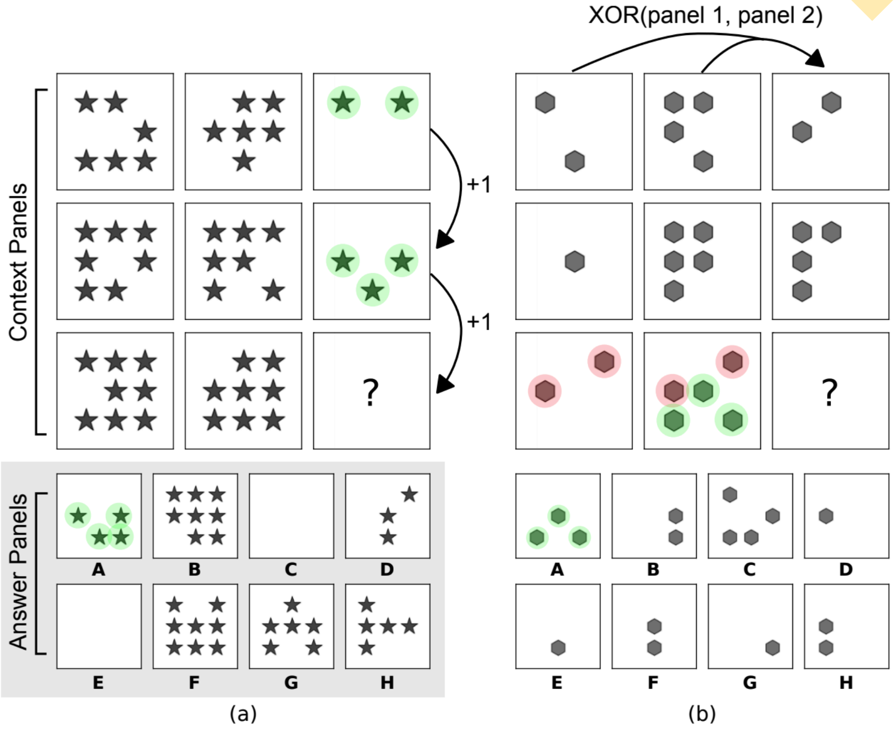
Efficient Abstract Reasoning with Dual-Contrast Network

Tao Zhuo, Mohan Kankanhalli

School of Computing, National University of Singapore

Abstract reasoning: measure machine intelligence

- Abstract reasoning is a crucial component of human intelligence.
  - Understanding patterns
  - Interpreting patterns
  - Solving problems
- Unknown logical rules in Raven's Progressive Matrices (RPM)
  - Shape, size, color, number, position, AND, OR, XOR, consistent, etc.
  - Row-wise or column-wise is unknown
  - Number of hidden rules is unknown
  - Logical rules in different problems are different



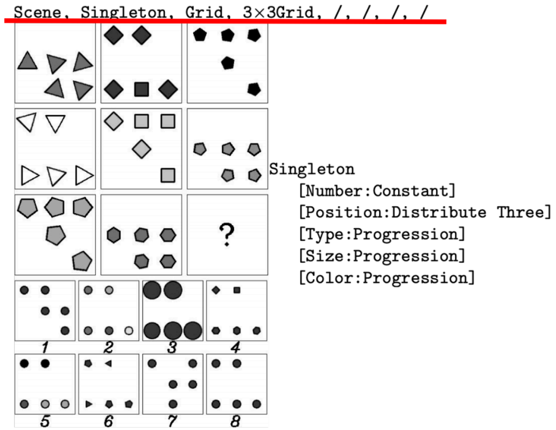
Barrett et al. Measuring abstract reasoning in neural networks, ICML 2018

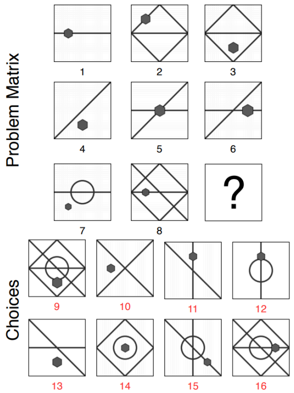
Learning with auxiliary annotations (logical rules)

