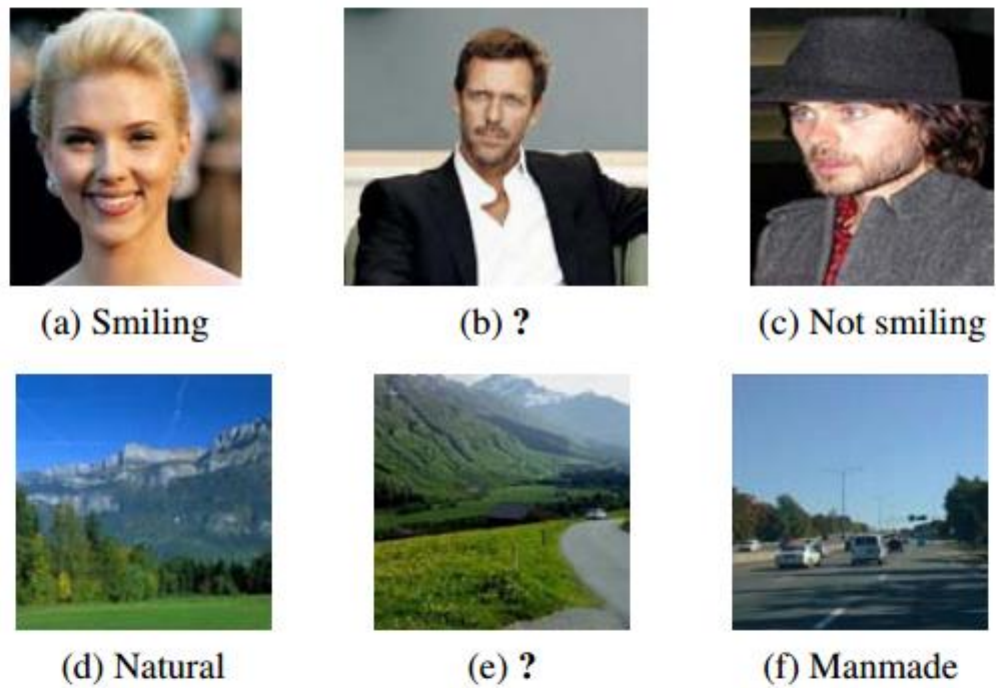
Relative Attributes

Devi Parikh, Kristen Grauman



(a) Smiling (b) ? (c) Not smiling

(d) Natural (e) ? (f) Manmade

Akanksha Saran

CS381V Paper Presentation

Outline

- Motivation
- Contributions
- Technical Details
- Experiments
- Discussion Points
- Extensions

Outline

- **Motivation**
- Contributions
- Technical Details
- Experiments
- Discussion Points
- Extensions

Horse

Slide Credit: Devi Parikh, Kristen Grauman


Donkey
Slide Credit: Devi Parikh, Kristen Grauman


Mule

Slide Credit: Devi Parikh, Kristen Grauman

Attributes

Is furry

Has four-legs

Legs shorter
than horses'

Tail longer
than donkeys'

Has tail

[Oliva 2001] [Ferrari 2007] [Lampert 2009] [Farhadi 2009] [Kumar 2009] [Wang 2009] [Wang 2010] [Berg 2010] [Branson 2010] [Parikh 2010] [ICCV 2011] …

Mule

Slide Credit: Devi Parikh, Kristen Grauman

Binary

Is furry

Has four-legs

Legs shorter than horses'

Tail longer than donkeys'

Has tail

Slide Credit: Devi Parikh, Kristen Grauman

Mule

Binary

Is furry

Has four-legs

Legs shorter than horses'

Tail longer than donkeys'

Has tail

Mule

Slide Credit: Devi Parikh, Kristen Grauman

Relative

Is furry

Has four-legs

Legs shorter than horses'

Tail longer than donkeys'

Has tail

Mule

Slide Credit: Devi Parikh, Kristen Grauman

Image Search

Slide Credit: Devi Parikh, Kristen Grauman

“Downtown Chicago”

