Machine Learning in Image Analysis
Day 3

Anirban Mukhopadhyay
Zuse Institute Berlin

Organization

- Recap
  - Day 1
  - Day 2
- General comments about the papers
- Decision Forest+GMM
- Marginal Space Learning
- Scale-Invariant Learning
- Relative Attributes

Recap Day 1 & 2

- Why ML in IA
- Intuitions behind choosing ML techniques
- Linear SVM and Cutting Plane to solve
- ML, MAP, Bayesian differences
- Derivation of EM for calculating ML
- Monte Carlo Integration and Importance Sampling
- MCMC
- Gibbs Sampling

List of Papers

- Medical Image Analysis
  - Decision Forest+GMM
  - Marginal Space Learning

- Computer Vision
  - Unsupervised Learning
  - Relative Attributes

Main Idea of each paper

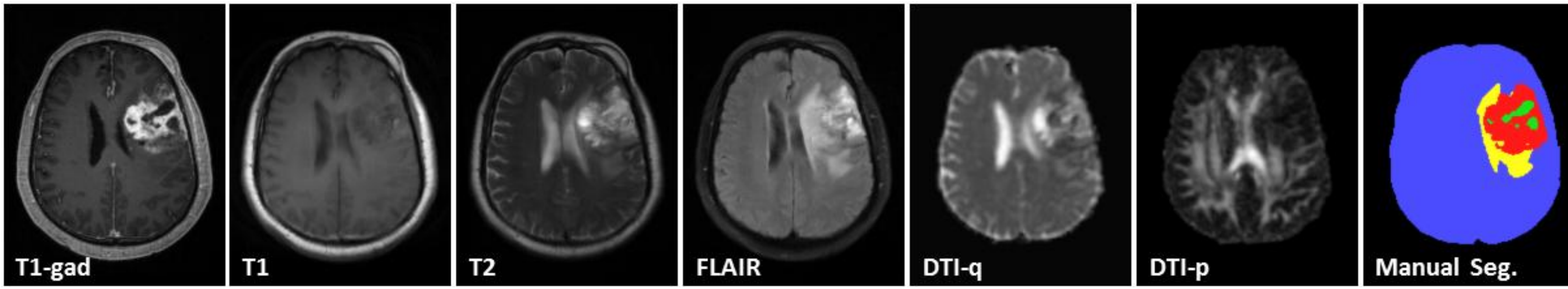
Decision Forest + GMM	Marginal Space Learning (MSL)	Relative Attributes	Scale-Invariant Learning
Multi Label classification using Decision Forest + Tissue specific GMM Posteriors	Localizing Heart chambers (pose estimation) using MSL	Learn ranking function per attribute -> relative strength of each property	Model objects using flexible constellation of parts + Expectation Maximization

Decision Forest+GMM

- Generative and Discriminative together to solve a multi label classification problem

Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR
D Zikic, B Glocker, E Konukoglu, A Criminisi... - ... Image Computing and ..., 2012 - Springer
Abstract We present a method for automatic segmentation of high-grade gliomas and their subregions from multi-channel MR images. Besides segmenting the gross tumor, we also differentiate between active cells, necrotic core, and edema. Our discriminative approach ...
Cited by 74   Related articles   All 21 versions   Cite   Save

Problem definition



- Automatic segmentation of high-grade gliomas and their subregions from multi-channel MR images
- Differentiate between
  - active cells
  - necrotic core
  - edema

Motivation of chosen method

- Most of the previous research focuses on segmentation of gross tumor
- Perform a tissue specific segmentation of three relevant tissues types
- Probability estimates based on Gaussian mixture models (GMM)
- Inherently multi-label classification using Decision Forest

Method

- Initial tissue probability estimate
  - Generative modeling using GMM

- Determination of class for spatial input point
  - Discriminative learning using Decision Forest

Basics of GMM

- A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters.
  - Think of mixture models as generalizing k-means clustering to incorporate information about the covariance structure of the data as well as the centers of the latent Gaussians.

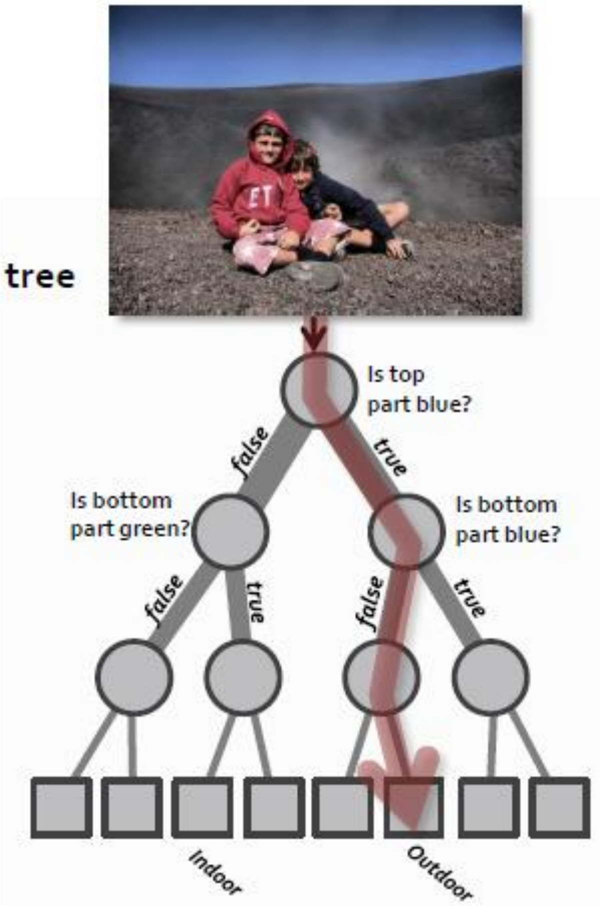
How GMM used here

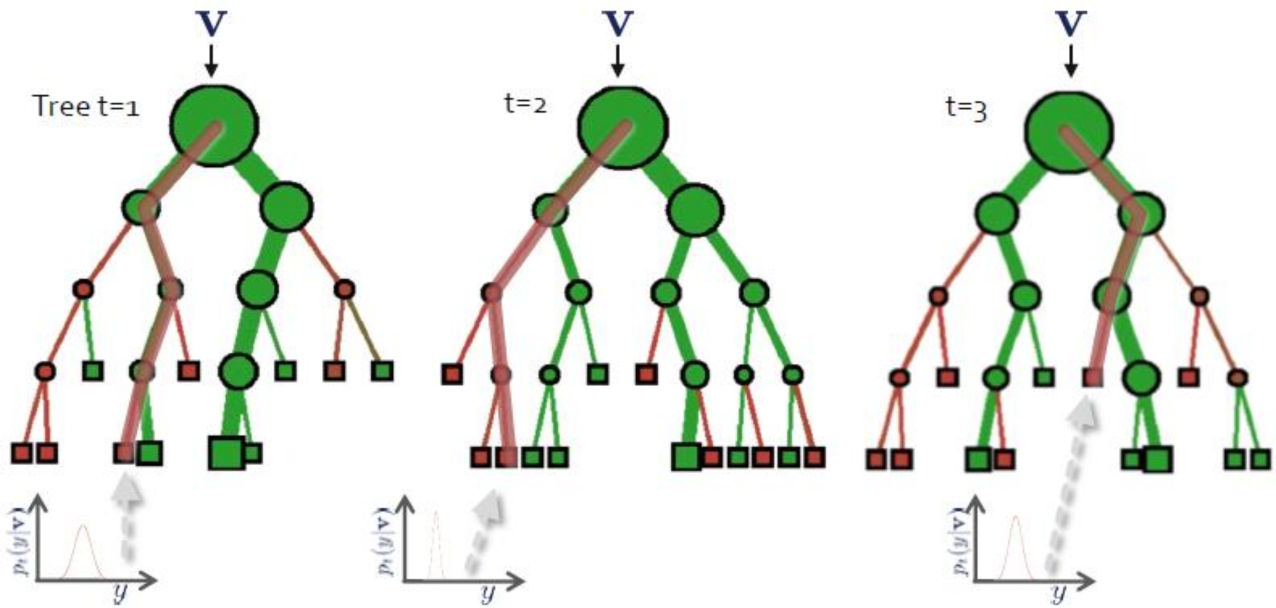
- Initial class probabilities for a given patient as posterior probabilities
  - based on likelihoods obtained by training a set of GMMs
- For each class c, a single GMM is trained,
  - captures the likelihood of the multi-dimensional intensity for this class.
- Use the probabilities directly as input for the decision forests, in addition to the multi-channel MR data.

$$I = (\text{T1-gad}, \text{T1}, \text{T2}, \text{FLAIR}, \text{DTI-q}, \text{DTI-p}, p_{\text{AC}}^{\text{GMM}}, p_{\text{NC}}^{\text{GMM}}, p_{\text{E}}^{\text{GMM}}, p_{\text{B}}^{\text{GMM}})$$

- Generate context-based features from I

Basics of Decision Forest





- Node: Training Examples, Predictor
- Successive splitting of the training examples at every node based on their feature
- Splits along randomly chosen dimensions of the feature space is considered -> maximizing the Information Gain

Decision Forest Training

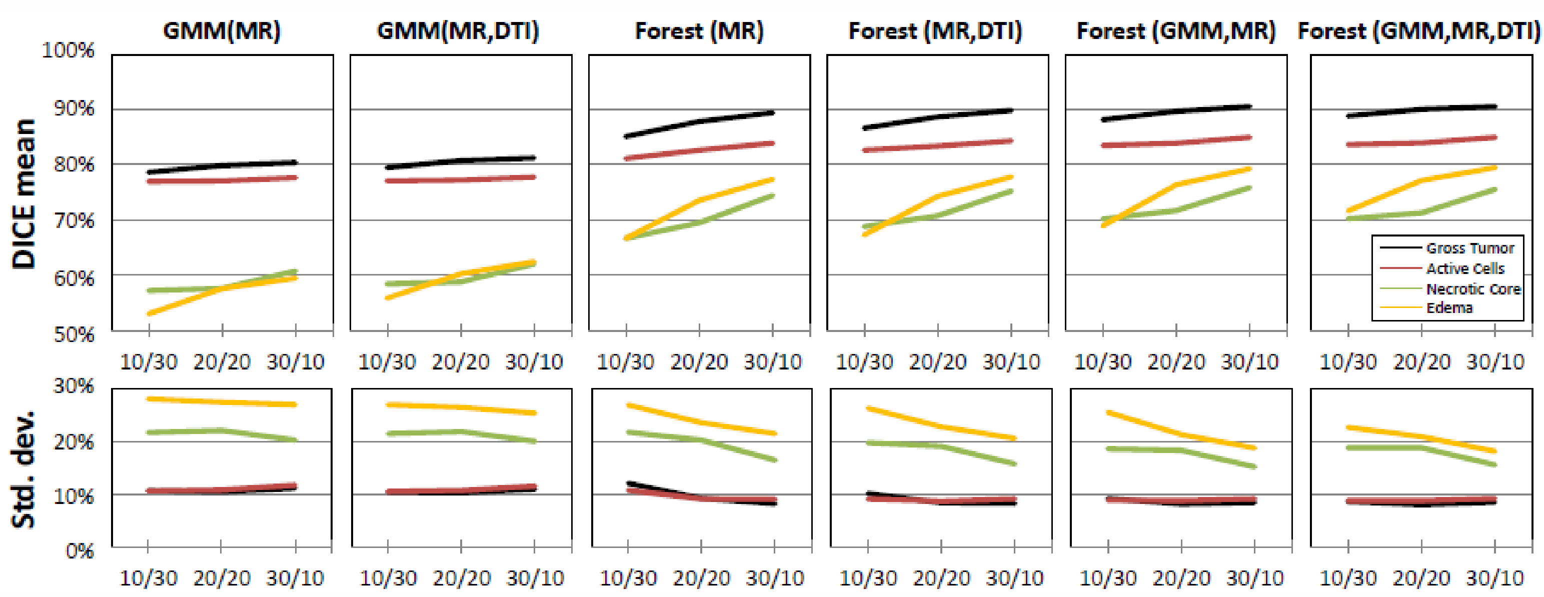
- Employ decision forests (DF) to determine a class c for a given spatial input point, based on the representation of x by the feature vector
- Training:
  - Each tree learns a weak classifier for the feature representation of a sample point
  - Split & Grow each tree
  - Tree growing is stopped at a certain tree depth

