

Seeing the Gaps: Improving Object Segmentation for Abstract Visual Reasoning in Julia

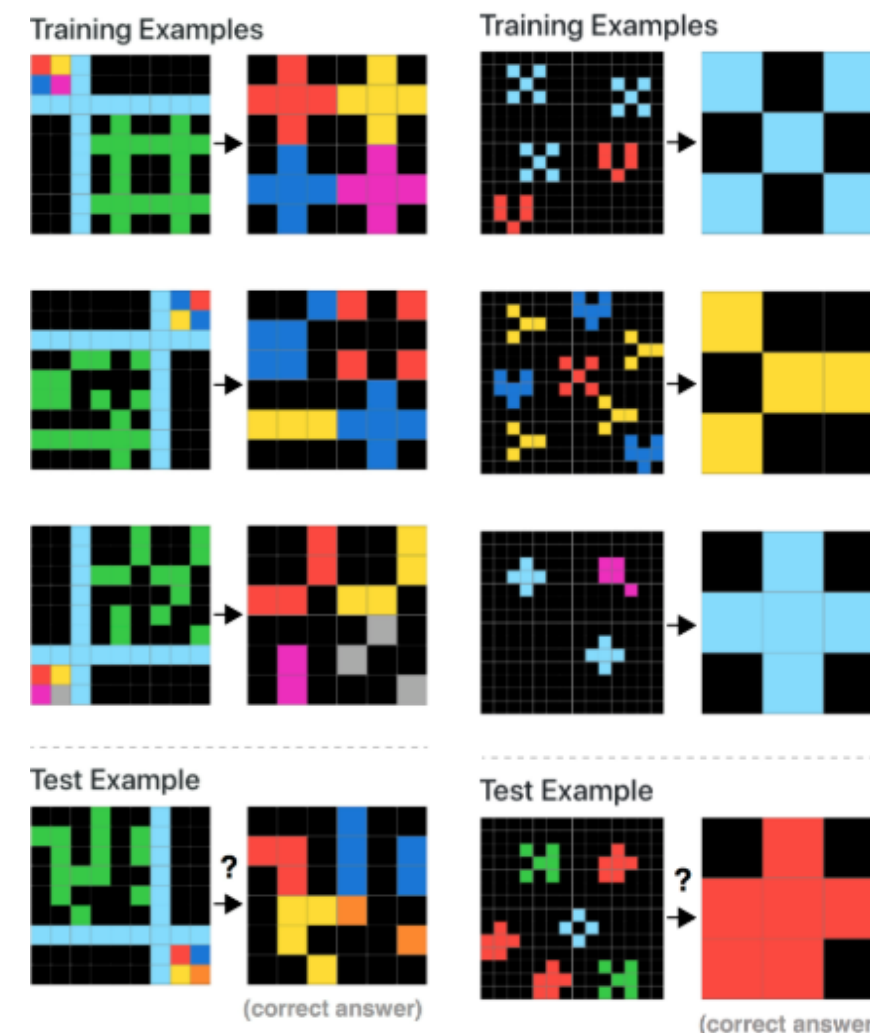
Grammar Extensions and Structural Ranking for the BEN Agent on the ARC Benchmark



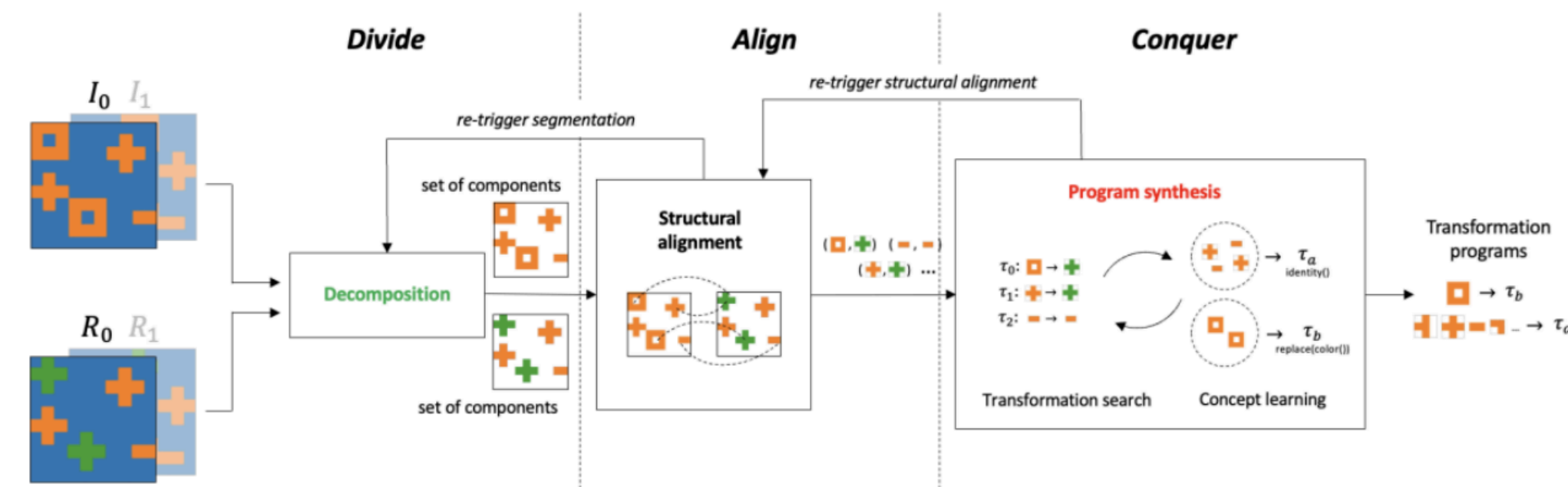
Author: Franciszek Howard
Responsible Professor: Sebastijan Dumancic
Supervisor: Dekel Zak

