Deep Learning the Relationship between Visual and Textual Data

Patrick Poirson
University of South Alabama

Last Week

- Wrote a python program to convert features in numpy to matlab
- Switched to a different deep learning toolbox
  - Due to lack of good documentation
- Ran first two experiments
- Rereading experimental results for "Multimodal Learning with Deep Boltzmann Machines" and "Multimodal Deep Learning"

First Experiment

- Given an image feature vector generate the tags vector
- Tags are 2,000 dim vector
- Was unimodal with two hidden units
- Trained on 20k images and tested on 5k
- Ran into memory management issues
  - Good experience for future experiments which will use more data
- Results
  - Only 39 correct

Second Experiment

- Given an image vector generate a class label for the image
- Labels are 38 dim vector. Ex. Bird, tree, people
  - Each image may belong to several classes
- Trained using 20k images and tested on 5k
- Results
  - 912 correct

Next Steps

- Write code for multimodal network
- Start training in the next few days
- Should move a lot faster now that all of the data is in the correct format and I am a lot more familiar with the deeplearning toolbox