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

Gang Liu
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. Neuron in ANNs is designed by the Knowledge about the biological neurons in the brain 70 years ago. Neuron in ANNs is expressed as $f(wx + b)$ or $f(WX)$. The design of this architecture does not consider the information processing capabilities of dendrites. However, recently, studies shows that dendrites participate in the pre-calculation of input data in the brain. Concretely, biological dendrites play a role in the pre-processing for the interaction information of input data. Therefore, it may be time to perfect the neuron of ANNs. According to our previous studies (Gang transform), this paper adds the dendrite processing section to neurons of ANNs. The dendrite processing section can be expressed as $W^{i,i-1}A^{i-1} \circ A^{0|1|2|\dots|i-1}$. The generalized new neuron can be expressed as $f(W(W^{i,i-1}A^{i-1} \circ A^{0|1|2|\dots|i-1}))$. The simplified new neuron be expressed as $f(\sum(WA \circ X))$. After perfecting the neuron, there are so many networks to try. This paper shows some basic architecture for reference in the future.

Interesting things: (1) The computational complexity of dendrite modules ($W^{i,i-1}A^{i-1} \circ A^{i-1}$) after being connected in series is far lower than Horner's method. Will this speed up the calculation of basic functions in computers? (2) The range of sight of animals has a gradient, but the convolution layer does not have this characteristic. This paper proposes receptive fields with gradient. (3) The networks using Gang neurons can delete the full connectional layer of traditional networks. In other words, the parameters of the full connectional layers are assigned to a single neuron, which reduces parameters of a network for the same mapping capacity.

Index Terms

Neural network, neuron, CR-PNN, Gang transform, Gang neuron, Receptive fields with gradient, Acceleration algorithm

CONTENTS

I	Introduction	2
I-A	Artificial neural networks	2
I-B	Biological neural network	2
II	Artificial Neuron	3
II-A	Traditional artificial neuron	3
II-B	Novel Gang artificial neuron	3
II-B1	Dendrite processing section ($\circ X$)	3
II-B2	Generalized Dendrite module ($\circ A^i$)	5
II-B3	Architecture-Gang neuron with one module	5
II-B4	Architecture-Gang "neuron" of unicellular organism	5
II-B5	Architecture-simplified Gang neuron	6
III	ANNs using Gang artificial neuron	6
III-A	Single-layer network	6
III-B	Multi-layer perceptron (MLP) using Gang artificial neuron	6
III-C	Information fusion network	7

A doctoral student at the Institute of Robotics and Intelligent Systems, School of Mechanical Engineering, Xi'an Jiao Tong University.

1. Do this research for interest.

2. The study will not be submitted to a journal. Thus, I retain copyright.

3. You can cite my research and apply Gang neuron to your study.

4. The study will continue to be updated in the future. I'm going to add a section: "Development and application of Gang neurons". If your research using Gang neuron interests me, I will cite yours and update the version.

5. Welcome to follow me. Some information can be found on my GitHub (later updated).

<https://github.com/liugang1234567/Gang-neuron>

Explanation video (I Keep updating it.): <https://space.bilibili.com/176499425/channel/detail?cid=137176>

Now that it is public, everyone can use it. Just cite and name it correctly, no need to email me. We are research people. Most of us are unknown and hard to do research. When we propose something that we feel proud of, we should tell people with pride. (Maybe only I think this is a good study, others do not think it is, but it is unimportant.) If we are not proud at this time, when should we be proud? Of course, our research may be like sand and flow away with time. (Most studies are like this.) However, at least I mentioned it before. I do not regret it. Please call it "Gang Neuron" when using it, please help me. Thank you. The early bird catches the worm.

IV	Some typical ANNs architecture using Gang neuron	9
IV-A	Convolution layer using Gang neuron	9
IV-B	Schematic diagram of RNN using Gang neuron	9
V	Biomimetic architecture using Gang neuron	9
V-A	Module 1: Retina,attention(focus),and convolution-like scan	9
V-B	Module 2: Population coding in brain	10
V-C	Application methods	10
VI	Discussion and conclusion	11
VII	Outlook	11
	Biographies	11
	Application letter	11
	References	12

I. 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 advances in many fields, 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 [2], [3].

Neurons are the fundamental units of 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 architecture. It may be time to perfect the neuron of the artificial neural network.

A brief introduction about the necessary knowledge and latest developments in artificial neural networks and biological neural networks is the following.

