Unsupervised Learning of Discriminative Relative Visual Attributes

Shugao Ma, Stan Sclaroff, Nazli Ikizler-Cinbis[*]

Department of Computer Science, Boston University

[*]Department of Computer Engineering, Hacettepe University

Overview

- Motivation
- Formulation
- Algorithm
- Experiments
- Conclusion and future work

Attributes: Binary vs. Relative

Attribute "furry"

Binary:  **yes**  **no**

Relative:  >  > 

It seems more natural to use the relative attribute for "furry"

Attributes: Category Level vs. Instance Level

- Category level: *Bears are furrier than giraffes*
- Instance level: ***This** bear is furrier than **that** bear*
- At category level, some attributes are not "relevant" to certain classes

Example: Attribute "open" in "Outdoor Scenes Recognition" dataset

Inside City

Mountain

- In this work, we learn relative attributes at the category level

Learning Attributes: Supervised vs. Unsupervised

- Supervised Learning
  - Attributes are defined and annotated on training data
  - Problems:

Attribute intuitive but not useful

Useful attributes may be overlooked

Annotations may be erroneous

Annotation is labor-intensive, not scalable

- Unsupervised learning methods can help discover useful attributes

Unsupervised Learning of Relative Attributes

- Large search space
  - For N classes, possible orderings are N!
  - Orderings in subsets of classes should also be considered
- Our contribution
  - A formulation for unsupervised relative attribute learning
  - Efficient heuristic algorithm for learning
  - Learned attributes are discriminative, can be used with unseen classes, and correlate well with human labeled relative attributes

Formulation

- Given set of images $I = \{i\}$ represented by feature vectors $\{x_i\}$ and class labels $\{c_a\}$ we learn rank function for attribute m:

$$r_m(\boldsymbol{x}_i^a) = \boldsymbol{w}_m^T \boldsymbol{x}_i^a, \quad s.t. \quad \boldsymbol{w}_m^T \boldsymbol{x}_i^a > \boldsymbol{w}_m^T \boldsymbol{x}_j^b \ , \ i \in c_a, j \in c_b, c_a \succ c_b$$

- Supervised learning formulation [Parikh&Grauman, ICCV 2011]:

$$\min_{\boldsymbol{w}_m, \boldsymbol{\xi}, \boldsymbol{\gamma}} \quad \frac{1}{2}||\boldsymbol{w}_m^T||_2^2 + C(\sum \xi_{ij,ab}^2 + \sum \gamma_{ij,ab}^2)$$

$$s.t. \quad \boldsymbol{w}_m^T(\boldsymbol{x}_i^a - \boldsymbol{x}_j^b) \geq 1 - \xi_{ij,ab}; \ \forall (i,j), i \in c_a, j \in c_b, c_a \succ c_b$$

$$|\boldsymbol{w}_m^T(\boldsymbol{x}_i^a - \boldsymbol{x}_j^b)| \leq \gamma_{ij,ab}; \ \forall (i,j), i \in c_a, j \in c_b, c_a \approx c_b$$

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

Formulation

- Given set of images $I = \{i\}$ represented by feature vectors $\{x_i\}$ and class labels $\{c_a\}$ we learn rank function for attribute m:

$$r_m(\boldsymbol{x}_i^a) = \boldsymbol{w}_m^T \boldsymbol{x}_i^a, \quad s.t. \quad \boldsymbol{w}_m^T \boldsymbol{x}_i^a > \boldsymbol{w}_m^T \boldsymbol{x}_j^b \ , \ i \in c_a, j \in c_b, c_a \succ c_b$$