Background & Motivation

- AI struggles with abstract reasoning despite strong performance in many other domains
- ARC** benchmarks this gap by testing skill acquisition from ~3 examples using only universal human cognitive priors.
- Solver must infer transformation rules from examples the way humans would



Divide Align & Conquer in BEN

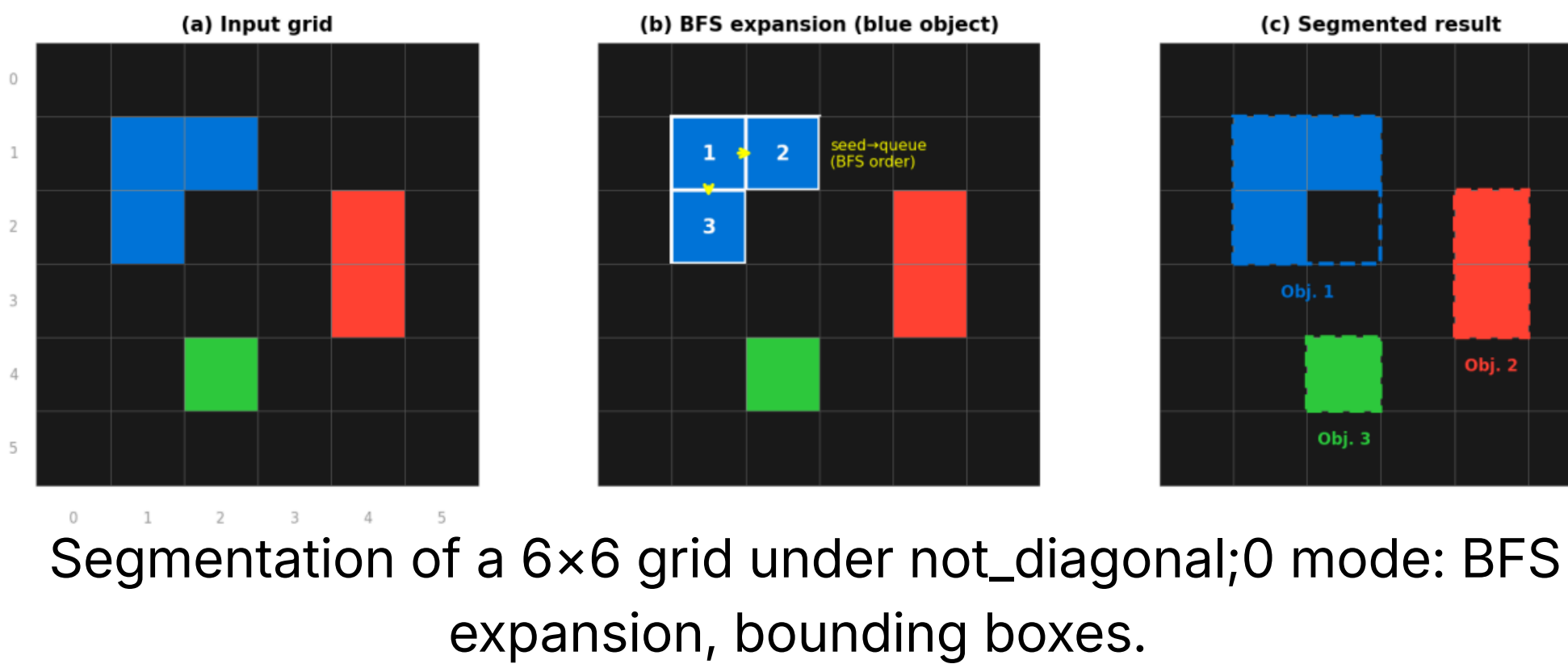


Divide, Align & Conquer (DA&C) decomposes ARC tasks into 3 stages:

