Deep Learning the Relationship Between Visual and Textual Data



REU: Patrick Poirson

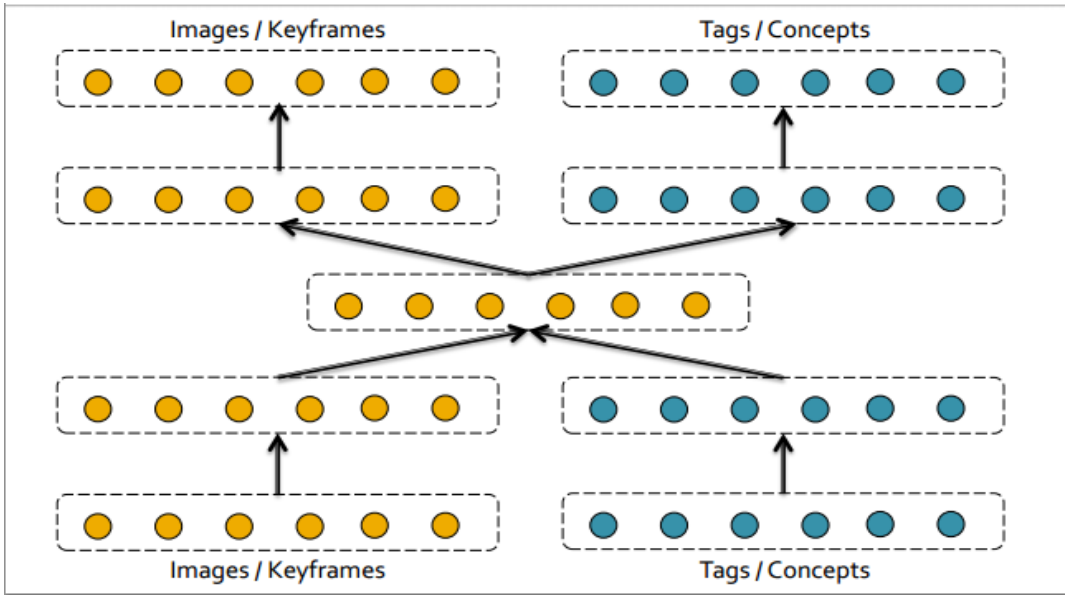


Mentor: Haroon Idrees

Problem

- Flickr dataset contains 1M images and their tags
- 25k of these images have been annotated for 38 classes
- Example classes bird, tree, and night
- Classify an image into the 38 classes
- Multimodal Classification
  - Given an image and its tags
  - Learn joint representation
- Propose Multimodal Stacked Denoising Autoencoders

Results



Method	MAP	Prec@50
Random	.124	.124
LDA	.492	.754
SVM	.475	.758
DBM	.609	.873
SDA (Proposed)	.533	.858