Relative Attributes

Shuangfei Fan

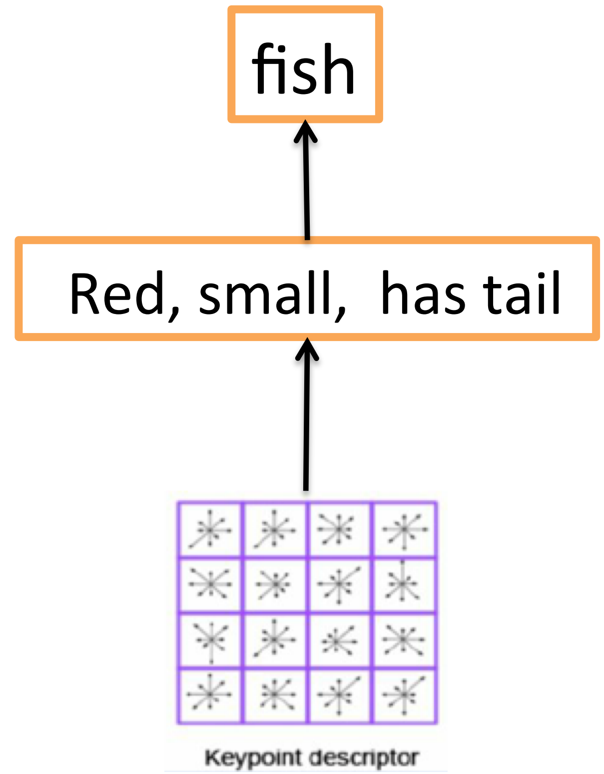
Electrical & Computer Engineering
Virginia Tech

OVERVIEW

- Introduction
- Learning Relative Attributes
- Relative Zero-shot Learning
- Automatic Relative Image Description
- Datasets
- Experiments
- Conclusion

What are Attributes

- ❖ Low-level concepts: features
- ❖ High-level concepts: labels, categories
- ❖ Mid-level concepts: attributes
    - ❖ Shared across categories
    - ❖ Have semantic meanings
    - ❖ Visual concepts (machine detectable)



Why Attributes?

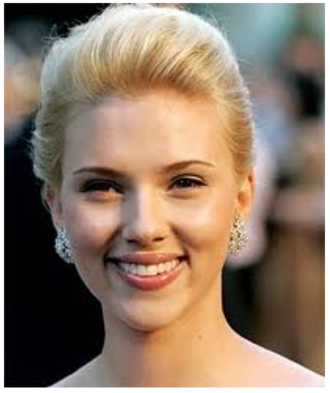
❖ How humans naturally describe natural concepts

❖ Image search

❖ Describe unknown objects

Has Horn
Has leg
Has Head
Has Wool

Relative Attributes

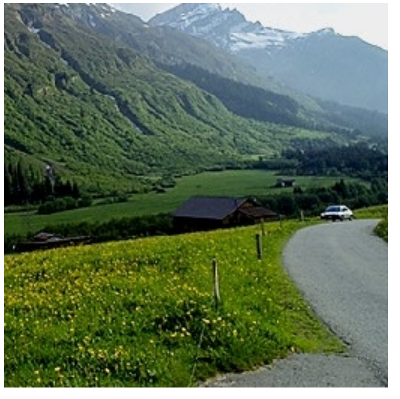


Smiling

???

Not smiling

Natural

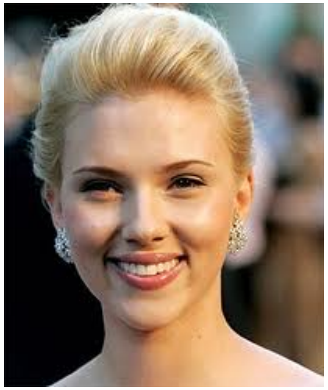


???

Not natural

Figure Credit: Devi Parikh

Relative Attributes

Smiling  >  > 

Natural  >  > 

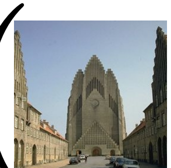
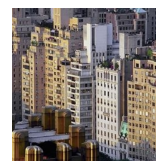
OVERVIEW

- Introduction
- **Learning Relative Attributes**
- Relative Zero-shot Learning
- Automatic Relative Image Description
- Datasets
- Experiments
- Conclusion

Learning Relative Attributes

For each attribute a_m, open

Supervision is

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

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

Slide Credit: Devi Parikh

Learning Relative Attributes

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

that best satisfies constraints:

