

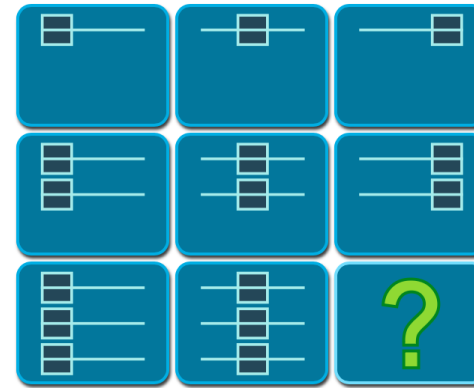
Haruo Hosoya (ATR / U-Tokyo)

書誌情報

- Mondal, S. S., Cohen, J. D., & Webb, T. W. (2024). Slot Abstractors: Toward Scalable Abstract Visual Reasoning. ICML.
- <http://arxiv.org/abs/2403.03458>

Abstract visual reasoning

- 少数の例から、抽象的なルールを推測し、一般化する課題
- 人間のintelligenceのテストベッド (IQテストなど)
- ニューラルネットは不得意



Which figure logically belongs
on the spot of the question mark?



Raven's Progressive Matrices

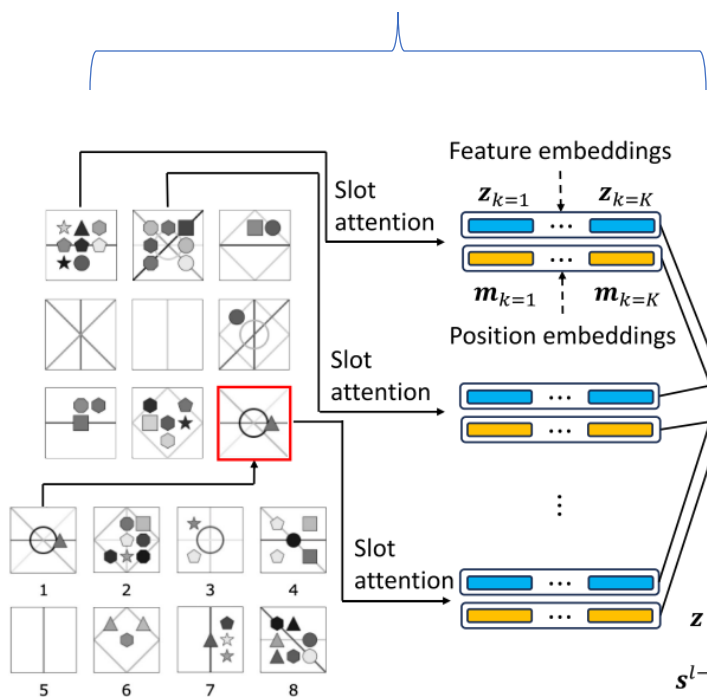
関係性ボトルネック [Webb+ 2023a]

- 抽象化を可能にする帰納バイアスを、アーキテクチャで実現
- 問題の画像（物体）特徴は抽出済みという前提で関係性ボトルネックを実現 [Webb+ 2020; 2021; Kerg+ 2022; Altabaa+ 2023]
- Object-Centric Relational Abstraction (OCRA) [Webb+ 2023b]
 - 物体中心表現（object-centric representation） [Greff+ 2019; Burgess+ 2019; etc] に関係性ボトルネックを導入
 - 計算量が高いという問題
- Abstractors [Altabaa+ 2023]
 - Transformerの拡張
 - relational cross-attention（後述）によって関係と物体特徴がdisentangleされる
- 本研究の提案Slot Abstractorsは、abstractorsと物体中心表現を統合