Slide Credit: Devi Parikh, Kristen Grauman

Slide Credit: Devi Parikh, Kristen Grauman

Relative Description

Slide Credit: Devi Parikh, Kristen Grauman

Outline

- Motivation
- **Contributions**
- Technical Details
- Experiments
- Discussion Points
- Extensions

Contributions

- Relative attributes
  - Allow relating images and categories to each other
  - Learn ranking function for each attribute
- Novel applications
  - Zero-shot learning from attribute comparisons
  - Automatically generating relative image descriptions

Slide Credit: Devi Parikh, Kristen Grauman

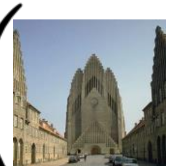
Outline

- Motivation
- Contributions
- **Technical Details**
- Experiments
- Discussion Points
- Extensions

Learning Relative Attributes

For each attribute a_m, open

Supervision is

$O_m : \{($  $\succ$  $), \dots \}$,

$S_m : \{\{$  $\sim$  $\}, \dots \}$

Slide Credit: Devi Parikh, Kristen Grauman

Learning Relative Attributes

Learn a scoring function $r_m(\boldsymbol{x_i}) = \boldsymbol{w_m^T x_i}$

Image features

Learned parameters

that best satisfies constraints:

$$\forall (i,j) \in O_m : \boldsymbol{w_m^T x_i} > \boldsymbol{w_m^T x_j}$$

$$\forall (i,j) \in S_m : \boldsymbol{w_m^T x_i} = \boldsymbol{w_m^T x_j}$$

Slide Credit: Devi Parikh, Kristen Grauman

Learning Relative Attributes

Max-margin learning to rank formulation

$$\min \quad \left(\frac{1}{2}||\boldsymbol{w_m^T}||_2^2 + C\left(\sum \xi_{ij}^2 + \sum \gamma_{ij}^2\right)\right)$$

$$\text{s.t} \quad \boldsymbol{w_m^T}(\boldsymbol{x_i} - \boldsymbol{x_j}) \geq 1 - \xi_{ij}, \forall (i,j) \in O_m$$

$$|\boldsymbol{w_m^T}(\boldsymbol{x_i} - \boldsymbol{x_j})| \leq \gamma_{ij}, \forall (i,j) \in S_m$$

$$\xi_{ij} \geq 0; \gamma_{ij} \geq 0$$

Based on [Joachims 2002]