- Divide:** segment grids into objects using a pixel-expansion grammar
- Align:** match input ↔ output objects via Structure-Mapping Engine (SME)
- Conquer:** synthesise transformation programs for each aligned pair

Current Implementation

- Segmentation** groups pixels into objects using a grammar of expansion modes
- Mode selection** is costly - each candidate requires the full pipeline to score
- Grammar gap** - cannot handle objects with internal or relational boundaries



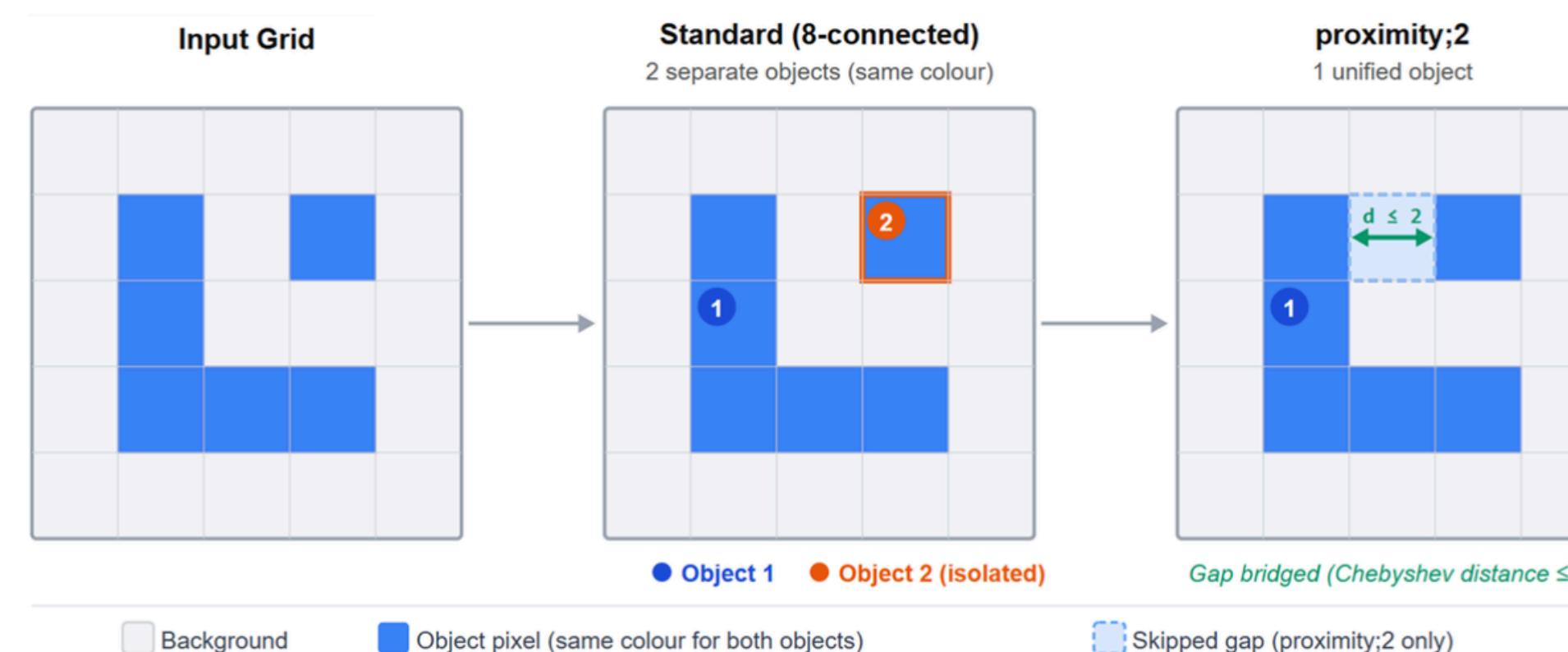
Segmentation of a 6×6 grid under not_diagonal;0 mode: BFS expansion, bounding boxes.

Research Question

Can BEN's segmentation stage be faithfully reimplemented in Julia and improved to increase both the range of addressable tasks and the efficiency of mode selection?

Contributions

- Julia Reimplementation:** Julia compiles to native machine code via LLVM, eliminating interpreter overhead and improving cache locality for pixel-level grid operations
- proximity;N mode** expands to same-colour pixels within Chebyshev distance N, bridging internal gaps that standard connectivity-based modes fragment into separate objects



- Lightweight Structural Ranking:** Replaces the original full-pipeline mode selection with a score_mode function that evaluates all 11 candidate modes on segmentation output alone (before any correspondence or synthesis search).

