Attributes

Ricardo Martin

Describable visual attributes

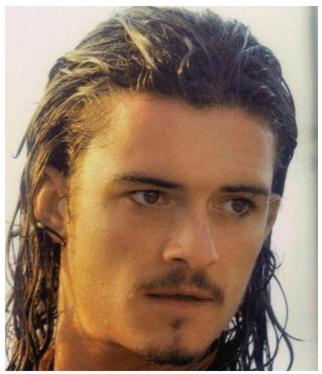
  

4-Legged	White	Male
Orange	Symmetric	Asian
Striped	Ionic columns	Beard
Furry	Classical	Smiling

Slide from N. Kumar

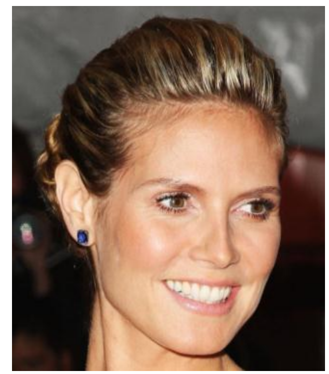
Attributes in Vision

- Face verification



male

≠



female

- Image description
  - Can be used for keyword search
- Zero-shot learning
  - Use descriptions for learning with no examples

Learn Attributes

- Label data on Mechanical Turk:

  - Is male?
  - Is asian?
  - Has glasses?

- Training SVM classifiers on features response
  - SIFT, HOG, LBP, RGB
- Learnt across many categories

Scalability issues

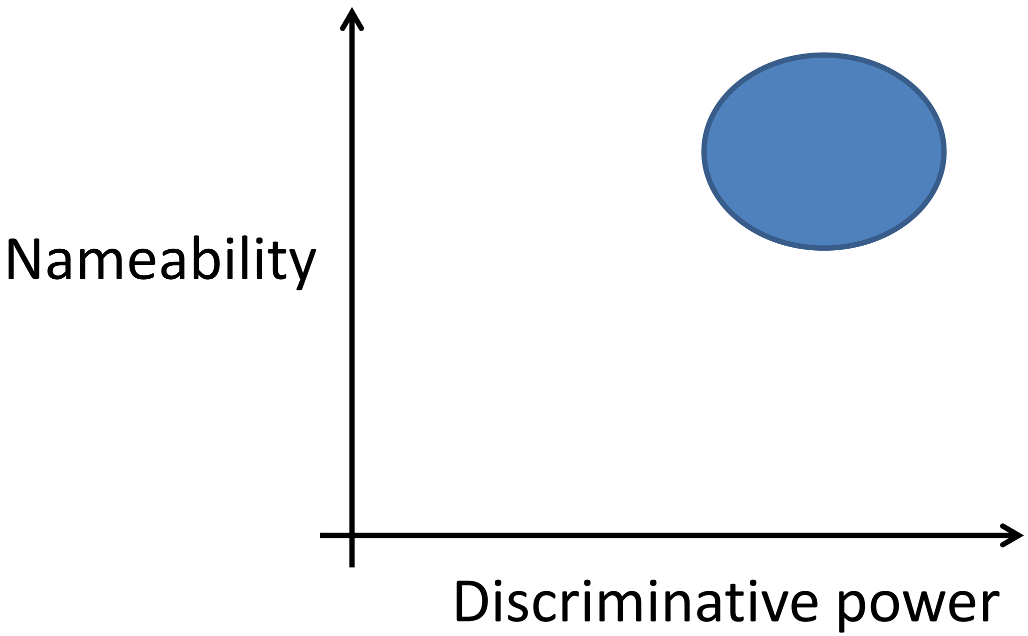
- Given an attribute it is easy to get labeled data on AMT.
- But, where do attributes come from?
  - Domain knowledge from experts
  - Expensive and not scalable

Attribute Discovery

- Nameable attribute discovery
- Discover attributes from product descriptions

Nameable Attribute Discovery

- Interested in nameable attributes, humans should understand them.



“Interactively Building a Discriminative Vocabulary of Nameable Attributes”, Devi Parikh and Kristen Grauman, CVPR 2011

Nameable Attribute Discovery

- Consider attributes as hyperplanes in feature space.
- Find attributes that separate best inter-class confusion.
- Send them to AMT to decide if an attribute is nameable or not.
  - There are too many attributes to consider
  - Model the manifold of nameable attributes and only send the probably nameable ones to AMT.

Nameable Attribute Discovery

“Congested”

“Smiling”

Attributes from Product Descriptions

The 12K pink and green gold leaves gently cascade down on these delicate beaded 10K earrings.

“Automatic Attribute Discovery and Characterization from Noisy Web Data”, Tamara L. Berg, Alexander C. Berg, and Jonathan Shih, ECCV 2010

Attributes from Product Descriptions

Rock and roll in these sexy, strapped heels from Report Signature. The smoldering Rockwell features a grey patent leather with pleated satin crossing at the open-toe atop a 1 inch platform, patent straps closing around the ankle with a gold buckled, and finally a 5 inch patent cone heel. Sizzle in these fierce mile-high shoes.

