Animals on the Web

• Authors
  – Tamara L. Berg, David A. Forsyth

• Presenter
  – Nitin Rastogi, Chen Ling
Animals on the Web

- Introduce a method for identifying images containing categories of animals.

- A classifier that uses word and image information to determine the category of any given test image.
Introduction
Motivation

• Animals are difficult to identify, (a wide variety of appearances, depictions and aspects)

• Google image search for “monkey” only 30 actual “monkey” pictures in first 100 results in 2006. (about 70 in 2012)

• Need more accuracy
Challenges
Challenges
Classifier

• Consists of two stages
  – Training(Preparing Dataset for Classifier)
  – Testing

• Computes following four Cues to determine category of the image.
  – Text
  – Color
  – Shape
  – Texture
Classifier Cont’d

• Training Stage
  – Text Representation
  – Image Representation
  – Exemplar Initialization (Dataset finalized!!)

• Testing Stage
  – Shape, Color, Texture and Word based Voting
Preparing Dataset for Classifier

- 10 animal categories: alligator, ant, bear, beaver, dolphin, frog, giraffe, leopard, monkey, penguin

- Google text search, find related webpages.

- From web pages extract images
Text Representation

- Nearby words are more likely to be relevant to an image on the web page.
- Parse HTML, get 100 words surrounding the image link (remove stop words)
- Get a bag of words, build words’ vector (word -> word’s count)
Researchers have identified a new species of African monkey, Cercopithecus lomamiensis (lesula).

Democratic Republic of Congo as the lesula, has a blond chin and upper chest, in contrast to its dark limbs. It has a reddish-colored lower back and tail.

“And adult males have a huge bare patch of skin in the buttocks, testicles and perianal area,” said John A. Hart, the researcher who described the coloring. “It’s a brilliant blue, really pretty spectacular.”

Dr. Hart is a field scientist with the Lukuru Foundation, a wildlife research group; his colleagues include his wife, Terese Hart. They report their findings in the journal PLoS One.

Text Representation

• Use **Latent Dirichlet allocation (LDA)** algorithm to find 10 latent topics for each category.

• Each topic contains the words’ probabilities.

• Select 50 top words to represent a topic.
LDA By Example - Input

• I ate a banana and spinach smoothie for breakfast
• I like to eat broccoli and bananas.
• Chinchillas and kittens are cute.
• My sister adopted a kitten yesterday.
• Look at this cute hamster munching on a piece of broccoli.
**LDA By Example - Output**

- **Sentences 1 and 2:** 100% Topic A
- **Sentences 3 and 4:** 100% Topic B
- **Sentence 5:** 60% Topic A, 40% Topic B
- **Topic A:** 30% broccoli, 15% bananas, 10% breakfast, 10% munching, ... (at which point, you could interpret topic A to be about food)
- **Topic B:** 20% chinchillas, 20% kittens, 20% cute, 15% hamster, ... (at which point, you could interpret topic B to be about cute animals)
Topics for category

• e.g., for the ”monkey” category, we get 10 topics, each topic contains top 50 words according to their probabilities:
  – Topic1 : \((w_{(1,1)}, w_{(1,2)}, w_{(1,3)}, \ldots w_{(1,50)})\)
  – Topic2 : \((w_{(2,1)}, w_{(2,2)}, w_{(2,3)}, \ldots w_{(2,50)})\)
  – Topic3 : \((w_{(3,1)}, w_{(3,2)}, w_{(3,3)}, \ldots w_{(3,50)})\)
  – ...
  – ...
  – Topic 10 : \((w_{(10,1)}, w_{(10,2)}, w_{(10,3)}, \ldots w_{(10,50)})\)
Image Representation

• 3 types of Image features are employed for each Image.
  – Shape based geometric blur features
    • 50-400 local shape features
  – Color features
    • 9 color features, RGB color histogram
  – Texture features
    • Texture histogram
Exemplar Initialization

- LDA gave us the likelihood of each image under each topic
- Assign each image to its most likely topic
- Select top 30 images as exemplars for each topic
Exemplar Initialization

• Because of the polysemy, (e.g, alligator can refer to animal or “alligator boots”), user is required to label each topic to “Relevant” or “Background”

• With this labelling we merge topics into a single relevant topic and a background topic (pooling their exemplars and likely words)
What all we did till now?

For Each category

One animal category (Ex: monkey)

LDA - 10 topics with label (positive or negative)

Topic 1

Topic 2 (One 50 words vector)

30 images

Image 1

Image 2

Image 3

......
• Topics from LDA

Topic 1

Topic 2

Topic 10

User Intervention

Relevant Topic

Background Topic
Categorization of Test Image by Voting

- Voting is based on
  - Shape
  - Color
  - Texture
  - Word

For each test image we compute the above mentioned features in a similar fashion as we did during training.
Voting Process Cont’d

• For each feature, we create two pool
  – Relevant exemplars (+)
  – Background exemplars (-)
  – 1 nearest neighbor classifier with similarity measured using normalized correlation
  – For each visual cue, Compute sum of similarities of features matching positive exemplars
  – Normalize each cue score for each image
Voting Process Cont’d

• We now have 3 scores/Cues belonging to
  – Shape
  – Color
  – Texture

• 4\textsuperscript{th} score is calculated for words, by ??
  – Hint: Similarity with relevant topic computed by
    LDA
Cue Combination and Its Advantage

• Combine 4 independent cue scores.
  – Weights??

• Advantage
  – Cues are used to rank image independently
  – Wider range of cues ??
  – By combining, much better result is achieved
    • Why??
    • Hint – 1 cue may work well for some images, not for all
Results

• Method outperforms Google text search in classification on top 100 images returned for all categories. (2006)
• Giraffe and frog classifiers are especially accurate, returning 74 and 83 true positives.
• Visual information makes substantial contribution.
• Discuss- Censored Vs Uncensored Exemplars.