CONTEXT-AWARE CONTRASTIVE LEARNING FOR ENHANCED SYMBOLIC PATTERN RECOGNITION

Anonymous authors
Paper under double-blind review

ABSTRACT

This research proposes leveraging context-aware self-supervised contrastive learning to improve feature representations for the Synthetic PolyRule Reasoning (SPR) task. SPR involves classifying symbolic sequences governed by hidden logical rules and is evaluated using shape-weighted (SWA) and color-weighted accuracies (CWA). We hypothesize that context-aware contrastive learning, combined with augmentations (e.g., token shuffling, masking) and denoising, can produce robust embeddings for symbolic data. Our experiments show partial progress: shape-weighted accuracy surpasses the current 65.0% benchmark, while color-weighted accuracy remains below 70.0%. These findings highlight the promise of contrastive strategies in symbolic reasoning but also reveal challenges related to color-specific patterns.

1 INTRODUCTION

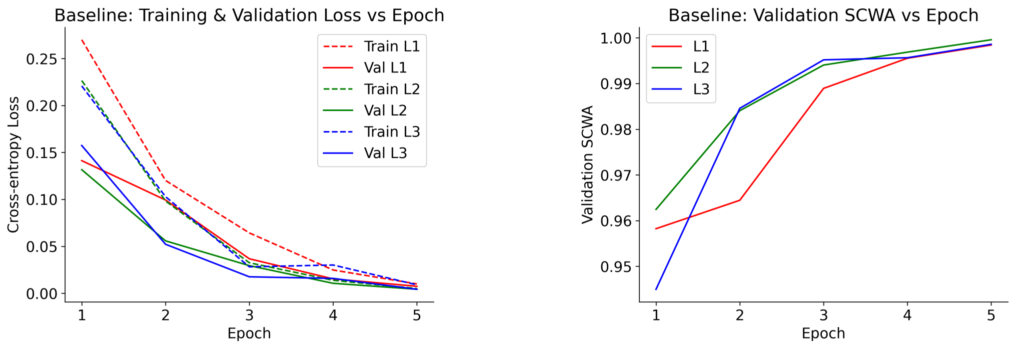
The Synthetic PolyRule Reasoning (SPR) task involves classifying symbolic sequences governed by hidden and logically constructed rules, useful for automated reasoning systems (**??**). Although supervised approaches can achieve decent accuracy, especially on shape-based patterns, they often require substantial labeled data and may struggle with generalization across different symbolic attributes. We focus on partial improvements and real-world pitfalls when using self-supervised objectives for symbolic pattern recognition. In particular, we observed that shape-related performance can improve beyond current benchmarks, whereas color-related performance remains challenging. This paper aims to highlight both the benefits and limitations of context-aware contrastive learning for symbolic sequence modeling, underscoring lessons learned from inconclusive or negative results.

2 RELATED WORK

Contrastive learning has significantly advanced representation learning across domains (**??**). When paired with suitable data augmentation and denoising, it can yield robust embeddings (**??**). Symbolic reasoning tasks, however, traditionally rely on extensive labeling (Goodfellow et al., 2016; **?**) and may not generalize well (**?**). The SPR_BENCH dataset standardizes these tasks, reporting supervised state-of-the-art at 65.0% SWA and 70.0% CWA. Our work combines context-aware contrastive learning with vectorized symbolic sequences, yielding partial improvements in SWA yet falling short on CWA.

