

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

- Reminder
- Milestone-3 tasks

Reminder

- **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 [\[1\]](#)
- **The Problem Statement** – Image Captioning
- **Base Implementation** – Deep Visual Semantic Alignments for Generating Image Descriptions [\[2\]](#)

Milestone – 3 tasks

- Learning the effects of peripheral vision [\[3\]](#) (part of report)
- Perform N-fold cross validation on spatial and color arrangements (work in progress)
- Work on the android app (work in progress)
- Work on adding voice feature (work in progress)
- Work on Project Report (work in progress)
- Work on Poster (work in progress)

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. Peripheral vision https://en.wikipedia.org/wiki/Peripheral_vision

Thank you