ICCES Online ICCES, vol.1, no.2, pp.36-39, 2019

New Activation Functions in CNN and Its Applications

Tomoyuki Enomoto, Kazuhiko Kakuda and Shinichiro Miura

Nihon University, Narashino, Chiba 275-8575, Japan.
*Corresponding Author: Tomoyuki Enomoto. Email: cito18002@g.nihon-u.ac.jp.

Abstract. In this paper, the nonlinear activation functions based on fluid dynamics are presented. We propose two types of activation functions by applying the so-called parametric softsign to the negative region. We apply the activation function to CNN (Convolutional Neural Network) which performs image recognition and approaches from multiple benchmark datasets such as MNIST, CIFAR-10. Numerical results demonstrate the workability and the validity of the present approach through comparison with other numerical performances.

Keywords: Deep learning; CNN; activation function; fluid dynamics; CIFAR-10

1 Introduction

The appropriate choice of the activation functions for neural networks is a key factor in the deep learning framework. Currently, the standard activation function is the rectified linear unit (ReLU) [1]. However, the ReLU has been pointed out several drawbacks. For example, there is a problem that so-called dying unit issue becase the slope of the negative region is always 0. Therefore, new activation functions have also been proposed such as PReLU [2] and ELU [3]. The PReLU mitigates this problem by applying a parametric slope to the negative region. In addition, the ELU is more robust against noise than ReLU and PReLU by nonlinearly extending the negative region.

The purpose of this paper is to propose the activation functions based on the concept of fluid dynamics framework. We present two types of activation functions by applying the so-called parametric softsign to the negative part of ReLU. Also, we apply the activation function to CNN (Convolutional Neural Network) which performs image recognition and approaches from multiple benchmark datasets such as MNIST, CIFAR-10. Numerical results demonstrate the workability and the validity of the present approach through comparison with other numerical performances.

2 Nonlinear Activation Functions

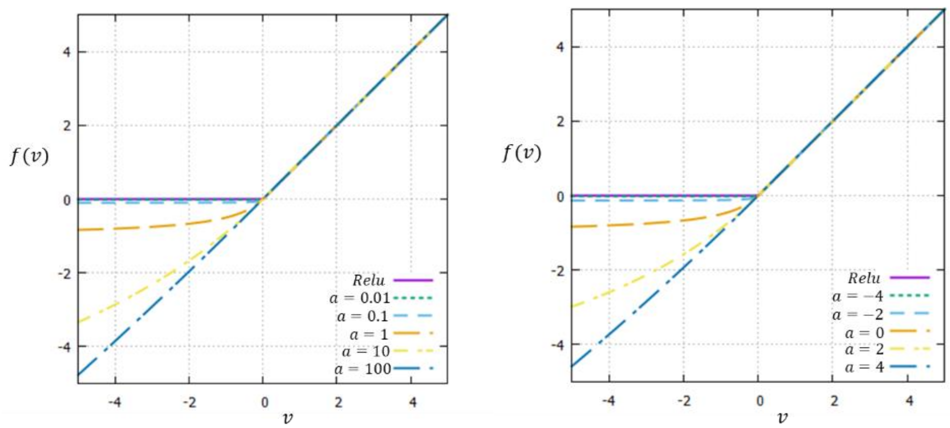
In this section, we present the activation functions on which is based the fluid dynamics framework. In order to avoid zero-gradients in the negative part, we propose two types of activation function involving parameter, a, as follows (see Fig. 1)

<u>Rational-type activation function</u>

$$f(v) = \begin{cases} v & (v \geq 0) \\ \frac{av}{a-v} & (v < 0) \end{cases} \tag{1}$$

<u>Exponential-type activation function</u>

$$f(v) = \begin{cases} v & (v \geq 0) \\ \frac{e^a v}{e^a - v} & (v < 0) \end{cases} \tag{2}$$



(a) Rational-type activation function (b) Exponential-type activation function

Figure 1: Nonlinear activation functions

The rational-type activation function converges to ReLU function at $a \to 0$ and converges to a linear function with slope 1 with $a \to +\infty$. The exponential-type activation function converges to the ReLU function with a $\to -\infty$, converging to a linear function with slope 1 with $a \to +\infty$.

We adopt the similar approach as the PReLU which can be trained using backpropagation and optimized simultaneously with other layers.

3 CNN Architecture

Fig. 2 shows the CNN architecture consisting of three convolutional layers with some max-pooling and one fully-connected softmax layer. The kernel size of each layer is 3 × 3, and the pooling size is 2 × 2. The number of kernels in the convolutional layer of each layer is 32.