Slot Abstractors : アーキテクチャ

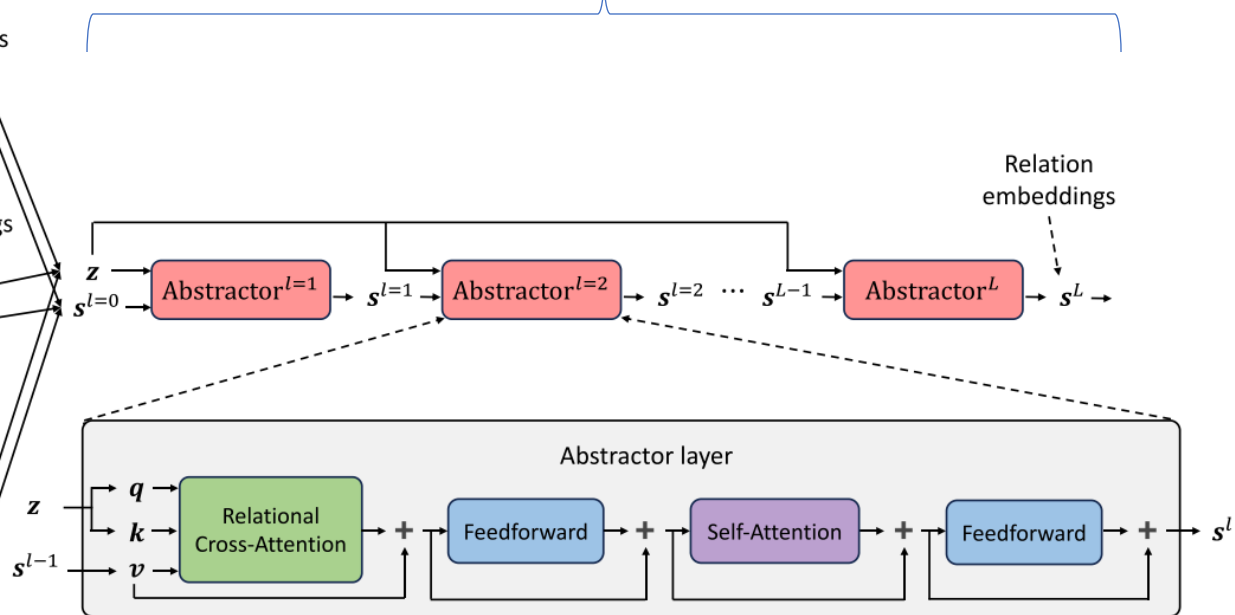
物体中心表現

[Locatello+ 2020; Webb+ 2023b]



物体表現から抽象化された関係性表現

[Altabaa+ 2023]



