Subword Sampling with Gaussian Noise vs. Adversarial Augmentations in CodeT5 MBPP Performance

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: How does combining subword sampling with Gaussian noise compare to other adversarial augmentation techniques in terms of CodeT5's performance on the MBPP benchmark for code generation tasks. 13 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.1/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Robustness Analysis of Video-Language Models Against Visual and Language Perturbations. Research question: How does combining subword sampling with Gaussian noise compare to other adversarial augmentation techniques in terms of CodeT5's performance on the MBPP benchmark for code generation tasks?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.1/10.

3 Results

16 papers retrieved. 13 claims extracted; 0 independently verified. Quality review score: 3.1/10.

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
Experiments were performed on the YouCook2-P and MSRVTT-P benchmarks.	×	0.04
For the MSRVTT-P benchmark, zero-shot models typically demonstrate higher absolute and relative robustness compared to o	×	0.05
For long, complex activities in the YouCook2-P benchmark, fine-tuned models are typically more relatively robust.	×	0.05
The FIT model has 180.9M parameters.	×	0.02
The FIT model uses a ViT video encoder.	×	0.03
The COOT model has 7.6M parameters.	×	0.02
On the MSRVTT-P benchmark, the FIT (zs) model achieved a video absolute robustness score (γa) of 0.96$\pm$0.06.	×	0.06
On the MSRVTT-P benchmark, the FIT (zs) model achieved a text absolute robustness score (γa) of 0.89$\pm$0.08.	×	0.06
On the YouCook2-P benchmark, the VideoClip (ft) model achieved a video absolute robustness score (γa) of 0.94$\pm$0.05.	×	0.06
In Table 3, the FIT (zs) model achieved a relative robustness score of 1.00$\pm$0.00 for the AddText perturbation category o	×	0.02
In Table 3, the COOT (scratch) model has no reported data (—) for the Bias perturbation category on YouCook2.	×	0.03
The UniVL model uses BERT as its text encoder and S3D as its video encoder.	×	0.05
The VideoClip model was pre-trained on the HowTo100M dataset.	×	0.06

References

- http://arxiv.org/abs/2207.02159v4
- http://arxiv.org/abs/2306.08568v2
- http://arxiv.org/abs/2208.10224v4