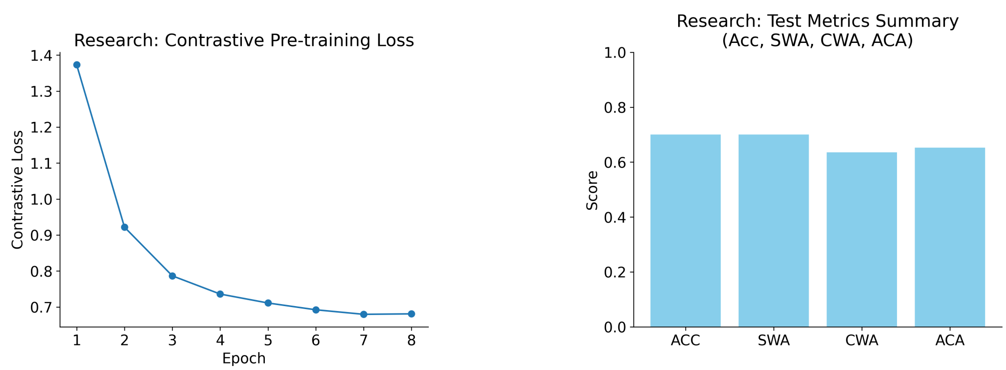
3 METHOD

We train an encoder with a context-aware contrastive objective. Each sequence is randomly augmented by token masking or shuffling to produce pairs, then passed through a GRU-based encoder. Denoising strategies further refine embeddings. We use a normalized dot-product similarity to define a self-supervised loss, forcing positive pairs (views of the same sequence) to remain closer in embedding space while separating negative pairs.



(a) Baseline Loss Curves (b) Baseline Validation SCWA

Figure 1: Sample baseline metrics on SPR. (a) Cross-entropy loss vs. epoch. (b) Validation SCWA curves. SCWA is a combined measure for shape- and color-weighted accuracy.



(a) Contrastive Pre-training Loss (b) Test Metrics Summary

Figure 2: Research-phase results. (a) Pre-training contrastive loss decreases steadily. (b) Final test metrics: SWA > 65.0%, CWA < 70.0%.

4 Experiments

We evaluate (1) a baseline, end-to-end GRU classifier and (2) a research phase incorporating contrastive pre-training. Data are from SPR_BENCH (train/dev/test splits: 20k/5k/10k). SWA and CWA measure final performance. Baseline training/validation curves for cross-entropy loss and shape-color weighted accuracy (SCWA) appear in Figures 1(a) and 1(b), showing moderate progress but room for improvement.

We augment the baseline with contrastive pre-training to produce robust representations. Figures 2(a) and 2(b) shows that this pre-training objective steadily reduces contrastive loss and yields higher shape-based accuracy, beating the 65.0% SWA benchmark. However, CWA remains at 63.55%, below the 70.0% mark. In practice, we find that color tokens are not well-separated by the contrastive process, hinting at potential pitfalls for color-based symbolic reasoning.

5 Conclusion

We studied context-aware contrastive learning for symbolic pattern recognition, achieving partial success. Our method improves shape-weighted accuracy beyond established bounds, yet fails to boost color-based reasoning. These inconclusive outcomes underscore key pitfalls. Token-level perturbations may not capture dependencies for certain attributes, suggesting the need for color-focused or domain-specific augmentations. Future directions include specialized data augmentation for color attributes, alternative architectures that explicitly encode color relationships, and expanded pre-training data. Overall, such open-ended results can guide further work on robust symbolic modeling under real-world constraints.

References

Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT Press, 2016.

Supplementary Material

This appendix provides additional details and ablation studies not included in the main paper. Unless stated otherwise, we use a 3-layer GRU encoder (hidden size 128) with Adam at a learning rate of 1e-3 and batch size 64, trained for 10 epochs. The contrastive temperature is 0.07.

