiao Tong University [Do this research for interest. The study will continue to be updated in the future. Welcome to follow me.]

E-mail: gangliu.6677@gmail.com

## Abstract

In recent years, artificial neural networks (ANNs) have won numerous contests in pattern recognition, machine learning, and artificial intelligence. The basic unit of an ANN is to mimic neurons in the brain. Neuron in ANNs is expressed as  $f(wx + b)$  or  $f(wx)$ . This structure does not consider the information processing capabilities of dendrites. However, recently, studies shown that dendrites participate in pre-calculation in the brain. Concretely, biological dendrites play a role in the pre-processing to the interaction information of input data. Therefore, it is time to perfect the neuron of the neural network. This paper, add dendrite processing section, presents a novel artificial neuron, according to our previous studies (CR-PNN or Gang transform). The dendrite processing section can be expressed as  $WA \circ X$ . Because I perfected the basic unit of ANNs— neuron, there are so many networks to try. This article gives the basic architecture for reference in future research.

Keywords: neural network, neuron, CR-PNN, Gang transform

## 1. Introduction

An artificial neural network (ANN) is an algorithmic architecture that imitates the biological brain [1, 2]. Today, neural networks are gradually changing our lives and making the world better [3]. First of all, I want to thank those who contributed to the present state of the art, e.g., Geoffrey Hinton, YannLeCun, Yoshua Bengio, and so on.

Recently, artificial neural networks (ANNs), particularly “deep learning” [4], have made some impressive recent advances, such as machine vision, speech recognition, autonomous vehicles, and machine translation. Thus, in the tech world today, optimism is high. Some people think ANNs are approaching human intelligence gradually, even it is still far away [1].

Neurons are the fundamental units in the biological brain. Accordingly, the fundamental units of ANNs are artificial neurons. With the development of technology, people have a

deeper understanding of the biological brain and neurons. However, the artificial neuron has maintained its original structure. It may be time to perfect the neuron of the artificial neural network.

The following presents a very brief introduction to the basic knowledge and latest developments in artificial neural networks and biological Neural Network.

### 1.1 Artificial neural networks

An ANN comprises many interconnected functional units or neurons and can learn the relationship of input space and output space. Thus, they usually are used to solve classification or regression problems [5]. Over the years, based on the application and the characteristics of the data involved, different architectures of ANNs have been developed. For example, Convolutional Neural Networks (CNN) [6, 7] in computer vision, Recurrent Neural Networks (RNN) [8] or Long Short Term Memory Network (LSTM) [9]

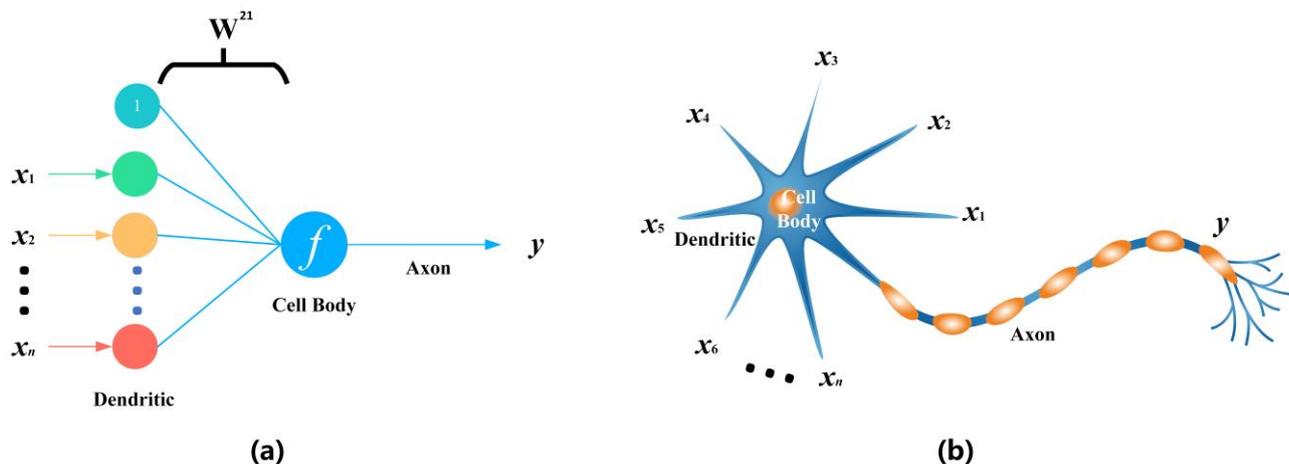


Fig. 1. Traditional neuron. (a) Traditional artificial neuron. (b) Traditional imitation to biological neuron.

in sequence and time series modeling, and Generative Adversarial Network (GANs) [10].