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

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

Learning Relative Attributes

Max-margin learning to rank formulation

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

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

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

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

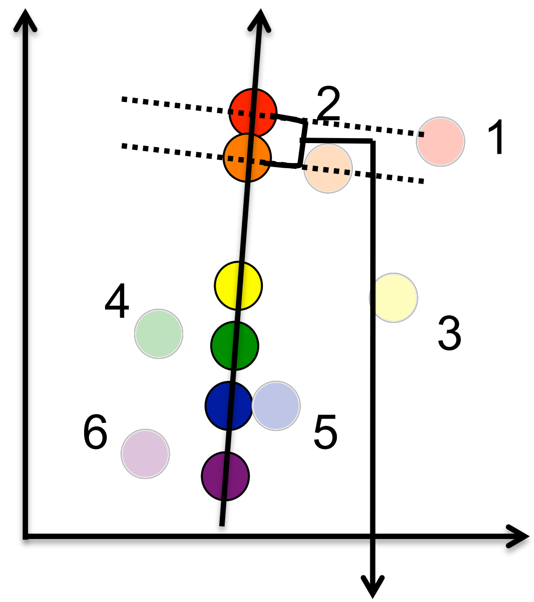


Image → Relative Attribute Score

Slide Credit: Devi Parikh

OVERVIEW

- Introduction
- Learning Relative Attributes
- **Relative Zero-shot Learning**
- Automatic Relative Image Description
- Datasets
- Experiments
- Conclusion

Zero-shot Learning

- Recognize the Wampimuk
  - Impossible?
- Solution: semantic tranfer
  - Wampimuk: small, horn, furry, cute
- Zero-Shot:
  - Pattern recognition with no training examples
  - Solved by semantic transfer

Slide Credit: Timothy Hospedales

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

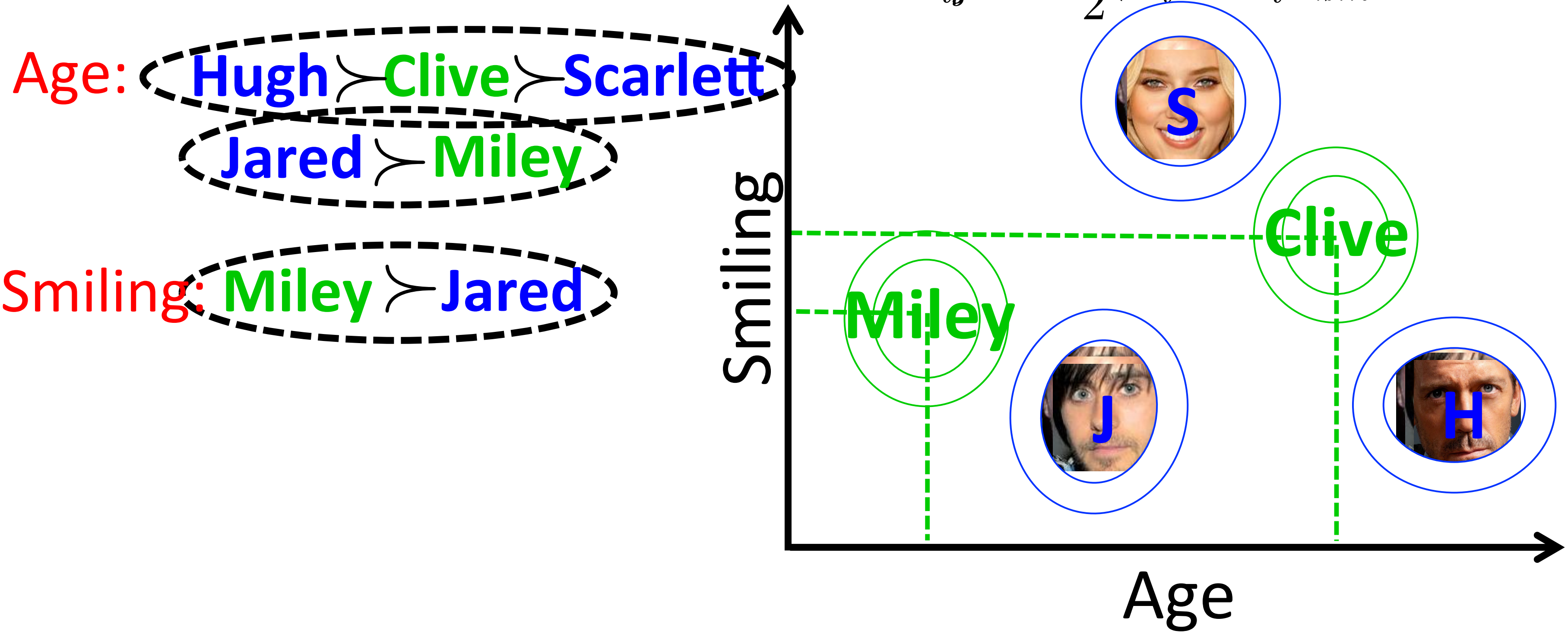
Testing: Categorize image into one of **S**+**U** categories

