



LUND  
UNIVERSITY

# Entity Linking in Images and Captions

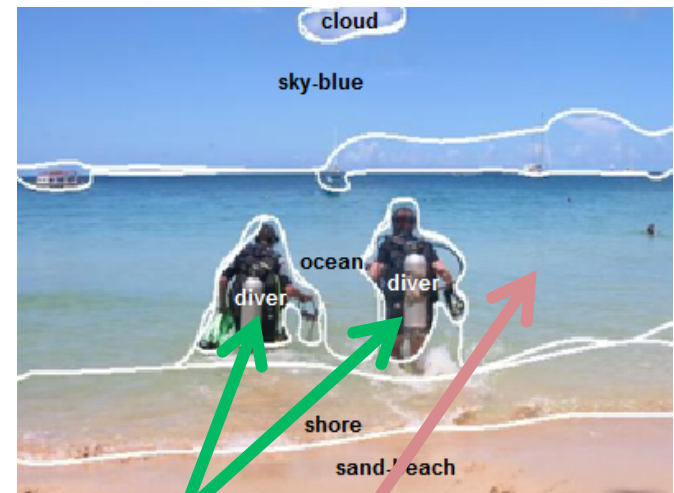
BY DUNCAN SOMMER AND SILAS PEDROSA



# The Big Picture

---

- Identify segments in images and entities in captions
- Then, find matches between the two so that useful information can be retrieved
- Several tasks involved:
  - Image Processing
  - Segment Classification
  - Natural Language Processing
  - Linking

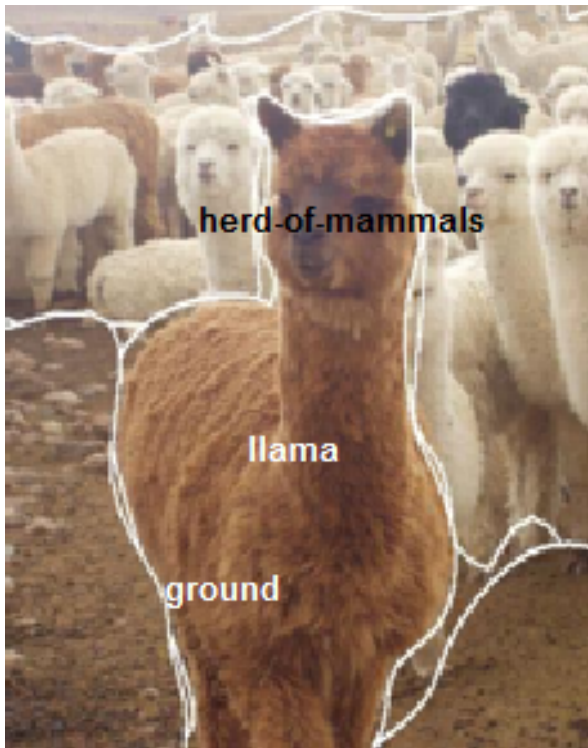


Two people with scuba equipment enter the blue water of a tropical bay.



