

# Growth Bound Matrix and Contrastive Adversarial Training for GLUE Robustness

Assignee Research

June 9, 2026

## Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: How does the Growth Bound Matrix approach compare to contrastive adversarial training in terms of robustness against semantic-preserving adversarial attacks on the GLUE benchmark. 4 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Unrestricted Adversarial Attacks on ImageNet Competition. Research question: How does the Growth Bound Matrix approach compare to contrastive adversarial training in terms of robustness against semantic-preserving adversarial attacks on the GLUE benchmark?.

## 2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.8/10.

## 3 Results

4 papers retrieved. 4 claims extracted; 1 independently verified. Quality review score: 4.8/10.

## 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
The competition is held on the TianChi platform as one of the series of AI Security Challengers Program.	✓	0.27
The final total subjective evaluation score is decided by two parts, i.e., attack success rate and image quality.	×	0.02
The competition aims to find the smallest perturbation $\delta^*$ such that $x + \delta^*$ is misclassified by the target model F under	×	0.06
Direct optimization of problem (8) is intractable, in part due to the lack of information about the target model F.	×	0.02

## References

- <http://arxiv.org/abs/1909.08072v2>
- <http://arxiv.org/abs/2208.00428v1>
- <http://arxiv.org/abs/2110.09903v2>