A. Artificial neural networks

An ANN consists of many interconnected functional units or neurons and can learn the relationship of input space and output space. Thus, they are usually used to solve classification and regression problems [5]. Over the years, based on the application and 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] in sequence and time series modeling, and Generative Adversarial Network (GANs) [10].

In the last decade, some newer architectures are developed to endure the need to develop human-like efficient machines in different application areas. From the perspective of informations, they 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].

B. 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.

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

In the field of biology, researchers have been studying the mechanism of the element of neurons over the years [27], [30]–[32]. Recently, a study 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] ***Computing exists in the intersection of dendrites. This means that dendrites participate in pre-calculation of a biological neuron or biological neural network.***

Reviewing the previous ANNs, I found few studies highlight perfecting the neuron of ANNs. Fortunately, it happens that Gang et al. [34] propose a module with low computational complexity named Gang transform. The module can process linearly nonseparable inputs by Hadamard product. Therefore, it can simulate the dendrites and perfect the neuron of ANNs.

II. ARTIFICIAL NEURON

A. Traditional artificial neuron

In 1943, McCulloch and Pitts proposed ANN models with adjustable weights [35]. More than a decade 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 in Dendrites. If X is the input and f is the nonlinear activation function, the output y can be represented by,

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

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

$$W_{new}^{21} = W^{21} - \eta \frac{\partial E}{\partial W^{21}} \quad (2)$$

Where W_{new}^{21} are the updated weight matrix for W^{21} , and E is the cost function or errors from the later layer in back-propagation and η is the learning rate.

B. 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/AND/OR operation means the pre-processing to the interaction information of input data. In the task of understanding a picture, it refers to the relationship between parts of an input-picture. It refers to the relationship between words of sentences in the task of understanding an article or a speech. However, traditional artificial neurons only simulate the soma.

1) *Dendrite processing section ($\circ X$):* In previous studies, inspired by Taylor series and Jiushao Qin’s or Horner’s method, we presented Gang transform and introduced interaction of inputs [34], [41]. The module of Gang transform is represented as follows.

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

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

Fig. 3. Generalized dendritic module. ' $\circ$ ' expresses 'or'. The dotted line represents any of them.

Where A^{i-1} and A^i are the input and output of the module, respectively, X is the original input. Coincidentally, it can simulate the dendrite processing.

In figure 2a, the dendrite is simulated by one module. The “dendrite” contains the interaction item between both inputs 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, the dendrite can also be simulated by using more modules. The “dendrite” in Fig.4 contains the interaction items between multiple inputs. The number of modules plus 1 is the number of interactive inputs. For example, if we use two modules to simulate the dendrite, the “dendrite” contains the interaction items among three inputs and can be represented as

Fig. 4. Gang artificial neuron with multiple dendrite modules.

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 consists of matrix multiplication and Hadamard product. Thus, the computational complexity is lower [34], [41]. Additionally, the mapping capacity of Gang neuron can be adjusted by the number of dendrite modules.

2) *Generalized Dendrite module ($\circ A^i$)*: Figure 3 shows a generalized dendritic module. The module is represented as follows.

$$A^i = W^{i,i-1}A^{i-1} \circ A^{0|1|2|\dots|i-1} \quad (6)$$

Where A^{i-1} and A^i are the input and output of the module, respectively, $A^{0|1|2|\dots|i-1}$ is any of A .

We can adjust the promotion degree or the number of interactive variables added for the single module by $A^{0|1|2|\dots|i-1}$. When we add a module, the degree of network/polynomial can increase by at least one and by i at most.

When $A^{0|1|2|\dots|i-1} = A^0$, the degree increases by 1 for one module, which is similar to Horner’s method. The degree increases with the number of modules linearly. For a certain degree of network, the computation complexity is $n - 1$, where n expresses the number of modules.

When $A^{0|1|2|\dots|i-1} = A^{i-1}$, the degree increases by i for one module. The degree increases with the number of modules exponentially. For a certain degree of network, the computation complexity is approximate $\log_2^n$, where n expresses the number of modules. This is similar to exponentiation by squaring. However, exponentiation by squaring is only used for calculation, and the modules in this paper are used for approximation. The weight matrix in Fig. 3 can be solved for by error backpropagation.