“Automatic Attribute Discovery and Characterization from Noisy Web Data”, Tamara L. Berg, Alexander C. Berg, and Jonathan Shih, ECCV 2010

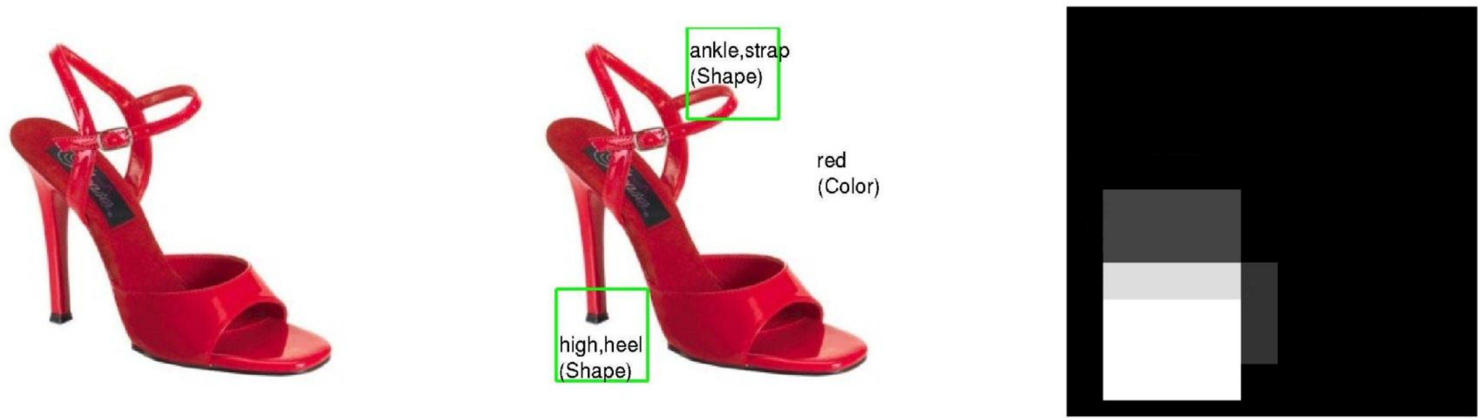
Attributes from Product Descriptions

- Find candidate attributes:
  - Learn visual models for certain words in the object description.
  - Take the ones that their visual models best predict the attribute in the description.
  - Cluster them based on mutual information
    - {stiletto, stiletto heel, sexy, traction, fabulous, styling}
    - {tote, handles, straps, lined, open}

Attributes from Product Descriptions

- Classify attributes:
  - Localizable:
    - Consider image as bag of regions.
    - Learn classifiers based on these bags with multiple instance learning.
    - If most probability of the learnt classifiers is in a small region, the attribute is localizable.
  - Attribute type:
    - Distinguish between color, texture and global attributes by learning on bags of feature types.

Attributes from Product Descriptions



Attribute limitations

- Many attributes are ambiguous

Is male?
Is asian?
Has long hair?
Has thin eyebrows?
Has round jaw?

Attribute limitations

- Many attributes are ambiguous

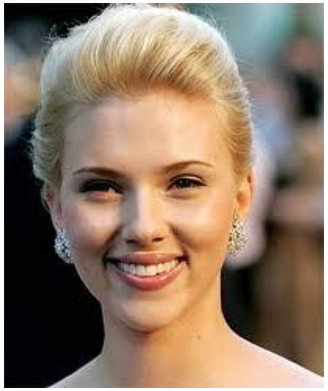
Is male?
Is asian?
Has long hair?
Has thin eyebrows?
Has round jaw?

Person A

Person B

Is A more male than B?
Is A more asian than B?
Has A longer hair than B?
Has A thinner eyebrows than B?
Has A a rounder jaw than B?

Relative attributes


Smiling


???


Not smiling


Natural


???


Not natural

“Relative Attributes”, Devi Parikh, Kristen Grauman, ICCV 2011

Slides from D. Parikh

Labeling data

Binary Attributes

Young: Yes
Smiling: No

Young: Yes
Smiling: Yes

Young: Yes
Smiling: Yes

Young: No
Smiling: Yes

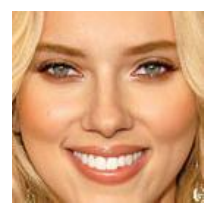
Young: Yes
Smiling: No

Relative Attributes

Young

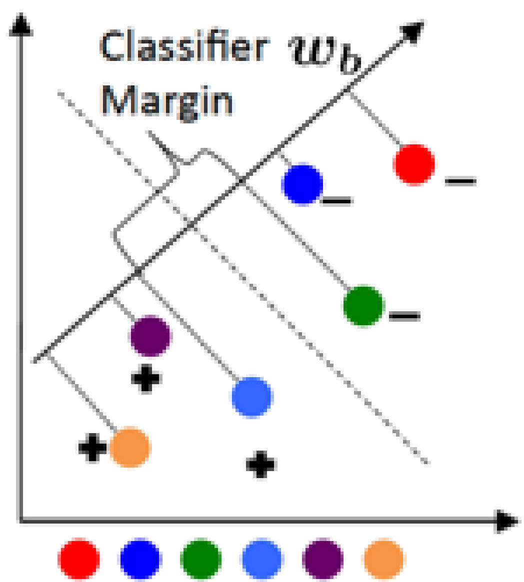
 $\succ$ 

Smiling

 $\sim$  $\succ$ 

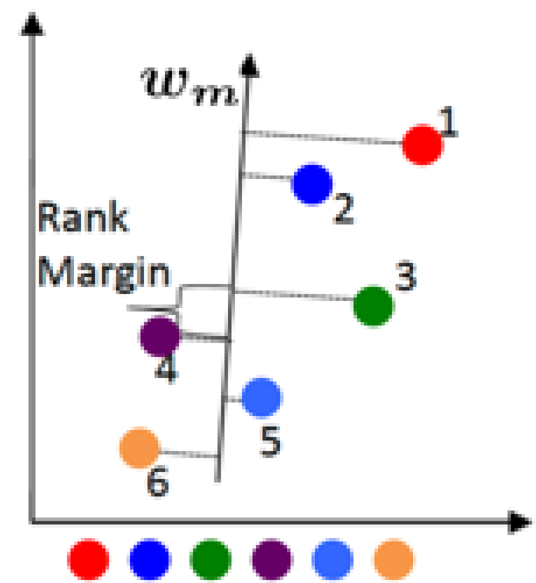
Learning relative attributes

Binary Attributes



Learn decision function
$$d_b(\boldsymbol{x}_i) = \boldsymbol{w}_b^T \boldsymbol{x}_i$$