# The Dataset

---

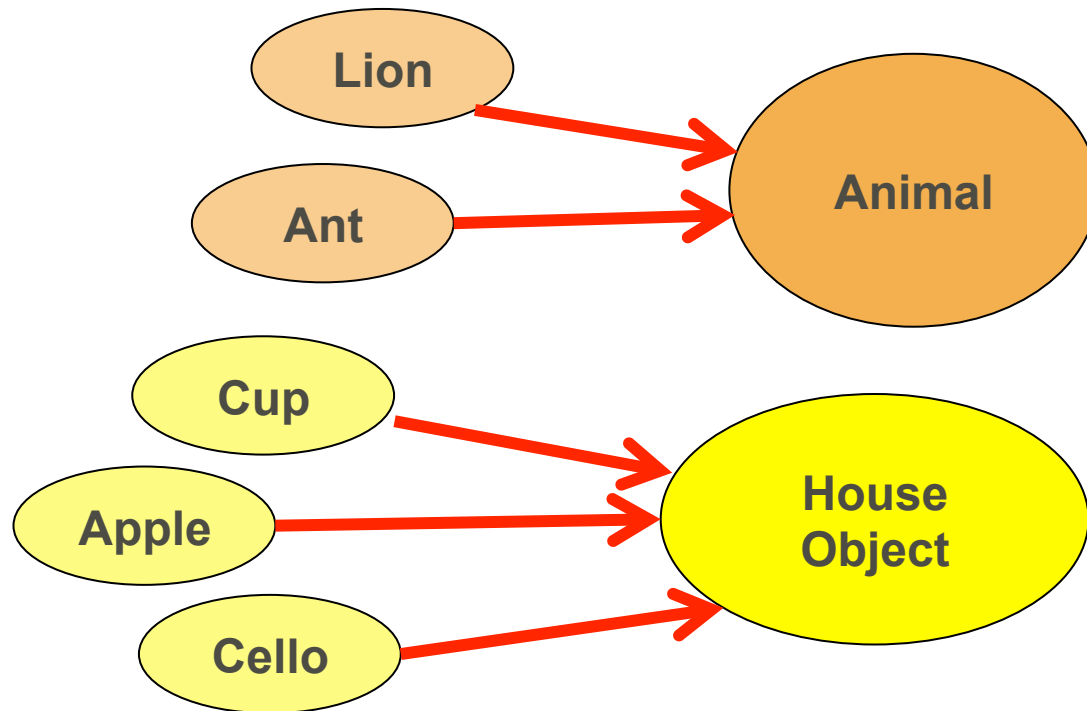
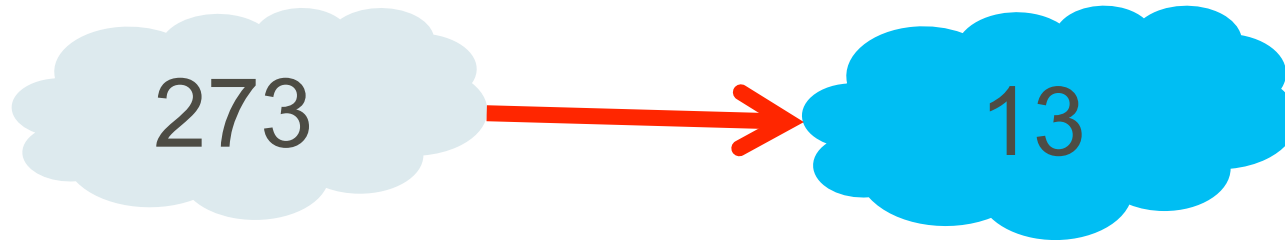


- SIAPR: Segmented and Annotated database
  - 99,535 segments, average 5/image
  - Manually segmented and classified
- Each segment has a feature vector representing color, convexity, area, etc.
- 273 possible classes