- Unsupervised learning formulation:

$$\begin{aligned} \min_{\boldsymbol{w}_m, \boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\mu}} \quad & \frac{1}{2}||\boldsymbol{w}_m^T||_2^2 + C_1 \sum \xi_{ij,ab}^2 + C_2(1 - \frac{1}{N} \sum \mu_a) \\ s.t. \quad & \delta_{ab} \boldsymbol{w}_m^T (\boldsymbol{x}_i^a - \boldsymbol{x}_j^b) \geq min(\mu_a, \mu_b) - \xi_{ij,ab}, \\ & \qquad \qquad \forall (i,j), i \in c_a, j \in c_b, a > b \\ & |\delta_{ab} - \delta_{bc}| \geq |\delta_{ab} - \delta_{ac}|, \quad \forall a > b > c, \mu_a = \mu_b = \mu_c = 1 \\ & |\delta_{ab}| = \mu_a, \quad \forall a \in \{2, \ldots, N\} \\ & |\delta_{ab}| = \mu_b, \quad \forall b \in \{1, 2, \ldots, N-1\} \\ & \xi_{ij,ab} \geq 0, \ \delta_{ab} \in \{-1, 0, 1\}, \ \mu_a \in \{0, 1\}. \end{aligned}$$

Formulation

- Given set of images $I = \{i\}$ represented by feature vectors $\{x_i\}$ and class labels $\{c_a\}$ we learn rank function for attribute m:

$$r_m(\boldsymbol{x}_i^a) = \boldsymbol{w}_m^T \boldsymbol{x}_i^a, \quad s.t. \quad \boldsymbol{w}_m^T \boldsymbol{x}_i^a > \boldsymbol{w}_m^T \boldsymbol{x}_j^b \;, \; i \in c_a, j \in c_b, c_a \succ c_b$$

- Unsupervised learning formulation:

$$\min_{\boldsymbol{w}_m, \boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\mu}} \quad \frac{1}{2}||\boldsymbol{w}_m^T||_2^2 + C_1 \sum \xi_{ij,ab}^2 + C_2(1 - \frac{1}{N}\sum \mu_a)$$

$$s.t. \quad \delta_{ab} \boldsymbol{w}_m^T (\boldsymbol{x}_i^a - \boldsymbol{x}_j^b) \geq min(\mu_a, \mu_b) - \xi_{ij,ab}, \quad \forall (i,j), i \in c_a, j \in c_b, a > b$$

$$|\delta_{ab} - \delta_{bc}| \geq |\delta_{ab} - \delta_{ac}|, \quad \forall a > b > c, \mu_a = \mu_b = \mu_c = 1$$

$$|\delta_{ab}| = \mu_a, \quad \forall a \in \{2, \ldots, N\}$$

$$|\delta_{ab}| = \mu_b, \quad \forall b \in \{1, 2, \ldots, N-1\}$$

$$\xi_{ij,ab} \geq 0, \; \delta_{ab} \in \{-1, 0, 1\}, \; \mu_a \in \{0, 1\}.$$

Decision variable δ_{ab} encodes class ordering: $\delta_{ab} = \begin{cases} 1 & c_a \succ c_b \\ -1 & c_a \prec c_b \\ 0 & \mu_a = 0 \vee \mu_b = 0 \end{cases}$

Decision variable $\mu_a \in \{0,1\}$ represents whether attribute *m* is relevant to class c_a

Formulation

- Given set of images $I = \{i\}$ represented by feature vectors $\{x_i\}$ and class labels $\{c_a\}$ we learn rank function for attribute m:

$$r_m(\boldsymbol{x}_i^a) = \boldsymbol{w}_m^T \boldsymbol{x}_i^a, \quad s.t. \quad \boldsymbol{w}_m^T \boldsymbol{x}_i^a > \boldsymbol{w}_m^T \boldsymbol{x}_j^b \ , \ i \in c_a, j \in c_b, c_a \succ c_b$$

