Crowdsourced Discovery of Fine-grained Attributes

Subhransu Maji

## Toyota Technological Institute at Chicago

Overview

Goals

Attribute discovery for fine-grained and detailed recognition of visual categories.

Approach

- A novel annotation framework to collect descriptions of a collection of images
- A novel analysis of the text to discover the structure of parts and attributes

Key idea

- A good set of attributes should enable easy *discrimination* between instances
- We also focus on the *communication* aspect which is important for human-computer interaction
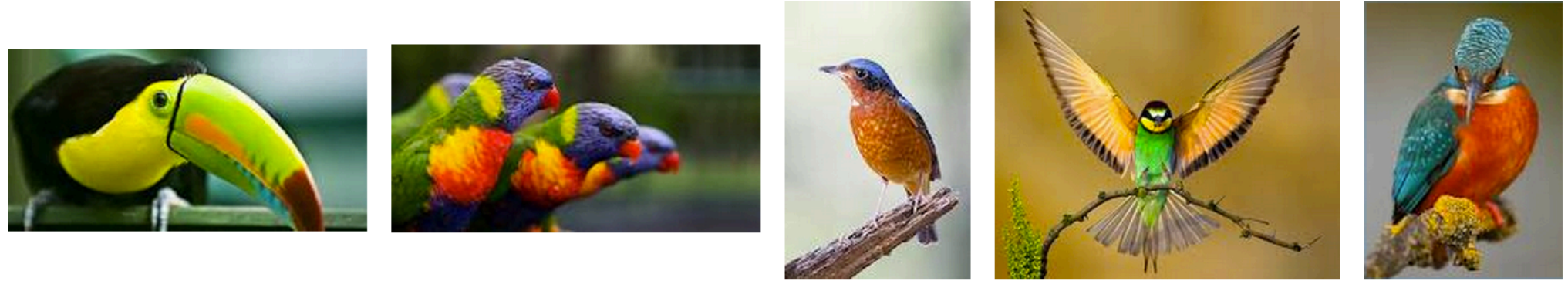
Sources of attributes

Field guides

- Can provide an exhaustive list of attributes when available, but may not be task specific, or jargon free
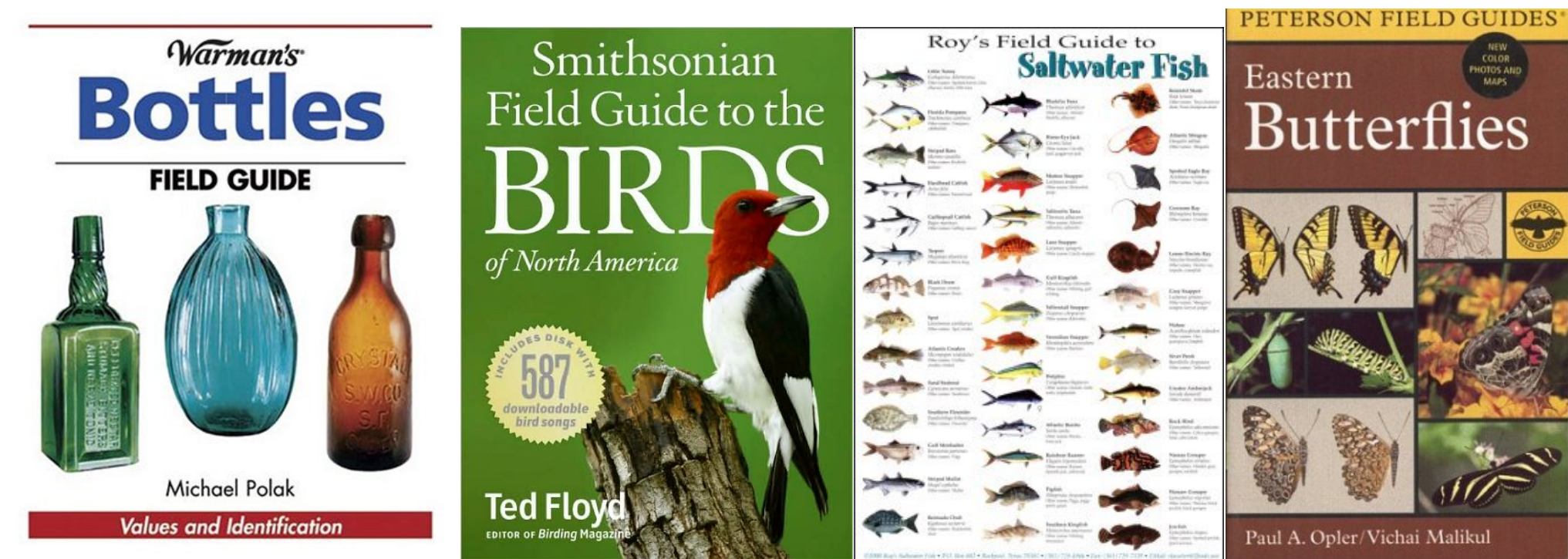
Captioned images on the web

