#### PERSPECTIVE

# Towards NeuroAI: introducing neuronal diversity into artificial neural networks

Feng-Lei Fan<sup>1</sup> · Yingxin Li<sup>2</sup> · Tieyong Zeng<sup>1</sup> · Fei Wang<sup>3</sup> · Hanchuan Peng<sup>2</sup>

Received: 12 April 2024 / Revised: 23 September 2024 / Accepted: 8 October 2024 © The Author(s) 2025

# Abstract

Throughout history, the development of artificial intelligence, especially artificial neural networks, has been continuously influenced by a deeper understanding of the brain. This influence includes the development of the neocognitron, considered a precursor to convolutional neural networks. The emerging field of NeuroAI posits that leveraging neuroscience knowledge could significantly advance AI by imbuing networks with enhanced capabilities. Unlike the human brain, which features a variety of morphologically and functionally distinct neurons, artificial neural networks typically rely on a homogeneous neuron model. In the human brain, the diversity of neurons facilitates a wide range of intelligent behaviors. Given that artificial networks aim to mimic the human brain, incorporating a diversity of neuron models could address key challenges in artificial intelligence, such as efficiency, interpretability, and memory capacity. This perspective begins by examining the basics of biological neuronal diversity and how biological neurons transmit and process information. We then explore research efforts to design novel neuron models for artificial networks and discuss the potential benefits of neuronal diversity, including applications in several critical areas. Finally, we address the challenges and future directions for integrating neuronal diversity into artificial networks, highlighting its potential to enrich NeuroAI.

# Highlights

- NeuroAI is an emerging field advocating that neuroscience can always serve as a think tank in the development of AI.
- This perspective summarizes recent works in introducing neuronal diversity, the widely observed biological phenomenon in our brain, into deep learning.
- This perspective ponders the future directions and challenges of integrating neuronal diversity into artificial networks.

Keywords NeuroAI  $\cdot$  Artificial neural networks  $\cdot$  Neuronal diversity  $\cdot$  New neurons

Tieyong Zeng zeng@math.cuhk.edu.hk

- Fei Wang few2001@math.cuhk.edu
- Hanchuan Peng h@braintell.org
- <sup>1</sup> Department of Mathematics, The Chinese University of Hong Kong, Hong Kong, China
- <sup>2</sup> SEU-ALLEN Joint Center, Institute for Brain and Intelligence, Southeast University, Nanjing, Jiangsu Province 210096, P.R. China
- <sup>3</sup> Department of Population Health Sciences, Weill Cornell Medicine, Cornell University, New York, NY 10065, USA

#### Published online: 09 January 2025

# Introduction

We are experiencing the third wave of the artificial intelligence revolution. Deep learning, represented by deep artificial neural networks, has been dominating numerous important research fields in the past decade [1-3]. The tale of artificial neural networks dates back to mimicking nervous systems [4] in 1943, where McCulloch and Pitts abstracted a nervous system as a net of neurons, and each neuron was modeled with a "threshold logic" because the neuron exhibits an all-or-none behavior. Then, Hebb was engaged by the connectionism idea and proposed the rule of Hebbian learning—neurons that fire together should be wired together [5].



The Hebbian rule now still plays an important role in neural networks and computational neuroscience to this day. Later, Rosenblatt extended McCulloch and Pitts's idea to propose the Perceptron [6] that computes the inner product of the input and the Perceptron's internal parameters followed by a nonlinear activation. The Perceptron is the simplest network structure whose weights and bias are automatically learned by contrasting the outputs of the Perceptron and the target. However, the Perceptron was suggested to have an extremely limited expressive ability in Minsky and Papert's book [7], *i.e.*, the Perceptron cannot even execute the XOR logic. Hence, the research on the Perceptron, unfortunately, was dramatically slowed. Until 1980s, the designs of Hopfield networks [8] and Boltzmann machines [9] greatly aided the resurgence of interest in neural networks. With the reinvention of the backpropagation algorithm [10, 11], studies of neural networks were rapidly advancing: many novel network models (LSTM [12], autoencoder [13], neocognitron [14], CNN [15]) were established, and finally caught the imagination of the world in 2012 by the AlexNet [16]'s top performance on the ImageNet [17].

The brain is the most intelligent system we have ever known so far. Throughout history, the development of artificial neural networks has been open to and constantly inspired by the increasingly deepened understanding of the brain [18]. Although after drawing inspiration from neuroscience, artificial neural networks would usually follow their own paths to fit the demands of real-world tasks. In the past decades, due to tens of billions of dollars invested into neuroscience such as NIH BRAIN initiative,<sup>1</sup> a great amount of knowledge regarding the brain has been amassed, which can provide an ample source for applying principles of brain intelligence to artificial neural networks. Therefore, our opinion is that neuroscience is still critical to the advances of artificial neural networks, given the incomparable capability of the human brain and an ever-growing understanding of brain intelligence. Our opinion well aligns with the premise of an emerging interdisciplinary field—NeuroAI [18] whose overarching goal is to catalyze the next generation of AI by endowing a network with more human-like capabilities. For the time being, although a human brain and an artificial neural network serve fundamentally different purposes, it is crystal clear that the existing artificial network still goes far behind our human brain from the following engineering perspectives:

#### Efficiency

With the advent of big models, the size of a neural network becomes increasingly larger. To train such a model to a

practical point, well-curated big data and considerable power need to be supplied [19]. For example, the well-performing language model GPT-3 has 175 billion parameters, and its training requires hundreds of GPUs running a few months on 45 TB text data [20]. In contrast, a human brain performs its incredible feat by managing billions of neurons and coordinating trillions of connections at extremely low power (< presumably 20W) [21]. This gap is because our brain is an efficient concept learner that can learn complex objects from just a few examples [22]. The efficiency issue gains more and more traction in the background of combating global warming and carbon neutrality [23], as the AI models, particularly big language models are widely recognized as a significant carbon emitter [24]. Environmentalists have raised the criticism that oftentimes, the exhaustive trialand-error fine-tuning only leads to little performance gain. According to the carbon footprint computation, the training of the BERT model has a carbon footprint close to a person's one round-trip trans-America flight.<sup>4</sup>

# Interpretability

A neural network is notoriously a black box [25]. Although a network performs quite well in real-world tasks, it is hard to explain the underlying mechanism. Questions are often asked what is the function of certain neurons, layers, blocks, etc. and how they impact the model's decision-making. However, only limited success is achieved for these questions. Interpretability studies are divided into two branches [25]: post hoc interpretation and ad hoc interpretable modeling. It was argued that post hoc interpretation cannot be completely faithful to the original model because if it can be, it becomes the original model [26]. What's worse is one can hardly know the nuance between the post hoc interpretation and the original model. But ad hoc interpretable modeling may suffer from model expressibility in accomplishing transparency. As opposed to deep models, the decision process in the human is highly tractable. Modern neuroscience attributes conscious acts to electrical and chemical changes within and across neurons. Visualization of working regions and neurons spatially and temporally can also be carried out well by modern brain imaging techniques such as fMRI [27].

#### Memory

Catastrophic forgetting is a common issue of connectionist models [28, 29], *i.e.*, artificial neural networks are incapable of learning new information without forgetting what is

<sup>&</sup>lt;sup>1</sup> https://braininitiative.nih.gov

<sup>&</sup>lt;sup>2</sup> https://www.technologyreview.com/2019/06/06/239031/training-asingle-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifet imes/

previously learned. When a network is trained to learn consecutive tasks, what was learned from the previous task is easy to be interrupted by what is learned from the current task. This is because the weights trained for the earlier task have to be changed to meet the objective of the new task. The existing solutions to overcome catastrophic forgetting either require explicit retraining using the old data (so-called interleaved learning [30]) or only show efficacy in a specific type of memory or network structures [31]. Compared to the artificial network, our human brain has evolved effective mechanisms to avoid catastrophic forgetting [32], e.g., sleeping [33] can consolidate the outcome of the awakestate learning. Although these mechanisms remain to be completely understood, more and more experiments on the neuronal level showed that sleeping facilitates the potential of target neurons to be evoked [34, 35].

#### Robustness

Lacking robustness is another Achilles' heel of a neural network [36]. It was often reported that a neural network is easy to trick [37]. Sometimes, adding noise that is indiscernible to a human can completely change a network's prediction [38]. Later, it was shown that a neural network can be severely interfered with by common perturbation [39], such as occlusion, blur in an image, etc. Due to the widespread deployment of AI models in mission-critical scenarios, the robustness issue of neural networks has

received lots of attention. Compared to a neural network, humans usually won't get confused despite small changes such as mask, shift, distortion, natural corruptions, and stylish changes in an image.

To promote neural networks to a higher level of intelligence, based on NeuroAI, we believe that at present an actionable way is to explicitly identify the differences between brain and artificial neural networks and then make efforts to mitigate these differences. Clearly, the current mainstream neural network models are remarkedly different from the biological neural system, and one primary distinction is that neural networks lack the neuronal diversity that is everywhere in the human brain [40]. Different from artificial networks that are built on a single universal primitive neuron type, the brain has numerous morphologically and functionally different neurons [41]. With no exaggeration, neuronal diversity is an enabling factor for all kinds of intelligent behaviors [42]. More and more works showed the biological effect of neuronal diversity. Padmanabhan and Urban examined the outputs from a single type of neuron and the mitral cells of the mouse olfactory bulb, and found that diverse populations were able to code for twofold more information than their homogeneous counterparts using this intrinsic heterogeneity [43]. Since the artificial neural network is a miniature of the biological neural network, introducing neuronal diversity should be able to shed light on the aforementioned problems of the artificial neural network.

**Fig. 1** The rapid growth of the number of articles on the research on new neurons in deep learning. The data are based on the search in the Web of Science on December 12, 2022, with the time range from 2000 to 2022 according to the keyword "Deep Learning New Neurons"



Fig. 2 Neurons exhibiting an extraordinary morphological diversity. The shape classifications of neurons include unipolar, bipolar, pseudounipolar, multipolar, and so on



# **Biological neuronal diversity**

The extraordinary neuronal diversity originates from neuron differentiation [44], a complicated process to obtain different types of neurons that motivates many molecular signals to drive electrophysiological, morphological, and transcriptional changes in a neuron. Neuronal diversity is reflected by different molecular, morphological, physiological, connectional features [41], and so on. Here, we briefly introduce morphological and functional diversity: the former is the most obvious diversity, and the latter is instrumental to understanding different functionalities of a brain.

#### Morphological diversity

As a defining characteristic of neuronal types, morphological diversity, as shown in Figs. 1 and 2, includes diversity in projection patterns, bifurcation patterns, the density of branches, etc. According to the structural differences found in microscopy, neuroscientists previously have roughly divided neurons into four categories: unipolar, bipolar, multipolar, and pseudo-polar. Unipolar and pseudo-polar neurons have only one neurite extending from the soma, while that of a pseudo-polar neuron will soon split into two branches with a T-shaped structure directing to peripheral receptors and central spinal cords, respectively. While bipolar neurons extend protrusions from each end of the soma and evolve into dendrites and axons, respectively, multipolar neurons have one axon and several dendrites, which are the most common neuron type, constituting many complex central neural networks. Furthermore, more precise morphology studies reveal more concrete diversity from morphometry features to projection patterns on the regional level [41, 45].

# **Functional diversity**

Functional diversity can be seen at structural, neuronal, systematic, and behavioral levels. Neuronal functional diversity is a natural result of evolution in order to conduct complex behaviors in human life. Let us take three basic functions as examples to explain.

