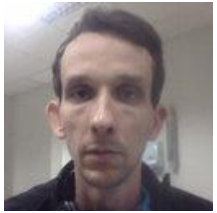
Commonsense Reasoning and Knowledge Acquisition to Guide Deep Learning on Robots

Tiago Mota
University of Auckland
New Zealand



Mohan Sridharan
University of Birmingham
UK

Problem

- Deep learning require lots of labeled data.
- Hard to understand the internal decisions.

Approach

- Do not train a deep net in an end-to-end fashion.
- Given an incomplete knowledge, learn new axioms with fewer samples.
- Try to predict the answer with the extended knowledge base.
- If that fails, use the deep net.

Result

- Increment in the accuracy
- Needs less labeled data
- Better decision making

Architecture

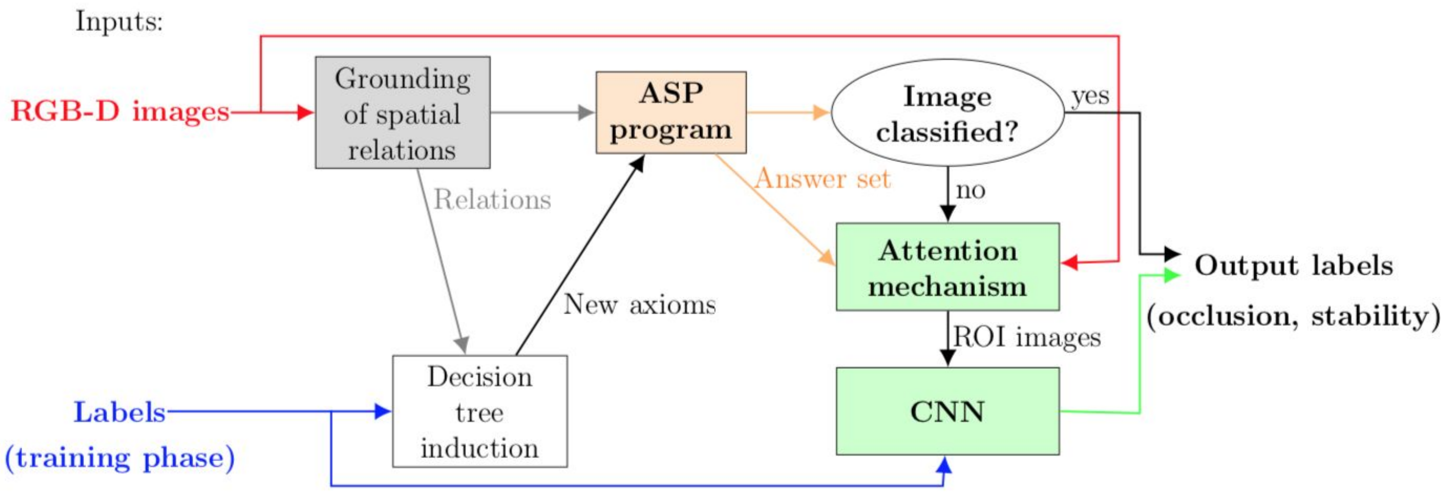


Fig. 2: Architecture combines the complementary strengths of non-monotonic logical reasoning, deep learning, and decision tree induction, to perform the scene understanding tasks reliably and efficiently.

Answer Set Prolog (ASP)

- They use non-monotonic logic.

Monotonic logic: adding a new premise to a theorem does not change the conclusion.

If (A->B) then (A,C->B)

Non-monotonic logic: we can update our beliefs based on new observations.

A->B : If A is true, B is probably true.

Answer Set Prolog (ASP)

- Sorts: can be thought as types.
  - object, robot, size, relation, surface, step
- Statics: like global variables.
  - obj_size(object, 5), obj_surface(object, slippery)
- Fluents: relations
  - obj_relation(above, A, B), in_hand(robot, object)

Answer Set Prolog (ASP)

$$holds(in\_hand(robot, object), I + 1) \leftarrow \qquad (2a)$$
$$occurs(pickup(robot, object), I)$$

$$holds(obj\_relation(above, A, B), I) \leftarrow \qquad (2b)$$
$$holds(obj\_relation(below, B, A), I)$$

$$holds(obj\_relation(infront, A, B), I) \leftarrow \qquad (2c)$$
$$holds(obj\_relation(behind, B, A), I)$$

$$\neg occurs(pickup(robot, object), I) \leftarrow \qquad (2d)$$
$$holds(in\_hand(robot, object), I)$$