Decision Forest Testing

- Testing
  - Point to be classified is pushed through each tree, by applying the learned split functions.
  - Upon arriving at a leaf node, the leaf probability is used as the tree probability
  - overall probability is computed as the average of tree probabilities
  - Actual class estimate is chosen as the most probable class

Results

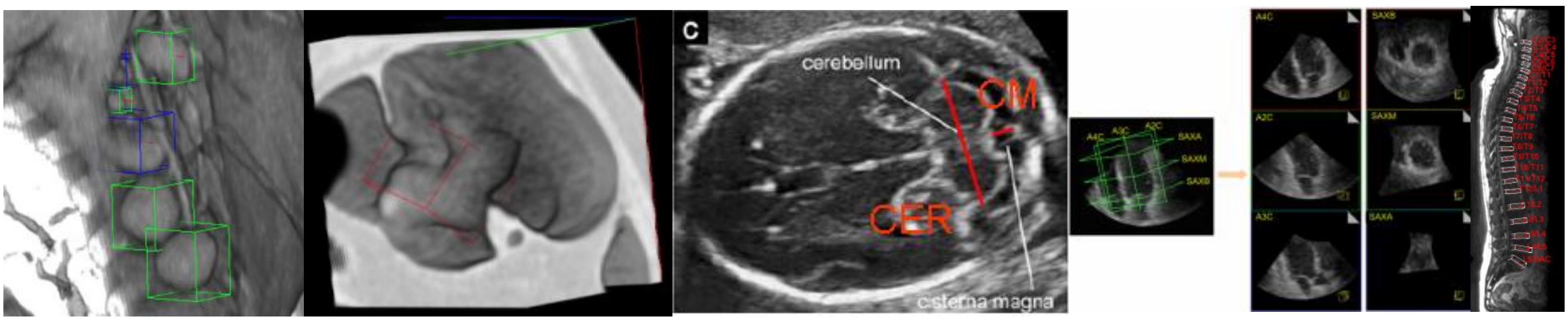
- 40 patients are randomly split into non-overlapping training and testing data sets
- perform experiments with following training/testing sizes: 10/30, 20/20, 30/10
- each of the three ratios, 10 tests are performed, by randomly generating 10 different training/testing splits.



Open Discussion

Marginal Space Learning

- Edison Award winning Patent for Marginal Space Learning



Four-chamber heart modeling and automatic **segmentation** for 3-D **cardiac** CT volumes using **marginal space learning** and steerable features
Y Zheng, A Barbu, B Georgescu... - Medical Imaging, ..., 2008 - ieeexplore.ieee.org
Abstract—We propose an automatic **four-chamber heart** seg-mentation system for the quantitative functional analysis of the **heart** from **cardiac** computed tomography (CT) volumes. Two topics are discussed: **heart** modeling and automatic model fitting to an **...**
Cited by 413 Related articles All 19 versions Cite Save

Problem definition

- Quantitative functional analysis of heart from 3D CT
- Automatic heart chamber segmentation
  - Heart Localization
  - Model modeling and fitting to unseen volumes

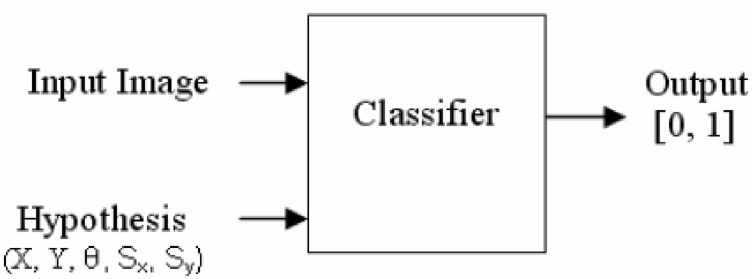
Motivation of chosen method

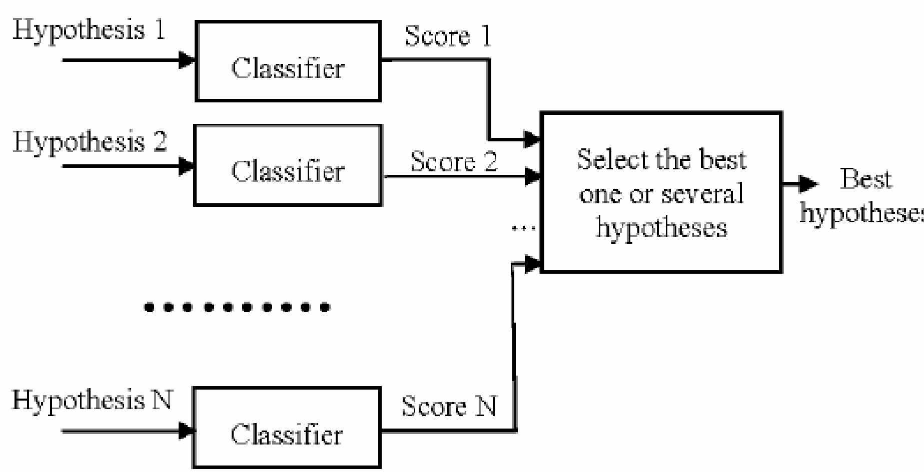
- Efficient 3D object detection based on Marginal Space Learning (MSL) and Steerable Features (SF).
- MSL: Incrementally learn classifiers on projected sample distributions
  – position estimation
  – position-orientation estimation
  – full similarity transformation estimation
- SF: Much fewer points are needed compared to the whole volume
  – sample a few points under a sampling pattern
  – extract a few local features (e.g., intensity and gradient)

Full Space Learning (FSL)

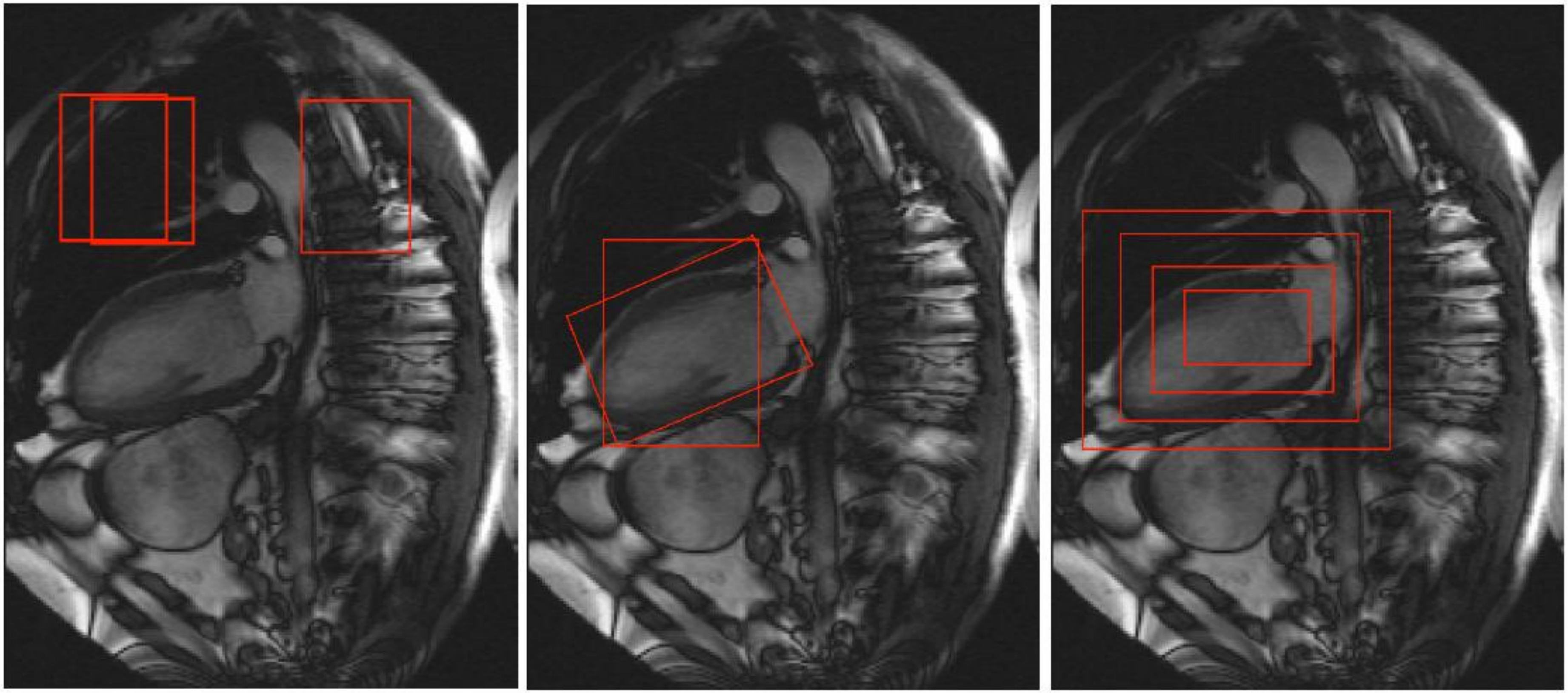
Learning based approach

- It is currently the state-of-the-art in 2D object detection.
- Learning: Whether an image block contains the target object or not.





FSL contd.



#Hypotheses: 1000 X 20 X 50 = 1,000,000

Full space learning tests all possible combinations of the transformations (over 1 million hypotheses) to pick the best one.

