

Novelty Assessment Report

Paper: Toward Principled Flexible Scaling for Self-Gated Neural Activation

PDF URL: <https://openreview.net/pdf?id=XGODWn7HeJ>

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-29

Abstract

Neural networks necessitate nonlinearities to achieve universal approximability. Traditional activation functions introduce nonlinearities through rigid feature rectifications. Recent self-gated variants improve traditional methods in fitting flexibility by incorporating learnable content-aware factors and non-local dependencies, enabling dynamic adjustments to activation curves via adaptive translation and scaling. While SOTA approaches achieve notable gains in conventional CNN layers, they struggle to enhance Transformer layers, where fine-grained context is inherently modeled, severely reducing the effectiveness of non-local dependencies leveraged in activation processes. We refer to this critical yet unexplored challenge as the **non-local tension** of activation. Drawing on a decision-making perspective, we systematically analyze the origins of the non-local tension problem and explore the initial solution to foster a more discriminative and generalizable neural activation methodology. This is achieved by rethinking how non-local cues are encoded and transformed into adaptive scaling coefficients, which in turn recalibrate the contributions of features to filter updates through neural activation. Grounded in these insights, we present **FleS**, a novel self-gated activation model for discriminative pattern recognition. Extensive experiments on various popular benchmarks validate our interpretable methodology for improving neural activation modeling.

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **Self-Gated Neural Activation with Flexible Scaling**

A total of **11 papers** were analyzed and organized into a taxonomy with **11 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Self-Gated Activation Function Design**
- **Architecture-Specific Gating Integration**
- **Adaptive Network Mechanisms**

Complete Taxonomy Tree

- Self-Gated Neural Activation with Flexible Scaling Survey Taxonomy
- Self-Gated Activation Function Design
 - Content-Aware Adaptive Scaling ★ (2 papers)
 - [0] Toward Principled Flexible Scaling for Self-Gated Neural Activation (Anon et al., 2026) [View paper](#)
 - [6] AdaShift: Learning Discriminative Self-Gated Neural Feature Activation With an Adaptive Shift Factor (Sudong Cai, 2024) [View paper](#)
 - Expanded-Range Gating Mechanisms (1 papers)
 - [5] Expanded Gating Ranges Improve Activation Functions (Huang, 2024) [View paper](#)
 - Stochastic Gating with Noise Enhancement (1 papers)
 - [2] Self-gating stochastic-resonance-based autoencoder for unsupervised learning. (Yuhao Ren, 2024) [View paper](#)
- Architecture-Specific Gating Integration
 - Vision-Specific Gated Architectures
 - Large Kernel and Cross-Gate Networks (1 papers)
 - [9] Large Kernel Modulation Network for Efficient Image Super-Resolution (Hu Quanwei, 2025) [View paper](#)
 - Mamba-Based Vision Models (1 papers)
 - [11] Frequency-Assisted Mamba for Remote Sensing Image Super-Resolution (Yi Xiao, 2024) [View paper](#)
 - Domain-Specific Vision Applications (1 papers)
 - [7] SAPS-ViM: Spatial Aggregation Prefix Synergistic Vision Mamba for Wheat Diseases Classification (Siyuan Qin, 2025) [View paper](#)
 - Sequence Modeling with Gating
 - Recurrent Representation Learning (1 papers)
 - [1] Recurrent Neural Networks Learn to Store and Generate Sequences using Non-Linear Representations (Csordás, 2024) [View paper](#)
 - Time-Frequency Gated Networks (1 papers)
 - [8] MSTFNet: A Multi-scale Time-Frequency Domain Network for Current Efficiency Prediction in Aluminum Electrolysis Process: Cen, Guo, Wu, and Chen (L Cen, 2025) [View paper](#)
- Adaptive Network Mechanisms
 - Dynamic Normalization with Rescaling (1 papers)
 - [10] SeeDNorm: Self-Rescaled Dynamic Normalization (Cai, 2025) [View paper](#)
 - Self-Building Skip Connections (1 papers)
 - [3] A step towards neuroplasticity: Capsule networks with self-building skip connections (Nikolai A. K. Steur, 2024) [View paper](#)

- Synaptic Scaling for Continual Learning (1 papers)
- [4] Adaptive Synaptic Scaling in Spiking Networks for Continual Learning and Enhanced Robustness. (Mingkun Xu, 2025) [View paper](#)