In the last decade, some newer architectures with newer learning algorithms are developed to endure the need to develop human-like efficient machines in different application areas. From the perspective of information dimension, it can be divided into a network focusing on spatial information [CNN, the representative field is computer vision (CV).] and a network focusing on temporal information [RNN, the representative field is natural language processing (NLP)]. *There are several newer CNN architectures and efficient mechanisms:* Alexnet [7], VGG [11], Googlenet [12], Inception-V3 [13], ResNet [14], ResNeXt [15], Convolutional Block Attention Module (CBAM) introduced by Woo et al. [16], and competitive squeeze-excitation (SE) mechanism introduced by Hu et al. [17]. *There are several newer RNN architectures and efficient mechanisms:* Deep Recurrent Attentive Writer (DRAW) [18], Grid Long Short-Term Memory [19], gating mechanism introduced by Jing et al. [20], and factorized recurrent network architecture introduced by Belletti et al. [21].

Besides, the learning of ANNs — by changing the distribution of weights to approximate the relationship between input and output space— has been studied. Lately, success was achieved in many techniques, such as L1 and L2 regularization [22], batch normalization [23], a good collection of weight initialization techniques [24, 25], and a good set of activation functions [5].

## 1.2 Biological Neural Network

The human brain has approximately 100 billion biological neurons, and neurons are connected via specialized structures known as synapses to sense stimulations and to pass signals to other neurons [26]. A neuron is the fundamental structural and functional unit in the neural information network and is composed of a cell body (soma), dendritic trees, and an axon

[27, 28]. The most extended parts of many neurons are dendrites [29], and the active electrical properties of dendrites shape neuronal input.

In the field of biology, researchers have been studying the mechanism of the element of neurons over the years [28, 30–32]. Recently, studies discovered a class of calcium-mediated dendritic action potentials (dCaAPs) [32, 33]. Here, we quote the original words in the literature: “These dCaAPs enabled the dendrites of individual human neocortical pyramidal neurons to classify linearly nonseparable inputs—a computation conventionally thought to require multilayered networks.” [32] The intersection of dendrites may exist computing. ***This means dendrites participate in pre-calculation in a biological neuron or biological neural network.***

Revisiting the previous ANNs, we found few studies highlight perfecting the neuron of ANNs. Fortunately, it happens that Gang et al. [34] proposed a module of fast calculation. The module can process linearly nonseparable inputs. Thus, it can simulate the function of dendrites and perfecting the neuron.

## 2. Artificial neuron

### 2.1 Traditional artificial neuron

In 1943, McCulloch and Pitts proposed ANN models with adjustable weights [35]. More than ten years later, Rosenblatt put forward the Perceptron Learning Algorithm [36]. In 1986, Rumelhart et al. proposed learning representations by back-propagating errors [37]. By then, a typical artificial neuron was established.

Figure 1 showed the traditional artificial neuron. This architecture of artificial neuron ignores the pre-calculation of Dendrites. If  $X$  is the input and  $f$  is the nonlinear activation function, the output  $y$  can be represented by,

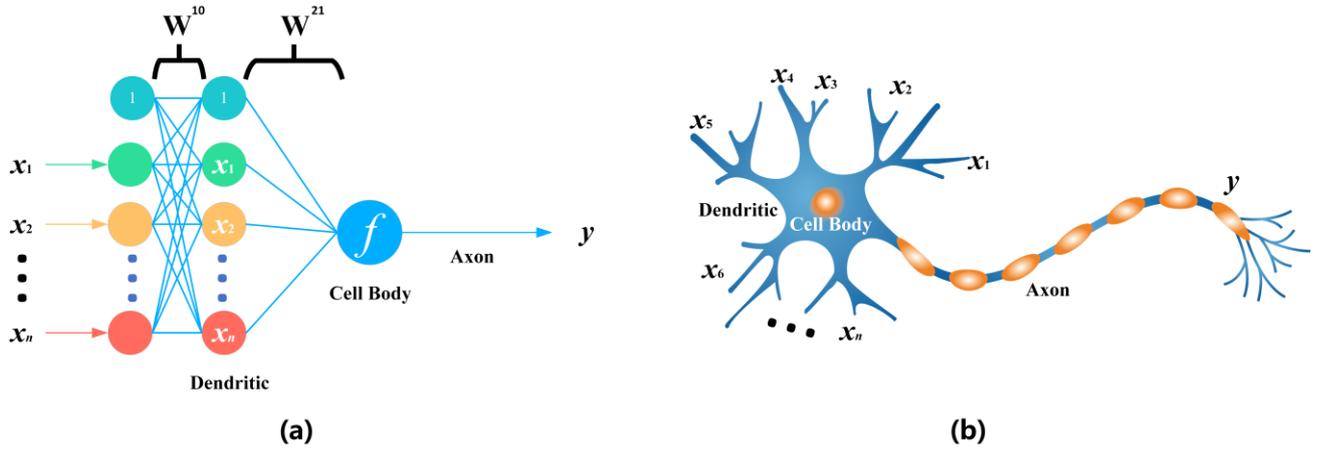


Fig. 2. Novel neuron. (a) Gang artificial neuron. (b) Gang imitation to biological neuron.

$$y = f(W^{21}X) \quad (1)$$

Where  $W^{21}$  is the weight matrix,  $\{1, x_1, x_2, \dots, x_n\} = X$ . In the backward pass, the weight matrix can be updated with,

$$W_{new} = W - \eta \frac{\partial E}{\partial W} \quad (2)$$

Where  $W_{new}$  are the updated weight matrix for  $W$ , and  $E$  is the cost function and  $\eta$  is the learning rate.

