

# Diverse Neuron Type Selection for Convolutional Neural Networks

Guibo Zhu<sup>1,2</sup>, Zhaoxiang Zhang<sup>1,2,3,4\*</sup>, Xu-Yao Zhang<sup>2</sup>, Cheng-Lin Liu<sup>2,3,4</sup>

<sup>1</sup>Research Center for Brain-inspired Intelligence,

Institute of Automation, Chinese Academy of Sciences (CASIA)

<sup>2</sup>National Laboratory of Pattern Recognition, CASIA

<sup>3</sup>Center for Excellence in Brain Science and Intelligence Technology, CAS

<sup>4</sup>University of Chinese Academy of Sciences

{guibo.zhu, zhaoxiang.zhang}@ia.ac.cn, {xyz, liucl}@nlpr.ia.ac.cn

## Abstract

The activation function for neurons is a prominent element in the deep learning architecture for obtaining high performance. Inspired by neuroscience findings, we introduce and define two types of neurons with different activation functions for artificial neural networks: excitatory and inhibitory neurons, which can be adaptively selected by self-learning. Based on the definition of neurons, in the paper we not only unify the mainstream activation functions, but also discuss the complementariness among these types of neurons. In addition, through the cooperation of excitatory and inhibitory neurons, we present a compositional activation function that leads to new state-of-the-art performance comparing to rectifier linear units. Finally, we hope that our framework not only gives a basic unified framework of the existing activation neurons to provide guidance for future design, but also contributes neurobiological explanations which can be treated as a window to bridge the gap between biology and computer science.

## 1 Introduction

Convolutional neural networks (CNNs) [Krizhevsky *et al.*, 2012] have recently been widely used in different computer vision tasks. Deep CNNs not only significantly improve the image recognition accuracy [Krizhevsky *et al.*, 2012; Simonyan and Zisserman, 2014b], but also play important roles in variety of vision tasks such as object detection [Girshick *et al.*, 2014], text detection [Shi *et al.*, 2016; Liao *et al.*, 2016], human activity recognition [Simonyan and Zisserman, 2014a], and air pollution [Zhang *et al.*, 2016]. To improve the performance of CNNs, there are so many techniques advancing in CNNs performance improvements which can be roughly divided into network structure design (*e.g.*, depth [Simonyan and Zisserman, 2014b], modified non-linear activation functions [Nair and Hinton, 2010]) and over-fitting preventing techniques (*e.g.*, normalization [Ioffe and Szegedy, 2015]).

Among these techniques, one key factor leading to the recent success of deep neural networks is the invention of the rectifier functions to model the neuron activation, *i.e.* Rectified Linear Unit (ReLU) [Nair and Hinton, 2010; Glorot *et al.*, 2011]. It accelerates the learning process [Nair and Hinton, 2010] and significantly improves the overall network performance compared with traditional sigmoid or tanh units [Glorot *et al.*, 2011] by suppressing the vanishing/exploding gradient problem in training deep networks. Due to the importance of the activation function, many researchers have proposed elaborate activation functions for improving the network performance. These works can be coarsely summarized as two categories: (1) modified ReLUs (*e.g.*, Leaky ReLU (LReLU) [Maas *et al.*, 2013], Parametric ReLU [He *et al.*, 2015], Random ReLU (RReLU) [Xu *et al.*, 2015], Exponential Linear Units (ELU) [Clevert *et al.*, 2015], Parametric ELU (PELU) [Trottier *et al.*, 2016], Multiple Parametric Exponential Linear Units (MPELU) [Li *et al.*, 2016]) which remove zero gradients with various functions in negative input, and (2) ensemble activation functions (*e.g.*, Maxout [Goodfellow *et al.*, 2013], Adaptive Piecewise Linear (APL) [Agostinelli *et al.*, 2014], Network in Network [Lin *et al.*, 2013], S-shaped ReLU (SReLU) [Jin *et al.*, 2015], Concatenated ReLU (CReLU) [Shang *et al.*, 2016]) which take advantage of the ensemble power.

Despite the prevalence of the activation functions, there are rarely proposals paying attention to their common properties in the perspective of neurological evidence. In the neuroscience domain, there are two main types of neurons in cerebral cortex (*i.e.*, the surface of the brain): the primary neuron is excitatory neuron and the relay neuron is inhibitory neuron [Markram *et al.*, 2004]. Under normal circumstances, two types of cells regulate themselves each other to keep brain in dynamic equilibrium of excitement and inhibition state [Pi *et al.*, 2013]. Therefore, they both play important roles in the neural network of brain. Inspired by the neurological knowledge, in this paper we propose two types of neurons with different activation functions for artificial neural networks: excitatory and inhibitory neurons, which are learned for the neuron type selection. Based on the proposed neurons, we not only unify the mainstream activation functions, but also discuss the complementariness among these neuron types in the view of neuroscience. Additionally, through the cooperation of excitatory and inhibitory neurons, experiments

\*Corresponding author.

demonstrate that the proposed activation functions work well for training neural networks and achieve state-of-the-art performance on several image recognition benchmarks.

