Towards Generative Abstract Reasoning: Completing Raven's Progressive Matrix via Rule Abstraction and Selection

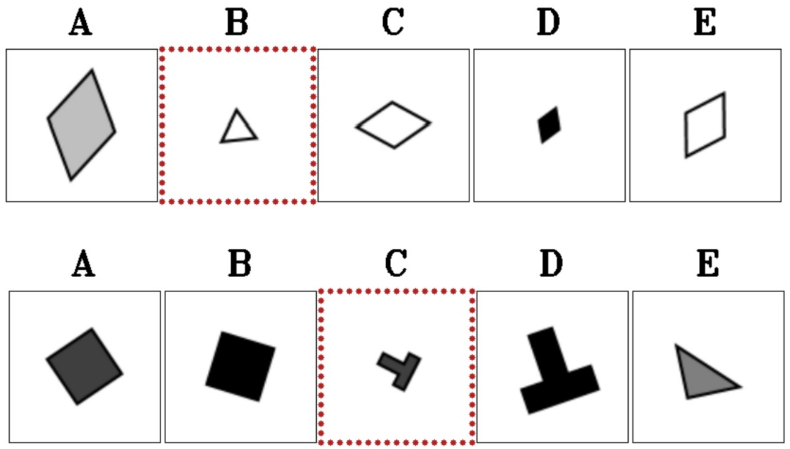
Fan Shi, Bin Li, Xiangyang Xue

Shanghai Key Laboratory of Intelligent Information Processing

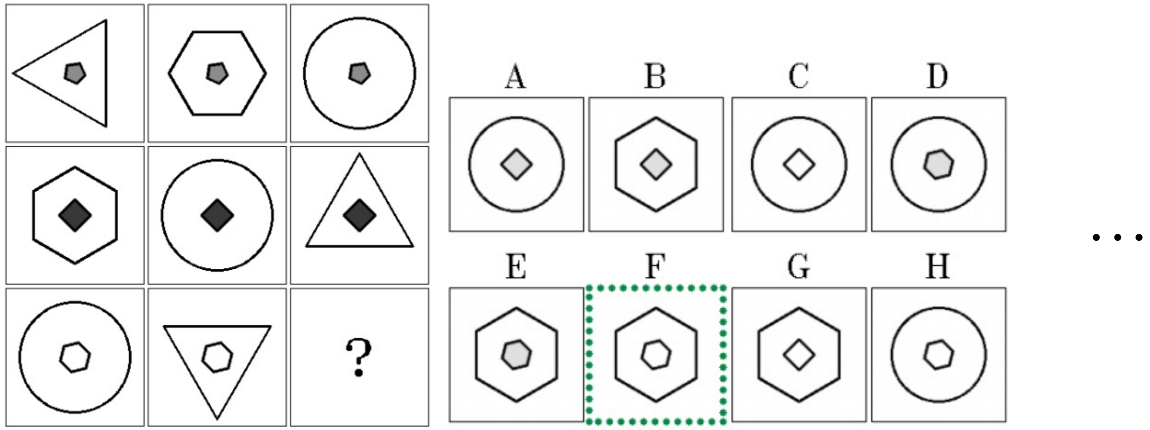
School of Computer Science, Fudan University

Background

Abstract Visual Reasoning Problems are analogical reasoning problems based on abstract visual concepts



Odd-One-Out Problems[1]
Find rule-breaking images on panels



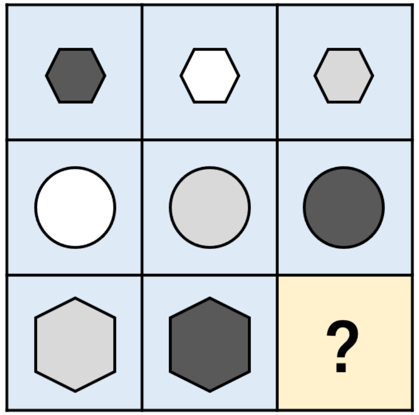
Raven's Progressive Matrices (RPMs)[1]
Find the missing image from the provided candidates

- [1] Mikołaj Małkin´ski and Jacek Man´dziuk. A review of emerging research directions in abstract visual reasoning. arXiv preprint arXiv:2202.10284, 2022b.