Med-X

(2025) 3:2

A human has five basic senses: sight, smell, touch, taste, and hearing, along with other senses including balance, proprioception, interior space emotion, time spiciness, etc. Hence, there exist sensory receptors that can specifically translate stimuli of different forms such as temperature, light, force, and sound, and then transmit impulses to the central nervous system. Once a sensory stimulus is applied to the exposed body part where the dendrite of the corresponding receptor neuron exists, the signal cascade and active signal pathway will be transferred, respectively. Then the functional sensory neural network starts to work. For example, as the human nose smells the flower, the olfactory sensory neurons located in the olfactory mucosa will be fired and evoke the olfactory network. Then a series of reactive events will happen.

Brain regions are often tied with functionalities. There are multiple brain regions whose neurons are involved in motion. Those motor neurons usually correspond to certain muscles. The more agile the body parts are, the more delicate the movement is, and the bigger will the corresponding neuron population grow to. Despite the target diversity, neurons from different motor regions also differ in function with regard to the process and degree of motion. For example, neurons in the primary motor area will lead to the simple movement of body muscles on the opposite side after the stimulus. The premotor area related to the gross shrinkage of muscle is all over the body, responsible for some aspects of motor control such as movement preparation, sensory instruction, or direct control of certain movements. The supplementary motor area serves a number of motor suggestion functions including internally generated motor planning, sequence planning of movements, and coordination of both sides of the body. More subtle mechanisms of delicate motor have been studied in a long term with technique on different levels [46].

The GPS system in our brain is attributed to place neurons, grid neurons, direction neurons, boundary neurons, speed neurons, etc. These neurons coordinate together so that our brain knows where we are, where we are heading, and how far we move. When entering a particular place, a place neuron [47] is fired. However, place neurons alone cannot explain how a human navigates the environment. A coordinate system is established after grid neurons [48] are respectively activated, as one traverses a set of small regions. These roughly equal regions are arranged in a periodic array to cover the entire open environment. Together with boundary neurons [49] encoding the space borderline, a mental map of the physical world is built, which allows for global precise positioning. In addition, neurons such as direction neurons [50], speed neurons [51], and angular head velocity neurons [52] inform the brain of characters of the motion in the environment and help modulate it. For example, when one makes a turn at a corner of a street, the direction neuron is fired to monitor the direction of movement. If one is in a hurry to arrive at the destination, the speed neuron is activated to supply the speed information.

The functional diversity of neurons is everywhere in the brain. In fact, a basic observation regarding the brain is that all complicated intelligence behaviors are a consequence of motivating different types of neurons. For example, different navigation neurons collaborate complementarily to realize all necessary functionalities for navigation.

# **Characteristics of bilogical neurons**

Per the premise of NeuroAI, one should retrospect how the information is transmitted and processed throughout a biological neuron [53] before introducing new neurons into artificial networks. To endow a network with more human-like capabilities, the neural mechanism should first be discussed. Roughly, it is divided into three steps: signal transduction (inbound), compartmentalized dendritic computation, and signal transduction (outbound). Our retrospection is centered around the macroscopic mechanisms and hallmarks relevant to neural computation. Unless necessary, we will not go into the level of cellular molecular biochemistry.

#### Signal Transduction (Inbound)

Neurons can transduce almost all types of physical signals into electrical signals, e.g., optical [54, 55], mechanical [56], biochemical [57]. Within the biological neural network, the most common is biochemical signal transduction. This transduction of a neuron is carried out by two categories of transducers of the current neuron: Membrane receptor-mediated type and nuclear receptor-mediated type. Two classic subcategories of the former are ligand-gated ion channels and transmembrane G-protein coupling receptors. The former [58] will open to allow the ion flux via a conformational change when the receptor binds a chemical messenger such as neurotransmitters; the latter [59] will bind with neurotransmitters and activate coupling G proteins on the membrane, which either directly causes the ion channel to open via protein-protein interaction or activates the enzyme that expedites the synthesis of the intracellular second messenger towards a lagged [60], cascaded [61], and longer-lasting [62] modulation<sup>3</sup> to ion channels. Finally, an electrical signal will be stimulated and further processed.

#### **Dendritic computation**

In the early days, dendritic trees were believed to solely receive and transport information to the axon (passive). The pioneer of modern neuroscience once asked "Why do dendritic trees even exist" [63]? However, with the progress of sharp electrodes [64], ever-growing evidence suggests that the dendritic branches are compartmentalized functional units (active). They play a fundamental role in many key computations such as coincidence detection, detection of motion direction, and storage of multiple input features (Chapter 15, [65]). Thus, a dendritic branch is a compartmentalized computation unit enabled and defined by the passive and active properties. The integrated dendritic computation, therefore, is more complicated and needs better understanding.

#### **Passive properties**

Due to the intrinsic resistance, the most salient passive property of dendrites is its attenuation effect whose role is to keep the signal local and sparse. The attenuation rate decreases with the diameter and increases with the length [66, 67]. A branchpoint, which is a bifurcation between the apical trunk and oblique dendrite, also serves as a strong attenuator to restrict the active signal transmission. In addition to the attenuation effect, dendritic morphology can

<sup>&</sup>lt;sup>3</sup> https://openbooks.lib.msu.edu/neuroscience/chapter/neurotransmitter-action-g-protein-coupled-receptors/

affect the firing rate and firing pattern as well [68, 69]. Furthermore, the signal propagates into distal dendritic arbors more readily than towards the soma (Chapter 14, [65]). Such an asymmetry can empower the neuron with the ability to infer the direction of motion based on the activation order of the somatic response. Finally, the dendritic structure is shown to facilitate the sparse coding in biological networks [70] that are associated with energy efficiency [71], feature discrimination [72], and memory capacity [73].

#### **Active properties**

A dendrite is embedded with a variety of ionic channels that enable powerful signal processing abilities and assist the information transmission intracellularly.

- i) Voltage-gated channels are nonlinear functional devices. For example, no action potential is generated when the input current is below a certain threshold. When the input current exceeds the action potential threshold, plenty of ion channels will open and generate an abrupt change in membrane potential. The action potential cancels the attenuation effect of the morphology and ensures the reliable propagation of the signal [74]. The dendritic spikes can exhibit complicated input-output relations other than the all-or-none relation. Recently, Gidon et al. [75] showed that the human layer 2/3 cortical neurons produce maximal activation when a stimulus is close to zero, and the activation is dampened when the stimulus gets stronger. Interaction between ion channels adds another layer of nonlinearity, e.g., calcium-dependent potassium channels [76] link potassium channels with calcium channels, which realizes a more precise regulation for neuronal excitability.
- ii) Voltage-gated channels are highly discriminatory, which enhances the dendrites' computation power. Typically, voltage-gated channels are only open to one kind of ions over another. As such, those channels are named after the most easily passed ions. Furthermore, for the same type of ion channels, there exist a variety of subtypes that differ in their voltage dependence, kinetics, singlechannel conductance, and so on. Different ion channels present different firing patterns [74].
- iii) Active dendrites respond in a location-dependent and time-dependent manner, implicating their spatiotemporal processing ability [77, 78]. For example, dendritic sodium spikes are propagatable throughout the dendritic tree [79], while the distal apical trunk tends to initialize calcium spikes that can only spread to the apical dendritic tree [80, 81]. In addition, both sodium and potassium spikes participate in coincidence detection [79, 82, 83]. For example, the potassium spike was activated

when distal and proximal dendritic regions of cortical neurons are synchronously activated [83].

#### Signal transduction (Outbound)

Transmitting information from a neuron's interior to its exterior has two fundamentally different modes. One is the direct electrical transfer via a specialized interconnected connection called the gap junction [84]. The transfer through gap junctions is very fast. The other mode is chemical transfer via synapses, *i.e.*, the presynaptic neuron releases the socalled neurohormone [85] or neurotransmitter [86] that will diffuse to the target neuron. The chemical transfer converts the electrical signal into the chemical release.

- The chemical transfer can exert different effects, depending on the types of released neurotransmitters [87]: The excitatory neurotransmitters fire the target neuron, while the inhibitory ones inhibit the target neuron, and modulatory neurotransmitters affect the effects of other chemical messengers. In most situations, one synapse can only be either excitatory or inhibitory.
- ii) The chemical transfer is subjected to random noise [74]. Even when the action potential is absent, a small depolarization is recorded in the target neuron due to the spontaneous random release of a small number of neurotransmitters from the presynaptic terminal.
- iii) The relation between the level of action potentials and the amount of the released neurotransmitter is nonlinear. When the voltage is low, only a few calcium channels are open. The high calcium concentration is achieved near those open channels, thereby only facilitating the fusion of nearby vesicles. When the voltage slightly increases, more channels open; therefore, a relatively small and uniform calcium concentration emerges. But the calcium level may not be sufficiently high to trigger subsequent processes [88].

Overall, despite the complicated information processing of a biological neuron, several salient characteristics should be noted: i) nonlinearity ubiquitously exists in every stage of information processing, not just in membrane channels; ii) a dendrite not only serves information transmission but also serve as compartmentalized computation unit; iii) neurons have the spatiotemporal information ability.

Figure 3 shows what kinds of computational properties of biological neurons remain unseen in the current mainstream artificial neurons. We divide the operations of artificial neurons into two stages: feature aggregation and nonlinear activation. It is widely believed that the nonlinear activation function in artificial neurons corresponds to the voltage/ligand-gated channels of biological neurons. However, in biological neurons, except for the outbound signal



**Fig.3** A schematic representation illustrating the computational properties of biological neurons that remain largely unseen in current mainstream artificial neurons. We categorize the operations of artificial neurons into two stages: feature aggregation and nonlinear activation. In biological neurons, in addition to outbound signal transduction, voltage and ligand-gated channels facilitate information transmission and processing intracellularly, rather than emitting information extracellularly. The outbound signal transduction of biological neurons corresponds to the nonlinear activation of artificial neurons, while the inbound signal transduction and dendritic computations collectively resemble the feature aggregation performed by artificial neurons

transduction, the roles of voltage/ligand-gated channels are to assist the information transmission and processing intracellularly, instead of emitting information extra-cellularly. Therefore, we think that the outbound signal transduction of biological neurons corresponds to the nonlinear activation of artificial neurons, while the inbound signal transduction and dendritic computation together are in analogy to the feature aggregation of artificial neurons.

# Neuronal diversity in artificial neural networks

In this section, we classify the studies of introducing neuronal diversity into four categories: activation design, polynomial neurons, dendritic neurons, and spiking neurons. Note that complex-valued neurons [89, 90] are not included because its idea is based on the need of addressing complex-valued data.

#### Activation design

Nowadays, the mainstream neuron type computes the inner product between the input and the connectivity parameters of upper neurons followed by a nonlinear activation function, which is mathematically formulated as

$$\sigma(\boldsymbol{x}^{\mathsf{T}}\boldsymbol{w}+\boldsymbol{b}),\tag{1}$$

where x is the input, w, b are parameters, and  $\sigma$  is the activation function. Past years have witnessed a plethora of novel activation functions being designed [91]. Three different properties are often considered in developing a novel activation function: i) As mentioned earlier, an activation function should be nonlinear, which is a necessity to guarantee a network has a sufficient discrimination ability. Nonlinearity is also bioplausible, *i.e.*, the outbound signal transduction in a biological neuron is nonlinear. ii) An activation function should allow a normal gradient flow across layers when a network is deep, *i.e.*, no gradient vanishment and explosion. iii) It should facilitate information transmission to expedite the extraction of useful features from data. Table 1 summarizes several representative activation function designs. Now, we illustrate them in detail.

