Project Management: Case Study (A3) NSCL (Neuro-Symbolic Concept Learner)

Dheeraj Chaudhary (17BCS009)

***Aim:** To develop a model that learns visual concepts, words, and semantic parsing of sentences without explicit supervision on any of them; instead, our model learns by simply looking at images and reading paired questions and answers.*

1. Define the purpose and scope of a project

A. Objective

Humans are capable of learning visual concepts by jointly understanding vision and language. Consider the example shown in Figure 1. Image someone with no prior knowledge of colors is presented with red and green cubes, paired with questions and answers. They can easily identify differences in an object's appearance and align it to corresponding words in the questions and answers. Other object attributes can be learned in similar fashion. Humans can inductively learn the correspondence between visual concepts and word semantics (Figure 2) and unravel compositional logic from complex questions assisted by the learned visual concepts (Figure 3).

Motivated by this, we want to develop NS-CL (neuro-symbolic concept learner), which jointly learns visual perception, words, and semantic language parsing from images and questions-answer pairs. NS-CL has three modules: a neural-based perception module that extracts object-level representations from the scene, a visually-grounded semantic parser for translating questions into executable programs, and a symbolic program executor that reads out the perceptual representation of objects, classifies their attributes/relations and executes the program to obtain an answer.

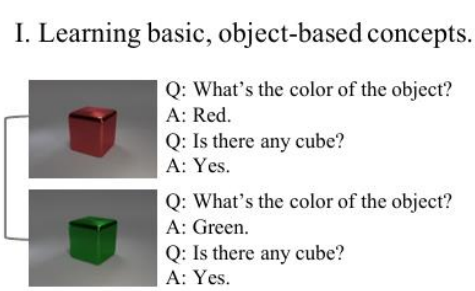


Figure 1

I. Learning visual concepts start from looking at simple scenes, reading simple questions, and reasoning over contrastive examples.

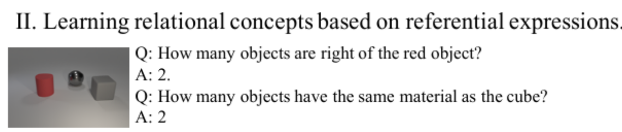


Figure 2

II. Afterward, we can interpret referential expressions based on the learned object-based concepts, and learn relational concepts.

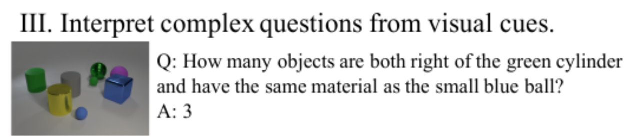


Figure 3

III. Finally, we can interpret complex questions from visual cues by exploiting the compositional structure.

B. Scope

We first use a visual perception module to construct an object-based representation for a scene and run a semantic parsing module to translate a question into an executable program. We then apply a quasi-symbolic program executor to infer the answer based on the scene representation. We use paired images, questions, and answers to jointly train the visual and language modules. As shown in Figure 4, given an input image, the visual perception module detects objects in the scene and extracts a deep, latent representation for each of them. The semantic parsing module translates an input question in natural language into an executable program given a domain-specific language (DSL). The generated programs have a hierarchical structure of symbolic, functional modules, each fulfilling a specific operation over the scene representation. The explicit program semantics