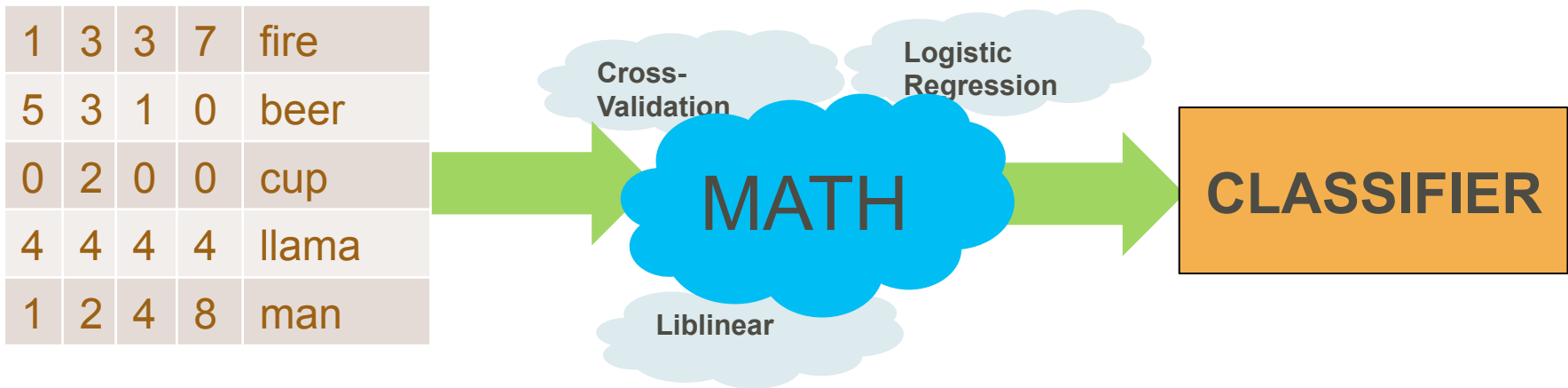
# Classes and Clustering

---



# Local Classifier

---



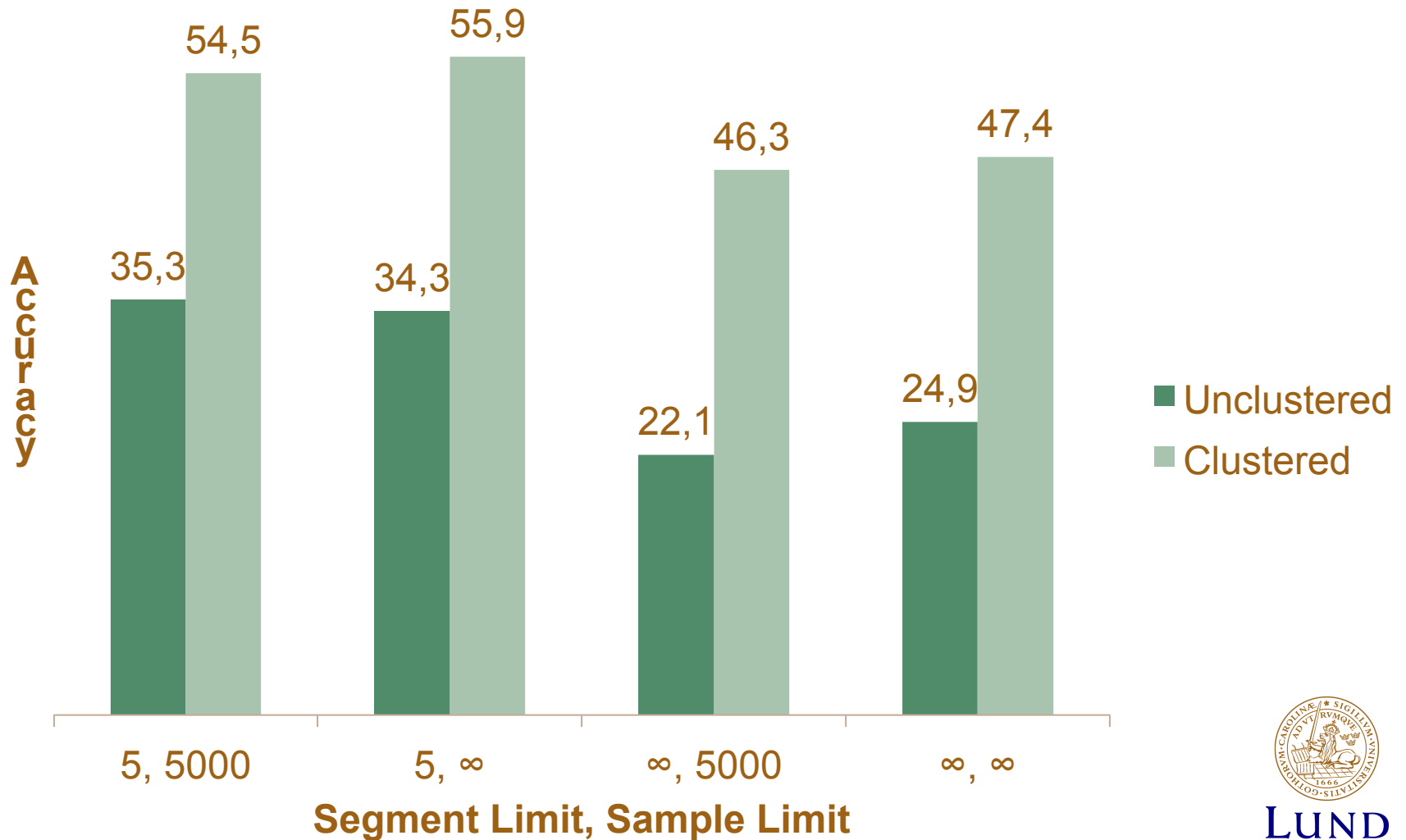
# Experimental Setup

---

- Sample Limit
  - Focus on a subset of the SIAPR data.  
Images will be more associated with one another.
- Segment Limit
  - Throw out images with too many entities.
- Clustering
  - How many classes?  
Fewer choices = better accuracy.



# Local Classifier Results



# Next Step: Reranker

---

- Think about which objects often appear together
- Make a new sample set to show this information

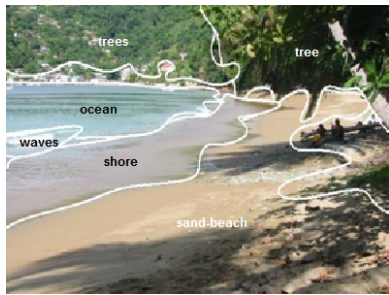
Beer	Hat	Dog	Squirrel	
1	1	0	1	True
0	1	1	0	False

- Train a new classifier!



