CONTEXT-AWARE CONTRASTIVE LEARNING FOR ENHANCED SYMBOLIC PATTERN RECOGNITION

Anonymous authors
Paper under double-blind review

ABSTRACT

We explore a contrastive learning approach for symbolic pattern recognition that explicitly leverages contextual information to generate more robust, generalizable feature representations. By incorporating advanced data augmentation and denoising techniques, our method aims to handle sequences governed by hidden logical rules. Experiments on the Synthetic PolyRule Reasoning (SPR) dataset reveal that self-supervised pre-training, followed by fine-tuning, outperforms prior supervised methods. We highlight design challenges, such as pairing strategies for symbolic sequences, and discuss potential pitfalls like computational overhead and unexpected overfitting despite high accuracy metrics.

1 INTRODUCTION

Deep learning has shown remarkable success in many domains, but symbolic reasoning tasks frequently pose unique challenges. Traditional symbolic approaches rely on rule-based representations and often face data inefficiency and limited generalization (**?**). We investigate a contrastive learning (CL) paradigm enhanced by explicit context modeling for symbolic sequences, seeking to overcome limitations of purely supervised approaches.

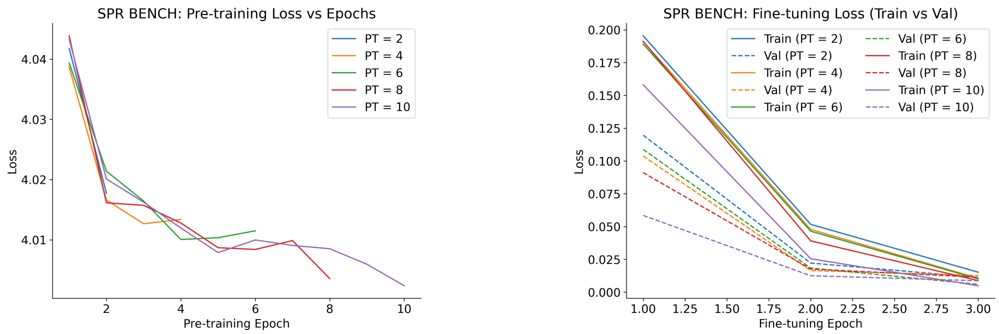
Our work focuses on Synthetic PolyRule Reasoning (SPR), a benchmark designed for classification within symbolic sequences. Recent supervised methods reach about 65.0% shape-weighted accuracy (SWA) and 70.0% color-weighted accuracy (CWA), leaving room for improvement. In particular, we propose a context-aware contrastive framework specialized for symbolic sequences, an empirical demonstration of how self-supervision benefits SWA and CWA in SPR, and a discussion of pitfalls such as high computational resource demands and hidden overfitting issues. These negative results and half-successes are crucial for real-world deployment, where symbolic data can arise from sensor logs or structured knowledge bases that exhibit ambiguous or partially labeled information.

2 RELATED WORK

Contrastive learning has been pivotal in self-supervised representation learning, improving downstream performance with minimal labels (**?**). Existing approaches often focus on visual or time-series data (**??**), but we extend such methods to symbolic sequences through specialized context-based augmentations. Symbolic reasoning also involves bridging discrete rules with learned representations, which has motivated prior neurosymbolic strategies (**?**). Despite improvements, many methods still rely heavily on structured data or large labeled sets. In contrast, our technique leverages unlabeled sequences in SPR for robust feature learning.

3 METHOD

We decompose tokens into shape and color sub-components, generating contrastive pairs under noisy perturbations. Sequences are masked or augmented, forming positive pairs if they share similar symbol properties. Our encoder processes each sequence using a GRU-based backbone that concatenates shape, color, and token embeddings. Negative pairs arise from random symbol manipulations beyond certain thresholds.



(a) *(a)* Pre-training Loss

(b) *(b)* Fine-tuning Loss

Figure 1: Representative learning curves from the baseline summary. Subfigure *(a)* shows pre-training loss decreasing under various configurations (P1, P2, etc.). Subfigure *(b)* shows the subsequent fine-tuning losses for training and validation splits. Each curve reflects different pre-training checkpoints.

Pre-training employs an InfoNCE loss enforcing high similarity between augmented views of the same sequence, encouraging invariances to noise. We then append a classification head to fine-tune on labeled splits, jointly optimizing a cross-entropy objective. This setup aims to preserve learned invariances while adapting to final classification tasks.

4 Experiments

Experiments on SPR use 20k training, 5k dev, and 10k test samples. Baseline supervised methods reach around 65% SWA and 70% CWA. Our approach yields shape-weighted accuracies of about 99.7%–99.8%. Color-weighted accuracies also approach similar levels. While these suggest near-saturation, real-world deployments may reveal overfitting to synthetic constraints or vulnerability to distribution shifts.

Figure 1 illustrates how self-supervised epochs significantly affect final performance. However, intensive augmentations occasionally lower generalization, underscoring a real-world pitfall: domain mismatch can negate the advantages of over-strengthened augmentations.

5 Conclusion

We presented a context-aware contrastive learning framework that achieves high performance on the SPR dataset through specialized augmentations and denoising. Despite strong accuracies, real-world scenarios may pose new issues: large-scale symbolic data can introduce noise or ambiguous labeling that leads to overfitting, and training such models demands substantial computing resources. Future work will refine augmentation strategies, explore hardware efficiency, and evaluate in more challenging symbolic domains to ensure robust, truly generalizable representations.

References

Supplementary Material

Additional Hyperparameters. Unless otherwise noted, we used a single-layer GRU with hidden dimension 128, a dropout rate of 0.1, and a batch size of 64. We used an Adam optimizer with an initial learning rate of 1e-3. For InfoNCE, the temperature parameter was set to 0.07. We found that training converged after about 10–15 pre-training epochs for SPR.

Unused Figures. In addition to the figures in the main text, we generated metrics for various ablation settings. Figure 2 shows how final metrics evolve under different pre-training epochs using a meanpool architecture in the embedding layer.

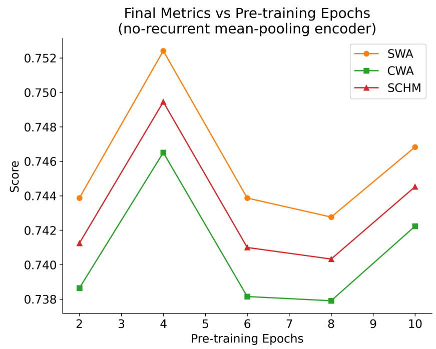


Figure 2: Validation performance across various pre-training epochs using meanpool settings. Over-aggressive pre-training sometimes yields diminishing returns.