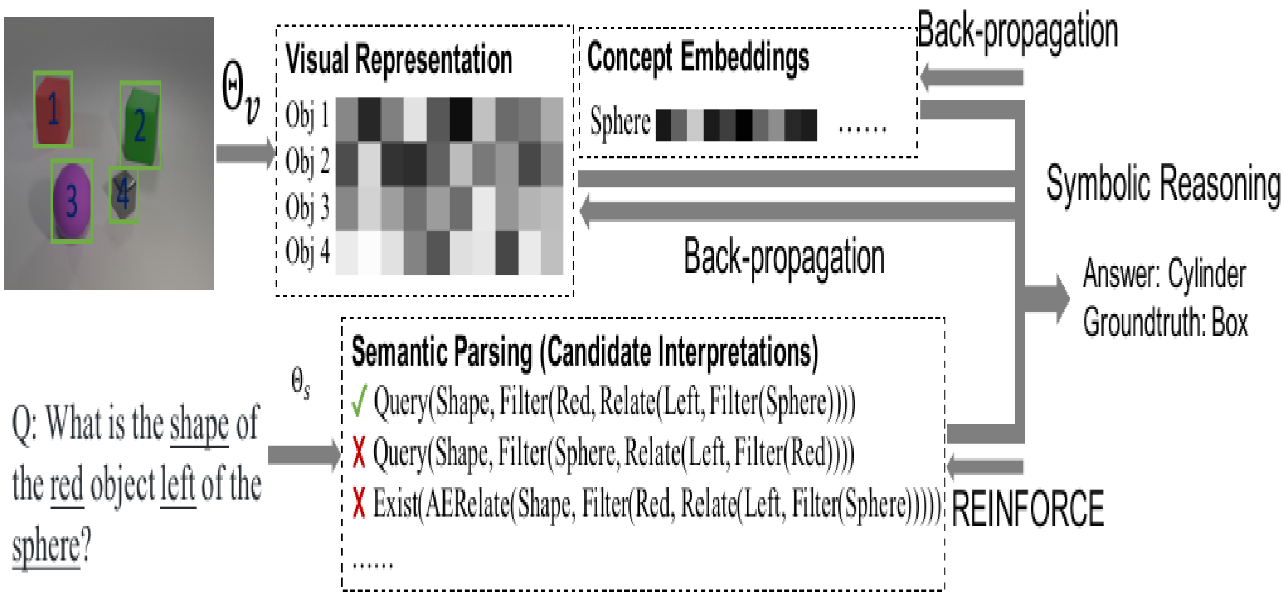


Figure 4

enjoys compositionality, interpretability, and generalizability. The program executor executes the program upon the derived scene representation and answers the question. Our program executor works in a symbolic and deterministic manner. This feature ensures a transparent execution trace of the program. Our program executor has a fully differentiable design w.r.t. the visual representations and the concept representations, which supports gradient-based optimization during training.

Different Models that will be used for the purpose are Visual perception, Concept quantization, DSL and semantic parsing, and Quasi-symbolic program execution.

We will train our model on the CLEVR dataset. We will be training the model on 5K images (<10% of CLEVR's 70K training images). We generate 20 questions for each image for the entire curriculum learning process. After validating our model, we will go for generalization. Once we get good accuracy in generalization, we will be ready to test for different images. After intensive testing on different images if we get a good result then we will extend our application domain to videos.

2. Generation and screening of Project idea

2.1 Generation of ideas and SWOT Analysis

The main idea is to develop an NSCL (Neuro-Symbolic Concept Learner) which can be used at various places for relational and non-relational reasoning. So far in the market, there is no such software available that can do relational reasoning. Development cost may be a little high as a high-end computer is required to build this project. The project is very innovative so cost is not much concern and investment in the a high-end pc will be beneficial in the future as well to develop other great projects.

SWOT Analysis :

	Positive	Negative
Internal	Strengths : 1. Team Work 2. Experienced in project domain 3. Hard Working Team member 4. Constant Efforts 5. Proper guidance 6. Good leadership skill	Weakness : 1. Resource limitation 2. Fewer researches on project domain. 3. New Dataset availability.
External	Opportunities : 1. First-ever relational reasoning software 2. Get a Deeper understanding of DL. 3. Reducing manual task for visual reasoning	Threats : 1. Hidden researches on the same project. 2. Unable to adopt new tech.

Table 1

2.2 Monitoring of environment

The research and development of these models lead to a lot of carbon emissions. In a study last year, researchers at the University of Massachusetts at Amherst estimated that training a large deep-learning model produces 626,000 pounds of planet-warming carbon dioxide, equal to the lifetime emissions of five cars. As models grow bigger, their demand for computing is outpacing improvements in hardware efficiency.

2.3 Corporate appraisal

The rapid improvement in computation power has led people to look for more human-like behaviors using deep learning. The reasoning is what differentiates a human from an AI.
So the reasoning power of an AI is on the rise with the increase in computation power. Microsoft, Google, and Facebook are pioneering the research in this field wild deep learning models having the reasoning ability on images and videos.

2.4 Porter model description on your project Idea

Competitive rivalry
There are different research groups within big companies that keep on working and upgrading the models. Every day new research is published and the technology is growing rapidly manyfold

Supplier power
Companies that have more budget for Research and Development and computational power hold an advantage over smaller research groups or independent researchers.

Buyer power
Buyer power is mostly the MNCs and other companies that want to use this technology in their product. The market is insensitive to price as most of these are not patented and can be used by companies for the development of their products

Threat of substitution
The threat of substitution is only posed by a new model or technology or algorithm which outperforms the current ones. They come after a long time after a significant amount of research. Every model has its strengths and weaknesses so utilized based on purpose hence threat of substitution depends on the number of weaknesses your model possesses.

