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
Our results

MAP
- 0.419

Prec@50
- 0.791
Unimodal Classification

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

<table>
<thead>
<tr>
<th>Unimodal Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Image-SVM [2]</td>
</tr>
<tr>
<td>Image-DBN</td>
</tr>
<tr>
<td>Image-DBM</td>
</tr>
<tr>
<td>DBM-ZeroText</td>
</tr>
<tr>
<td>DBM-GenText</td>
</tr>
</tbody>
</table>
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