Narrative

Core task: self-gated neural activation with flexible scaling. The field explores how neural networks can dynamically modulate their own activations through gating mechanisms that adapt to input content or learned parameters. The taxonomy organizes this landscape into three main branches. Self-Gated Activation Function Design focuses on novel activation functions that incorporate internal gating or scaling logic, often drawing inspiration from biological neurons or information-theoretic principles. Architecture-Specific Gating Integration examines how gating mechanisms are woven into particular network architectures—such as recurrent networks, autoencoders, or vision models—where the gating serves specialized roles like memory control or feature selection. Adaptive Network Mechanisms encompasses broader strategies for dynamic adjustment, including normalization schemes and parameter-free modulation techniques that respond to data statistics or task demands. Representative works illustrate these themes: RNN Nonlinear Representations[1] and Self-Gating Stochastic Autoencoder[2] exemplify architecture-specific integration, while Capsule Skip Connections[3] and Adaptive Synaptic Scaling[4] highlight adaptive mechanisms that extend beyond single activation functions.

Several active lines of work reveal contrasting design philosophies and open questions. One thread emphasizes learnable, content-aware scaling—where gating parameters are derived from the input itself—balancing expressiveness against computational overhead. Another explores parameter-free or biologically inspired modulation, as seen in Expanded Gating Ranges[5] and AdaShift[6], which aim for efficiency and interpretability. Recent efforts like SAPS-ViM[7], MSTFNet[8], and Large Kernel Modulation[9] integrate gating into modern vision architectures, while SeeDNorm[10] and Frequency-Assisted Mamba[11] address normalization and frequency-domain adaptivity. Within this landscape, Principled Flexible Scaling[0] sits in the Content-Aware Adaptive Scaling cluster, closely aligned with works like AdaShift[6] that prioritize input-driven modulation. Compared to AdaShift[6], which focuses on shift-based operations, Principled Flexible Scaling[0] emphasizes a more general framework for scaling, offering flexibility in how gating signals are computed and applied across diverse network contexts.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. AdaShift: Learning Discriminative Self-Gated Neural Feature Activation With an Adaptive Shift Factor

Authors: Sudong Cai | **Year/Venue:** 2024 • Computer Vision and Pattern Recognition | **URL:** [View paper](#)

Abstract

N/A

Relationship Analysis

Both papers belong to the Content-Aware Adaptive Scaling category, focusing on self-gated activations that use content-dependent factors for dynamic curve adjustment. They overlap in addressing adaptive scaling mechanisms for neural activation functions, with both incorporating learnable parameters to modulate activation behavior based on input features. The key difference is that the original paper (FleS) specifically addresses the 'non-local tension' problem in Transformers through coupled vertical and horizontal scaling coefficients derived from effective mean responses, while AdaShift focuses on learning an adaptive shift factor for discriminative feature activation without explicitly targeting the non-local tension challenge.

Contributions Analysis

Overall novelty summary. The paper proposes a self-gated activation mechanism addressing what it terms 'non-local tension'—the challenge that existing self-gated activations struggle to enhance Transformer layers where context is already modeled. It sits in the Content-Aware Adaptive Scaling leaf, which contains only two papers total. This is a sparse research direction within the broader taxonomy of eleven papers across eleven leaf nodes, suggesting the specific focus on content-dependent scaling for activation functions remains relatively unexplored compared to architecture-specific gating or normalization-based adaptation.

The taxonomy reveals neighboring work in Expanded-Range Gating and Stochastic Gating, both exploring alternative scaling strategies but without the content-aware focus. Architecture-Specific Gating Integration branches show how gating mechanisms are applied in vision models and sequence processing, yet these emphasize architectural integration rather than fundamental activation design. The sibling paper in the same leaf likely addresses content-aware scaling but may not tackle the Transformer-specific tension problem. The taxonomy's scope notes clarify that this leaf excludes fixed-range and stochastic methods, positioning the work at the intersection of adaptive scaling and architectural generalization.

Among thirty candidates examined, none clearly refute the three core contributions: formalizing non-local tension, proposing the FleS activation model, and introducing a decision-making-inspired framework. Each contribution was assessed against ten candidates with zero refutable overlaps found. The identification of non-local tension as a distinct problem appears novel within this search scope, as does the specific flexible scaling mechanism. The decision-making perspective for analyzing activation behavior shows no direct precedent among the examined papers, though the limited search scale means broader literature may contain related theoretical frameworks not captured here.