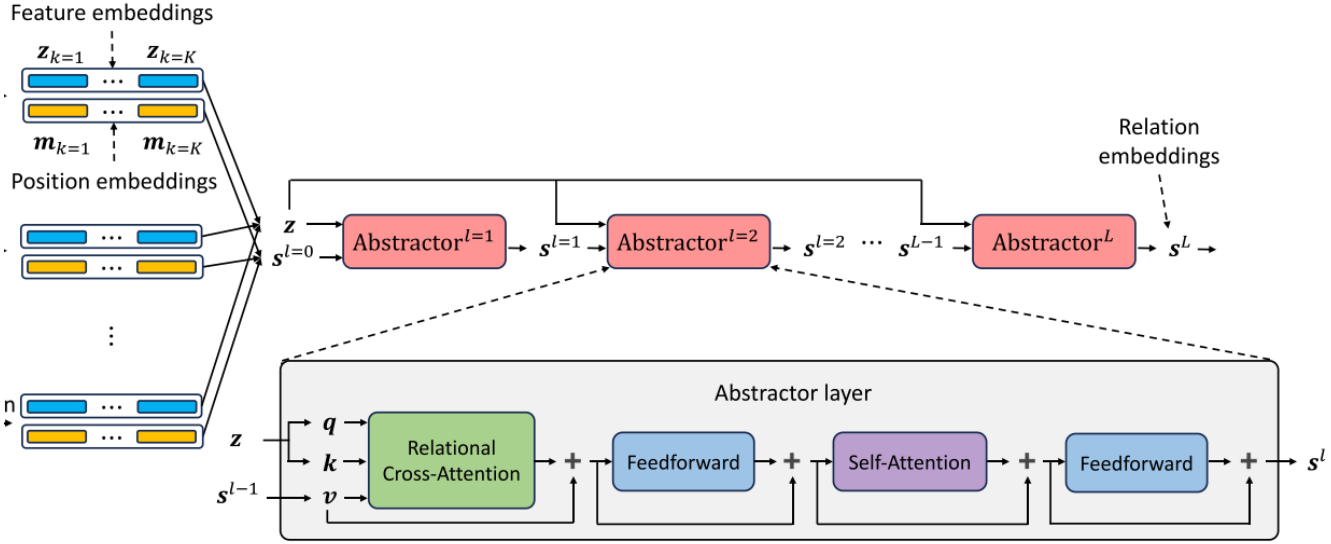
関係性表現の学習

- Multi-head relational cross-attention
 - 標準のcross-attentionとの違い
 - queryとkeyはfeature embeddingから、valueはposition embeddingから作成
 - downstreamがpositionに関する情報にのみアクセス→関係性ボトルネック

$$s^l = \text{RCA}(z, s^{l-1})$$

$$\text{RCA}(z, s^{l-1}) = \text{concat}(\tilde{s}^{h=1}, \dots, \tilde{s}^{h=H}) W_o$$

$$\tilde{s}^h = \text{softmax} \left(\frac{(z W_q^h)^T (z W_k^h)}{\sqrt{D}} \right) s^{l-1} W_v^h$$



実験：ART

- ARTデータセット [Webb+ 2021]

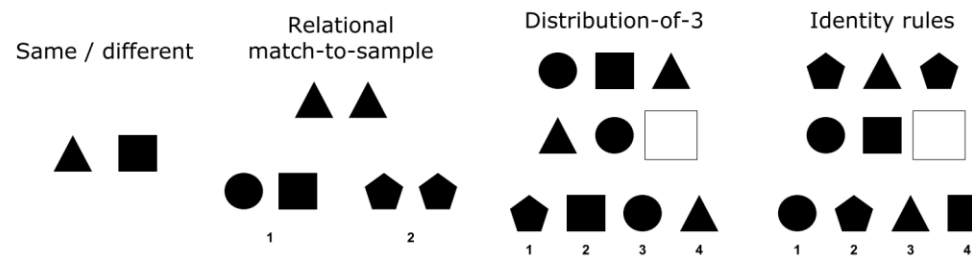


Table 1. Results on the four tasks of the ART dataset. Results reflect test accuracy averaged over 10 trained networks (\pm standard error).

MODEL	SAME/DIFFERENT	RELATIONAL-MATCH-TO-SAMPLE	DISTRIBUTION-OF-3	IDENTITY RULES
RESNET	66.60 \pm 1.5	49.89 \pm 0.2	50.07 \pm 1.3	54.84 \pm 2.4
SLOT-CORELNET	50.50 \pm 0.2	49.82 \pm 0.2	26.80 \pm 0.8	43.50 \pm 5.2
SLOT-ESBN	50.02 \pm 0.2	49.99 \pm 0.2	25.56 \pm 0.1	50.33 \pm 2.8
SLOT-GAMR	62.98 \pm 1.4	59.55 \pm 2.7	32.77 \pm 1.0	61.92 \pm 0.9
SLOT-RN	77.26 \pm 1.9	61.62 \pm 1.1	52.10 \pm 0.7	65.96 \pm 1.1
SLOT-IN	59.23 \pm 2.3	56.93 \pm 0.8	49.48 \pm 1.8	72.82 \pm 1.6
SLOT-TRANSFORMER	68.46 \pm 2.0	73.99 \pm 3.0	60.61 \pm 1.9	78.32 \pm 1.8
GAMR	83.49 \pm 1.4	72.20 \pm 3.0	68.62 \pm 1.8	66.23 \pm 4.8
OCRA	87.95 \pm 1.3	85.31 \pm 2.0	86.42 \pm 1.3	92.8 \pm 0.3
SLOT-ABTRACTOR	96.36\pm 0.4	91.64\pm 1.6	95.22\pm 0.4	96.41\pm0.4

