Deep Learning the Relationship between Visual and Textual Data

Patrick Poirson
University of South Alabama

Multimodal Classification

- New results
- MAP
  - .522
- Prec@50
  - .858
- Previous Resu
- 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

How to Improve

- Logistic Regression
- Right now using regression svm with no tuning of parameters

Unimodal Classification

- Last time used zero vector for tags
- This time generated tags using given image
- New Results
- MAP
  - .321
- Prec@50
  - .556
- Previous 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**

Things to do

- Logistic Regression
  - Should be able to find good code to use
- Improve method for generating tags