Decision Tree

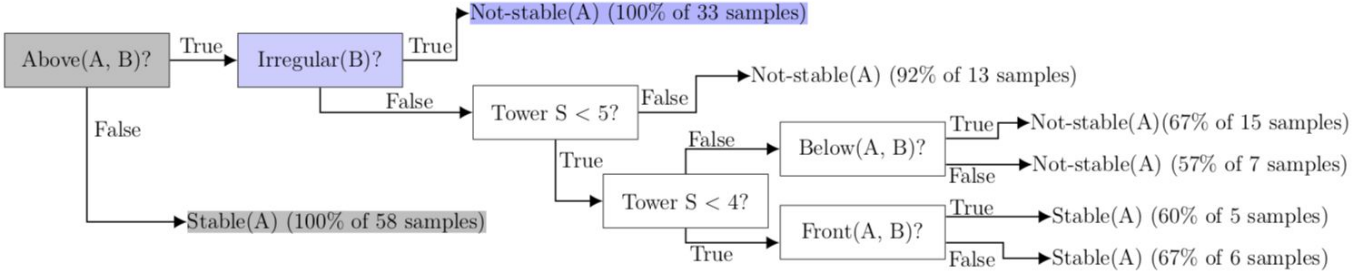


Fig. 3: Example of a decision tree constructed for stability estimation using some labeled examples. Highlighted branches are used to construct previously unknown axioms.

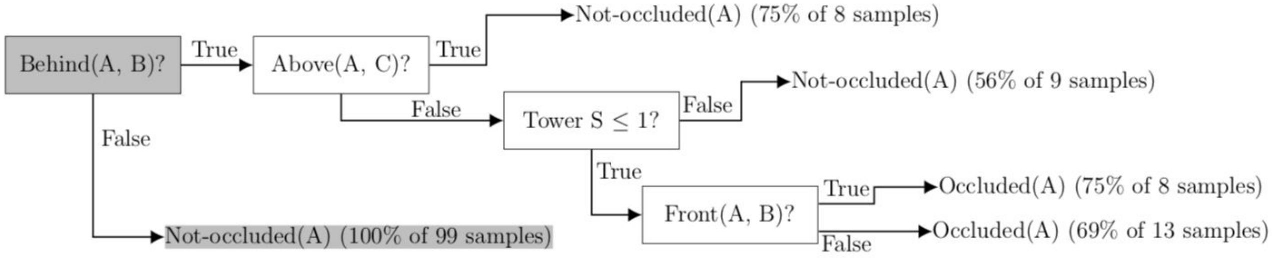


Fig. 4: Example of a decision tree constructed for occlusion estimation using some labeled examples. Highlighted branch is used to construct previously unknown axiom.

Reasoning with commonsense domain knowledge and the attention mechanism:

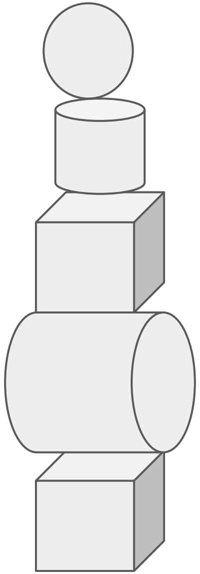
Hypothesis 1: improves the accuracy of deep networks.

Hypothesis 2: reduces sample complexity and time complexity of training deep networks.

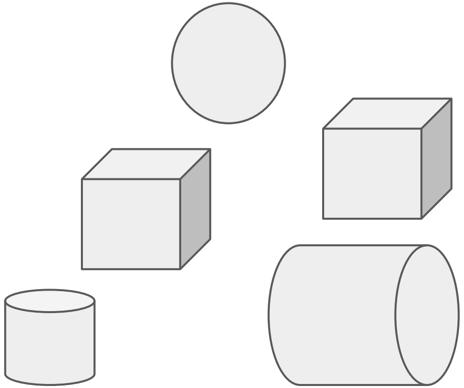
Hypothesis 3: able to learn new axioms that will improve decision making.