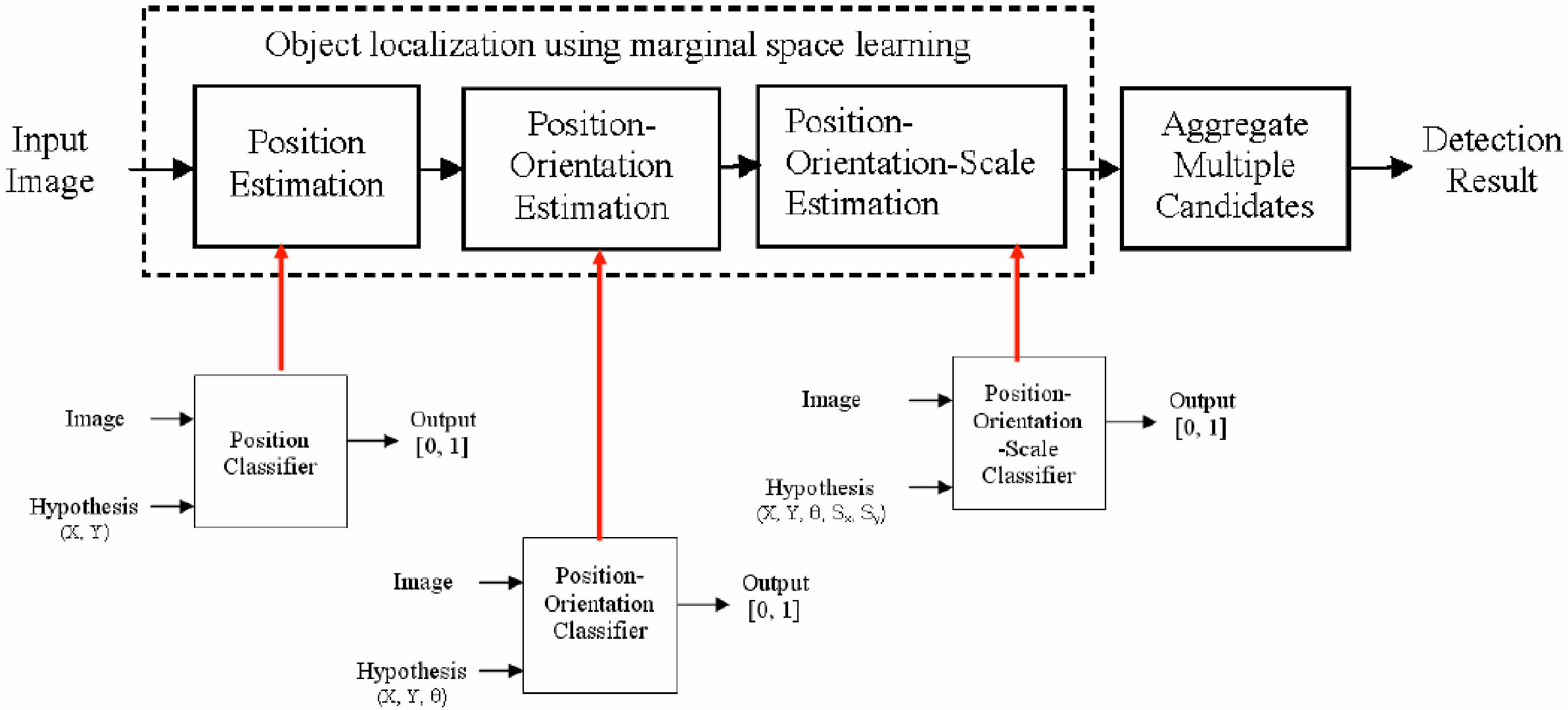
3D challenges of FSL

hypotheses increases exponentially w.r.t. the dimensionality of the parameter space.

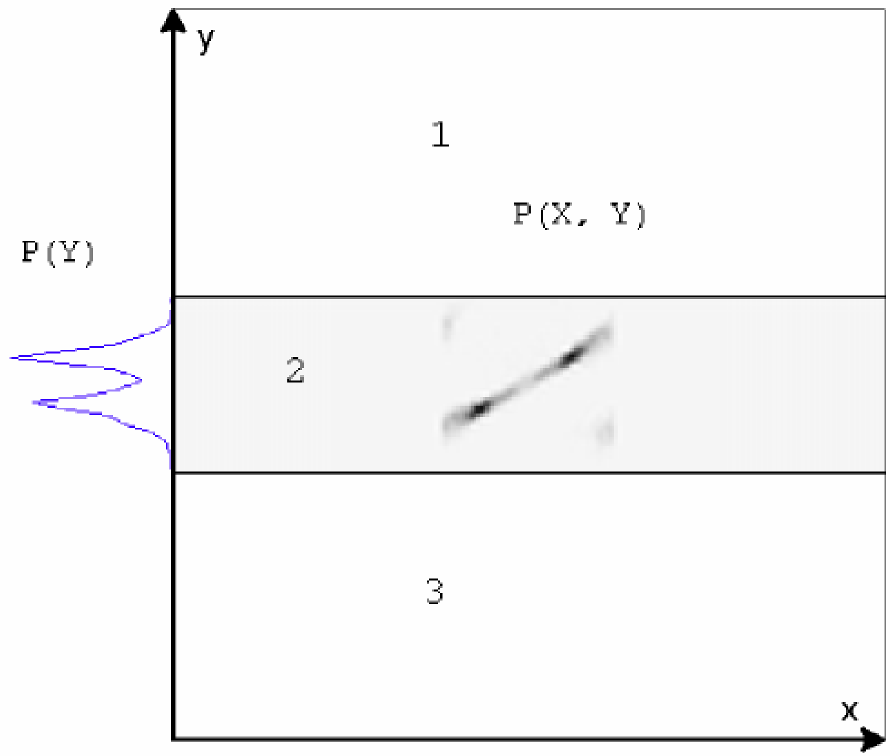
- 9 degrees of freedom for the similarity transformation (3 translations, 3 rotation angles, and 3 anisotropic scales).
- For a small n=10, # hypotheses is n^9 = 1,000,000,000.
- Need to develop an efficient method to explore the parameter space.
- Solution: Marginal Space Learning

Marginal Space Learning Details

- Efficiently detect position, orientation, and scaling of an object
- Train 3 classifiers instead of 1 monolithic classifier
- Perform learning/detection in marginal spaces of increasing dimensions.



Why MSL is efficient?

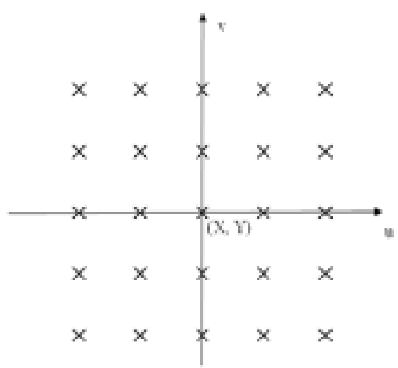


A 2D example: A classifier trained on p(y) can quickly eliminate a large portion (regions 1 and 3) of the search space.

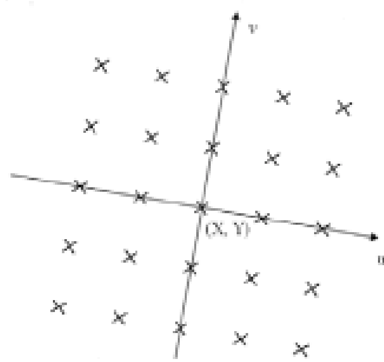
Steerable Features

- Steerable features combine advantages of global and local features (for orientation/scale estimation)
  - Global featers, (e.g., 3D Haar wavelet features), are effective to capture the global orientation and scale information of an object.
  - Local features are fast to evaluate but lose the global information.
  - Sampling patterns to incorporate orientation and scale information.
  - Local features (voxel intensity and gradient).
  - Flexible framework.

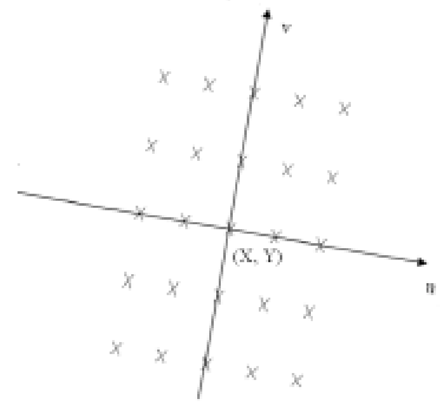
Given a hypothesis (X, Y, alpha, Sx, Sy)



Put center of the sampling pattern to (X, Y)

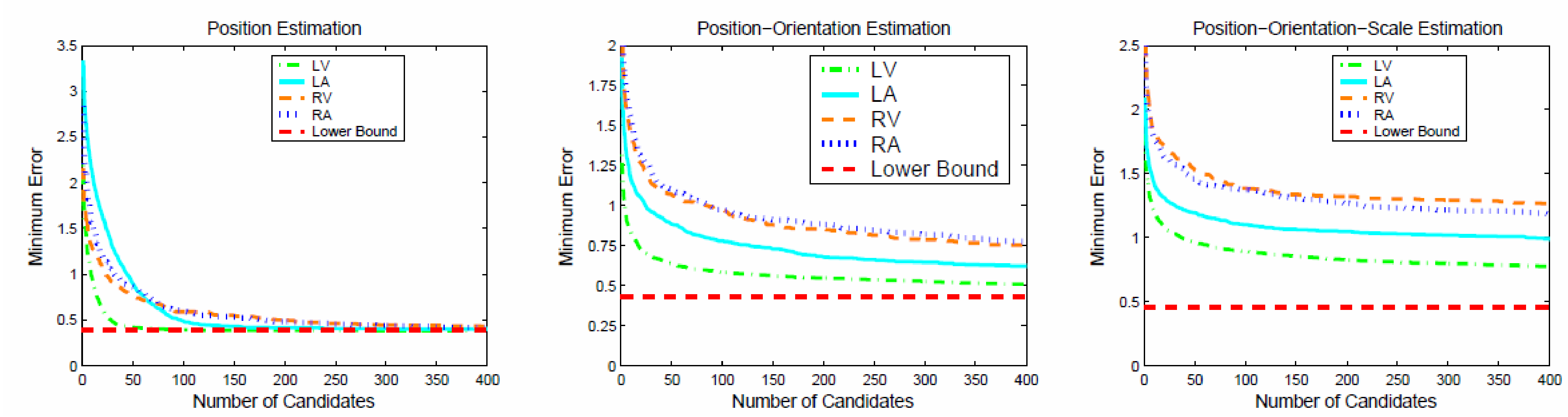


Align sampling pattern to the orientation (alpha).



Scale sampling pattern to incorporate scaling (Sx, Sy).

Results



- # of cands vs. Average Error of best candidate
  – only need to preserve a small number of candidates after each step, without deteriorating accuracy much.

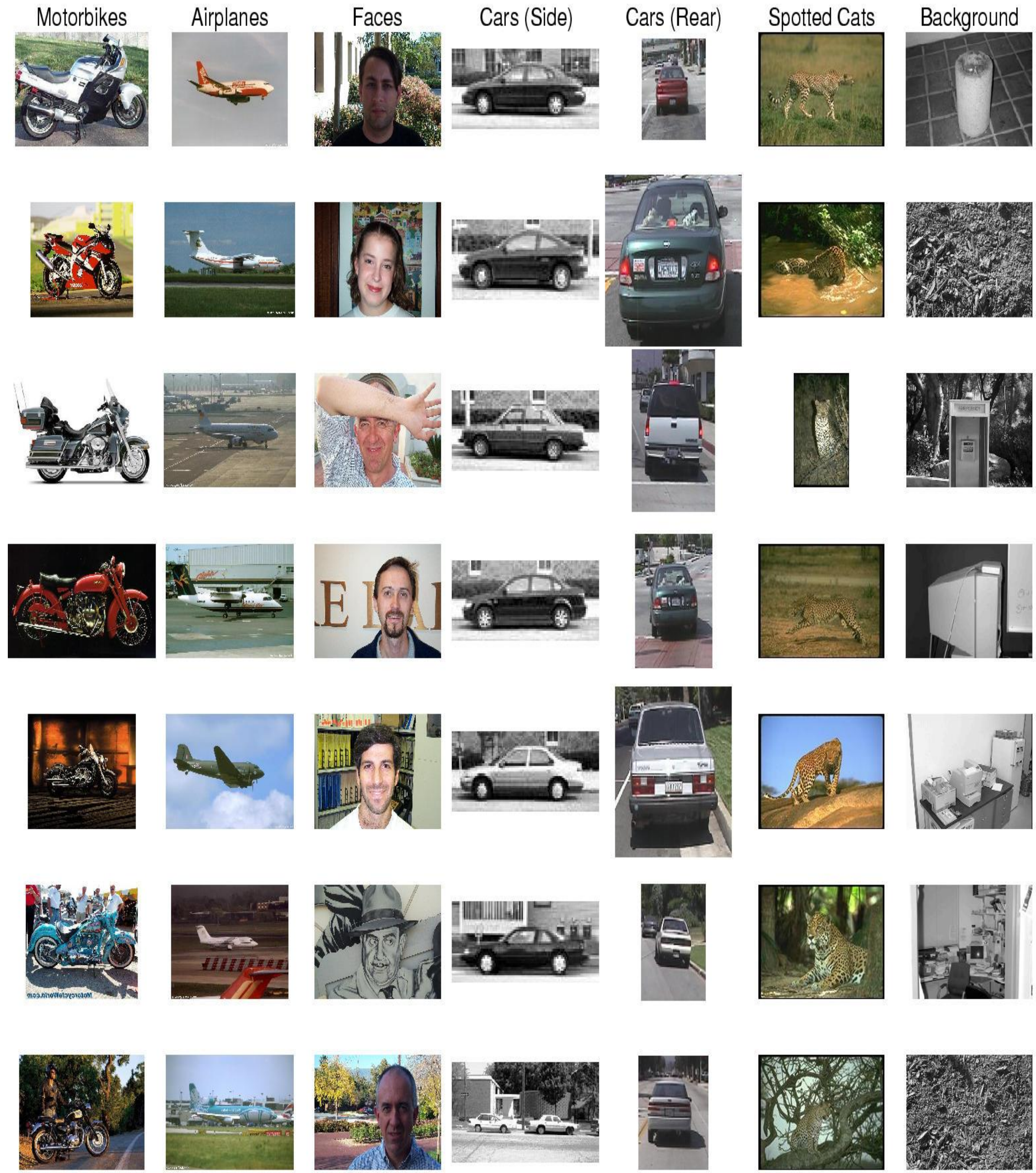
Open Discussion

Unsupervised Learning (weakly supervised)