- Descriptions are often not visual
- Detailed captions are available only for few categories

Image from Berg et al., ECCV'10

We crowdsource attribute discovery

- Using a simple interface that we found to be very effective

Our approach

Annotation task

- Describe the visual differences between the two images
- The annotators list 5 differences between the images in short sentences separated by v.s.

list properties
- plane
- has engine
- red color
- has rudder

vs.

list differences
- propeller plane vs. passenger plane
- one engine vs. four engines
- red color vs. white color
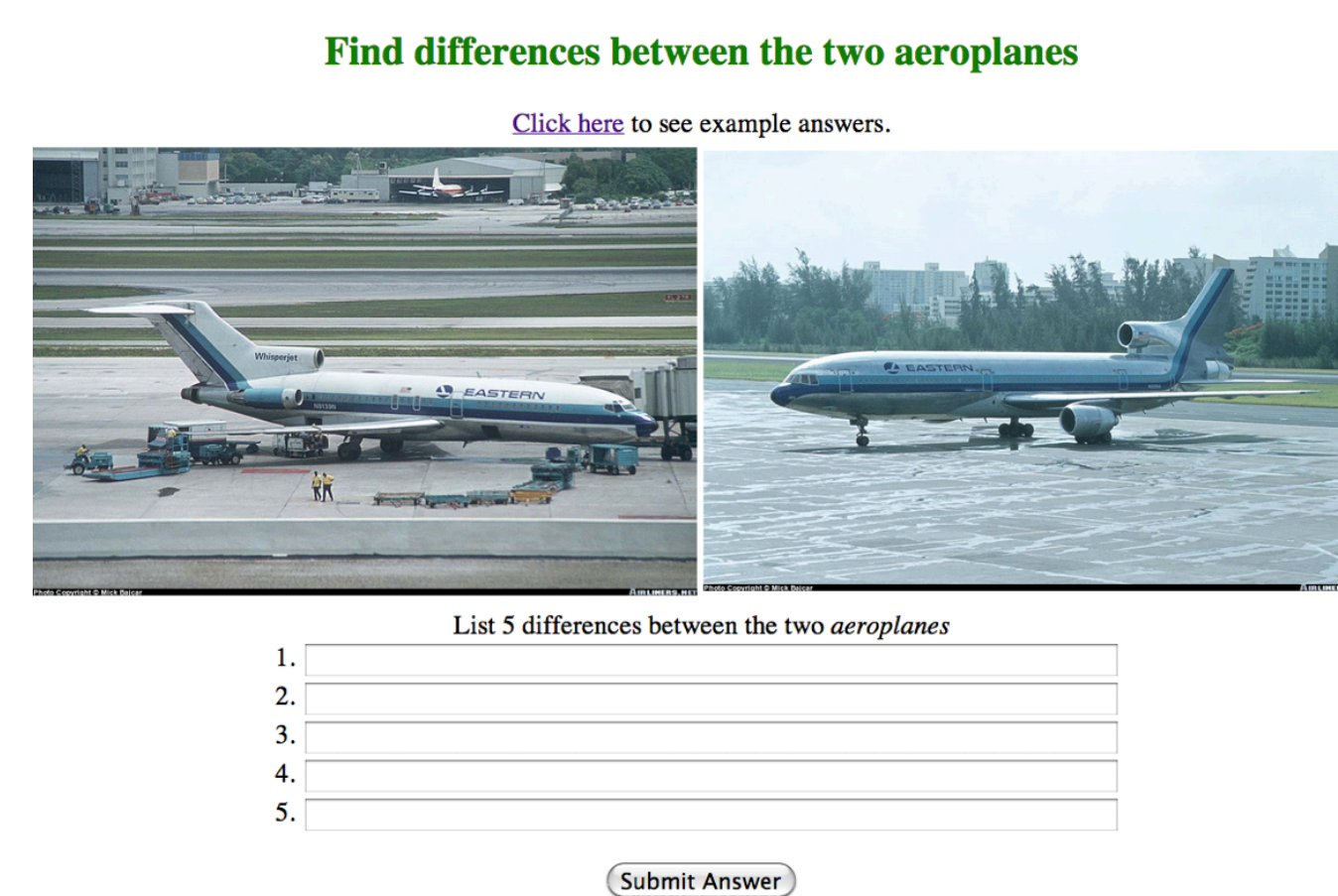- round rudder vs. pointy rudder