- Unsupervised learning formulation:

$$\begin{aligned}
\min_{\boldsymbol{w}_m, \boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\mu}} \quad & \frac{1}{2}||\boldsymbol{w}_m^T||_2^2 + C_1 \sum \xi_{ij,ab}^2 \boxed{+ C_2(1 - \frac{1}{N}\sum \mu_a)} \\
s.t. \quad & \delta_{ab}\boldsymbol{w}_m^T(\boldsymbol{x}_i^a - \boldsymbol{x}_j^b) \geq min(\mu_a, \mu_b) - \xi_{ij,ab}, \\
& \qquad \forall (i,j), i \in c_a, j \in c_b, a > b \\
& |\delta_{ab} - \delta_{bc}| \geq |\delta_{ab} - \delta_{ac}|, \quad \forall a > b > c, \mu_a = \mu_b = \mu_c = 1 \\
& |\delta_{ab}| = \mu_a, \quad \forall a \in \{2, \ldots, N\} \\
& |\delta_{ab}| = \mu_b, \quad \forall b \in \{1, 2, \ldots, N-1\} \\
& \xi_{ij,ab} \geq 0, \ \delta_{ab} \in \{-1, 0, 1\}, \ \mu_a \in \{0, 1\}.
\end{aligned}$$

Favor those attributes that are relevant to more training classes.

Formulation

- Given set of images $I = \{i\}$ represented by feature vectors $\{x_i\}$ and class labels $\{c_a\}$ we learn rank function for attribute m:

$$r_m(\boldsymbol{x}_i^a) = \boldsymbol{w}_m^T \boldsymbol{x}_i^a, \quad s.t. \quad \boldsymbol{w}_m^T \boldsymbol{x}_i^a > \boldsymbol{w}_m^T \boldsymbol{x}_j^b \;, \; i \in c_a, j \in c_b, c_a \succ c_b$$

- Unsupervised learning formulation:

$$\min_{\boldsymbol{w}_m, \boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\mu}} \quad \frac{1}{2} ||\boldsymbol{w}_m^T||_2^2 + C_1 \sum \xi_{ij,ab}^2 + C_2 (1 - \frac{1}{N} \sum \mu_a)$$

$$s.t. \quad \delta_{ab} \boldsymbol{w}_m^T (\boldsymbol{x}_i^a - \boldsymbol{x}_j^b) \geq min(\mu_a, \mu_b) - \xi_{ij,ab},$$

$$\forall (i,j), i \in c_a, j \in c_b, a > b$$

$$|\delta_{ab} - \delta_{bc}| \geq |\delta_{ab} - \delta_{ac}|, \quad \forall a > b > c, \mu_a = \mu_b = \mu_c = 1$$

$$|\delta_{ab}| = \mu_a, \quad \forall a \in \{2, \ldots, N\}$$

$$|\delta_{ab}| = \mu_b, \quad \forall b \in \{1, 2, \ldots, N-1\}$$

$$\xi_{ij,ab} \geq 0, \; \delta_{ab} \in \{-1, 0, 1\}, \; \mu_a \in \{0, 1\}.$$

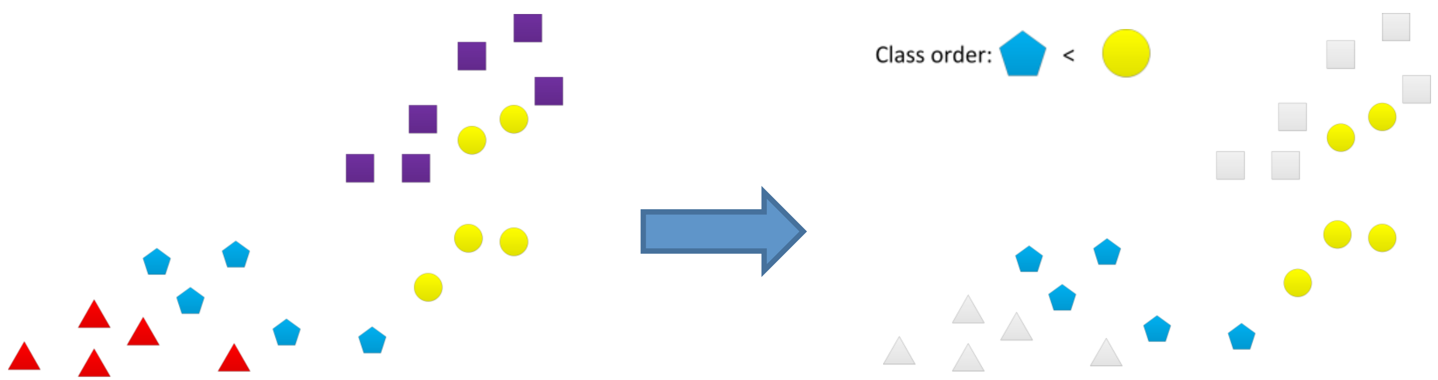
Enforce strict ordering among classes and make sure order is not contradictory.