## 2 Related Works

We review the existing works related to activation functions, which are the focus of our research in this paper. They can mainly be divided into two categories: rectified unit family and ensemble activation function.

### 2.1 Rectified Unit Family

Since activation functions help to bring non-linearity, they play an important role in deep neural networks for their significant effect on the overall performance. ReLU [Nair and Hinton, 2010] used linear activation function to replace the saturated counterpart of the convolutional sigmoid or tanh-like units for positive inputs and set the output of negative inputs as zeros, *i.e.*,  $f(x) = \max(0, x)$ . It alleviated the exploding/vanishing gradient problem and expedited the convergence rate. Different from ReLU, LReLU [Maas *et al.*, 2013], PReLU and RReLU incorporated a positive slope parameter for negative input in activation units to remove zero gradients. They are different from each other in the acquisition methods for the slope parameter, where LReLU is pre-defined before the training process, PReLU is adaptively learned and RReLU is randomly sampling based on stochastic regularization in the training process. ELU [Clevert *et al.*, 2015] was defined as an identify function for positive inputs and  $\exp(x) - 1$  for negative ones, which can handle the largen variance and bias shift problems by allowing the mean activations closer to zeros. Similar to PReLU, MPELU [Li *et al.*, 2016] introduced a non-zero slope parameter for the negative parts of ELU and proposed a weight initialization strategy for ELU series. PELU [Trottier *et al.*, 2016] introduced several slope parameters in positive/negative parts and adjusted them in training by modifying ELU with differentiability near zero.

### 2.2 Ensemble Activation Function

Ensemble activation functions have powerful approximate ability for any smooth functions in theory. Maxout [Goodfellow *et al.*, 2013] used the maxout activation unit to output the maximum over several learned transformation functions for each neuron input. Adaptive Piecewise Linear (APL) [Agostinelli *et al.*, 2014] was a weight sum of parameterized hinge-shaped functions whose parameters are learned during training. Network in Network [Lin *et al.*, 2013] approximated the nonlinear rectified activation function in convolution networks with a multilayer perceptron in which the parameters are learned from data. S-shaped Rectified Linear Activation Units (SReLU) [Jin *et al.*, 2015] formulated three piecewise linear functions with four learnable parameters which can imitate complex convex and non-convex functions in psychophysics and neural sciences. With the observed property of filters with opposite phase in lower layers, Concatenated ReLU (CReLU) [Shang *et al.*, 2016] was designed by making an identical copy of the output values of the convolution layer, negating them, concatenating both the negated part and the original one, and then applying ReLU altogether.

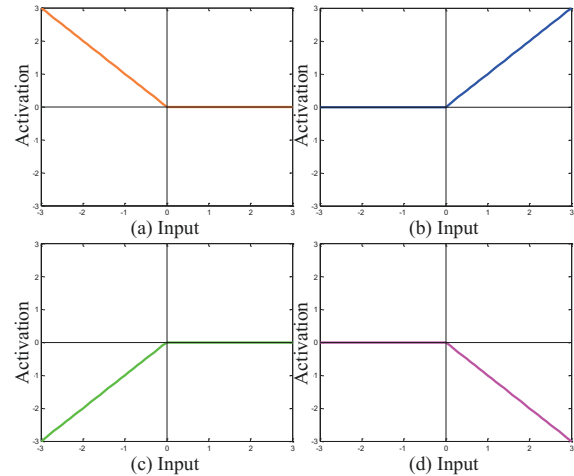


Figure 1: Four simple excitatory activation functions.

## 3 Excitatory and Inhibitory Activation Function

In this section, we first give the definitions of excitatory and inhibitory neurons for artificial neural networks. Then we propose the compositional activation function of the two neurons. Finally, we discuss the intrinsic relationships among the mainstream activation functions and give some explanations.

### 3.1 Two Types of Neurons

Inspired by the neuroscience [Markram *et al.*, 2004], we can define two types of neurons for artificial neural networks. Not every signal is useful for a particular task in human decision-making, therefore, there exit two types of different neurons. The first is “**excitatory neuron**” which is a function preserving the input signal well. The second is “**inhibitory neuron**” that will suppress the input signal. In this way, the neural networks can select the signals dynamically according to the requirement of the task in a data-driven manner. In previous researches, most widely-used neurons are excitatory, and the inhibitory neuron is not well-studied in the literature. In this paper, we will show that both of these two types of neurons are important and complementary for neural networks.