- Mother technique to a volume of Computer Vision papers

Object class recognition by **unsupervised scale-invariant learning**
R Fergus, P Perona, A Zisserman - ... and Pattern **Recognition**, ..., 2003 - ieeexplore.ieee.org
Abstract We present a method to learn and recognize object class models from unlabeled and unsegmented cluttered scenes in a scale invariant manner. Objects are modeled as flexible constellations of parts. A probabilistic representation is used for all aspects of the ...
Cited by 2313 Related articles All 62 versions Cite Save

Problem Definition



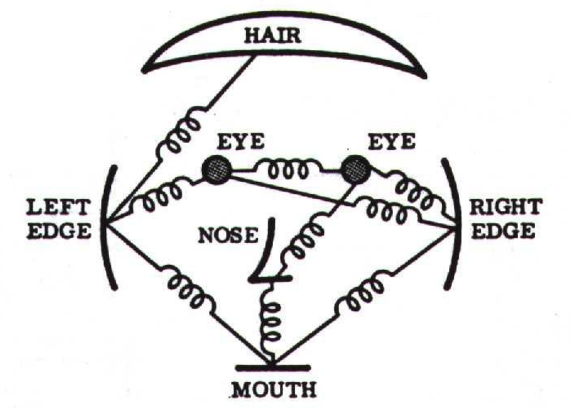
Learn from examples

Difficulties:

- Size variation
- Background clutter
- Occlusion
- Intra-class variation

Motivation of chosen method

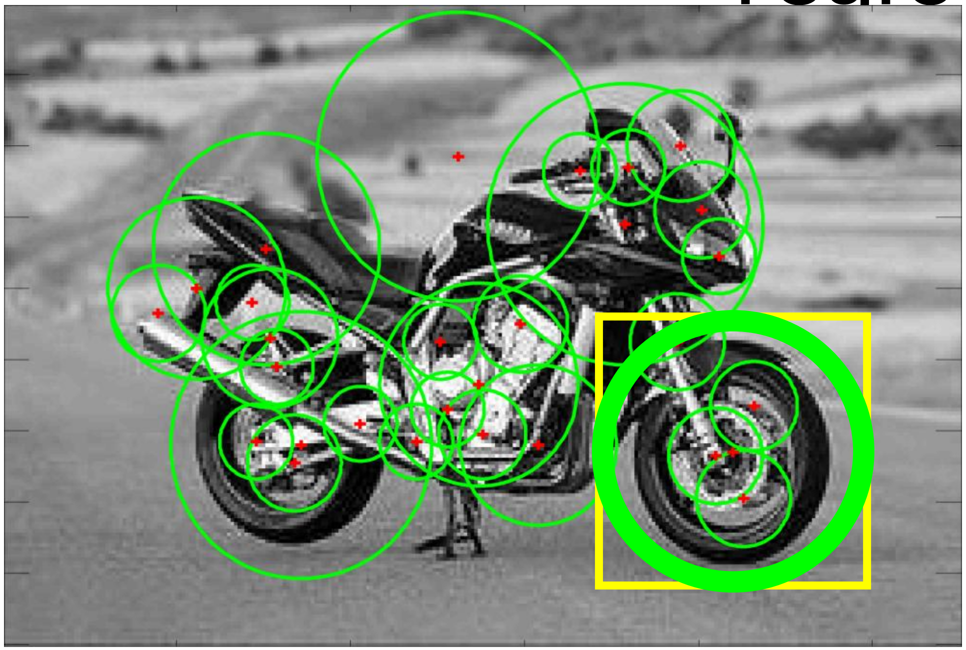
- Model objects as flexible constellation of parts
- Probabilistic model of the object
  - Shape
  - Appearance
  - Occlusion
  - Relative Scale
- EM for learning, Bayesian for classification



Fischler & Elschlager 1973

Detection & Representation of regions

© slide VGG



- Find regions within image
- Use salient region operator

Location

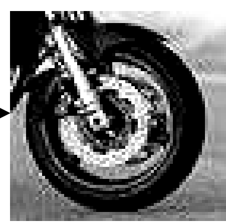
(x,y) coords. of region center

Scale

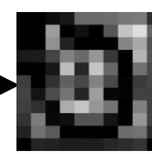
Diameter of region (pixels)

Appearance

Normalize →



11x11 patch →



Projection onto PCA basis →

$$\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{15} \end{pmatrix}$$

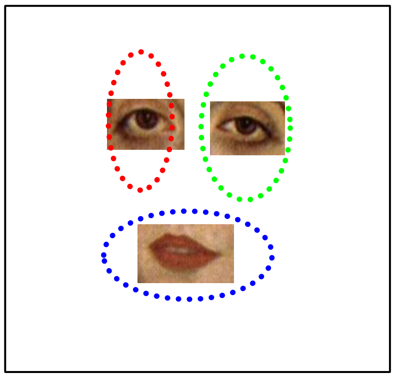
Gives representation of appearance in low-dimensional vector space

Generative probabilistic model

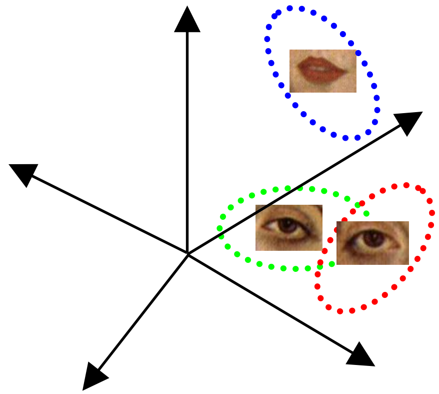
© slide VGG

Foreground model

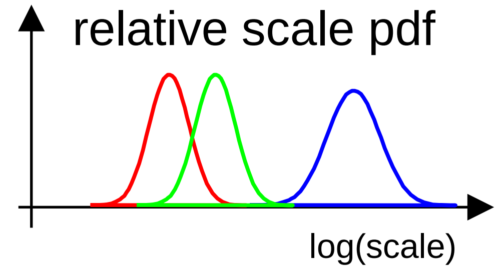
Gaussian shape pdf



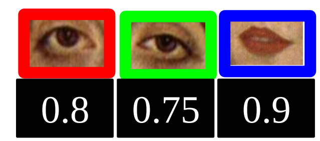
Gaussian part appearance pdf



Gaussian relative scale pdf

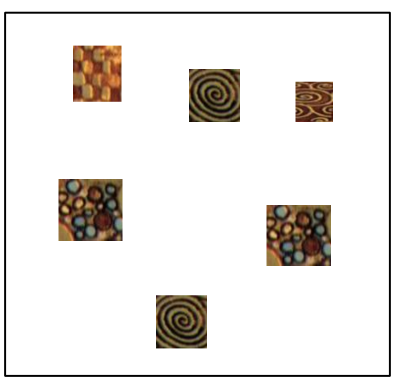


Prob. of detection

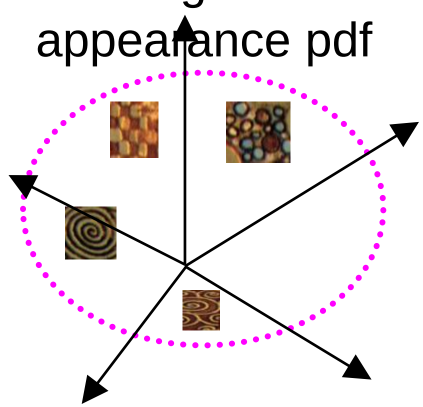


Clutter model

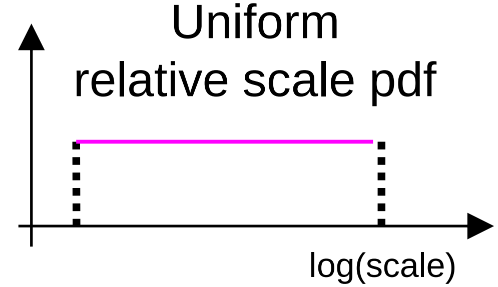
Uniform shape pdf



Gaussian background appearance pdf



Uniform relative scale pdf



Poission pdf on # detections

Formally

- Model Structure

$$
\begin{aligned}
R &= \frac{p(\text{Object}|\mathbf{X},\mathbf{S},\mathbf{A})}{p(\text{No object}|\mathbf{X},\mathbf{S},\mathbf{A})} \\
&= \frac{p(\mathbf{X},\mathbf{S},\mathbf{A}|\text{Object})\; p(\text{Object})}{p(\mathbf{X},\mathbf{S},\mathbf{A}|\text{No object})\; p(\text{No object})} \\
&\approx \frac{p(\mathbf{X},\mathbf{S},\mathbf{A}|\,\theta)\; p(\text{Object})}{p(\mathbf{X},\mathbf{S},\mathbf{A}|\theta_{bg})\; p(\text{No object})}
\end{aligned}
$$

- Likelihood

$$
p(\mathbf{X},\mathbf{S},\mathbf{A}|\,\theta) = \sum_{\mathbf{h}\in H} p(\mathbf{X},\mathbf{S},\mathbf{A},\mathbf{h}|\,\theta) =
$$