#### Logistic sigmoid/tanh

The logistic sigmoid [92] was extensively used in the early stage of neural networks. However, its employment in a deep network suffers gradient vanishment and poor convergence. The gradient is killed when the pre-activation value is super high or low, while the non-zero-centered nature forces the convergence trajectory to go zig-zag [91]. Tanh is of zero-centric nature, but its computational complexity is high, and still subjected to the gradient issue. Several variants of Tanh were proposed to enlarge the range of output function [15] and overcome the gradient vanishment problem [93, 94, 106].

 Table 1
 A summary of different types of activation functions

/pe	Representative Example	Mathematical Expression
gmoid/Tanh Type	Sigmoid [92] PSF [93]	$\frac{\frac{1}{1+e^{-x}}}{\frac{1}{(1+e^{-x})^m}}$
	ISigmoid [94]	ISigmoid(x) = $\begin{cases} a(x-a) + 1/(1 + e^{-a}), x \le -a \\ 1/(1 + e^{-x}), -a < x < a \\ a(x+a) + 1/(1 + e^{-a}), x \ge a \end{cases}$
	Tanh [95]	$(e^x - e^{-x})/(e^x + e^{-x})$
	scaled Tanh [15]	$A(e^{Bx}-e^{-Bx})/(e^{Bx}+e^{-Bx})$
EU Type	ReLU [96]	$\operatorname{ReLU}(\mathbf{x}) = \begin{cases} x, x \le 0\\ 0, x < 0 \end{cases}$
	Leaky-ReLU [95]	Leaky ReLU(x) = $\begin{cases} x, x \ge 0\\ \alpha \cdot x, x < 0 \end{cases}$
	Concatenate-ReLU [97]	[ReLU(x), ReLU(-x)]
	GenLU [98]	GenLU(x) = $sgn(x)max\{ x +b,0\}$
	Randomly Translational ReLU [99]	$RTReLU(x) = \begin{cases} x + a, x + a \ge 0\\ 0, x + a \le 0 \end{cases}$
	ELU [100]	$ELU(x) = \begin{cases} x, x \ge 0\\ \alpha(e^x - 1), x < 0 \end{cases}$
	SReLU [101]	SReLU(x) = $\begin{cases} t^{r} + a^{r}(x - t^{r}), x \ge t^{r} \\ x, t^{l} \le x \le t^{r} \\ t^{l} + a^{l}(x - t^{l}), x \le t^{l} \end{cases}$
	Natural Logarithm ReLU [102]	$NTReLU(x) = \ln(\beta \max\{0, x\} + 1)$
adial Basis Function Activation	Gaussian [103]	$\rho(x) = \exp(\beta \ x - c\ ^2)$
oplausible Activation	apical dendrite activation (ADA) [104]	$ADA(x) = \min\{0, x\} + \max\{0, x\}\exp(-\alpha x + c)$
	Bionodal root unit (BRU) [105]	BRU(x) = $\begin{cases} (r^2 x + 1)^{\frac{1}{r}}, x \ge 0\\ e^{rz} - \frac{1}{r}, x < 0 \end{cases}$
oisy Activation	[106]	$\phi(x) = h(x) + s(x)$ , where $s(x)$ is random noise
idial Basis Function Activation oplausible Activation bisy Activation	GenLU [98] Randomly Translational ReLU [99] ELU [100] SReLU [101] Natural Logarithm ReLU [102] Gaussian [103] apical dendrite activation (ADA) [104] Bionodal root unit (BRU) [105] [106]	GenLU(x) = sgn(x)max {lxl+b,0} RTReLU(x) = $\begin{cases} x + a, x + a \ge 0\\ 0, x + a \le 0 \end{cases}$ ELU(x) = $\begin{cases} x, x \ge 0\\ \alpha(e^x - 1), x < 0 \end{cases}$ SReLU(x) = $\begin{cases} t^r + a^r(x - t^r), x \ge t^r\\ x, t^1 \le x \le t^r\\ t^1 + a^l(x - t^l), x \le t^l \end{cases}$ NTReLU(x) = ln( $\beta$ max {0,x} + 1) $\rho(x) = \exp(\beta \ x - c\ ^2)$ ADA(x) = min {0,x} + max {0,x} exp(- $\alpha x + \alpha$ BRU(x) = $\begin{cases} (r^2 x + 1)^{\frac{1}{r}}, x \ge 0\\ e^{rz} - \frac{1}{r}, x < 0 \end{cases}$ $\phi(x) = h(x) + s(x)$ , where $s(x)$ is random noi

#### **Rectified linear unit (ReLU)**

ReLU [96] is currently the most popular activation function thanks to its excellent scalability. However, some researchers argued that ReLU causes the loss of useful information in negative parts. To address this issue, several variants that allow the passage of negative parts were proposed such as Leaky-ReLU [95], Concatenate-ReLU [97], GenLU [98], Randomly Translational ReLU [99], ELU [100]. Besides, ReLU was believed to have a limited discriminative ability because it just assumes a linear relation in the positive range. Therefore, several variants were proposed to modify ReLU towards an enhanced non-linearity such as SReLU [101] and Natural Logarithm ReLU [102].

Radial Basis Function (RBF): A radial basis function activation is usually used in the radial basis function network [103], which is formulated as

$$\psi(\mathbf{x}) = \sum_{i}^{N} a_{i} \rho(||\mathbf{x} - c_{i}||), \qquad (2)$$

where  $\rho$  is often taken as a Gaussian function  $\rho(||\mathbf{x} - c_i||) = \exp(-\beta_i ||\mathbf{x} - c_i||^2)$ . The radial basis function network can be regarded as a fuzzy logic system such as the Takagi–Sugeno rule system [107] whose rule is of the format: "if  $\mathbf{x} \in \text{set } A$  and  $\mathbf{y} \in \text{set } B$ , then z = f(x, y)" [108].

#### **Bioplausible activation**

One notable class of activation functions are biologicallyinspired. Georgescu et al. [104] proposed a bioplausible activation function by mathematically modeling the newlydiscovered action potential pattern in a study [75] published in Science. Bhumbra [105] introduced a bionodal root activation based on the input–output relation acquired from physiological measurement. Electrophysiological recordings show that only a moderate increment in inputs is required to drive an action potential for some neurons [105]. Then, after an initial linear relation, the input–output curve partially saturates because the voltage-gated channels become less sensitive.

#### **Noisy activation**

The idea of noisy activation is to inject noise into the activation function when it is saturated. Such an injection can make gradients flow easily [106].

#### **Polynomial neuron**

Compared to the activation function, altering feature aggregation is much less explored. Per our earlier analysis, the nonlinearity in biological neurons is embedded not only in the outbound signal transduction but also in the inbound signal transduction and dendritic computation. But the inner product is a linear operation, which does not align with nonlinearity of the inbound signal transduction and dendritic computation. The use of polynomial neurons is to endow an artificial neuron with the nonlinear feature processing ability in the aggregation phase to better fit strongly nonlinear information. The story of polynomial neurons originates from the Group Method of Data Handling (GMDH [109]), which takes a high-order polynomial as a feature extractor:

$$Y(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \cdots,$$
(3)

where  $x_i$  is the *i*-th input, and  $a_i$ ,  $a_{ij}$ ,  $a_{ijk}$  are coefficients for interaction terms. Usually, only quadratic terms are retained in this model to avoid nonlinear explosion for high-dimensional inputs. Furthermore, with GMDH, the so-called

higher-order unit was defined in [110–112] which is mathematically formulated as

$$y = \sigma(Y(x_1, x_2, \cdots, x_n)), \tag{4}$$

where  $\sigma(\cdot)$  is a nonlinear activation function. To achieve a balance between maintaining the power of high-order units and parameter efficiency, Milenkoiv et al. [113] only utilized linear and quadratic terms and proposed to use an annealing technique to find optimal parameters.

Recently, high-order, particularly quadratic units were revisited [114–122]. In the work by Chrysos et al. [119], the complexity of higher-order units as described by Eq. (5) were greatly reduced via tensor decomposition and factor sharing; therefore, they scaled polynomial networks into a very deep paradigm to achieve the cutting-edge performance on several tasks. In [114, 123, 124], a quadratic convolutional filter of the complexity  $O(n^2)$  was proposed to replace the linear filter. In [116], a parabolic neuron:  $\sigma((x^Tw_1 + b_1)(x^Tw_2 + b_2))$  was proposed for deep learning, while in [125],  $\sigma((x \odot x)^T w)$  was proposed. Fan et al. [126]. proposed a simplified quadratic neuron with O(3n) parameters:  $\sigma((x^Tw_1 + b_1)(x^Tw_2 + b_2) + (x \odot x)^Tw_3)$  and further argued that higher-order neurons are not necessary because the fundamental theorem of algebra suggests that any polynomial can be factorized into a product of linear and quadratic terms [127]. Neuron designs in [116, 125] are special cases of that in [126]. Bu et al. [128] utilized the quadratic neuron, which is equivalent to [116] when combining  $x^T w_3$ into $\sigma((x^T w_1)(x^T w_2))$ . Xu et al.'s design [118] is the same as [128]. Liu and Wang [129] defined the so-called Gang neuron, which is essentially a polynomial neuron under a particular tensor decomposition.



**Fig.4** Three exemplary new types of neuron designs: polynomial neurons, dendritic neurons, and spiking neurons. Polynomial neurons, dendritic neurons, and spiking neurons conform to the three salient characteristics of biological neurons A) ubiquitous nonlinearity in information processing; B) the active dendrite;C) the spatiotemporal information ability, respectively

#### **Dendritic neuron**

The current mainstream neuron type is a point neuron which computes a single weighted sum of all synapses. Such modeling actually ignores the active computation of the dendritic fibers and only takes them as a passive transmission medium. However, as aforementioned, dendrites are active and serve as compartmentalized information processing units to assist a neuron to perform different kinds of tasks. Thus, can we go from a linear and passive neuron (point neuron) to a nonlinear and active neuron with extensive dendrite branching and extra nonlinear *computation?* (Chapter 14, [65]). Along this direction, a feasible design is to first compute a subset of nonlinear terms within a subset of computational units, then combine the responses of these units and nonlinearly map their total response (Chapter 16, [65]), as shown in Fig. 4. For example, dendritic branches were simulated to compute the sum of products (pi-sigma units [130, 131]) and to implement Boolean logic networks [132-134]. Shin et al. reported the so-called pi-sigma unit [135],

$$h = \sum_{k} w_{kji} x_k + \theta_{ji}, and \ y_i = \sigma \left( \prod_{j} h_{ji} \right), \tag{5}$$

where  $h_{ji}$  is the output of the *j*-th sigma unit for the *i*-th output element  $y_i$ , and  $w_{kji}$  is the weight of the *j*-th sigma unit associated with the input element  $x_k$ . The cluster-sensitive phenomenon was discovered that given a fixed number of synaptic inputs, concentrating the activated synapses of intermediate size can lead to the largest post-synaptic response [136–138]. Based on this phenomenon, the "clusteron" was proposed in [136] whose output is given by

$$y = g(\sum_{i}^{N} a_{i}), \tag{6}$$

where  $a_i = w_i x_i (\sum_{j \in D_i} w_j x_j)$  is the net excitatory input at synapse *i*, and  $D_i = \{i - r, i - r + 1, \dots, i, \dots, i + r - 1, i + r\}$  is a set of neighbors of the synapse *i*. It can be seen that the clusteron is a constrained sigma-pi unit with products of neighboring synapses. Furthermore, Jadi et al. [139] proposed a two-stage dendritic neuron:

$$y = FI(\sum_{j} W_{j} d_{j}), \tag{7}$$

where FI is an experimentally-determined frequency-current relation, and  $d_j = \sigma(\sum_i w_{ij}x_i)$ . However, such a dendritic neuron is essentially isomorphic to a two-layer network. Hawkins and Ahmad [140] and Grewal et al. [141] developed a type of dendritic neurons, as shown in Fig. 4, where each dendritic branch contains multiple synapses such that each branch can detect multiple input patterns. At the same time, Grewal et al. showed that such a type of dendritic neurons can alleviate catastrophic forgetting. In the aspect of hardware, Li et al. [142] experimentally demonstrated that neural networks with artificial dendrites are power-efficient. Spiking neuron

The current mainstream neuron types are regarded as the second-generation neurons that can only process static amplitude information. The neurons that have the spatiotemporal information ability are referred to as the third-generation neurons, which are primarily of the "integrate-and-fire" type [143–145] via spikes [146].