Experimental setup

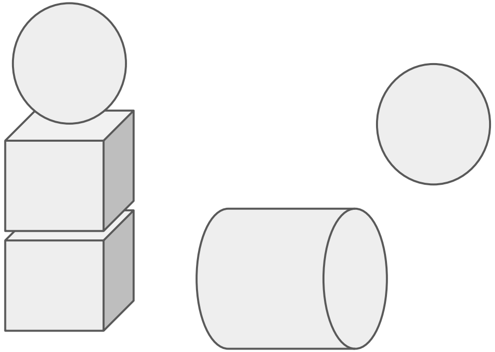
6000 labeled images. Each image contain five objects. Occlusion and stability prediction.



Towers



Spread



Intersection

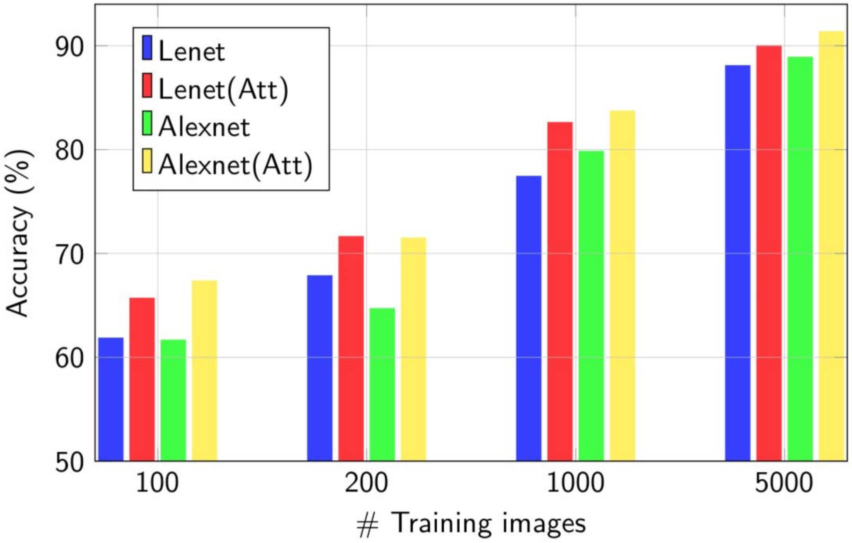
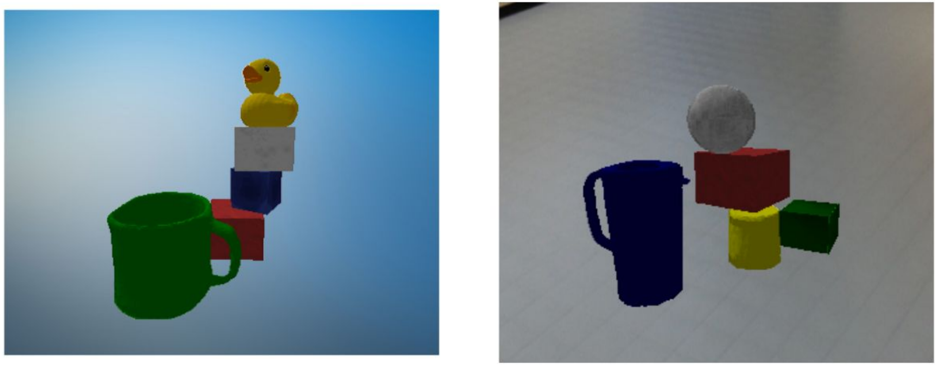


Fig. 6: Accuracy of Lenet and Alexnet with and without commonsense reasoning and the attention mechanism. The number of background images were fixed at 100. Our architecture improves accuracy in comparison with the baselines.

Supports H1.



(a) (b)

Fig. 7: Examples of test images for Lenet and Lenet(Att): (a) both detected the occlusion of the red cube by the green mug, but only the latter correctly estimated the tower's instability; and (b) both predicted the instability of the tower, but only Lenet(Att) detected the obstruction of the green cube by the yellow cylinder.

Supports H1.

