SpikingResformer vs. Transformers in Zero-Shot ImageNet Classification Performance

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does the inference latency of deep residual architectures compare to transformer-based models in zero-shot image classification on ImageNet. The remarkable success of Vision Transformers in Artificial Neural Networks (ANNs) has led to a growing interest in incorporating the self-attention mechanism and transformer-based architecture into Spiking Neural Networks (SNNs). While existing methods propose spiking. 12 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: SpikingResformer: Bridging ResNet and Vision Transformer in Spiking Neural Networks. Research question: How does the inference latency of deep residual architectures compare to transformer-based models in zero-shot image classification on ImageNet?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.7/10.

3 Results

12 papers retrieved. 12 claims extracted; 2 independently verified. Quality review score: 5.7/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
SpikingResformer achieves higher accuracy with fewer parameters and lower energy consumption than other spiking Vision T	✓	0.36
SpikingResformer-L achieves 79.40% top-1 accuracy on ImageNet with 4 time-steps.	✓	0.22
SpikingResformer-L achieves 79.40% accuracy when the input size is enlarged to 288$\times$288.	×	0.10
SpikingResformer-Ti achieves 74.34% accuracy with 11.14M parameters and 2.73G SOPs (2.46mJ).	×	0.03
SpikingResformer-M achieves 77.24% accuracy with 35.52M parameters and 6.07G SOPs (5.46mJ).	×	0.05
SpikingResformer-Ti outperforms Spike-driven Transformer-8-384 by 2.06%.	×	0.04
SpikingResformer-Ti saves 5.67M parameters compared to Spike-driven Transformer-8-384.	×	0.04
SpikingResformer-Ti saves 1.44mJ energy compared to Spike-driven Transformer-8-384.	×	0.05
SpikingResformer-M outperforms Spike-driven Transformer-8-768 by 0.92%.	×	0.04
SpikingResformer-M saves 30.82M parameters compared to Spike-driven Transformer-8-768.	×	0.04
SpikingResformer outperforms other transfer learning methods on CIFAR10 and CIFAR100 datasets with fewer parameters.	×	0.06
SpikingResformer-Ti achieves 84.53% accuracy on static datasets.	×	0.03

References

- http://arxiv.org/abs/2403.14302v2
- http://arxiv.org/abs/2206.10552v2
- http://arxiv.org/abs/2506.20967v2