For example, the leaky integrator neuron is shown in Fig. 4. Its neuronal dynamics consists of two parts: (i) an equation that describes the change of the transmemebrane potential difference; (ii) a mechanism to generate a spike. We use the law of current conservation to derive the integration equation of spiking neurons. The current is split into two components:

$$I_t = I_R + I_C = \frac{\Delta u_t}{R} + C \frac{d}{dt} \Delta u_t, \tag{8}$$

where  $\Delta u_t$  is the transmembrane potential difference,  $I_t$  is the current from synapses or the external injection independent of the membrane potential, R is the leaky resistance, and C is the membrane capacitance. When  $\Delta u_t$  hits a threshold, it forms a spike and then set to zero. Assuming a constant external current injection and no synaptic current, we have the following:

$$\Delta u_t = IR(1 - e^{-\frac{1}{RC}}). \tag{9}$$

Next, the mechanism of generating a spike is via a spike response model [147], which simulates the refractory properties of a neuron because the membrane potential depends on the time of the last spike. There are other variants of the integrate-and-fire model: the linear integrate-and-fire neuron [148] which replaces  $\Delta u_t / R$  in Eq. (8) with a constant term; the quadratic integrate-and-fire neuron [149-151] which adds a quadratic term  $u^2$  to the right side of Eq. (8); the exponential integrate-and-fire neuron [152] which adds an exponential term regarding  $u_i$  to the right side of Eq. (8). As mentioned earlier, the information transmission in a neuron is coupled with noise. For example, the voltage-gated channels randomly open and close, and the vesicles randomly fuse with a neuron's membrane to release neurotransmitters. Thus, a diffusion variant of spiking neurons [153, 154] was established by incorporating the stochastic nature of the current:

$$I_t = \mu + \sqrt{2RC}\xi_t = \frac{\Delta u_t}{R} + C\frac{d}{dt}\Delta u_t,$$
(10)

where  $\mu$  is the average synaptic current, and  $\xi_t$  is a Gaussian noise.

The second-generation neurons are trained with gradient descent, whereas the training of spiking neurons is difficult due to the non-differentiability. Trainability has been a major bottleneck in the development of spiking networks. Now there are roughly three approaches to train a spiking network.

The idea of conversion-based methods is to map a wellperforming conventional network into a spiking network, which requires parameter recalibration and activation rescaling [155-157]. This method helps a spiking network to showcase the state-of-the-art performance on the ImageNet. However, such a conversion may be unfaithful when time steps are low.

The second method is to employ surrogate gradients [158–160]. The spiking network is hard to train because spiking neurons fire discrete spikes that are non-differentiable. To solve this problem, differentiable activation functions are employed as surrogates. Moreover, unrolling a spiking neuron in a discrete, recursive form perfectly corresponds to a recurrent neural network. As a result, the training of a spiking network can be done via backpropagation through time (BPTT) [161]. The caveat of surrogate gradient methods is the high computation overhead given a large time step. Spike-timing-dependent plasticity (STDP) approaches [162–164] gain great interest recently. SDTP methods are with local learning, following the Hebbian learning rule and updating weights in an unsupervised manner. However, STDP methods are hard to scale to multilayer spiking networks, and have limited performance on large-scale datasets.

#### Remarks

Polynomial neurons, dendritic neurons, and spiking neurons conform to the three salient characteristics of biological neurons: i) ubiquitous nonlinearity in information processing; ii) the active dendrite; iii) the spatiotemporal information ability, respectively. But polynomial neurons, dendritic neurons, and spiking neurons are not exactly the same as the ways of biological neurons. We argue that artificial networks should draw inspiration instead of copying from neuroscience. In other words, designing new neurons should prioritize the need for real-world problems instead of blindly imitating biological neurons.

#### **Potential gains**

So far, we have discussed new types of neurons, with the motivation of filling the gap between the current mainstream neuron type and the real-world biological neuron. Next, more importantly, what can we gain from these new neurons? *i.e.*, how can neuronal diversity truly bring benefits to the aforementioned critical issues in artificial networks? In the following, let us illustrate the potential gains of incorporating neuronal diversity into artificial networks.

#### Efficiency

In the information world, most learning tasks establish a nonlinear mapping. Intuitively, it is more efficient to incorporate a nonlinear computation unit to learn a nonlinear function. Although the existing mainstream neurons can do the universal approximation when connected to a network, it is computationally heavy to use mainstream neurons to represent other neuron types because such a representation needs far more neurons, compared to directly adopting other neuron types in building a model. It was proved that there exists a class of functions that can be approximated by a heterogeneous network made of both quadratic and conventional neurons with a polynomial number of neurons, but is hard to approximate by a purely conventional or quadratic network unless an exponential number of neurons are used [165]. Moreover, regarding the training cost, using the homogeneous type of neurons takes the extra learning cost to wire neurons specifically and orchestrate the learning process relative to using different neurons beforehand. In other words, involving neuronal diversity in artificial networks is an embodiment of modularization at the neuronal level. The development of modern industry has confirmed the superiority of modularization in efficiency. Table 2 summarizes the recent work that addresses the efficiency of introducing neuronal diversity.

Due to the spatiotemporal information processing ability, spiking neurons are highly energy-efficient. Unlike the conventional neuron that keeps the working status all the time, the spiking neuron idles unless it receives a spike from some events.

Tab	le 2	Studies	regarding	the efficient	ncy of c	quadratic	neurons
-----	------	---------	-----------	---------------	----------	-----------	---------

Work	Description
[1 <mark>66</mark> ]	This paper investigates the superiority of quadratic over conventional neural networks for classification of gaussian mixture data
[167]	This paper investigates the efficiency of quadratic networks over a wide variety of tasks including the image classification on the ImageNet, image segmentation, point cloud segmentation, and so on
[168]	This paper investigates the efficiency of mixing quadratic and conventional neurons in the task of anomaly detection
[169]	This paper proposes to design task-based neurons and evaluate its efficiency compared to conventional and quadratic neurons

#### Memory

The reason why a connectionist model suffers catastrophic forgetting is that training such a model for a new task catastrophically interferes with the knowledge amassed in the previous task. In contrast, although humans tend to gradually forget previous information as getting old, learning new knowledge while catastrophically forgetting old knowledge rarely happens. Achieving lifelong learning is difficult because of the stability-plasticity dilemma: the model has to maintain both plasticity to acquire new knowledge and stability to prevent the consolidated knowledge from being dismantled. Roughly, three types of computational approaches have been proposed to address the catastrophic forgetting [170]: i) imposing constraints on the level of plasticity to protect the consolidated knowledge [171]; ii) allocating additional neural resources for new tasks [172]; iii) using two complementary learning systems dedicated to learning new information and replaying old experiences, respectively [173]. Inspired by the second type of approaches, we find that polynomial neurons can offer a novel view to achieve lifelong learning by enabling a network to be internally progressive. Thus, the knowledge of old tasks is stored instead of destroyed when facing new tasks. As shown in Fig. 5(a), traditionally, given a new task, a new sub-network is created, and lateral links with the old tasks are learned. We refer to such a network as an externally progressive network. In contrast, assuming a polynomial neuron is used, as Fig. 5(b), tensor decomposition is doable to rearrange a polynomial neuron [119] into an internally compositional structure which encodes knowledge of a sequence of tasks into a hierarchy. Retaining old terms and adding new terms provides flexibility to balance old knowledge and the new. The internally progressive mechanism may apply to situations where new tasks and old tasks are somewhat relevant.

#### Interpretability

One way to derive an interpretation from a model is to understand its components, as the entire complex system can be usually decomposed into a combination of many functional modules [174]. For example, Bau et al. [175] analyzed a CNN trained on the scene classification task and discovered via the receptive field analyses that each neuron matches certain object concepts. In the same vein, an exciting point about a polynomial neuron or other nonlinear neurons is that the neuron per se contains an internal attention mechanism. The following derivation shows how to cast the attention mechanism from a quadratic neuron [176]:

$$\sigma(\left(\boldsymbol{x}^{\top}\boldsymbol{w}^{r}+\boldsymbol{b}^{r}\right)\left(\boldsymbol{x}^{\top}\boldsymbol{w}^{g}+\boldsymbol{b}^{g}\right)+\left(\boldsymbol{x}\odot\boldsymbol{x}\right)^{\top}\boldsymbol{w}^{b}+\boldsymbol{c}\right)$$

$$=\sigma(\boldsymbol{x}^{\top}\boldsymbol{w}^{g}\left(\boldsymbol{x}^{\top}\boldsymbol{w}^{r}+\boldsymbol{b}^{r}\right)+\boldsymbol{b}^{g}\boldsymbol{x}^{\top}\boldsymbol{w}^{r}+\boldsymbol{b}^{g}\boldsymbol{b}^{r}+\left(\boldsymbol{x}\odot\boldsymbol{x}\right)^{\top}\boldsymbol{w}^{b}\right)$$

$$=\sigma(\boldsymbol{x}^{\top}\left(\boldsymbol{w}^{g}\left(\boldsymbol{x}^{\top}\boldsymbol{w}^{r}+\boldsymbol{b}^{r}\right)\right)+\boldsymbol{x}^{\top}\boldsymbol{w}^{r}\boldsymbol{b}^{g}+\boldsymbol{x}^{\top}(\boldsymbol{x}\odot\boldsymbol{w}^{b})\right)$$

$$=\sigma(\boldsymbol{x}^{\top}(\boldsymbol{x}\odot\boldsymbol{w}^{b}+\boldsymbol{w}^{g}\boldsymbol{x}^{\top}\boldsymbol{w}^{r}+\boldsymbol{w}^{g}\boldsymbol{b}^{r}+\boldsymbol{w}^{r}\boldsymbol{b}_{g})),$$
(11)

where  $x \odot w^b + w^g x^\top w^r$  can reflect where a neuron deems as important regions, in analogy to the attention mechanism.