Description vs. Discriminative description

- Annotators are forced to reveal detailed attributes which are useful for fine-grained discrimination by design
- Also provides us with sentence pairs that can be analyzed to discover the structure of attributes, e.g., parts, modifiers and their relations.

Collecting annotations

Interface on Amazon Mechanical Turk

- Users respond in free form text

Find differences between the two aeroplanes
Click here to see example answers.

List 5 differences between the two aeroplanes
1.
2.
3.
4.
5.

Submit Answer

We also provide a few example annotations

- yellow and red rudder vs blue rudder
- two wings vs one wing
- yellow body vs blue body
- facing right vs facing left

- left facing vs right facing
- red rudder vs white rudder
- passenger plane vs propeller plane
- wings near bottom vs wings on top
- two side engines vs one side engine

We collect annotations for random pairs of images

- If an attribute is present with probability p, then it is likely to be revealed in a pair with probability $p(1-p)$, which is highest for $p=0.5$, i.e. the most discriminative attribute

Example annotations collected

yellow black body
pointy beak
short tail
black spot over head
short leg

orange brown body
shape beak
long tail
brown stripe over head
long leg

propeller to the body
one rudder
thin body
low wings
facing towards left side

propeller to the wing
two rudders
fat body
high wings
facing slightly towards

Analyzing the text

Instance specific properties

- Different properties are revealed in different comparisons
- The frequency of an attribute is a measure of its discriminative ability

sorted by frequency

5: blue wings
4: black beak
4: long tail
3: blue head
3: short beak
2: blue white body
2: blue white fur
2: sharp beak
2: short tail
1: big body
1: black leggs
1: black legs
1: black wings
1: blue body
1: blue color head
1: gray body
1: long bird
1: orange white body
1: short wings
1: slim body
1: white and blue fur

[From Wikipedia] **Florida jay:** It has a strong black bill, blue head and nape without a crest, a whitish forehead and *supercilium*, blue bib, blue wings, grayish underparts, gray back, long blue tail, black legs and feet.