Based on the top-thirty semantic matches and taxonomy structure, the work appears to occupy a relatively underexplored niche—content-aware activation scaling that explicitly addresses Transformer limitations. The sparse population of its taxonomy leaf and absence of refuting candidates suggest novelty, though the analysis cannot confirm whether larger-scale searches or domain-specific venues might reveal closer prior work. The contribution's distinctiveness hinges on the non-local tension framing and its proposed solution rather than the general concept of adaptive activation.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Identification and formalization of non-local tension problem in self-gated activation

Description: The authors identify and formalize a previously unexplored challenge called non-local tension, which occurs when self-gated activation functions fail to effectively leverage non-local cues in Transformer layers. They analyze its origins through a decision-making lens, tracing it to the convergence limitation and trivially discriminative gating weights phenomenon.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. LoFormer: Local Frequency Transformer for Image Deblurring

URL: [View paper](#)

Brief Assessment

LoFormer[18] focuses on image deblurring using local frequency transformers and does not address self-gated activation functions or the non-local tension problem in neural network activations.

2. Multi-behavior hypergraph-enhanced transformer for sequential recommendation

URL: [View paper](#)

Brief Assessment

Multi-behavior Hypergraph Transformer[13] focuses on sequential recommendation with multi-behavior dynamics in e-commerce systems, not on self-gated activation functions or non-local tension in Transformers. The candidate addresses a completely different problem domain (recommendation systems vs. neural activation functions).

3. Long Range Language Modeling via Gated State Spaces

URL: [View paper](#)

Brief Assessment

Gated State Spaces[12] focuses on state space models for language modeling and does not address self-gated activation functions or the non-local tension problem in Transformers. The paper discusses gating mechanisms in the context of state space architectures, not activation function design.

4. Context-Aware Token Selection and Packing for Enhanced Vision Transformer

URL: [View paper](#)

Brief Assessment

Context-Aware Token Selection[15] focuses on token selection and packing mechanisms for vision transformers to improve efficiency, not on analyzing self-gated activation functions or non-local tension phenomena in neural activation processes.

5. Enhancing Skin Cancer Diagnosis Using Swin Transformer with Hybrid Shifted Window-Based Multi-head Self-attention and SwiGLU-Based MLP

URL: [View paper](#)

Brief Assessment

Swin Transformer Skin Cancer[21] focuses on medical image analysis for skin cancer diagnosis using vision transformers, not on analyzing non-local tension in self-gated activation functions or neural activation mechanisms in general deep learning architectures.

6. Translution: A Hybrid Transformer-Convolutional Architecture with Adaptive Gating for Occupancy Detection in Smart Buildings

URL: [View paper](#)

Brief Assessment

Translution[17] focuses on occupancy detection in smart buildings using a hybrid transformer-convolutional architecture with adaptive gating for feature selection. It does not address self-gated activation functions, non-local tension, or convergence limitations in neural activation mechanisms.

7. MossFormer: Pushing the Performance Limit of Monaural Speech Separation Using Gated Single-Head Transformer with Convolution-Augmented Joint Self-Attentions

URL: [View paper](#)

Brief Assessment

MossFormer[14] addresses speech separation using gated transformers with joint self-attentions, not self-gated activation functions or the non-local tension problem in neural activation mechanisms.

8. Branchformer: Parallel MLP-Attention Architectures to Capture Local and Global Context for Speech Recognition and Understanding

URL: [View paper](#)

Brief Assessment

Branchformer[16] addresses parallel architectures for local and global context modeling in speech recognition, not self-gated activation functions or non-local tension in neural activations. The candidate focuses on combining attention mechanisms with MLP modules for speech tasks, which is orthogonal to the original paper's analysis of activation function behavior.

9. VisionTwinNet: Gated Clarity Enhancement Paired With Light-Robust CD Transformers

URL: [View paper](#)

Brief Assessment

VisionTwinNet[19] focuses on change detection in remote sensing images using gated clarity enhancement and transformers for computer vision tasks, not on analyzing self-gated activation functions or non-local tension in neural network activation mechanisms.