Slide Credit: Devi Parikh

Relative Zero-shot Learning

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

$$\boldsymbol{\mu}_{jm}^{(u)} = \frac{1}{2}\left(\boldsymbol{\mu}_{im}^{(s)} + \boldsymbol{\mu}_{km}^{(s)}\right)$$



Infer image category using max-likelihood

$$c^* = \underset{j \in \{1,\ldots,N\}}{\operatorname{argmax}} \; P\left(\tilde{\boldsymbol{x}}_i \mid \boldsymbol{\mu}_j, \Sigma_j\right)$$

Slide Credit: Devi Parikh

OVERVIEW

- Introduction
- Learning Relative Attributes
- Relative Zero-shot Learning
- **Automatic Relative Image Description**
- Datasets
- Experiments
- Conclusion

Automatic Relative Image Description

Density 

Novel image 

Conventional binary description: *not dense*

Dense: 

Not dense: 

Slide Credit: Devi Parikh

Automatic Relative Image Description

Density

Novel image

more dense than

less dense than

Slide Credit: Devi Parikh

Automatic Relative Image Description

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

Slide Credit: Devi Parikh

OVERVIEW

- Introduction
- Learning Relative Attributes
- Relative Zero-shot Learning
- Automatic Relative Image Description
- **Datasets**
- Experiments
- Conclusion

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

Datasets

Category level annotation

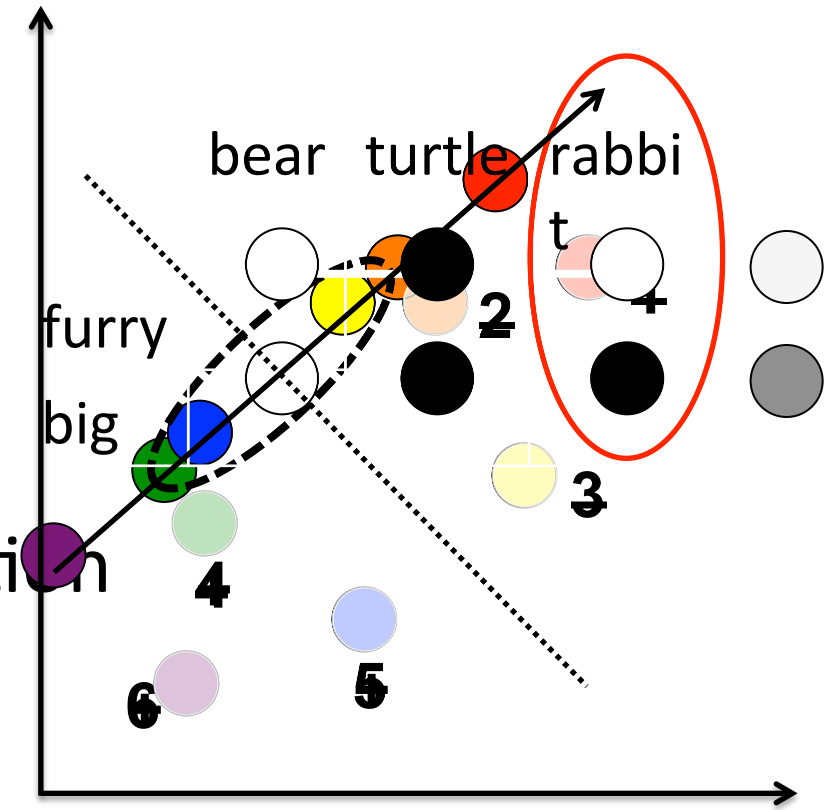
	Binary	Relative
OSR	TI SHC OMF	
natural	0 0 0 0 1 1 1 1	$T \prec I \sim S \prec H \prec C \sim O \sim M \sim F$
open	0 0 0 1 1 1 1 0	$T \sim F \prec I \sim S \prec M \prec H \sim C \sim O$
perspective	1 1 1 1 0 0 0 0	$O \prec C \prec M \sim F \prec H \prec I \prec S \prec T$
large-objects	1 1 1 0 0 0 0 0	$F \prec O \sim M \prec I \sim S \prec H \sim C \prec T$
diagonal-plane	1 1 1 1 0 0 0 0	$F \prec O \sim M \prec C \prec I \sim S \prec H \prec T$
close-depth	1 1 1 1 0 0 0 1	$C \prec M \prec O \prec T \sim I \sim S \sim H \sim F$
PubFig	ACHJ MS V Z	
Masculine-looking	1 1 1 1 0 0 1 1	$S \prec M \prec Z \prec V \prec J \prec A \prec H \prec C$
White	0 1 1 1 1 1 1 1	$A \prec C \prec H \prec Z \prec J \prec S \prec M \prec V$
Young	0 0 0 0 1 1 0 1	$V \prec H \prec C \prec J \prec A \prec S \prec Z \prec M$
Smiling	1 1 1 0 1 1 0 1	$J \prec V \prec H \prec A \sim C \prec S \sim Z \prec M$
Chubby	1 0 0 0 0 0 0 0	$V \prec J \prec H \prec C \prec Z \prec M \prec S \prec A$
Visible-forehead	1 1 1 0 1 1 1 0	$J \prec Z \prec M \prec S \prec A \sim C \sim H \sim V$
Bushy-eyebrows	0 1 0 1 0 0 0 0	$M \prec S \prec Z \prec V \prec H \prec A \prec C \prec J$
Narrow-eyes	0 1 1 0 0 0 1 1	$M \prec J \prec S \prec A \prec H \prec C \prec V \prec Z$
Pointy-nose	0 0 1 0 0 0 0 1	$A \prec C \prec J \sim M \sim V \prec S \prec Z \prec H$
Big-lips	1 0 0 0 1 1 0 0	$H \prec J \prec V \prec Z \prec C \prec M \prec A \prec S$
Round-face	1 0 0 0 1 1 0 0	$H \prec V \prec J \prec C \prec Z \prec A \prec S \prec M$

