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

Pink, purse, bowknot...

Has Horn
Has leg
Has Head
Has Wool
Relative Attributes

- Smiling
- Not smiling
- Natural
- Not natural

Figure Credit: Devi Parikh
Relative Attributes

- Smiling
- Natural

Images show examples of smiling and natural attributes.
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

\[
O_m: \left\{ \left( \begin{array}{c} \mbox{\includegraphics[width=0.1\textwidth]{image1}} \\ \mbox{\includegraphics[width=0.1\textwidth]{image2}} \end{array} \right), \ldots \right\},
\]

\[
S_m: \left\{ \left( \begin{array}{c} \mbox{\includegraphics[width=0.1\textwidth]{image3}} \\ \mbox{\includegraphics[width=0.1\textwidth]{image4}} \end{array} \right), \ldots \right\}
\]

Slide Credit: Devi Parikh
Learning Relative Attributes

Learn a scoring function \( r_m(x_i) = w_m^T x_i \)

that best satisfies constraints:

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

\[
\forall (i, j) \in S_m : w_m^T x_i = w_m^T x_j
\]
Learning Relative Attributes

Max-margin learning to rank formulation

$$\begin{align*}
\min & \quad \left( \frac{1}{2} \| w_m^T \|_2^2 + C \left( \sum \xi_{ij}^2 + \sum \gamma_{ij}^2 \right) \right) \\
\text{s.t} & \quad w_m^T (x_i - x_j) \geq 1 - \xi_{ij}, \forall (i, j) \in O_m \\
& \quad |w_m^T (x_i - x_j)| \leq \gamma_{ij}, \forall (i, j) \in S_m \\
& \quad \xi_{ij} \geq 0; \gamma_{ij} \geq 0
\end{align*}$$

Image → Relative Attribute Score
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 transfer
  - 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 > Clive > Scarlett
Jared > Miley

Smiling: Miley > 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

\[ p_{\text{ijm}} = \frac{1}{2} (\mu_{\text{im}} + \mu_{\text{km}}) \]

Infer image category using max-likelihood

\[ c^* = \text{argmax}_{j \in \{1, \ldots, N\}} P (\tilde{x}_i \mid \mu_j, \Sigma_j) \]

Age: Hugh \(\rightarrow\) Clive \(\rightarrow\) Scarlett

Jared \(\rightarrow\) Miley

Smiling: Miley \(\rightarrow\) Jared

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
Density

more dense than

less dense than

Slide Credit: Devi Parikh
Density

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
## Category level annotation

<table>
<thead>
<tr>
<th></th>
<th>Binary</th>
<th>Relative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OSR</strong></td>
<td>TI S HC OMF</td>
<td>T &lt; I ~ S &lt; H &lt; C ~ O ~ M ~ F</td>
</tr>
<tr>
<td>natural</td>
<td>0 0 0 0 1 1 1 1</td>
<td>T ~ F &lt; I ~ S &lt; M &lt; H ~ C ~ O</td>
</tr>
<tr>
<td>open</td>
<td>0 0 0 1 1 1 0 0</td>
<td>O ~ C &lt; M ~ F &lt; H ~ I ~ S &lt; T</td>
</tr>
<tr>
<td>perspective</td>
<td>1 1 1 1 0 0 0 0</td>
<td>F &lt; O ~ M &lt; I ~ S &lt; H ~ C &lt; T</td>
</tr>
<tr>
<td>large-objects</td>
<td>1 1 1 0 0 0 0 0</td>
<td>F &lt; O ~ M &lt; C &lt; I ~ S &lt; H &lt; T</td>
</tr>
<tr>
<td>diagonal-plane</td>
<td>1 1 1 1 0 0 0 1</td>
<td>C &lt; M &lt; O ~ T ~ I ~ S &lt; H ~ F</td>
</tr>
<tr>
<td>close-depth</td>
<td>1 1 1 1 0 0 0 1</td>
<td>A C H J M S V Z</td>
</tr>
</tbody>
</table>

**PubFig**

| Masculine-looking | 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 1 1 0 1 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 0 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 |
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
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
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?

Secret Image

Description

Slide Credit: Devi Parikh
Experiments: Describe images

Binary: Smiling, Young
    Smiling
    Not Smiling

    Young
    Not Young

Relative
    More Smiling than
    Less Smiling than

    Younger 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
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
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.
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.