The threat of new entry
The threat of new entry is very less as it will have to start on some completely fresh new fundamentals and will take time to reach the level of accuracy and development we have reached with the current systems.

2.5 Project Rating Index

The project Rating index is shown in table 2.

3. Project Analysis for Capital Budgeting

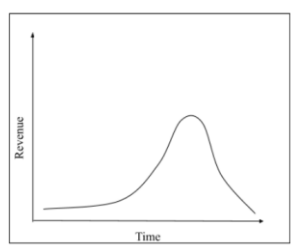
Discount	10%			
Year	0	1	2	3
Estimated Capital Investment	Rs. 5 Cr		Rs. 0	Rs. 0
Cash Flows	Rs. 0	Rs. 1 Cr	Rs. 10Cr	Rs. 1000Cr
The total cash flow of the year	Rs. 5 Cr	Rs. 1 Cr	Rs. 10 Cr	Rs. 1000 Cr
Net Present Value	Rs. 686 Cr approx			

Table 3

4. Risk Analysis

Risk analysis is shown in table 4.

5. Detailed Cost-benefit analysis



As depicted by the curve above, our product is more research-oriented and requires a significant amount of resources to get a head start. The revenue curve increases gradually over time as the domain which we have adopted is a very advanced tech and will require a considerate amount of time people could digest. As soon as people start understanding the utility of the product, our revenue will take huge leaps. But as time goes by, every product has a downfall as they get competition in the market which in turn affects our revenue.

No.	Factor	Factor Weight	Rating					Factor score
			Very Good (5)	Good (4)	Average (3)	Poor (2)	Very Poor (1)	
1.	Understanding of project requirements	0.2		Y				0.8
2.	Input availability	0.1		Y				0.4
3.	Difficulty	0.05	Y					0.25
4.	Technical know-how	0.2		Y				0.8
5.	Stability	0.1		Y				0.4
6.	Success	0.05		Y				0.2
7.	Performance w.r.t cost	0.05			Y			0.15
8.	Effective communication	0.1	Y					0.5
9.	Performance w.r.t product quality	0.05			Y			0.15
10.	Risk	0.1		Y				0.4
	Total	1						**4.05**

Table 2 (Project Rating Index)

ID	Risk Description	Likelihood of the risk occurring	Impact if the risk occurs	Owner	Mitigation Action	Status
1	Project Purpose and need is not well-defined	Medium	High	Mentor/Sponsor	Analyze a pre-existing model and ensure that the purpose is well defined	Closed
2	Project design and deliverable definition is incomplete.	Low	High	Mentor/Sponsor	Define the scope in the detail via design workshop with input from subject matter experts.	Closed
3	Estimating or scheduling errors	Medium	High	Group Members	Track schedule daily and include schedule review as an agenda item in every project team meeting. Flag forecast errors and/or delays early.	Closed

Table 4 (Risk Analysis)

6. Project Plan Schedule

6.1 Gantt Chart for Project Development

6.2 Major Milestones

The Major Milestones of our project are as follows:

1. Project Background study: Detail study of project background through reading various research papers. Study different case study and implement research. After implementing different types of researches, discuss High-Level Design, and start the project.

2. Project Execution: Project is divided into different modules and implemented separately. Later all the modules are integrated and the first version of the product is released.

3. Project Improvisation: After releasing the first version bug testing and fixing is done to improvise the project. Accuracy of the project is monitored and once the desired accuracy is reached final product will be released.

6.3 Critical Path Method Analysis to estimate the time frame for the project

Activity	Description	Predecessor Activity	Duration (days)
A	Project Start		
B	Background Research	A	7
C	Research Implementation	B	10
D	High-Level Design	B,C	5
E	Data collection	C	10
F	Prepare data	E	10
G	Choose model	D, F	10
H	Train Model	G	11
I	Evaluate Model	H	15
J	Hyperparameter Tuning	I	10
K	Prediction	J	5
L	Demonstration	K	2
M	Project Finish	L	0

7. Conclusion

This project aimed to provide the AI models the part of reasoning which would bring them closer to human-like intelligence. The reasoning part over images and videos is done by the model also known as neuro symbolic reasoning. Currently, we use the dataset of CLEVR (for images) and CLEVRER (for video) reasoning.

The questions are understood using RNNs and object detection in images and videos is done using CNNs. These different modules are used together to make a reasoning model. More research is being conducted in the field and our work adds to the research and gains a formidable step towards the development and enhancement of AI.