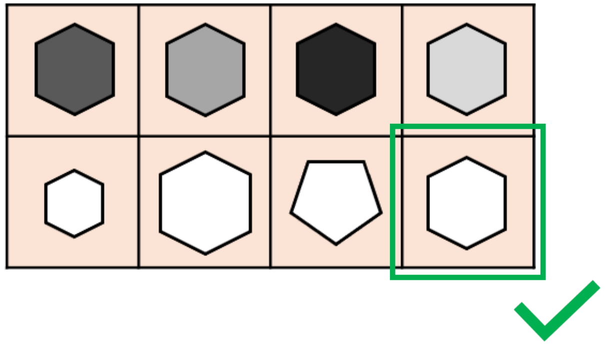
Background

The core challenges of abstract visual reasoning tasks:

- Discover abstract concepts from images automatically
- Categorize and understand different types of rules on abstract concepts



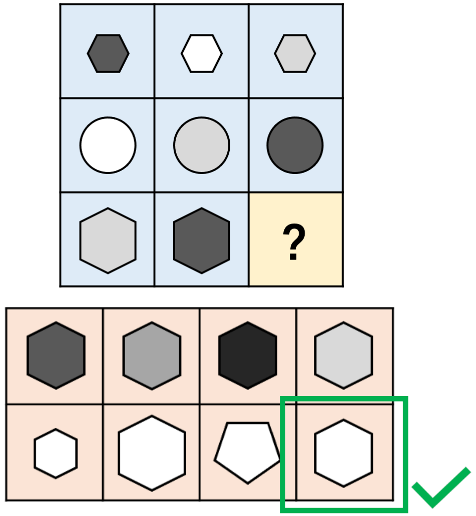
Concepts	Rules
Size	Constant
Color	Permutation
Shape	Constant
Rotation	Constant



Background

Selective Reasoning Tasks

Selecting correct answers from provided candidates



Shortcut learning on selective tasks

Statistics of candidate attributes

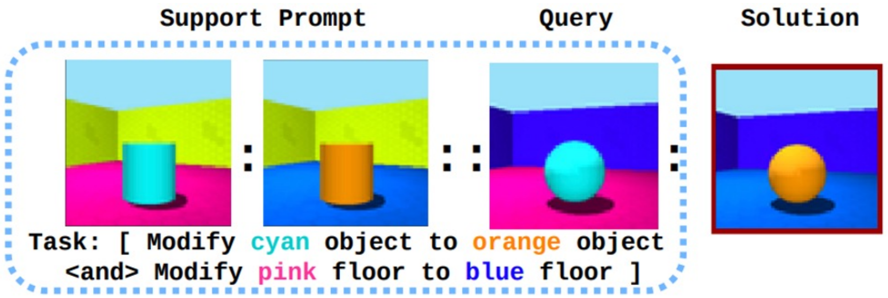
- Shape: **hexagon 7/8**, pentagon 1/8
- Size: small 1/8, **middle 6/8**, large 1/8
- Color: **white 4/8**, light gray 1/8, gray 1/8, dark gray 1/8， black 1/8

Guess attributes of the answer:

=> white, middle, hexagon?

Generative Reasoning Tasks

Generating answers from context panels



Visual in-context learning task[1]

- require in-depth understanding of abstract concepts and concept-changing rules

Generative reasoning tasks can better reflect abstract reasoning ability of intelligent systems.[2]

- [1] Dedhia, Bhishma, et al. "Im-Promptu: In-Context Composition from Image Prompts." Advances in Neural Information Processing Systems 36 (2024).
- [2] Mitchell, Melanie. "Abstraction and analogy-making in artificial intelligence." Annals of the New York Academy of Sciences 1505.1 (2021): 79-101.

Background

Existing methods have not thoroughly shown such generative reasoning ability in realistic reasoning problems or benchmarks, e.g., RAVEN/I-RAVEN

- GCA[1] cannot parse interpretable abstract concepts and concept-changing rules
- ALANS[2] and PrAE[3] introduce artificial priors when designing perception and reasoning processes
- LGPP[4] and CLAP[5] can hardly generate answers for realistic RPM problems

- [1] Pekar, Niv, Yaniv Benny, and Lior Wolf. "Generating correct answers for progressive matrices intelligence tests." Advances in Neural Information Processing Systems 33 (2020): 7390-7400.
- [2] Chi Zhang, Sirui Xie, Baoxiong Jia, Ying Nian Wu, Song-Chun Zhu, and Yixin Zhu. Learning algebraic representation for systematic generalization in abstract reasoning. arXiv preprint arXiv:2111.12990, 2021b.
- [3] Chi Zhang, Baoxiong Jia, Song-Chun Zhu, and Yixin Zhu. Abstract spatial-temporal reasoning via probabilistic abduction and execution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9736–9746, 2021a.
- [4] Fan Shi, Bin Li, and Xiangyang Xue. Raven's progressive matrices completion with latent gaussian process priors. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pp. 9612–9620, 2021.
- [5] Fan Shi, Bin Li, and Xiangyang Xue. Compositional law parsing with latent random functions. In International Conference on Learning Representations, 2023.

