

Multimodal Intermediate-Task Training for Zero-Shot Cross-Lingual Performance on XTREME-R

Assignee Research

July 26, 2026

Abstract

Intermediate-task training—fine-tuning a pretrained model on an intermediate task before fine-tuning again on the target task—often improves model performance substantially on language understanding tasks in monolingual English settings. We investigate whether English intermediate-task training is still helpful on non-English target tasks. Using nine intermediate language-understanding tasks, we evaluate intermediate-task transfer in a zero-shot cross-lingual setting on the XTREME benchmark. We see large improvements from intermediate training on the BUCC and Tatoeba sentence retrieval tas

1 Introduction

This paper examines: Throughput Efficiency Trade-off in Multimodal vs. Text-Only Models for Zero-Shot Cross-Lingual Transfer. Research question: Does multimodal intermediate-task training (e.g., combining text and image tasks) improve zero-shot cross-lingual performance on the XTREME-R benchmark compared to text-only intermediate-task training, and how much does it affect throughput during inference?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.0/10.

3 Results

11 papers retrieved. 3 claims extracted; 3 independently verified. Quality review score: 8.0/10.

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
Intermediate-task training often improves model performance substantially on language understanding tasks in monolingual	✓	0.40
Using nine intermediate language-understanding tasks, improvements from intermediate training were observed on the BUCC	✓	0.30
The research goal is to determine the throughput efficiency trade-off (in tokens/second) of using intermediate-task trai	✓	0.59

References

- <https://doi.org/10.1613/jair.1.13083>
- <https://doi.org/10.5281/zenodo.21321991>
- <https://doi.org/10.5281/zenodo.21321990>