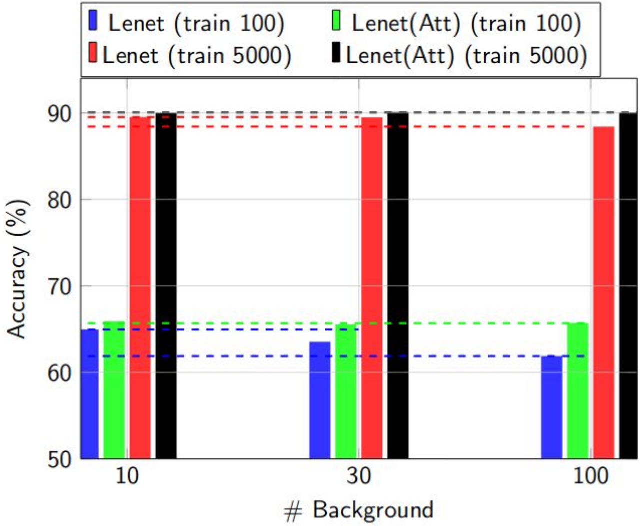


Fig. 8: Effect of changing the number of backgrounds on the accuracy of the Lenet and Lenet(Att) networks for 100 and 5000 training images. Without the attention mechanism and commonsense reasoning, increasing the number of backgrounds reduces the classification accuracy.

Supports H2.

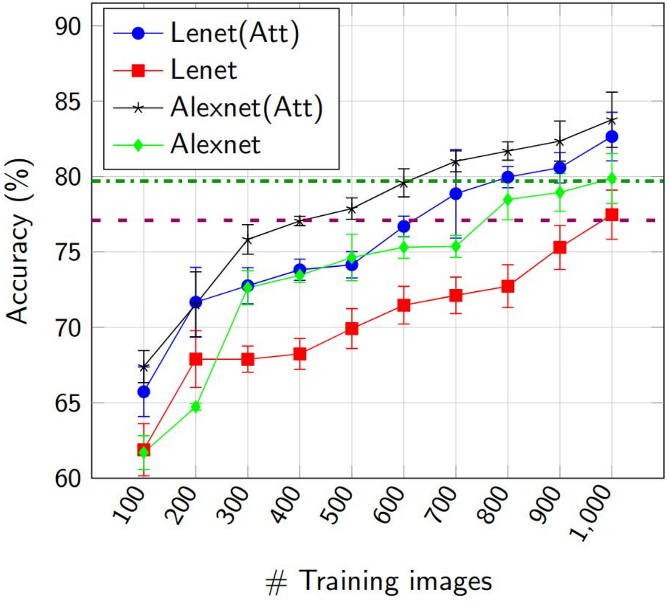


Fig. 9: Accuracy of Lenet and Alexnet with and without the attention mechanism and commonsense reasoning. The number of background images was fixed at 100. Any desired accuracy is achieved with a smaller training set.

Supports H2.

Axiom type	Precision	Recall
Unknown (normal)	98%	100%
Unknown (default)	78%	62%

TABLE I: Precision and recall for previously unknown axioms (normal, default) using decision tree induction.

Unknown normal - leafs with prediction accuracy of %95.

$$stable(A) \leftarrow \neg obj\_relation(above, A, B) \tag{4a}$$

$$\neg occluded(A) \leftarrow \neg obj\_relation(behind, A, B) \tag{4b}$$

$$\begin{aligned} \neg stable(A) \leftarrow\ & obj\_relation(above, A, B), \\ & obj\_surface(B, irregular) \end{aligned} \tag{5}$$

Unknown default - leafs with prediction accuracy of %70.

$$\begin{aligned} \neg holds(stable(A), I) \leftarrow\ & holds(obj\_relation(above, A, B), I), \\ & size(A, large),\ size(B, small), \\ & not\ holds(stable(A), I) \end{aligned} \tag{3}$$

Supports H3.

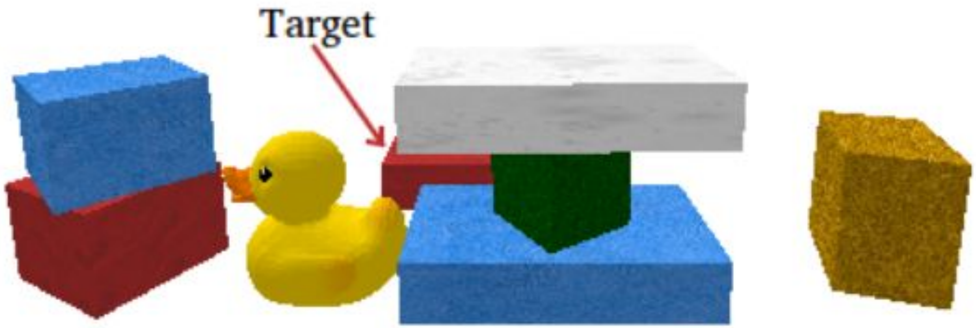


Fig. 10: Illustrative image used for planning experiments with and without the learned axioms.