Contributions

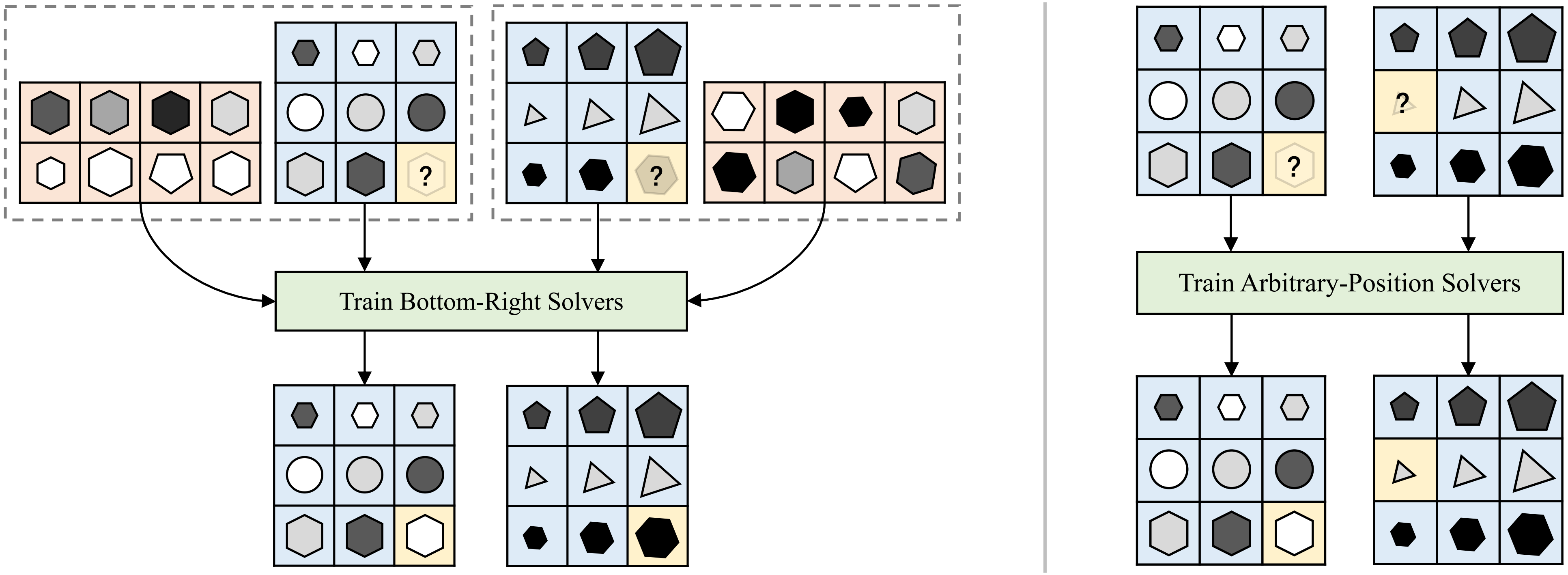
We propose a novel deep latent variable model RAISE to solve generative RPM problems.

- RAISE learns independent **latent concepts** from RPMs automatically
- RAISE infers concept-changing rules and abstracts **learnable atomic rules**
- RAISE can generate answers at **arbitrary or multiple positions** of an RPM
- RAISE can **detect rule-breaking images** in odd-one-out problems and solve RPMs with **unseen rule-attribute combinations**

RAISE

Design 1: RAISE is trained by predicting answers at **arbitrary positions**.

- Training via arbitrary-position answer generation requires only original RPM panels, eliminating the influence of biases in candidate answers



RAISE

Design 2: **Explicit** definitions of latent concepts and atomic rules in the answer generation process

- The latent concepts and the acquired atomic rules are composed to represent **large amounts of rule-attribute combinations** and even **out-of-distribution combinations** in RPMs

- Color In: Gray
- Type In: Circle
- Size In: Small
- Type Out: Pentagon
- Size Out: Middle

Latent concepts of an image

Latent Concepts
- Size In & Size Out： small, middle, large
- Color In: white, gray, black
- Type In & Type Out: triangle, square, circle

