

Relative Attributes

Devi Parikh, Kristen Grauman
ICCV 2011

Bora Çelikkale

Purpose



(a) Smiling



(b) ?



(c) Not smiling

Purpose



(a) Smiling



(b) ?



(c) Not smiling

Relative Attributes instead of Binary Attributes

↙
Degree of attribute's presence

Related Work

BINARY ATTRIBUTES

Learning Visual Attributes – NIPS 2007



Object or Face Recognition (CVPR 2009, ICCV 2009, ECCV 2010)

RELATIVE INFORMATION

Explicit Similarity-based Supervision (CVPR 2010) - category dependent similarity
Attribute & Simile Classifiers for Face Verification (ICCV 2009) - comparative facial attributes

LEARNING TO RANK

Optimizing Search Engines using Clickthrough Data (2002)

Learning to Rank for Information Retrieval (2009)

Learning Relative Attributes

Image Set $I = \{i\}$

Attribute Set $A = \{a_m\}$

Goal : Learn M ranking functions:

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

Satisfy maximum number of constraints

$$\forall (i, j) \in O_m : \mathbf{w}_m^T \mathbf{x}_i > \mathbf{w}_m^T \mathbf{x}_j$$

$$\forall (i, j) \in S_m : \mathbf{w}_m^T \mathbf{x}_i = \mathbf{w}_m^T \mathbf{x}_j.$$

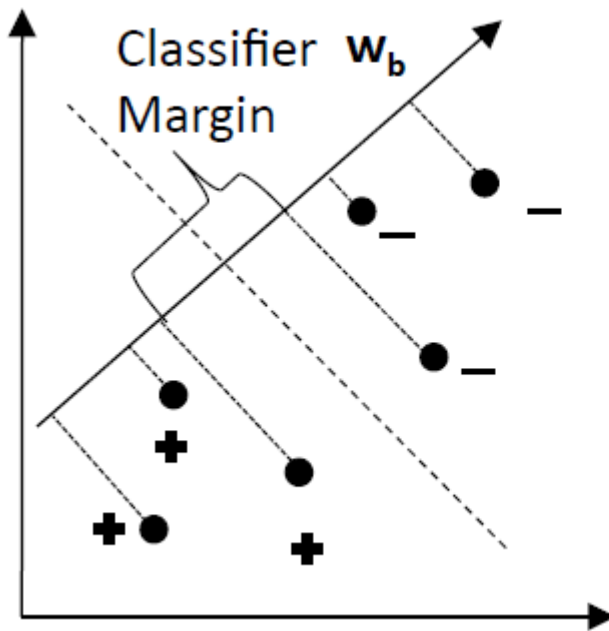
Learning Relative Attributes

NP Hard Problem

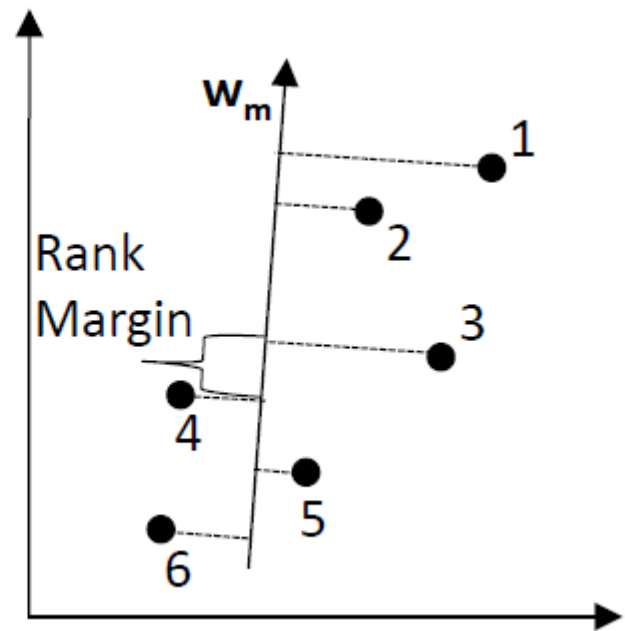
Idea from SVM - **Slack Variables** : Degree of misclassification