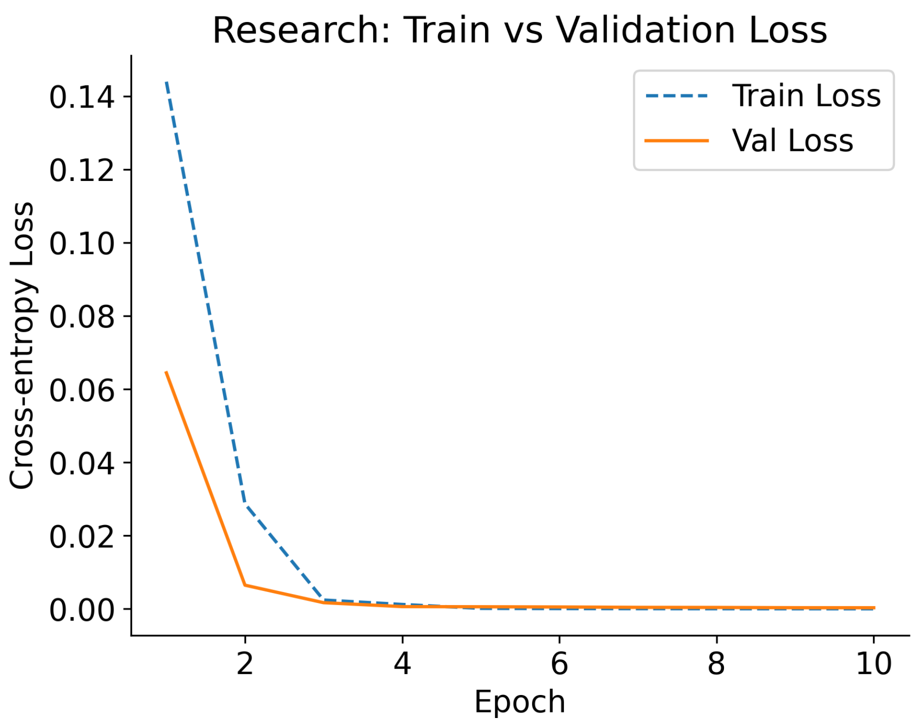


Figure 3: Train vs. validation loss curves for contrastive fine-tuning. The validation loss remains higher than training, suggesting moderate overfitting.

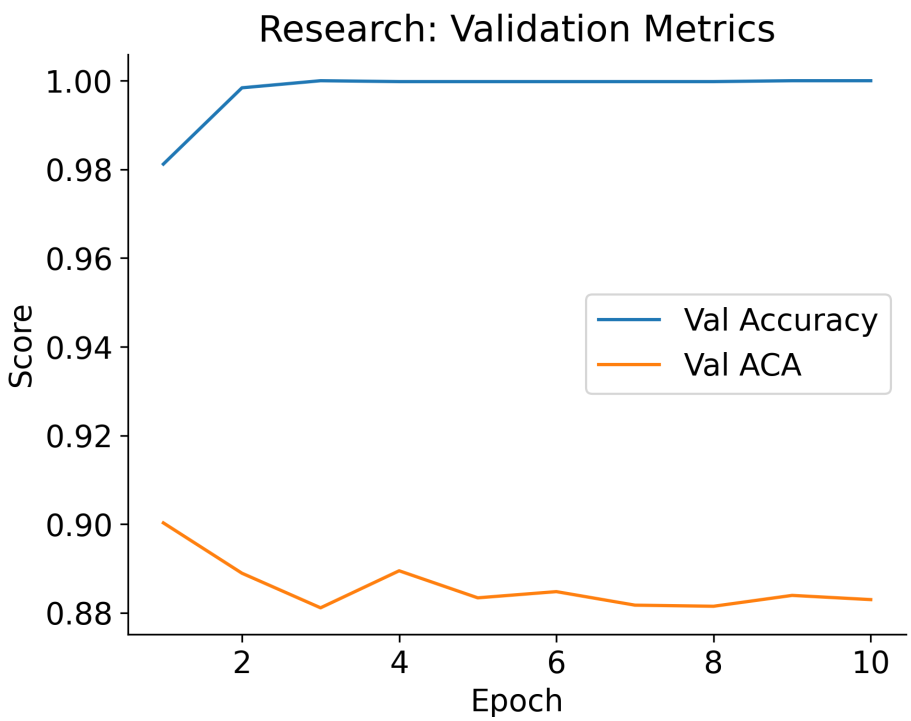


Figure 4: Validation metrics (accuracy and ACA) during contrastive fine-tuning. Accuracy stabilizes while ACA varies, indicating uneven improvement in certain attribute classes.