×

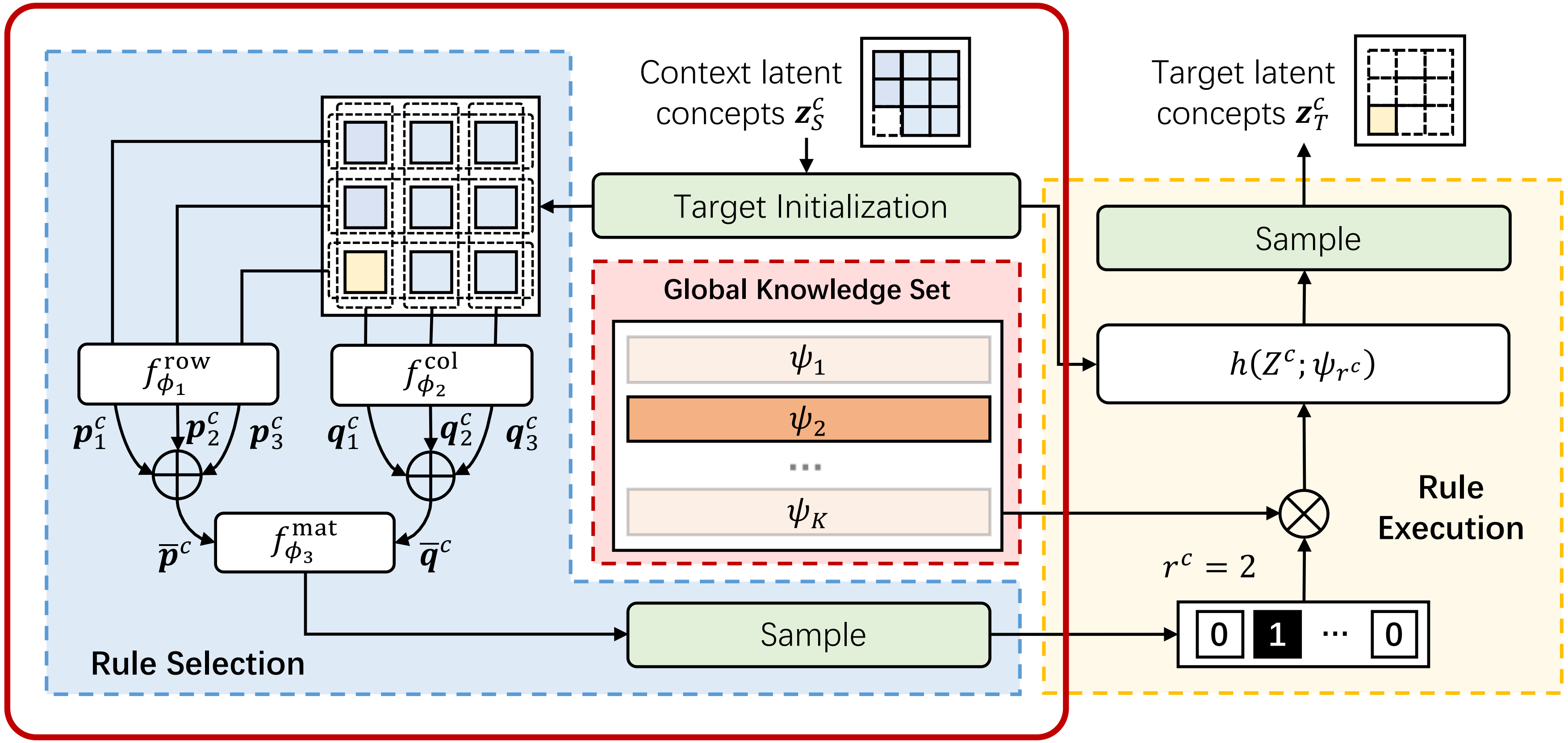
Atomic Rules
- Constant
- Progression
- Permutation
- …

Composition of latent concepts and atomic rules

RAISE

The Rule Selection Stage:

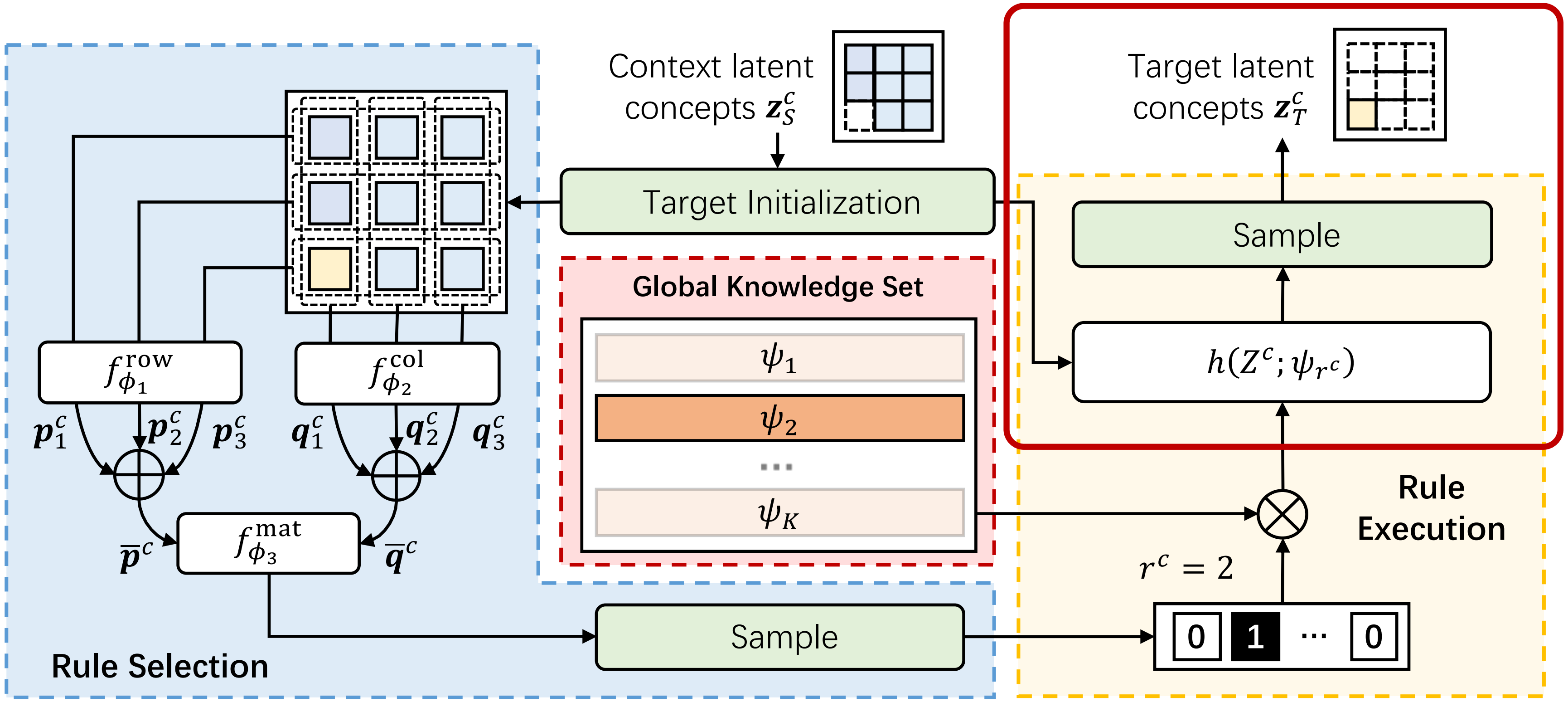
- extract row-wise and column-wise representations $\overline{\boldsymbol{p}}^c$ and $\overline{\boldsymbol{q}}^c$ from the context latent concepts $\boldsymbol{z}_S^c$
- predict an indicator r^c to selection the r^c-th atomic rule ψ_{r^c} for the c-th latent concept