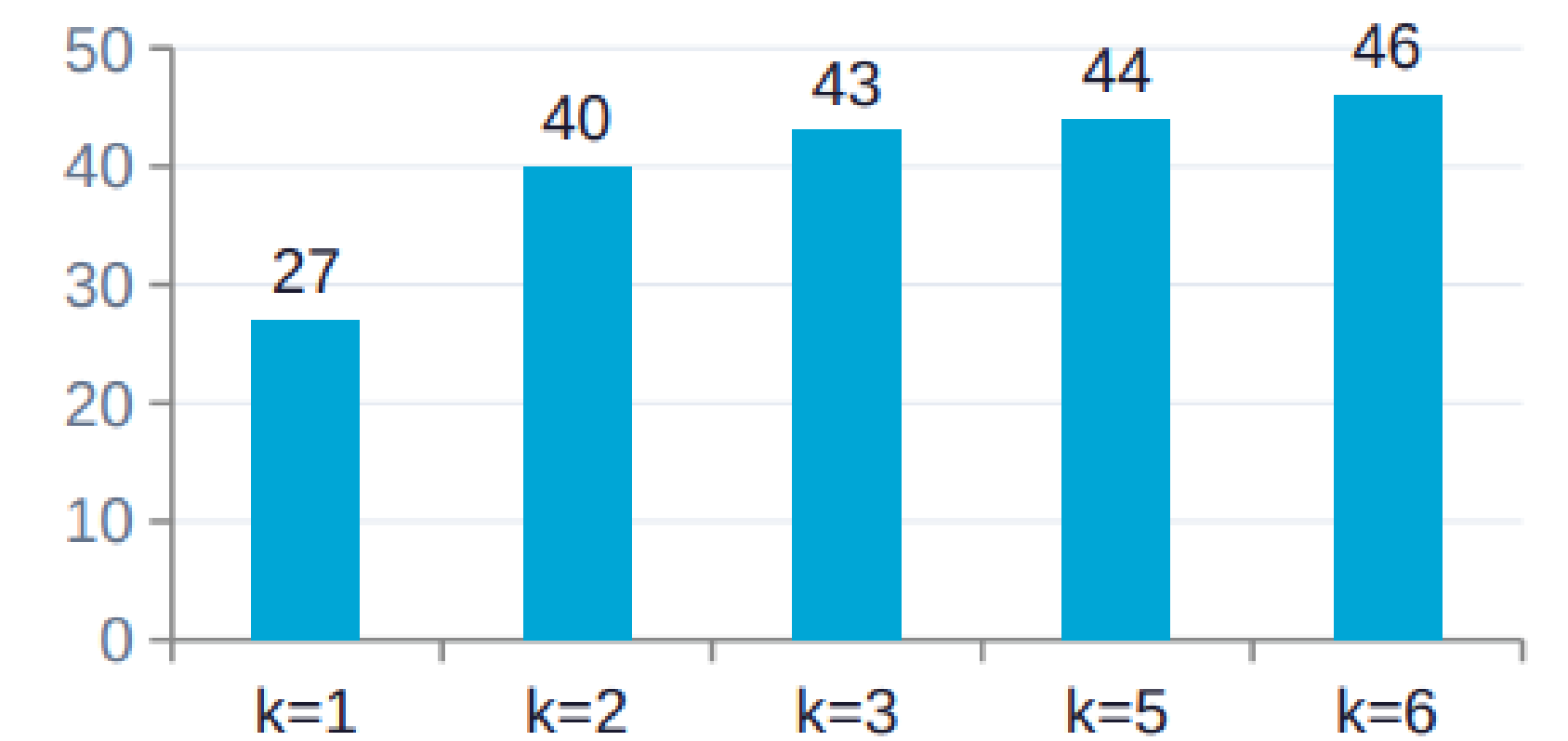
Score combines:

- Count match:** +5 if input object count = output object count, -0.5 per mismatch
- Fragmentation penalty:** discourages >30 objects
- Consistency penalty:** standard deviation across training pairs (key addition)

The most reliable base mode (**not_diagonal;0**) is always guaranteed to be tried second, ensuring a safe fallback regardless of the heuristic's output.

Results

Selected mode	Tasks solved	Distribution of solving segmentation mode across the 46 passing tasks.
not_diagonal;0	31	Distribution of solving segmentation mode across the 46 passing tasks.
not_diagonal;epsilon	3	
diagonal;0	3	
proximity;2	3	
proximity;3	2	
multicolor_not_diagonal;epsilon	2	
diagonal;inferrepeat	2	
Total	46	



Cumulative solves recovered by top-k ranked modes.

Conclusions & Future Work

- ✓ Julia reimplementation integrates cleanly with Herb.jl
- ✓ 5 tasks solved via proximity modes absent from original grammar
- ✓ Structural ranking completes in milliseconds vs. seconds per mode
- ▲ Only 59% of solvable tasks recovered by top-ranked mode, count-match term overranks multicolor modes
- ▲ Grammar cannot express relational boundaries (contours, symmetry)
- Future:** Penalty for multicolor overranking, connected-contour mode, symmetry-based mode, controlled Python ablation.