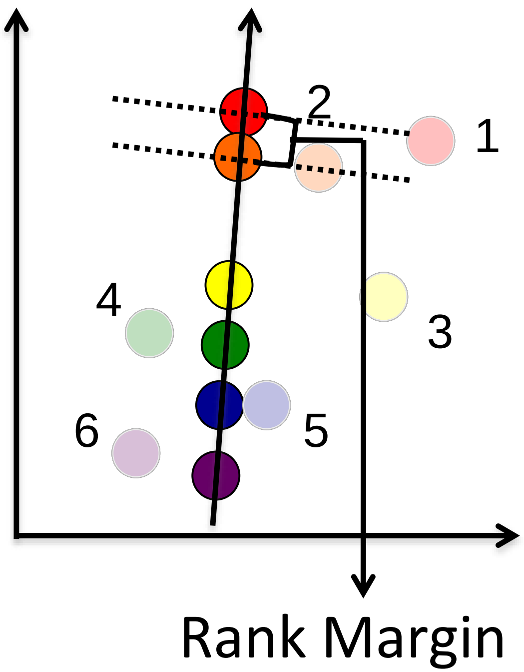


Image → Relative Attribute Score

Slide Credit: Devi Parikh, Kristen Grauman

Relative Zero-shot Learning

Training: Images from **S seen** categories and

Descriptions of **U unseen** categories

Age: **Hugh** $\succ$ **Clive** $\succ$ **Scarlett**

Jared $\succ$ **Miley**

Smiling: **Miley** $\succ$ **Jared**

Need not use all attributes, or all seen categories

Testing: Categorize image into one of **S**+**U** categories

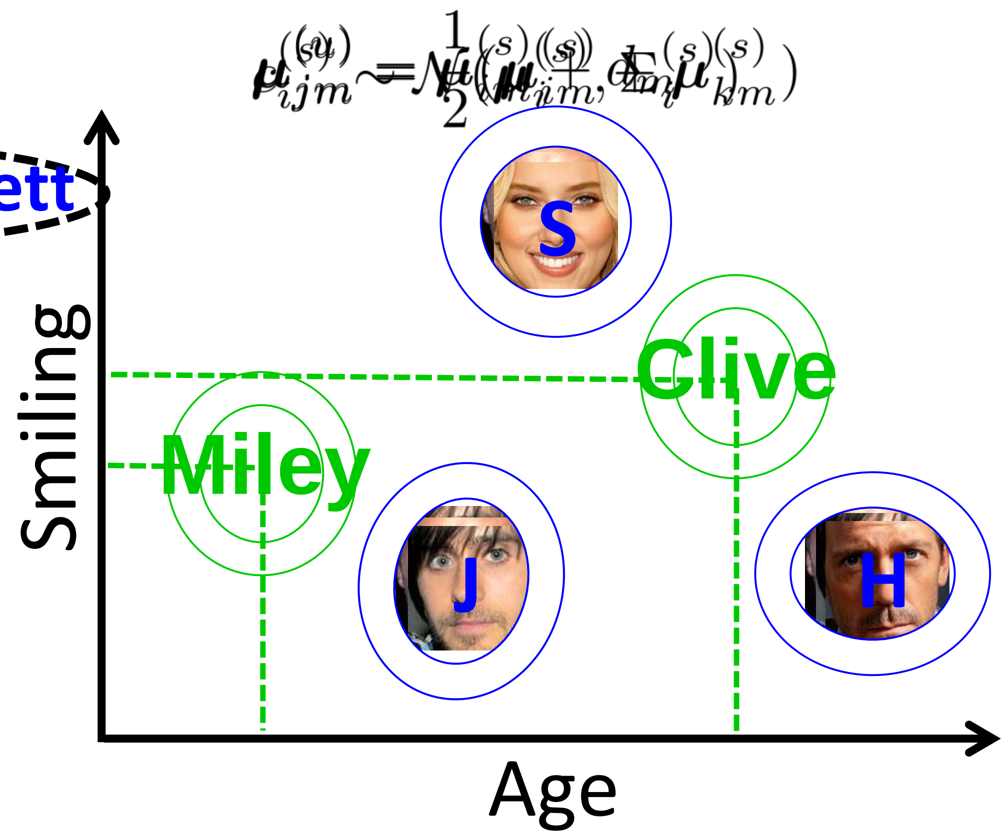
Slide Credit: Devi Parikh, Kristen Grauman