10. What Comes After Transformers? A Selective Survey Connecting Ideas in Deep LearningGPT

URL: [View paper](#)

Brief Assessment

After Transformers Survey[20] focuses on architectural alternatives to transformers (state-space models, capsule networks, etc.) and does not discuss self-gated activation functions or the non-local tension problem in neural activations.

Contribution 2: FleS activation model with flexible scaling mechanism

Description: The authors propose FleS, a novel self-gated activation function that addresses non-local tension through adaptive vertical and horizontal scaling coefficients. These coefficients are derived from channel-wise statistical cues (effective mean responses) and enable discriminative recalibration of feature contributions even under convergence limitation.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. EEG-based Auditory Attention Switch Detection with Multi-scale Gated Attention and Multi-task Learning based Hierarchical Spatiotemporal Networks.

URL: [View paper](#)

Brief Assessment

EEG Auditory Attention[36] focuses on EEG-based auditory attention detection using multi-scale gated attention and hierarchical spatiotemporal networks. This is a completely different domain (biomedical signal processing) with different objectives (attention detection) compared to the original paper's self-gated neural activation functions for general pattern recognition.

2. Role of spike-frequency adaptation in shaping neuronal response to dynamic stimuli

URL: [View paper](#)

Brief Assessment

Spike-Frequency Adaptation[38] focuses on neuronal spike-frequency adaptation in biological neurons responding to dynamic stimuli, not on self-gated activation functions with adaptive scaling for neural networks in pattern recognition.

3. Dynamic Fusion of Multi-Scale Perception and Adaptive Discrimination for Compressed GANs

URL: [View paper](#)

Brief Assessment

Dynamic Multi-Scale Fusion[37] focuses on GAN compression through dual-network feature extraction (VGG-19 and ResNet-50) for knowledge distillation, not on self-gated activation functions with adaptive scaling coefficients for neural networks.

4. Self-Gating: An Adaptive Center-of-Mass Approach for Respiratory Gating in PET.

URL: [View paper](#)

Brief Assessment

Self-Gating PET[39] addresses respiratory gating in PET imaging using center-of-mass approaches for medical imaging, which is entirely unrelated to neural network activation functions or discriminative pattern recognition in deep learning.

5. Expanded Gating Ranges Improve Activation Functions

URL: [View paper](#)

Brief Assessment

Expanded Gating Ranges[5] focuses on expanding gating ranges beyond [0,1] for activation functions in transformers, while the original paper addresses non-local tension through adaptive vertical/horizontal scaling coefficients derived from channel-wise statistics in CNNs and transformers.

6. Attention-Based Gated Scaling Adaptive Acoustic Model For Ctc-Based Speech Recognition

URL: [View paper](#)

Brief Assessment

The candidate paper (Gated Scaling Acoustic[33]) focuses on acoustic modeling for CTC-based speech recognition, not on self-gated activation functions for general discriminative pattern recognition. No full text context was provided to assess technical overlap.

7. Learning Discriminative Neural Representations for Visual Recognition

URL: [View paper](#)

Brief Assessment

Discriminative Neural Representations[35] focuses on general visual recognition with self-gated activation, but the extremely limited candidate text provides insufficient detail to assess whether it addresses non-local tension, convergence limitation, or the specific dual-scaling (vertical/horizontal) mechanism that defines FleS.

8. Adaptive recurrent vision performs zero-shot computation scaling to unseen difficulty levels

URL: [View paper](#)

Brief Assessment

Adaptive Recurrent Vision[32] focuses on adaptive computation scaling in recurrent vision models for zero-shot generalization to harder problem difficulties, not on self-gated activation functions with adaptive scaling coefficients for discriminative pattern recognition.

9. AdaShift: Learning Discriminative Self-Gated Neural Feature Activation With an Adaptive Shift Factor

URL: [View paper](#)

Brief Assessment

AdaShift[6] focuses on learning discriminative self-gated neural feature activation with an adaptive shift factor, which is a different mechanism from FleS's flexible vertical and horizontal scaling coefficients derived from channel-wise statistical cues. The candidate paper's full text context is not available (marked as 'n/a'), making detailed comparison impossible.