- To learn good feature representations on logical rules.

$$\mathcal{L} = \mathcal{L}_{target} + \alpha \mathcal{L}_{struct}$$

$$\mathcal{L} = \mathcal{L}_{target} + \beta \mathcal{L}_{meta\_target}$$





(shape, line, color, number, position, size, type, progression, XOR, OR, AND, consistent union)

Zhang et al. RAVEN: A Dataset for Relational and Analogical Visual rEasoNing, CVPR 2019

Barrett et al. Measuring abstract reasoning in neural networks, ICML 2018

Learning with auxiliary annotations

- Human can learn how to solve RPM without auxiliary annotations
- Auxiliary annotations might be not provided in some datasets
- Auxiliary annotations do not always boost the performance

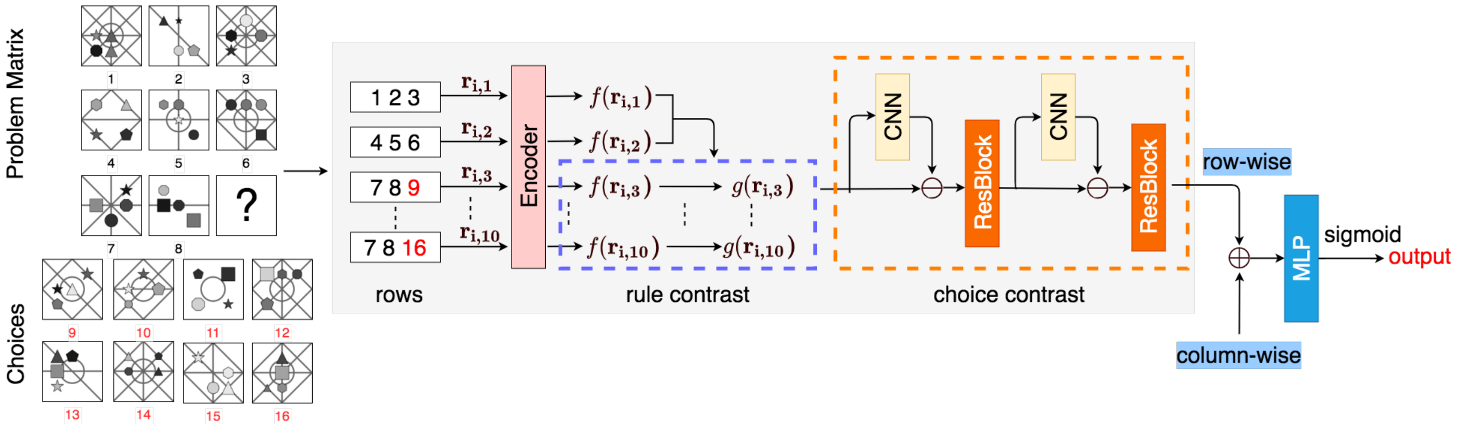
$$\mathcal{L} = \mathcal{L}_{target} + \alpha\mathcal{L}_{struct} \qquad \mathcal{L} = \mathcal{L}_{target} + \beta\mathcal{L}_{meta\_target}$$

- **Therefore, we aim to learn the model with only ground truth target (answer)**

$$\mathcal{L} = \mathcal{L}_{target}$$

Our method: dual-contrast

- **Rule contrast:** the three rows/columns must share the same rules
  - Represent the relative rule differences of different rows
- **Choice contrast:** the correct answer has to best satisfy the problem matrix than other choices
  - Increase the relative differences among the candidate



Testing accuracy on RAVEN and PGM dataset

- Aux denotes auxiliary annotations (logical rules)
- Avg denotes average accuracy
- **Human performance on RAVEN dataset is 84.41%**