Algorithm

Basic idea: Alternate between learning $\boldsymbol{w}_m$ and $\boldsymbol{\delta}, \boldsymbol{\mu}$

Initialization: Pick pair of classes c_a and c_b, let $c_a \succ c_b$, and

$$\mu_k = \begin{cases} 1 & k = a \vee k = b \\ 0 & otherwise \end{cases} \qquad \delta_{kh} = \begin{cases} 1 & k = a \wedge h = b \\ 0 & otherwise \end{cases}$$

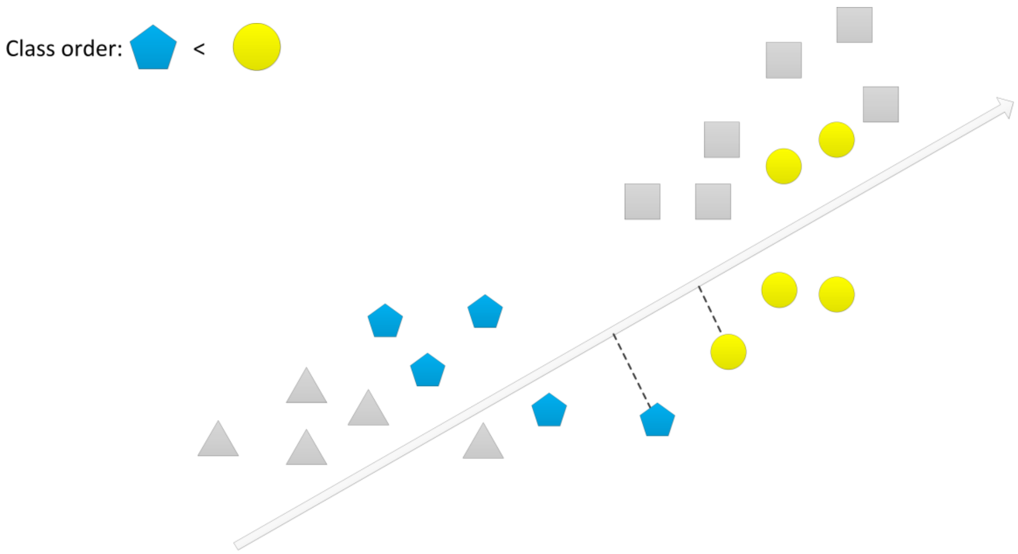


Make as few assumptions about the class ordering as possible; for each pair of classes, we run the algorithm once so that the training data are effectively explored while the search space is not huge: $O(n^2)$.

Algorithm

Updating $\boldsymbol{w}_m$:

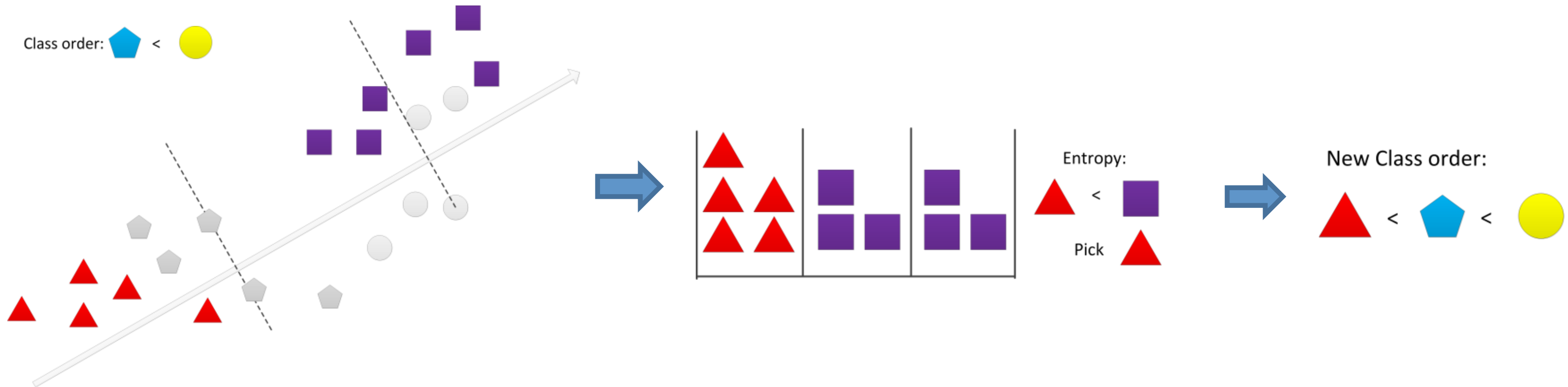
When $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$ are fixed, $\boldsymbol{w}_m$ can be learned via SVM solver.



Algorithm

Updating $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$:

- With $\boldsymbol{w}_m$ fixed the update of $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$ is a mixed integer programming problem, so we use a heuristic method.
- Idea: greedily pick a class that introduces small additional loss if labeled as relevant at this iteration.



Algorithm

Updating $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$ (continued):

- After selecting a class c_d to add to the list of relevant classes, update $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$:

$$\mu_k^t = \begin{cases} \mu_k^{t-1}, & k \neq d \\ 1, & k = d \end{cases} \qquad \delta_{kh}^t = \begin{cases} \delta_{kh}^{t-1}, & k \neq d \wedge h \neq d \\ 1, & (k = d \ \wedge \ m_d^t > m_h^t) \ \vee \ (h = d \ \wedge \ m_k^t > m_d^t) \\ -1, & (k = d \ \wedge \ m_d^t < m_h^t) \ \vee \ (h = d \ \wedge \ m_k^t < m_d^t) \\ 0, & (k = d \ \wedge \ \mu_h^t = 0) \ \vee \ (h = d \ \wedge \ \mu_k^t = 0) \end{cases}$$

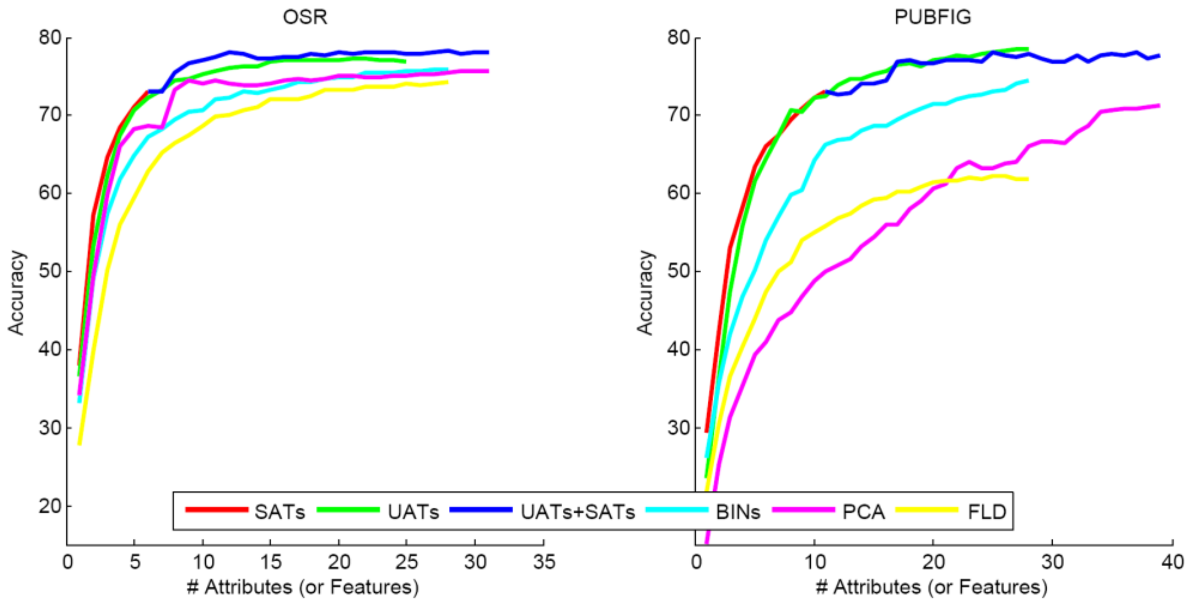
Where m_k^t = median attribute value of class k at the current iteration.