Actually, the excitatory and inhibitory neurons can be defined in many different formulations. In this paper, inspired from the Occam’s razor theory (*i.e.*, entities should not be multiplied unnecessarily), we give simple and straightforward definitions for them. **Excitatory neuron:** The neuron can be defined as:

$$x \rightarrow +\infty \quad \text{or} \quad x \rightarrow -\infty, |f(x)| \rightarrow +\infty \quad (1)$$

which means the output  $f(x)$  preserves the absolute value of the input  $x$  without suppression under some conditions.

For an excitatory neuron, the simplest linear activation function is  $f(x) = x$ . However, it can not make an artificial neural network possess the ability of nonlinear mapping property. For better nonlinear data fitting, the simplest non-linear activation function is  $f(x) = \max(x, 0)$ , *i.e.*,

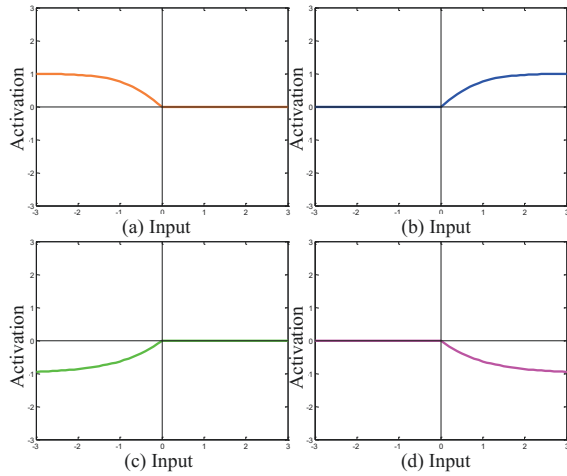


Figure 2: Four simple inhibitory activation functions.

ReLU [Nair and Hinton, 2010]. Because it satisfies the condition of Eq. (1), we think it is a representative case of excitatory neuron. Since there are more than one excitatory neurons in neural networks, here we only present four excitatory activation neurons in different quadrants as shown in Fig. 1. They are:  $f_1(x)$ ,  $f_2(x)$ ,  $f_3(x)$  and  $f_4(x)$  where  $f_1(x) = \max(x, 0)$  in the first quadrant as shown in Fig. 1(b),  $f_2(x) = \max(-x, 0)$  in the second quadrant as shown in Fig. 1(a),  $f_3(x) = -\max(-x, 0)$  in the third quadrant Fig. 1(c) and  $f_4(x) = -\max(x, 0)$  in the fourth quadrant as shown in Fig. 1(d), respectively. These activation neurons are excitatory because they all satisfy Eq. (1).

**Inhibitory Neuron:** Not only are the excitatory neurons important, but also the inhibitory neurons play a key role in neural network [van Vreeswijk and Sompolinsky, 1996]. With the inhibitory neurons, one neural network can keep in equilibrium by restraining various neuron responses in different layers. Here we also give the definition of the inhibitory neuron as:

$$x \rightarrow +\infty \quad \text{or} \quad x \rightarrow -\infty, |f(x)| \rightarrow \text{const} \quad (2)$$

which means that the output of inhibitory neuron  $f(x)$  does not exceed a “constant” or tend towards stability even though the absolute value of the input signal  $x$  tends to infinity.

There are many inhibitory neurons with different activation functions. To simulate the inhibitory activation neurons which restrain the input signal, in this paper we just introduce four inhibitory activation functions in each quadrant as shown in Fig. 2 corresponding to the excitatory neurons. They are:  $f_5(x)$ ,  $f_6(x)$ ,  $f_7(x)$  and  $f_8(x)$  where  $f_5(x) = \tanh(\max(x, 0))$  in the first quadrant as shown in Fig. 2(b),  $f_6(x) = \tanh(\max(-x, 0))$  in the second quadrant as shown in Fig. 2(a),  $f_7(x) = \exp(-\max(-x, 0)) - 1$  in the third quadrant as shown in Fig. 2(c) and  $f_8(x) = \exp(-\max(x, 0)) - 1$  as shown in Fig. 2(d), respectively.

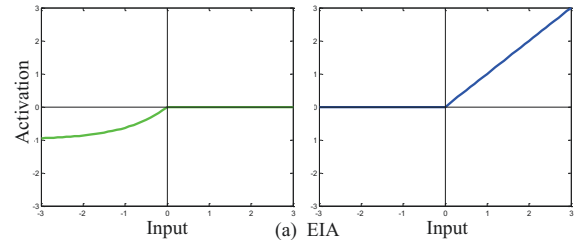


Figure 3: Diagram of the compositional excitatory and inhibitory activation (EIA) function. EIA:  $(e_1(x), e_2(x))$ , where  $e_1(x) = \tanh(\max(x, 0))$  in the first quadrant,  $e_2(x) = \alpha \exp(\min(\frac{x}{\beta}, 0)) - 1$  in the third quadrant.