OVERVIEW

- Introduction
- Learning Relative Attributes
- Relative Zero-shot Learning
- Automatic Relative Image Description
- Datasets
- **Experiments**
- Conclusion

Experiments: Baselines

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



Slide Credit: Devi Parikh

Experiments: 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

Experiments: Zero-shot learning

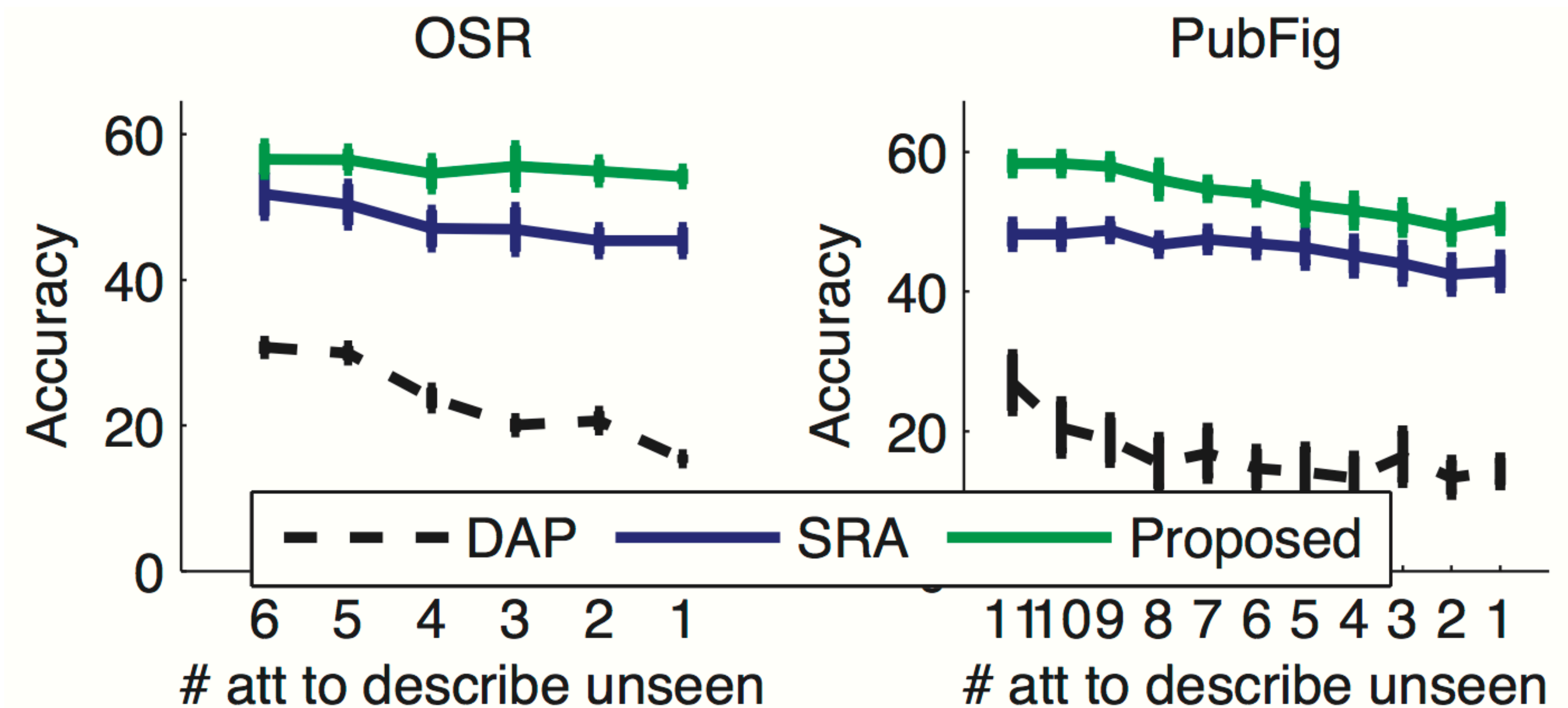


Figure 5. Zero-shot learning performance as fewer attributes are used to describe the unseen categories.