実験：SVRT

- SVRTデータセット [Fleuret+ 2011]

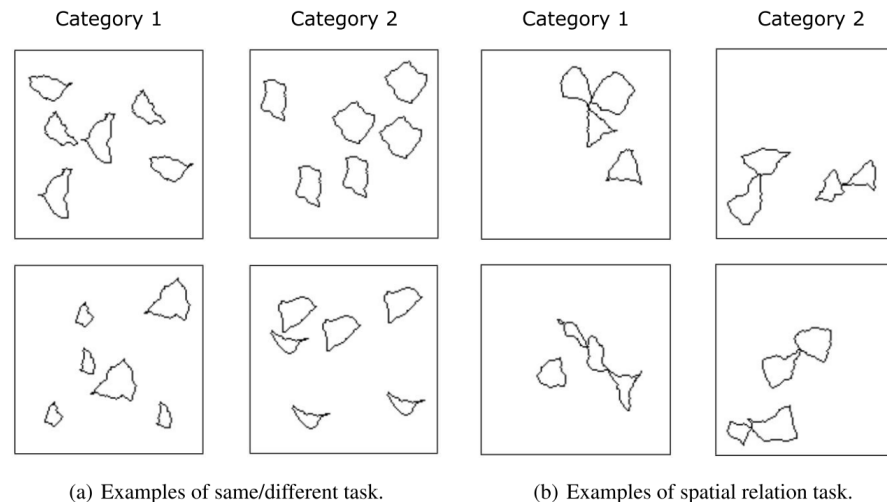


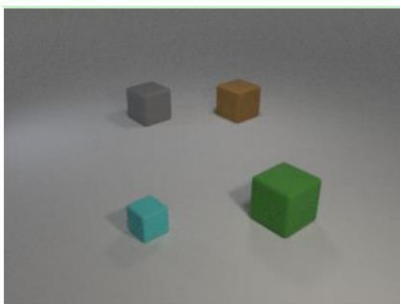
Table 2. Results on the two task categories of the SVRT dataset. Results reflect test accuracy averaged over different tasks from each category (\pm standard error), for 1 trained network for each task.

MODEL	SAME/DIFFERENT		SPATIAL RELATIONS	
	DATASET SIZE =0.5K	DATASET SIZE =1K	DATASET SIZE =0.5K	DATASET SIZE = 1K
RESNET	54.97 \pm 2.2	56.88 \pm 2.5	85.18 \pm 4.3	94.87 \pm 1.6
ATTN-RESNET	62.30 \pm 3.5	68.83 \pm 4.4	94.80 \pm 1.4	97.66 \pm 0.7
SLOT-CORELNET	52.95 \pm 1.4	57.13 \pm 2.6	60.95 \pm 3.7	74.59 \pm 3.7
SLOT-ESBN	53.83 \pm 1.1	51.67 \pm 1.1	61.30 \pm 2.3	62.69 \pm 2.4
SLOT-GAMR	63.06 \pm 3.7	66.87 \pm 3.2	84.90 \pm 2.4	86.99 \pm 2.2
SLOT-RN	71.48 \pm 4.8	81.79 \pm 4.4	91.73 \pm 1.8	96.20 \pm 1.4
SLOT-IN	68.23 \pm 4.8	74.99 \pm 4.9	90.23 \pm 2.0	94.86 \pm 1.4
SLOT-TRANSFORMER	76.54\pm5.1	89.85\pm4.2	94.06 \pm 1.6	97.86\pm0.9
GAMR	76.80\pm4.9	82.05 \pm 4.4	97.40\pm0.7	98.74\pm0.3
OCRA	79.89\pm4.5	90.30\pm4.1	89.25 \pm 2.5	95.02 \pm 2.4
SLOT-ABTRACTOR	82.20\pm4.7	91.86\pm4.0	91.74 \pm 2.2	97.26 \pm 1.1

