Journal of Electrical and Electronics Engineering (IOSR-JEEE)
e-ISSN: 2278-1676, p-ISSN: 2320-3331, Volume 20, Issue 3 Ser. 1 (May. – June. 2025), PP 57-59
www.iosrjournals.org

A Survey on Different Self-Gating Techniques of Activation Functions in Neural Networks

Israt Jahan

(Department of Electrical and Electronic Engineering, Noakhali Science and Technology University, Noakhali 3814, Bangladesh)

Corresponding Author: Israt Jahan (israt.eee@nstu.edu.bd)

Abstract:
Activation functions are considered the heart of the neural networks, as they enable models to learn and extract complex non-linear data patterns from diverse types of data. A significant number of activation functions have been proposed in the literature considering their own advantages and limitations. To further improve the performance of the activation functions, the self-gating technique has been introduced. In this mechanism, the activation output is modified by a learnable gating function. This technique enhances gradient flow and learning efficiency to ensure more robust training and improved generalization. To the best of our knowledge, this is the first survey dedicated to self-gated activation functions. In this paper, we present a comprehensive overview of self-gated activation functions, discussing their evolution, current challenges, and future research directions.

Keywords: *Activation function, neural network, self-gating technique.*

Date of Submission: 25-05-2025 Date of Acceptance: 05-06-2025

I. Introduction

In neural networks, a very common and basic component is the activation function. Generally, the activation function introduces non-linearity which enables to learn complex configurations in deep neural networks. Traditionally, sigmoid, Tanh, and Rectified Linear Unit (ReLU) are the activation functions used in neural networks. Sigmoid and Tanh suffer from the vanishing gradient problem because both of the function gets saturated at some point and eventually their derivates becomes very small whereas ReLU suffers from the dying ReLU problem which occurs because the gradient becomes zero in the negative region [1].

In order to overcome all these limitations, self-gated activation functions came to light. The self-gated mechanism was first introduced in Gated Linear Units (GLUs) [2], which is controlled by a separate gating function that is basically a sigmoid transformation. This mechanism increases gradient flow and learning efficiency. Swish [3], Sigmoid Linear Unit (SiLU) [4], Self-gated ReLU (SGReLU) [1], Mish [5], Meta-Activate or Not (Meta-ACON) [6], Gated Tanh Unit (GTU) [2], Gaussian Error Linear Unit (GELU) [7], SNAKE [8] are few upgrades of GLU. Because of the complex pattern of the self-gating mechanism, these activation functions show improved performance than traditional activation functions in deep neural networks. For instance, self-gated activation functions provide improved gradient flow, prevention of Dead Neurons, context-aware activation scaling, better generalization.

The contributions of this survey paper are as follows:

- To classify different gating techniques in different activation functions and summarize them.
- To summarize the contributions of existing survey papers related to activation functions and make a clear difference where this survey paper is different from the existing ones.

The rest of the paper is organized as follows: Section II presents a literature review of various self-gated activation functions and existing surveys on related topics. Section III provides the classifications of self-gated activation functions based on gating mechanisms, learnable parameters, structural complexity, and application context. Finally, Section IV concludes the paper and outlines the future research directions.

II. Literature Review

In recent years, a remarkable number of surveys on activation functions have been published, focusing on various aspects such as systematic analysis, classification and benchmarking, practical usage trends, trainable activation functions, and recent trends. To the best of our knowledge, however, no survey paper has focused on self-gated activation functions. A brief overview of the existing works is presented in Table 1.

Table 1. List of survey papers on activation functions in deep learning

Ref.	Year	Focus Areas	Key Contributions
[9]	2024	Extensive cataloging of activation functions	Presents an extensive survey of 400 activation functions, providing a detailed overview and framework of already published functions with references to their primary sources.
[10]	2022	Classification and benchmarking	Offers a comprehensive overview and survey of activation functions in deep neural networks for deep learning, covering various classes and providing performance comparisons among 18 state-of-the-art functions across different networks and datasets.
[11]	2021	Trainable activation functions	Discusses the organized framework of trainable activation functions, highlighting common and unique properties, and analysing their advantages and limitations.
[12]	2021	Systematic analysis	Offers an in depth yet latest overview of well-known activation functions and their properties, aiming to clarify the landscape for researchers and practitioners.
[13]	2018	Practical usage trends	Compiles an organized survey on documented activation functions used in deep learning applications, highlighting recent trends and comparing practical deployments against research findings.
[14]	2022	Recent developments	Explores the core ideas related to activation functions in neural networks, particularly their properties, types, challenges, limitations, and alternative solutions.
This Survey	NA	Gating Techniques	Describes different gating mechanisms of activation functions used in neural networks.