Furthermore, in this light, all neurons with a nonlinear aggregation function are self-explanatory. Suppose that a neuron is  $\sigma(g(x))$ , we conduct the Taylor expansion around 0 for g(x) and remove the third and higher-order terms:

$$\sigma(g(\mathbf{x})) = \sigma(g(0) + \mathbf{x}^{\mathsf{T}} Dg(0) + \mathbf{x}^{\mathsf{T}} Hg(0)x)$$
  
=  $\sigma(g(0) + \mathbf{x}^{\mathsf{T}} (Dg(0) + Hg(0)x)),$  (12)

where Dg(0) is the partial derivative of  $g(\mathbf{x})$  at zero, and Hg(0) is the Hessian matrix of  $g(\mathbf{x})$  at zero. Clearly, the mainstream neuron type  $\sigma(\mathbf{x}^{\mathsf{T}}\mathbf{w} + b)$  does not enjoy such a kind of self-interpretability, which necessitates the involvement of new type of neurons in artificial networks for better interpretability.



(a) Externally Progressive Networks

(b) Internally Progressive Networks

Fig. 5 Polynomial networks can work progressively to avoid catastrophic forgetting

#### **Representative applications**

Since introducing neuronal diversity is a fundamental change instead of a slight modification for a neural network, it has a global impact on the research and development of artificial networks, with the promise of pushing a wide spectrum of applications. Here, we discuss representative realworld applications in different important fields to illustrate the practical value of introducing neuronal diversity.

# **Medical imaging**

X-ray computed tomography (CT) is one of the most popular and important imaging modalities in hospitals and clinics. Although CT can offer critical clinical information, patients have to bear potential risks because X-rays may cause genetic changes and cancer [177]. Therefore, reducing the radiation dose is an important problem in the CT field. However, images reconstructed from the lower dose suffer from noise and other kinds of artifacts. In clinics, noise removal, texture preservation, and structure fidelity are three key aspects concerning radiologists. Algorithms should achieve a reasonable balance between these three aspects for a better clinical diagnosis.

Autoencoders [178] are a class of networks that consist of encoding and decoding parts. The encoding part attempts to learn a new representation, and the decoding part regenerates the input from the learned representation. In [179], a quadratic autoencoder (Q-AE) was proposed to process the low-dose CT images, in hope that the quality of processed images can reach the level of images reconstructed from the normal-dose CT. This Q-AE model employs ReLU as activation functions for all neurons, and has 5 quadratic convolutional layers in the encoding and 5 quadratic deconvolutional layers in the decoding, where each layer has 15 quadratic filters of  $5 \times 5$ , and symmetric layers are aggregated by residual connections. The anonymous reader study on the Mayo low-dose CT dataset revealed the superior performance of the quadratic autoencoder in terms of image denoising and model efficiency than other state-of-the-art models.

#### Industrial informatics

The bearing faults are the most common source of faults in rotating machines such as wind turbines and aircraft engines [180]. To enhance the reliability of rotating machines and avoid economic loss, accurately diagnosing bearing faults is of great importance. Currently, a common and viable diagnosis method is to first measure vibration signals by attaching the measurement instrument to the rotating bearing [181], and then use an artificial network to classify the faulty signals from the normal. Despite great successes in bearing

fault detection, artificial networks lack interpretability, *i.e.*, it is hard to know if the model conforms to the physics principle when classifying fault signals out.

A convolutional neural network made of quadratic neurons (QCNN) was proposed for bearing fault diagnosis [176]. With the qttention mechanism derived earlier, the feature extraction process of QCNN and the physics principle explaining why the model can deliver good classification performance are deciphered to a large extent. For example, by visualizing faulty bearing signals and the qttention maps, it was found that all the faulty areas were captured by QCNN. By comparing the raw signal and the qttention map in the frequency domain, it was discovered that QCNN favors bearing fault frequency elements over the shaft frequency elements.

#### Numerical computing

A general form of a non-linear partial differential equation (PDE) can be expressed as  $f(u, \gamma) = 0$ , where *f* is a non-linear operator performed on partial derivatives of the target variable *u*, and  $\gamma$  represents the parameters of the PDE. In the realm of PDEs and numerical computing, two classes of problems are mainly concerned: (a) forward problems, which are to solve for the target variable u prescribed by the PDE, and (b) inverse problems, which involve learning the unknown parameters,  $\gamma$ , of the PDE, given the observations of *u* at different timestamps.

Physics-informed neural networks (PINNs) are a type of neural networks that consider the physical laws and prior knowledge governing the problem in model design and training [182]. The prior knowledge of general physical laws is in the form of PDEs, and can be employed as a regularization for the training of a network, e.g., formulating the PDE equation into a supervised loss function. Motivated by the nonlinear approximation ability of quadratic networks, Bu et al. [128] proposed a quadratic residual network to solve the forward and inverse problems in PDEs. Following the original PINN framework, empirical results demonstrated that QResNet exhibits consistent advantages over conventional networks in terms of parameter efficiency and approximation accuracy. Let us take the Allen-Cahn equation [183] as an example to show why QResNet fits. The task is to predict u based on the PDE as follows:

$$\frac{\partial u}{\partial t} - 0.0001 \cdot \frac{\partial^2 u}{\partial x^2} + 5u^3 - 5u = 0, \tag{13}$$

and the initial condition. Discretizing this equation leads to

$$u_{t+1} = u_t + 0.0001 \cdot \frac{\partial^2 u_t}{\partial x^2} - 5u_t^3 - 5u_t.$$
(14)

Since the right-hand side of Eq. 14 consists of a polynomial function regarding  $u_t$ . Therefore, it is more suitable to use a quadratic neuron to learn it than a conventional neuron. Moreover, Eq. 14 contains a residual relation, which well fits a residual network.

#### **Computer vision**

Computer vision is an interdisciplinary field that enables a computer to derive a meaningful understanding from digital images, videos, or other visual inputs [184]. Computer vision tasks concern acquiring, processing, analyzing, and understanding digital inputs such as image restoration, face recognition, and video tracking. Neural network models have been dominating computer vision since the tremendous success in classifying approximately 1.2 million images into 1,000 classes in the ImageNet challenge [16]. As a drop-in replacement, polynomial networks [119] have been applied to a plethora of computer vision tasks. It was empirically demonstrated that the polynomial networks produce competitive results in a large variety of computer vision tasks such as image recognition, face recognition, and image generation.

# **Challenges and outlooks**

#### Neuronal synergy

To unleash the potential of neuronal diversity, the core problem is how to make different neurons synergize together to maximize their strengths, as Fig. 6 shows. In the human brain, the activities of a large number of neurons are well coordinated. For example, parallel activities of neurons are observed to lie in a low-dimensional manifold [185, 186]. Although the underlying coordination mechanism remains unclear, we know that the coordination of different neurons in the brain is highly artful and efficient, *e.g.*,

Fig. 6 Different neurons should be synergized together to maximize their strengths. Despite the diversity, neurons have several basic types (denoted by various shapes), and each type of neurons have similar but slightly different variants (denoted by various colors) the coordination of distant neurons is enabled not only via long-range connections but also through heterogeneity in local connectivity [187]. Generally speaking, the learning of an artificial network is governed by the loss function and the optimization algorithm. There is no explicit algorithm suggesting how to synergize different types of neurons for the same task. Inspired by neural architecture search [188], a brute-force means to accommodate this problem is neuronal cell search, which takes the neuronal type as the model's hyperparameters to optimize in the framework of AutoML [189]. However, it is more desirable if neuroscience findings can shed light on some rule-of-thumb or useful inductive bias to guide neuronal synergy. For example, in the population coding theory, the collective responses of a population of neurons are to maximize the amount of information [190].

#### Task-based neuron design

The past 10 years have witnessed a surge of many outstanding architectures, such as U-Net [191], the pyramidal structure [192], and shortcuts [1, 193]. The central principle behind these studies is designing a network architecture according to the needs of a task. In light of neuronal diversity, the neuronal type is also critical to the power of a neural network. Thus, we ask the following question: Can the network design go from the task-based architecture design to the task-based neuron design? Our brain is exactly a taskbased neuron designer, i.e., biological neurons have abundant functional diversity, which is a necessity for the brain to execute different tasks. The advantage of the task-based neuron design over the task-based architecture design is that task-specific neurons contain useful implicit bias for the task. Thus, the network of these task-specific neurons can integrate the task-driven forces of all these neurons, which should be much stronger than the network of generic neurons with the same structure. Our conjecture is that the task-based neuron design will escalate neural network research into a new stage and make advances in many previously-believed





Fig. 7 A roadmap for designing and deploying task-based neurons: i) build an elementary neuronal model by the symbolic regression; ii) parameterize the acquired elementary neuron to make its parameters learnable; iii) employ task-based neurons in a network for validation and feedback



difficult tasks. Figures 7 and 8 showcases a three-step roadmap for the task-based neuron design:

- Build an elementary neuronal model via symbolic regression. Encouraged by the concept of linear regression or polynomial regression, symbolic regression [194] is to search over the space of all possible mathematical formulas regarding the input variables, starting from base functions such as logarithmic, trigonometric, and exponential functions. Furthermore, the search space should be regularized to entitle the established neuron with the desirable properties, *e.g.*, no gradient vanishment or explosion when connected to a network. For example, we can split the search space into two parts: the aggregation function and the activation function.
- 2) Parameterize the learned elementary neuron to make its parameters trainable. Such a parameterization can be straightforwardly made by casting all coefficients in the

elementary neuron as learnable parameters. However, for better efficiency and expressivity, selecting which coefficients to parameterize can be optimized based on the final performance.

3) Employ task-based neurons in a network for validation and feedback. For example, according to the performance of the network, some base formulas in a taskbased neuron might be pruned. Task-based neurons can also facilitate multi-task learning. With increasingly many task-based neurons designed, a warehouse of neurons can be gradually established in which a knowledge graph and neuron-based informatics could be developed.

# **Theoretical issues**

Since the rise in 2012, deep learning has been criticized for lacking a fundamental theory. This embarrassment

Med-X (2025) 3:2

Resource	Туре	Description
QuadraLib	Library	QuadraLib is a library for the efficient optimization and design exploration of quadratic networks. The paper of QuadraLib won MLSys2022's best paper award
Dr. Fenglei Fan's GitHub Page	Code	Dr. Fenglei Fan's GitHub Page summarizes a series of papers and associated code on quadratic net- works, including quadratic autoencoder and the training algorithm <i>ReLinear</i>
Polynomial Network	Code	This repertoire shows how to build a deep polynomial network and sparsify it with tensor decomposition
Dendrite	Book	A comprehensive book covering all aspects of dendritic computation

Table 3 A resource summary for neuronal diversity in artificial networks

has been greatly alleviated in recent years with the rapid development of deep learning theory. Expressivity-wise, why a deep network performs superbly [195, 196] is well addressed by characterizing the complexity of a function expressed by a neural network, *i.e.*, increasing depth can greatly maximize such a complexity measure compared to increasing width; the power of shortcuts are demonstrated as well [1], *i.e.*, a network with shortcuts can express a far more complicated network than a network without shortcuts. Optimization-wise, the neural tangent kernel theory suggests that the training of an infinitely wide network is equivalent to a kernel ridge regression [197]. The understanding of the generalization ability of deep networks is also deepened by the discovery of the double descent phenomenon [198].

