Compare and Contrast: Learning Prominent Visual Differences

Steven Chen and Kristen Grauman

University of Texas at Austin

TEXAS The University of Texas at Austin

2018 CVPR SALT LAKE CITY • JUNE 18-22

Main Idea

- When people compare images, certain prominent differences **stick out**, are **described first**

Left shoe is ____ than right shoe:

Less colorful
Less comfortable
More rugged
More shiny
Less feminine
More stylish
More formal

Prominent Differences

More formal
More shiny
Less comfortable
Less feminine
Less colorful
More rugged
More stylish

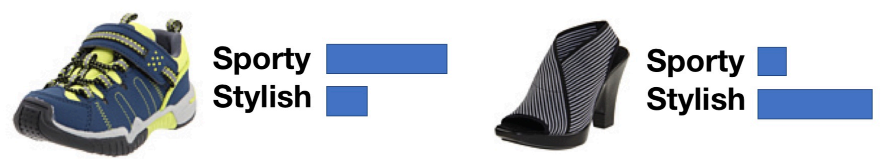
Goal: Learn and use prominent differences

Important to our perception (ex. what stands out when comparing shopping items?)

Influences human input (ex. labeling, feedback to an interactive system)

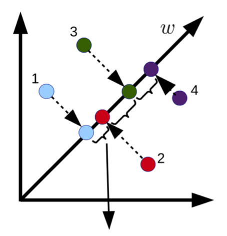
Relative Attributes

Express an image's attribute strength with respect to other images:

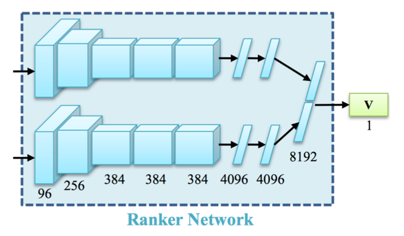


Learn more/less labels using a ranker

Ranker models used:



Ranking SVM
[Parikh and Grauman 2011]



Deep CNN + STN
[Singh and Lee 2016]

Learning Prominent Differences

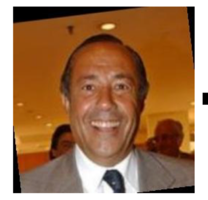
Given pair y_{ij}, get relative attribute scores for each image:
$x_i \rightarrow r_1^i, r_2^i, \ldots, r_M^i,\ x_j \rightarrow r_1^j, r_2^j, \ldots, r_M^j$

Create symmetric representation $\phi(y_{ij})$ for pair using attribute scores

Train multiclass classifier on $\phi(y_{ij})$ using labeled prominence pairs

Given new image pair, predict prominent difference(s)