- Repeat the two updating steps, until the objective value stops decreasing or $\mu_a = 1$ for all classes.

Experiments

- Datasets: provided by [Parikh & Grauman, ICCV 2011]
  - Outdoor Scene Recognition (OSR) : 2688 images, 8 categories, 512-D gist as features
  - Subset of the Public Figure Face Database (PUBFIG): 772 images, 8 identities, 512-D gist + 45-D lab color histogram as features
- Three experiments:
  - Multi-class classification
  - K-Shot classification
  - Correlation analysis between automatically learned class orderings and human labeled class orderings

Multiclass Classification

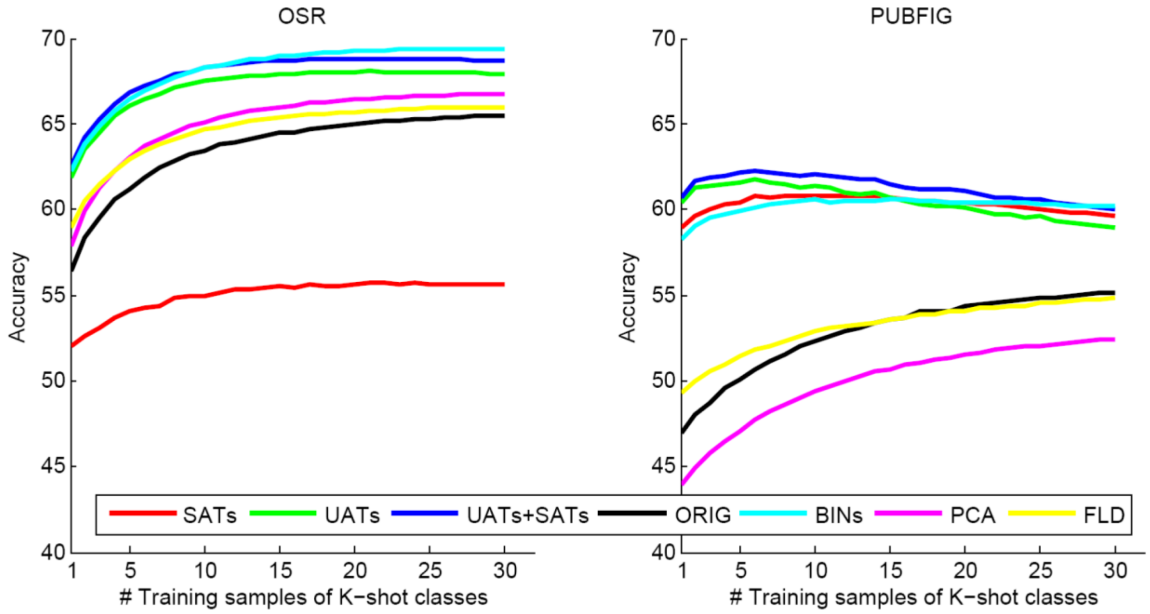


SATs: relative attributes learned by [Parikh&Grauman]
UATs: relative attributes learned by our method
UATs+SATs: all SATs with UATs added one by one

BINs: linear SVM learned between pair of classes
PCA: principal components
FLD: Fisher's Linear Discriminant between pair of classes

- Multi-class SVM with RBF kernel are learned using attributes (or features)
- For same number of attributes, UATs perform similar to SATs.
- However, UATs outnumber SATs, and capture some discriminative attributes that may be overlooked by humans when labeling relative attributes.

