1. Main Idea

Motivation:
Categorical (binary) attributes are restrictive and can be unnatural

Proposed idea: Relative Attributes
- Richer communication between humans and machines
- Describe images or categories relatively e.g. “dogs are furrier than giraffes”, “find less congested downtown Chicago scene than street”, “learn a ranking function for each attribute”

Enables new applications
- Novel zero-shot learning from attribute comparisons
- Precisely automatically generated textual descriptions of images

2. Learning Relative Attributes

For each attribute $\psi_m$, Supervision is $O_m=\{\{0\},\{1\}\}$, $S_m=\{\{0\},\{1\}\}$

Learn a scoring function $w_m(x) = w_m^T x_i$ that best satisfies constraints:

\[
\forall (i,j) \in O_m : w_m^T x_i > w_m^T x_j \quad \forall (i,j) \in S_m : w_m^T x_i = w_m^T x_j
\]

Max-margin learning to rank formulation

\[
\min \left( \frac{1}{2} |w_m|_2^2 + C \sum_i \xi_i + \sum_j \gamma_j \right)
\]

Adapted objective from [Joachims, 2002]

s.t.

$w_m^T (x_i - x_j) \geq 1 - \xi_j, \forall (i,j) \in O_m$

$|w_m^T (x_i - x_j)| \leq \gamma_j, \forall (i,j) \in S_m$

$\xi_i \geq 0; \gamma_j \geq 0$

3. Ranking Function vs. Binary Classifier Score

How do learned ranking functions differ from classifier outputs?

% correctly ordered pairs: Classifier 80% Ranker 89%

Outdoor scenes: 80% 80% Celebrity faces: 67% 82%

4. Relative Zero-shot Learning

Learned attributes Density: ...

Maximization

Relative attributes space: 1/8 dataset

Relative description:

“more dense than Tallbuildings”, less dense than People,

“more dense than Highways”, less dense than Forests”

Validation: Density: Not dense: Dense

5. Describing Images Relatively

Relative attributes space: 1/8 dataset

Relative descriptions:

“more dense than Tallbuildings”, less dense than People,

“more dense than Highways”, less dense than Forests”

Relative: Not dense: Dense

6. Datasets

- Outdoor Scene Recognition (OSR): 1688 images, 8 categories: corner (C), forest (F), highway (H), inside-city (I), mountain (M), open-country (O), street (S) and tallbuilding (T), get features.
- Public figure Face (PubFig): 630 images, 5 categories: Alex Rodriguez (A), Chuck Norris (C), Hugh Laurie (H), Jared Leto (J), Sylvester Young (Y), Smiling: Smiling, Not Smiling.

7. Image Description Results

Human subject experiment: Which image is...

Example descriptions:

Not natural: More natural than Tallbuildings, less natural than Forests.

An attribute is more discriminative when used relatively

8. Zero-shot Learning Results

Baseline:
- Direct Attribute Prediction (DAP) [Lampert et al. 2009] (binary)
- Classifier instead of ranker (SRA)

Number of unseen categories

Am. of labeled data to learn attributes

Amount of description

Quality of description

Looseness of constraints: Relative attributes jointly carve out space for unseen category.

Relative attributes

Reference