Acquiring Complex Concepts with Comparative Learning

Paper #10

D. Calanzone, DISI, University of Trento
F. Merlo, CIMEC, University of Trento

UNIVERSITÀ DI TRENTO
CIMeC Center for Mind/Brain Sciences

Motivation

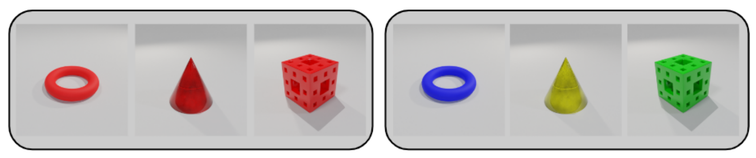
- VLMs present **sparse concepts and weak compositional relations** [5, 6].
- **Comparison** allows humans to attend to relevant features in inputs, to highlight differences and thus to learn abstractions.
- **RQ1:** *can we teach logical concepts through comparison*?
- **RQ2:** *can we find an efficient multi-task architecture for this task*?

Experimental setup

- We base on **S**imulated **O**bjects for **L**anguage **A**quisition (**SOLA**) [1].
- As in [1], we teach concepts by grouping objects that express them (**similarity batch**) and objects that do not (**dissimilarity batch**).

$$\text{REP}_{l_i} = \text{SIM}_{l_i}(\{a^{l_i} \in \mathcal{B}_s\})$$
$$\text{REP}_{l_i} = \text{DIFF}_{l_i}(a^{l_i}, \{b^{l_j} \in \mathcal{B}_d\})$$

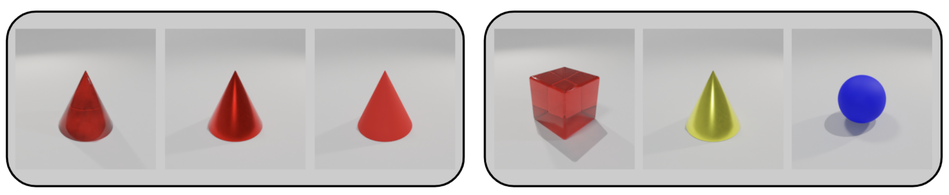
RED



- Similarly, we construct such batches with respectively the worlds satisfying or falsifying a logical formula.

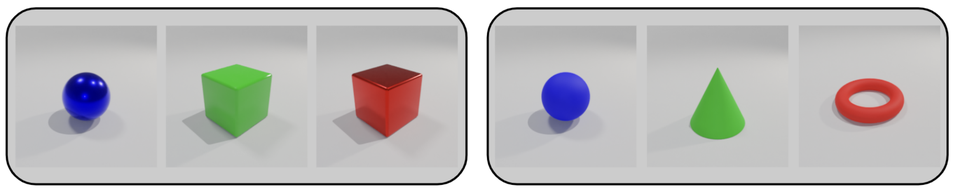
AND

$$\mathcal{B}_s = \{a \,|\, \text{red AND cone}\}$$
$$\mathcal{B}_d = \{a \,|\, \text{red AND NOT cone} \oplus \text{NOT red AND cone} \oplus \text{NOT red AND NOT cone}\}$$



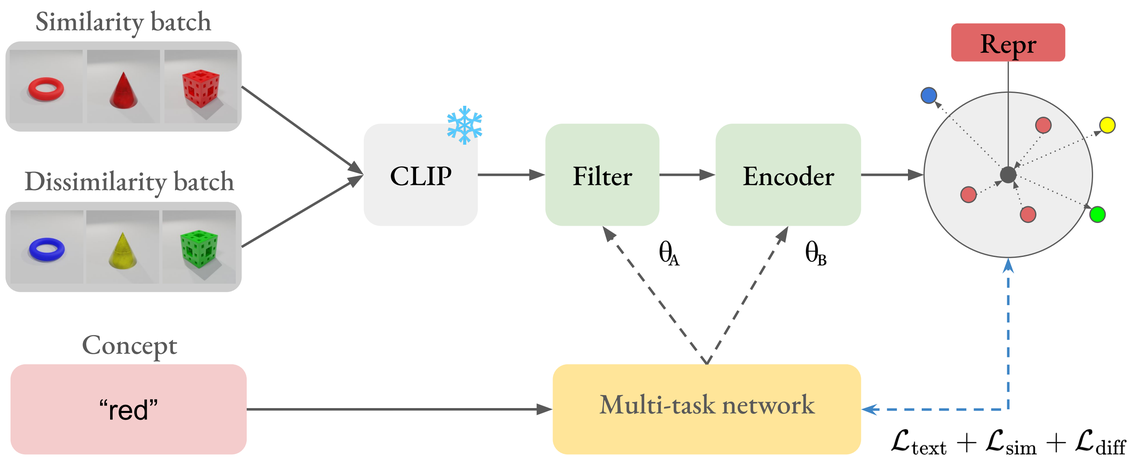
OR

$$\mathcal{B}_s = \{a \,|\, \text{metallic AND NOT cube} \oplus \text{NOT metallic AND cube} \oplus \text{metallic AND cube}\}$$
$$\mathcal{B}_d = \{a \,|\, \text{NOT metallic AND NOT cube}\}$$