Experiments: Describe images

Binary attribute:

Not natural
Not open
Has perspective

Relative attribute:

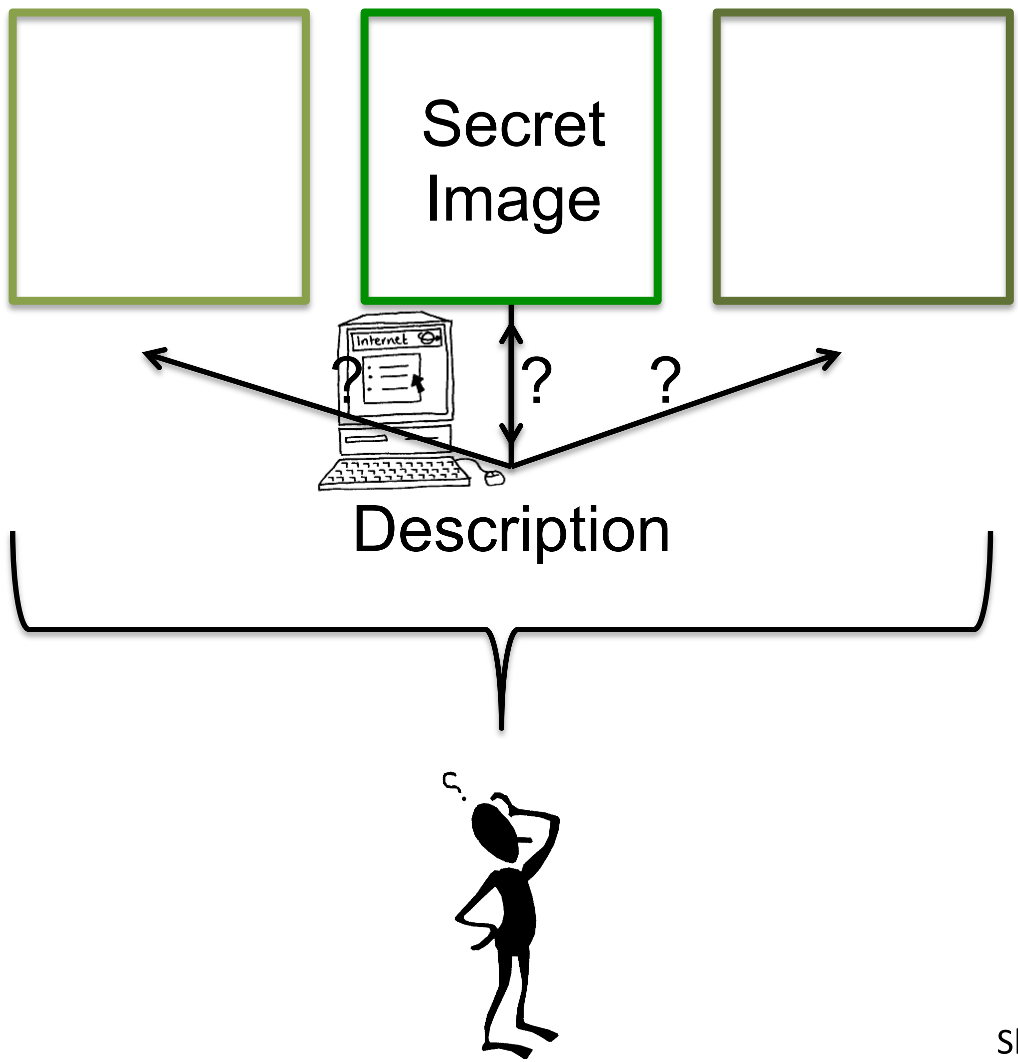
More natural than insidecity
Less natural than highway

More open than street
Less open than coast

Has more perspective than highway
Has less perspective than insidecity

Experiments: Describe images

Human Studies: Which Image is Being Described?