III. Gating Mechanism Classification

The general equation of the self-gated activation functions can be expressed as

$$f(x) = x.g(x), \tag{1}$$

where $g(x)$ is a gating function that modulates the activation of deep neural networks [15]. Nine commonly used activation functions, along with their mathematical formulations, gating mechanisms, and key characteristics, are presented in Table 2.

Table 2. Different Self-gated activation functions.

Activation Functions	Equation	Gating Mechanism	Key Characteristics
GLU [2]	$f(x) = (x.w + b) \otimes \sigma(x.v + c)$	Sigmoid	Modulates feature selection.
Swish [3]	$f(x) = x.\sigma(\beta x)$	Sigmoid	Smooth, better generalization, non-monotonic.
SiLU [4]	$f(x) = x.\sigma(x)$	Sigmoid	Less complex, efficient.
SGReLU [1]	$f(x) = \begin{cases} x, for\ x \geq 0 \\ \sigma(x), for\ x < 0 \end{cases}$	Sigmoid	Better generalization, better gradient flow.
Mish [5]	$f(x) = x.\tanh(\ln(1 + e^x))$	Tanh	Smooth, easy to find the global minima, introduces soft gating mechanism.
Meta-ACON [6]	$f(x) = (p_1x - p_2x).\sigma(\beta(p_1x - p_2x))$	Sigmoid	Highly flexible, input-adaptive.
GTU [2]	$f(x) = \tanh(x).\sigma(x)$	Hybrid	Smooth, bounded output.
GELU [7]	$f(x) = 0.5x(1 + \tanh[\sqrt{\frac{2}{\pi}}(x + 0.044715x^3)]$	Tanh	Includes the benefit of drop-out regularization.
SNAKE [8]	$f(x) = x + a.tanh^2(bx)$	Tanh	Periodic, captures every detail

From Table 2, it can be seen that different gating mechanisms contribute differently to shaping activation functions, which eventually helps in learning hidden patterns in neural networks and makes them more efficient for real-world problems. Sigmoid gating generally introduces better gradient flow and smooths the activation shape, whereas Tanh gating integrates the benefits of dropout regularization, which helps enhance model robustness. Hybrid gating incorporates the benefits of both sigmoid and Tanh gating.

In this paper, we have categorized self-gated activation functions based on their gating function, use of learnable parameters, structural complexity, and application context, as presented in Table 3, 4, 5, and 6, respectively.

In Table 3, a more summarized classification based on different gating mechanisms has been presented, and from this table, it is more evident why applying a gating mechanism is necessary. Most of the recent activation functions use these gating mechanisms to enhance the overall performance of a model.

Table 3. Classification of self-gated activation functions based on gating mechanisms.

Type	Gating Function	Range	Key Characteristics	Examples
Sigmoid-gated	$\sigma(x) = \frac{1}{1 + e^{-x}}$	(0, 1)	Smooth, continuous gating, used for feature modulation	Swish, SiLU, Meta-ACON
Tanh-gated	$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	(-1, 1)	Symmetric, allows negative gating	Mish, SNAKE
Hybrid gated	Combination of $\sigma(x)$ and $\tanh(x)$	both	Merges strengths of both sigmoid and tanh	GTU, Tanh-Swish

Table 4. Classification of self-gated activation functions based on learnable parameters.

Type	Equation Form	Characteristics	Examples
Parameter-free	Depends only on the input value without any trainable elements.	Lightweight, stable	SiLU, Mish
Parameterized	Involves trainable weights that modulate and combine input terms.	More expressive, adaptive	Swish, Meta-ACON

In Table 4, we have separated the activation functions into two categories: one is parameter-free, and the other is parameterized, which can learn during the model training. The parameter-free self-gated activation functions offer the benefit of lower time complexity, which is useful when real-time fast processing is needed. However, the parameterized self-gated activation functions are more accurate and adaptive.

