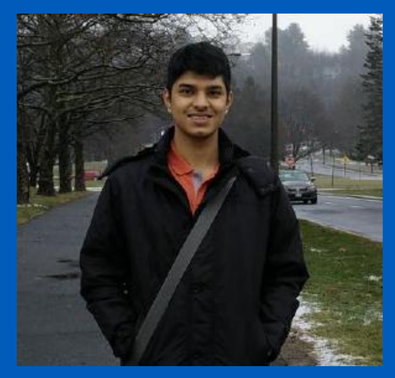
Visual QA for Relational Reasoning

Improving models for spatial and semantic relations between objects

Microsoft Research | MathWorks | SenseHawk

- Grad ML Research - Vision, QA | 5 months
- ML-Vision Intern - Semantic Seg | 8 months
- ML Engineer - Classification | 7 Months

ML | Vision
QA | Deep Learning



apimpley@cs.umass.edu | anishpimpley.github.io

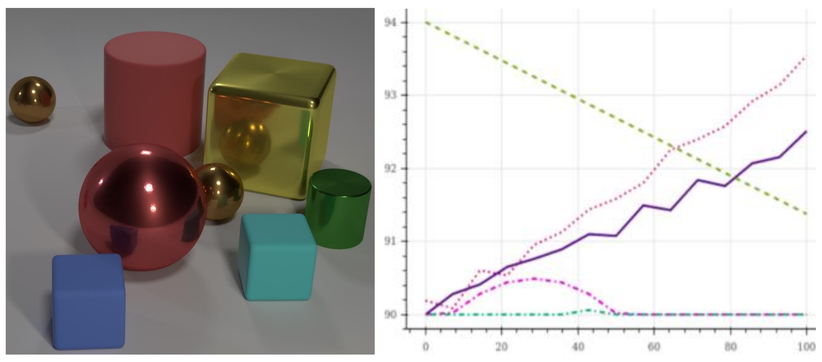
425-362-8258 | UMASS MS - CS (2017-19)

ANISH PIMPLEY

Problem :

- *How are different objects in a scene related?*
- *How to decode complex questions about semantic and spatial relationships between objects?*
- *How to achieve computational efficiency at scale?*

***What size** is the **cylinder** that **is left** of the **brown metal** thing that **is right** of the **big sphere**? → Large*

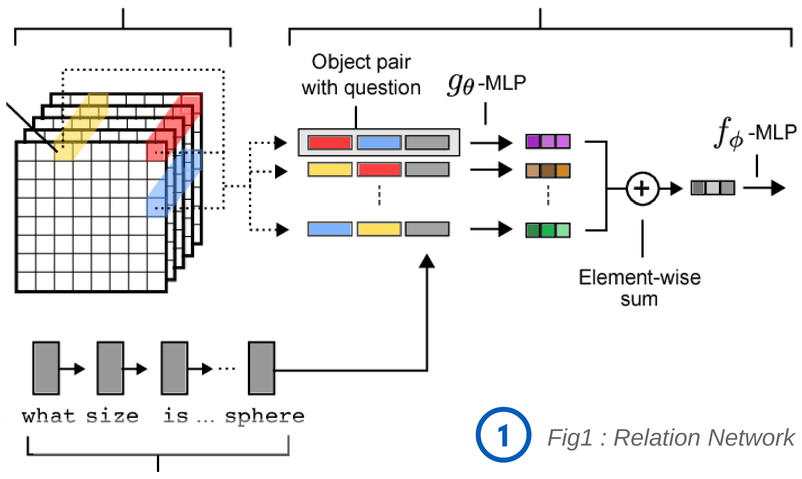


***Does medium orchid have the minimum area under the curve ?** → No*

Related Approaches :

Relation Networks – Pairwise score wrt. Qn.