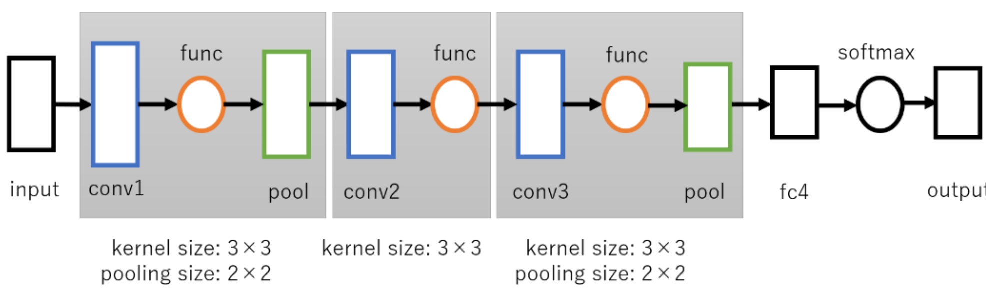
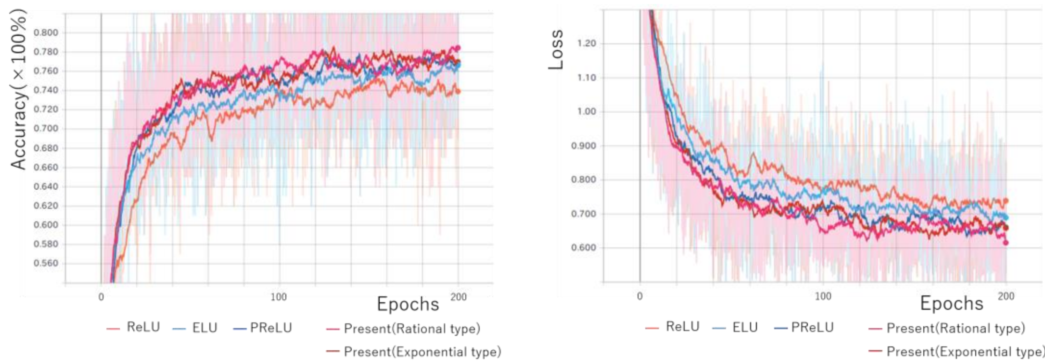


Figure 2: CNN architecture

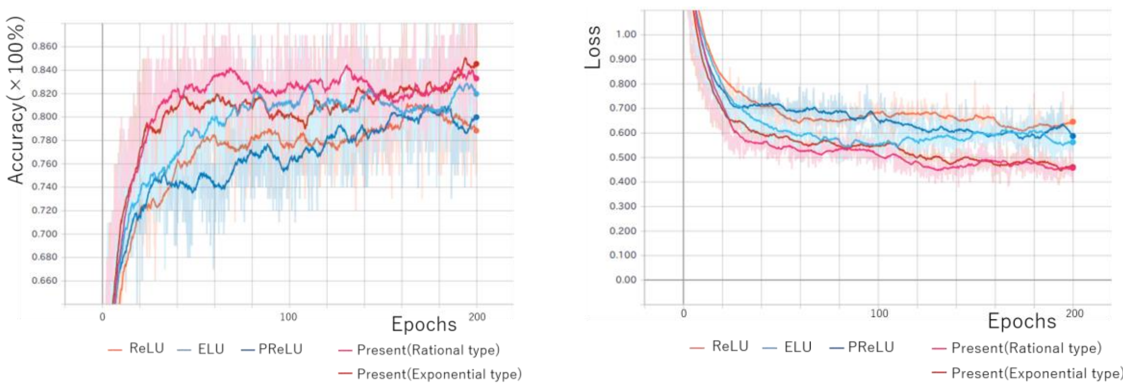
4 Numerical Experiment on CIFAR-10

In this section, we present numerical performances obtained from applications of the above-mentioned CNN approach to CIFAR-10. The model is previously trained for 200 epochs on mini-batches of size 100 with the learning rate, $\varepsilon = 10^{-3}$, and the momentum, $\mu = 0$.



(a) Training accuracy (b) Training loss

Figure 3: Training accuracy and loss behaviors for the CIFAR-10



(a) Validation accuracy (b) Validation loss

Figure 4: Validation accuracy and loss behaviors for the CIFAR-10

Table 1: Accuracy rate and loss for the CIFAR-10

	ReLU	ELU	PReLU	Present Rat.-type	Present Exp.-type
Accuracy [%]	77.68	80.40	79.76	80.23	80.76
Loss	0.56802	0.50304	0.55520	0.38574	0.39621

Fig. 3 shows the behaviors of training accuracy and loss obtained by using various activation functions for the CIFAR-10. The corresponding validation accuracy and loss behaviors are shown in Fig. 4. The accuracy rate and loss are given in Tab. 1.

4 Conclusions

We have proposed new activation functions which were based on the advection diffusion system in fluid dynamics framework. The learning performances demonstrated that our approaches were capable of recognizing accurately the object images in comparison with other ones.

References

1. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 2278-2324.

ICCES, vol.1, no.2, pp.36-39, 2019

2. Nair, V., Hinton, G. E. (2010). Rectified linear units improverestricted boltzmann machines. *ICML,* 807-814.
3. He, K. M., Zhang, X. Y., Ren, S. Q., Sun, J. (2015). Delving deep into rectifiers. *Proceedings of the 2015 IEEE International Conference on Computer Vision,* 1026-1034.