## 2.2 Novel Gang artificial neuron

For the biological neurons, the XOR operation is performed in the dendrites with dCaAPs [32, 33], and AND/OR operations are performed at the soma and at tuft and basal dendrites with sodium and NMDA spikes [38-40]. The XOR operation means or contains the brain pre-processing to the interaction information of input data. For understanding a picture task, It refers to the relationship across parts in an input-picture. For understanding an article or a speech task, It refers to the relationship across parts in an input-word. However, traditional artificial neurons only imitate the soma.

### 2.2.1 Dendrite processing section

In previous studies, inspired by Taylor series and Jiushao Qin or Horner Algorithm, we presented a module that upgrades the input order and introduces the interaction of input [34, 41]. The module is represented as follows.

$$A^i = W^{i,i-1} A^{i-1} \circ X \quad (3)$$

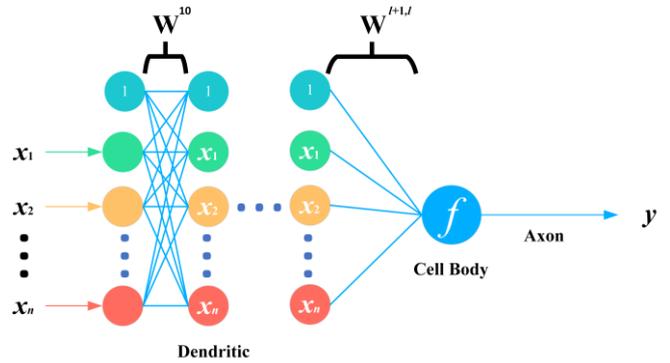


Fig. 3. Gang artificial neuron with multiple interactions.

Where  $A^{i-1}$  and  $A^i$  are the input and output of the previous layer, respectively,  $X$  is the original input. Coincidentally, It can simulate the dendrite processing section.

In figure 2a, we use one module to simulate the dendrite. The "dendrite" contains the interaction item across both input variables and can be represented as follows.

$$A = W^{10} X \circ X \quad (4)$$

Where  $A$  is the output of dendrite or the input of the cell body.

Additionally, we also can use more modules to simulate the dendrite (see Fig. 3). The "dendrite" contains the interaction item across multiple input variables (the number of modules plus 1). For example, if we use two modules to simulate the dendrite. The "dendrite" contains the interaction item across three input variables and can be represented as follows.

$$\begin{cases} A^1 = W^{10} X \circ X \\ A^2 = W^{21} A^1 \circ X \end{cases} \quad (5)$$

Where  $A^2$  is the output of dendrite or the input of the cell body.

The operation of the “dendrite” only uses matrix multiplication and Hadamard product, and due to Jiushao Qin or Horner Algorithm, and upgrading interaction item only need to add one module; thus, the computational complexity is lower [34, 41].

## 2.2.2 architecture of Gang neuron

### 2.2.2.1 Gang neuron with one module

Figure 2 shown the Gang artificial neuron with one module. The architecture of Gang artificial neuron can be represented as

$$\begin{cases} A = W^{10} X \circ X \\ y = f(W^{21} A) \end{cases} \quad (6)$$

Where  $W^{10}$  and  $W^{21}$  are the weight matrix,  $\{1, x_1, x_2, \dots, x_n\} = X$ . In the backward pass, the weight matrix can be updated with,

$$W_{new}^{21} = W^{21} - \eta^{21} \frac{\partial E}{\partial W^{21}} \quad (7)$$

$$W_{new}^{10} = W^{10} - \eta^{10} \frac{\partial A}{\partial W^{10}} \frac{\partial E}{\partial A} \quad (8)$$

Where  $W_{new}^{21}$  and  $W_{new}^{10}$  are the updated weight matrix for  $W^{21}$  and  $W^{10}$  respectively, and  $E$  is the cost function, and  $\eta^{21}$  and  $\eta^{10}$  are the learning rate. The learning rate of “dendrite” and cell body can be different.

### 2.2.2.2 Gang neuron of unicellular organism

In nature, there are unicellular organisms, such as bacteria and fungus. The first organisms to appear on Earth were presumably single-celled [42]. Thus, figure 3 shown the Gang artificial neuron with multiple interactions to imitate unicellular organisms. The architecture has been shown to be able to realize strong nonlinear expression [34, 41] and can be represented as follows.

$$\begin{cases} A^1 = W^{10} X \circ X \\ A^2 = W^{21} A^1 \circ X \\ \vdots \\ A^l = W^l A^{l-1} \circ X \\ y = f(W^{l+1,l} A^l) \end{cases} \quad (9)$$

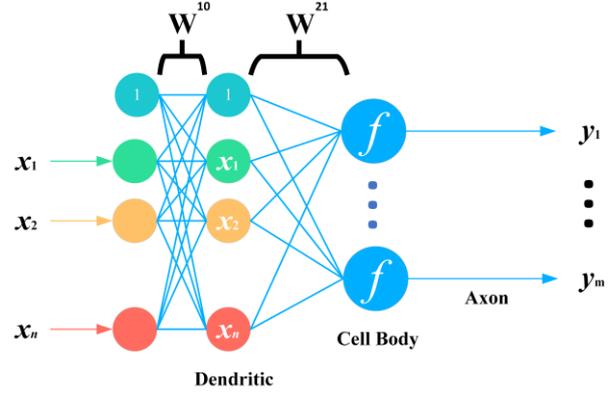


Fig. 4. Single-layer network.