Here, I give a off-topic special USES. when this module is connected in series, it is used for polynomial calculation (Fig. 3). When $A^{0|1|2|\dots|i-1} = A^{i-1}$, there is no doubt that its computation complexity is far lower than Horner’s method. This can change many things.

The typical functions (e.g., $\sin$) are stored in the form of polynomial coefficients in the computer. The value of the function is calculated by polynomial when we invoke these functions.

We can solve for the polynomial coefficients of my new module ($A^{0|1|2|\dots|i-1} = A^{i-1}$) and store these coefficients in the computer. The value of the function is calculated by the network in Fig. 3 when we invoke these functions.

For the ease of description, this paper describes the Gang neuron with the best-represented architecture ($\circ X$). In use, you can replace ($\circ X$) by ($\circ A^{0|1|2|\dots|i-1}$).

3) *Architecture-Gang neuron with one module*: Figure 2 shows the Gang artificial neuron with one module. The architecture of Gang artificial neuron can be represented as

$$\begin{cases} A = W^{10}X \circ X \\ y = f(W^{21}A) \end{cases} \quad (7)$$

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

$$W_{new}^{21} = W^{21} - \eta^{21} \frac{\partial E}{\partial W^{21}} \quad (8)$$

$$W_{new}^{10} = W^{10} - \eta^{10} \frac{\partial A}{\partial W^{10}} \frac{\partial E}{\partial A} \quad (9)$$

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 or errors from the later layer in back-propagation. η^{21} and η^{10} are the learning rates. The learning rates of “dendrite” and cell body can be different.

4) *Architecture-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 4 shows the Gang artificial “neuron” with multiple interactions to imitate unicellular organisms. The architecture can realize strong nonlinear expression [34], [41] and can be represented as follows.

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

Where l represents the number of the dendrite modules.

Fig. 5. Single-layer network.

5) *Architecture-simplified Gang neuron*: The weight matrix of the cell body W^{21} in Fig.2 or $W^{l+1,l}$ in Fig. 4 is defined as a constant matrix where all its elements are 1. [Considering batch training, we describes it as a matrix rather than a vector.] Thus, the architecture in figure 2 can be represented as

$$y = f(W_{ones}(W^{10}X \circ X)) \quad (11)$$

And the architecture in figure 4 can be represented as

$$\begin{cases} A^1 = W^{10}X \circ X \\ A^2 = W^{21}A^1 \circ X \\ \vdots \\ A^l = W^{l,l-1}A^{l-1} \circ X \\ y = f(W_{ones}A^l) \end{cases} \quad (12)$$

III. ANNS USING GANG ARTIFICIAL NEURON

A. Single-layer network

Figure 5 shows the single-layer network. This architecture imitates the intricate dendrite of multiple neurons. Multiple neurons may obtain the same signal [32], [33], [42]. " $f(\dots)$ " in Fig. 5 can be set as the function that you plan to use for application. For example, when " $f(\dots)$ " is set as *softmax*, it can be used for classification.

B. Multi-layer perceptron (MLP) using Gang artificial neuron

MLP is the most typical neural network. Here, Fig. 6 shows a MLP using Gang artificial neuron with one module. The forward propagation can be represented as follows.

$$\begin{cases} A_h = W_h^{10}X \circ X \\ H_h = f(W_h^{21}A_h) \end{cases} \quad (13)$$

$$\begin{cases} A_o = W_o^{10}H_h \circ H_h \\ Y_o = f(W_o^{21}A_o) \end{cases} \quad (14)$$

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 can be represented according to formula 8 and 9.

Additionally, when the simplest form of Gang artificial neuron is used, MLP can be represented as follows.

$$\begin{cases} H_h = f(\sum(W_h^{10}X \circ X)) \\ Y_o = f(\sum(W_o^{10}H_h \circ H_h)) \end{cases} \quad (15)$$

Fig. 6. MLP using Gang artificial neuron.

Fig. 7. Information fusion network.

C. Information fusion network

Figure 7 shows the information fusion of different neurons. This information interaction is common in the biological brain [38]–[40]. The architecture in Fig. 7 can be expressed as follows.

