Batch-Ensemble Mechanisms in BE-SNNs vs. Gradient Masking for Adversarial Robustness

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: How does the batch-ensemble mechanism in BE-SNNs compare to gradient masking techniques in standard SNNs when evaluated on adversarial robustness benchmarks like MNIST or CIFAR-10 using accuracy. 9 claims were extracted from source literature; 4 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: Enhancing Adversarial Robustness in SNNs with Sparse Gradients. Research question: How does the batch-ensemble mechanism in BE-SNNs compare to gradient masking techniques in standard SNNs when evaluated on adversarial robustness benchmarks like MNIST or CIFAR-10 using accuracy degradation metrics?.

2 Methodology

Systematic literature search across multiple databases yielded 9 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

9 papers retrieved. 9 claims extracted; 4 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
Adversarial examples generated under l_infinity attacks tend to be more destructive compared to those generated under l_	×	0.05
SNNs exhibit greater resilience to random perturbations compared to adversarial perturbations, even at larger scales.	✓	0.31
The performance gap between SNNs under adversarial and random perturbations is upper bounded by the gradient sparsity of	✓	0.37
The proposed approach was validated through experiments on both image-based and event-based datasets.	✓	0.16
The results demonstrate notable improvements in the robustness of SNNs using the proposed gradient sparsity regularizati	✓	0.25
The study conducted a small-scale experiment with a primary focus on l_infinity attacks to analyze the disparity in vuln	×	0.06
The experiment evaluated the performance of a well-trained SNN with the VGG-11 architecture.	×	0.05
Random vulnerability is defined as the expected value of the squared difference between the true label probability of a	×	0.05
Adversarial vulnerability is defined as the supremum of the squared difference between the true label probability of a p	×	0.05

References

- http://arxiv.org/abs/2110.11417v1
- http://arxiv.org/abs/2405.20355v1

- http://arxiv.org/abs/2404.17092v2