Discovering Latent Structure in Abstract Visual Reasoning Tasks Using Unsupervised Learning

IBM Unsupervised Machine Learning – Final Project

1 Main Objective of the Analysis

The primary objective of this analysis is to assess whether **unsupervised learning techniques** can uncover **meaningful latent structure** in abstract visual reasoning tasks without relying on labeled supervision. The analysis focuses on both **dimensionality reduction** and **clustering**, with the goal of identifying shared transformation patterns across tasks.

From a business and stakeholder perspective, this work demonstrates how unsupervised learning can:

- Reduce dependence on costly manual labeling,
- Reveal reusable task structures in complex, non-tabular data,
- Support downstream applications such as task categorization, benchmarking, and dataset diagnostics.

The guiding question of this project is:

> *Can unsupervised models discover latent transformation patterns that align with human-defined reasoning categories?*

2 Dataset Description

This study uses the **ConceptARC** dataset, a structured collection of abstract visual reasoning tasks derived from the Abstraction and Reasoning Corpus (ARC). Each task consists of:

- An input grid,
- A corresponding output grid,
- A conceptual category describing the underlying transformation.

Each task is annotated with one of **16 conceptual categories**, including: *AboveBelow*, *Center*, *CleanUp*, *CompleteShape*, *Copy*, *Count*, *ExtendToBoundary*, *ExtractObjects*, *FilledNotFilled*, *HorizontalVertical*, *InsideOutside*, *MoveToBoundary*, *Order*, *SameDifferent*, *TopBottom2D*, and *TopBottom3D*.

Key dataset characteristics:

- 427 training input–output pairs were used in this analysis,

- Variable-sized grids padded to a fixed resolution of 32×32,
- Discrete color values (0–9) plus background,
- Category labels used *only for evaluation*, not for model training.

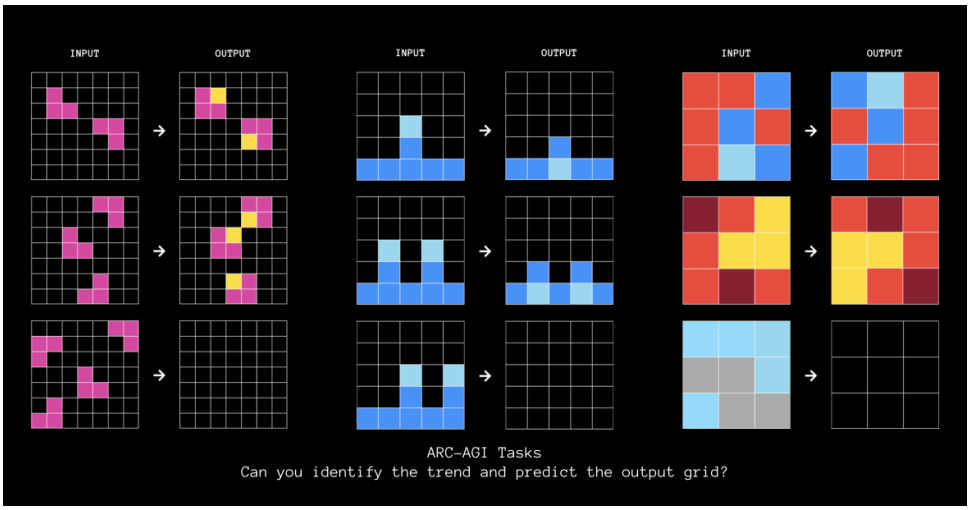

Figure 1: Examples of three ConceptARC tasks illustrating abstract visual transformations between input grids (left) and their corresponding output grids (right).

The objective of this analysis is not to reproduce the dataset taxonomy, but to evaluate whether unsupervised learning methods recover **coherent latent structure** that is broadly consistent with these conceptual categories.

Although synthetic, ConceptARC captures core challenges present in enterprise vision systems: compositional structure, sparse signals, and ambiguity under weak supervision.

3 Data Exploration and Feature Engineering

Several preprocessing and feature engineering steps were applied:

1. **Image normalization**: All grids were padded to a fixed size to enable vectorization.
2. **One-hot encoding**: Each grid was transformed into a high-dimensional, non-negative vector representation over color channels.
3. **Latent representations**: Both input and output images were embedded into a shared latent space.
4. **Transformation vectors**: For each task, a latent transformation vector (Δ) was computed as the difference between output and input embeddings.

These transformation vectors represent *how a task changes an image*, rather than the appearance of the image itself. No label information was used during this process.

4 Unsupervised Models Evaluated

To satisfy the requirement of evaluating multiple unsupervised approaches, three model families were explored.

4.1 Dimensionality Reduction with Non-Negative Matrix Factorization

Non-Negative Matrix Factorization (NMF) was applied to the vectorized images. Multiple component sizes were evaluated, and **64 components were selected as optimal**.

This choice was motivated by:

- Significantly lower reconstruction error compared to random baselines,
- Fewer visually hard-to-reconstruct images,
- Improved stability and interpretability of learned components.

Visual inspection confirmed that NMF with 64 components preserved essential task structure while maintaining part-based interpretability.