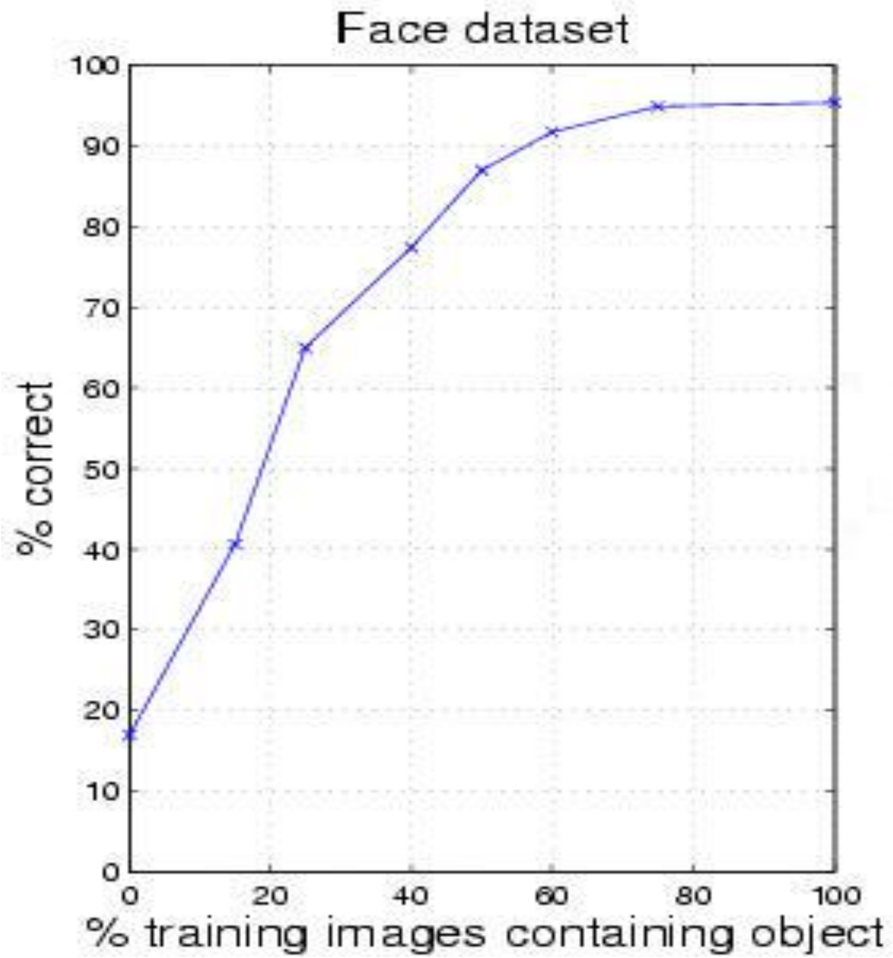
$$
\sum_{\mathbf{h}\in H} \underbrace{p(\mathbf{A}|\mathbf{X},\mathbf{S},\mathbf{h},\theta)}_{Appearance}\,\underbrace{p(\mathbf{X}|\mathbf{S},\mathbf{h},\theta)}_{Shape}\,\underbrace{p(\mathbf{S}|\mathbf{h},\theta)}_{Rel.\ Scale}\,\underbrace{p(\mathbf{h}|\theta)}_{Other}
$$

Hypothesis h: vector of length P (# of parts), each entry in between 1…N (# Feature regions). Background = unassigned feature regions

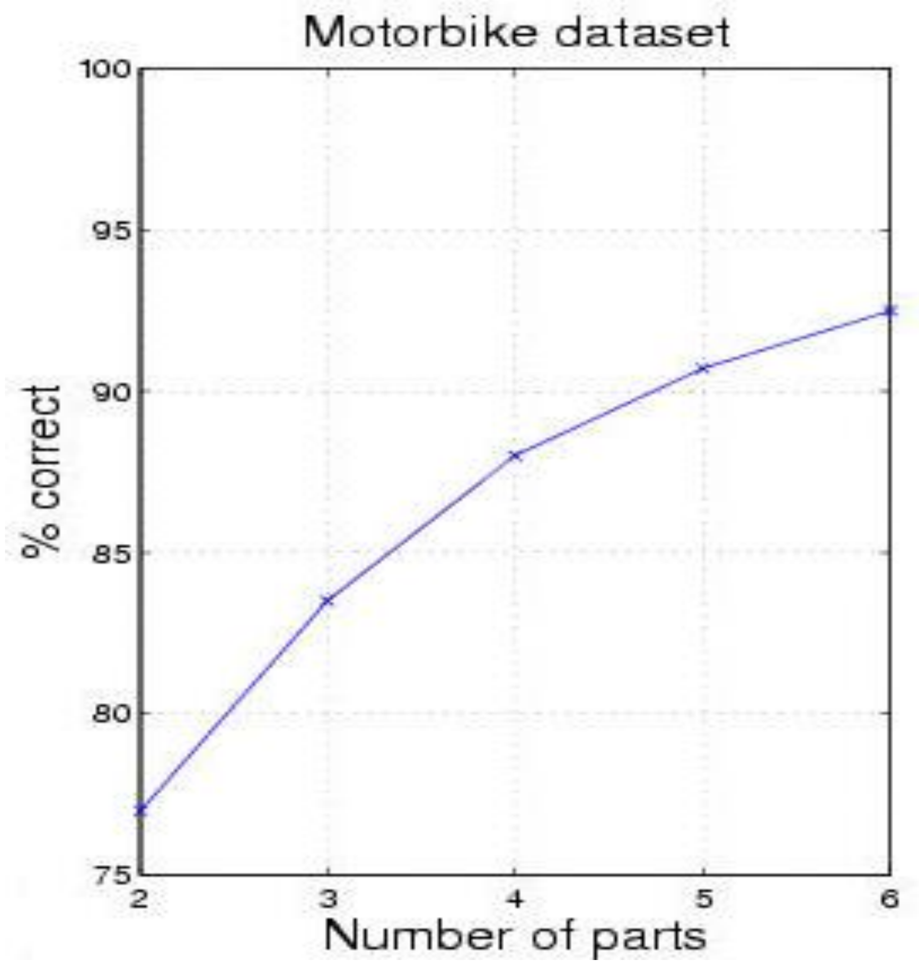
Recognition

- Detect Feature Regions
- Evaluate feature regions using model structure R
- If R> T
  - Presence
- Else
  - Absence
- End If

Results

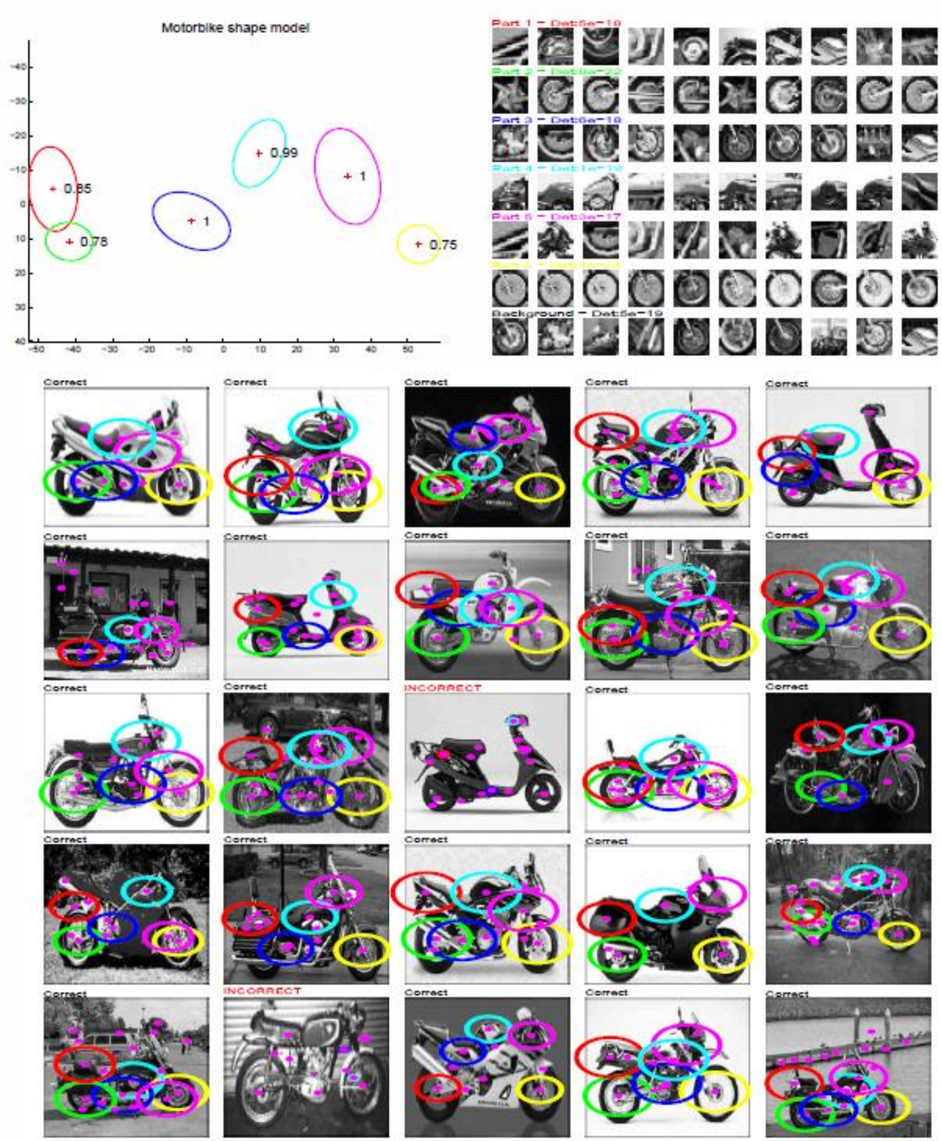


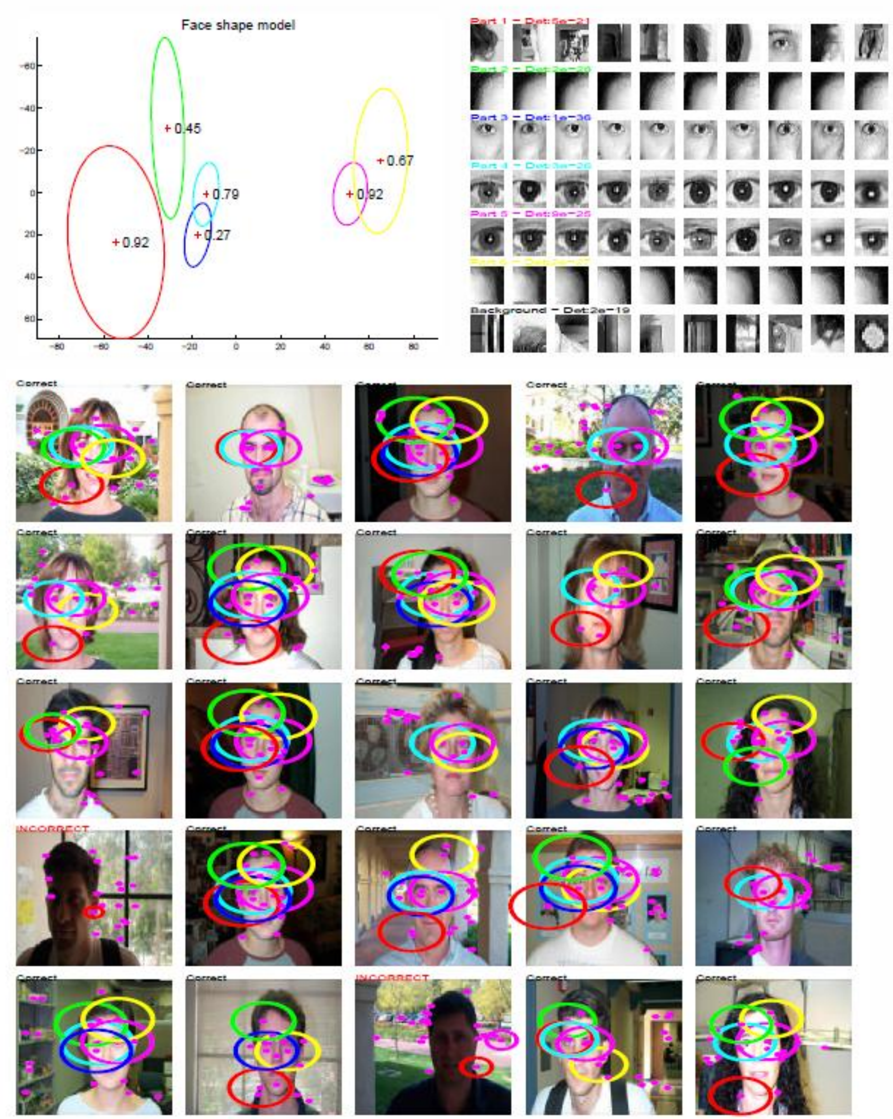
Mixing BG images in training data



Performance drop off with reduced # of parts

Results contd.