Relative Attributes



Learn ranking function:
$$r_m(\boldsymbol{x}_i) = \boldsymbol{w}_m^T \boldsymbol{x}_i$$

Learning relative attributes

Binary Attributes

Conditions that hold:

- $d_b(\boldsymbol{x}_i) \geq 1$ for positive examples of attribute b (P_b).
- $d_b(\boldsymbol{x}_i) \leq -1$ for negative examples of attribute b (N_b).

Relative Attributes

Conditions that hold:

- $r_m(\boldsymbol{x}_i) > r_m(\boldsymbol{x}_j)$ for pairs (i, j) where attribute m is stronger in i (O_m).
- $r_m(\boldsymbol{x}_i) = r_m(\boldsymbol{x}_j)$ for pairs (i, j) where attribute m is similar in i and j (S_m).

Learning relative attributes

Binary Attributes

$$\begin{aligned} \min \quad & \tfrac{1}{2}\left\|\boldsymbol{w}_b^T\right\|_2^2 + C\sum(\xi_i + \gamma_i) \\ s.t. \quad & \boldsymbol{w}_b^T\boldsymbol{x}_i \geq 1 - \xi_i, \qquad \forall i \in P_b \\ & \boldsymbol{w}_b^T\boldsymbol{x}_i \leq -1 + \gamma_i, \quad \forall i \in N_b \\ & \xi_i \geq 0, \gamma_i \geq 0 \end{aligned}$$

- ξ_i and γ_i are slack variables
- C is a trade-off between maximizing the margin and satisfying the conditions.

Relative Attributes

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|\boldsymbol{w}_m^T\|_2^2 + C\sum\left(\xi_{ij}^2 + \gamma_{ij}^2\right) \\ \mathrm{s}.t. \quad & \boldsymbol{w}_m^T\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right) \geq 1 - \xi_{ij} \ \forall(i,j) \in O_m \\ & \left|\boldsymbol{w}_m^T\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)\right| \leq \gamma_{ij}, \quad \forall(i,j) \in S_m \\ & \xi_{ij} \geq 0, \gamma_{ij} \geq 0 \end{aligned}$$

- ξ_{ij} and γ_{ij} are slack variables
- C is a trade-off between maximizing the margin and satisfying the conditions.

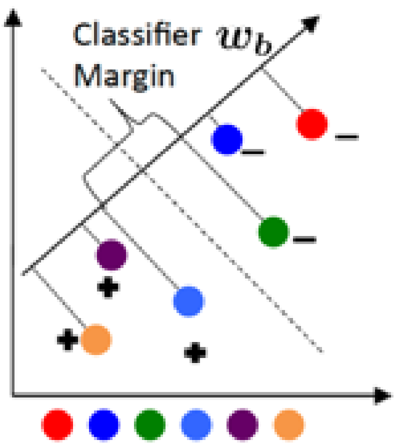
Learning relative attributes

Two possible ranks:

- Score-based Relative Attribute rank:

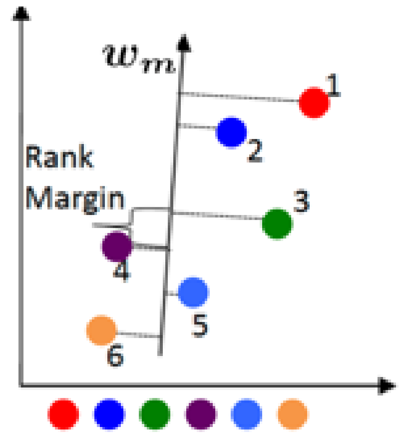
  $$r_b(\boldsymbol{x}_i) = d_b(\boldsymbol{x}_i) = \boldsymbol{w}_b^T \boldsymbol{x}_i$$

  Use the score of the classifier to measure the relative strength of the attribute



- Relative attribute rank:

  $$r_m(\boldsymbol{x}_i) = \boldsymbol{w}_m^T \boldsymbol{x}_i$$



Learning relative attributes

- Test on two datasets (scenes in OSR and faces in PubFig).

Scarlett Johansson | **Hugh Laurie** | **Miley Cyrus**

Binary Attributes

Young: **M, S**

Masculine: **H**

Smiling: **M, S**

Transfer from Category to Images

Not Smiling | Smiling | Smiling

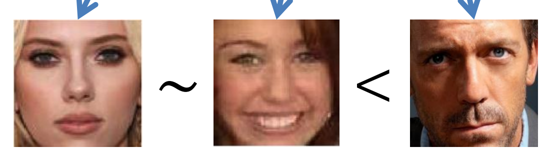
Relative Attributes

Young: **H** < **S** < **M**

Smiling: **H** < **S** < **M**

Masculine: **S** ~ **M** < **H**

Transfer from Category to Images



Learning relative attributes

- 8 categories in both datasets, 8 attributes in OSR and 11 attributes in PubFig
- Learn binary and relative attributes.
- Compare how well they rank images based on the defined partial ordering.

