Scaling Laws of Model Size and Performance on the HLCE Benchmark

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: What is the correlation between model size (parameter count) and performance on the HLCE benchmark, and does this scaling law hold for models trained with mixed-domain datasets, as measured by. 10 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Testing the Efficient Network TRaining (ENTR) Hypothesis: initially reducing training image size makes Convolutional Neural Network training for image recognition tasks more efficient. Research question: What is the correlation between model size (parameter count) and performance on the HLCE benchmark, and does this scaling law hold for models trained with mixed-domain datasets, as measured by accuracy and computational efficiency?.

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 3.5/10.

3 Results

16 papers retrieved. 10 claims extracted; 0 independently verified. Quality review score: 3.5/10.

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 standardized accuracy response to training regimes, where image size was increased was consistently higher across da	×	0.10
For the most efficient training on the Steel dataset, a stepwise image size increase yielded the best result with the Re	×	0.10
In the UCM dataset, a gradual image size increase with the ResNet18 architecture was most efficient.	×	0.08
The most efficient result for the Bee dataset was achieved with a gradual image size increase in the smallest ResNet18 n	×	0.09
The accuracy response to training regimes was different across datasets and – except from the UCM dataset also within ar	×	0.06
The ResNet50 network trained on the Steel dataset yielded the highest accuracy with no image size modification regardles	×	0.08
In the UCM dataset, the highest accuracies could be achieved with the ResNet18 network, which was consistent for pre (91	×	0.05
In the Bees dataset, a stepwise image size increase yielded the highest accuracy pre and post pre-training layers of res	×	0.08
The Steel dataset contains 1460 images in 7 classes.	×	0.04
The UCM dataset contains 2100 images in 21 classes.	×	0.03

References

- http://arxiv.org/abs/2402.04177v3

- http://arxiv.org/abs/2408.11029v2
- http://arxiv.org/abs/1807.11583v1