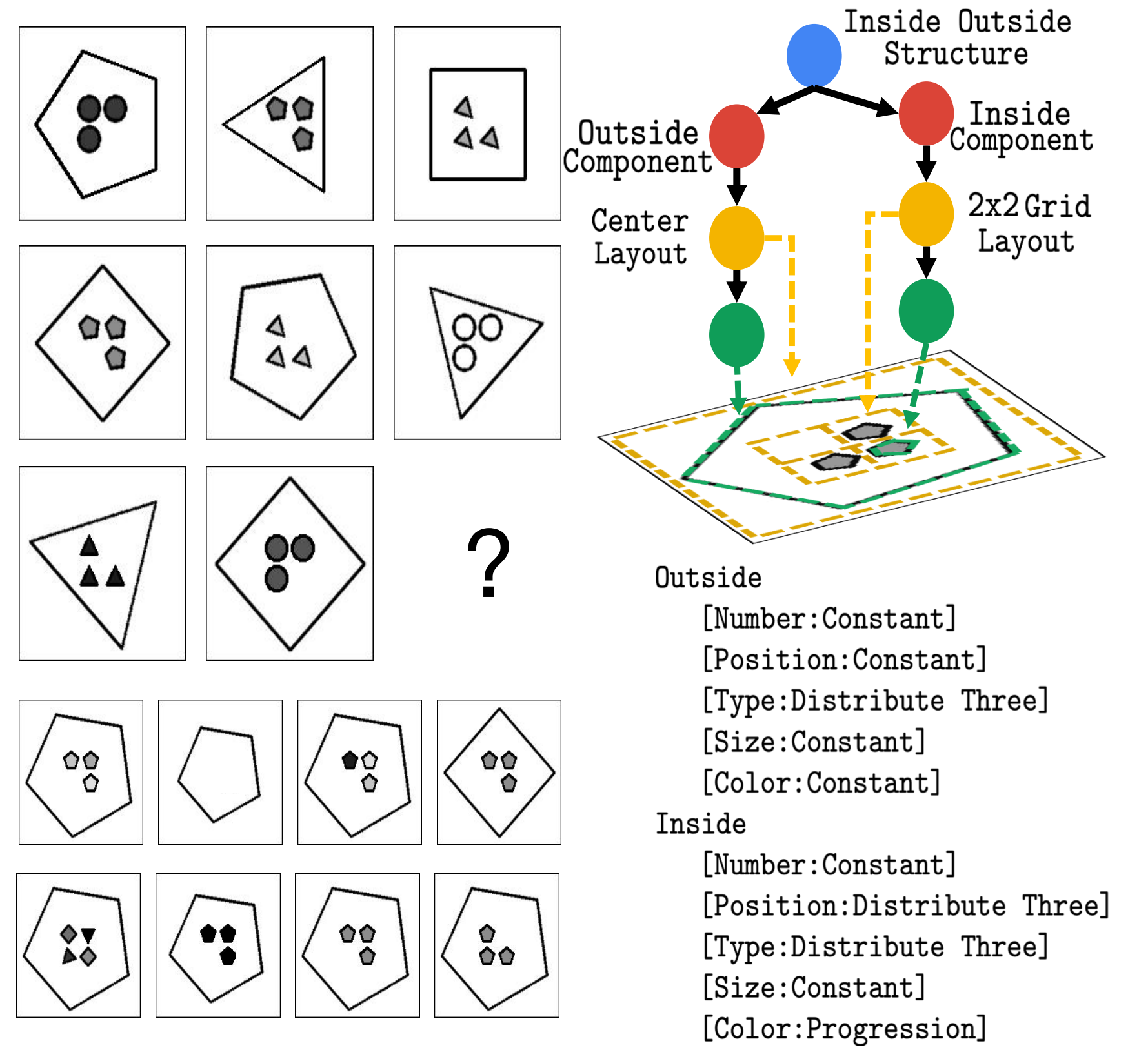
LONG BEACH
CALIFORNIA
June 16-20, 2019

RAVEN: A Dataset for Relational and Analogical Visual rEasoNing

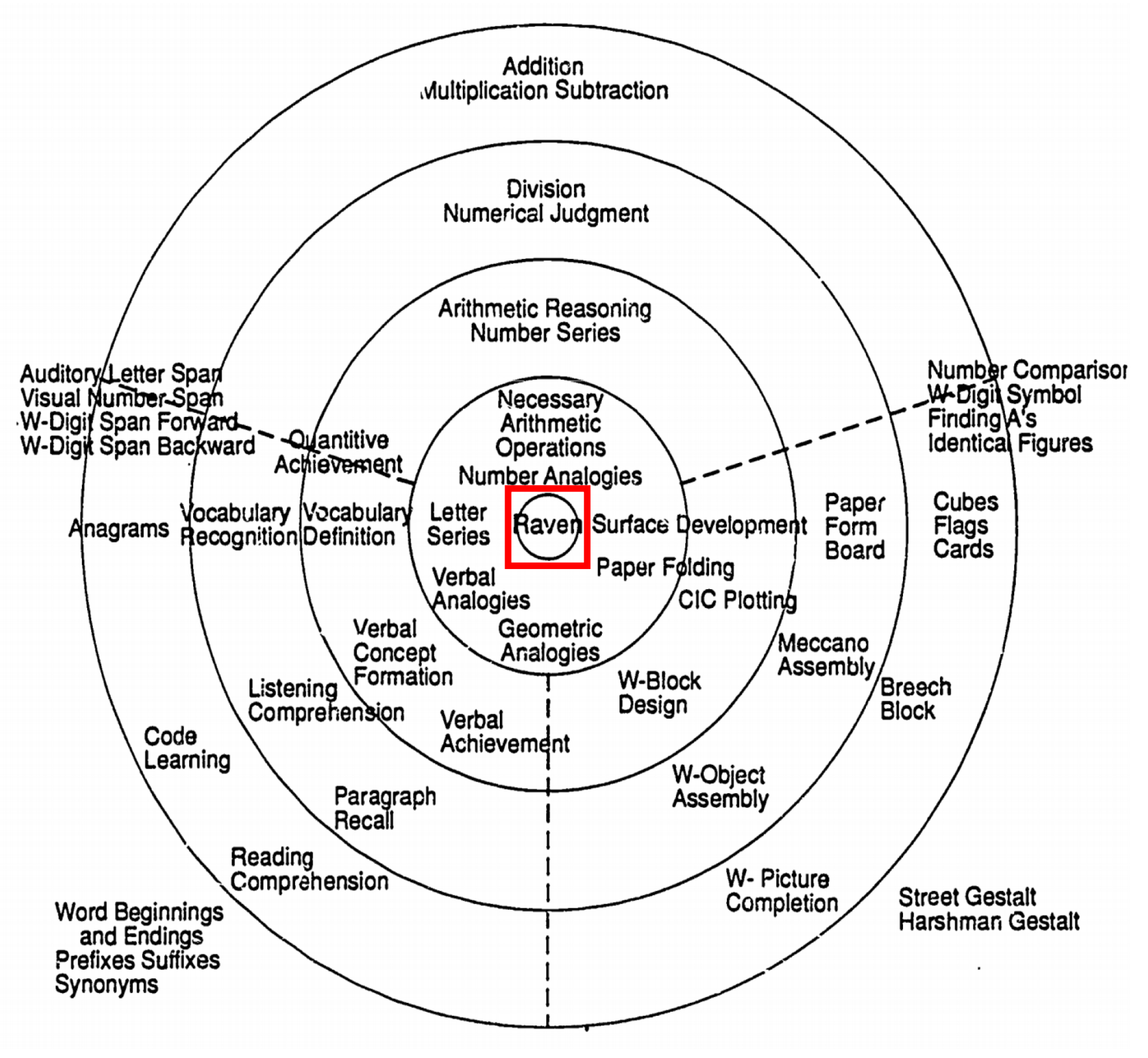
Chi Zhang★,1,2 Feng Gao★,1,2 Baoxiong Jia[1] Yixin Zhu[1,2] Song-Chun Zhu[1,2]
[1] UCLA Center for Vision, Cognition, Learning and Autonomy [2] International Center for AI and Robot Autonomy
{chi.zhang, f.gao, baoxiongjia, yixin.zhu}@ucla.edu, sczhu@stat.ucla.edu

Motivation

> "The study of vision must therefore include not only the study of how to extract from images various aspects of the world … but an inquiry into the nature of the *internal representations* … and make it available as a **basis** for *decisions about our thoughts and actions*."
>
> – David Marr



