Robust Accuracy and Latency Trade-offs in ViT and MLP-Mixer Under Adaptive Attacks

Assignee Research

June 3, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How do inference latency and robust accuracy trade-offs differ between ViT and MLP-Mixer architectures under adaptive white-box attacks. 10 claims were extracted from source literature; 9 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 8.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: MIMIR: Masked Image Modeling for Mutual Information-based Adversarial Robustness. Research question: How do inference latency and robust accuracy trade-offs differ between ViT and MLP-Mixer architectures under adaptive white-box attacks?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.5/10.

3 Results

13 papers retrieved. 10 claims extracted; 9 independently verified. Quality review score: 8.5/10.

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
Vision Transformers (ViTs) have emerged as a fundamental architecture and serve as the backbone of modern vision-languag	✓	0.27
ViTs exhibit notable vulnerability to evasion attacks.	✓	0.21
Existing AT methods such as Generalist (CVPR 2023) and DBAT (USENIX Security 2024) have significant incompatibilities wi	✓	0.23
The paper presents a novel theoretical Mutual Information (MI) analysis in its autoencoder-based self-supervised pre-tra	✓	0.35
MI between the adversarial example and its latent representation in ViT-based autoencoders should be constrained via der	✓	0.30
MIMIR is a self-supervised AT method that employs an MI penalty to facilitate adversarial pre-training by masked image m	✓	0.37
Extensive experiments on CIFAR-10, Tiny-ImageNet, and ImageNet-1K show that MIMIR can consistently provide improved natu	✓	0.32
MIMIR outperforms SOTA AT results on ImageNet-1K.	✓	0.24
MIMIR demonstrates superior robustness against unforeseen attacks and common corruption data.	✓	0.26
MIMIR can withstand adaptive attacks where the adversary is aware of the defense mechanism.	×	0.11

References

- https://doi.org/10.1016/j.imed.2022.07.002

- https://doi.org/10.1109/access.2021.3140175
- https://doi.org/10.14722/ndss.2026.241813