	Scene OSR	Faces PubFig
Score-based Relative Attribute rank	80%	67%
Relative attribute rank	89%	82%

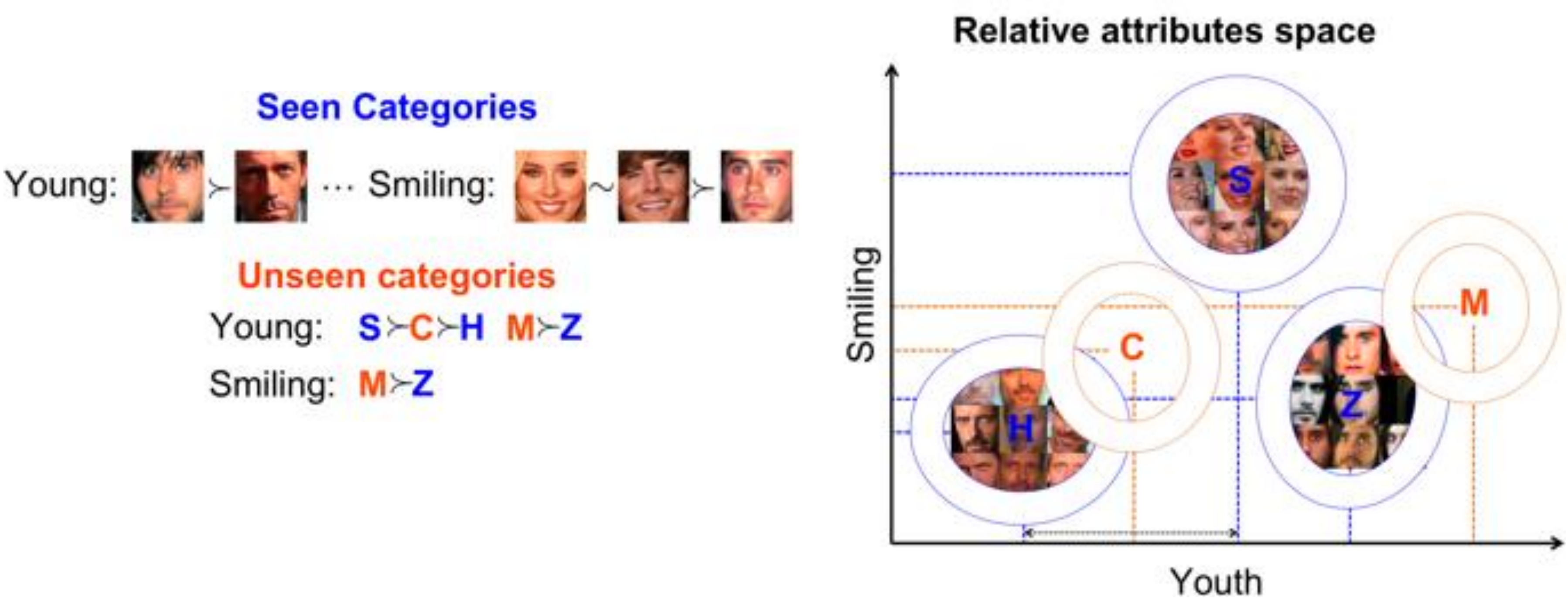
Applications

- Zero-shot learning
- Image description

Zero-shot learning

- Train relative attributes on a set of categories.
- Describe unseen categories with comparisons:
  - “bears are furrier than giraffes but less furry than rabbits”
  - “lions are larger than dogs, as large as tigers, but less large than elephants”
- Build a model based on the attributes for the unseen categories.
- Test the accuracy of that model.