Relative Zero-shot Learning

Age: Hugh $\succ$ Clive $\succ$ Scarlett

Jared $\succ$ Miley

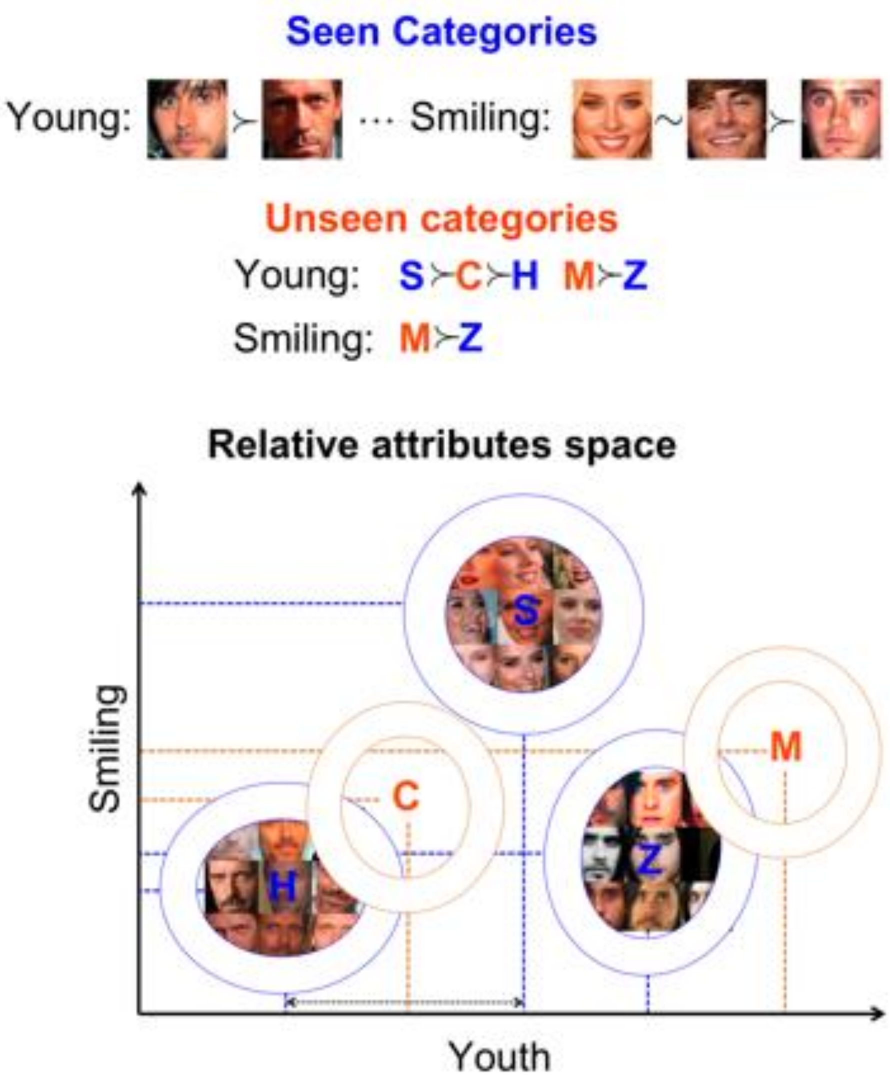
Smiling: Miley $\succ$ Jared



Infer image category using max-likelihood

Slide Credit: Devi Parikh, Kristen Grauman

Relative zero-shot learning



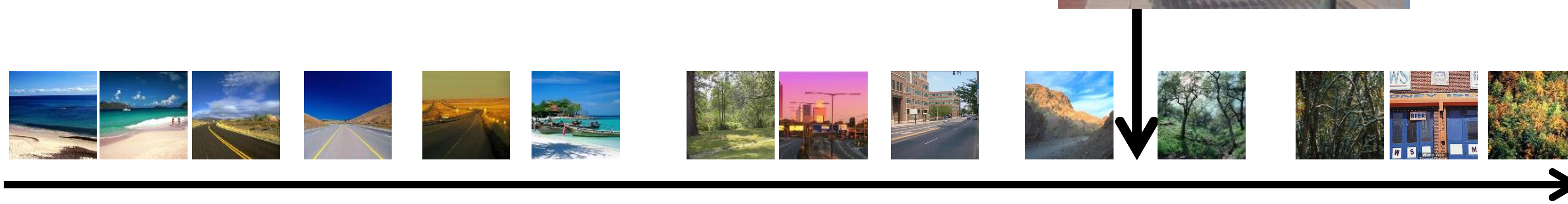
Can predict new classes based on their relationships to existing classes – without training images

Slide Credit: Devi Parikh, Kristen Grauman

Automatic Relative Image Description

Density  $\succ$ 

Novel image 



Conventional binary description: *not dense*

Dense: 

Not dense: 

Slide Credit: Devi Parikh, Kristen Grauman

Automatic Relative Image Description

Density

Novel image

more dense than

less dense than

Slide Credit: Devi Parikh, Kristen Grauman

Automatic Relative Image Description

Density

Novel image

more dense *than* ***Highways***, *less dense* *than* ***Forests***

Slide Credit: Devi Parikh, Kristen Grauman

Outline

- Motivation
- Contributions
- Technical Details
- **Experiments**
- Discussion Points
- Extensions

Datasets

Outdoor Scene Recognition (OSR)
[Oliva 2001]

8 classes, ~2700 images, Gist

6 attributes: open, natural, etc.

