Scaling Laws of Patchout Audio Transformers on Edge Devices: Latency-Throughput Tradeoffs

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 7 peer-reviewed papers addressing the following research question: How does the inference efficiency of Patchout Audio Transformers scale with model size when deployed on edge devices, as measured by latency-throughput tradeoffs on the AVE-AV benchmark. The deployment of transformer-based models on resource-constrained edge devices represents a critical challenge in enabling real-time artificial intelligence applications. This comprehensive survey examines lightweight transformer architectures specifically designed for edge. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Lightweight Transformer Architectures for Edge Devices in Real-Time Applications. Research question: How does the inference efficiency of Patchout Audio Transformers scale with model size when deployed on edge devices, as measured by latency-throughput tradeoffs on the AVE-AV benchmark?.

2 Methodology

Systematic literature search across multiple databases yielded 7 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.2/10.

3 Results

7 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 2.2/10.

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/2311.02772v2
- http://arxiv.org/abs/2309.10787v2
- http://arxiv.org/abs/2601.03290v1