RAISE

The Rule Execution Stage:

- parameterize the prediction net $h(*;\psi_{r^c})$ according to the selected atomic rule ψ_{r^c}
- predict the target representations in the matrix Z^c via convolutional blocks in $h(*;\psi_{r^c})$



Bottom-Right Answer Selection

Table 1: **The accuracy (%) of selecting bottom-right answers on different configurations (i.e., *Center*, *L-R*, etc) of RAVEN/I-RAVEN**. The table displays the average results of ten trials.

Models	Average	Center	L-R	U-D	O-IC	O-IG	2×2Grid	3×3Grid
GCA-I	12.0/24.1	14.0/30.2	7.9/22.4	7.5/26.9	13.4/32.9	15.5/25.0	11.3/16.3	14.5/15.3
GCA-R	13.8/27.4	16.6/34.5	9.4/26.9	6.9/28.0	17.3/37.8	16.7/26.0	11.7/19.2	18.1/19.3
GCA-C	32.7/41.7	37.3/51.8	26.4/44.6	21.5/42.6	30.2/46.7	33.0/35.6	37.6/38.1	43.0/32.4
ALANS	54.3/62.8	42.7/63.9	42.4/60.9	46.2/65.6	49.5/64.8	53.6/52.0	70.5/66.4	75.1/65.7
PrAE	80.0/85.7	97.3/**99.9**	96.2/97.9	96.7/97.7	95.8/98.4	68.6/76.5	**82.0/84.5**	23.2/45.1
LGPP	6.4/16.3	9.2/20.1	4.7/18.9	5.2/21.2	4.0/13.9	3.1/12.3	8.6/13.7	10.4/13.9
ANP	7.3/27.6	9.8/47.4	4.1/20.3	3.5/20.7	5.4/38.2	7.6/36.1	10.0/15.0	10.5/15.6
CLAP	17.5/32.8	30.4/42.9	13.4/35.1	12.2/32.1	16.4/37.5	9.5/26.0	16.0/20.1	24.3/35.8
Transformer	40.1/64.0	98.4/99.2	67.0/91.1	60.9/86.6	14.5/69.9	13.5/57.1	14.7/25.2	11.6/18.6
RAISE	**90.0/92.1**	**99.2**/99.8	**98.5/99.6**	**99.3/99.9**	**97.6/99.6**	**89.3/96.0**	68.2/71.3	**77.7/78.7**