10. Radar Signal Modulation Recognition Using Self-Enhanced Multidimensional Taylor Network

URL: [View paper](#)

Brief Assessment

Radar Multidimensional Taylor[34] focuses on radar signal modulation recognition using polynomial expansion and feature importance weighting, not on self-gated neural activation functions with adaptive scaling for general pattern recognition tasks.

Contribution 3: Decision-making-inspired theoretical framework for activation analysis

Description: The authors develop a theoretical framework that interprets neural activation through multi-criteria decision-making principles, treating filters as ideal alternatives and features as realistic alternatives. This perspective enables them to identify convergence limitation as the root cause of non-local tension and motivates their flexible scaling solution.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Review and comparison of commonly used activation functions for deep neural networks

URL: [View paper](#)

Brief Assessment

Activation Functions Review[22] is a survey paper comparing existing activation functions (sigmoid, ReLU, tanh, etc.) based on empirical performance metrics. It does not present any decision-making framework for analyzing neural network activations or theoretical interpretations of activation mechanisms.

2. Mathematical analysis and performance evaluation of the gelu activation function in deep learning

URL: [View paper](#)

Brief Assessment

GELU Mathematical Analysis[25] focuses on mathematical properties (differentiability, boundedness, stationarity, smoothness) and empirical performance comparisons of the GELU activation function. It does not present a decision-making framework for analyzing activation functions or interpret neural activation through multi-criteria decision-making principles.

3. On the impact of the activation function on deep neural networks training

URL: [View paper](#)

Brief Assessment

Activation Function Impact[26] analyzes activation functions through dynamical systems and Gaussian process theory, focusing on variance/correlation propagation and the 'edge of chaos' regime. This differs fundamentally from the original paper's multi-criteria decision-making interpretation that treats filters as ideal alternatives and features as realistic alternatives.

4. On the selection of initialization and activation function for deep neural networks

URL: [View paper](#)

Brief Assessment

Initialization Activation Selection[30] focuses on Gaussian process approximations and variance/correlation propagation in neural networks, not on decision-making frameworks for analyzing activation functions. The theoretical approaches are fundamentally different.

5. Effective activation functions for homomorphic evaluation of deep neural networks

URL: [View paper](#)

Brief Assessment

Homomorphic Activation Functions[31] focuses on polynomial approximation methods for homomorphic encryption in CNNs, not on decision-making frameworks for analyzing activation functions. The candidate addresses computational constraints in encrypted neural networks rather than theoretical interpretation through multi-criteria decision-making principles.

6. A comparative performance analysis of different activation functions in LSTM networks for classification

URL: [View paper](#)

Brief Assessment

LSTM Activation Comparison[27] focuses on empirical performance evaluation of 23 activation functions in LSTM networks for classification tasks, without proposing any theoretical framework or decision-making perspective for analyzing neural activations.

7. Optimizing Convolutional Neural Network Architectures with Optimal Activation Functions for Pediatric Pneumonia Diagnosis Using Chest X-Rays

URL: [View paper](#)

Brief Assessment

Optimal Activation Pneumonia[28] focuses on empirical evaluation of activation functions (Mish, Swish, ReLU) for pediatric pneumonia classification using CNNs. It does not present any decision-making theoretical framework or multi-criteria decision-making principles for analyzing neural activation.

8. The role of neural network activation functions

URL: [View paper](#)

Brief Assessment

Neural Activation Functions Role[23] analyzes activation functions through spline theory and infinite-dimensional optimization, not through multi-criteria decision-making principles. The theoretical frameworks are fundamentally different.

9. Kolmogorov-arnold graph neural networks

URL: [View paper](#)

Brief Assessment

Kolmogorov-Arnold GNN[29] focuses on integrating spline-based activation functions into graph neural networks for improved generalization and interpretability, not on decision-making frameworks for analyzing neural activation functions in general networks.