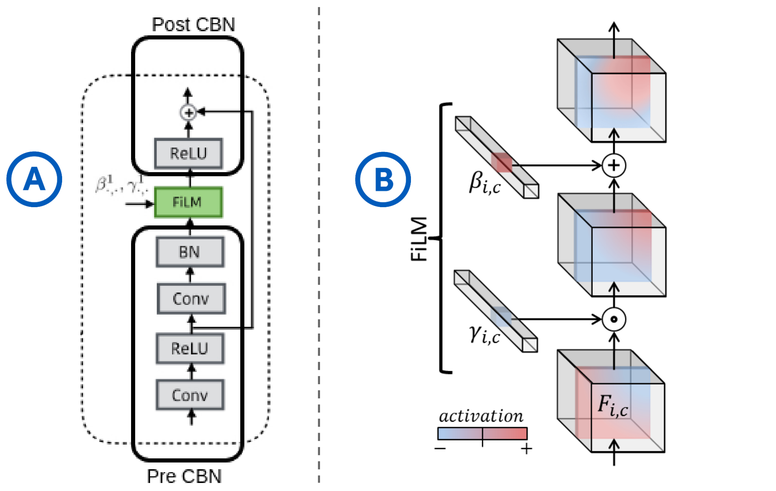
- n^(4) number of relations → ***SLOW !***
- After Conditioning, lot of redundant relations



1 *Fig1 : Relation Network*

Film – Question based affine transforms

- Repeated modulation of image feature maps
- Attention like; most features maps = 0

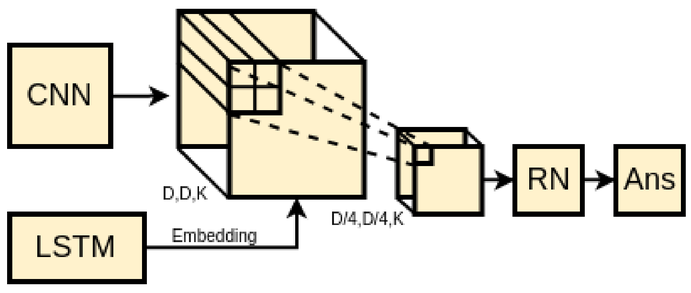


2 *Fig 2 : (a) Film block ; (b) Conditional Batch Norm*

Approach :

(A) Group Attention :

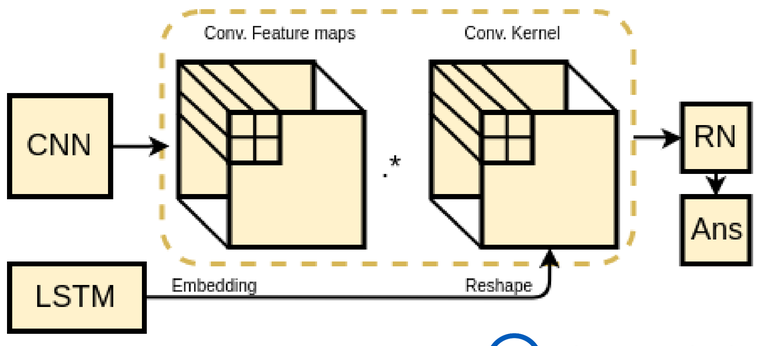
Most features after transformation based on question are redundant. Use Attention on groups (g) of objects and implicitly discard irrelevant objects. **Reduces number of computations in RN by: g^(4)**



3 *Fig 3 :group attention*

(B) Embedding as Convolutional kernel :

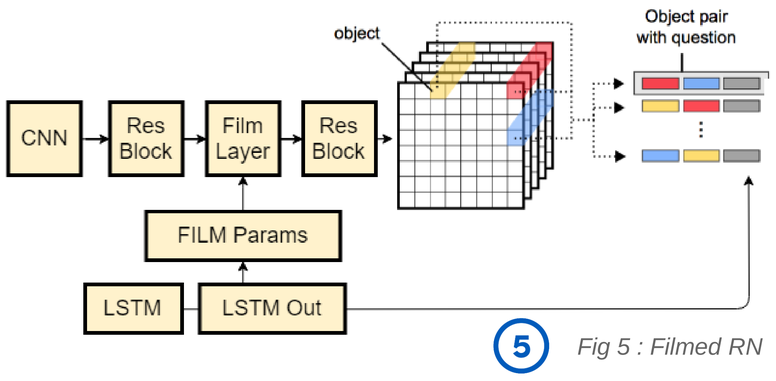
Reshape question embedding into a conv kernel. Let it implicitly perform attention on the image feature map convolutionally.