Public Figures Face (PubFig)
[Kumar 2009]

8 classes, ~800 images, Gist+color

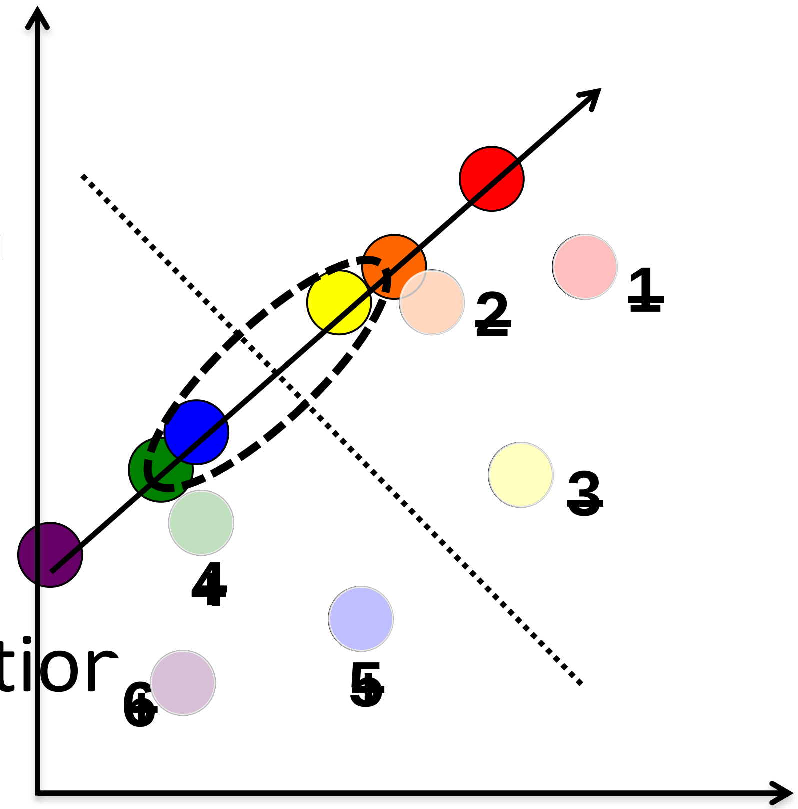
11 attributes: white, chubby, etc.

Attributes labeled at category level

Slide Credit: Devi Parikh, Kristen Grauman

Baselines

- Zero-shot learning
  - Binary attributes: Direct Attribute Prediction [Lampert 2009]
  - Relative attributes via classifier scores
- Automatic image-description
  - Binary attributes



Slide Credit: Devi Parikh, Kristen Grauman

Relative Zero-shot Learning

- Robustness:
  - Fewer comparisons to train relative attributes
  - More unseen (fewer seen) categories
- Flexibility in supervision:
  - ‘Looseness’ in description of unseen
  - Fewer attributes used to describe unseen

