Scaling Contrastive Learning Models and Inference Efficiency in Time-Series Foundation Models

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: What is the impact of scaling contrastive learning model size on the inference efficiency of time-series foundation models, evaluated using the Latency-Throughput score on the UEA multivariate. 13 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Scaling-laws for Large Time-series Models. Research question: What is the impact of scaling contrastive learning model size on the inference efficiency of time-series foundation models, evaluated using the Latency-Throughput score on the UEA multivariate time-series dataset?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.3/10.

3 Results

15 papers retrieved. 13 claims extracted; 0 independently verified. Quality review score: 3.3/10.

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
The study fits a power law of the form ln L(A) = -B0 ln A + B0 ln A0 to scaling relations, where L is the objective func	×	0.08
Fitted parameter values for the scaling laws are provided in Table 7 in Appendix D.	×	0.11
In cases of broken power-law scaling, the reported fit applies only to the regime after the break.	×	0.05
Parameter scaling shows approximately power-law behavior in minimum in-sequence test loss (MSE, CRPS, and log-likelihood	×	0.08
A mild break in power-law behavior is observed in both MSE and CRPS test losses as a function of parameter count.	×	0.06
Little or no break is observed in log-likelihood scaling as a function of parameter count.	×	0.05
A constant factor of two was added to the log-likelihood values to ensure they are always positive for power-law scaling	×	0.05
To extract reliable data scaling behavior, the data diversity was kept fixed by maintaining each dataset's relative cont	×	0.06
For time-series significantly longer than the context length, a randomly chosen portion (fd) of each series was used dur	×	0.05
For time-series that would become shorter than the context length when cut, the entire series was randomly dropped with	×	0.04
Data scaling results were generated using a model with approximately 20 million parameters.	×	0.04
Power-law scaling was observed across four orders of magnitude in dataset size for all three performance measures (MSE,	×	0.13
Compute (C) at any given training stage is calculated using the formula C = 6BNpLseq, where B is batch size, Np is param	×	0.05

References

- http://arxiv.org/abs/2512.03307v1
- http://arxiv.org/abs/2405.13867v2
- http://arxiv.org/abs/2406.06518v1