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

Learning Relative Attributes



Binary (SVM)



Ranking

Zero-Shot Learning From Relationships

N Categories (obj class or scene type)

S: seen categories (training images provided)

U = $N - S$: unseen categories (training images not provided)

Zero-Shot Learning From Relationships

Rank-valued attribute vector for each image:

$$\mathbf{x}_i \in \mathbb{R}^n \rightarrow \tilde{\mathbf{x}}_i \in \mathbb{R}^M$$

Zero-Shot Learning From Relationships

Rank-valued attribute vector for each image:

$$\mathbf{x}_i \in \mathbb{R}^n \rightarrow \tilde{\mathbf{x}}_i \in \mathbb{R}^M$$

Category definition:

$$c_i^{(s)} \sim \mathcal{N}(\mu_i^{(s)}, \Sigma_i^{(s)})$$

Zero-Shot Learning From Relationships

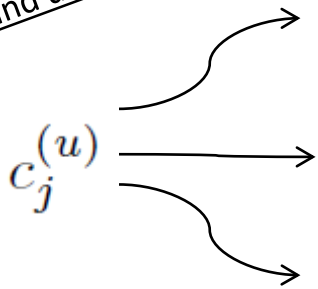
Rank-valued attribute vector for each image:

$$\mathbf{x}_i \in \mathbb{R}^n \rightarrow \tilde{\mathbf{x}}_i \in \mathbb{R}^M$$

Category definition:

$$c_i^{(s)} \sim \mathcal{N}(\boldsymbol{\mu}_i^{(s)}, \boldsymbol{\Sigma}_i^{(s)})$$

find unseen parameters



$$c_i^{(s)} \succ c_j^{(u)} \succ c_k^{(s)}$$

$$c_i^{(s)} \succ c_j^{(u)}$$

$$c_j^{(u)} \succ c_k^{(s)}$$

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

$$\boldsymbol{\mu}_{jm}^{(u)} = \boldsymbol{\mu}_{im}^{(s)} - d_m$$

$$\boldsymbol{\mu}_{jm}^{(u)} = \boldsymbol{\mu}_{im}^{(s)} + d_m$$

Zero-Shot Learning From Relationships

For a test image i :

Calculate attribute vector $\tilde{\mathbf{x}}_i \in \mathbb{R}^M$

$$c^* = \operatorname{argmax}_{j \in \{1, \dots, N\}} P(\tilde{\mathbf{x}}_i \mid \mu_j, \Sigma_j)$$

Describing Images In Relative Terms

Goal : relate any new example image to other images according to different properties

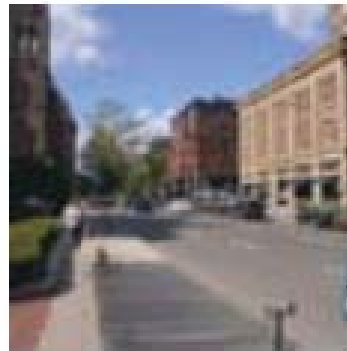


Image is open, ...
Not open Open



Image is
More open than Less open than



Experiments

Outdoor Scene Recognition (OSR) Dataset

2688 images, 8 categories

512 dimensional Gist descriptor

Public Figure Face (PubFig) Database

800 images, 11 categories

Gist + 45 dimensional Lab color histogram

Experiments

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

T: Tall-building

I: Inside-city

S: Street

H: Highway

C: Coast

O: Open-country

M: Mountain

F: Forest

A: Alex Rodriguez

C: Clive Owen

H: Hugh Laurie

J: Jared Leto

M: Miley Cyrus

S: Scarlett Johansson

V: Viggo Mortensen

Z: Zac Efron

Experiments

Compare with:

1. Direct Attribute Prediction (DAP, 2009) – Binary atts

$$c^* = \operatorname{argmax}_{c \in \{1, \dots, N\}} \prod_{m=1}^M P(a_m = b_m^c \mid \mathbf{x})$$