Open Discussion

Relative Attributes

- Marr Prize 2011 winner

Relative attributes
D Parikh, K Grauman - Computer Vision (ICCV), 2011 IEEE ..., 2011 - ieeexplore.ieee.org
Abstract Human-nameable visual "attributes" can benefit various recognition tasks. However, existing techniques restrict these properties to categorical labels (for example, a person is 'smiling'or not, a scene is 'dry'or not), and thus fail to capture more general ...
Cited by 345 Related articles All 23 versions Cite Save

Horse

Horse


Horse

Donkey


DreamWorks
Shrek
Shrek Forever After TM & © 2010 DreamWorks Animation LLC
All Rights Reserved.

Donkey

Mule

Problems within Binary Attributes

Some tags are binary while some are relative.

Is furry

Has four-legs

Legs shorter than horses'

Mule

Tail longer than donkeys'

Binary

relative

Has tail

What is visual attributes?

- Attributes are properties observable in images that have human-designated names, such as ‘Orange’, ‘striped’, or ‘Furry’.

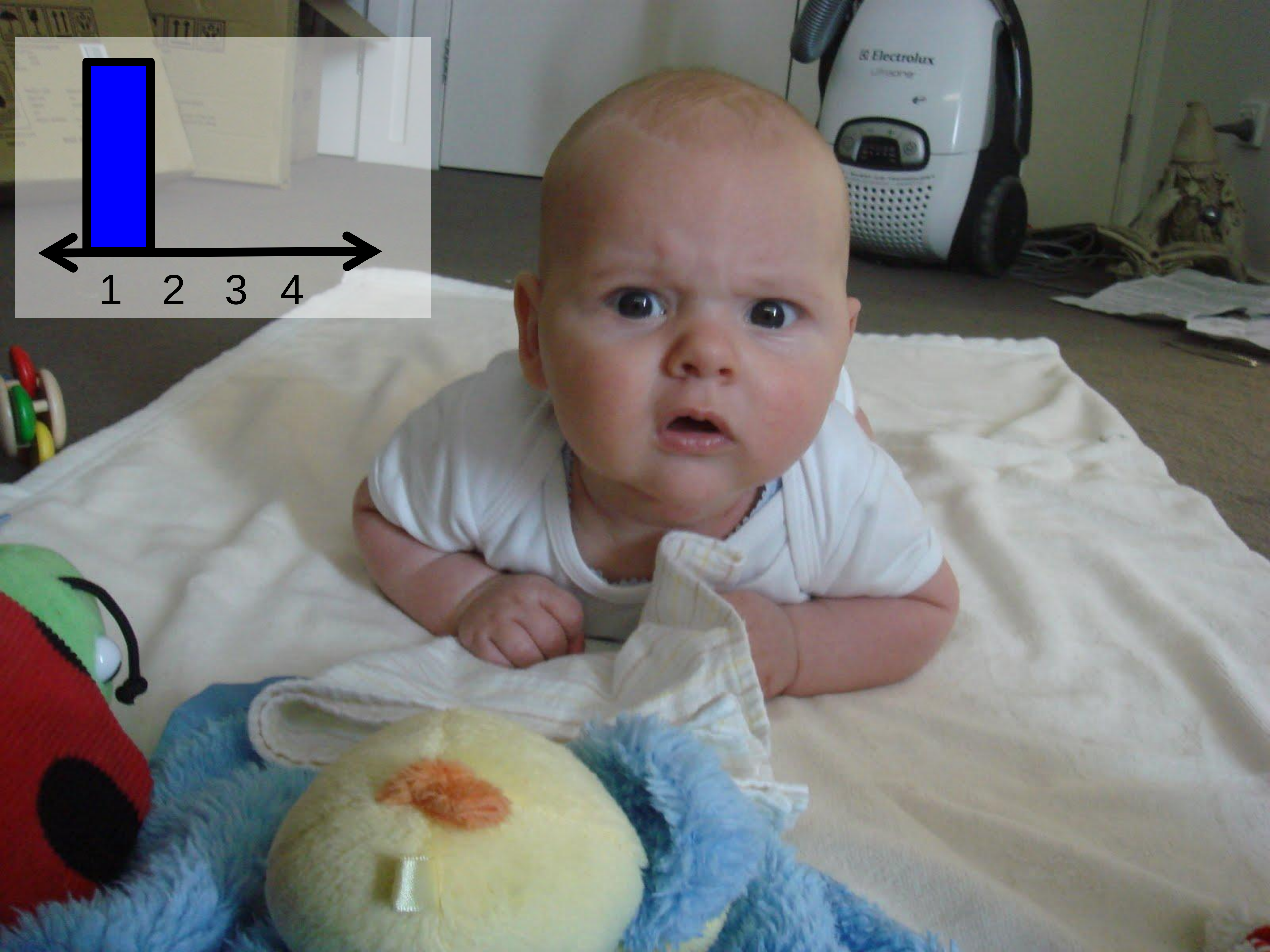
4-Legged	White	Male
Orange	Symmetric	Asian
Striped	Ionic columns	Beard
Furry	Classical	Smiling

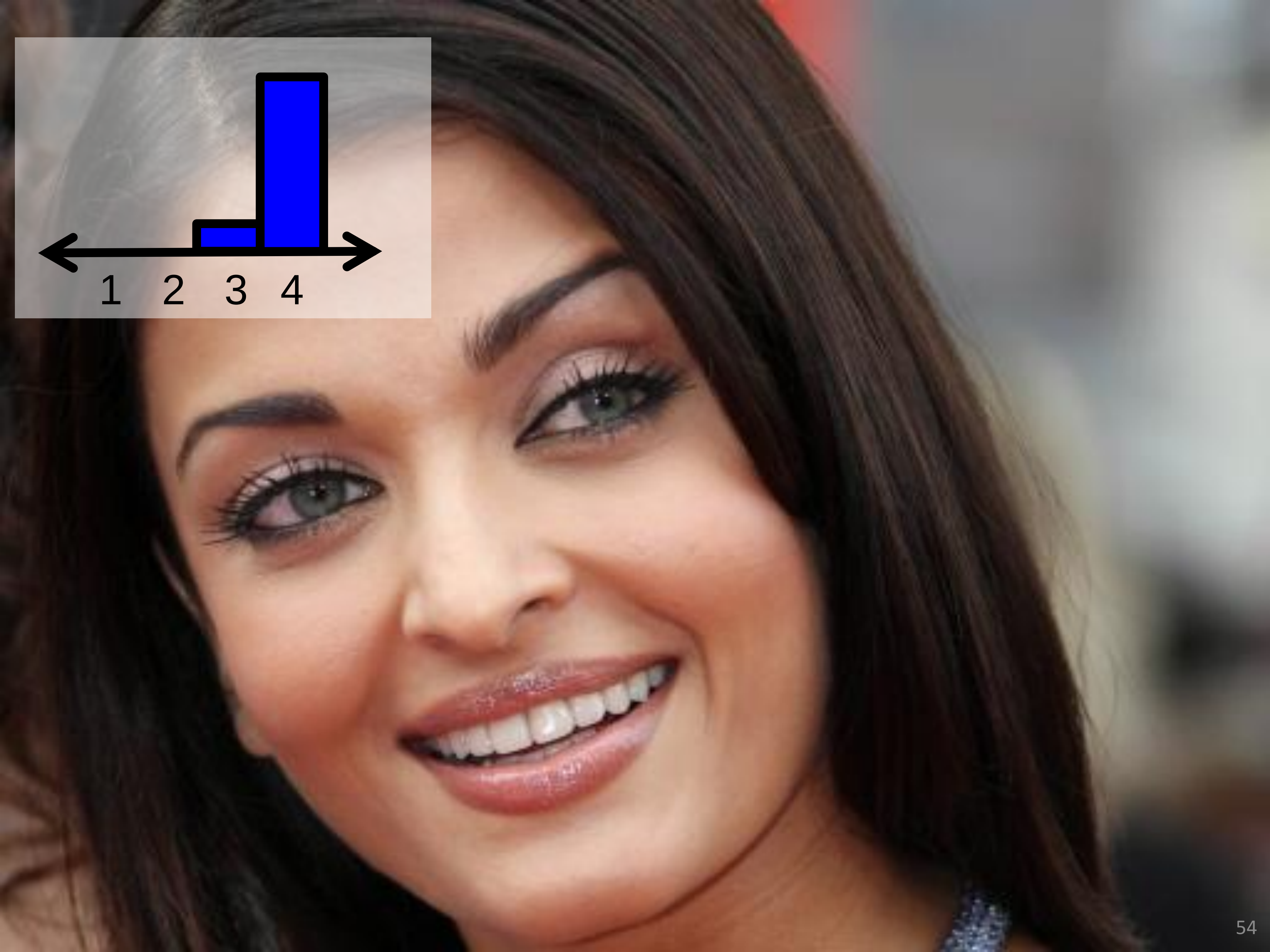
“Downtown Chicago”

