Deep Learning the Relationship between Visual and Textual Data

Patrick Poirson
University of South Alabama

Progress

- Finished pretraining
- Multimodal Classification
- Unimodal Classification
- Close to finishing retrieval code

Multimodal Classification

- Our results
- MAP
  - .419
- Prec@50
  - .791

Multimodal Inputs		
Model	MAP	Prec@50
Random	0.124	0.124
LDA [2]	0.492	0.754
SVM [2]	0.475	0.758
DBM-Lab	0.526	0.791
DBM-Unlab	0.585	0.836
DBN	0.599	0.867
Autoencoder (based on [5])	0.600	**0.875**
DBM	**0.609**	0.873

Unimodal Classification

- Our Results
- MAP
  - .183
- Prec@50
  - .279

Unimodal Inputs		
Model	MAP	Prec@50
Image-SVM [2]	0.375	-
Image-DBN	0.463	0.801
Image-DBM	0.469	0.803
DBM-ZeroText	0.522	0.827
DBM-GenText	**0.531**	**0.832**

Retrieval Tasks

- Create database of images by randomly selecting 5k images
- Select 1k disjoint set of images to use as queries
- All queries and images in the database were mapped to the joint hidden representation
- Use cosine similarity function to match queries
- Assume that if any of the 38 classes overlap than the image is relevant to the query
- Measure Precision and Recall
- Perform multimodal and unimodal

Next Week

- Finish code for retrieval
- Run author's code for autoencoder classification
- Improve classification results