

# What are the benchmark performance scores of SALMONN-7B on reasoning mathematics coding and language understanding

Assignee Research

June 10, 2026

## Abstract

Vision language models (VLMs) achieve unified modeling of images and text, enabling them to accomplish complex real-world tasks through perception, planning, and reasoning. Among these tasks, reasoning is particularly representative, with mathematical reasoning serving as a prominent example. It highlights the high-level capability of VLMs to comprehend mathematical information in images and to perform sophisticated reasoning. Recently, numerous visual mathematical reasoning benchmarks have been proposed, but they are often restricted to geometry, lack coverage of math word problems, and rarel

## 1 Introduction

This paper examines: GSM8K-V: Can Vision Language Models Solve Grade School Math Word Problems in Visual Contexts. Research question: What are the benchmark performance scores of SALMONN-7B on reasoning mathematics coding and language understanding 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 7.2/10.

## 3 Results

16 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 7.2/10.

## 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.

## References

- <http://arxiv.org/abs/2312.17080v4>
- <http://arxiv.org/abs/2509.25160v1>
- <http://arxiv.org/abs/2410.12381v3>