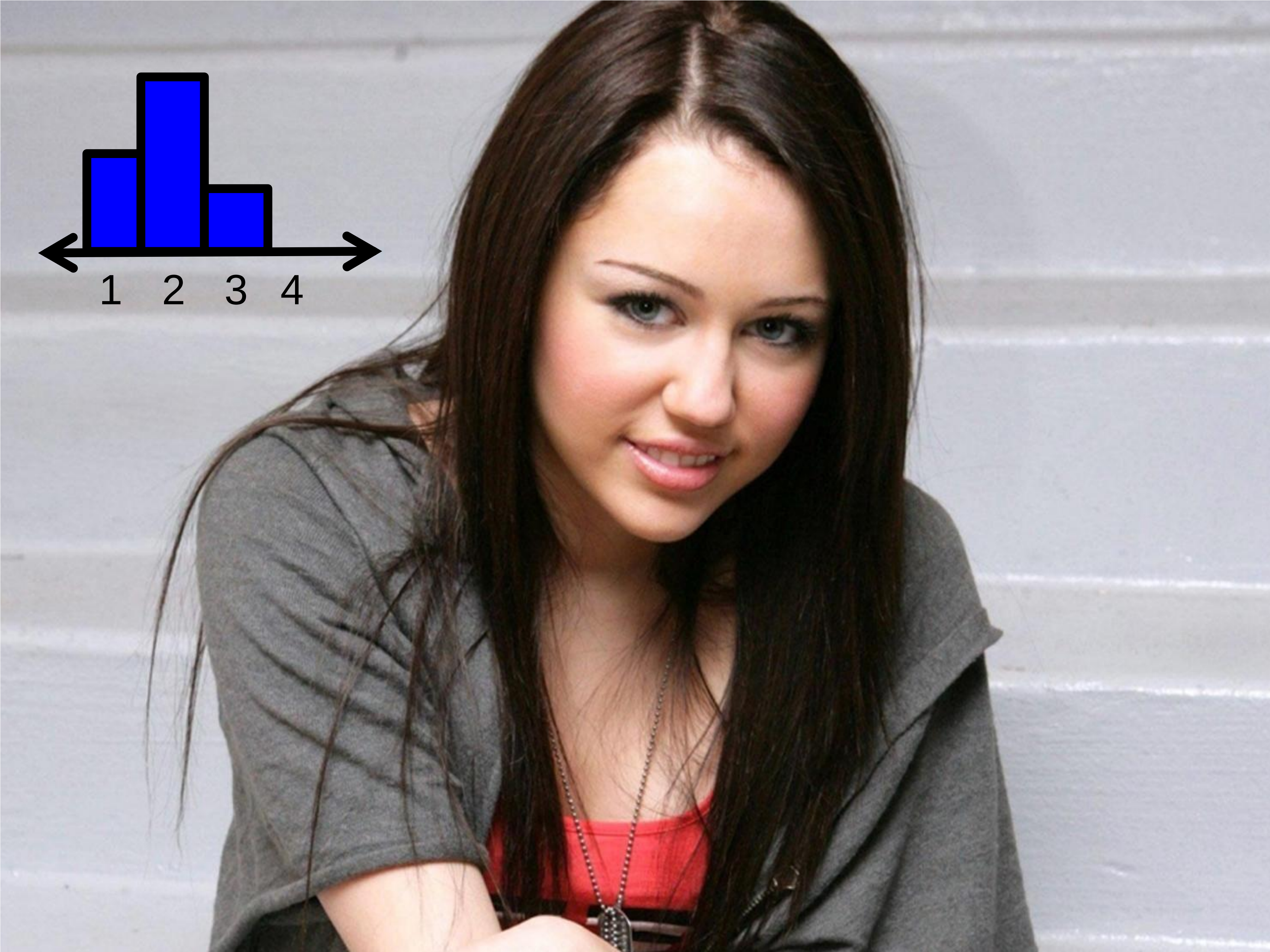

DO NOT ENTER
S Michigan Av
citibank
DO NOT ENTER

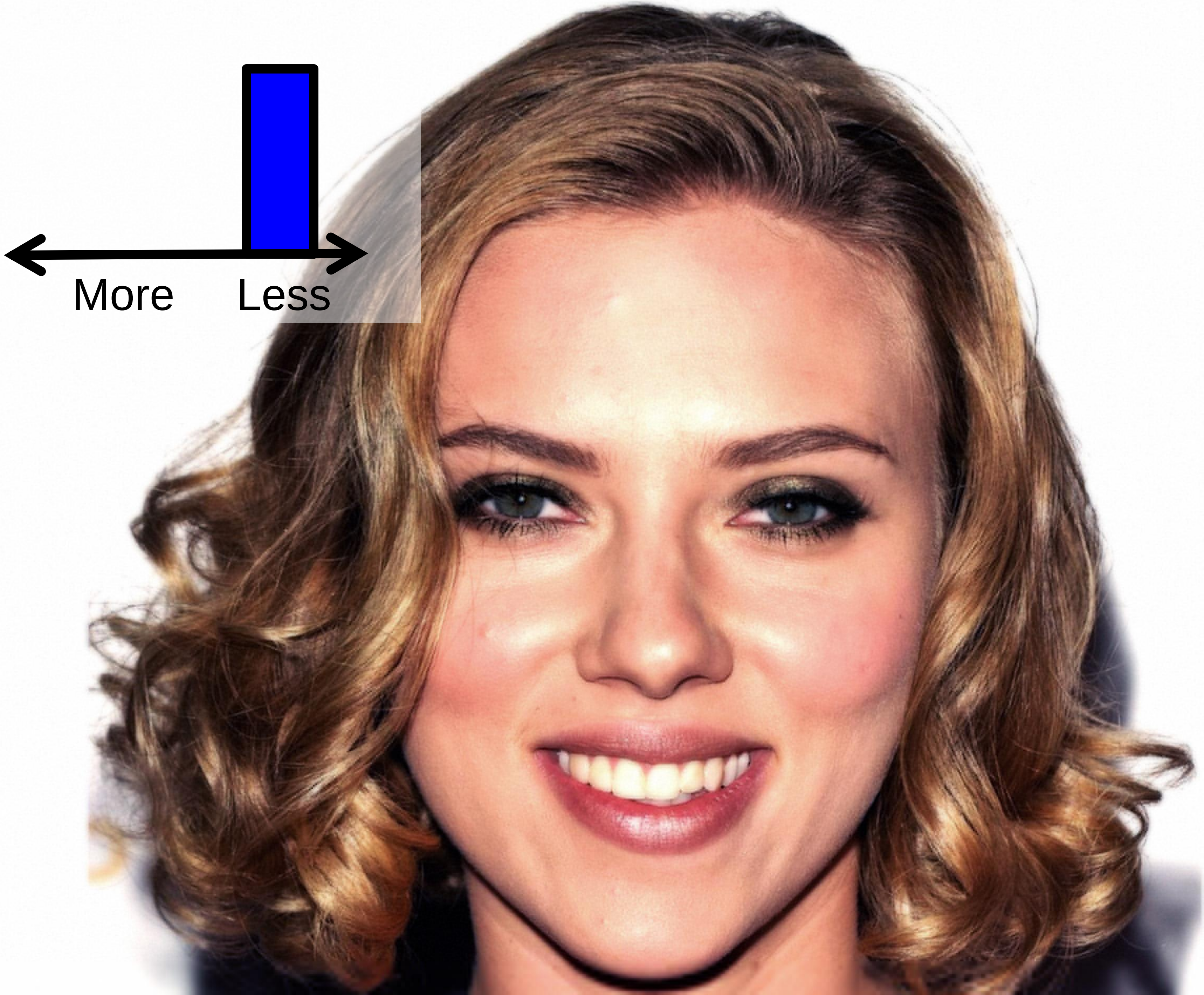
Relative Description


1 2 3 4


1 2 3 4
Electrolux


1 2 3 4


1 2 3 4


More
Less

Labeling data

Binary Attributes

Young:	Yes
Smiling:	No
Young:	Yes
Smiling:	Yes
Young:	Yes
Smiling:	Yes
Young:	No
Smiling:	Yes
Young:	Yes
Smiling:	No

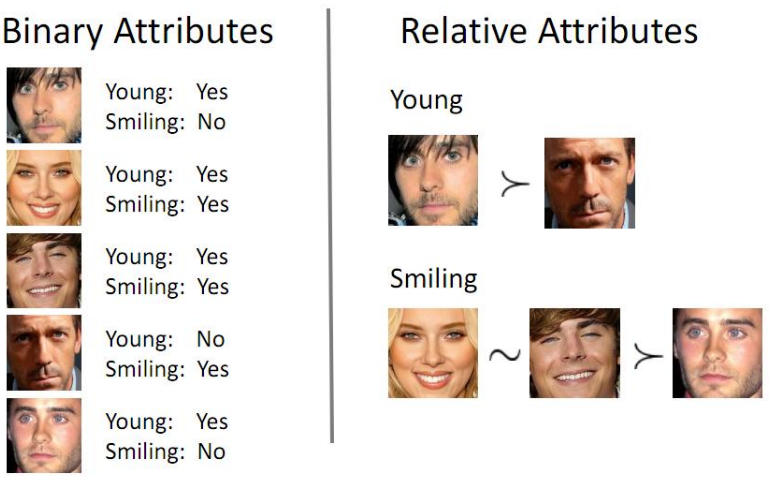
Relative Attributes

Young

Smiling

What is relative attributes?

- Relative attribute indicates **the strength of an attribute** in an image **with respect to other image** rather than simply predicting the presence of an attribute.



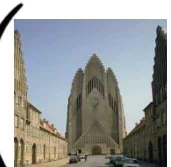
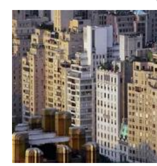
Advantages of Relative Attributes

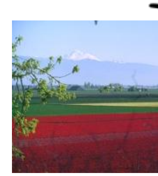
- Enhanced human-machine communication
- More informative
- Natural for humans

Learning Relative Attributes

For each attribute a_m, open

Supervision is

$O_m : \{ ($  $\succ$  $), \dots \},$

$S_m : \{ \{$  $\sim$  $\}, \dots \}$

Learning Relative Attributes

Learn a scoring function $r_m(\boldsymbol{x_i}) = \boldsymbol{w_m^T x_i}$

Image features

Learned parameters

that best satisfies constraints:

$$\forall (i,j) \in O_m : \boldsymbol{w_m^T x_i} > \boldsymbol{w_m^T x_j}$$

$$\forall (i,j) \in S_m : \boldsymbol{w_m^T x_i} = \boldsymbol{w_m^T x_j}$$

Learning Relative Attributes

Max-margin learning to rank formulation

$$\min \quad \left(\frac{1}{2}||\boldsymbol{w_m^T}||_2^2 + C\left(\sum \xi_{ij}^2 + \sum \gamma_{ij}^2\right)\right)$$

$$\text{s.t} \quad \boldsymbol{w_m^T}(\boldsymbol{x_i} - \boldsymbol{x_j}) \geq 1 - \xi_{ij}, \forall (i,j) \in O_m$$

$$|\boldsymbol{w_m^T}(\boldsymbol{x_i} - \boldsymbol{x_j})| \leq \gamma_{ij}, \forall (i,j) \in S_m$$

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

Based on [Joachims 2002]

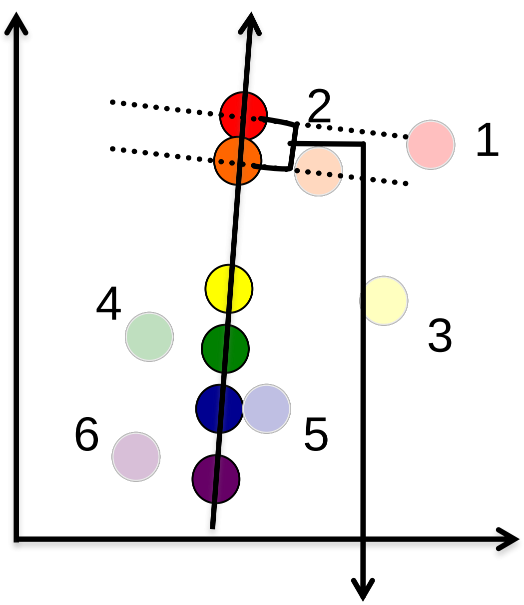
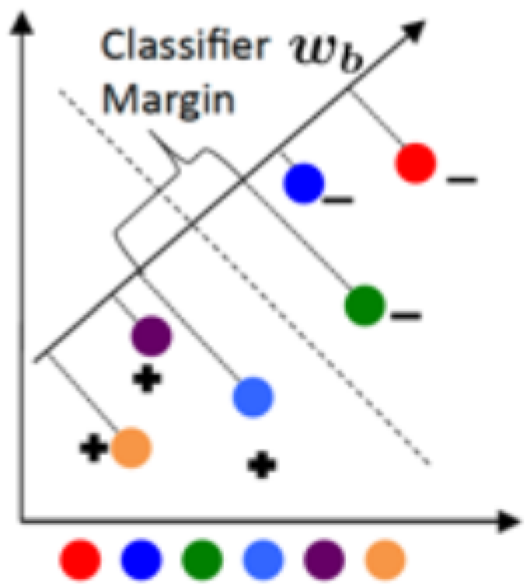