input: $y_{uv} = (x_u, x_v)$



x_u → Relative Attribute Rankers → $r^u_{1 \cdots M}$

x_v → Relative Attribute Rankers → $r^v_{1 \cdots M}$

$\phi(y_{uv})$ → Prominence Classifier → $\mathcal{A}^{uv}$

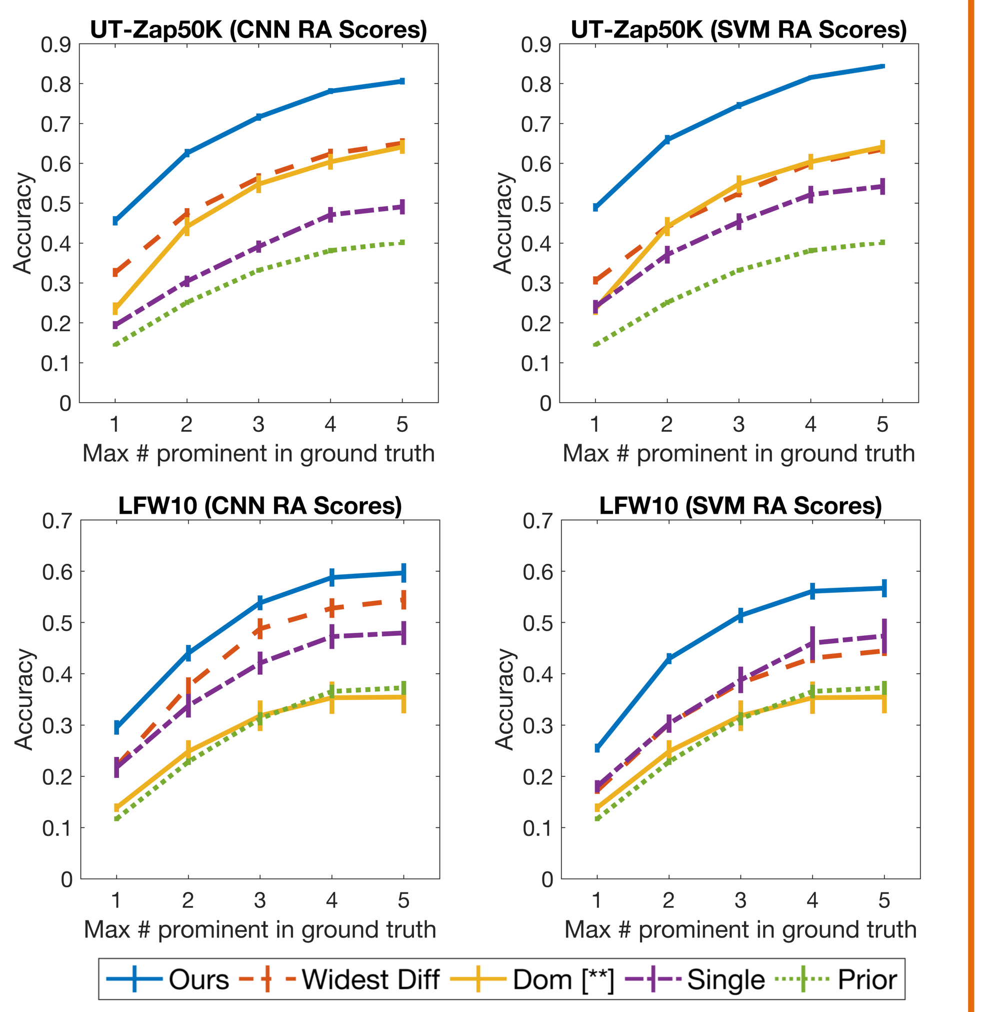
Prominent Difference: **Visible Teeth**

Predicting Prominent Visual Differences

Datasets:
- UT-Zap50K Shoes [Yu and Grauman 2014]: 50,025 shoe images, 10 attributes (sporty, formal, etc.)
- LFW10 Faces [Sandeep et al. 2014]: 2,000 face images, 10 attributes (smiling, bald head, etc.)

Annotations:
- Collect 5,000 pairs / dataset, label prominent difference

Evaluation:
- Benchmark accuracy compared to prior work [Turakhia and Parikh 2013 **] and baseline approaches

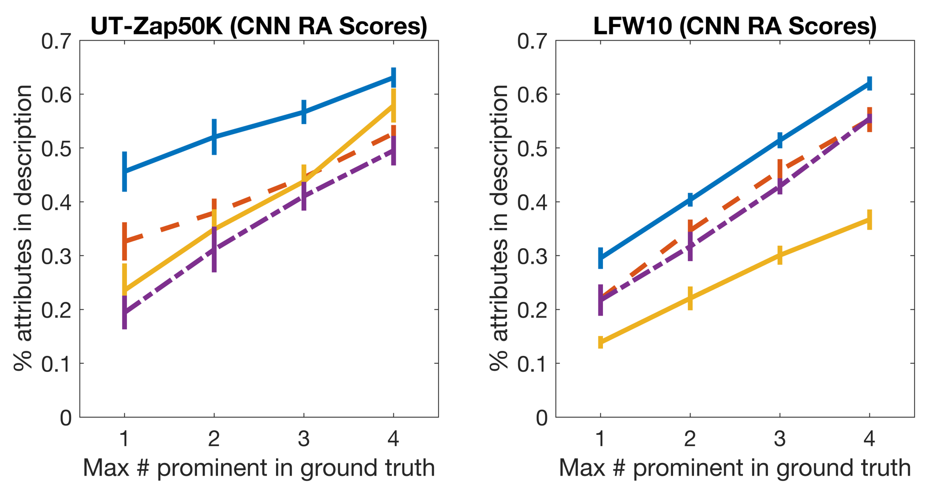


We outperform all baselines on both datasets and for both ranking algorithms

Impact on Description Generation

Intuition: People describe images using prominent differences
Our Approach: Name predicted prominent attributes first

Our descriptions contain more prominent differences than other approaches

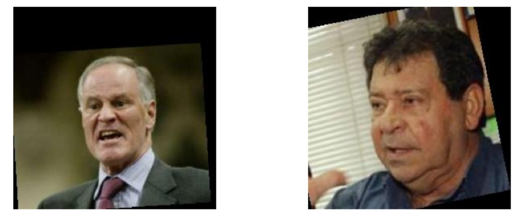


Human judges perceive our descriptions as more natural, appropriate

UT-Zap50K	Ours: 69%	Baseline: 31%
	Ground Truth: 69%	Baseline: 31%
LFW10	Ours: 61%	Baseline: 39%
	Ground Truth: 70%	Baseline: 30%

Ours: Left is **more tall, less sporty, and less rugged** than the right.

Baseline: Less colorful, more shiny, more feminine



Ours: Left has **less dark hair, more bald, and more open mouth** than right.

Baseline: More good looking, more mouth open, less young

Impact on Image Search

Apply prominence to **WhittleSearch** [Kovashka et al. 2012 ~], an interactive image search framework

Less colorful than this

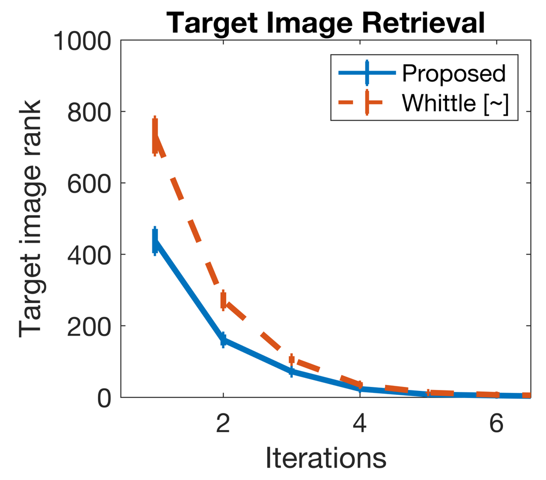
More shiny than this

User provides feedback:
more formal / comfortable, etc. than this

System ranks images by feedback satisfied

Problem: Many images satisfy all feedback, and appear equally relevant to the system

Intuition: People choose prominent differences between images to tell the system
Our approach: Order images by their prominence difference with user feedback



User's Target Image:

Whittle Top Results (two iterations):

Our Top Results (two iterations):

User's Target Image:

Whittle Top Results (two iterations):

Our Top Results (two iterations):

Our approach produces more relevant results, that are more similar to the user's target, yet requires no additional user feedback

Contributions

- Introduce prominent differences, a new functionality for understanding and expressing visual comparisons
- Model and predict prominent differences
- Demonstrate impact on visual search and natural language image description

www.stevenzc.com/comparecontrast