Notwithstanding, the existing theory provides an explanation for the simplest neuron that is based on an inner product and an activation function, which may not be applicable to other kinds of nonlinear neurons. In the context of neuronal diversity, we ask the following question for theorists to brainstorm: to what extent are the current theories scalable to heterogeneous networks? Addressing this question is highly nontrivial in two senses. On the one hand, the philosophy behind homogeneous and heterogeneous networks varies greatly. The former implicitly assumes that a universal type of neurons can solve a wide class of complicated nonlinear problems, simply referred to as "one-for-all". Such a philosophy is well supported by the universal approximation theorem [199]. However, the problem of this philosophy is practicality and efficiency, *i.e.*, it may suffer the curse of dimensionality. In contrast, the latter assumes different types of neurons solve a specific problem, which is referred to as "all-for-one". The loss of universality in heterogeneous networks adds a layer of complication to the theoretical analyses. The low-hanging fruits may come from the efficiency analysis first by showing that the mode of "all-for-one" is more efficient than that of "one-for-all". Then, the optimization and generalization properties can be further analyzed.

#### Neuroinformatics of artificial neurons and networks

Neuroinformatics is an interdisciplinary field that introduces informatics into neuroscience for data mining and information processing. With the large-scale deployment of those task-specific neurons into networks, methodologically, we think it is highly necessary to introduce tools of informatics into artificial neurons and networks to further extract information from different neurons and networks, referred to as neuroinformatics of artificial neurons. The goals of this kind of neuroinformatics are to make contributions to the next generation of AI by drawing insights into the information processing of biological networks, supporting brain-inspired intelligence, and fertilizing the interpretability of artificial networks. To realize this goal, on the one hand, tools for analyzing and standardizing artificial neurons and networks should be developed; on the other hand, the database and knowledge graph can be built for various artificial neurons to supply ontology information about neuron fitness and relations between different neurons, to advance the connectivity organization, and to instruct the multi-module synergy.

#### Remark

Future research should particularly emphasize the identification of impactful applications of neuronal diversity or leverage this concept to address real-world challenges, thereby enhancing the relevance and success of neuronal diversity. Additionally, in light of the remarkable achievements of large language models, there is a pressing need to develop a large model that incorporates diverse types of neurons. Such a model could be computationally efficient, promoting widespread adoption in resource-constrained environments, owing to the robust representational capabilities of innovative neurons at a fundamental level.

# **Resource summary**

We list the useful resources about neuronal diversity in Table 3 for readers' reference.

# Conclusions

In this perspective, we have systematically introduced neuronal diversity into artificial networks, as a practice of NeuroAI, including biological background, the existing studies, challenges, and future directions. We believe that neuronal diversity, the insight from NeuroAI, has the potential of elevating artificial networks into a new age. Future efforts can be invested to demonstrate more and more killer applications of new neurons.

#### Acknowledgements No.

Authors' contributions Conceptualization: Fenglei Fan, Yingxin Li, Tieyong Zeng, Fei Wang, Hanchuan Peng. Original draft preparation: Fenglei Fan and Yingxin Li. Review and editing: Tieyong Zeng, Fei Wang, Hanchuan Peng.

**Funding** Dr. Fenglei Fan would like to acknowledge that this paper was supported by the Direct Grant for Research from the Chinese University of Hong Kong and ITS/173/22FP from the Innovation and Technology Fund of Hong Kong.

**Data availability** This draft is semi-perspective-semi-review paper; therefore, no data needs to be shared.

#### Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Yes.

Competing interests The authors declare no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

# References

- 1. LeCun Y, Bengio Y and Hinton G. Deep learning. Nature. 2015;521(7553):436–44.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Tunyasuvunakool. Highly accurate protein structure prediction with alphafold. Nature. 2021;596(7873):583–9.
- Floridi L, Chiriatti M. Gpt-3: Its nature, scope, limits, and consequences. Mind Mach. 2020;30(4):681–94.
- Warren SC, Walter P. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics. 1943;5(4):115–33.

- Hebb DO. The organization of behavior: A neuropsychological theory. London, United Kingdom: Psychology Press; 2005.
- Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev. 1958;65(6):386.
- Minsky M, A Papert S. Perceptrons. 6th ed. Cambridge: MA: MIT Press; 1969. p. 318–62.
- Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the national academy of sciences. 1982;79(8):2554–8.
- 9. Hinton GE, Sejnowski TJ, et al. Learning and relearning in boltzmann machines. Parallel distributed processing: Explorations in the microstructure of cognition. 1986;1(282–317):2.
- Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. nature. 1986;323(6088):533-6.
- Werbos PJ. The roots of backpropagation: from ordered derivatives to neural networks and political forecasting, volume 1. Oxford: John Wiley & Sons; 1994.
- 12. Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997;9(8):1735–80.
- Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. California Univ San Diego La Jolla Inst for Cognitive Science: Technical report; 1985.
- Fukushima K and Miyake S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In Competition and cooperation in neural nets. New York City, NY: Springer; 1982. p 267–85.
- LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998;86(11):2278–324.
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems. 2012;25:1097–105.
- Deng J, Dong W, Socher R, Li J-L, Li K, Li F-F. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. 2009. Ieee.
- Zador A, Richards B, Ölveczky B, Escola S, Bengio Y, Boahen K, Botvinick M, Chklovskii D, Churchland A, Clopath C, et al. Toward next-generation artificial intelligence: Catalyzing the neuroai revolution. arXiv preprint arXiv:2210.08340, 2022.
- García-Martín E, Rodrigues CF, Riley G, Grahn HK. Estimation of energy consumption in machine learning. J Parallel Distrib Comput. 2019;134:75–88.
- Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, et al. Language models are few-shot learners. Adv Neural Inf Process Syst. 2020;33:1877–901.
- Cox DD, Dean T. Neural networks and neuroscience-inspired computer vision. Curr Biol. 2014;24(18):R921–9.
- Lake B, Salakhutdinov R, Gross J, Tenenbaum J. One shot learning of simple visual concepts. In: Proceedings of the annual meeting of the cognitive science society. volume 33. 2011.
- Dhar P. The carbon impact of artificial intelligence. Nat Mach Intell. 2020;2(8):423–5.
- Strubell E, Ganesh A, and McCallum A. Energy and policy considerations for deep learning in nlp. arXiv preprint arXiv:1906. 02243, 2019.
- Fan F-L, Xiong J, Li M, Wang Ge. On interpretability of artificial neural networks: A survey. IEEE Transactions on Radiation and Plasma Medical Sciences. 2021;5(6):741–60.
- Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence. 2019;1(5):206–15.

- 27. Heeger DJ, Ress D. What does fmri tell us about neuronal activity? Nat Rev Neurosci. 2002;3(2):142–51.
- McCloskey M and Cohen NJ. Catastrophic interference in connectionist networks: The sequential learning problem. Psychol Learn Motiv. 1989;24:109–165
- Ratcliff R. Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. Psychol Rev. 1990;97(2):285.
- Hasselmo ME. Avoiding catastrophic forgetting. Trends Cogn Sci. 2017;21(6):407–8.
- Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, Milan K, Quan J, Ramalho T, Grabska-Barwinska A, et al. Overcoming catastrophic forgetting in neural networks. Proceedings Nat Academy Sci. 2017;114(13):3521–6.
- González OC, Sokolov Y, Krishnan GP, Delanois JE, and Bazhenov M. Can sleep protect memories from catastrophic forgetting? Elife. 2020;9:e51005.
- Paller KA, Voss JL. Memory reactivation and consolidation during sleep. Learning & Memory. 2004;11(6):664–70.
- Wilson MA, McNaughton BL. Reactivation of hippocampal ensemble memories during sleep. Science. 1994;265:676–9.
- Foster DJ, Wilson MA. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature. 2006;440:680–3.
- Carlini N, Wagner D. Towards evaluating the robustness of neural networks. In: 2017 ieee symposium on security and privacy (sp). 2017. p. 39–57 Ieee.
- 37. Tencent Keen Security Lab. Experimental security research of tesla autopilot. 2019.
- Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, and Fergus R. Intriguing properties of neural networks. arXiv preprint arXiv:1312.6199, 2013.
- Hendrycks D and Dietterich T. Benchmarking neural network robustness to common corruptions and perturbations. arXiv preprint arXiv:1903.12261, 2019.
- Stevens CF. Neuronal diversity: too many cell types for comfort? Current Biology. 1998;8(20):R708–10.
- Peng H, Xie P, Liu L, Kuang X, Wang Y, Lei Qu, Gong H, Jiang S, Li A, Ruan Z, et al. Morphological diversity of single neurons in molecularly defined cell types. Nature. 2021;598(7879):174–81.
- Thivierge J-P. Neural diversity creates a rich repertoire of brain activity. Communicative & integrative biology. 2008;1(2):188–9.
- Padmanabhan K, Urban NN. Intrinsic biophysical diversity decorrelates neuronal firing while increasing information content. Nat Neurosci. 2010;13:1276–82.
- 44. Flames N, Hobert O. Gene regulatory logic of dopamine neuron differentiation. Nature. 2009;458(7240):885–9.
- 45. Wang Y, Hu P, Shan Q, Huang C, Huang Z, Chen P, Li A, Gong H, Zhou JN. Single-cell morphological characterization of crh neurons throughout the whole mouse brain. BMC Biol. 2021;19:47.
- Ferreira-Pinto MJ, Kanodia H, Falasconi A, Sigrist M, Esposito MS, Arber S. Functional diversity for body actions in the mesencephalic locomotor region. Cell. 2021;184:4564–78.
- O'Keefe J, Burgess N, Donnett JG, Jeffery KJ, Maguire EA. Place cells, navigational accuracy, and the human hippocampus. Philos Trans R Soc Lond B Biol Sci. 1998;353(1373):1333–40.
- Fyhn M, Hafting T. Menno P Witter, Edvard I Moser, and May-Britt Moser. Grid cells in mice Hippocampus. 2008;18(12):1230–8.
- Lever C, Burton S, Jeewajee A, O'Keefe J, Burgess N. Boundary vector cells in the subiculum of the hippocampal formation. J Neurosci. 2009;29:9771–7.

