Week 8
11-777

Due: Thursday, Oct 19, by 4:30 pm

Instructions

A small yet significant part of the course is answering a few questions based on a paper you just read. By design, some of these questions are open-ended and the purpose is to start a thinking process. Some of those questions may already have been answered in recent literature (or in the given paper itself). We encourage you to, both, come up with your own ideas and survey recent papers before answering these questions.

A few key points to take care of:

- Answer two out of the four questions (you are welcome to answer all 4), unless stated otherwise.
- A lot of questions ask you to suggest changes or make alterations. It would be nice to support it with logical / mathematical arguments. Cite all sources used in the process of coming up with your answer. Figures and equations, if they support your arguments will be appreciated.

Questions

Learning to Reason: End-to-End Module Networks for Visual Question Answering

1. How is the work presented in this paper different from other prior work on Neural Module Networks? Give specific examples. What are the main contributions of this work?

2. How is the N2NMN trained end to end? More specifically, how do the authors optimize the loss function, given that it is non-differentiable in certain places?

3. The authors claim: "Compared with MCB (the VQA 2016 challenge winner method) trained on the same ResNet-152 image features, our model achieves slightly higher performance while being more interpretable." What do you think makes this model more interpretable than MCB?

4. Observe that the "policy search after cloning" method gives significantly better results than the "cloning expert" method for CLEVR, but the difference is very small for VQA when trained on the same ResNet-152 image features. Why do you think this is the case?