Where  $l$  represents the number of modules to simulate the dendrite.

### 2.2.2.3 simplified Gang neuron

The weight matrix of the cell body  $W^{21}$  in Fig.2 or  $W^{l+1,l}$  is defined as a constant matrix  $W_{ones}$  where all its elements are 1. [Considering batch training, we described it as a matrix rather than a vector.]

Thus, the architecture in figure 2 can be represented as

$$y = f(W_{ones} (W^{10} X \circ X)) \quad (10)$$

And the architecture in figure 3 can be represented as

$$\begin{cases} A^1 = W^{10} X \circ X \\ A^2 = W^{21} A^1 \circ X \\ \vdots \\ A^l = W^l A^{l-1} \circ X \\ y = f(W_{ones} A^l) \end{cases} \quad (11)$$

## 3. ANNs using Gang artificial neuron

### 3.1 Single-layer network

Figure 4 shows the single-layer network. This architecture imitates the intricate dendrite of multiple neurons, and one information can be obtained by multiple neurons [32, 33, 42].

### 3.2 BPNN using Gang artificial neuron

The backpropagation neural network (BPNN) is the most typical neural network. Here, we show a BPNN using Gang artificial neuron with one module (see Fig. 5). The forward propagation can be represented as follows.

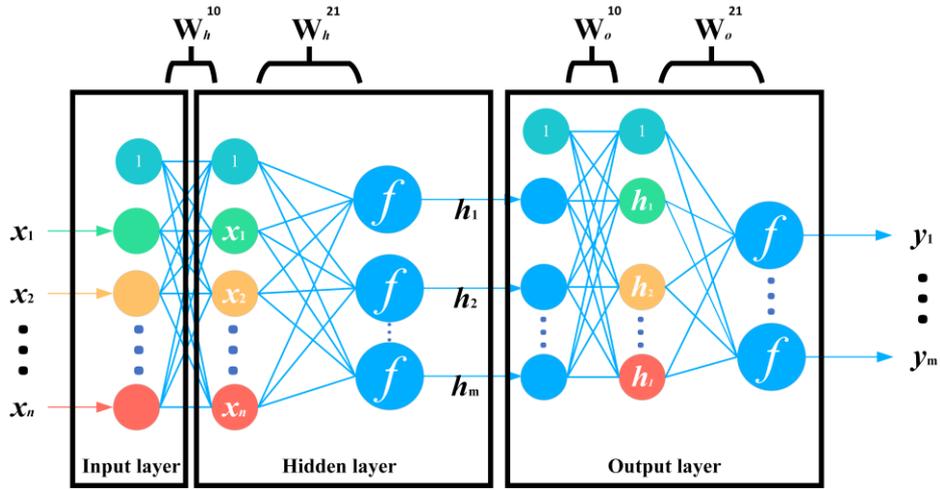


Fig. 5. BPNN using Gang artificial neuron.