実験：CLEVR-ART

- CLEVR-ART [Webb+ 2023b]

Relational-match-to-sample



Identity rules

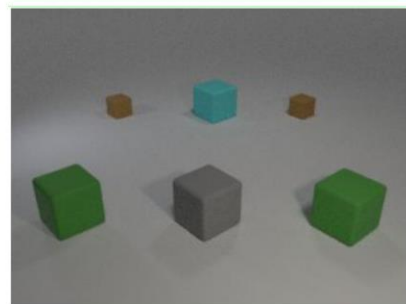


Table 3. Results on the two tasks (relational-match-to-sample (RMTS) and identity rules (ID)) of the CLEVR-ART dataset. Results reflect test accuracy averaged over 5 trained networks (\pm standard error).

MODEL	RMTS	ID
SLOT-CORELNET	49.87 \pm 0.2	24.80 \pm 0.3
SLOT-ESBN	62.53 \pm 0.1	28.87 \pm 0.7
SLOT-GAMR	52.56 \pm 0.5	39.83 \pm 0.9
SLOT-RN	64.79 \pm 0.5	60.27 \pm 0.6
SLOT-IN	66.72 \pm 3.7	67.22 \pm 1.7
SLOT-TRANSFORMER	87.54 \pm 0.7	78.81 \pm 1.6
GAMR	70.40 \pm 5.8	74.15 \pm 4.0
OCRA	93.34 \pm 1.0	77.06 \pm 0.7
SLOT-ABTRACTOR	96.34\pm0.5	91.61\pm0.2

実験：PGM

- PGMデータセット [Barrett+ 2018]

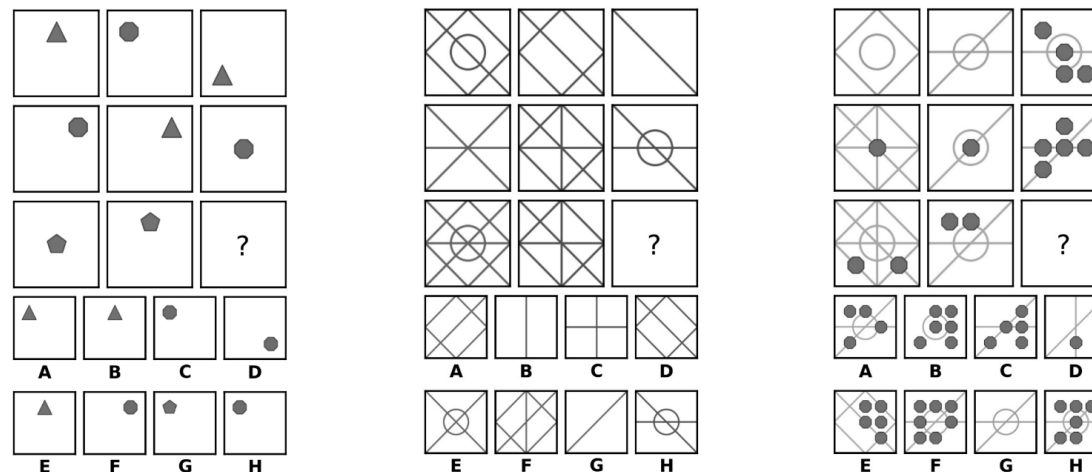


Table 4. Results on different generalization regimes of the PGM dataset. Results reflect test accuracy for 1 trained model for each regime.