### 3.2 Composition of Excitatory and Inhibitory Neurons

Excitatory neuron preserves the input signal while the inhibitory neuron suppresses the signal. They play different roles in neural networks. There are three most possible compositions among excitatory and inhibitory neurons: (1) excitatory and excitatory neurons; (2) excitatory and inhibitory neurons; (3) inhibitory and inhibitory neurons.

**Excitatory and Excitatory Neurons:** CReLU [Shang *et al.*, 2016] is the representative one which is concatenated by the excitatory neurons or activation functions with Fig. 1(a) and Fig. 1(b) in parallel. Meanwhile, the importance of the ways to combine excitatory neurons is different. Specially, not any two excitatory neurons to combine is a good choice because of their relevance. If they are independent, they help each other; otherwise, the coupling relationship between them may lead to the oscillation of the learning process, *e.g.*, the composition of Fig. 1(b) and Fig. 1(d).

**Excitatory and Inhibitory Neurons:** We need to explore which composition between excitatory activation function and inhibitory activation function has the best complementarity. After investigating the well-known rectifiers, *e.g.*, LReLU [Maas *et al.*, 2013], PReLU [He *et al.*, 2015], ELU [Clevert *et al.*, 2015] and PELU [Trottier *et al.*, 2016], we argue that the parallel composition of the excitatory neuron in Fig. 1(b) and the inhibitory neuron in Fig. 2(c) has better complementarity with each other.

Following the complementary direction of neuron compositions, we design an excitatory and inhibitory activation (EIA) function with different neurons as shown in Fig. 3, represented by  $\delta_n : \Theta \rightarrow \Theta^2$  activation as follows:  $\forall x \in \Theta, \delta_n \triangleq (e_1(x), e_2(x))$ . Here,  $e_1(x) = \max(x, 0)$ ,  $e_2(x) = \alpha \exp(\min(\frac{x}{\beta}, 0)) - 1$ . The optimization algorithm of the parameters  $\alpha$  and  $\beta$  is back-propagation similar to [Trottier *et al.*, 2016]. The motivation of our EIA method is not only to reduce the bias shift of layer-wise convolution, but also keep the filter to be activated in an adaptive way. (*e.g.*, in the positive and negative direction or the antipodal inhibited way).

**Inhibitory and Inhibitory Neurons:** There exists the composition of inhibitory and inhibitory neurons [Kvitsiani *et al.*, 2013]. Tanh [Hinton *et al.*, 1986], *i.e.*,  $f(x) = \frac{e^{2x}-1}{e^{2x}+1}$ , is the addition of two inhibitory neurons  $f_5(x)$  and  $f_6(x)$ .

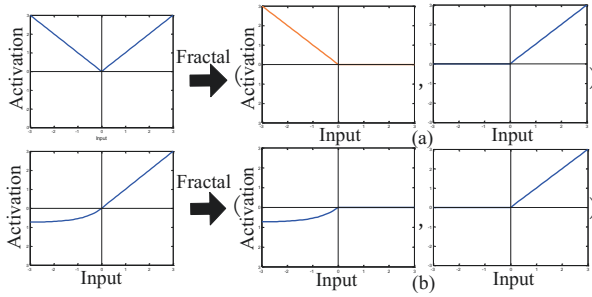


Figure 4: Diagram of fractal assumption. The original activation function is divided into two sub-activation functions with different quadrants. For example, (a) AVR,  $f(x) = \max(x, 0) + \max(-x, 0)$ , is transformed into the composition of  $f_1(x)$  (i.e. ReLU) and  $f_2(x)$ . (b) ELU generates the parallel structure of  $f_1(x)$  and  $f_7(x)$ .

**Neurological Guess:** The proportion of inhibitory neurons in total neurons is almost 10% [Markram *et al.*, 2004] while the other is mainly excitatory neurons. Suppose the negative input needs to be suppressed by inhibitory neurons and the positive output transmitted by the excitatory activation, the weight ratio between them should be 1 : 9. Given an input signal  $\mathbf{x}$  after preprocessing of batch normalization [Ioffe and Szegedy, 2015], it satisfies the normal gaussian distribution  $\mathbf{x} \sim N(\mu, \delta^2)$ . We introduce a general activation function:

$$g(x) = \begin{cases} g_1(x), & \text{if } x \geq 0, \text{ excitatory} \\ g_2(x), & \text{if } x < 0, \text{ inhibitory} \end{cases} \quad (3)$$

According to the ratio of excitatory activation and inhibitory activation, we guess there exists the expression as follows,

$$\int_{-\infty}^0 |g_2(x)|p(x)dx : \int_0^{+\infty} |g_1(x)|p(x)dx \approx 1 : 9 \quad (4)$$

where  $\int$  is the integral operator,  $|\cdot|$  gets the absolute value and  $p(x)$  is the probability density function. According to Eq. (4) and numerical integration algorithm, given  $g_1(x) = x$  and  $g_2(x) = \alpha x$ , the parameter  $\alpha$  is optimally initialized as  $\frac{1}{9}$ .