Fig. 8. Convolution layer using Gang neuron or traditional neuron.

$$\left\{ \begin{array}{l} A_x^1 = W_x^{10} X \circ X \\ A_x^2 = W_x^{21} A_x^1 \circ X \\ \vdots \\ A_x^l = W_x^{l,l-1} A_x^{l-1} \circ X \\ h_1 = f(W_x^{l+1,l} A_x^l) \\ \vdots \end{array} \right. \quad (16)$$

$$\left\{ \begin{array}{l} A_z^1 = W_z^{10} X \circ X \\ A_z^2 = W_z^{21} A_z^1 \circ X \\ \vdots \\ A_z^l = W_z^{l,l-1} A_z^{l-1} \circ X \\ h_e = f(W_z^{l+1,l} A_z^l) \end{array} \right. \quad (17)$$

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

Where $\{1, x_1, x_2, \dots, x_n\} = X$, $\{1, z_1, z_2, \dots, z_k\} = Z$, and $\{1, h_1, h_2, \dots, h_e\} = H$. e is the number of neurons. The network architecture can be trained by error backpropagation according to formula 8 and 9.

In fact, these network architectures using Gang neuron can be trained by *pytorch* expediently. [Note that the input and output are normalized to $[-1, 1]$.] These architectures in this paper are like some plugins and can be used to other big architectures where you want to design to improve the mapping capacity of networks.

Fig. 9. Schematic diagram of RNN using Gang neuron.

IV. SOME TYPICAL ANNS ARCHITECTURE USING GANG NEURON

A. Convolution layer using Gang neuron

Figure 8 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.

The network using Gang neurons can delete the full connectional layers of traditional networks. In other words, the parameters of the full connectional layer are assigned to a single neuron, which reduces parameters of network for the same mapping capacity. **The full connectional layers do not exist in human brain.**

For example, the number of parameters in the full connectional layer is 10000×100 . When the parameters are assigned to a single neuron, the number of parameters can be $100 \times 10 \times 10$. [**100** expresses the number of Gang neurons.]

B. Schematic diagram of RNN using Gang neuron

Figure 9 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 about time. Where x represents the current input and h represents the output of the previous time.

V. BIOMIMETIC ARCHITECTURE USING GANG NEURON

A. Module 1: Retina, attention(focus), and convolution-like scan

Fig. 10 shows a retina module. This module imitates the focus mechanism of the human eye when we observe things. At the focus, our eye extracts the multiple relationships of the picture, and then, the extracted texture is more and more rough along the direction of focus diffusion. Additionally, **Attention** in Fig. 10 can scan the entire picture, which happens to be similar to convolution. However, my architecture contains gradients related to focus, and the maximum receptive field can be set as the entire picture (Of course, you can also set a small range of receptive field with gradient). I shows a 4th-order gradient in Fig. 10. The number of gradients can be adjusted by dendrite modules appropriately. For specific details, please see the specific formula in the Fig. 10.

Fig. 10. Module 1: Retina, attention(focus), and convolution-like scan. The gap between the 'squares' only aims to distinguish the 'squares' in drawing. The calculation of each color block consists of two steps (see Fig. 8): 1) Transform X into a 1-dimensional input; 2) use the dendritic modules. In terms of the activation function, you can use what you want, such as *relu*. Besides, linear activation function (e.g., $Y = WA$) also can be used because the dendritic module has the capacity of nonlinear mapping. Additionally, as an example, the figure shows the calculation process of one channel. For more channels, each channel can be calculated separately. Then, the outputs of channels are averaged in the same position.

Fig. 11. Module 2: Population coding in brain. Refer to Fig. 10 for the calculation of neurons. Besides, as an example, the figure shows the calculation process of one channel. For more channels, each channel can be calculated separately. Then, the outputs of channels are averaged in the same position.

B. Module 2: Population coding in brain

The dendrites of neurons are criss-crossed. In population coding of biological brain, the primary afferent fibre whose receptive field centre is closest to the point of stimulation will produce more action potentials than those on the periphery (Fig. 11, the green fibre has a greater action potential frequency than the blue fibres) [43].

Fig. 11 shows a biomimetic population coding module. Two points are important to emphasize.

1) Biomimetic dendrite is staggered; 2) When we design a network by multilayer Module 2 in Fig. 11, the dimension of output decreases as the layer.

C. Application methods

This paper shows three usage.

- 1) In the existing networks, the convolutional layers can be replaced by module 1. (Compared with the existing convolution, module 1 has a larger receptive field and can extract the position information.)
- 2) Multi-layer modules 2 are interconnected to build a new neural network.
- 3) Module 1 and Module 2 are used together. There is a special architecture by imitating the biological brain. The first layer is designed by module 1, and all the later layers are designed by module 2.