2. Score-based Relative Attributes (SRA)

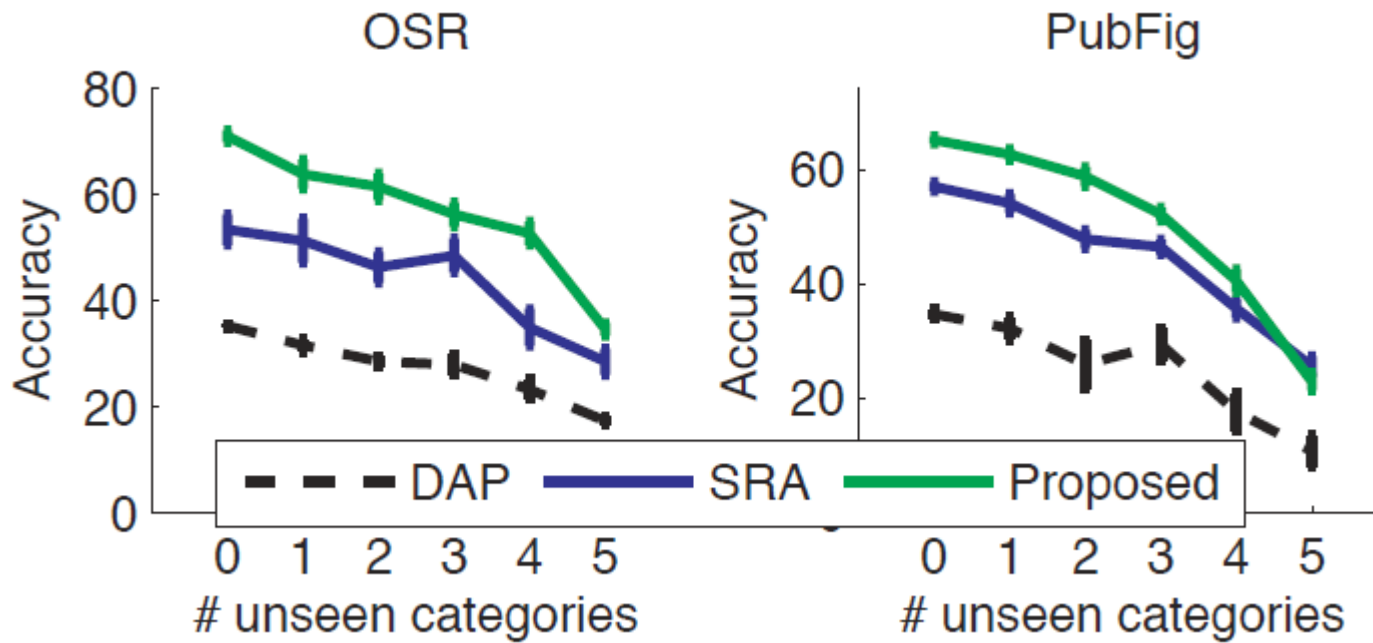
Stronger than DAP

Replace rank values with binary classifier output score

Not limited by binary description of categories

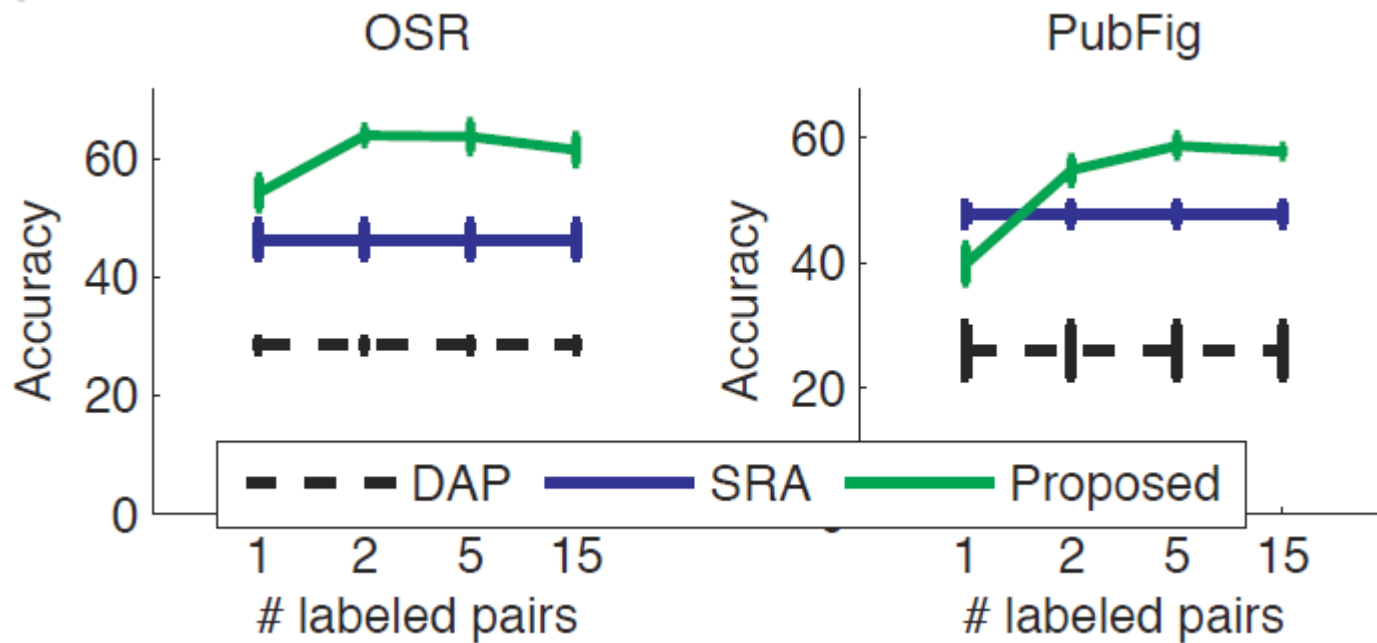
