Deep Learning for Image Captioning

By
Siddesh Pillai (srp4698@rit.edu)

Advisor : Professor Jeremy Brown (jsb@cs.rit.edu)
Colloquium Advisor : Professor Leon Reznik (lr@cs.rit.edu)

Agenda

- Recap
- Milestone-1 tasks
- Challenges

Recap

- **What is Deep Learning?**
- A branch of machine learning based on set of algorithms that attempts to model high level abstractions in data by using multiple processing layers with complex structures otherwise known as non-linear transformations [1]
- Last time we talked about some applications of Deep Learning
- **The Problem Statement** – Image Captioning
- **Base implementation paper** – Deep Visual Semantic Alignments for Generating Image Descriptions [2]

Milestone – 1 tasks

- Data collection – MSCOCO dataset[3]
- Setting up Torch[4] and Lua
- Understand the neural talk framework
- Test the pre-trained model

Challenges

- Adapting to Lua was a good experience
- Implementation point of view - most of the code mere functions to initiate the framework
- Configuring the machine with neural talk was a bit tricky
- Tested on 40,505 images

Next Steps

- Expose the model and try to add features
- Retrain the model
- Start working on client application

References

1. Deep Learning https://en.wikipedia.org/wiki/Deep_learning
2. Andrej Karpathy and L. Fei-Fei. Deep Visual-Semantic Alignments for Generating Image Descriptions, 2012
3. MSCOCO https://github.com/tylin/coco-caption
4. Torch http://torch.ch/

Thank you