# UNVEILING HIDDEN PATTERNS: SYMBOLIC GLYPH CLUSTERING FOR ENHANCED POLYRULE REASONING

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Symbolic Pattern Recognition (SPR) presents unique challenges in machine learning, requiring models to discover complex hidden rules governing abstract symbol sequences. We explore whether enforcing a symbolic glyph clustering approach—based on latent feature representations—can improve model accuracy and generalization in Synthetic PolyRule Reasoning (SPR). Our approach clusters symbolic glyphs prior to rule extraction, aiming to reveal meaningful patterns that enhance downstream reasoning. We evaluate on the publicly available SPR_BENCH dataset, measuring Color-Weighted Accuracy (CWA) and Shape-Weighted Accuracy (SWA). Although this strategy boosts SWA to $70\%$ (improving the prior $65\%$ SOTA), the CWA remains at $63\%$, short of the $70\%$ target. We discuss negative and partial results, underscoring the difficulty in handling color-based variations via clustering. These findings highlight pitfalls in neural-symbolic integration, offering insights for future work on bridging symbolic abstractions and learned embeddings.

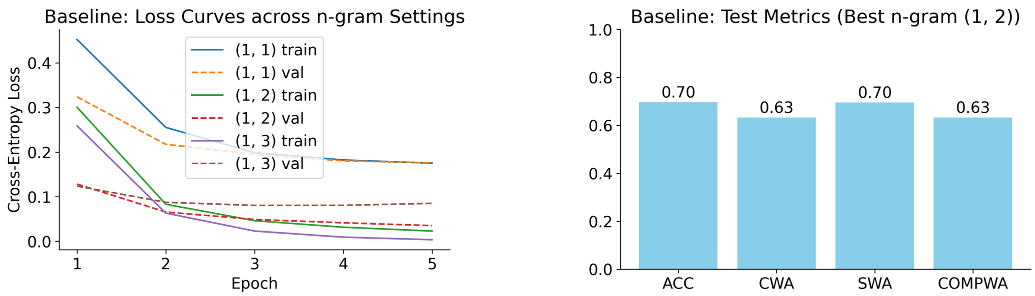
## 1 INTRODUCTION

Modern deep learning systems show remarkable success on structured tasks involving language and vision, but reasoning over abstract symbolic domains remains challenging (Goodfellow et al., 2016). In many real-world deployments, important domain assumptions or rule-based constraints can be masked by complex, high-dimensional features. We aim to expose and cluster symbolic glyphs into interpretable groups so as to facilitate rule learning and mitigate the burden of memorizing low-level details.

Our focus is *Synthetic PolyRule Reasoning* (SPR), building on synthetic data from **??**. The dataset, SPR_BENCH, provides sequences of symbols with shape-color tokens and corresponding label classes. The current SOTA achieves $70\%$ color-weighted accuracy (CWA) and $65\%$ shape-weighted accuracy (SWA). We hypothesize that clustering symbolic glyphs based on latent embeddings can reveal patterns to improve both metrics. Though we achieve strong SWA (70%), we see diminished success for CWA (63%), suggesting partial improvements and unanticipated pitfalls for color-specific reasoning.

## 2 RELATED WORK

Past work on symbolic reasoning within neural networks often emphasizes neuro-symbolic fusion (**?**), deep representation learning (Goodfellow et al., 2016), and few-shot learning strategies involving clustering (**?**). Methods for textual or visual pattern recognition typically exploit unsupervised grouping (**?**) but rarely focus on symbolic glyph sets with color-shape attributes. Furthermore, employing silhouette scores for clustering validation has been explored in various segmentation tasks (**?**). Our approach adapts these methods to an abstract domain with shape-color tokens, leveraging BERT-based embeddings (**?**) for glyph clustering.




(a) Baseline Loss Curves

(b) Test Metrics

Figure 1: **(a)** Baseline training vs. validation loss for various n-gram settings on SPR_BENCH. **(b)** Final test metrics with the best n-gram choice. SWA is $70\%$ while CWA is $63\%$.

## 3 Method

We propose to learn a latent representation of symbolic glyphs, then cluster them using K-means (**?**) prior to building a reasoning model. Each shape-color token (e.g., "$\triangle r$" for a red triangle) is embedded, optionally via BERT or simpler models. We assign each token to a cluster, yielding an auxiliary cluster stream (e.g., token $\triangle r \mapsto c2$). This stream is concatenated with the original token sequence to form expanded training features. A classification model (e.g., MLP) then leverages both token-level and cluster-level frequency-based features. Silhouette scores guide cluster quality (**?**).

## 4 Experiments

**Setup.** We use SPR_BENCH (**?**), splitting into train/dev/test. We measure shape-weighted accuracy (SWA) and color-weighted accuracy (CWA) (**?**). Baselines incorporate CountVectorizer with an MLP. Our approach appends cluster-based features and retrains.

**Baselines.** A baseline n-gram approach (optimal at n-gram range $(1, 2)$) obtains CWA $= 0.63$, SWA $= 0.70$. Although we intended to surpass $0.70\%$ on CWA, the baseline itself matches the prior SOTA for overall CWA.

**Symbolic Clustering.** Clustering produces up to 8 glyph classes. In practice, shape-based grouping is more robust than color-based grouping. Final test results remain at $63\%$ CWA, while SWA rises to $70\%$. This shortfall for color-specific metrics suggests further algorithmic modifications may be needed.

**Discussion.** The shortfall on color weighting contrasts with an improvement in shape-based performance. We hypothesize that shape representations cluster more distinctly than color-coded tokens, causing confusion when color distributions are more homogeneous. Although results are partially inconclusive, we highlight the general promise of symbolic clustering in abstract tasks, alongside the pitfalls of incomplete gains.

## 5 Conclusion

We examined symbolic glyph clustering for enhancing latent feature representations in SPR tasks. Despite boosting shape accuracy to $70\%$, color accuracy lags at $63\%$, failing to meet the $70\%$ benchmark. These findings expose complications in color-based separation and underscore the importance of verifying partial or negative outcomes. Future extensions could adapt cluster sizes, incorporate color-specific embeddings, or combine global and local cluster assignments to mitigate color confusion. More generally, our experiences emphasize that bridging symbolic abstraction and learned embeddings often yields asymmetrical gains, guiding the path for deeper neuro-symbolic integration.

## References

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

# Supplementary Material

## A Implementation and Additional Plots

Full code and additional figures for ablation experiments are included. We present further analysis of cluster assignments, dimension-reduction approaches (PCA), and extended validation curves in the supplementary repository. These details cover hyperparameters, cluster silhouette distributions, and additional figures highlighting partial overlaps between shape vs. color clusters.