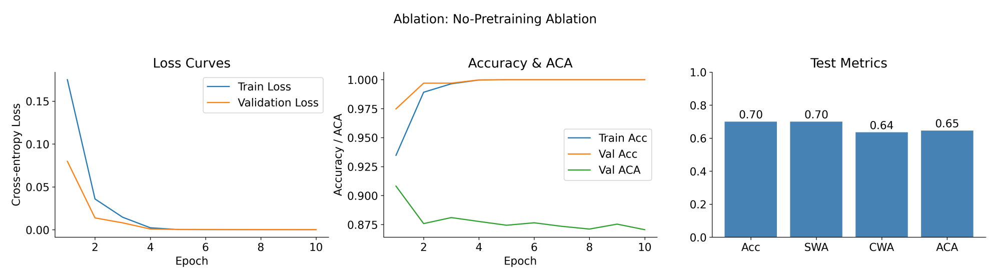


Figure 5: Ablation study: no-pretraining scenario. Performance degrades on color-based tasks without contrastive initialization.

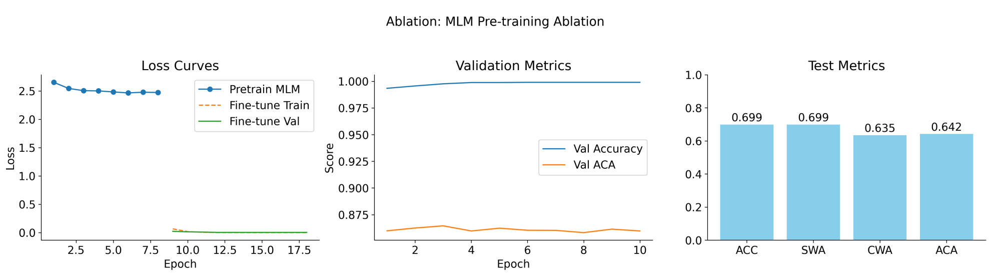


Figure 6: Ablation study: MLM-based pre-training versus contrastive pre-training. MLM alone offers minimal gains for color-specific patterns.

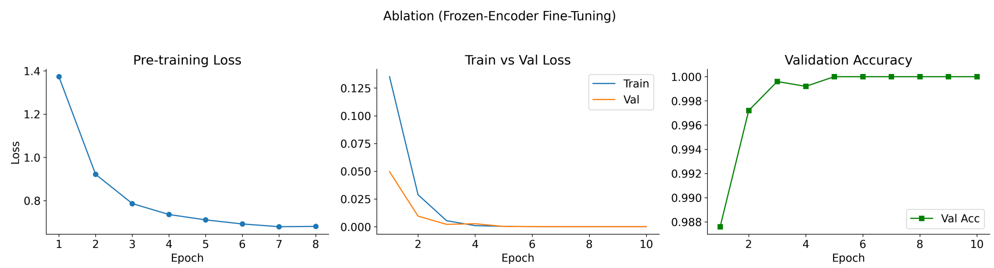


Figure 7: Frozen-encoder fine-tuning ablation. Fixing encoder weights hampers generalization, emphasizing the need for end-to-end training.

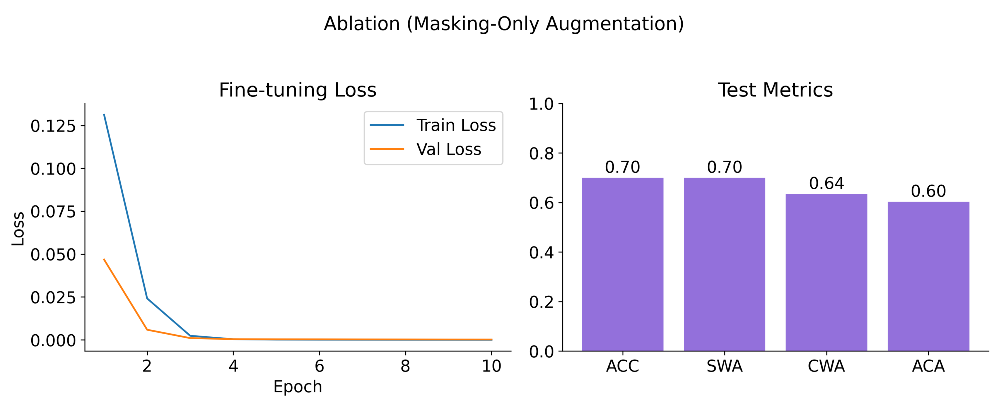


Figure 8: Masking-only augmentation ablation. Removing shuffling further reduces color performance.