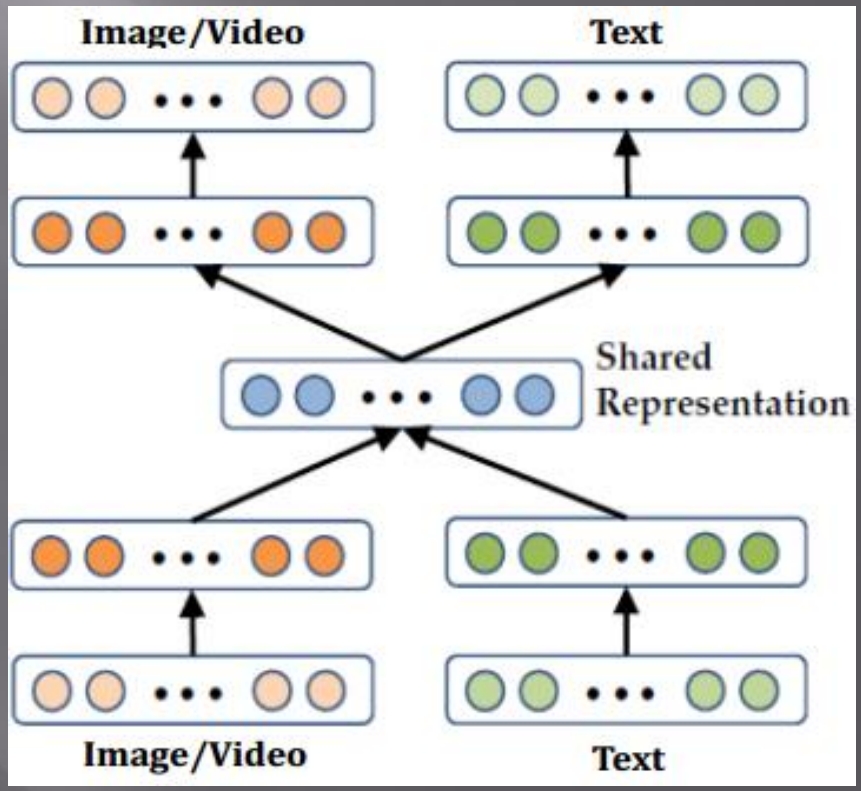
Deep Learning the Relationship between Visual and Textual Data

Patrick Poirson
University of South Alabama

Multimodal SDA

- Wrote code to train a multimodal SDA
- Pretraining a model with the code and flickr data
- 60% complete



Multimodal Classification

- Use image and text to classify
- 38 classes exs. Dog, Car, People, Sky
- Extract the fused representation and feed to a separate logistic regression for each of the 38 topics
- Could use a SVM

Unimodal Classification

- Use only images to classify
- Same 38 topics
- With the DBM they do two unimodal classification experiments
  - Let the model generate the text
  - Keep the text input clamped at zero

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 these experiments
- Read paper *Multimodal semi-supervised learning for image classification* CVPR 2010
  - Got slightly better results than DBM
  - Use multiple kernel learning