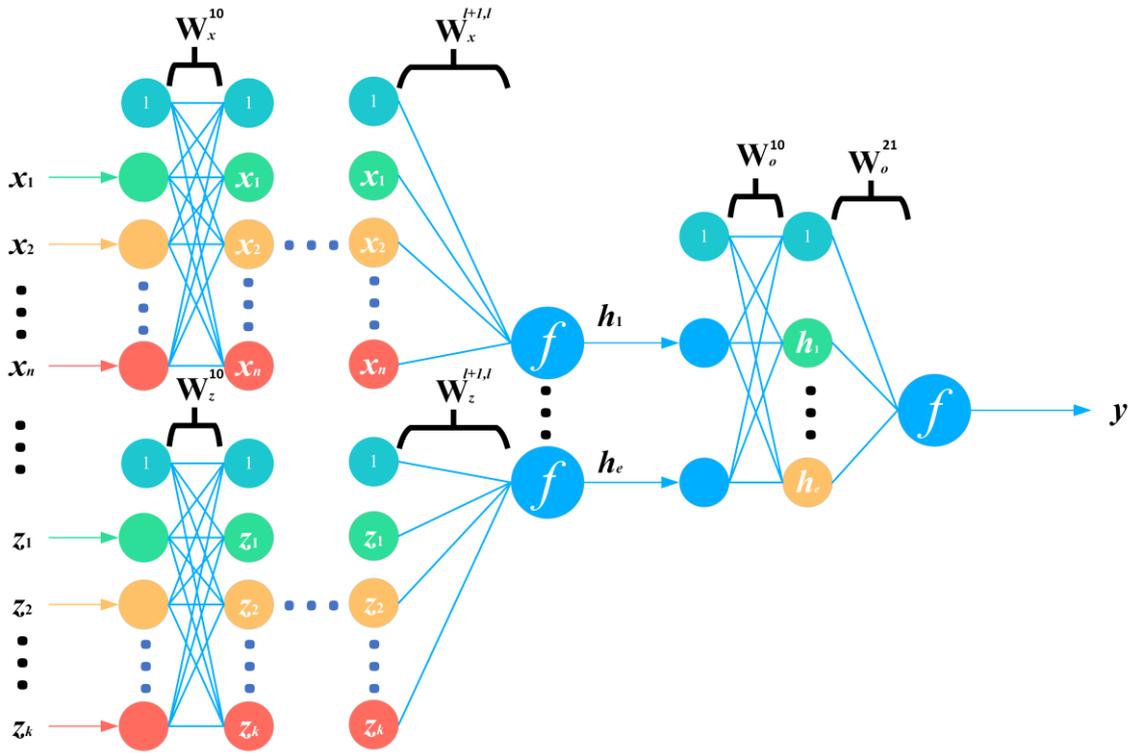


Fig. 6. Information fusion network.

$$\begin{cases} A_h = W_h^{10} X \circ X \\ H_h = f(W_h^{21} A_h) \end{cases} \quad (12)$$

### 3.3 Information fusion network

$$\begin{cases} A_o = W_o^{10} H \circ H \\ Y = f(W_o^{21} A_o) \end{cases} \quad (13)$$

Figure 6 shows the information fusion from different neurons. This kind of information interaction is common in the biological brain [38-40]. This architecture in Fig. 6 can be expressed as follows.

Where  $H=[1; H_h]$ ,  $W_h^{10}$ ,  $W_h^{21}$ ,  $W_o^{10}$ , and  $W_o^{21}$  are the weight matrix,  $\{1, x_1, x_2, \dots, x_n\}=X$ . The backpropagation is represented according to formula 7 and 8.

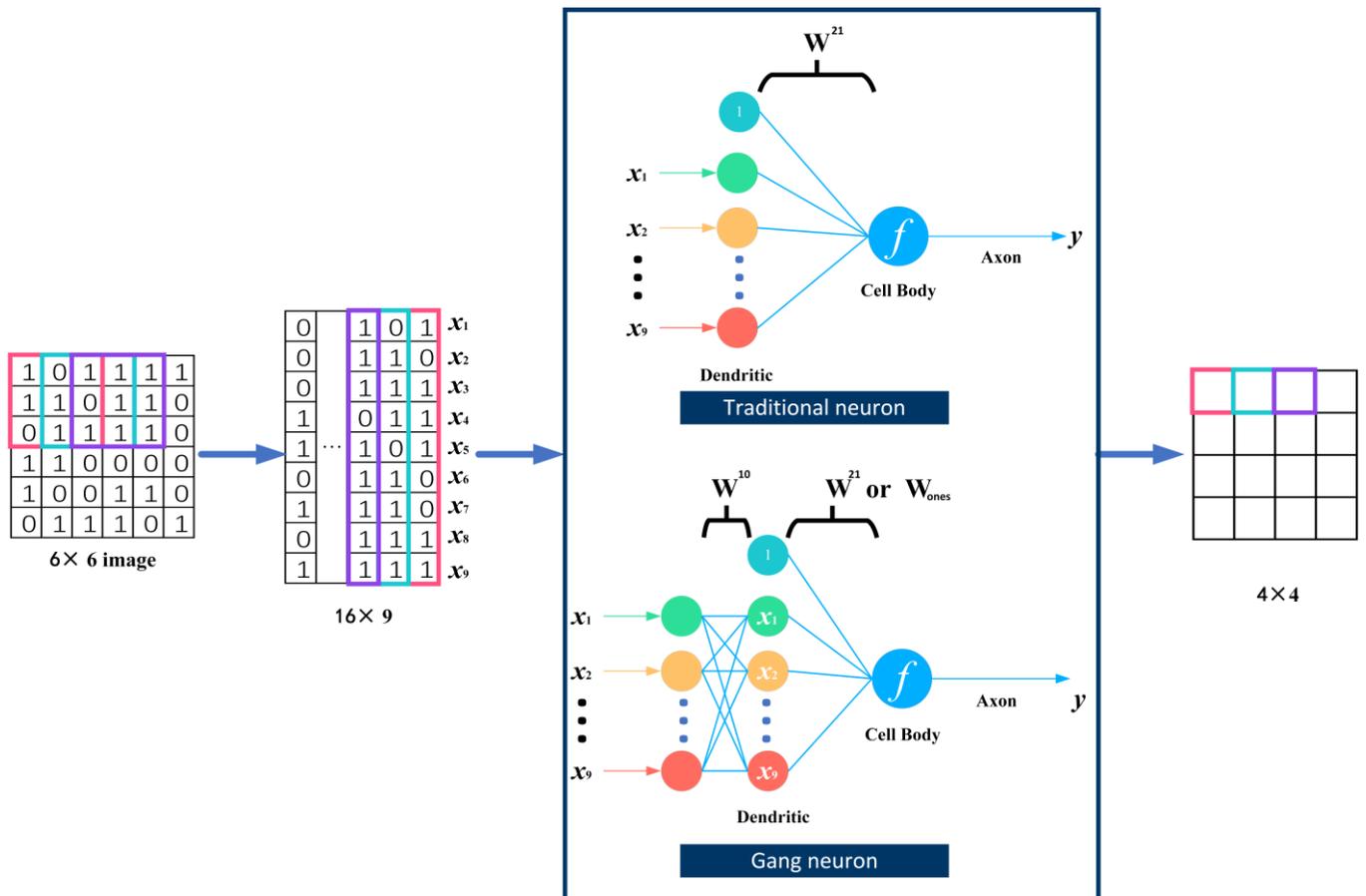


Fig. 7. Convolution layer using Gang neuron or traditional neuron.