Slide Credit: Devi Parikh, Kristen Grauman

Relative Zero-shot Learning

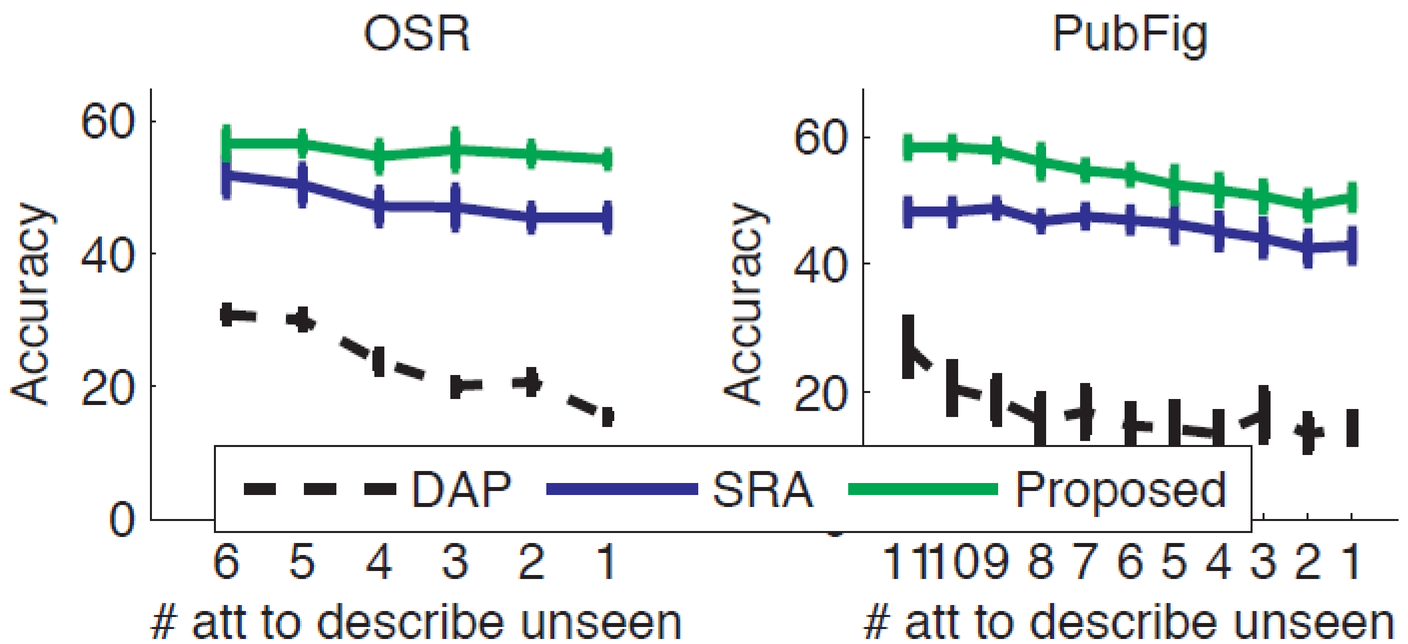


Figure 5. Zero-shot learning performance as fewer attributes are used to describe the unseen categories.

An attribute is more discriminative when used relatively

Slide Credit: Devi Parikh, Kristen Grauman

Automatic Relative Image Description

Binary (existing):

Not natural

Not open

Has perspective

Relative (proposed):

More natural than insidecity
Less natural than highway

More open than street
Less open than coast

Has more perspective than highway
Has less perspective than insidecity

Slide Credit: Devi Parikh, Kristen Grauman

Automatic Relative Image Description

Binary (existing):

Not natural

Not open

Has perspective

Relative (proposed):