- Taube JS, Muller RU, Ranck JB. Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis. J Neurosci. 1990;10(2):420–35.
- Góis ZHTD, Tort ABL. Characterizing speed cells in the rat hippocampus. Cell reports. 2018;25(7):1872–84.
- Stackman RW, Taube JS. Firing properties of rat lateral mammillary single units: head direction, head pitch, and angular head velocity. J Neurosci. 1998;18(21):9020–37.
- Gardner D. Where are the cores in this thing?... and what are they computing?(with apologies to larry abbott). J Comput Neurosci. 2022;50(2):133–8.
- Bi A, Cui J, Ma Y-P, Olshevskaya E, Pu M, Dizhoor AM, Pan Z-H. Ectopic expression of a microbial- type rhodopsin restores visual responses in mice with photoreceptor degeneration. Neuron. 2006;50(1):23–33.
- 55. Deisseroth K. Optogenetics. Nat Methods. 2011;8(1):26-9.
- Suchyna TM, Sachs F. Mechanosensitive channel properties and membrane mechanics in mouse dystrophic myotubes. J Physiol. 2007;581(1):369–87.
- 57. Rosenbaum DM, Rasmussen SGF, Kobilka BK. The structure and function of g-protein-coupled receptors. Nature. 2009;459(7245):356–63.
- Hucho F, Weise C. Ligand-gated ion channels. Angew Chem Int Ed. 2001;40(17):3100–16.
- Gilman AG. G proteins: transducers of receptor-generated signals. Annu Rev Biochem. 1987;56(1):615–49.
- Lohse MJ, Hein P, Hoffmann C, Nikolaev VO, Vilardaga J-P, Bünemann M. Kinetics of g-protein-coupled receptor signals in intact cells. Br J Pharmacol. 2008;153(S1):S125–32.
- Lamb TD, Pugh EN Jr. G-protein cascades: gain and kinetics. Trends Neurosci. 1992;15(8):291–8.
- 62. Sheng M, McFadden G, Greenberg ME. Membrane depolarization and calcium induce c-fos transcription via phosphorylation of transcription factor creb. Neuron. 1990;4(4):571–82.
- Branco T, Häusser M. The single dendritic branch as a fundamental functional unit in the nervous system. Curr Opin Neurobiol. 2010;20(4):494–502.
- 64. Ling G. History of the membrane (pump) theory of the living cell from its beginning in mid-19th century to its disproof 45 years ago-though still taught worldwide today as established truth. Physiol Chem Phys Med NMR. 2007;39(1):1–68.
- 65. Stuart G, Spruston N, and Häusser M. Dendrites. Oxford: Oxford University Press; 2016.
- Rall W. Branching dendritic trees and motoneuron membrane resistivity. Exp Neurol. 1959;1(5):491–527.
- 67. Rall W. Membrane potential transients and membrane time constant of motoneurons. Exp Neurol. 1960;2(5):503–32.
- Mainen ZF, Sejnowski TJ. Influence of dendritic structure on firing pattern in model neocortical neurons. Nature. 1996;382(6589):363-6.
- Ferrante M, Migliore M, Ascoli GA. Functional impact of dendritic branch-point morphology. J Neurosci. 2013;33(5):2156–65.
- De Sousa G, Maex R, Adams R, Davey N, Steuber V. Dendritic morphology predicts pattern recognition performance in multicompartmental model neurons with and without active conductances. J Comput Neurosci. 2015;38(2):221–34.
- 71. Wang G, Wang R, Kong W, and Zhang J. The relationship between sparseness and energy consumption of neural networks. Neural Plast. 2020;(1):8848901.
- Chavlis S, Petrantonakis PC, Poirazi P. Dendrites of dentate gyrus granule cells contribute to pattern separation by controlling sparsity. Hippocampus. 2017;27(1):89–110.
- Brunel N, Hakim V, Isope P, Nadal J-P, Barbour B. Optimal information storage and the distribution of synaptic weights: perceptron versus purkinje cell. Neuron. 2004;43(5):745–57.

- Levitan IB, Kaczmarek LK, et al. The neuron: cell and molecular biology. USA: Oxford University Press; 2002.
- Gidon A, Zolnik TA, Fidzinski P, Bolduan F, Papoutsi A, Poirazi P, Holtkamp M, Vida I, Larkum ME. Dendritic action potentials and computation in human layer 2/3 cortical neurons. Science. 2020;367(6473):83–7.
- Vergara C, Latorre R, Marrion NV, Adelman JP. Calcium-activated potassium channels. Curr Opin Neurobiol. 1998;8(3):321–9.
- Cook EP, Johnston D. Active dendrites reduce location-dependent variability of synaptic input trains. J Neurophysiology. 1997;78(4):2116–28.
- Das A, Narayanan R. Active dendrites regulate spectral selectivity in location-dependent spike initiation dynamics of hippocampal model neurons. J Neurosci. 2014;34(4):1195–211.
- Ariav G, Polsky A, Schiller J. Submillisecond precision of the input-output transformation function mediated by fast sodium dendritic spikes in basal dendrites of ca1 pyramidal neurons. J Neurosci. 2003;23(21):7750–8.
- Schiller J, Schiller Y, Stuart G, Sakmann B. Calcium action potentials restricted to distal apical dendrites of rat neocortical pyramidal neurons. J Physiol. 1997;505(3):605–16.
- Polsky A, Mel BW, Schiller J. Computational subunits in thin dendrites of pyramidal cells. Nature Neurosci. 2004;7(6):621–7.
- Losonczy A, Magee JC. Integrative properties of radial oblique dendrites in hippocampal ca1 pyramidal neurons. Neuron. 2006;50(2):291–307.
- Larkum ME, Zhu JJ, Sakmann B. A new cellular mechanism for coupling inputs arriving at different cortical layers. Nature. 1999;398(6725):338–41.
- 84. Faber DS, Pereda AE. Two forms of electrical transmission between neurons. Front Mol Neurosci. 2018;11:427.
- Nelson RJ. An introduction to behavioral endocrinology. Sunderland, MA: Sinauer Associates; 2005.
- Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, and Darnell J. Neurotransmitters, synapses, and impulse transmission. In Molecular Cell Biology. 4th edition. New York City, NY: WH Freeman; 2000.
- Eccles JC. The physiology of synapses. Cambridge, MA: Academic Press; 2013.
- Poage RE, Meriney SD. Presynaptic calcium influx, neurotransmitter release, and neuromuscular disease. Physiol Behavior. 2002;77(4–5):507–12.
- 89. Bassey J, Qian L, and Li X. A survey of complex-valued neural networks. arXiv preprint arXiv:2101.12249, 2021.
- Zhang S-Q, Gao W, Zhou Z-H. Towards understanding theoretical advantages of complex-reaction networks. Neural Netw. 2022;151:80–93.
- Dubey SR, Singh SK, and Chaudhuri BB. A comprehensive survey and performance analysis of activation functions in deep learning. arXiv preprint arXiv:2109.14545, 2021.
- Han J and Moraga C. The influence of the sigmoid function parameters on the speed of backpropagation learning. In International workshop on artificial neural networks. 1995:195–201.
- Chandra P, Singh Y. An activation function adapting training algorithm for sigmoidal feedforward networks. Neurocomputing. 2004;61:429–37.
- 94. Qin Yi, Wang X, Zou J. The optimized deep belief networks with improved logistic sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines. IEEE Trans Industr Electron. 2018;66(5):3814–24.
- Maas AL, Hannun AY, Ng AY, et al. Rectifier nonlinearities improve neural network acoustic models. In Proc Icml. 2013;30:3.
- 96. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: Icml. 2010.

- Shang W, Sohn K, Almeida D, Lee H. Understanding and improving convolutional neural networks via concatenated rectified linear units. In: International conference on machine learning. 2016. p. 2217–25 PMLR.
- Fan F, Li M, Teng Y, Wang Ge. Soft autoencoder and its wavelet adaptation interpretation. IEEE Transactions on Computational Imaging. 2020;6:1245–57.
- Cao J, Pang Y, Li X, Liang J. Randomly translational activation inspired by the input distributions of relu. Neurocomputing. 2018;275:859–68.
- Clevert D-A, Unterthiner T, and Hochreiter S. Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289, 2015.
- 101. Jin X, Xu C, Feng J, Wei Y, Xiong J, Yan S. Deep learning with s-shaped rectified linear activation units. In: Proceedings of the AAAI Conference on Artificial Intelligence, volume 30. 2016.
- 102. Liu Y, Zhang J, Gao C, Qu J, Ji L. Natural-logarithm-rectified activation function in convolutional neural networks. In: 2019 IEEE 5th International Conference on Computer and Communications (ICCC). 2019. p. 2000–8.
- Moody J, Darken CJ. Fast learning in networks of locally-tuned processing units. Neural Comput. 1989;1(2):281–94.
- Georgescu MI, Ionescu RT, Ristea N-C, and Sebe N. Non-linear neurons with human-like apical dendrite activations. arXiv preprint arXiv:2003.03229, 2020.
- Bhumbra GS. Deep learning improved by biological activation functions. arXiv preprint arXiv:1804.11237, 2018.
- Gulcehre C, Moczulski M, Denil M, Bengio Y. Noisy activation functions. In: International conference on machine learning. 2016. p. 3059–68 PMLR.
- Takagi T, Sugeno M. Fuzzy identification of systems and its applications to modeling and control. IEEE Trans Syst Man Cybern. 1985;1:116–32.
- Park J, Sandberg IW. Universal approximation using radial-basisfunction networks. Neural Comput. 1991;3(2):246–57.
- Alexey Grigorevich Ivakhnenko. Polynomial theory of complex systems. IEEE Trans Syst Man Cybern. 1971;4:364–78.
- Poggio T. On optimal nonlinear associative recall. Biol Cybern. 1975;19(4):201–9.
- Giles CL, Maxwell T. Learning, invariance, and generalization in high-order neural networks. Appl Opt. 1987;26(23):4972–8.
- Lippmann RP. Pattern classification using neural networks. IEEE Commun Mag. 1989;27(11):47–50.
- Milenkovic S, Obradovic Z, and Litovski V. Annealing based dynamic learning in second-order neural networks. In Proceedings of International Conference on Neural Networks (ICNN'96). 1996;1:458–63.
- Zoumpourlis H, Doumanoglou A, Vretos N, Daras P. Non-linear convolution filters for cnn-based learning. In: Proceedings of the IEEE International Conference on Computer Vision. 2017. p. 4761–9.
- Cheung KF and Leung CS. Rotational quadratic function neural networks. In [Proceedings] 1991 IEEE International Joint Conference on Neural Networks. 1991:869–874.
- 116. Tsapanos N, Tefas A, Nikolaidis N, Pitas I. Neurons with paraboloid decision boundaries for improved neural network classification performance. IEEE transactions on neural networks and learning systems. 2018;30(1):284–94.
- 117. Redlapalli S, Gupta MM, and Song K-Y. Development of quadratic neural unit with applications to pattern classification. In Fourth International Symposium on Uncertainty Modeling and Analysis, 2003. ISUMA 2003. 2003:141–6.
- 118. Zirui Xu, Fuxun Yu, Xiong J, Chen X. Quadralib: A performant quadratic neural network library for architecture optimization and design exploration. Proceedings of Machine Learning and Systems. 2022;4:503–14.

- Chrysos, G. G., Moschoglou, S., Bouritsas, G., Deng, J., Panagakis, Y., & Zafeiriou, S. (). Deep polynomial neural networks. IEEE Trans Pattern Anal Mach Intell. 2021;44(8):4021–34.
- 120. Livni R, Shalev-Shwartz S, and Ohad Shamir. On the computational efficiency of training neural networks. arXiv preprint arXiv:1410.1141, 2014.
- Krotov D, Hopfield J. Dense associative memory is robust to adversarial inputs. Neural Comput. 2018;30(12):3151–67.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, and Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst. 2017;30:6000–10.
- Jiang Y, Yang F, Zhu H, Zhou D, Zeng X. Nonlinear cnn: improving cnns with quadratic convolutions. Neural Comput Appl. 2020;32(12):8507–16.
- Mantini P and Shah SK. Cqnn: Convolutional quadratic neural networks. In 2020 25th International Conference on Pattern Recognition (ICPR). 2021:9819–26.
- Goyal M, Goyal R, and Lall B. Improved polynomial neural networks with normalised activations. In 2020 International Joint Conference on Neural Networks (IJCNN). 2020:1–8.
- 126. Fan F, Cong W, Wang Ge. A new type of neurons for machine learning. International journal for numerical methods in biomedical engineering. 2018;34(2): e2920.
- 127. Remmert R. The fundamental theorem of algebra. In Numbers. New York City, NY: Springer: 1991. p 97–122.
- 128. Bu J and Karpatne A. Quadratic residual networks: A new class of neural networks for solving forward and inverse problems in physics involving pdes. In Proceedings of the 2021 SIAM International Conference on Data Mining (SDM). 2021:675–83.
- 129. Liu G and Wang J. Dendrite net: A white-box module for classification, regression, and system identification. IEEE Transactions on Cybernetics, 2021.
- Mel B. Murphy: A robot that learns by doing. In: Neural information processing systems. 1987.
- Durbin R, Rumelhart RE. Product units: A computationally powerful and biologically plausible extension to backpropagation networks. Neural Comput. 1989;1(1):133–42.
- Koch C, Poggio T, Torre V. Retinal ganglion cells: a functional interpretation of dendritic morphology. Philos Trans R Soc Lond B Biol Sci. 1982;298(1090):227–63.
- Shepherd GM, Brayton RK. Logic operations are properties of computer-simulated interactions between excitable dendritic spines. Neurosci. 1987;21(1):151–65.
- Zador A, Claiborne B, and Brown T. Nonlinear pattern separation in single hippocampal neurons with active dendritic membrane. Adv Neural Inf Process Syst. 1991:51–8.
- 135. Shin Y and Ghosh J. The pi-sigma network: An efficient higherorder neural network for pattern classification and function approximation. In IJCNN-91-Seattle international joint conference on neural networks. 1991;1:13–18.
- 136. Mel B. The clusteron: toward a simple abstraction for a complex neuron. Adv Neural Inf Process Syst. 1991;4:35–42.
- Gasparini S, Migliore M, Magee JC. On the initiation and propagation of dendritic spikes in ca1 pyramidal neurons. J Neurosci. 2004;24(49):11046–56.
- Polsky A, Mel B, Schiller J. Encoding and decoding bursts by nmda spikes in basal dendrites of layer 5 pyramidal neurons. J Neurosci. 2009;29(38):11891–903.
- 139. Jadi MP, Behabadi BF, Poleg-Polsky A, Schiller J, Mel BW. An augmented two-layer model captures nonlinear analog spatial integration effects in pyramidal neuron dendrites. Proc IEEE. 2014;102(5):782–98.
- Hawkins, J., & Ahmad, S. Why neurons have thousands of synapses, a theory of sequence memory in neocortex. Front Neural Circuits. 2016;10:174222.

