Architectural Innovations Enhancing Transformer Performance in Multi-Step Logical Reasoning

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: What architectural innovations improve transformer performance on multi-step logical reasoning v11. 15 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Vision Transformer with Quadrangle Attention. Research question: What architectural innovations improve transformer performance on multi-step logical reasoning v11.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.7/10.

3 Results

12 papers retrieved. 15 claims extracted; 0 independently verified. Quality review score: 3.7/10.

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
ImageNet-1k contains 1 million images for training and 50,000 for testing, with 1,000 and 50 images per category, respec	×	0.01
ADE20k comprises 27k images with 150 labeled classes, including 25k images for training and 2k for validation.	×	0.01
MS COCO includes 200k images from 80 object classes, with 150k human instances each annotated with 17 different keypoint	×	0.02
For classification, Top-1 and Top-5 accuracy are reported as evaluation metrics.	×	0.04
For object detection, instance segmentation, and human pose estimation tasks, mean average precision (mAP) over all clas	×	0.06
For pose estimation, object keypoint similarity (OKS) is used to determine the threshold.	×	0.04
For semantic segmentation tasks, intersection over union (IoU) is used as the evaluation metric.	×	0.04
The input image resolution is set to 224 $\times$ 224 for image classification.	×	0.03
The Top-1 accuracy for the model with scale and shift transformations is 81.2.	×	0.02
The Top-1 accuracy for the model with scale, shift, shear, and rotation transformations is 82.9.	×	0.01
The Top-1 accuracy for the model with 0 attention number is 81.2.	×	0.04
The Top-1 accuracy for the model with 6 attention number is 83.3.	×	0.04
The epoch time for the Window model is 6:20.	×	0.04
The epoch time for the Quadrangle model is 8:19.	×	0.05
The Top-1 accuracy for the model with λ=0.1 is 82.9.	×	0.03

References

- http://arxiv.org/abs/2506.22084v1
- http://arxiv.org/abs/2407.04973v1
- http://arxiv.org/abs/2303.15105v1