Contrastive Learning and Masked Autoencoders in Tabular Data Benchmark Performance

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does integrating contrastive learning with masked autoencoder objectives affect downstream task performance on tabular benchmarks like TabTime compared to using either objective alone. 16 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Understanding Masked Autoencoders From a Local Contrastive Perspective. Research question: How does integrating contrastive learning with masked autoencoder objectives affect downstream task performance on tabular benchmarks like TabTime compared to using either objective alone?.

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 4.2/10.

3 Results

12 papers retrieved. 16 claims extracted; 1 independently verified. Quality review score: 4.2/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
LC-MAE preserves MAE's state-of-the-art performance.	×	0.13
LC-MAE uses three explicit token-based loss functions: a reconstruction loss, a cross-view contrastive loss, and an in-v	×	0.07
The cross-view loss ensures that, for a given image, token features in the same position but different random masks are	×	0.05
The in-view loss ensures the consistency of the distribution of the output features and the input image.	×	0.05
The decoder primarily utilizes positional information in the shallow layers, and gradually learns semantic information a	×	0.02
Random masking strategy serves two purposes: (1) data augmentation; (2) restricting the effective receptive field of Vis	✓	0.17
Restricting the receptive field is enough to improve downstream finetuning performance.	×	0.07
LC-MAE achieves a finetuning accuracy of 83.0% on ImageNet.	×	0.06
The weighted average decoder version achieves a finetuning accuracy of 82.9% on ImageNet.	×	0.02
The original Transformer decoder version of MAE achieves a finetuning accuracy of 82.2% on ImageNet.	×	0.05
The cross-view contrastive loss alone achieves a finetuning accuracy of 82.8% on ImageNet.	×	0.04
The in-view contrastive loss alone achieves a finetuning accuracy of 82.5% on ImageNet.	×	0.04
The combination of LMAE and Lin achieves a finetuning accuracy of 83.1% on ImageNet.	×	0.02
LC-MAE is pretrained on ImageNet with a mask ratio of 75% for 100 epochs.	×	0.04
LC-MAE is supervised finetuned on ImageNet for an additional 100 epochs.	×	0.06
The depth of the decoder in LC-MAE is set to 2 to shorten the pretraining duration.	×	0.06

References

- http://arxiv.org/abs/2407.05862v1
- http://arxiv.org/abs/2310.01994v2
- http://arxiv.org/abs/2402.01204v4