- Grewal K, Forest J, Cohen BP, and Ahmad S. Going beyond the point neuron: Active dendrites and sparse representations for continual learning. bioRxiv, 2021.
- 142. Li X, Tang J, Zhang Q, Gao B, Yang JJ, Song S, Wu W, Zhang W, Yao P, Deng N. Power-efficient neural network with artificial dendrites. Nat Nanotechnol. 2020;15(9):776–82.
- 143. Izhikevich EM. Simple model of spiking neurons. IEEE Trans Neural Netw. 2003;14(6):1569–72.
- 144. Hodgkin AL, Huxley AF. A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol. 1952;117(4):500.
- 145. Burkitt AN. A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. Biol Cybern. 2006;95(1):1–19.
- 146. Roy K, Jaiswal A, Panda P. Towards spike-based machine intelligence with neuromorphic computing. Nature. 2019;575(7784):607–17.
- 147. Gerstner W. Time structure of the activity in neural network models. Phys Rev E. 1995;51(1):738.
- Fusi S, Mattia M. Collective behavior of networks with linear (vlsi) integrate-and-fire neurons. Neural Comput. 1999;11(3):633-52.
- 149. Ermentrout B. Type i membranes, phase resetting curves, and synchrony. Neural Comput. 1996;8(5):979–1001.
- Brunel N, Latham PE. Firing rate of the noisy quadratic integrate-and-fire neuron. Neural Comput. 2003;15(10):2281–306.
- 151. Clusella P, Pietras B, Montbrió E. Kuramoto model for populations of quadratic integrate-and-fire neurons with chemical and electrical coupling. Chaos. 2022;32(1):013105.
- Fourcaud-Trocmé N, Brunel N. Dynamics of the instantaneous firing rate in response to changes in input statistics. J Comput Neurosci. 2005;18(3):311–21.
- Tuckwell HC. Nonlinear and stochastic theories. Introduction to Theoretical Neurobiology, vol 2, Cambridge, England: Cambridge University Press; 1988.
- Tuckwell HC, Cope DK. Accuracy of neuronal interspike times calculated from a diffusion approximation. J Theor Biol. 1980;83(3):377–87.
- Cao Y, Chen Y, Khosla D. Spiking deep convolutional neural networks for energy-efficient object recognition. Int J Comput Vision. 2015;113(1):54–66.
- 156. Sengupta A, Ye Y, Wang R, Liu C, Roy K. Going deeper in spiking neural networks: Vgg and residual architectures. Front Neurosci. 2019;13:95.
- 157. Rueckauer B, Lungu I-A, Yuhuang Hu, Pfeiffer M, Liu S-C. Conversion of continuous-valued deep networks to efficient event-driven networks for image classification. Front Neurosci. 2017;11:682.
- 158. Lee JH, Delbruck T, Pfeiffer M. Training deep spiking neural networks using backpropagation. Front Neurosci. 2016;10:508.
- Gu P, Xiao R, Pan G, Tang H. Stca: Spatio-temporal credit assignment with delayed feedback in deep spiking neural networks. In: IJCAI. 2019. p. 1366–72.
- Li Y, Guo Y, Zhang S, Deng S, Hai Y, Shi Gu. Differentiable spike: Rethinking gradient-descent for training spiking neural networks. Adv Neural Inf Process Syst. 2021;34:23426–39.
- Werbos PJ. Generalization of backpropagation with application to a recurrent gas market model. Neural Netw. 1988;1(4):339–56.
- 162. Bi G-Q, Poo M-M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci. 1998;18(24):10464–72.
- 163. Iakymchuk T, Rosado-Muñoz A, Guerrero-Martínez JF, Bataller-Mompeán M, Francás-Víllora JV. Simplified spiking neural network architecture and stdp learning algorithm applied to image classification. EURASIP J Image Video Process. 2015;2015(1):1–11.

- 164. Lobov SA, Mikhaylov AN, Shamshin M, Makarov VA, Kazantsev VB. Spatial properties of stdp in a self-learning spiking neural network enable controlling a mobile robot. Front Neurosci. 2020;14:88.
- Liao J-X, Hou B-J, Dong H-C, Zhang H, Ma J, Sun J, Zhang S, and Fan F-L. Heterogeneous autoencoder empowered by quadratic neurons. arXiv preprint arXiv:2204.01707, 2022.
- 166. Qi T, Wang Ge. Superiority of quadratic over conventional neural networks for classification of gaussian mixture data. Visual Computing for Industry, Biomedicine, and Art. 2022;5(1):23.
- Fan F-L, et al. "One neuron saved is one neuron earned: On parametric efficiency of quadratic networks." arXiv preprint arXiv: 2303.06316, 2023.
- 168. Liao J-X, et al. Quadratic Neuron-empowered Heterogeneous Autoencoder for Unsupervised Anomaly Detection. IEEE Transactions on Artificial Intelligence, in press, 2024. Available at https://ieeexplore.ieee.org/abstract/document/10510400.
- Fan F-L, et al. "No One-Size-Fits-All Neurons: Task-based Neurons for Artificial Neural Networks." arXiv preprint arXiv:2405. 02369, 2024.
- Parisi GI, Kemker R, Part JL, Kanan C, Wermter S. Continual lifelong learning with neural networks: A review. Neural Netw. 2019;113:54–71.
- 171. Li Z, Hoiem D. Learning without forgetting. IEEE Trans Pattern Anal Mach Intell. 2017;40(12):2935–47.
- 172. Rusu AA, Rabinowitz NC, Desjardins G, Soyer H, Kirkpatrick J, Kavukcuoglu K, Pascanu R, and Hadsell R. Progressive neural networks. arXiv preprint arXiv:1606.04671, 2016.
- Shin H, Lee JK, Kim J, and Kim J. Continual learning with deep generative replay. Advances in neural information processing systems. 2017;30:2994-003.
- 174. Lipton ZC. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue. 2018;16(3):31–57.
- 175. Bau D, Zhu J-Y, Strobelt H, Lapedriza A, Zhou B, Torralba A. Understanding the role of individual units in a deep neural network. Proc Natl Acad Sci. 2020;117(48):30071–8.
- 176. Liao J-X, Dong H-C, Sun Z-Q, Sun J, Zhang S, and Fan F-L. Attention-embedded quadratic network (qttention) for effective and interpretable bearing fault diagnosis. arXiv preprint arXiv: 2206.00390, 2022.
- MacMahon B. Prenatal x-ray exposure and childhood cancer. J Natl Cancer Inst. 1962;28(5):1173–91.
- 178. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A, and Bottou L. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion.J Mach Learn Res. 2010;11(12):3371–408.
- Fan F, Shan H, Kalra MK, Singh R, Qian G, Getzin M, Teng Y, Hahn J, Wang G. Quadratic autoencoder (q-ae) for low-dose ct denoising. IEEE Trans Medical Imaging. 2019;39(6):2035–50.
- Bonnett AH, Yung C. Increased efficiency versus increased reliability. IEEE Ind Appl Mag. 2008;14(1):29–36.
- McFadden PD, Smith JD. Model for the vibration produced by a single point defect in a rolling element bearing. J Sound Vib. 1984;96(1):69–82.
- George Em Karniadakis. Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning Nature Reviews Physics. 2021;3(6):422–40.

- Shen J, Yang X. Numerical approximations of allen-cahn and cahn-hilliard equations. Discrete Contin Dynam Systems. 2010;28(4):1669.
- Szeliski R. Computer vision: algorithms and applications. Berlin, Germany: Springer Nature; 2022.
- Gallego JA, Perich MG, Miller LE, Solla SA. Neural manifolds for the control of movement. Neuron. 2017;94(5):978–84.
- Gallego JA, Perich MG, Naufel PN, Ethier C, Solla SA, Miller LE. Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. Nat Commun. 2048;9(1):1–13.
- 187. Dahmen D, Layer M, Deutz L, Dabrowska PA, Voges N, von Papen M, Brochier T, Riehle A, Diesmann M, Grün S, et al. Global organization of neuronal activity only requires unstructured local connectivity. Elife. 2022;11:e68422.
- 188. Elsken T, Metzen JH, Hutter F, et al. Neural architecture search: A survey. J Mach Learn Res. 2019;20(55):1–21.
- He X, Zhao K, Chu X. Automl: A survey of the state-of-the-art. Knowl-Based Syst. 2021;212: 106622.
- Tkacik G, Prentice JS, Balasubramanian V, Schneidman E. Optimal population coding by noisy spiking neurons. Proceed National Academy Sci. 2010;107(32):14419–24.
- 191. Ronneberger O, Fischer P, and Brox T. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention. New York City, NY: Springer; 2015. p 234–41.
- Han D, Kim J, Kim J. Deep pyramidal residual networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 5927–35.
- 193. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770–8.
- Schmidt M, Lipson H. Distilling free-form natural laws from experimental data. Science. 2009;324(5923):81–5.
- Cohen N, Sharir O, and Shashua A. On the expressive power of deep learning: A tensor analysis. In Conference on learning theory. 2016:698–728.
- 196. Poole B, Lahiri S, Raghu M, Sohl-Dickstein J, and Ganguli S. Exponential expressivity in deep neural networks through transient chaos. Adv Neural Inf Process Syst. 2016;29:3368-76.
- 197. Jacot, A., Gabriel, F., & Hongler, C. Neural tangent kernel: convergence and generalization in neural networks. In Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing. 2021:6–6.
- Belkin M, Hsu D, Ma S, Mandal S. Reconciling modern machine-learning practice and the classical bias-variance tradeoff. Proc Natl Acad Sci. 2019;116(32):15849–54.
- 199. Hornik K, Stinchcombe M, White H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Netw. 1990;3(5):551–60.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.