$$\left\{ \begin{array}{l} A_x^1 = W_x^{10} X \circ X \\ A_x^2 = W_x^{21} A_x^1 \circ X \\ \vdots \\ A_x^l = W_x^{l,l-1} A_x^{l-1} \circ X \\ h_l = f(W_x^{l+1,l} A_x^l) \\ \vdots \\ A_z^1 = W_z^{10} Z \circ Z \\ A_z^2 = W_z^{21} A_z^1 \circ Z \\ \vdots \\ A_z^l = W_z^{l,l-1} A_z^{l-1} \circ Z \\ h_e = f(W_z^{l+1,l} A_z^l) \end{array} \right.$$

$$\left\{ \begin{array}{l} A_h = W_h^{10} H \circ H \\ y = f(W_h^{21} A_h) \end{array} \right. \quad (15)$$

Where  $\{1, x_1, x_2, \dots, x_n\} = X$ ,  $\{1, z_1, z_2, \dots, z_n\} = Z$ , and  $\{1, h_1, h_2, \dots, h_e\} = H$ .  $e$  is the number of neurons. The network

can learn by backpropagation is represented according to formula 7 and 8.

#### 4. Some typical ANNs architecture using Gang neuron

##### 4.1 Convolution layer using Gang neuron

(14) Figure 7 shows the comparison of using Gang neuron or traditional neuron in the convolution layer (an example). The Gang neuron extracts information about the interactions between inputs of space.

##### 4.2 RNN schematic diagram using Gang neuron

Figure 8 shows the comparison of using Gang neuron or traditional neuron in schematic diagram of RNN. The Gang neuron extracts information about the interactions between inputs of time. Where  $x$  represents the current input and  $h$  represents the output of the previous time.