K-Shot Classification



- 2 classes are left out when training attributes
- 1-NN classifiers' accuracy is plotted as a function of number of images of the left out classes in the database.
- Attributes learned by the unsupervised algorithm have good generalizability and they can complement the attributes learned via the supervised algorithm.

Correlation Analysis

- Compute Kendal Tau correlation $\tau = \frac{2(n_c - n_d)}{n(n-1)}$ where n_c and n_d are concordant and discordant pairs between two orderings.
- Considering anti-correlation, we use $\hat{\tau} = |\tau|$
- For all human labeled relative attributes, there are highly correlated automatically learned relative attributes.

OSR			
Attr. Name	Sem. Attr.	Auto. Learned Attr.	$\hat{\tau}$
natural	T≺I∼S≺H≺C∼O∼M∼F	S≺I≺H≺F≺O	0.89
open	T∼F≺I∼S≺M≺H∼C∼O	T≺F≺S≺O≺C≺H	0.86
perspective	O≺C≺M∼F≺H≺I≺S≺T	O≺F≺H≺I≺S	1
large-objects	F≺O∼M≺I∼S≺H∼C≺T	F≺M≺S≺H≺C≺T	0.97
diagonal-plane	F≺O∼M≺C≺I∼S≺H≺T	F≺O≺M≺I≺H≺S	0.79
close-depth	C≺M≺O≺T∼I∼S∼H∼F	M≺O≺F≺I≺S	0.84
PUBFIG			
Attr. Name	Sem. Attr.	Auto. Learned Attr.	$\hat{\tau}$
Masculine-looking	S≺M≺Z≺V≺J≺A≺H≺C	S≺M≺Z≺A≺H≺C	1
White	A≺C≺H≺Z≺J≺S≺M≺V	A≺Z≺H≺J≺S	0.80
Young	V≺H≺C≺J≺A≺S≺Z≺M	V≺H≺C≺J≺A≺M	1
Smiling	J≺V≺H≺A∼C≺S∼Z≺M	J≺H≺C≺A≺Z	0.95
Chubby	V≺J≺H≺C≺Z≺M≺S≺A	J≺H≺C≺Z≺A≺M	0.87
Visible-forehead	J≺Z≺M≺S≺A∼C∼H∼V	J≺Z≺M≺C≺A≺H	0.89
Bushy-eyebrows	M≺S≺Z≺V≺H≺A≺C≺J	S≺M≺Z≺A≺H≺C	0.73
Narrow-eyes	M≺J≺S≺A≺H≺C≺V≺Z	M≺A≺J≺H≺C	0.80
Pointy-nose	A≺C≺J∼M∼V≺S≺Z≺H	A≺M≺V≺J≺H	0.84
Big-lips	H≺J≺V≺Z≺C≺M≺A≺S	H≺J≺V≺M≺A	1
Round-face	H≺V≺J≺C≺Z≺A≺S≺M	V≺J≺Z≺A≺S	1

OSR classes include: coast (C), forest (F), highway (H), inside-city (I), mountain (M), open-country (O), street (S) and tall-building (T)

PUBFIG classes include: Alex Rodriguez (A), Clive Owen (C), Hugh Laurie (H), Jared Leto (J), Miley Cyrus (M), Scarlett Johansson (S), Viggo Mortensen (V) and Zac Efron (Z)

Conclusion and Future Work

- Our method automatically discovers useful relative attributes that correlate well with human labeled relative attributes.
- The formulation also considers an attribute's relevance to each training class.
- An interesting direction for future work is learning relative attributes at the instance level.