To make the excitatory neurons play a leading role, we need to penalize the inhibitory neurons with adaptive slope. In the literature, to handle the zero gradient problem during optimization, Maas *et al.* [Maas *et al.*, 2013] proposed the non-zero slope for the activation function of the negative input counterpart which is potentially more robust. Inspired by PReLU [He *et al.*, 2015], ELU [Clevert *et al.*, 2015] and PELU [Trottier *et al.*, 2016], the parameter with adaptive learning is a good strategy which reduces the overfitting risk of human interventions and alleviates the dynamic gradient propagation. More punishment reduces the noise effect of smaller negative input. In addition, the adaptive slope can be learned itself through the end-to-end network training.

**Fractal Assumption:** Divide an activation function into different quadrants and then conquer them as shown in Fig. 4, which are complementary.

The objectives of deep neural networks [Bengio, 2009; Glorot *et al.*, 2011] is learning to disentangle and represent

the factors of variation in the data. While concatenating with independent and appropriate quadrant functions, there are many advantages of sparsity (e.g., information disentangling, efficient variable-size representation, linear separability and distributed but sparse) [Glorot *et al.*, 2011] for the deep neural networks. The importance of different quadrants is statistically mining the collective knowledge in the literature and our empirical experiments.

### 3.3 Neurological Explanation

We will introduce the advantages of excitatory and inhibitory neurons from three aspects: using only one type of excitatory neuron, the collaboration of two types of excitatory neurons and the complementarity of excitatory and inhibitory neurons. Since excitatory neurons take a dominant role for neuroscience, only one type of excitatory neuron can still work very well in the neural networks for multiple tasks, such as the most widely used ReLU activation function [Nair and Hinton, 2010]. On the other hand, there are a variety of neurons which can construct different activation functions. For example, since CReLU [Shang *et al.*, 2016] is the collaboration of two types of excitatory neurons, its performance is satisfactory. In fact, we know that there are also many inhibitory neurons in the neural networks from neurological knowledge. Although the excitatory neuron is very important, it is better to add inhibitory neurons into excitatory neurons for the high-level semantic cognition task where the different neurons are complementary. Therefore, we propose the compositional excitatory and inhibitory activation (EIA) function which leads to much better performance in the following experiments.

## 4 Experiments

To evaluate the effects of our proposed activation functions, we conducted experiments on three datasets including CIFAR-10 and CIFAR-100 [Krizhevsky and Hinton, 2009] with the most common used deep neural network VGG [Simonyan and Zisserman, 2014b] and the ImageNet benchmark dataset [Deng *et al.*, 2009] with the AlexNet [Krizhevsky *et al.*, 2012] network architecture. The adopted VGG was modified with 19 layers and batch normalization for implementing on CIFAR-10/100. We chose VGG and CIFAR-10/100 to perform an in-depth investigation because the computational cost of a large network on the datasets is moderate and about ten hours on a K40 GPU. Note that the implementation of the VGG architecture was following to [Zagoruyko and Komodakis, 2016] who had released their source code. Except only changing the types of the activation functions for fair comparison, we kept all the hyperparameters remaining the same for all experiments. Additionally, we selected AlexNet with batch normalization using Torch for evaluating our proposed EIA activation function in ImageNet.

### 4.1 Datasets

The CIFAR-10/100 datasets [Krizhevsky and Hinton, 2009] contain 50,000 training and 10,000 testing samples with size of  $32 \times 32$  color images from 10 and 100 classes, respectively. The data preprocessing techniques follow the method used in [Goodfellow *et al.*, 2013] ImageNet includes 1.3M training and 100K testing images for 1000 categories with one crop.

Table 1: Comparison of one type of neurons.

Excitatory Neuron Type		Inhibitory Neuron Type	
Name	CIFAR-10/100	Name	CIFAR-10/100
E1	<b>93.95/73.17</b>	I1	93.58/72.42
E2	93.67/72.5	I2	93.83/72.09
E3	93.13/71.17	I3	93.09/70.65
E4	93.26/71.29	I4	92.91/70.64

### 4.2 Evaluation on Single Neuron Type

To evaluate different kinds of excitatory neuron types, we designed several experiments with these neurons. First, we tested four types of excitatory neurons. *ReLU* [Nair and Hinton, 2010] was treated as the first type of excitatory neuron *E1*. *E2*, *E3*, and *E4* were the remaining types of excitatory neurons corresponding to the activation functions  $f_2(x)$ ,  $f_3(x)$  and  $f_4(x)$  respectively. Similarly, four types of inhibitory neurons (*i.e.*, *I1*, *I2*, *I3*, and *I4* corresponding to activation functions  $f_5(x)$ ,  $f_6(x)$ ,  $f_7(x)$  and  $f_8(x)$ ) were also compared. However, these inhibitory neurons were not convergent with the original VGG model. Even with batch normalization, their performance still had significant difference and worse than the excitatory neurons. As shown in Table 1, the first type of excitatory neuron achieved the best performance in these neuron types.