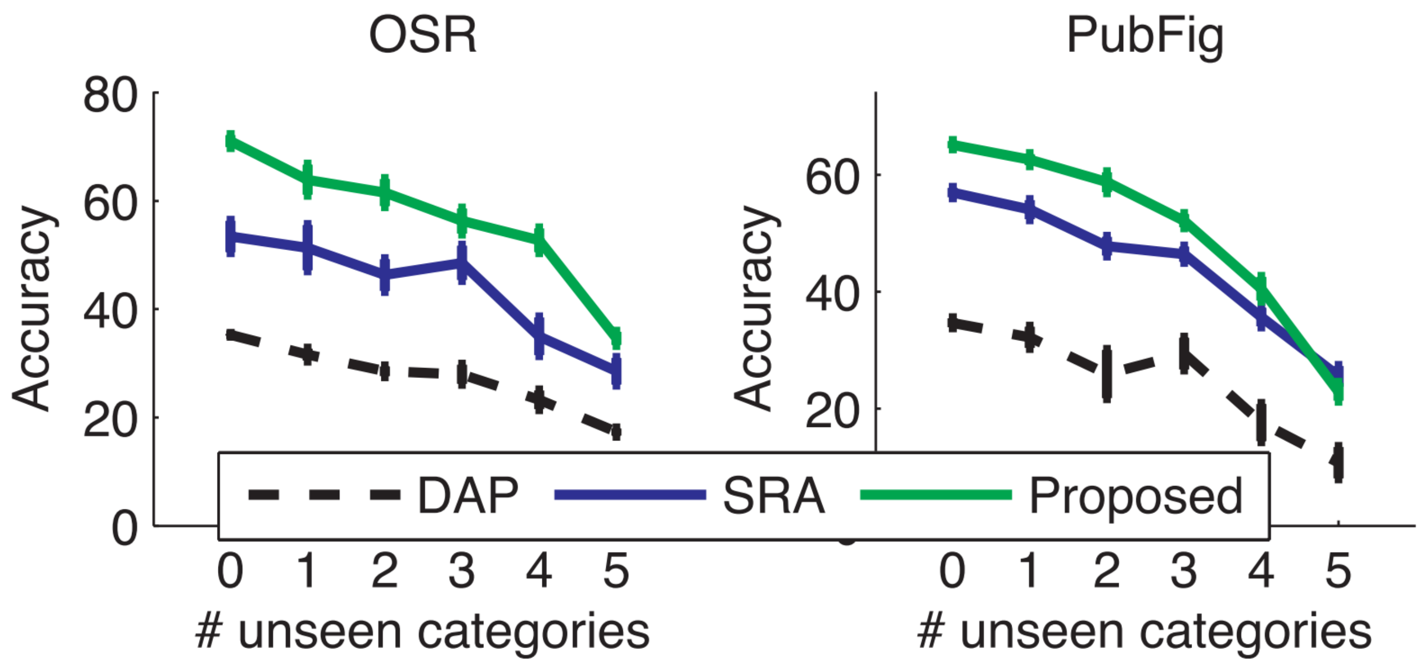
Zero-shot learning



Unseen categories are modeled as Gaussian distributions in attribute space constrained by the category definition.

Zero-shot learning

Proportion of unseen categories

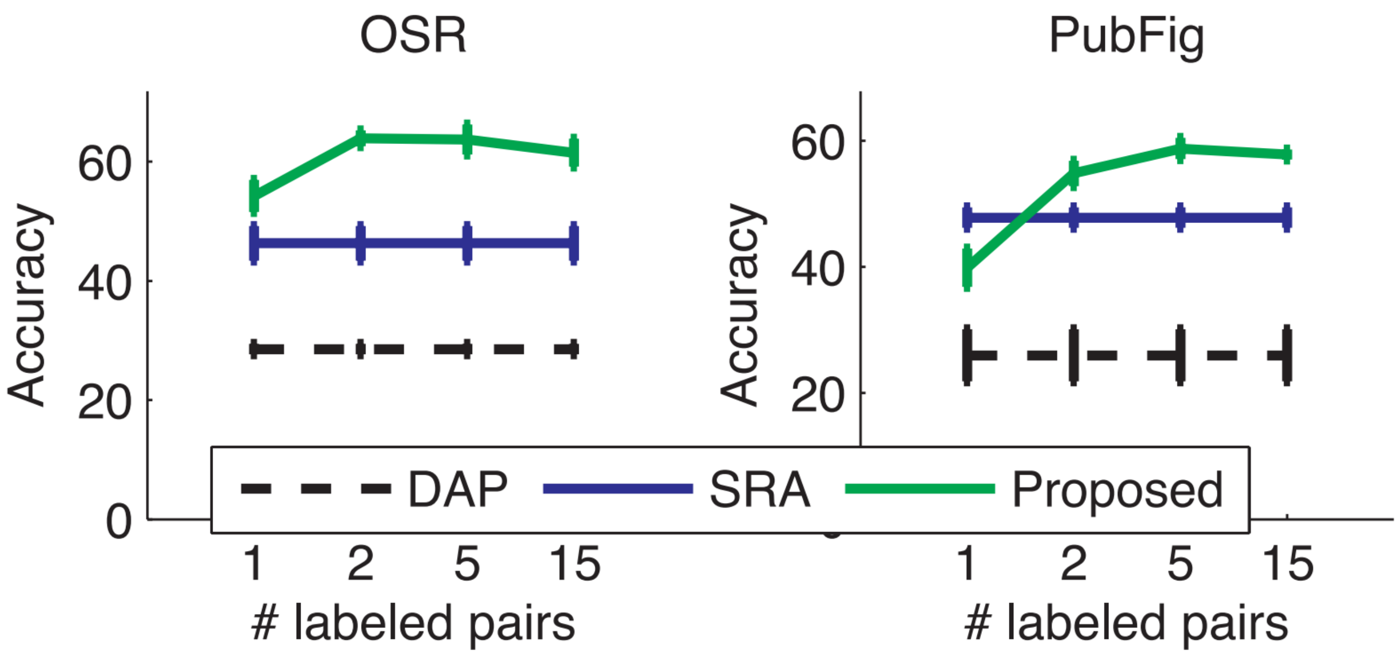


Baselines:

- SRA: Score-based Relative Attributes
- DAP: Direct Attribute Prediction
  "Learning to Detect Unseen Object Classes by Between-Class Attribute Transfer", Lampert et. al., CVPR 2009