Results

Accuracy change as number of unseen categories increases



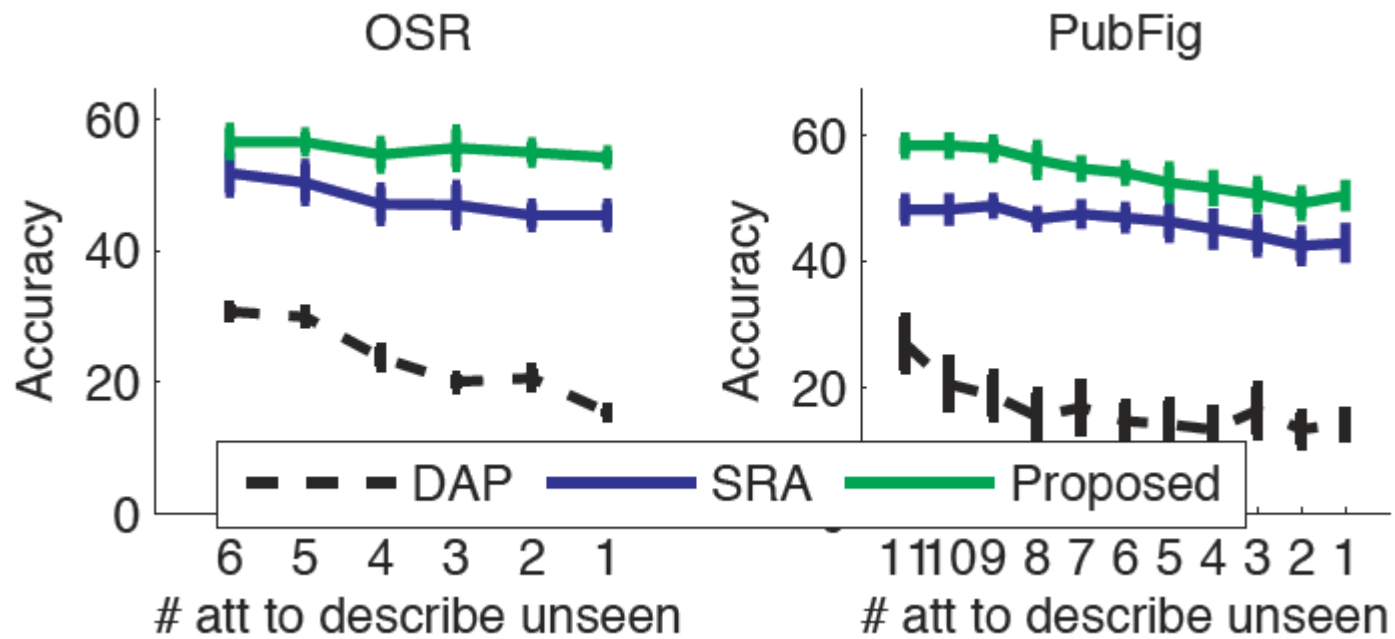
Results

Accuracy change as number of seen categories increases (amount of supervision)



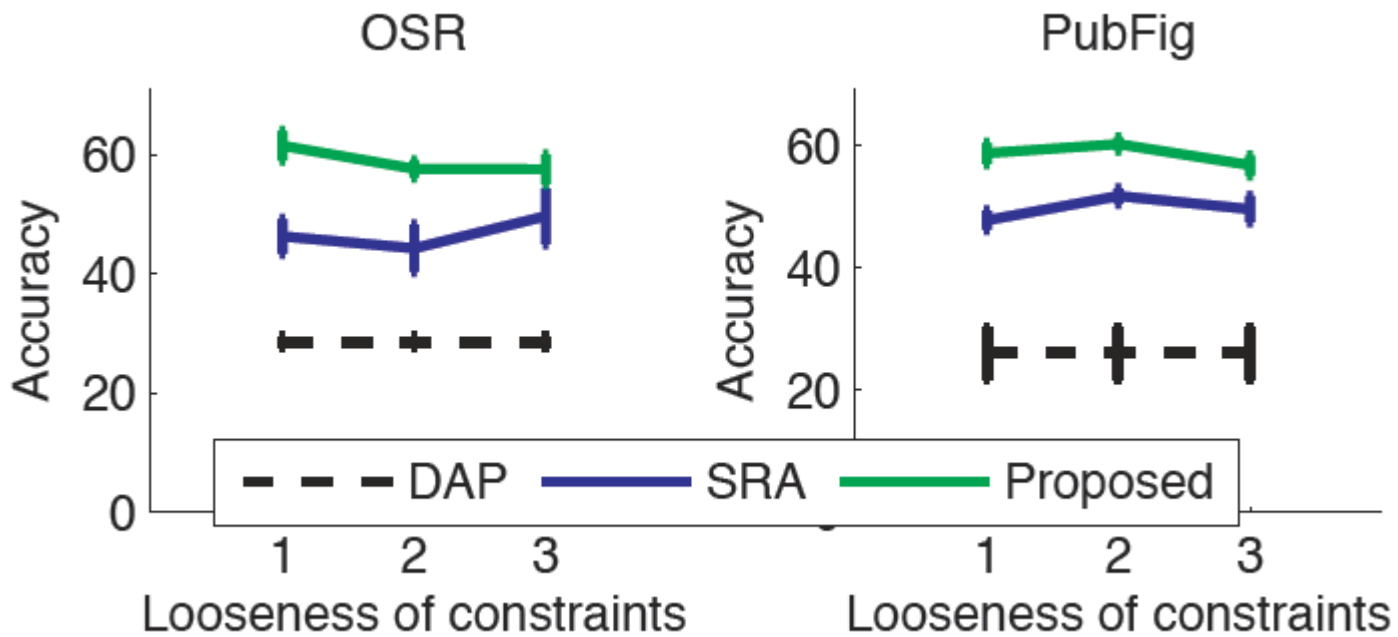
Results

Accuracy change as number of attributes of unseen categories decreases




Results

Accuracy change as unseen categories are described via looser relationships




Describing Image Results

Which image is?