Analyzing the sentence pairs

- Sentence pairs provide additional structure over the words used
  - *parts*: words that repeat across a pair
  - *modifiers*: words that are different across a pair
  - Can also be obtained using part-of-speech taggers
- A sentence pair is more informative than two sentences in isolation
  - *Additional constraint*: the modifiers belong from the same category
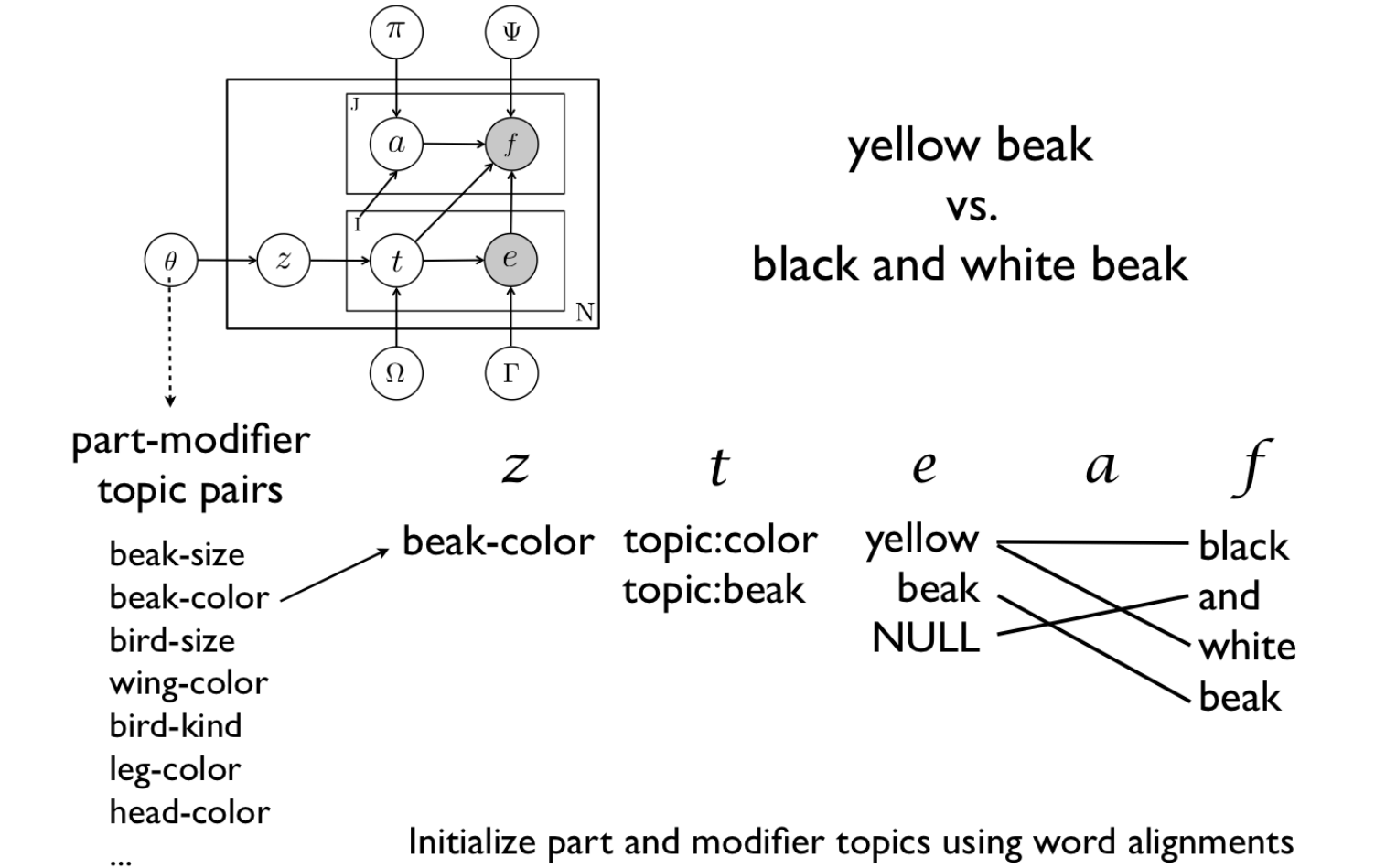 - In the example below, red and white must be two kinds of the same thing (color in this case).

Example

red rudder vs. white rudder → rudder {red, white}
pointy nose vs. round nose → nose {pointy, round}