Zero-shot learning

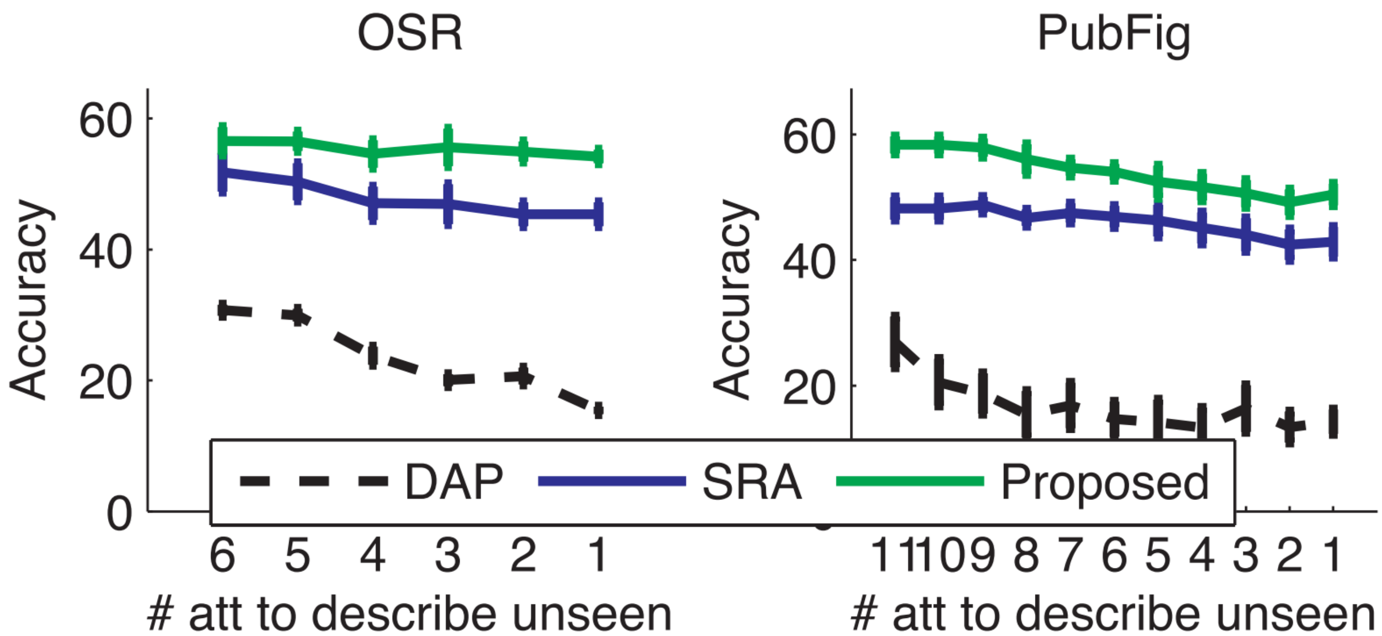
Amount of supervision



Unseen categories are described with respect to N labeled pairs of seen categories

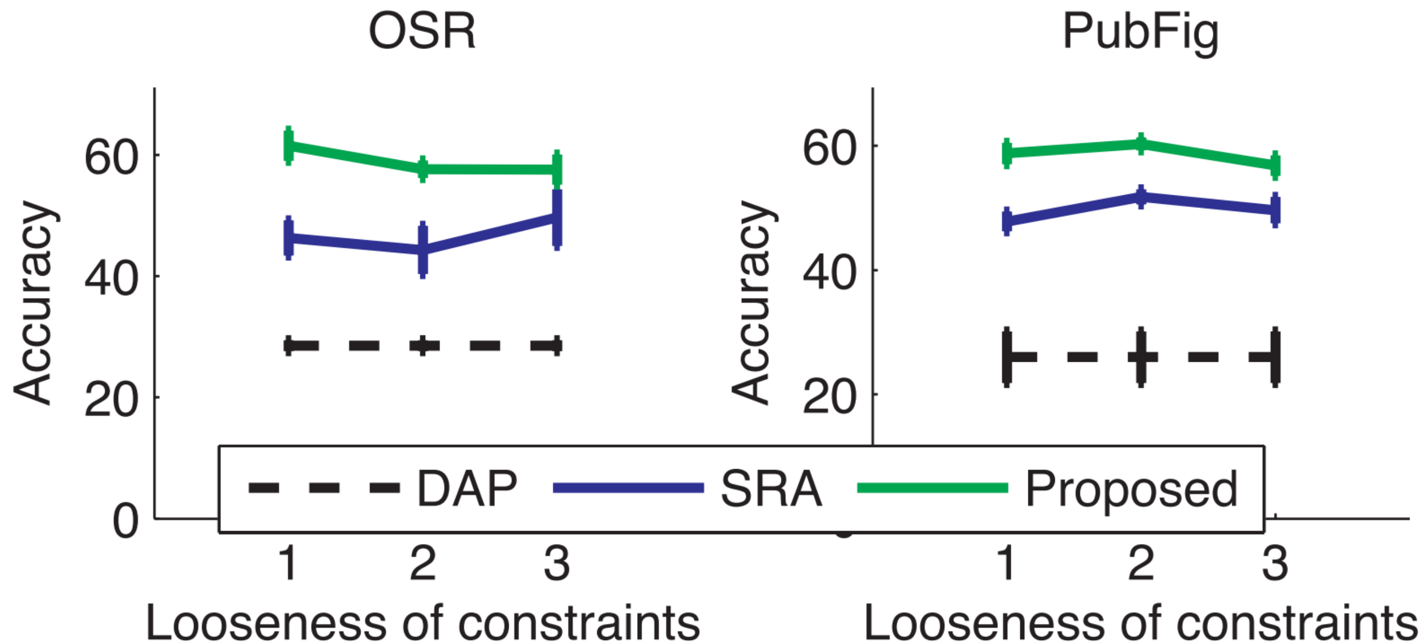
Zero-shot learning

Amount of supervision



Zero-shot learning

Amount of supervision



Attribute: Density

Image description

- Use relative attributes for image description:
  - For each attribute, look for two images that bound the image strength in the attribute
- Example: Generate description of: 