10. Review of adaptive activation function in deep neural network

URL: [View paper](#)

Brief Assessment

Adaptive Activation Review[24] focuses on empirical comparison of activation function categories (saturated, unsaturated, adaptive) without proposing any decision-making framework for theoretical analysis. The paper is a review of existing adaptive activation functions, not a theoretical contribution.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Toward Principled Flexible Scaling for Self-Gated Neural Activation [View paper](#)
- [1] Recurrent Neural Networks Learn to Store and Generate Sequences using Non-Linear Representations [View paper](#)
- [2] Self-gating stochastic-resonance-based autoencoder for unsupervised learning. [View paper](#)
- [3] A step towards neuroplasticity: Capsule networks with self-building skip connections [View paper](#)
- [4] Adaptive Synaptic Scaling in Spiking Networks for Continual Learning and Enhanced Robustness. [View paper](#)
- [5] Expanded Gating Ranges Improve Activation Functions [View paper](#)
- [6] AdaShift: Learning Discriminative Self-Gated Neural Feature Activation With an Adaptive Shift Factor [View paper](#)
- [7] SAPS-ViM: Spatial Aggregation Prefix Synergistic Vision Mamba for Wheat Diseases Classification [View paper](#)
- [8] MSTFNet: A Multi-scale Time-Frequency Domain Network for Current Efficiency Prediction in Aluminum Electrolysis Process: Cen, Guo, Wu, and Chen [View paper](#)
- [9] Large Kernel Modulation Network for Efficient Image Super-Resolution [View paper](#)
- [10] SeeDNorm: Self-Rescaled Dynamic Normalization [View paper](#)
- [11] Frequency-Assisted Mamba for Remote Sensing Image Super-Resolution [View paper](#)
- [12] Long Range Language Modeling via Gated State Spaces [View paper](#)
- [13] Multi-behavior hypergraph-enhanced transformer for sequential recommendation [View paper](#)
- [14] MossFormer: Pushing the Performance Limit of Monaural Speech Separation Using Gated Single-Head Transformer with Convolution-Augmented Joint Self-Attentions [View paper](#)
- [15] Context-Aware Token Selection and Packing for Enhanced Vision Transformer [View paper](#)

- [16] Branchformer: Parallel MLP-Attention Architectures to Capture Local and Global Context for Speech Recognition and Understanding [View paper](#)
- [17] Translusion: A Hybrid Transformer-Convolutional Architecture with Adaptive Gating for Occupancy Detection in Smart Buildings [View paper](#)
- [18] LoFormer: Local Frequency Transformer for Image Deblurring [View paper](#)
- [19] VisionTwinNet: Gated Clarity Enhancement Paired With Light-Robust CD Transformers [View paper](#)
- [20] What Comes After Transformers? A Selective Survey Connecting Ideas in Deep Learning [View paper](#)
- [21] Enhancing Skin Cancer Diagnosis Using Swin Transformer with Hybrid Shifted Window-Based Multi-head Self-attention and SwiGLU-Based MLP [View paper](#)
- [22] Review and comparison of commonly used activation functions for deep neural networks [View paper](#)
- [23] The role of neural network activation functions [View paper](#)
- [24] Review of adaptive activation function in deep neural network [View paper](#)
- [25] Mathematical analysis and performance evaluation of the gelu activation function in deep learning [View paper](#)
- [26] On the impact of the activation function on deep neural networks training [View paper](#)
- [27] A comparative performance analysis of different activation functions in LSTM networks for classification [View paper](#)
- [28] Optimizing Convolutional Neural Network Architectures with Optimal Activation Functions for Pediatric Pneumonia Diagnosis Using Chest X-Rays [View paper](#)
- [29] Kolmogorov-arnold graph neural networks [View paper](#)
- [30] On the selection of initialization and activation function for deep neural networks [View paper](#)
- [31] Effective activation functions for homomorphic evaluation of deep neural networks [View paper](#)
- [32] Adaptive recurrent vision performs zero-shot computation scaling to unseen difficulty levels [View paper](#)
- [33] Attention-Based Gated Scaling Adaptive Acoustic Model For Ctc-Based Speech Recognition [View paper](#)
- [34] Radar Signal Modulation Recognition Using Self-Enhanced Multidimensional Taylor Network [View paper](#)
- [35] Learning Discriminative Neural Representations for Visual Recognition [View paper](#)
- [36] EEG-based Auditory Attention Switch Detection with Multi-scale Gated Attention and Multi-task Learning based Hierarchical Spatiotemporal Networks. [View paper](#)
- [37] Dynamic Fusion of Multi-Scale Perception and Adaptive Discrimination for Compressed GANs [View paper](#)
- [38] Role of spike-frequency adaptation in shaping neuronal response to dynamic stimuli [View paper](#)
- [39] Self-Gating: An Adaptive Center-of-Mass Approach for Respiratory Gating in PET. [View paper](#)