VI. DISCUSSION AND CONCLUSION

In this paper, I compared the development of artificial neural networks with biological neural networks. As expected, the direction of their research fields is different. Researchers in ANNs pay attention to the architectures for different applications, while researchers in biology focus on exploring the neural mechanisms. Both directions achieve 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, we have a better understanding of how the work of neurons, especially dendrites.

The traditional artificial neurons ignore 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 pre-processing to the interaction information of input data. Two examples can be mentioned in this regard. In the task of understanding a picture, biological dendrites play a role in extracting the relationship between parts of an input-picture. In the task of understanding an article, biological dendrites play a role in extracting the relationship between parts of an article.

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

VII. OUTLOOK

This paper shows some basis architecture using Gang neuron. In fact, these network architectures using Gang neuron can be trained by *pytorch* expediently. [Note that the input and output are normalized to $[-1, 1]$.] These architectures in this paper are like some plugins. They can be used to other big architectures where you want to design to improve the mapping capacity of networks.

There is an interesting phenomenon. Some data augmentation methods, such as flipping and zooming in or out pictures, improve networks' performance. I believe this shows that the existing networks can not fully extract some simple information, such as flipping and zooming in or out pictures. However, our brains do this well.

Application letter :

I'm Gang Liu, 26 years old, a 2nd year Ph.D. candidate from the Institute of Robotics and Intelligent Systems, Xi'an Jiao Tong University, P. R. China. I am crazy about math and computer. I plan to apply for a Chinese National Fund program, which will support those Ph.D. candidates in China to study in some world-class Universities. This program is a short time (6-24 months) co-education experiences.

If you're interested in my work and give me a chance to study in your team, please contact me.