Relative attributes space

Density 

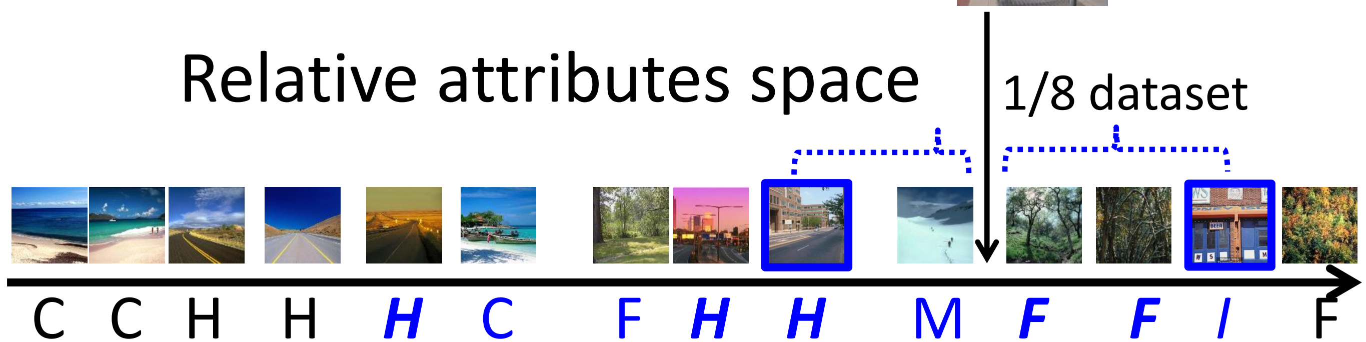
“more dense than , less dense than ”

Image description

- Contrast to binary description:

Dense

- Use categories instead of images:



Density

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

Image description

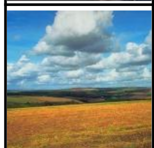
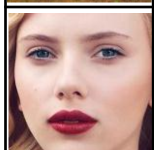
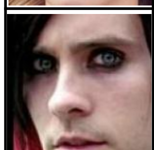
Image	Binary descriptions	Relative descriptions
	not natural, not open, perspective	more natural than tallbuilding; less natural than forest; more open than tallbuilding; less open than coast; more perspective than tallbuilding;
	not natural, not open, perspective	more natural than insidecity; less natural than highway; more open than street; less open than coast; more perspective than highway; less perspective than insidecity
	natural, open, perspective	more natural than tallbuilding; less natural than mountain; more open than mountain; less perspective than opencountry;
	White, not Smiling, VisibleForehead	more White than AlexRodriguez; more Smiling than JaredLeto; less Smiling than ZacEfron; more VisibleForehead than JaredLeto; less VisibleForehead than MileyCyrus
	White, not Smiling, not VisibleForehead	more White than AlexRodriguez; less White than MileyCyrus; less Smiling than HughLaurie; more VisibleForehead than ZacEfron; less VisibleForehead than MileyCyrus
	not Young, BushyEyebrows, RoundFace	more Young than CliveOwen; less Young than ScarlettJohansson; more BushyEyebrows than ZacEfron; less BushyEyebrows than AlexRodriguez; more RoundFace than CliveOwen; less RoundFace than ZacEfron

Image description

- User study to evaluate descriptions:

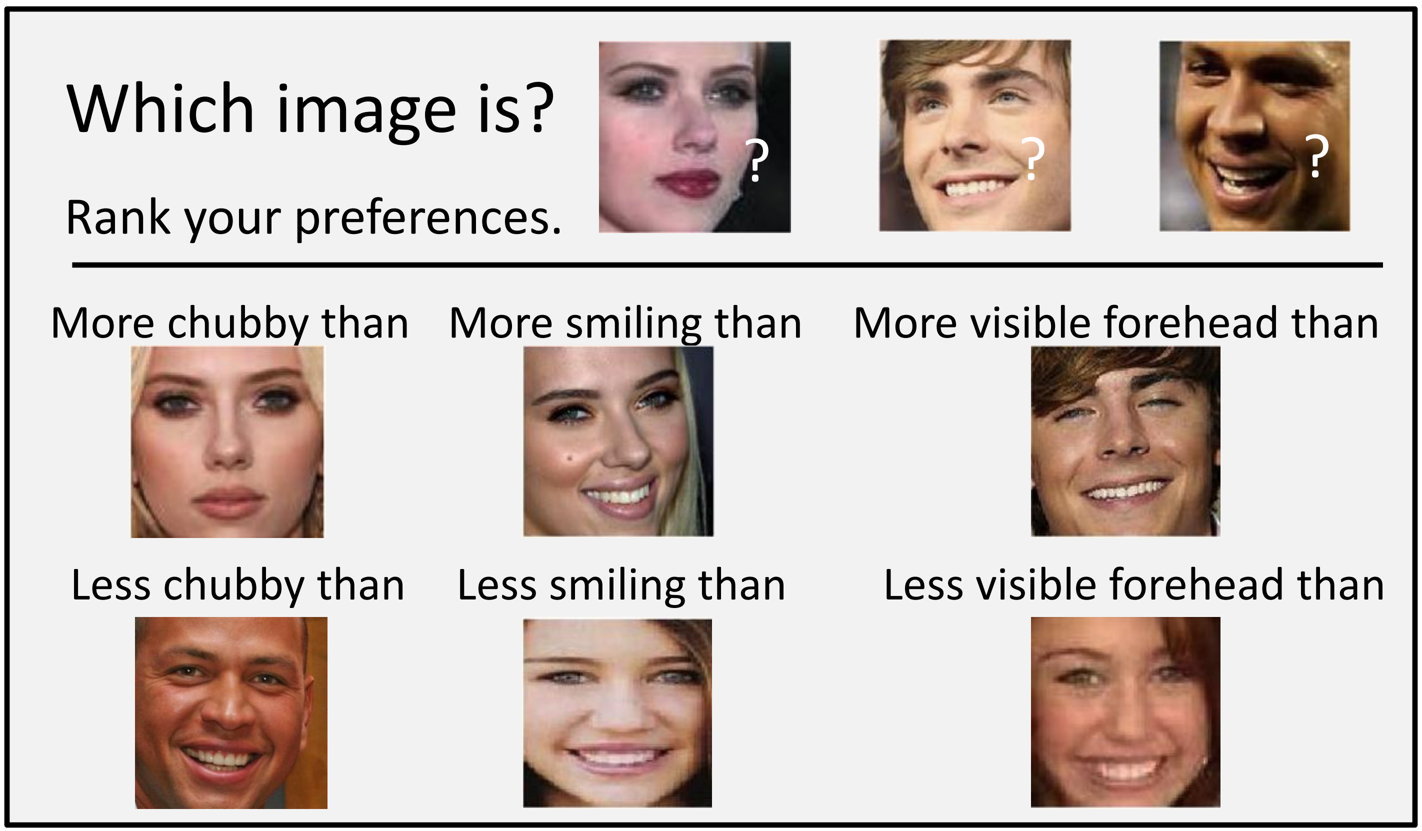


Image description

- User study to evaluate descriptions:

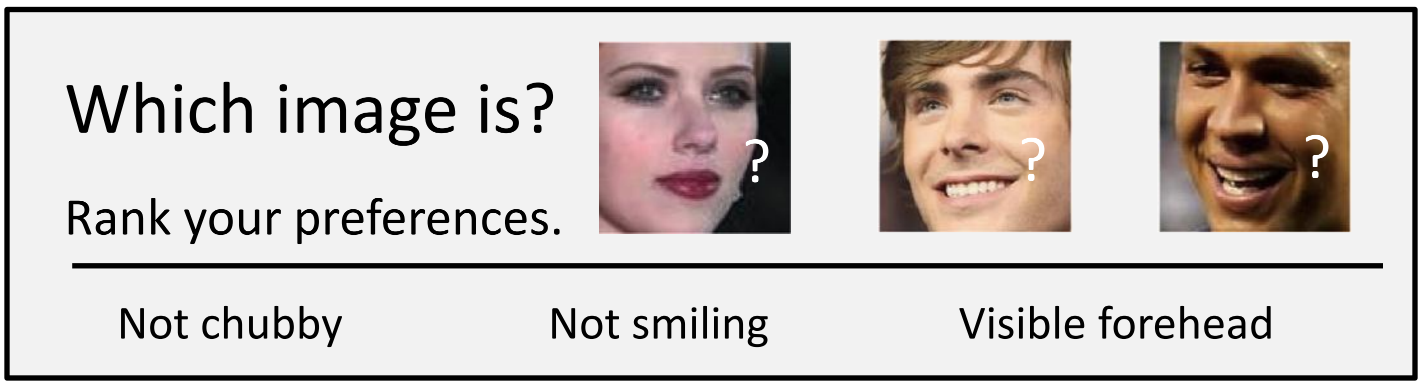
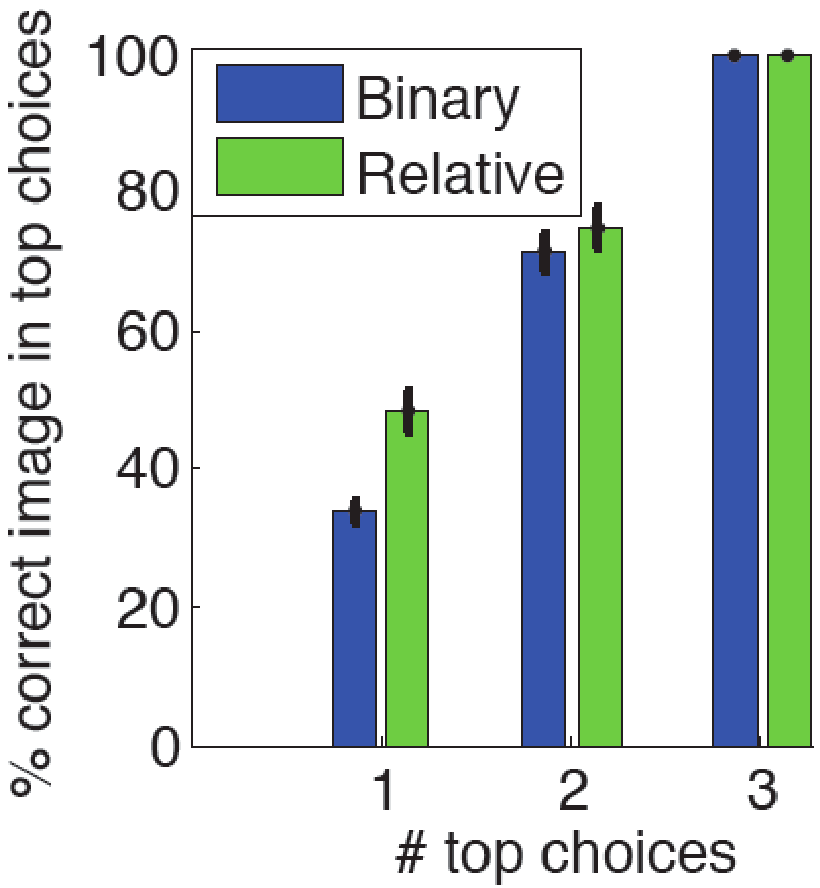


Image description



Conclusion & Discussion

- Relative attributes allow richer supervision and description than categorical attributes.
- Is it valid to transfer attribute labels from categories to images?
- When should an attribute be relative and when categorical?
- Are relative attributes unambiguous enough?