sentence pairs → parts → modifiers

A generative model of sentence pairs

Discovering parts, modifiers and attributes

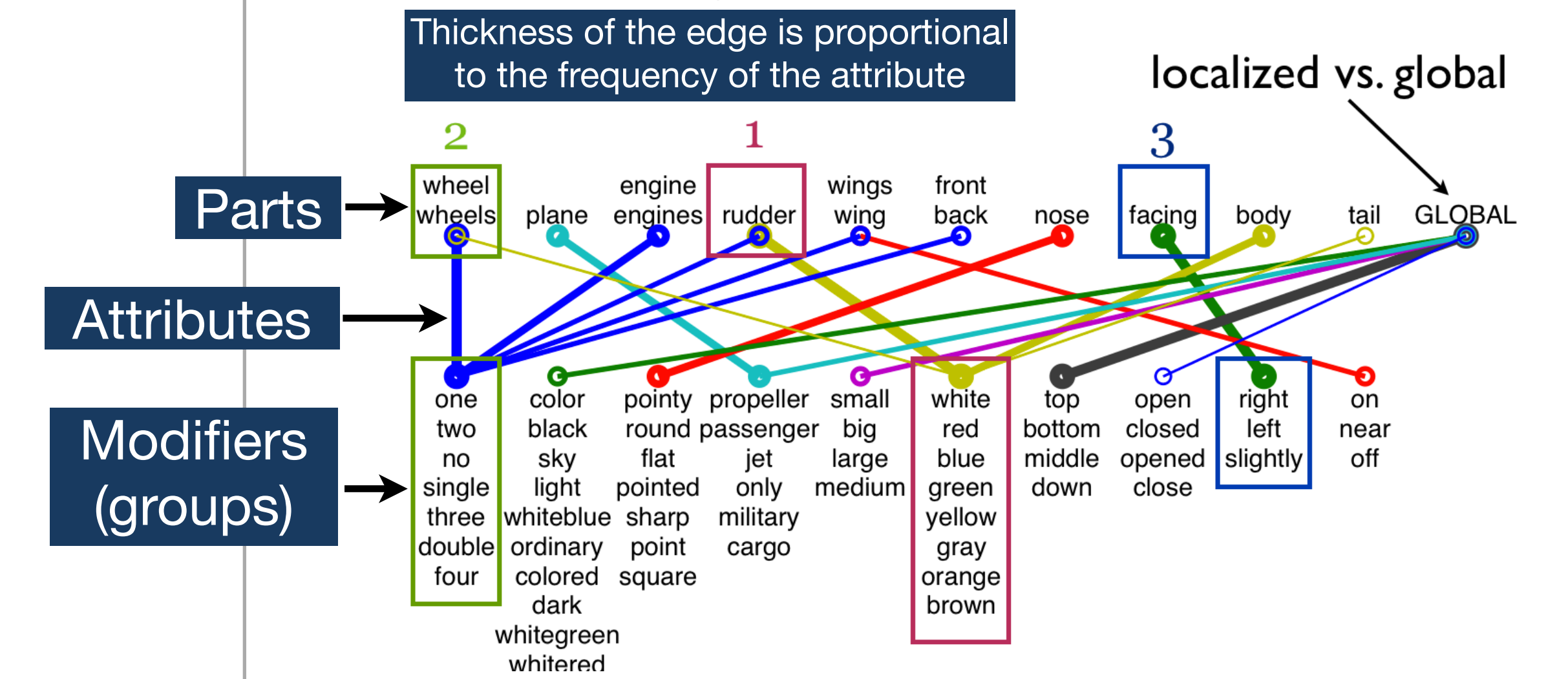- The constraints described earlier can be captured using a generative bipartite-topic model.

yellow beak
vs.
black and white beak

part-modifier topic pairs

beak-size
beak-color
bird-size
wing-color
bird-kind
leg-color
head-color
...

beak-color   topic:color   yellow   black
                topic:beak   beak   and
                               NULL   white
                                          beak

Initialize part and modifier topics using word alignments

- Parameter estimation using variational EM algorithm

Experiments

FGVC-Aircraft dataset (see our other poster)

- 200 random images, 1000 random pairs

Thickness of the edge is proportional to the frequency of the attribute

localized vs. global

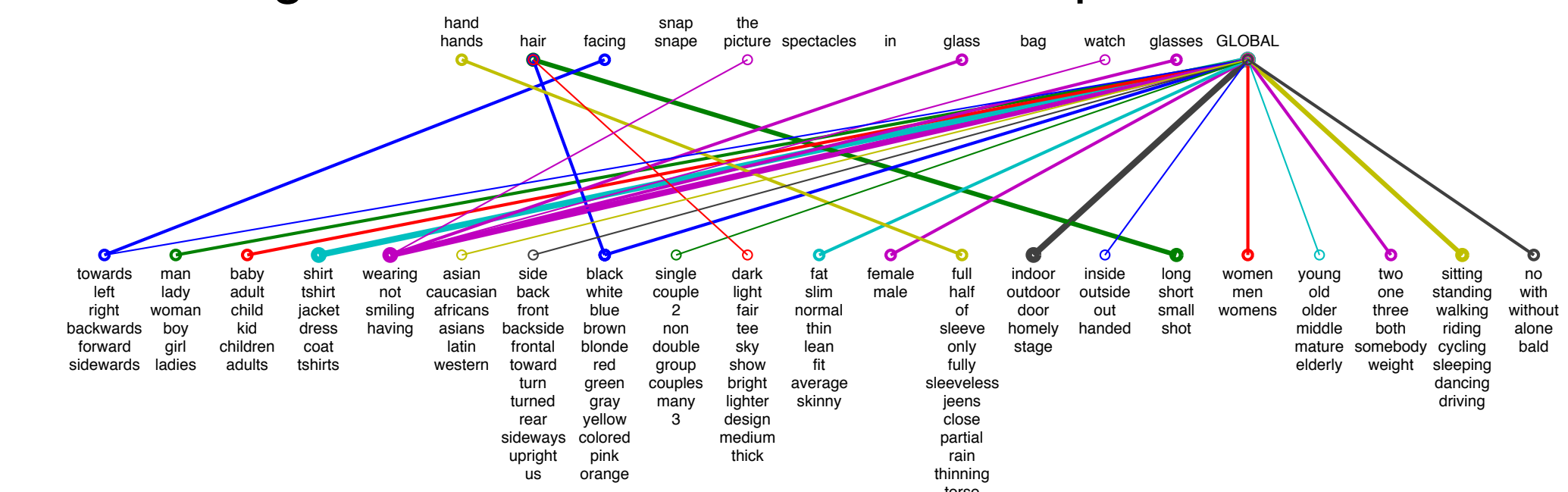Parts / Attributes / Modifiers (groups)