MODEL	NEUTRAL	INTP.	H.O.A.P.	H.O.T.P.	H.O.TRIPLES	H.O.L-T	H.O.S-C	EXTP.
WREN $\beta = 0$	62.6	64.4	27.2	41.9	19.0	14.4	12.5	17.2
VAE-WREN	64.2	-	36.8	43.6	24.6	-	-	-
MXGNET $\beta = 0$	66.7	65.4	33.6	43.3	19.9	16.7	16.6	18.9
DCNET	68.6	59.7	-	-	-	-	-	17.8
REL-BASE	85.5	-	-	-	-	-	-	22.1
MRNET	93.4	68.1	38.4	55.3	25.9	30.1	16.9	19.2
MLRN	98.0	57.8	-	-	-	-	-	14.9
STSN	98.2	78.5	-	-	-	-	-	20.4
ARII	88.0	72.0	50.0	64.1	32.1	16.0	12.7	29.0
PREDRNET	97.4	70.5	63.4	67.8	23.4	27.3	13.1	19.7
SLOT-ABTRACTOR	91.5	91.6	63.3	78.3	20.4	16.7	14.3	39.3

まとめ

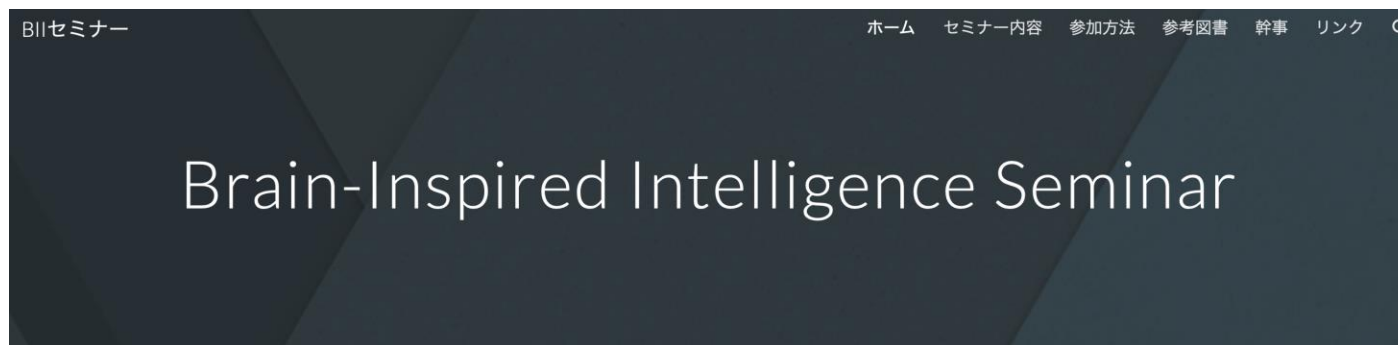
- 関係性ボトルネックを導入したアーキテクチャSlot Abstractorsを提案
 - 物体中心表現とAbstractorsの統合
- 様々なAbstract visual reasoningのタスクで高成績

将来展望

- 抽象化や関係性の学習は、現在主流の人工知能は不得意
- 少数例から抽象的な規則を発見して一般化するのは、人間ならではの能力
 - アブダクション（仮説の思いつき）
- 脳科学では、海馬や前頭前野が関係しているかも、という知見も
- 汎用人工知能への糸口？

Brain-Inspired Intelligence (BII) セミナー

- 脳とAIの接点における最新論文の輪読会
- Weblab Slackチャンネル：#brain-inspired-intelligence
- HP: <https://sites.google.com/view/bii-seminar>
- 月1度 水曜 16:00~



概要

深層学習はもともと脳から学んで発展してきましたが、今やコンピュータビジョン・ロボティクス・大規模言語モデルなど、幅広い分野で活用されるまでに進化を遂げ、AI研究は最盛期を迎えています。しかし、脳にはまだまだ未知の可能性が眠っているはずで

Brain-Inspired Intelligence (BII) セミナーは、そんな脳の潜在的なアイデアを発掘し、さらに未来的なAIの扉を開くことを目的に、2023年秋に発足しました。月に一度、神経科学から学んだAIモデル、AIを使った計算神経科学、新しいAIのヒントになりそうな神経科学など、BII関連の最新論文を輪読形式で議論しています。

機械学習に基づくAIの基礎技術を持つ研究者や大学院生の参加を募集中です。セミナーでの議論を通じて脳の潜在力を解き放ち、AI研究の新たな地平を切り拓きましょう！