The first panel contains three portrait images stacked vertically. The top image is a woman with dark hair and red lips. The middle image is a man with light brown hair smiling. The bottom image is a man with light brown hair smiling.


More chubby than Less chubby than



More smiling than Less smiling than



More VisibleForehead than Less VisibleForehead than



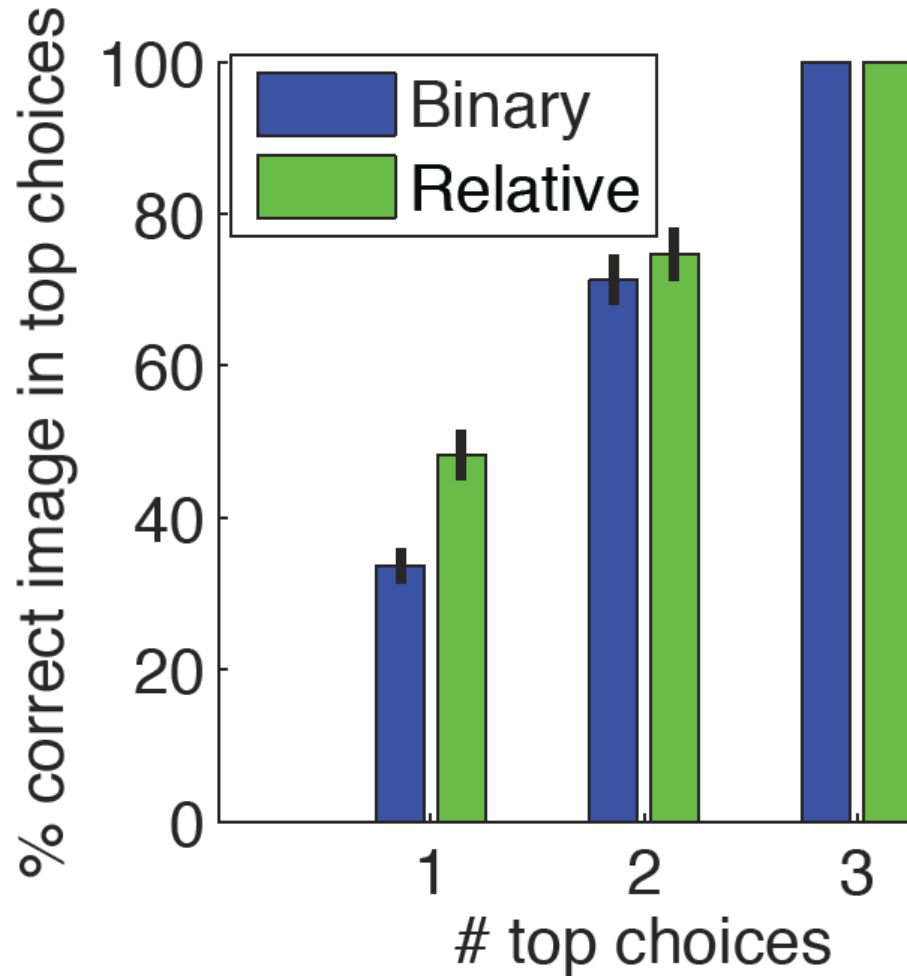
The second panel is divided into three horizontal sections. The top section is titled 'More chubby than' and 'Less chubby than' and shows a woman's face on the left and a man's face on the right. The middle section is titled 'More smiling than' and 'Less smiling than' and shows a woman's face on the left and another woman's face on the right. The bottom section is titled 'More VisibleForehead than' and 'Less VisibleForehead than' and shows a man's face on the left and a woman's face on the right.

Move these three "labels" onto the three images above according to your choices. →

Best Fit Second Fit Worst Fit

The labels are presented in three colored boxes: a green box for 'Best Fit', a blue box for 'Second Fit', and a red box for 'Worst Fit'.

Describing Image Results



Thank You