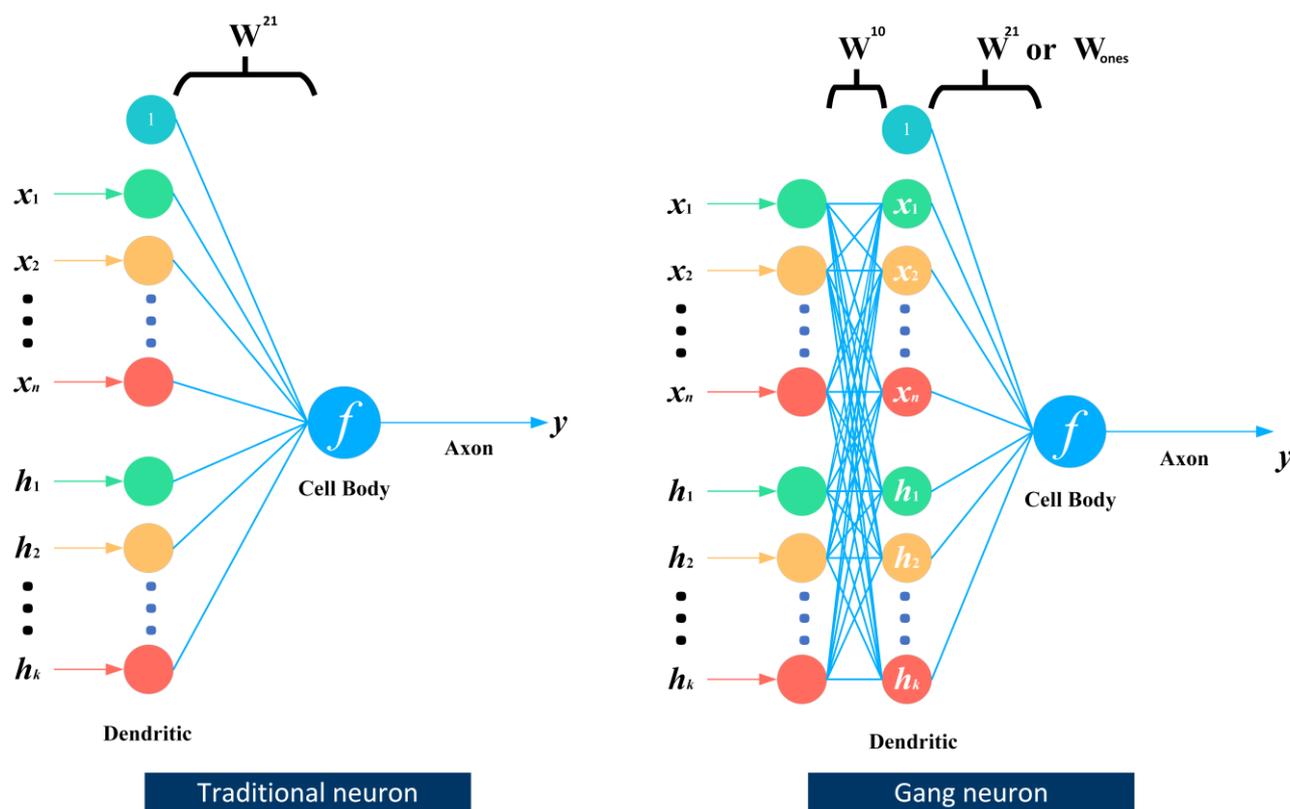


Fig. 8. Schematic diagram of RNN using Gang neuron

## 5. Discussion and conclusion

In this paper, we compared the development of artificial neural networks with biological neural networks. As expected, development directions are different. Researchers in ANNs paid more attention to the architectures aiming to different application areas, while researchers in biology focused on exploring the neural mechanisms. Both directions achieved good success. It may be time to learn from each other.

As we know, artificial neural networks are inspired by biological neural networks. Seventy years ago, people designed artificial neurons by imitating the knowledge about biological neurons at that time. Today, due to the development of biology, we have a relatively good understanding of how the work of neurons, especially dendrites.

We found, at the time of design, the traditional artificial neurons ignored a fact that dendrites participate in pre-calculation in a biological neuron or biological neural network. More specifically, biological dendrites play a role in the brain pre-processing to the interaction information of input data. This can be illustrated briefly by two tasks in life. For understanding a picture task, biological dendrites play a role in extracting the relationship across parts of an input-picture. For understanding an article or a speech task, biological dendrites play a role in extracting the relationship across parts in an input-word.

This paper perfected the neuron of ANNs by introducing the dendritic module and presented a novel neuron, named Gang neuron.

## 6. Outlook

In this paper, we have shown some basis architecture using Gang neuron. These architectures can be implemented easily. However, because I perfected the basic unit of ANNs—neuron, there are a lot of networks to try. I show my guess: traditional artificial neurons are insufficient in extracting the interaction information of input data. Thus we have designed a lot of convolutional layers. Gang neurons maybe reduce the number of layers in an existing network for the same task.

## References

- [1] A.M. Zador, A critique of pure learning and what artificial neural networks can learn from animal brains, *Nat. Commun.* 10 (2019) 7.
- [2] T.P. Lillicrap, A. Santoro, L. Marris, C.J. Akerman, G. Hinton, Backpropagation and the brain, *Nat. Rev. Neurosci.* 21(6) (2020) 335-346.
- [3] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Networks* 61 (2015) 85-117.
- [4] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521(7553) (2015) 436-444.