Answer Selection at Arbitrary Positions

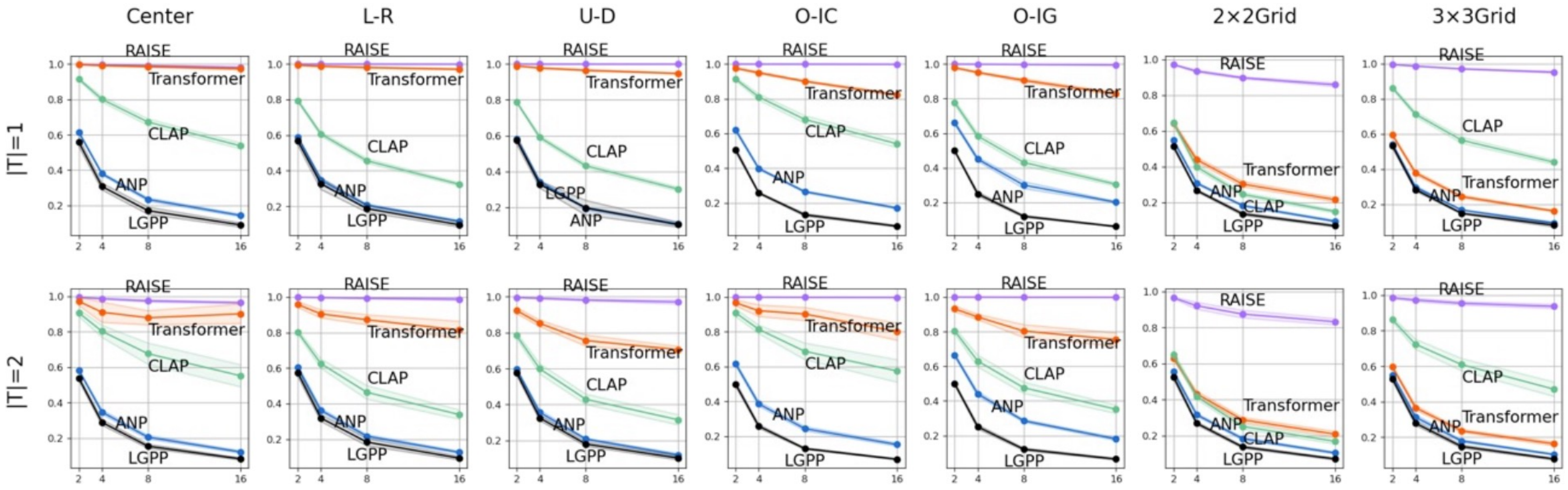


Figure 2: **Selection accuracy at arbitrary positions**. The selection accuracy of RAISE (purple), Transformer (orange), CLAP (green), ANP (blue), and LGPP (black) in arbitrary positions. The x-axis of each plot indicates the number of candidates, and the y-axis is the selection accuracy.

