CS671 - Deep Learning and Applications
Hackathon 2019

GROUP NAME : LearnOrBurn
GROUP MEMBERS
Aj R Laddha B16004
Kaustubh Verma B16021
Nikhil T R B16066

Date: 17th May 2019

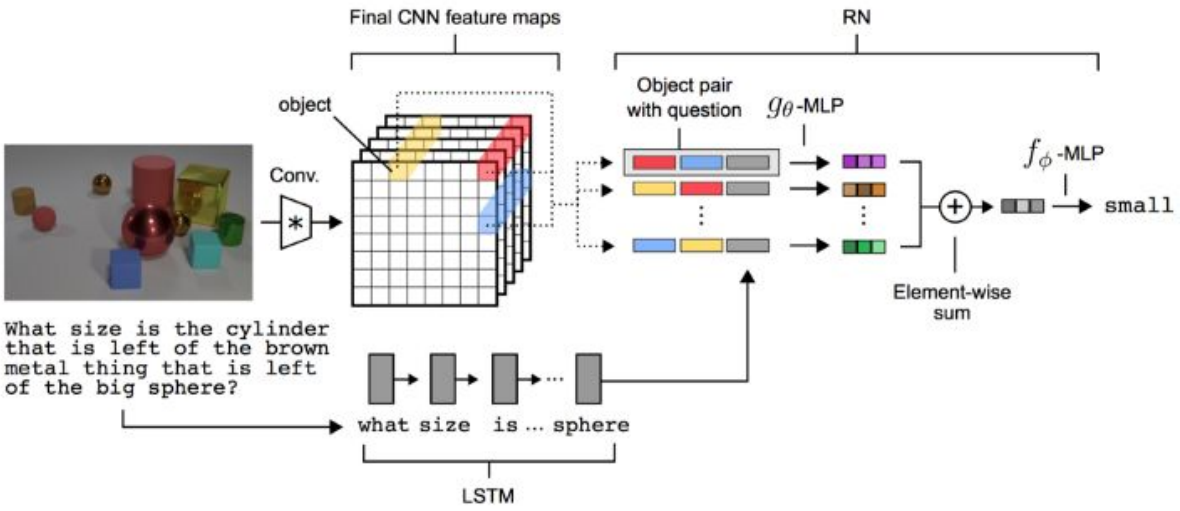
Problem Statement: Relational Reasoning is a difficult thing to achieve in picture based question and answer. With use of traditional MLPs, there is a requirement to learn n^2 functions. This will quickly get intractable as the number of objects increases. This needs to be somehow brought under control.

Dataset: We will use CLEVR v1.0 dataset with the pixel version, in which images are represented in standard 2D pixel form. It consists of -

- A **training set** of 70,000 images and 699,989 questions
- A **validation set** of 15,000 images and 149,991 questions
- A **test set** of 15,000 images and 14,988 questions

Existing Literature: The paper titled “A simple neural network module for relational reasoning” was submitted in 2017, with a goal to improve the relational reasoning using a specific formula for modeling generalised logic for relational reasoning.

Procedure: The architecture we are building is an improvement over the previous Visual question answering model by introducing a Reasoning network which works on pairs of object to develop reasoning between them to be able to answer questions that require reason logic.
In simple form Relational network is a composite function over a function of object pairs and question word embedding. This relational network would be data-efficient since it is a single equation of multiple object pairs and question words.



Questions are processed with LSTM to produce a question embedding, and images are processed with a CNN to produce a set of objects for the RN. Objects are constructed using feature-map vectors from the convolved image. The RN considers relations across all pairs of objects, conditioned on the question embedding, and integrates all these relations to answer the question.

Final Deliverable: Our model would be able to efficiently answer the visual questions present in CLEVR dataset with the added capability of answering reasoning questions such as shown below:

Non-relational question

Q: What is the shape of the gray object?
A: circle

Relational question

Q: What is the shape of the object
that is furthest from the gray object?
A: square

Above result obtained on Sort-of-CLEVR dataset.
Furthermore we may apply Siamese Network to get reasoning similarities between two such images.