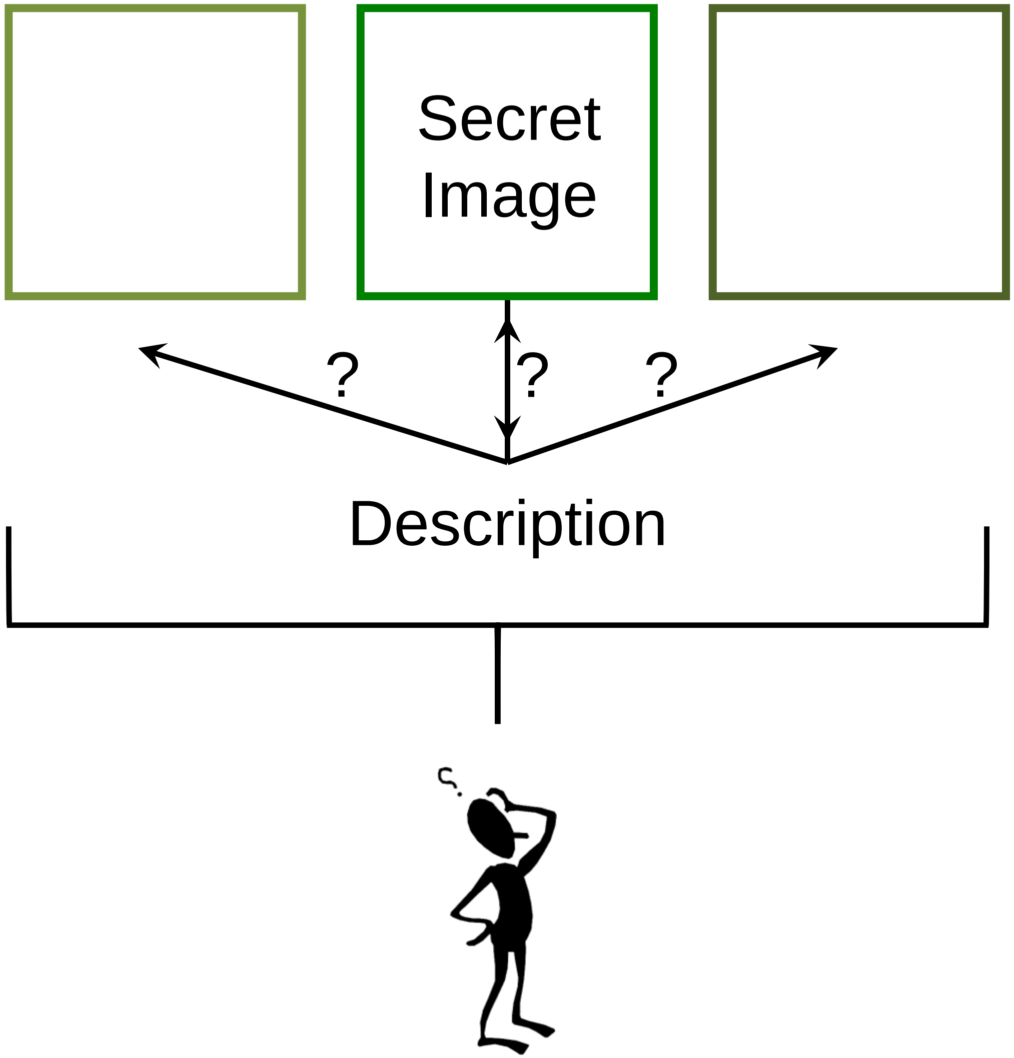
More natural than tallbuilding
Less natural than forest

More open than tallbuilding
Less open than coast

Has more perspective than tallbuilding

Slide Credit: Devi Parikh, Kristen Grauman

Human Studies:
Which Image is Being Described?

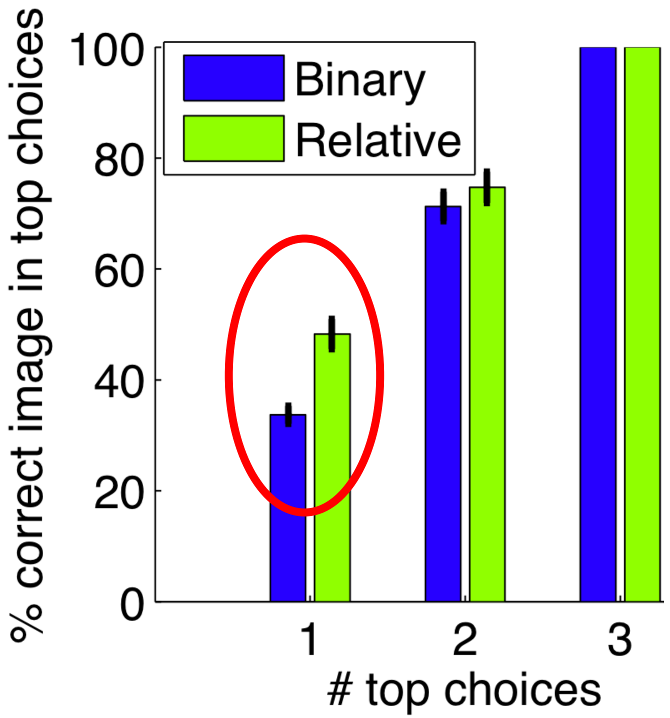


Slide Credit: Devi Parikh, Kristen Grauman

Automatic Relative Image Description

18 subjects

Test cases:
10OSR, 20 PubFig



Slide Credit: Devi Parikh, Kristen Grauman

Outline

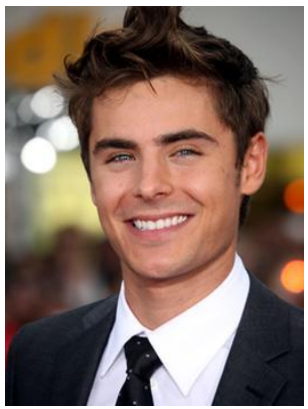
- Motivation
- Contributions
- Technical Details
- Experiments
- **Discussion Points**
- Extensions