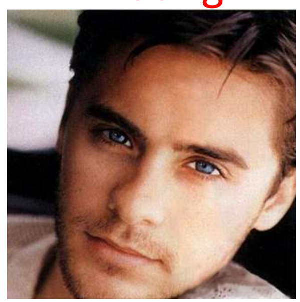
Slide Credit: Devi Parikh

Experiments: Describe images

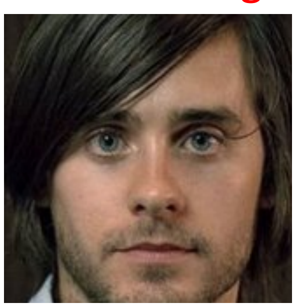
Binary: Smiling, Young

Smiling

Young



Not Smiling



Not Young

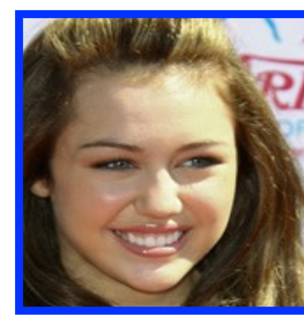
Relative

More Smiling than

Younger than

Less Smiling than

Older than



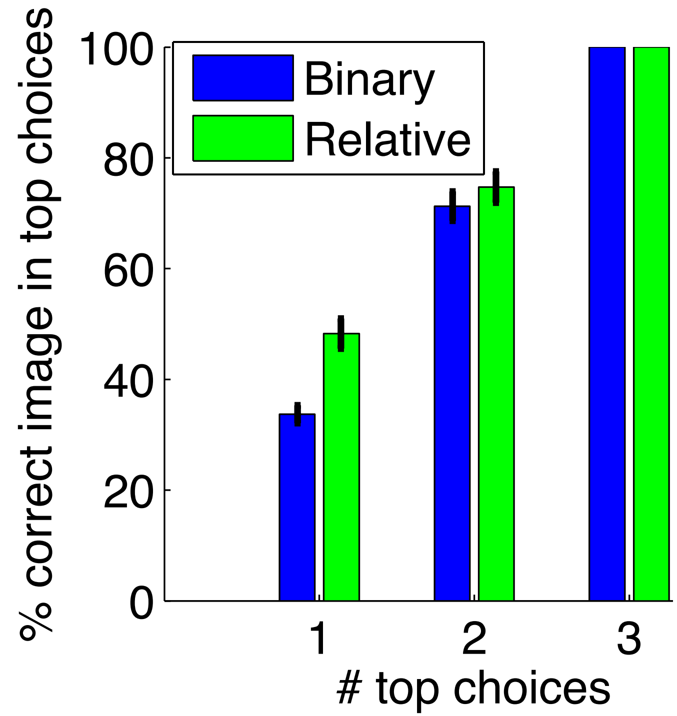
Slide Credit: Devi Parikh

Experiments: Describe images

Human Studies: Which Image is Being Described?

18 subjects

Test cases:
10 OSR, 20 PubFig



Slide Credit: Devi Parikh

OVERVIEW

- Introduction
- Learning Relative Attributes
- Relative Zero-shot Learning
- Automatic Relative Image Description
- Datasets
- Experiments
- **Conclusion**

Conclusion

- Relative attributes
  - Allow relating images and categories to each other
  - Learn ranking function for each attribute

- Novel applications
  - Natural and accurate zero-shot learning from attribute comparisons
  - Automatically generating precise relative image descriptions for human interpretation

Questions?