4 *Fig 4 : question kernel*

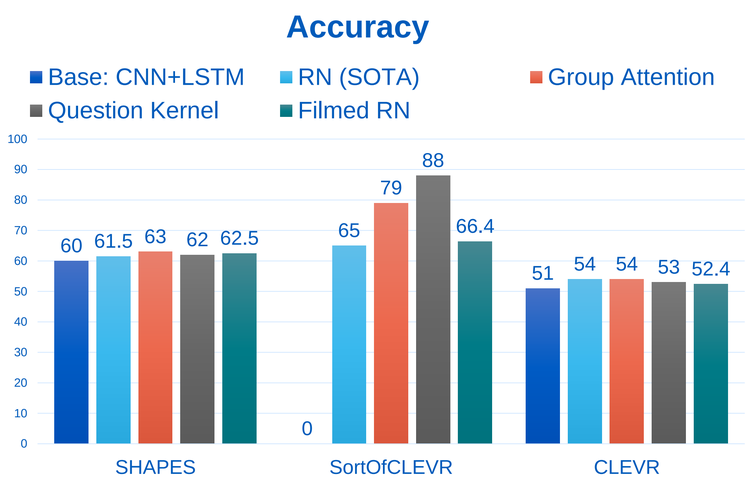
(C) Filmed RN - Apply CBN to CNN features :

Learn a FiLM transform and apply to intermediate CNN feature maps.

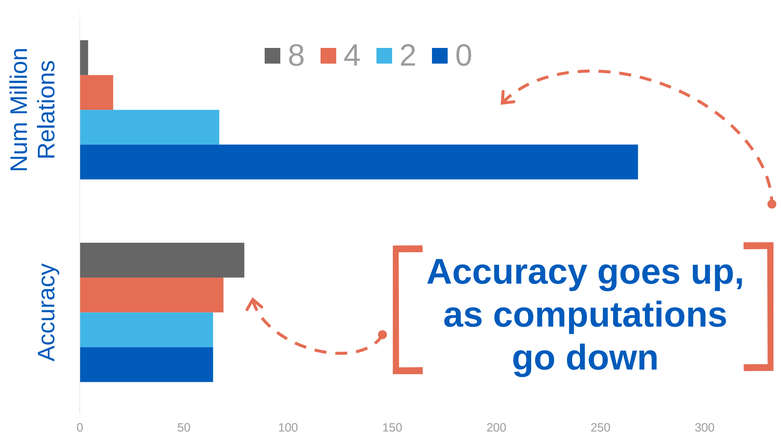


5 *Fig 5 : Filmed RN*

Results :



Interesting Observations :



Conclusion :

- Massive amount of feature redundancy
- Making neurons compete → improves results

Future :

- Explicit feature competition to eliminate features enmasse
- Rigorous experimentation

More about me:

MathWorks — Vision Intern

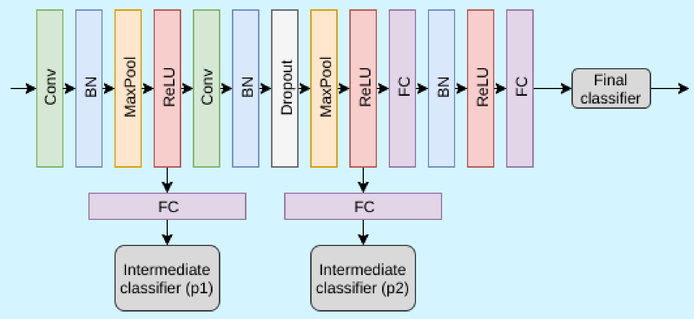
- State-of-the-art semantic segmentation model
- Extended concept to other networks
- Sole responsibility for architectural choices, implementation, function design & testing
- Converting hand-drawn diagrams to Simulink

SenseHawk — ML Engineer

- Extracting information from 3D point clouds : buildings, trees, foliage, cliffs, rocks, rivers
- Elevation Model construction & ground detection

UMASSCS SCHOOL OF COMPUTER SCIENCE — Grad Research

- Used for unbalanced datasets
- Deep Cascade classifiers for early prediction
- Replaced loss function with firm cascade loss



Seeking Full Time Opportunities

Starting : **May 2019**

- Machine Learning
- Data Science
- Computer Vision

References :

E. Perez et al. Film:Visual reasoning with a general conditioning layer.
A. Santoro et al.A simple neural network module for relational reasoning.
S.E.Kahou et al.,FigureQA:An annotated figure dataset for visual reasoning.
J. Johnson et al. Clevr:A diagnostic dataset for compositional language and elementary visual reasoning.

Special thanks to our mentors at IESL Umass & Microsoft Research Montreal : Samira Kahou, Adam Atkinson, Adam Trischler for this research opportunity.

Equal Contribution: Anish, Srideepika, Shruti, Srikanth