Advantages

- Natural Descriptions: Leverages a natural mode of description
- Flexibility: Allows use of as many attributes for defining relations among as many classes

Image based based Attribute Ranking

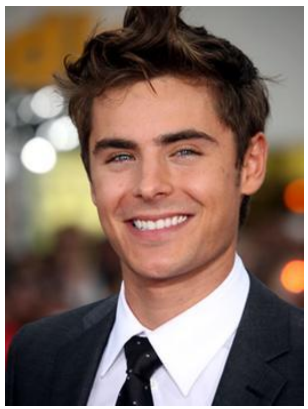
Relative ordering for attributes are transferred to all images in a category



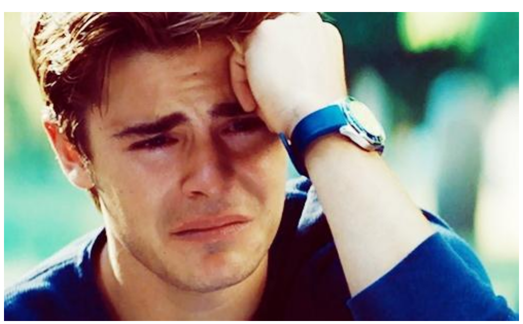
$\succ$

Image based based Attribute Ranking

Relative ordering for attributes are transferred to all images in a category



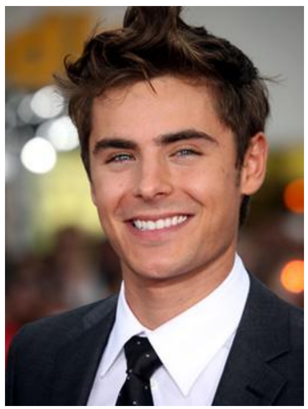
$\succ$



$\prec$

Image based based Attribute Ranking

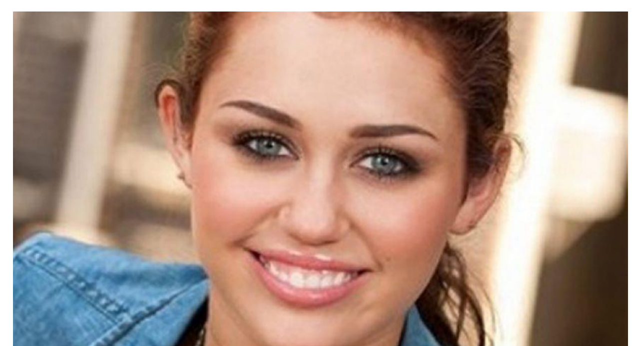
Relative ordering for attributes are transferred to all images in a category



$\succ$

Image Search

$\prec$



Gaussian distribution in joint attribute space

- Underlying distributions may be multi-modal

Fine-grained differences?

Can retaining the ranks for two very similar images/categories help identify them ?

male russet sparrow

male spanish sparrow

Outline

- Motivation
- Contributions
- Technical Details
- Experiments
- Strengths and Weaknesses
- **Extensions**

Extensions

- Relative attributes learned per image
  “Image Search with Interactive Feedback: Whittle Search”, A. Kovashka, D. Parikh, K. Grauman
- Active Learning of Discriminative Classifiers through feedback from users
  ”Simultaneous Active Learning of Classifiers & Attributes via Relative Feedback”, A. Biswas, D.Parikh
- Use of binary and relative attributes together
  ’ A horse has 4 legs’
- More expressive features instead of global features
  To discriminate a large set of image categories
  “Discovering Spatial Extent of Relative Attributes”, F.Xiao, Y.J. Lee
- Scalability to more categories and attribute labels

Thank you!