Relative Attributes
Devi Parikh, Kristen Grauman

(a) Smiling
(b) ?
(c) Not smiling
(d) Natural
(e) ?
(f) Manmade

Akanksha Saran
CS381V Paper Presentation
Outline

• Motivation
• Contributions
• Technical Details
• Experiments
• Discussion Points
• Extensions
Outline

• Motivation
  • Contributions
  • Technical Details
  • Experiments
  • Discussion Points
  • Extensions
Donkey

Slide Credit: Devi Parikh, Kristen Grauman
Mule

Slide Credit: Devi Parikh, Kristen Grauman
Attributes

- Is furry
- Has four-legs
- Legs shorter than horses’
- Tail longer than donkeys’
- Has tail

[Oliva 2001] [Ferrari 2007] [Lampert 2009] [Farhadi 2009] [Kumar 2009] [Wang 2009] [Wang 2010] [Berg 2010] [Branson 2010] [Parikh 2010] [ICCV 2011] ...
Binary

- Is furry
- Has four-legs
- Legs shorter than horses’
- Tail longer than donkeys’
- Has tail

Mule

Slide Credit: Devi Parikh, Kristen Grauman
Binary

Is furry

Has four-legs

Legs shorter than horses’

Tail longer than donkeys’

Has tail

Slide Credit: Devi Parikh, Kristen Grauman
Relative

- Is furry
- Has four-legs
- Legs shorter than horses’
- Tail longer than donkeys’
- Has tail

Slide Credit: Devi Parikh, Kristen Grauman
Image Search
“Downtown Chicago”
Relative Description
Outline

• Motivation

• **Contributions**
  • Technical Details
  • Experiments
  • Discussion Points
  • Extensions
Contributions

• Relative attributes
  – Allow relating images and categories to each other
  – Learn ranking function for each attribute

• Novel applications
  – Zero-shot learning from attribute comparisons
  – Automatically generating relative image descriptions
Outline

• Motivation
• Contributions

• **Technical Details**
  • Experiments
  • Discussion Points
  • Extensions
Learning Relative Attributes

For each attribute $a_m$, open supervision is

$$O_m: \{(\sim), \ldots\},$$

$$S_m: \{\{\sim\}, \ldots\}.$$
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

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

Based on [Joachims 2002]

Image \rightarrow Relative Attribute Score

Slide Credit: Devi Parikh, Kristen Grauman
Relative Zero-shot Learning

Training: Images from S seen categories and descriptions of U unseen categories

Age: Hugh > Clive > Scarlett > Jared > Miley

Smiling: Jared > Miley

Need not use all attributes, or all seen categories

Testing: Categorize image into one of S+U categories

Slide Credit: Devi Parikh, Kristen Grauman
Relative Zero-shot Learning

Infer image category using max-likelihood

Slide Credit: Devi Parikh, Kristen Grauman
Relative zero-shot learning

Can predict new classes based on their relationships to existing classes – without training images

Slide Credit: Devi Parikh, Kristen Grauman
Automatic Relative Image Description

Density

Dense: Novel image

Not dense:

Conventional binary description: not dense

Slide Credit: Devi Parikh, Kristen Grauman
Automatic Relative Image Description

Density

Novel image

more dense than

less dense than

Slide Credit: Devi Parikh, Kristen Grauman
more dense than Highways, less dense than Forests
Outline

• Motivation
• Contributions
• Technical Details

• Experiments
  • Discussion Points
  • Extensions
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, Kristen Grauman
Baselines

• Zero-shot learning
  – Binary attributes:
    Direct Attribute Prediction [Lampert 2009]
  – Relative attributes via classifier scores

• Automatic image-description
  – Binary attributes
Relative 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
Relative Zero-shot Learning

An attribute is more discriminative when used relatively.
Automatic Relative Image Description

Binary (existing):
- Not natural
- Not open
- Has perspective

Relative (proposed):
- 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

Slide Credit: Devi Parikh, Kristen Grauman
**Relative (proposed):**

More natural than tallbuilding
Less natural than forest

More open than tallbuilding
Less open than coast

Has more perspective than tallbuilding

**Binary (existing):**

Not natural
Not open
Has perspective
Human Studies:
Which Image is Being Described?

Slide Credit: Devi Parikh, Kristen Grauman
Automatic Relative Image Description

18 subjects

Test cases:
10OSR, 20 PubFig

% correct image in top choices

# top choices

Slide Credit: Devi Parikh, Kristen Grauman
Outline

• Motivation
• Contributions
• Technical Details
• Experiments
• Discussion Points
• Extensions
Advantages

• Natural Descriptions: Leverages a natural mode of description

• Flexibility: Allows use of as many attributes for defining relations among as many classes
Image based Attribute Ranking

Relative ordering for attributes are transferred to all images in a category.
Image based based Attribute Ranking

Relative ordering for attributes are transferred to all images in a category.
Image based Attribute Ranking

Relative ordering for attributes are transferred to all images in a category

Image Search
Gaussian distribution in joint attribute space

• Underlying distributions may be multi-modal
Fine-grained differences?

Can retaining the ranks for two very similar images/categories help identify them?
Outline

• Motivation
• Contributions
• Technical Details
• Experiments
• Strengths and Weaknesses

• Extensions
Extensions

- Relative attributes learned per image

- Active Learning of Discriminative Classifiers through feedback from users
  ”Simultaneous Active Learning of Classifiers & Attributes via Relative Feedback”, A. Biswas, D. Parikh

- Use of binary and relative attributes together
  ‘A horse has 4 legs’

- More expressive features instead of global features
  To discriminate a large set of image categories
  “Discovering Spatial Extent of Relative Attributes”, F. Xiao, Y.J. Lee

- Scalability to more categories and attribute labels
Thank you!