REFERENCES

- [1] T. P. Lillicrap, A. Santoro, L. Marris, C. J. Akerman, and G. Hinton, "Backpropagation and the brain," *Nature Reviews Neuroscience*, vol. 21, no. 6, pp. 335–346, 2020.
- [2] A. M. Zador, "A critique of pure learning and what artificial neural networks can learn from animal brains," *Nature Communications*, vol. 10, p. 7, 2019.
- [3] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [5] S. Sengupta, S. Basak, P. Saikia, S. Paul, V. Tsilavoutis, F. Atiah, V. Ravi, and A. Peters, "A review of deep learning with special emphasis on architectures, applications and recent trends," *Knowledge-Based Systems*, vol. 194, p. 33, 2020.
- [6] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the Ieee*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the Acm*, vol. 60, no. 6, pp. 84–90, 2017.
- [8] T. Mikolov, M. Karafiát, L. Burget, J. Cernocký, and S. Khudanpur, "Recurrent neural network based language model," in *conference of the international speech communication association*, pp. 1045–1048, 2010.
- [9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Wardefarley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *neural information processing systems*, pp. 2672–2680, 2014.
- [11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *computer vision and pattern recognition*, 2014.
- [12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *computer vision and pattern recognition*, pp. 1–9, 2015.
- [13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *computer vision and pattern recognition*, pp. 2818–2826, 2016.
- [14] Y. Bengio, "Learning deep architectures for ai," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.
- [15] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *computer vision and pattern recognition*, pp. 5987–5995, 2017.
- [16] S. Woo, J. Park, J. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *european conference on computer vision*, pp. 3–19, 2018.
- [17] Y. Hu, G. Wen, M. Luo, D. Dai, and 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, and D. Wierstra, "Draw: A recurrent neural network for image generation," *arXiv: Computer Vision and Pattern Recognition*, 2015.
- [19] N. Kalchbrenner, I. Danihelka, and 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. Soljacic, and Y. Bengio, "Gated orthogonal recurrent units: On learning to forget," *Neural Computation*, vol. 31, no. 4, pp. 765–783, 2019.
- [21] F. Belletti, A. Beutel, S. Jain, and E. H. Chi, "Factorized recurrent neural architectures for longer range dependence," in *international conference on artificial intelligence and statistics*, pp. 1522–1530, 2018.
- [22] Y. Bengio, N. Boulanger-Lewandowski, and R. Pascanu, "Advances in optimizing recurrent networks," in *international conference on acoustics, speech, and signal processing*, pp. 8624–8628, 2013.
- [23] G. E. Dahl, T. N. Sainath, G. E. Hinton, and Ieee, *IMPROVING DEEP NEURAL NETWORKS FOR LVCSR USING RECTIFIED LINEAR UNITS AND DROPOUT*, pp. 8609–8613. International Conference on Acoustics Speech and Signal Processing ICASSP, New York: Ieee, 2013.
- [24] D. Mishkin and J. Matas, "All you need is a good init," in *international conference on learning representations*, 2015.
- [25] D. Sussillo and L. F. Abbott, "Random walk initialization for training very deep feedforward networks," *arXiv: Neural and Evolutionary Computing*, 2014.
- [26] Z. Shi, S. Zhang, J. Yuan, B. Zhu, Y. Jiang, X. Shen, and Y. Wang, "Spatiotemporal summation and correlation mimicked in a four-emitter light-induced artificial synapse," *Scientific Reports*, vol. 8, no. 1, pp. 2159–2159, 2018.
- [27] J. Defelipe, L. Alonsonanclares, and J. I. Arellano, "Microstructure of the neocortex: Comparative aspects," *Journal of Neurocytology*, vol. 31, no. 3, pp. 299–316, 2002.
- [28] S. Sardi, R. Vardi, A. Sheinin, A. Goldental, and I. Kanter, "New types of experiments reveal that a neuron functions as multiple independent threshold units," *Scientific Reports*, vol. 7, no. 1, pp. 18036–18036, 2017.
- [29] Y. Timofeeva, S. Coombes, and D. Michieletto, "Gap junctions, dendrites and resonances: A recipe for tuning network dynamics," *Journal of Mathematical Neuroscience*, vol. 3, no. 1, pp. 15–15, 2013.
- [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, and C. Groot, "Dendritic and axonal architecture of individual pyramidal neurons across layers of adult human neocortex," *Cerebral Cortex*, vol. 25, no. 12, pp. 4839–4853, 2015.
- [31] Y. Deitcher, G. Eyal, L. Kanari, M. B. Verhoog, G. A. A. Kahou, H. D. Mansvelder, C. P. J. De Kock, and I. Segev, "Comprehensive morpho-electronic analysis shows 2 distinct classes of I2 and I3 pyramidal neurons in human temporal cortex," *Cerebral Cortex*, vol. 27, no. 11, pp. 5398–5414, 2017.
- [32] A. Gidon, T. A. Zolnik, P. Fidzinski, F. Bolduan, A. Papoutsi, P. Poirazi, M. Holtkamp, I. Vida, and M. E. Larkum, "Dendritic action potentials and computation in human layer 2/3 cortical neurons," *Science*, vol. 367, no. 6473, pp. 83–, 2020.
- [33] P. Poirazi and A. Papoutsi, "Illuminating dendritic function with computational models," *Nature Reviews Neuroscience*, vol. 21, no. 6, pp. 303–321, 2020.
- [34] G. Liu and J. Wang, "A Polynomial Neural Network with Controllable Precision and Human-Readable Topology for Prediction and System Identification," *arXiv e-prints*, p. arXiv:2004.03955, Apr. 2020.
- [35] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.
- [36] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958.
- [37] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [38] M. E. Larkum, T. Nevian, M. Sandler, A. Polsky, and J. Schiller, "Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: A new unifying principle," *Science*, vol. 325, no. 5941, pp. 756–760, 2009.
- [39] A. Polsky, B. W. Mel, and J. Schiller, "Computational subunits in thin dendrites of pyramidal cells," *Nature Neuroscience*, vol. 7, no. 6, pp. 621–627, 2004.
- [40] P. Poirazi, T. Brannon, and B. W. Mel, "Pyramidal neuron as two-layer neural network," *Neuron*, vol. 37, no. 6, pp. 989–999, 2003.
- [41] G. Liu and J. Wang, "A Polynomial Neural network with Controllable Precision and Human-Readable Topology II: Accelerated Approach Based on Expanded Layer," *arXiv e-prints*, p. arXiv:2006.02901, June 2020.
- [42] A. V. Ritchie, S. Van Es, C. Fouquet, and P. Schaap, "From drought sensing to developmental control: Evolution of cyclic amp signaling in social amoebas," *Molecular Biology and Evolution*, vol. 25, no. 10, pp. 2109–2118, 2008.

- [43] J. Fitzakerley, "Sensory physiology," <https://www.d.umn.edu/~jfitzake/Lectures/DMED/SensoryPhysiology/GeneralPrinciples/LateralInhibition.html>, 2014.