### 4.3 Evaluation on Coupling Excitatory Neurons

To investigate the cooperation with excitatory and excitatory neurons, we compared various compositions with different types of excitatory neurons. These compositions are as follows: *AVR* is  $f(x) = f_1(x) + f_2(x)$ , *EE-1-2* is  $f(x) = (f_1(x), f_2(x))$ , *EE-1-3* is  $f(x) = (f_1(x), f_3(x))$ , *EE-1-4* is  $f(x) = (f_1(x), f_4(x))$ , *EE-2-3* is  $f(x) = (f_2(x), f_3(x))$ , *EE-2-4* is  $f(x) = (f_2(x), f_4(x))$  and *EE-3-4* is  $f(x) = (f_3(x), f_4(x))$ . As shown in Table 2, we found that the performance of compositions with Index 2-7 was all better than *AVR* which meant that the concatenated structure in parallel was better. That was because the concatenated structure in parallel not only utilized the advantage of sparsity, but also explored the complementary information among independent quadrants which consisted of the phase information. While evaluating the compositions with Index 2-7, we found that concatenations with independent quadrant functions in the diagonal line (*e.g.*, *EE-1-2*, *EE-1-3*, *EE-3-4*) were better than the coupling quadrant functions at only one side of axis (*e.g.*

Table 2: Comparison among different excitatory neuron composition. EE: excitatory and excitatory. EE-1-2 is CReLU.

Index	Quad.	Name	CIFAR-10	CIFAR-100
1	1, 2	AVR	92.62	65.44
2	1, 2	EE-1-2	93.98	<b>73.22</b>
3	1, 3	EE-1-3	93.84	73.16
4	1, 4	EE-1-4	93.69	72.96
5	2, 3	EE-2-3	93.68	73.08
6	2, 4	EE-2-4	<b>94.08</b>	72.73
7	3, 4	EE-3-4	93.92	73.03

Table 3: Comparison among compositions of different excitatory and inhibitory neurons. EI: excitatory and inhibitory.

Name	CIFAR-10	CIFAR-100
EI-1-3	93.76	73.16
EI-2-4	93.51	73.32
EIA	<b>94.22</b>	<b>73.86</b>

, *EE-2-3* and *EE-1-4*). Additionally, the best composition excitatory and excitatory neurons was better than the best type of neuron in Table 1 which showed their collaboration.

### 4.4 Evaluation on Pairs of One Excitatory Neuron and One Inhibitory Neuron

To evaluate the complementariness of excitatory and inhibitory neurons, we conducted several compositions among them. They were *EI-1-3*, *EI-2-4* and *EIA* where *EI-1-3* was constructed by the first type of excitatory neuron  $f_1(x)$  and the third inhibitory neuron  $f_7(x)$ , *EI-2-4* was  $f(x) = (f_2(x), f_8(x))$  and *EIA* was our proposed activation function. The main difference between *EI-1-3* and *EIA* was that we penalized the inhibitory neurons with adaptive slopes from self-learning. Table 3 showed that *EIA* got the best performance which verified the importance of diverse neuron type selection coming from self-learning in the end-to-end convolutional neural network structure.

Besides, we also compared paired inhibitory and inhibitory neurons with *EIA*. Their performance was also not as good as *EIA*. For simplicity, we did not give the results here.

### 4.5 Evaluation of Fractal Assumption

In commercial applications, we need to consider the limitation of device memory and time complexity. With the fractal assumption of different excitatory and inhibitory neurons, we can select various neurons to design new activation functions. For example, *AVR* can be constructed by two excitatory neurons, *i.e.*,  $f_1(x)$  and  $f_2(x)$ , *Linear* can be generated by  $f_1(x)$  and  $f_3(x)$  and *ELU* can be divided into one excitatory neuron  $f_1(x)$  and one inhibitory neuron  $f_7(x)$ . There are the fractal and concatenated structures among them. Specifically, *AVR*,  $f(x) = \max(x, 0) + \max(-x, 0)$ , equals to the inverse composition result of *EE-1-2*. There was the same relation between *Linear*,  $f(x) = \max(x, 0) + \min(x, 0)$ , and *EE-1-3*,  $f(x) = (\max(x, 0), \min(x, 0))$ . As show in Table 4, we found that the fractal activation functions (*i.e.*, *CReLU* and *EE-1-3*) had a significant improvement over their corresponding original activation functions (*i.e.*, *AVR* and *Linear*). Therefore, the fractal and concatenated structure played an important role in activation functions.