Method	Aux	Avg	RAVEN	PGM
ResNet-18+DRT Zhang et al. (2019a)	✓	-	59.56	-
WReN+Aux Santoro et al. (2018)	✓	55.44	33.97	76.90
LEN+Aux Zheng et al. (2019)	✓	70.85	59.40	82.30
MXGNet+Aux Wang et al. (2020)	✓	-	-	**89.60**
ACL Kim et al. (2020)	✓	-	**93.71**	-
LSTM Zhang et al. (2019b)		24.44	13.07	35.80
CNN Zhang et al. (2019b)		34.99	36.97	33.00
WReN Santoro et al. (2018)		40.10	17.94	62.60
Wild-ResNet Santoro et al. (2018)		-	-	48.00
ResNet-50 Santoro et al. (2018)		64.13	86.26#	42.00
MLRN Jahrens & Martinetz (2020)		55.33	12.50*	**98.03**
LEN Zheng et al. (2019)		70.50	72.90	68.10
CoPINet Zhang et al. (2019b)		73.90	91.42	56.37
MXGNet Wang et al. (2020)		75.31	83.91	66.70
DCNet-RC		78.10	92.74	63.45
DCNet-CC		47.12	36.47	57.76
DCNet		**81.08**	**93.58**	68.57

Few-shot results on RAVEN

Training Set Size	CoPINet	**DCNet**
658 (1.57%)	44.48	**60.09** (+15.61)
1, 316 (3.13%)	57.69	**71.91** (+14.22)
2, 625 (6.25%)	65.55	**79.75** (+14.20)
5, 250 (12.5%)	74.53	**84.42** (+9.89)
10, 500 (25.0%)	80.92	**87.95** (+7.03)
21, 000 (50.0%)	86.43	**91.31** (+4.88)
42, 000 (100%)	91.42	**93.87** (+2.45)

Few-shot results on PGM

Training Set Size	CoPINet	**DCNet**
293 (0.25%)	14.73	**15.94** (+1.21)
1, 172 (0.10%)	15.48	**18.76** (+3.32)
4, 688 (0.39%)	18.39	**25.78** (+7.39)
18, 750 (1.56%)	22.07	**34.04** (+11.97)
75, 000 (6.25%)	32.39	**43.10** (+10.71)
300, 000 (25%)	43.89	**50.26** (+6.37)
12,000,000 (100%)	56.37	**68.57** (+12.20)

Generalization test results

Generalization test on RAVEN, each model is trained on one row-wise configuration and tested on all column-wise configurations

Method	Avg	Figure	*Center*	*2*2Grid*	*3*3Grid*	*L-R*	*U-D*	*O-IC*	*O-IG*
CoPINet	66.71	*Center*	98.05	53.05	49.81	78.65	87.20	61.40	38.35
	68.19	*2*2Grid*	64.55	71.20	67.65	87.25	83.00	62.25	41.45
	63.76	*3*3Grid*	48.90	57.95	76.85	63.95	51.80	71.70	75.20
	69.83	*L-R*	55.50	51.90	54.55	98.25	92.35	76.40	59.85
	73.95	*U-D*	63.25	56.30	55.75	81.90	67.40	94.70	98.35
	68.81	*O-IC*	51.50	46.30	54.60	88.85	64.35	88.90	87.15
	68.74	*O-IG*	44.00	48.60	54.75	75.45	84.30	82.95	91.10
DCNet	**70.76**	*Center*	97.25	54.10	51.80	84.75	86.40	72.90	48.10
	74.09	*2*2Grid*	69.35	76.60	66.55	86.85	88.60	72.45	58.20
	68.75	*3*3Grid*	49.20	56.45	81.75	78.20	78.60	74.80	62.25
	73.21	*L-R*	60.30	53.35	59.30	98.50	91.70	83.85	65.45
	75.38	*U-D*	69.30	54.10	56.75	93.25	98.75	85.65	69.85
	68.11	*O-IC*	46.75	40.35	47.30	84.60	88.00	97.80	72.00
	66.36	*O-IG*	48.25	44.60	51.80	79.50	75.75	77.05	87.55

Generalization test on the neutral regime of PGM dataset

Method	neutral	interpolation	extrapolation
WReN	62.6	**64.4**	17.2
MLRN	**98.0**	57.8	14.9
DCNet	68.6	59.7	**17.8**

Generalization ability still needs further improvement!

THANK YOU