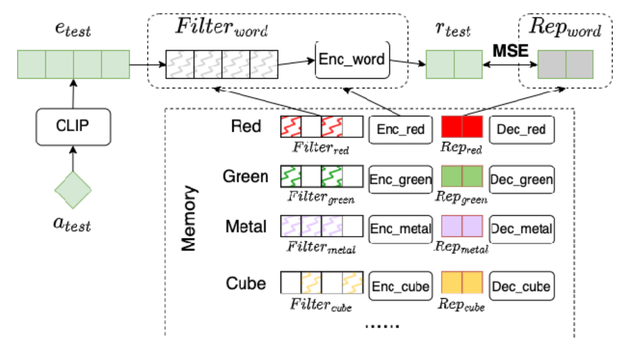
Proposed methodology

- We adopt for **RQ1** concept-specific networks as in [1].
- Fine-grained Contrastive Learning with additional attention maps.
- For **RQ2**, we introduce multi-task adapters for CLIP modulated by the input text: a hyper-network [4] and modular skill sharing [3], in comparison.



Results

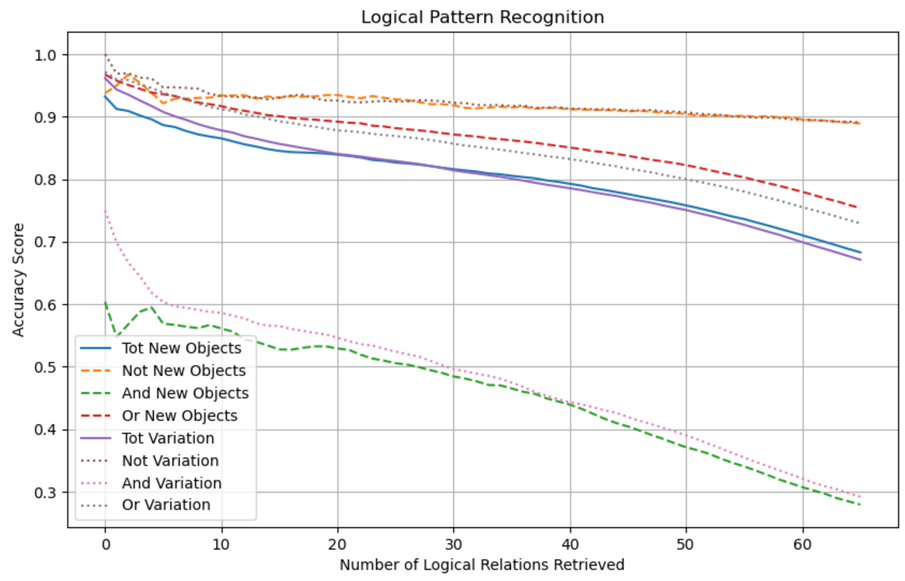
- For **primitive concepts**, we test our networks on Multi-Attribute Recognition (MAR) [1].
- For **complex logical expressions**, we modify MAR and thus define Logical Pattern Recognition (LPR).



Purple Plastic Torus

1. Aqua or Torus,
2. Red or Torus,
3. Purple or Glass,
4. Brown or Torus,
5. Purple and Plastic,
6. Plastic and Torus,
7. Not Plastic,
8. Purple or Gear,
9. Purple or Rubber,
10. Purple and Torus

- **RQ1:** Comparative learning enables fine-grained understanding to correctly retrieve visual examples describing a logical formula.



- **RQ2:** Modular skill sharing (**MSS**) [3] is comparable in performance to **hyper-networks**, but much more parameter efficient ($\sim$ 3x).
- Learning in sequence requires experience replay (DER++ [2]).
- Materials have subtle visual features and thus are harder to learn.

Split	Model	Color	Material	Shape
D_{test_nc}	Concept-specific	0.96	0.48	0.98
	HyperNet	0.37	0.25	0.73
	HyperNet (DER++)	0.71	0.28	0.89
	Modular Skills Sharing	0.72	0.21	0.67

Conclusions & Further Work

- **RQ1:** concept-specific networks can learn worlds satisfying a logical rules with CL. *Next: testing noisy environments.*
- **RQ2:** multi-task adapters well adapt CLIP for the task [4]. Reducing the skills could nudge learning more generic patterns. *Next: extending to logical pattern recognition.*

KEY REFERENCES

[1] Human Inspired Progressive Alignment and Comparative Learning for Grounded Word Acquisition. Bao et al. 2023.
[2] Dark Experience for General Continual Learning: a Strong, Simple Baseline. Buzzega et al. 2020.
[3] Combining Modular Skills in Multitask Learning. Ponti et al. 2022.
[4] Parameter-efficient Multi-task Fine-tuning for Transformers via Shared Hypernetworks. Mahabadi et al. 2021.
[5] Do Vision-Language Pretrained Models Learn Composable Primitive Concepts? Yun et al. 2022.
[6] When and why vision-language models behave like bags-of-words, and what to do about it? Yuksekgonul et al. 2022.

MORE INFORMATION

Diego Calanzone
University of Trento
DISI
diego.calanzone@studenti.unitn.it
https://halixness.github.io