Table 4: Comparison for fractal assumption.

Index	Quad.	Name	CIFAR-10	CIFAR-100
1	1, 2	AVR	92.62	65.44
2	1, 2	CReLU	<b>93.98</b>	<b>73.22</b>
3	1, 3	Linear	88.36	56.37
4	1, 3	EE-1-3	<b>93.84</b>	<b>73.16</b>

Table 5: Evaluation on the recognition performance with different ratio of excitatory and inhibitory neurons.

Name	$\alpha$	CIFAR-10	CIFAR-100
ELU	0.11	<b>93.97</b>	<b>73.68</b>
ELU	0.2	93.67	73.49
ELU	0.3	93.42	73.09
ELU	0.5	93.01	72.18
ELU	0.6	92.56	70.85
ELU	0.7	92.21	63.16
ELU	1	91.32	67.28

To further evaluate the effects of excitatory and inhibitory neurons with different ratios, we adopted the addition of the first excitatory neuron and the third inhibitory neuron, *i.e.*, ELU [Clevert *et al.*, 2015] whose activation function is  $f(x) = f_1(x) + \alpha f_7(x)$ , for the baseline. Here, the adaptive penalized parameter  $\alpha$  was set to different values. As shown in Table 5, we found that as the parameter  $\alpha$  changes, ELU had significant performance differences in CIFAR-10 and CIFAR-100 datasets. While  $\alpha = 0.11$ , ELU achieved the best performance. This was corresponding with the neurological guess in Section 3.2. In some extent, it showed the relevance between computer science and neurological evidences.

#### 4.6 Comparison with other State-of-the-Art Activation Functions

We compared the excitatory and inhibitory activation (EIA) function with other state-of-the-art activation functions, *i.e.*, ReLU [Nair and Hinton, 2010], PReLU [He *et al.*, 2015], CReLU [Shah *et al.*, 2016] and PELU [Trottier *et al.*, 2016]. VGG-Large-2 was increased the filter number as fair baseline for comparison with CReLU and EIA. From Table 6, we found that the EIA function achieved the best performance. Specifically, EIA not only took advantage of the phase information and modulus information of different fractal quadrants, but also utilized the complementarity of excitatory and inhibitory neurons.

#### 4.7 ImageNet Challenge Dataset

To evaluate the performance of our proposed activation functions in large dataset benchmark, we performed the activation function of EIA on the 1000-class ImageNet dataset, which consists of about 1.3M training color images. For this task, we adopted AlexNet [Krizhevsky *et al.*, 2012] with batch normalization for test. In Table 7, we found that our proposed activation function achieved the best performance (*i.e.*

Table 6: Comparison of different activation functions. (.) corresponds to the denoted name in the paper.

Model	CIFAR-10	CIFAR-100
ReLU(E1)	93.95	73.17
PReLU(adaptive EE-1-3)	93.97	73.06
PELU(adaptive EI-1-3)	93.61	73.16
CReLU(EE-1-2)	93.98	73.22
VGG-Large-2	94.05	73.14
<b>Ours</b>	<b>94.22</b>	<b>73.86</b>

Table 7: Comparison of different activation functions in the large ImageNet benchmark dataset. The filter number of PELU-large is  $\sqrt{2}$  times as more as PELU's.

Model	ReLU	CReLU	PELU	PELU-large	Ours
Top-1(%)	58.05	58.64	59.55	61.32	<b>62.35</b>

, the highest testing accuracy) compared with the state-of-the-art activation functions ReLU [Nair and Hinton, 2010] and CReLU [Shang *et al.*, 2016] in the large ImageNet benchmark dataset. The experiments also demonstrated the effectiveness of the proposed EIA function.

### 5 Discussion

In this section, we discussed the relationship between our findings and neuroscience. For parallel and distributed biological neurons, the expected firing rate is simulating the activation function while there are signals at synapse incoming [Dayan and Abbott, 2001; Glorot *et al.*, 2011]. The biological activation function consists of features: unilateral inhibition, sparse activation, and antisymmetric inhibitory or symmetric excitatory states. Inspired by biological knowledge, the proposed EIA function had relations with the above features. Fractal assumption was similar to unilateral inhibition, while the concatenated structure in parallel used sparse activation and inhibitory/excitatory states potentially. In some extent, the compositional activation functions provided understanding and guidance for future neurological findings.

### 6 Conclusion

In this paper, we proposed two types of neurons with different activation functions for artificial neural networks: excitatory and inhibitory neurons. With the help of different types of neurons, we designed the excitatory and inhibitory activation function which enabled the networks to be more sparse and alleviated the bias shift and vanishing gradients problems. Comprehensive experiments were not only in agreement with neuroscience findings, *e.g.*, the importance of excitatory and inhibitory neurons, but also demonstrated the proposed method had achieved state-of-the-art performance.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 61375036, 61511130079, 61403380, 61602481, and the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDB02060009. Zhaoxiang Zhang is the corresponding author of the paper.