- [5] S. Sengupta, S. Basak, P. Saikia, S. Paul, V. Tsalavoutis, F. Atiah, V. Ravi, A. Peters, A review of deep learning with special emphasis on architectures, applications and recent trends, *Knowledge-Based Syst.* 194 (2020) 33.
- [6] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86(11) (1998) 2278-2324.
- [7] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, *Commun. ACM* 60(6) (2017) 84-90.
- [8] T. Mikolov, M. Karafiat, L. Burget, J. Cernocký, S. Khudanpur, Recurrent neural network based language model, conference of the international speech communication association, 2010, pp. 1045-1048.
- [9] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9(8) (1997) 1735-1780.
- [10] I. Goodfellow, J. Pougetabadie, M. Mirza, B. Xu, D. Wardefarley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Nets, neural information processing systems, 2014, pp. 2672-2680.
- [11] K. Simonyan, A. Zisserman, VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION, computer vision and pattern recognition, 2014.
- [12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, computer vision and pattern recognition, 2015, pp. 1-9.
- [13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the Inception Architecture for Computer Vision, computer vision and pattern recognition, 2016, pp. 2818-2826.
- [14] Y. Bengio, Learning Deep Architectures for AI, *Found. Trends Mach. Learn. (USA)* 2(1) (2009) 1-127.
- [15] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated Residual Transformations for Deep Neural Networks, computer vision and pattern recognition, 2017, pp. 5987-5995.
- [16] S. Woo, J. Park, J. Lee, I.S. Kweon, CBAM: Convolutional Block Attention Module, european conference on computer vision, 2018, pp. 3-19.
- [17] Y. Hu, G. Wen, M. Luo, D. Dai, J. Ma, Competitive Inner-Imaging Squeeze and Excitation for Residual Network, arXiv: Computer Vision and Pattern Recognition (2018).
- [18] K. Gregor, I. Danihelka, A. Graves, D.J. Rezende, D. Wierstra, DRAW: A Recurrent Neural Network For Image Generation, arXiv: Computer Vision and Pattern Recognition (2015).
- [19] N. Kalchbrenner, I. Danihelka, A. Graves, Grid Long Short-Term Memory, arXiv: Neural and Evolutionary Computing (2015).
- [20] L. Jing, C. Gulcehre, J. Peurifoy, Y. Shen, M. Tegmark, M. Soljagic, Y. Bengio, Gated Orthogonal Recurrent Units: On Learning to Forget, *Neural Comput.* 31(4) (2019) 765-783.
- [21] F. Belletti, A. Beutel, S. Jain, E.H. Chi, Factorized Recurrent Neural Architectures for Longer Range Dependence, international conference on artificial intelligence and statistics, 2018, pp. 1522-1530.
- [22] Y. Bengio, N. Boulangerlewandowski, R. Pascanu, Advances in optimizing recurrent networks, international conference on acoustics, speech, and signal processing, 2013, pp. 8624-8628.
- [23] G.E. Dahl, T.N. Sainath, G.E. Hinton, Ieee, IMPROVING DEEP NEURAL NETWORKS FOR LVCSR USING RECTIFIED LINEAR UNITS AND DROPOUT, 2013 Ieee International Conference on Acoustics, Speech and Signal Processing, Ieee, New York, 2013, pp. 8609-8613.
- [24] D. Sussillo, L.F. Abbott, Random Walk Initialization for Training Very Deep Feedforward Networks, arXiv: Neural and Evolutionary Computing (2014).
- [25] D. Mishkin, J. Matas, All you need is a good init, international conference on learning representations, 2016.
- [26] Z. Shi, S. Zhang, J. Yuan, B. Zhu, Y. Jiang, X. Shen, Y. Wang, Spatiotemporal summation and correlation mimicked in a four-emitter light-induced artificial synapse, *Sci Rep* 8(1) (2018) 2159-2159.
- [27] S. Sardi, R. Vardi, A. Sheinin, A. Goldental, I. Kanter, New Types of Experiments Reveal that a Neuron Functions as Multiple Independent Threshold Units, *Sci Rep* 7(1) (2017) 18036-18036.
- [28] J. Defelipe, L. Alonsonanclares, J.I. Arellano, Microstructure of the neocortex: Comparative aspects, *Journal of Neurocytology* 31(3) (2002) 299-316.
- [29] Y. Timofeeva, S. Coombes, D. Michieletto, Gap Junctions, Dendrites and Resonances: A Recipe for Tuning Network Dynamics, *Journal of Mathematical Neuroscience* 3(1) (2013) 15-15.
- [30] H. Mohan, M.B. Verhoog, K.K. Doreswamy, G. Eyal, R. Aardse, B. Lodder, N.A. Goriounova, B. Asamoah, A.B.C.B. Brakspear, C. Groot, Dendritic and Axonal Architecture of Individual Pyramidal Neurons across Layers of Adult Human Neocortex, *Cerebral Cortex* 25(12) (2015) 4839-4853.
- [31] Y. Deitcher, G. Eyal, L. Kanari, M.B. Verhoog, G.A.A. Kahou, H.D. Mansvelder, C.P.J. De Kock, I. Segev, Comprehensive Morpho-Electrotonic Analysis Shows 2 Distinct Classes of L2 and L3 Pyramidal Neurons in Human Temporal Cortex, *Cerebral Cortex* 27(11) (2017) 5398-5414.
- [32] A. Gidon, T.A. Zolnik, P. Fidzinski, F. Bolduan, A. Papoutsi, P. Poirazi, M. Holtkamp, I. Vida, M.E. Larkum, Dendritic action potentials and computation in human layer 2/3 cortical neurons, *Science* 367(6473) (2020) 83-+.
- [33] P. Poirazi, A. Papoutsi, Illuminating dendritic function with computational models, *Nat. Rev. Neurosci.* 21(6) (2020) 303-321.
- [34] G. Liu, J. Wang, A Polynomial Neural Network with Controllable Precision and Human-Readable Topology for Prediction and System Identification, arXiv e-prints, 2020, p. arXiv:2004.03955.
- [35] W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, *The bulletin of mathematical biophysics* 5(4) (1943) 115-133.
- [36] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain, *Psychological Review* 65(6) (1958) 386-408.
- [37] D.E. Rumelhart, G.E. Hinton, R.J. Williams, LEARNING REPRESENTATIONS BY BACK-PROPAGATING ERRORS, *Nature* 323(6088) (1986) 533-536.

- [38] M.E. Larkum, T. Nevian, M. Sandler, A. Polsky, J. Schiller, Synaptic Integration in Tuft Dendrites of Layer 5 Pyramidal Neurons: A New Unifying Principle, *Science* 325(5941) (2009) 756-760.
- [39] A. Polsky, B.W. Mel, J. Schiller, Computational subunits in thin dendrites of pyramidal cells, *Nature Neuroscience* 7(6) (2004) 621-627.
- [40] P. Poirazi, T. Brannon, B.W. Mel, Pyramidal neuron as two-layer neural network, *Neuron* 37(6) (2003) 989-999.
- [41] G. Liu, J. Wang, A Polynomial Neural network with Controllable Precision and Human-Readable Topology II: Accelerated Approach Based on Expanded Layer, arXiv e-prints, 2020, p. arXiv:2006.02901.
- [42] A.V. Ritchie, S. Van Es, C. Fouquet, P. Schaap, From Drought Sensing to Developmental Control: Evolution of Cyclic AMP Signaling in Social Amoebas, *Molecular Biology and Evolution* 25(10) (2008) 2109-2118.