Structured Understanding



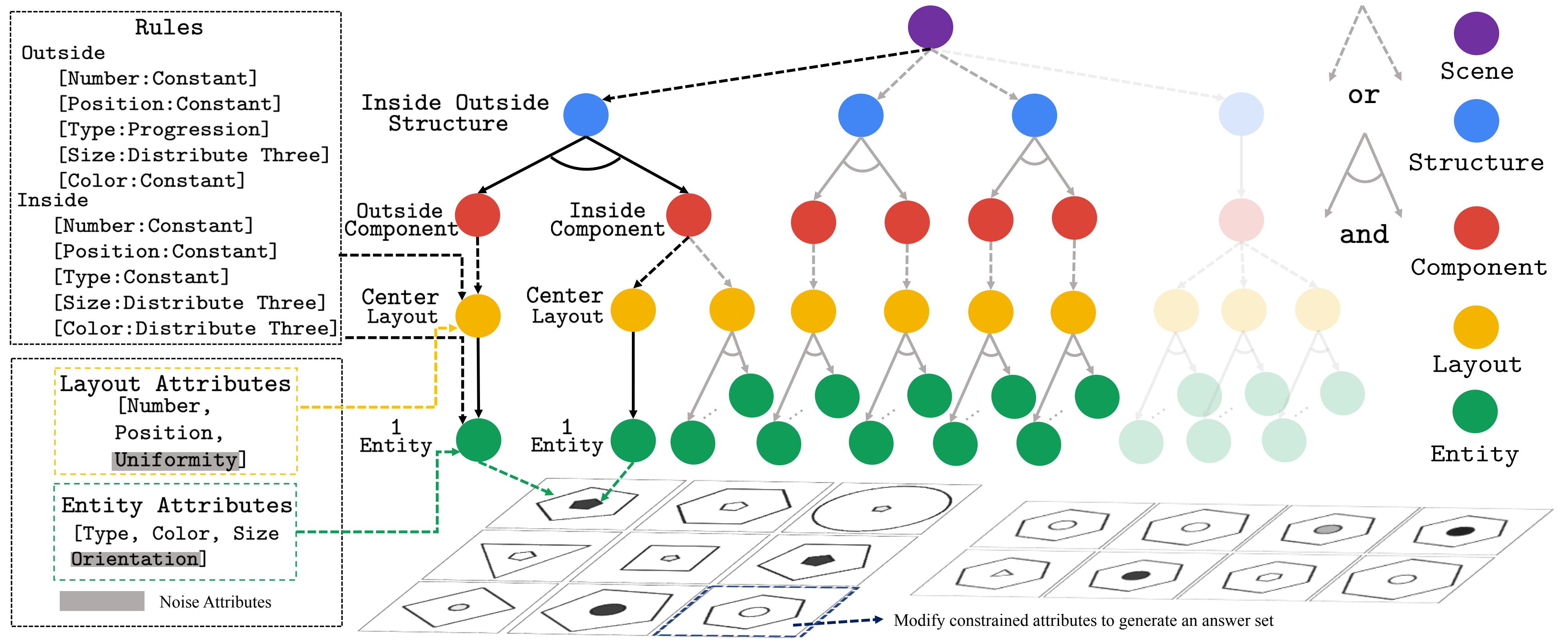
Cognitive Ability Test

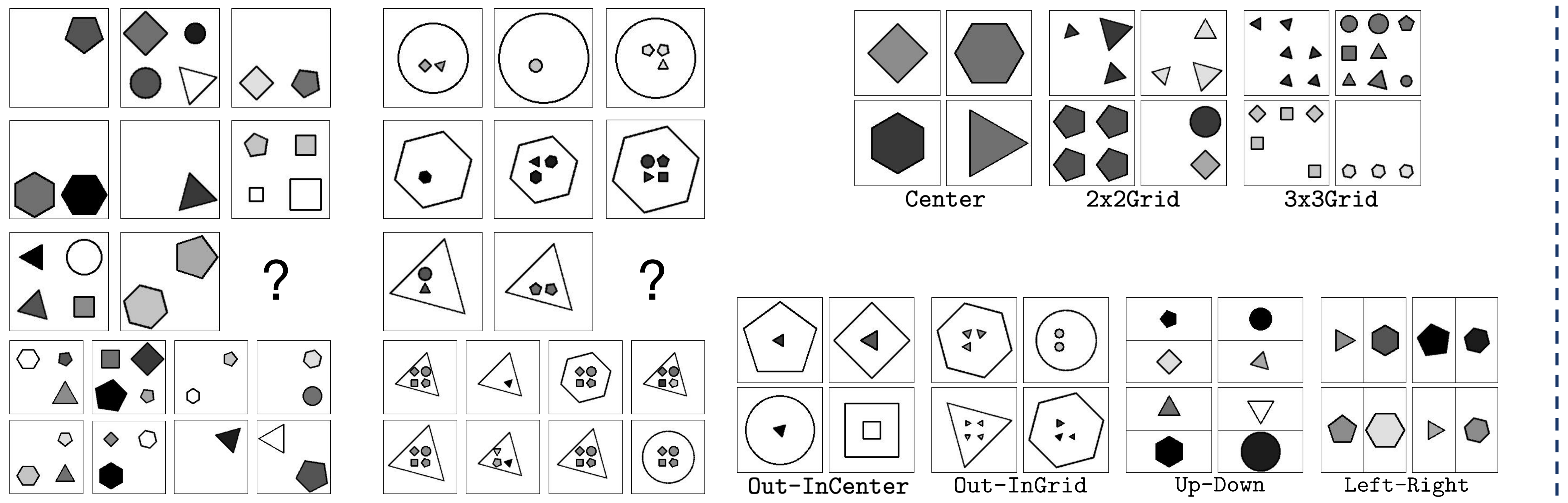
Barrett *et al.* ? Carpenter *et al.*

Hill *et al.* Forbus *et al.*

Neural Data-Driven Rule-Based Symbolic

Generating RAVEN





Comparison and Analysis

	AvgRule	RuleIns	Struct	FigConfig	StructAnno	HumanPerf
PGM	1.37	5	1	3	0	-
RAVEN	6.29	8	4	7	1,120,000	✓

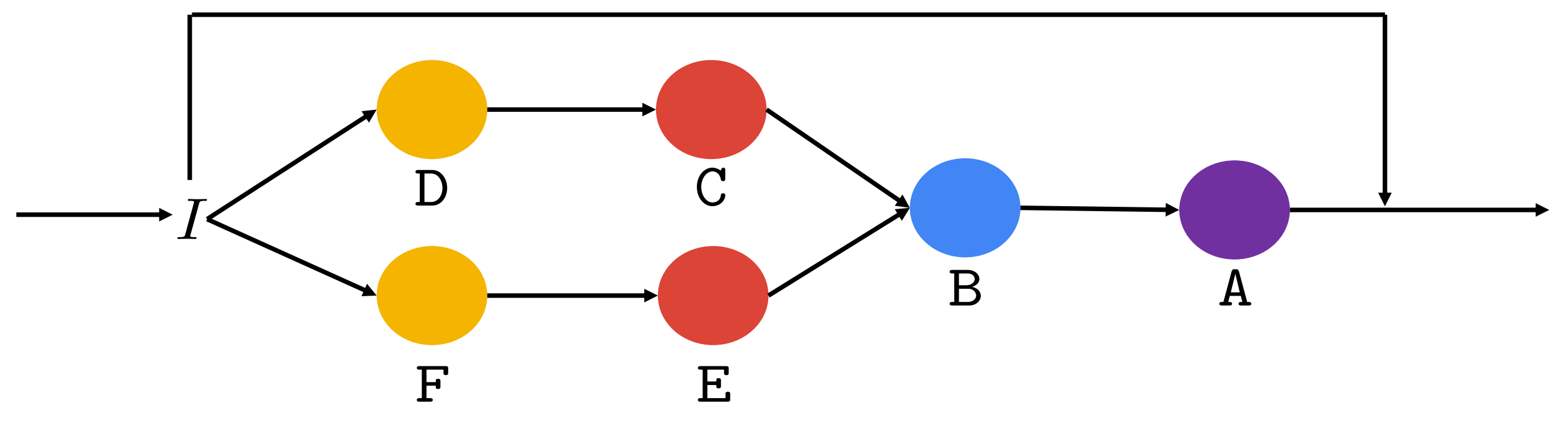
A Structured Module

Dynamic Residual Tree (DRT)

- Generate node sequences from images (RNN/LSTM)

  `A, B, C, D, /, /, E, F, /, /, /, /`

- Assemble nodes into trees



Benchmarking RAVEN

Method	Acc	Center	2x2Grid	3x3Grid	L-R	U-D	O-IC	O-IG
LSTM	13.07%	13.19%	14.13%	13.69%	12.84%	12.35%	12.15%	12.99%
WReN	14.69%	13.09%	28.62%	28.27%	7.49%	6.34%	8.38%	10.56%
CNN	36.97%	33.58%	30.30%	33.53%	39.43%	41.26%	43.20%	37.54%
ResNet	53.43%	52.82%	41.86%	44.29%	58.77%	60.16%	63.19%	53.12%
LSTM+DRT	13.96%	14.29%	15.08%	14.09%	13.79%	13.24%	13.99%	13.29%
WReN+DRT	15.02%	15.38%	23.26%	29.51%	6.99%	8.43%	8.93%	12.35%
CNN+DRT	39.42%	37.30%	30.06%	34.57%	45.49%	45.54%	45.93%	37.54%
ResNet+DRT	**59.56%**	**58.08%**	**46.53%**	**50.40%**	**65.82%**	**67.11%**	**69.09%**	**60.11%**
Human	84.41%	95.45%	81.82%	79.55%	86.36%	81.81%	86.36%	81.81%
Solver★	100%	100%	100%	100%	100%	100%	100%	100%

Future Work

- How to formulate visual reasoning
- Better way of structured reasoning
- Top-down bottom-up method