Answer Selection at Arbitrary Positions

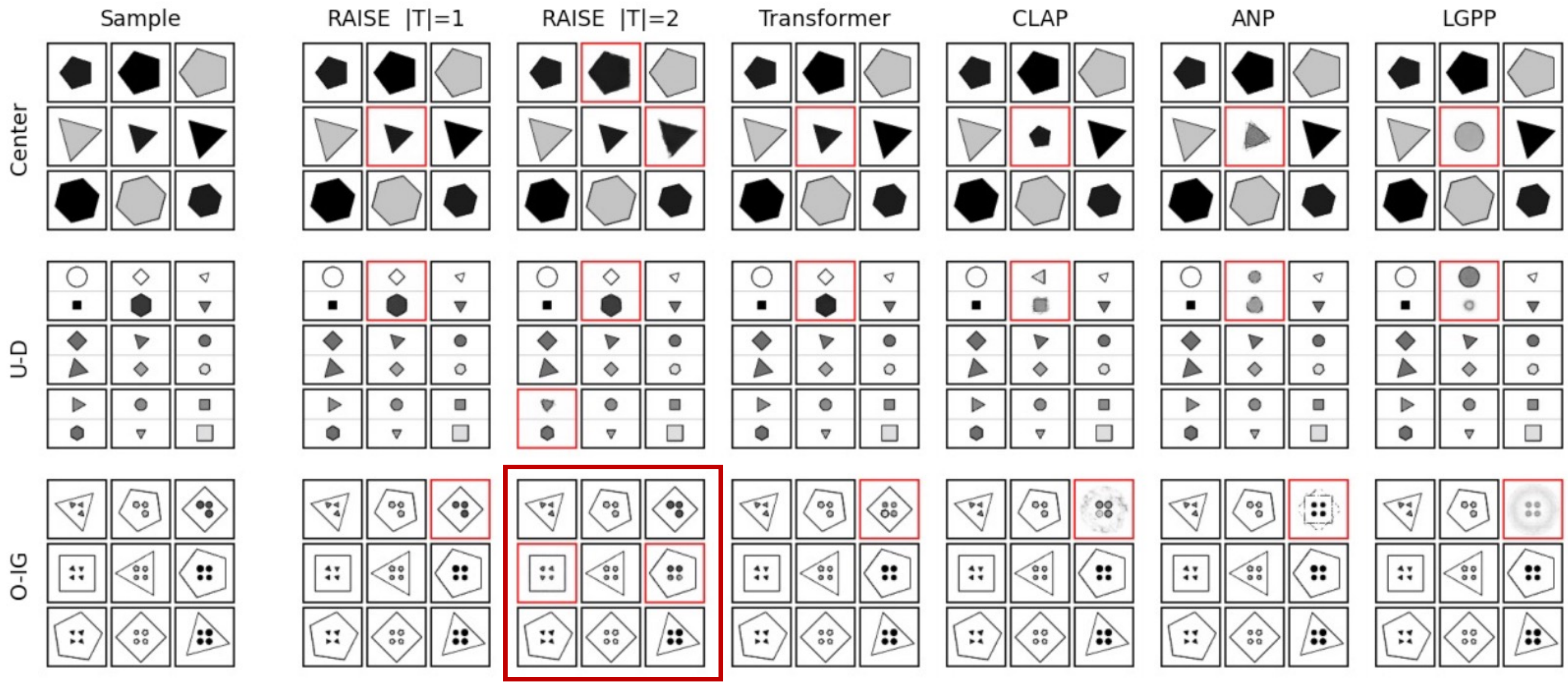
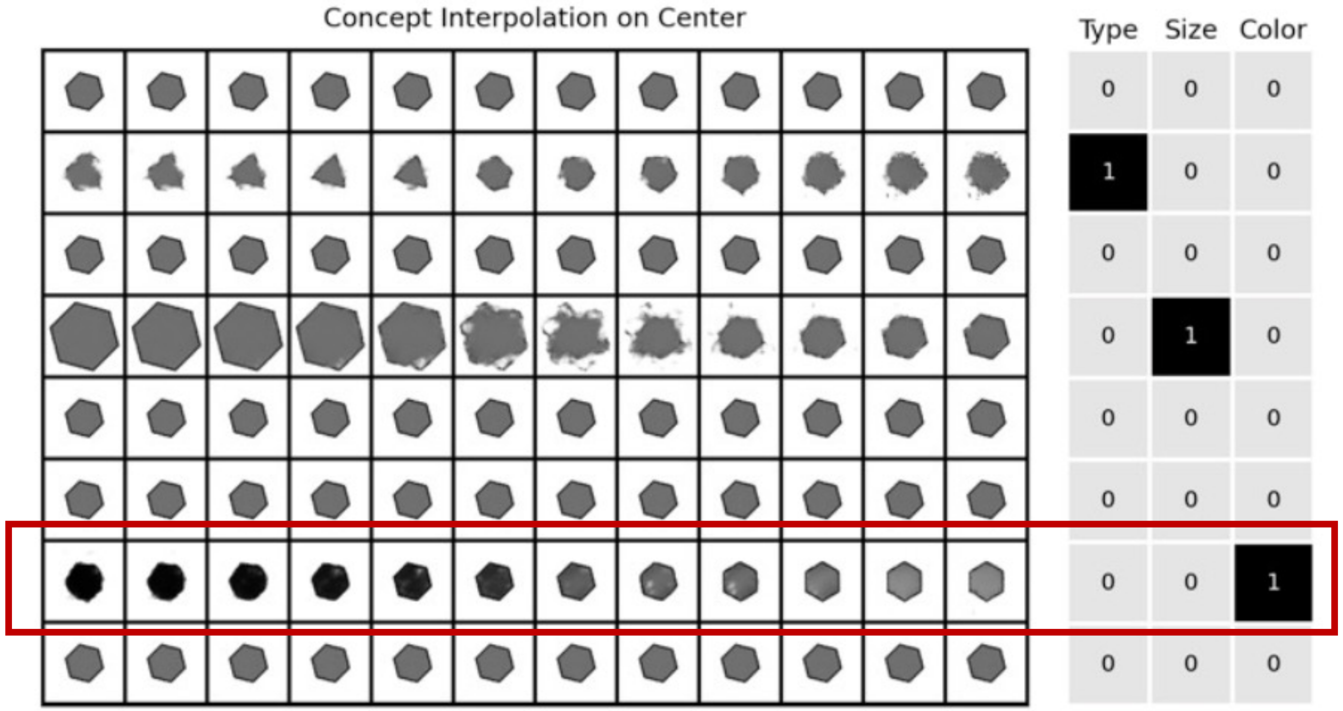
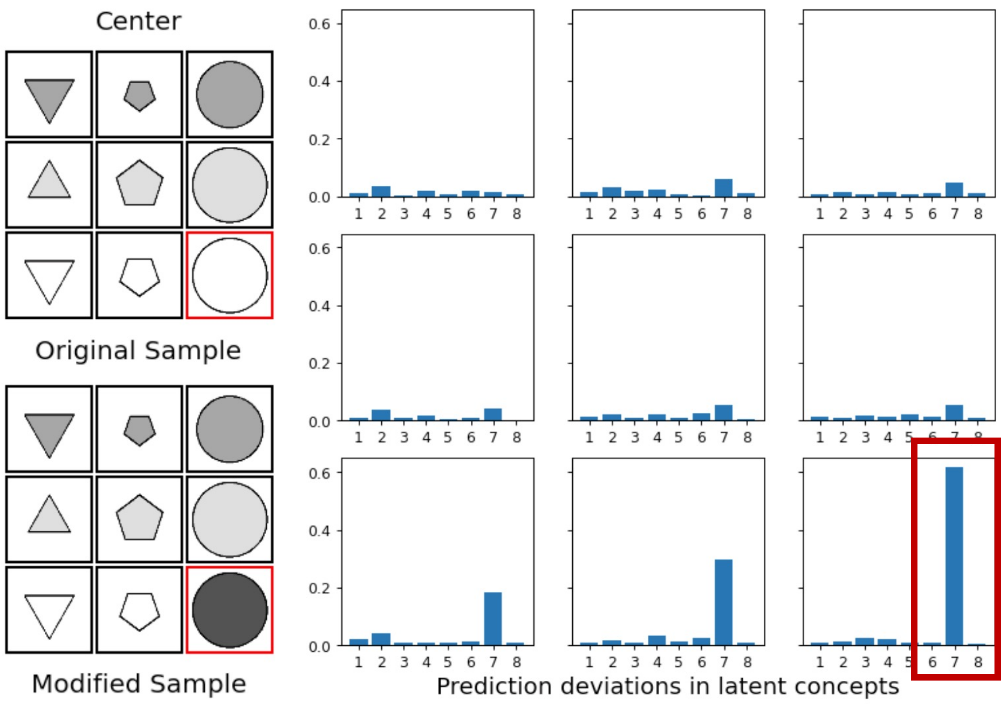