BACKUP

Experiments: Zero-shot learning

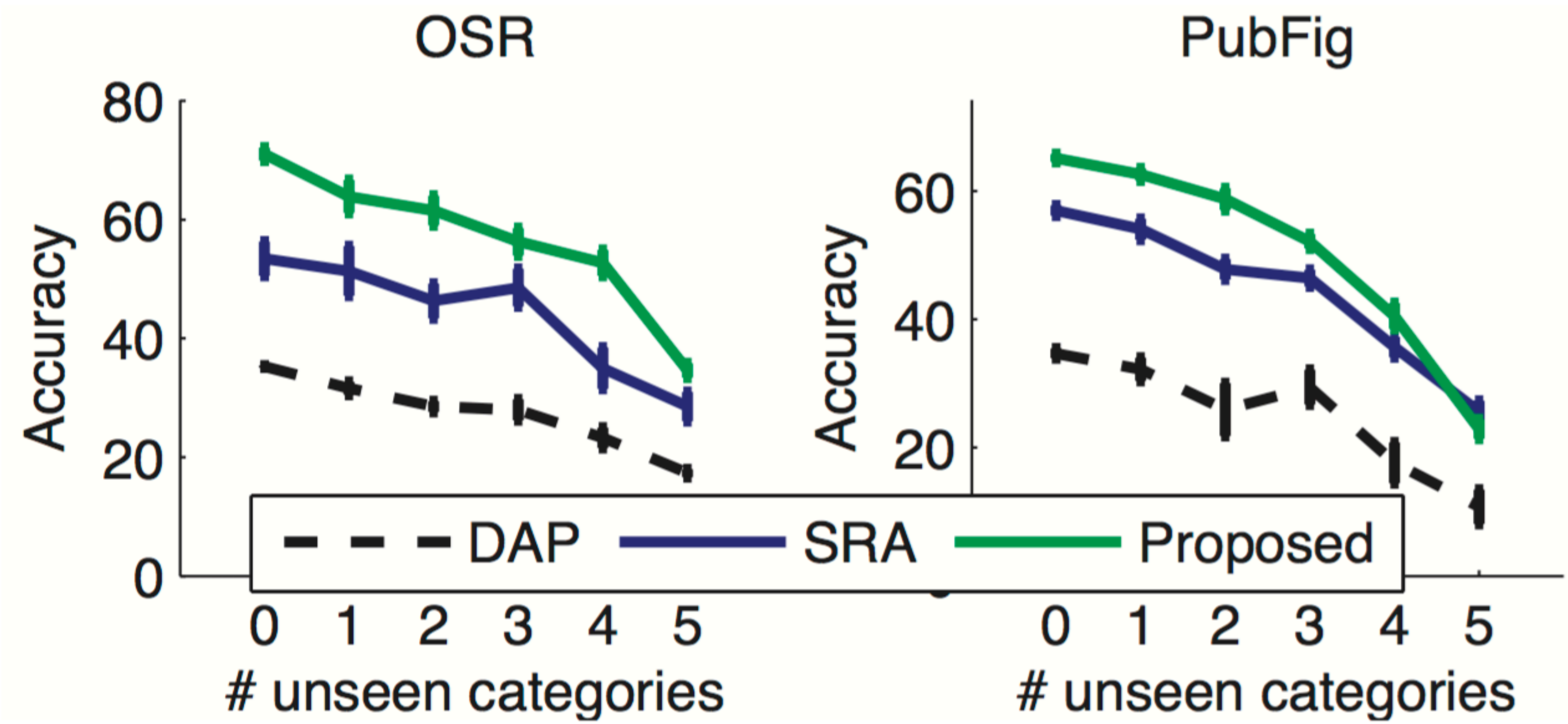


Figure 3. Zero-shot learning performance as the proportion of unseen categories increases. Total number of classes N remains constant at 8.

Experiments: Zero-shot learning

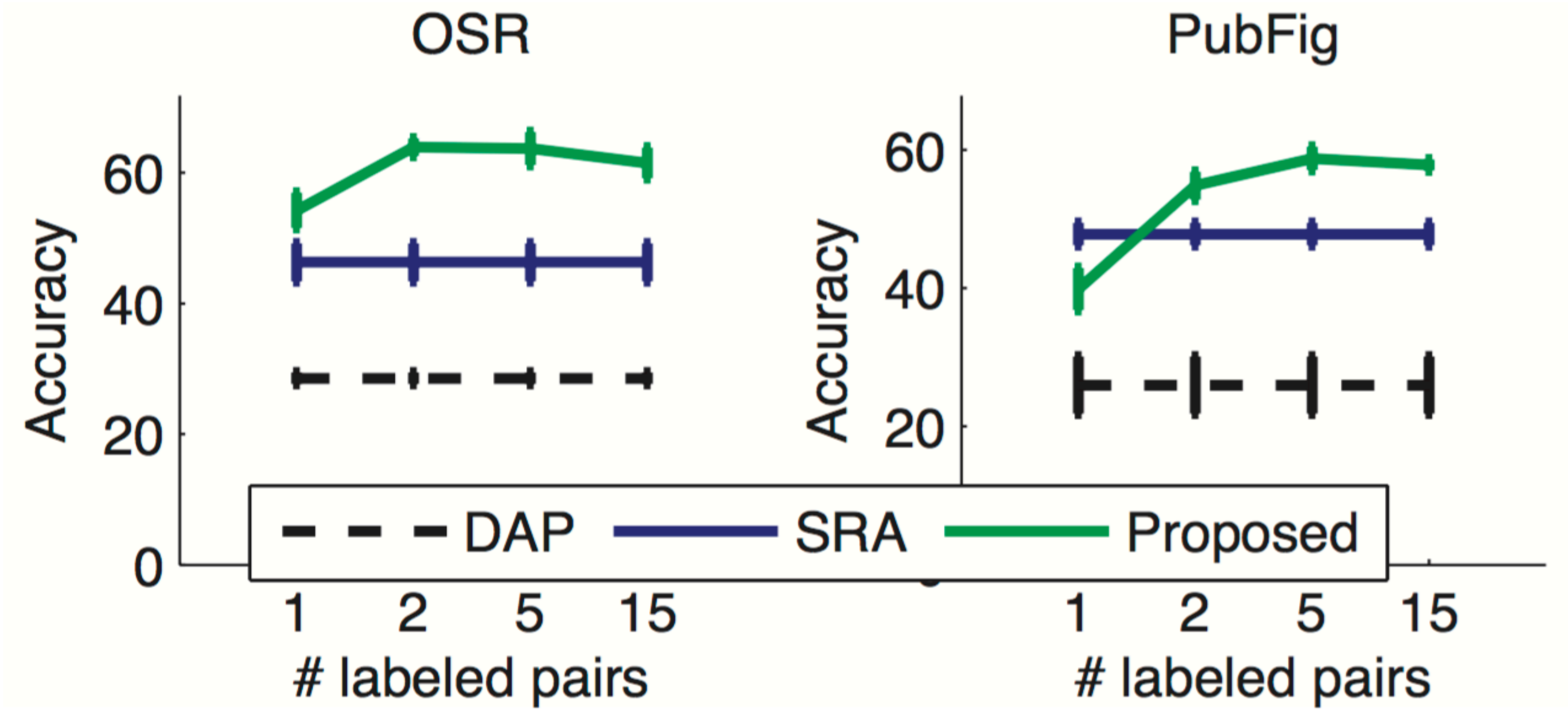


Figure 4. Zero-shot learning performance as more pairs of seen categories are related (*i.e.* labeled) during training.