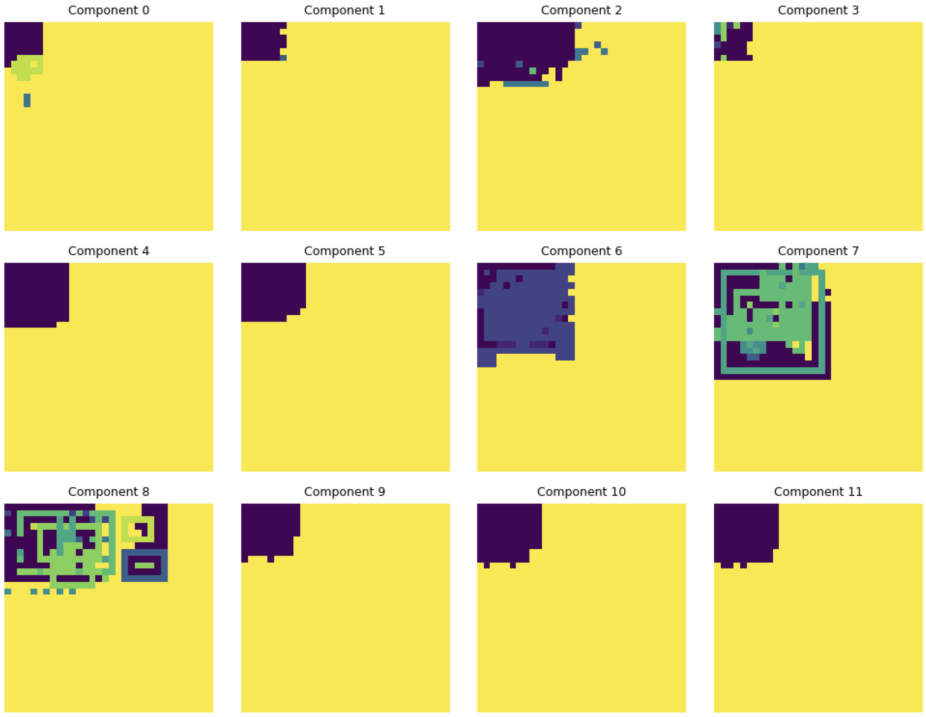

Figure 2: First few components out of 64.

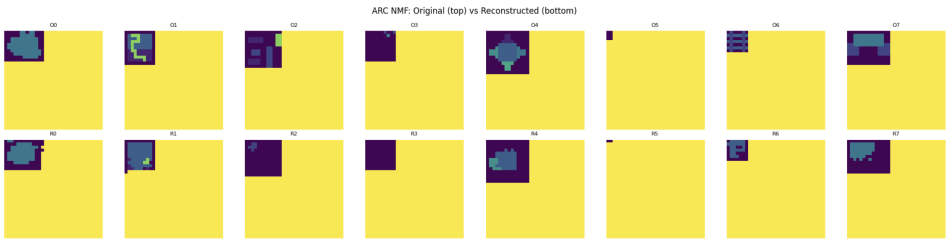

Figure 3: Reconstructions.

4.2 Clustering Latent Transformation Vectors

Clustering was performed on the latent transformation (Δ) vectors using multiple techniques:

- DBSCAN with cosine distance,
- K-Means clustering,
- Hierarchical clustering.

DBSCAN was explored across multiple neighborhood radius (ε) values to understand cluster stability:

ε	Number of Clusters
0.2	4
0.3	8
0.4	12
0.5	7
0.6	4

Moderate ε values (0.3–0.4) produced stable cluster structures and the strongest alignment with known ConceptARC categories. In particular, DBSCAN at $\varepsilon = 0.4$ achieved a Normalized Mutual Information (NMI) score of approximately **0.35**, outperforming alternative approaches. By comparison, K-Means clustering with $k = 16$ achieved a lower NMI of approximately **0.21**, while hierarchical clustering variants yielded NMI scores approximately **0.22**. These results indicate that density-based clustering more effectively captures shared transformation structure than methods that impose a fixed number of clusters or hierarchical partitions.

4.3 Joint Input–Output Embedding Clustering

As an additional baseline, input and output embeddings were concatenated and clustered jointly. This approach was less effective than clustering latent transformation vectors, reinforcing the importance of modeling task behavior by capturing how an image changes rather than relying on static appearance alone.

5 Recommended Final Model

Based on quantitative performance and interpretability, the recommended approach is:

> **Non-Negative Matrix Factorization with 64 components, followed by DBSCAN clustering on latent transformation vectors.**

This model was selected because it:

- Does not require specifying the number of clusters in advance,
- Naturally handles overlapping and ambiguous tasks,
- Achieves the highest alignment with known categories,
- Reveals shared transformation families rather than forcing artificial separation.

The fact that DBSCAN identifies fewer clusters than labeled categories is an important insight, indicating that multiple human-defined categories share common underlying transformations.

6 Key Findings and Insights

The main findings of this analysis are:

- Latent transformation vectors cluster more coherently than raw image representations.
- Several ConceptARC categories overlap significantly in latent space.
- A subset of tasks does not belong to any dominant transformation cluster, indicating ambiguity or unique structure.
- NMF produces interpretable components that support explainability and trust.

These results show that unsupervised learning can both organize tasks and diagnose limitations in existing taxonomies.

7 Limitations and Next Steps

Limitations

- ConceptARC categories are coarse and overlapping by design.
- NMF assumes linear additive structure.
- Clustering results are sensitive to distance metrics and hyperparameters.

Next Steps

- Explore neural network–based embeddings, such as Variational Autoencoders (VAEs),
- Incorporate spatial priors into representations,
- Combine unsupervised clustering with lightweight supervision,
- Evaluate robustness across random seeds and dataset subsets.

Conclusion

This project demonstrates that unsupervised learning can uncover meaningful latent structure in abstract visual reasoning tasks. By focusing on transformations rather than appearances, the analysis reveals reusable reasoning patterns and provides insights into both the strengths and limitations of human-defined task categories.