### References

[Agostinelli *et al.*, 2014] Forest Agostinelli, Matthew Hoffman, Peter Sadowski, and Pierre Baldi. Learning activation functions to improve deep neural networks. *arXiv:1412.6830*, 2014.

[Bengio, 2009] Yoshua Bengio. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

- [Clevert *et al.*, 2015] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv:1511.07289*, 2015.
- [Dayan and Abbott, 2001] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience*, volume 10. Cambridge, MA: MIT Press, 2001.
- [Deng *et al.*, 2009] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, pages 248–255. IEEE, 2009.
- [Girshick *et al.*, 2014] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, pages 580–587, 2014.
- [Glorot *et al.*, 2011] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Aistats*, volume 15, page 275, 2011.
- [Goodfellow *et al.*, 2013] Ian J Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron C Courville, and Yoshua Bengio. Maxout networks. *ICML (3)*, 28:1319–1327, 2013.
- [He *et al.*, 2015] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV*, pages 1026–1034, 2015.
- [Hinton *et al.*, 1986] Geoffrey E Hinton, James L McClelland, and David E Rumelhart. Distributed representations, parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations, 1986.
- [Ioffe and Szegedy, 2015] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv:1502.03167*, 2015.
- [Jin *et al.*, 2015] Xiaojie Jin, Chunyan Xu, Jiashi Feng, Yunchao Wei, Junjun Xiong, and Shuicheng Yan. Deep learning with s-shaped rectified linear activation units. *arXiv:1512.07030*, 2015.
- [Krizhevsky and Hinton, 2009] Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. 2009.
- [Krizhevsky *et al.*, 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012.
- [Kvitsiani *et al.*, 2013] D Kvitsiani, S Ranade, B Hangya, H Taniguchi, JZ Huang, and A Kepecs. Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature*, 498(7454):363–366, 2013.
- [Li *et al.*, 2016] Yang Li, Chunxiao Fan, Yong Li, and Qiong Wu. Improving deep neural network with multiple parametric exponential linear units. *arXiv:1606.00305*, 2016.
- [Liao *et al.*, 2016] Minghui Liao, Baoguang Shi, Xiang Bai, Xinggang Wang, and Wenyu Liu. Textboxes: A fast text detector with a single deep neural network. *arXiv preprint arXiv:1611.06779*, 2016.
- [Lin *et al.*, 2013] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv:1312.4400*, 2013.
- [Maas *et al.*, 2013] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. ICML*, volume 30, 2013.
- [Markram *et al.*, 2004] Henry Markram, Maria Toledo-Rodriguez, Yun Wang, Anirudh Gupta, Gilad Silberberg, and Caizhi Wu. Interneurons of the neocortical inhibitory system. *Nature Reviews Neuroscience*, 5(10):793–807, 2004.
- [Nair and Hinton, 2010] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML*, pages 807–814, 2010.
- [Pi *et al.*, 2013] Hyun-Jae Pi, Balázs Hangya, Duda Kvitsiani, Joshua I Sanders, Z Josh Huang, and Adam Kepecs. Cortical interneurons that specialize in disinhibitory control. *Nature*, 503(7477):521–524, 2013.
- [Shah *et al.*, 2016] Anish Shah, Eashan Kadam, Hena Shah, and Sameer Shinde. Deep residual networks with exponential linear unit. *arXiv:1604.04112*, 2016.
- [Shang *et al.*, 2016] Wenling Shang, Kihyuk Sohn, Diogo Almeida, and Honglak Lee. Understanding and improving convolutional neural networks via concatenated rectified linear units. *arXiv:1603.05201*, 2016.
- [Shi *et al.*, 2016] Baoguang Shi, Xiang Bai, and Cong Yao. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.
- [Simonyan and Zisserman, 2014a] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in Neural Information Processing Systems*, pages 568–576, 2014.
- [Simonyan and Zisserman, 2014b] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*, 2014.
- [Trottier *et al.*, 2016] Ludovic Trottier, Philippe Giguère, and Brahim Chaib-draa. Parametric exponential linear unit for deep convolutional neural networks. *arXiv:1605.09332*, 2016.
- [van Vreeswijk and Sompolinsky, 1996] Carl van Vreeswijk and Haim Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724, 1996.
- [Xu *et al.*, 2015] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv:1505.00853*, 2015.
- [Zagoruyko and Komodakis, 2016] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv:1605.07146*, 2016.
- [Zhang *et al.*, 2016] Chao Zhang, Junchi Yan, Changsheng Li, Xiaoguang Rui, Liang Liu, and Rongfang Bie. On estimating air pollution from photos using convolutional neural network. In *MM*, pages 297–301. ACM, 2016.