Experiments: Zero-shot learning

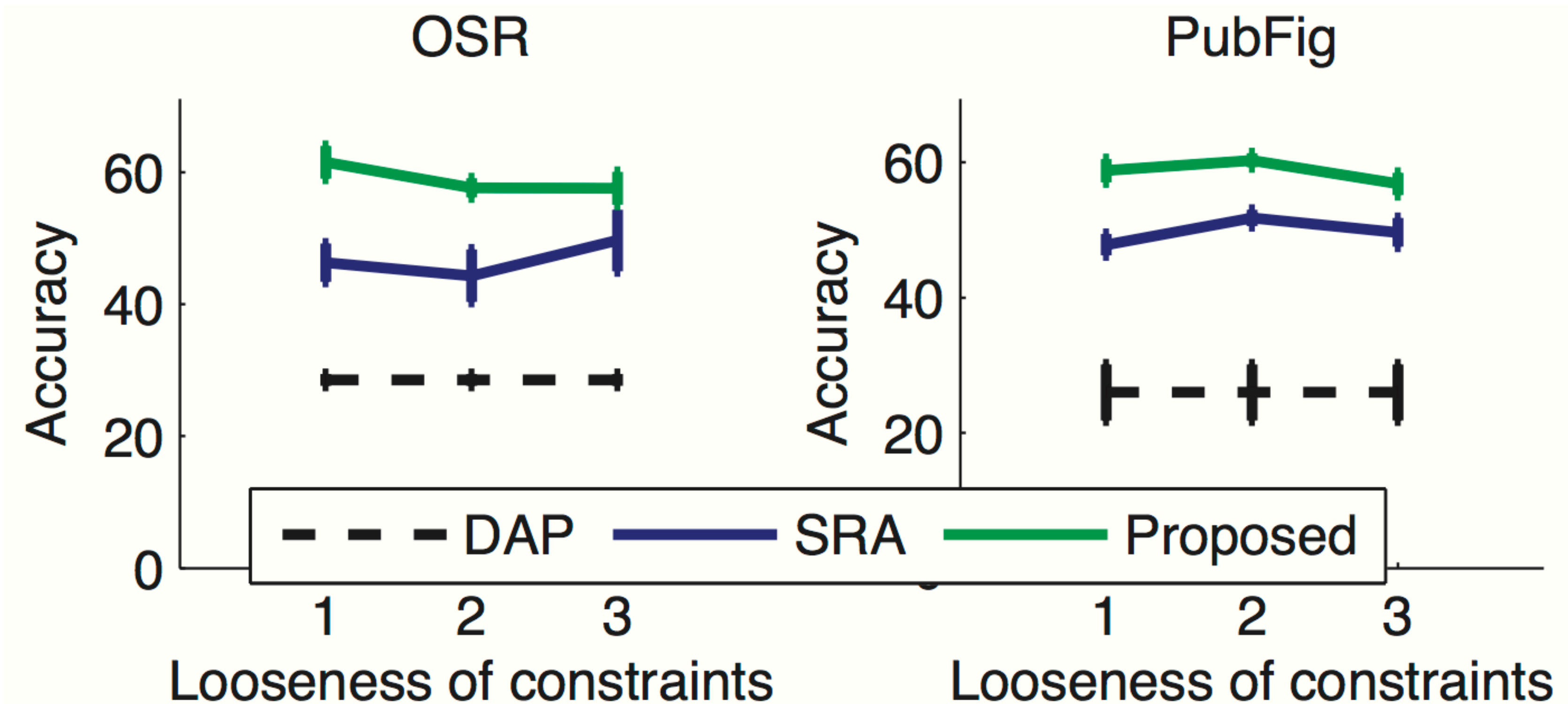


Figure 6. Zero-shot learning performance as the unseen categories are described via looser relationships.

GIST

- GIST is a steerable filter (Gabor filter) response of an image.
- Any image has 1 GIST descriptor of 512 dimensions.
- GIST was developed to provide a holistic descriptor that provides a simpler representation.
- Compared to SIFT features:
  - SIFT is a localized image patch descriptor. A typical image has a few thousand SIFT descriptors, each of 128 dimensions.
  - SIFT was designed for scale and affine invariance in wide baseline image matching tasks, which were part of stereo vision.