Table 5. Classification of self-gated activation functions based on structural complexity.

Type	Characteristics	Examples
Simple	Input modulated by simple gating	Swish, SiLU
Self-learned	Involves multiple learnable pathways or gates	MetaACON
Periodic	Nonlinear and periodic self-gating	SNAKE

In Table 5, we have categorized the self-gated activation functions based on structural complexity into three categories: simple, self-learned, and periodic. The structural patterns also help the activation functions adopt different characteristics, which makes them suitable for different models—essential for solving various types of real-world problems.

Finally, from Table 6, it can be seen that different real-life application areas require different types of self-gated activation functions. In the vision area, Mish, Swish, and Meta-ACON are generally used, whereas GTU and SiLU are more suitable for attention-based models and Edge-AI-based models, respectively. So, it is essential to know which activation function is suitable for which purpose in order to achieve better results with less hassle.

Table 6. Classification of self-gated activation functions based on application context.

Area	Used Functions	Reasons
Vision	Mish, Swish, Meta-ACON	Better generalization and gradient flow
Attention	GTU	Compatibility with gating in RNNs and Transformers
Edge AI	SiLU	Efficient and hardware friendly

IV. Conclusion

This paper exclusively focuses on self-gated activation functions, which play a crucial role in the design of neural networks. These techniques include sigmoid gating, tanh gating, and hybrid gating mechanisms. We believe that this survey will assist researchers in understanding, developing, and applying self-gated activation functions in a broader context. However, the advantages of these methods often come at the cost of increased computational overhead. Therefore, further research is needed to develop lightweight and hardware-friendly variants that maintain performance while improving efficiency.

References

[1] I. Jahan, M. F. Ahmed, M. O. Ali, And Y. M. Jang, Self-Gated Rectified Linear Unit for Performance Improvement of Deep Neural Networks, ICT Express, 9 (3), 2023, 320–325.
[2] Y. N. Dauphin, A. Fan, M. Auli, And D. Grangier, Language Modeling with Gated Convolutional Networks, *arXiv [cs.CL]*, 2016.
[3] P. Ramachandran, B. Zoph, And Q. V. Le, Searching for Activation Functions, *arXiv [cs.NE]*, 2017.
[4] S. Elfwing, E. Uchibe, And K. Doya, Sigmoid-Weighted Linear Units for Neural Network Function Approximation in Reinforcement Learning, Neural Netw., 107, 2018, 3–11.
[5] D. Misra, Mish: A Self Regularized Non-Monotonic Activation Function, Proc. British Machine Vision Conference (BMVC), 2020.
[6] N. Ma, X. Zhang, M. Liu, And J. Sun, Activate or Not: Learning Customized Activation, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
[7] D. Hendrycks and K. Gimpel, Gaussian Error Linear Units (GELUs), *arXiv [cs.LG]*, 2016.
[8] L. Ziyin, T. Hartwig, And M. Ueda, Neural Networks Fail to Learn Periodic Functions and How to Fix It, *arXiv [cs.LG]*, 2020.
[9] V. Kunc and J. Kléma, Three Decades of Activations: A Comprehensive Survey Of 400 Activation Functions for Neural Networks, *arXiv [cs.LG]*, 2024.
[10] S. R. Dubey, S. K. Singh, And B. B. Chaudhuri, Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark, Neurocomputing, 503, 2022, 92–108.
[11] A. Apicella, F. Donnarumma, F. Isgrò, And R. Prevete, A Survey on Modern Trainable Activation Functions, Neural Netw., 138, 2021, 14–32.
[12] J. Lederer, Activation Functions in Artificial Neural Networks: A Systematic Overview, *arXiv [cs.LG]*, 2021.
[13] C. Nwankpa, W. Ijomah, A. Gachagan, And S. Marshall, Activation Functions: Comparison of Trends in Practice and Research for Deep Learning, *arXiv [cs.LG]*, 2018.
[14] M. Gustineli, A Survey on Recently Proposed Activation Functions for Deep Learning, *arXiv [cs.LG]*, 2022.
[15] A. H. Huang, Expanded gating ranges improve activation functions, *arXiv [cs.NE]*, 2024.