Caltech-UCSD birds 200 dataset

- 200 images (1 per category), 1600 pairs
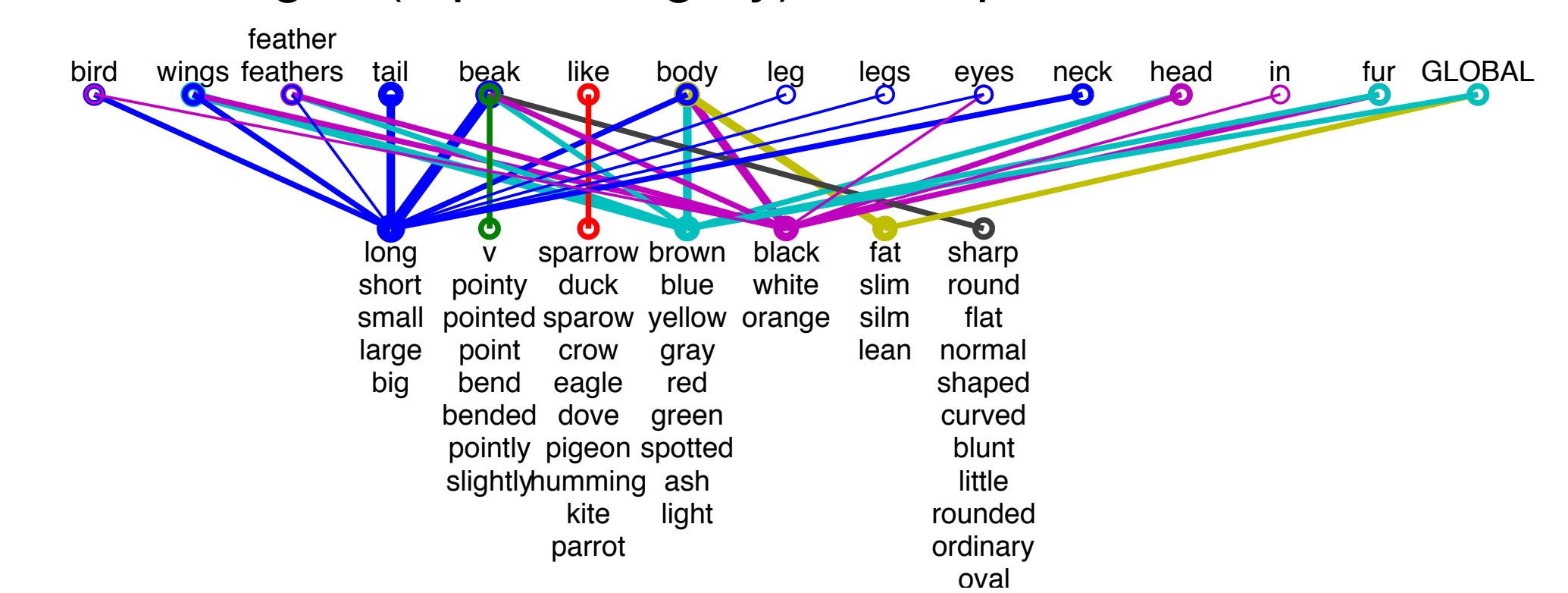
PASCAL VOC 2011 person dataset

- 200 images from trainval subset, 1600 pairs

Reference

Discovering a Lexicon of Parts and Attributes, S. Maji, Workshop on Parts and Attributes, ECCV, 2012