Figure 3: **Answer generation at arbitrary positions**. The prediction results on RAVEN are highlighted (red box) to illustrate the arbitrary-position generation ability. Due to the existence of noise, some predictions may differ from the original sample, but they still follow the correct rules.

Latent Concepts Visualization and Odd-One-Out Tests



(a) Interpolation results of latent concepts and the correspondence between the concepts and the real attributes



(b) An example of odd-one-out tests and the prediction errors of latent concepts

Out-Of-Distribution Configurations

Selection Accuracy on Out-Of-Distribution Configurations

	Type	Color	Size
Const	✓	✓	○
Prog	✓	✓	✓
Arith	✓	✓	✓
Dist3	✓	✓	✓

Center-Held-Out

	Type Out	Size Out	Type In	Size In	Color In
Const	✓	✓	✓	✓	✓
Prog	✓	✓	✓	✓	✓
Arith	✓	✓	○	○	○
Dist3	✓	✓	○	○	○

O-IC-Held-Out

legend

✓ training and validation rules

○ test rules

Rules

Const: Constant

Prog: Progression

Arith: Arithmetic

Dist3: Distribute Three

OOD Settings	RAISE	PrAE	ALANS	GCA-C	GCA-R	GCA-I	Transformer	ANP	LGPP	CLAP-NP
Center-Held-Out	99.2	**99.8**	46.9	35.0	14.4	12.1	12.1	10.6	8.6	19.5
O-IC-Held-Out	**56.1**	40.5	33.4	10.1	5.3	4.9	15.8	7.5	4.6	8.6

Limitations

Noise in data. The noise of object attributes in grids will influence the selection accuracy of generative solvers trained without distractors, e.g., RAISE and Transformer.

Models	O-IG-Uni	2×2Grid-Uni	3×3Grid-Uni
GCA-I	21.2/36.7	19.5/23.3	20.6/21.6
GCA-R	20.7/36.3	21.9/28.1	25.9/25.2
GCA-C	53.8/37.7	58.8/35.6	67.0/27.5
PrAE	29.1/45.1	85.4/85.6	26.8/47.2
ALANS	29.7/41.5	66.2/55.3	84.0/73.3
LGPP	3.4/12.3	4.1/13.0	4.0/13.1
ANP	31.5/34.0	10.0/15.6	12.0/16.3
CLAP	14.4/31.7	22.5/39.1	12.1/32.9
Transformer	70.6/57.9	73.3/73.0	34.2/37.0
RAISE	**95.8/99.0**	**87.6/97.9**	**95.3/93.2**

Configurations without object-level noise

O-IG	2×2Grid	3×3Grid
15.5/25.0	11.3/16.3	14.5/15.3
16.7/26.0	11.7/19.2	18.1/19.3
33.0/35.6	37.6/38.1	43.0/32.4
53.6/52.0	70.5/66.4	75.1/65.7
68.6/76.5	**82.0/84.5**	23.2/45.1
3.1/12.3	8.6/13.7	10.4/13.9
7.6/36.1	10.0/15.0	10.5/15.6
9.5/26.0	16.0/20.1	24.3/35.8
13.5/57.1	14.7/25.2	11.6/18.6
89.3/96.0	68.2/71.3	**77.7/78.7**

Configurations with full noise

use distractors

w/o distractors

Thanks for watching!

E-mail: fshi22@m.fudan.edu.cn {libin,xyxue}@fudan.edu.cn