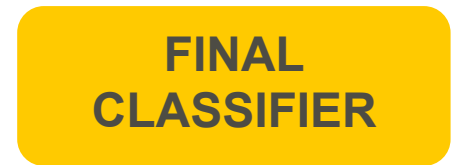


# Data Flow

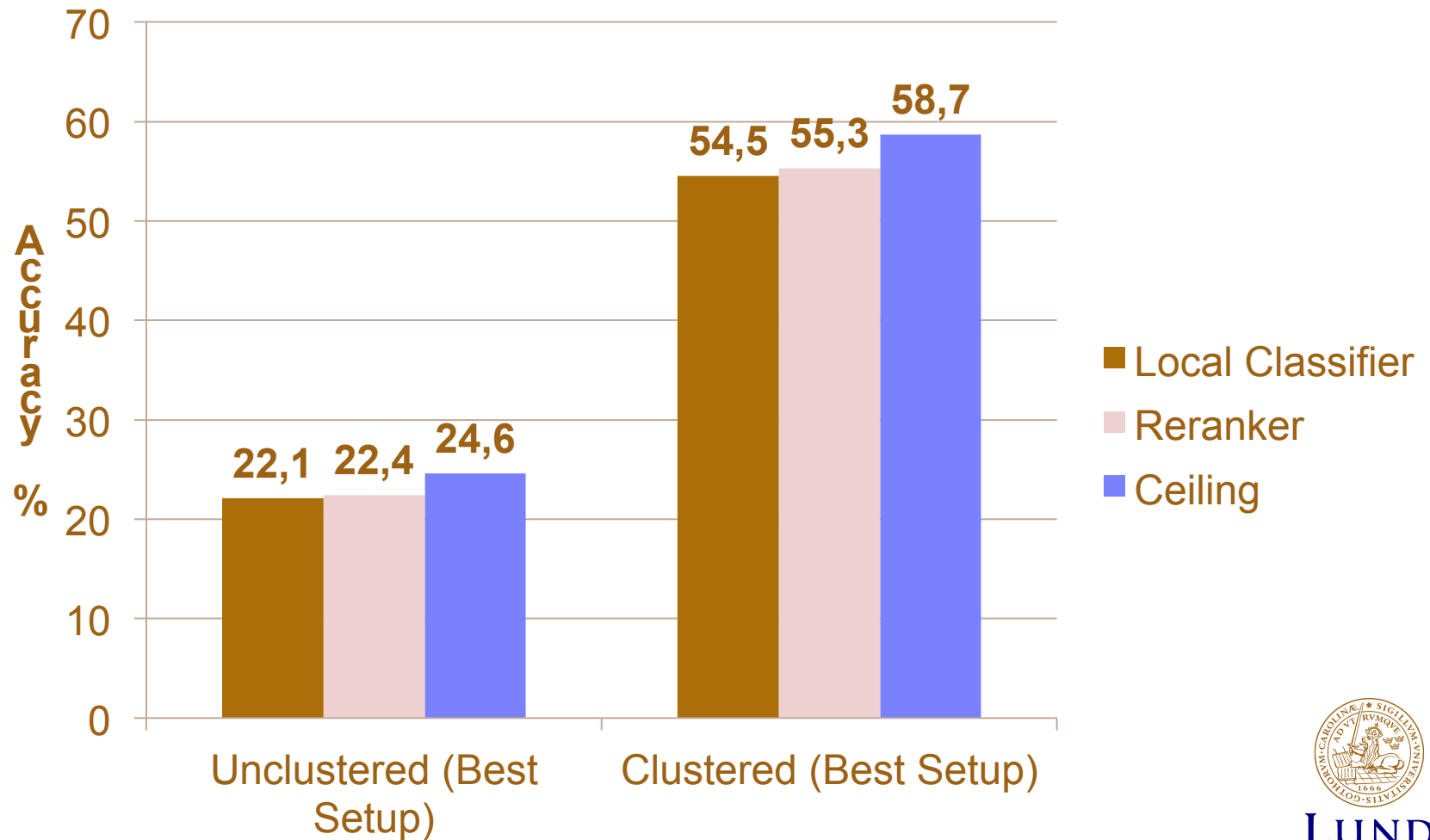


9	0	0	1	dog
1	3	3	7	cat

0	0	0	1	True
1	0	1	0	False



# Reranker Results



# Final Thoughts

---

- Better classes and clustering
- Other approaches to the reranker
- Our classifier is only part of the bigger project





LUND  
UNIVERSITY