Image → Relative Attribute Score

Learning binary attributes v.s. Learning relative attributes

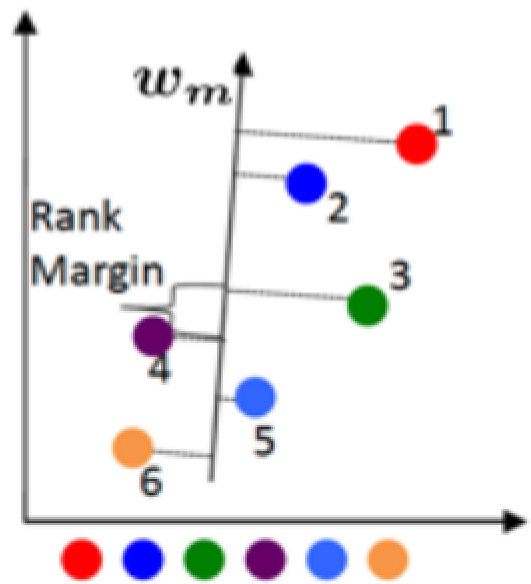
Binary Attributes



Learn decision function

$$d_b(\boldsymbol{x}_i) = \boldsymbol{w}_b^T \boldsymbol{x}_i$$

Relative Attributes



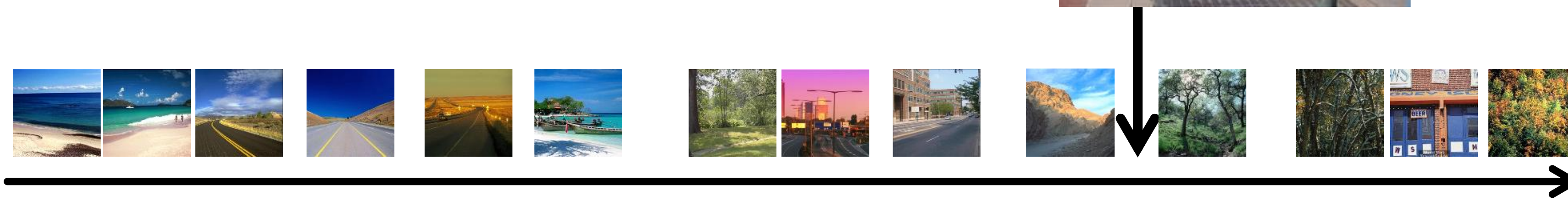
Learn ranking function:

$$r_m(\boldsymbol{x}_i) = \boldsymbol{w}_m^T \boldsymbol{x}_i$$

Automatic Relative Image Description

Density 

Novel image 

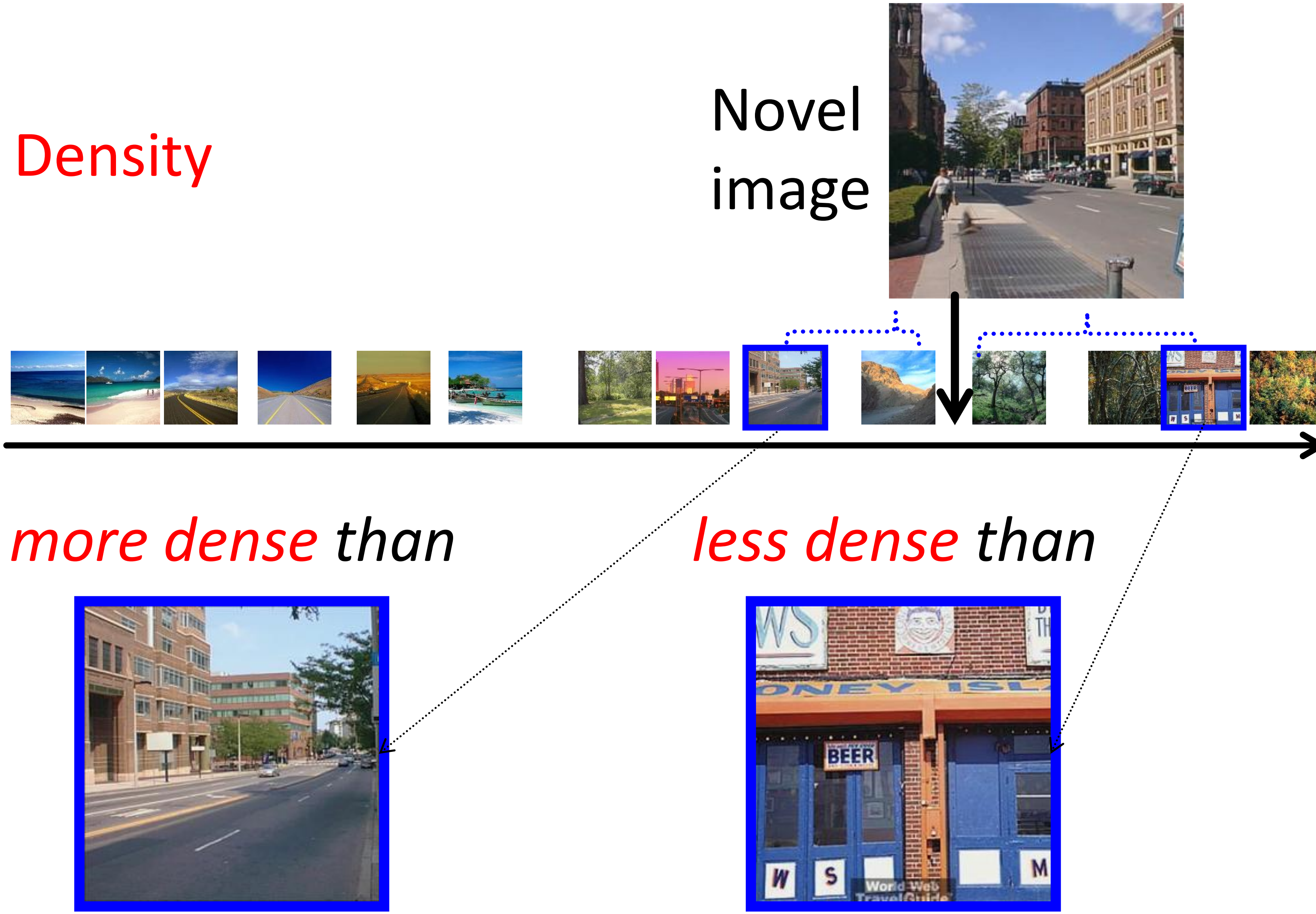


Conventional binary description: *not dense*

Dense: 

Not dense: 

Automatic Relative Image Description



Automatic Relative Image Description

more dense than ***Highways***, *less dense* than ***Forests***

Automatic Relative Image Description

Binary (existing):

Not natural

Not open

Has perspective

Relative (ours):

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

Automatic Relative Image Description

Binary (existing):

Not natural

Not open

Has perspective

Relative (ours):

More natural than tallbuilding
Less natural than forest

More open than tallbuilding
Less open than coast

Has more perspective than tallbuilding

Automatic Relative Image Description

Binary (existing):

Not Young

BushyEyebrows

RoundFace

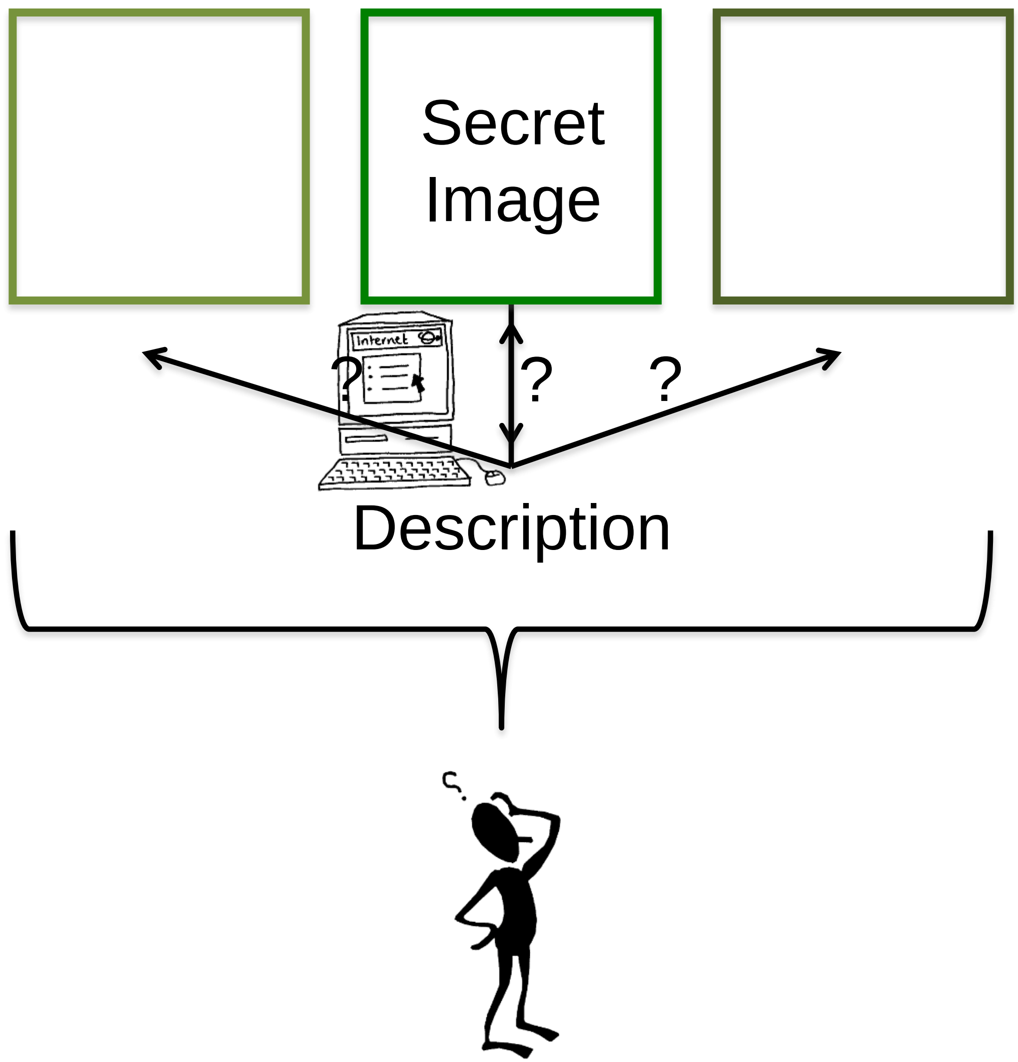
Relative (ours):

More Young than CliveOwen
Less Young than ScarlettJohansson

More BushyEyebrows than ZacEfron
Less BushyEyebrows than AlexRodriguez

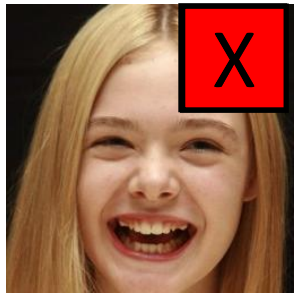
More RoundFace than CliveOwen
Less RoundFace than ZacEfron

Human Studies:
Which Image is Being Described?



Human Studies:

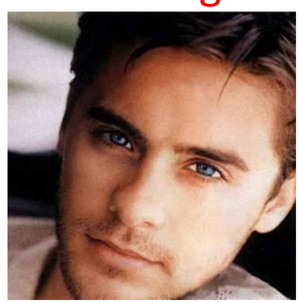
Which Image is Being Described?



Binary: Smiling, Young

Smiling

Young



Not Smiling

Not Young

Relative

More Smiling than

Younger than

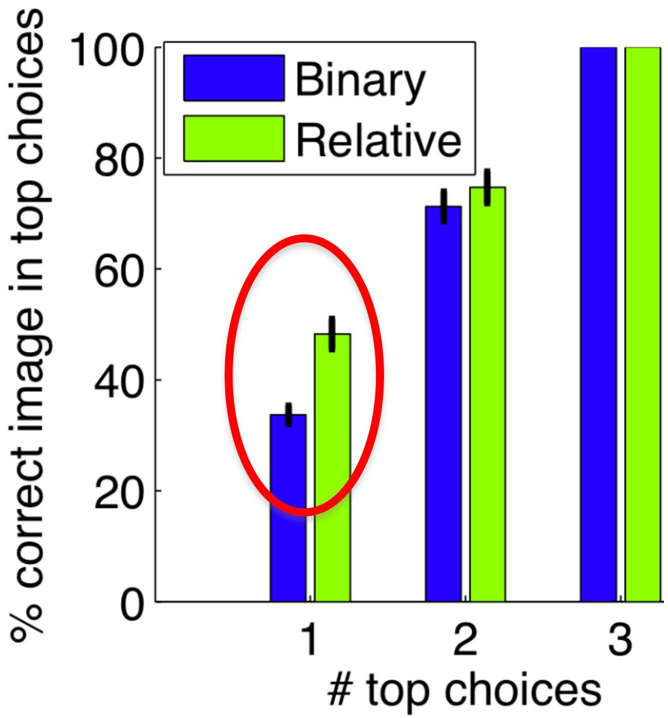
Less Smiling than

Older than

Automatic Relative Image Description

18 subjects

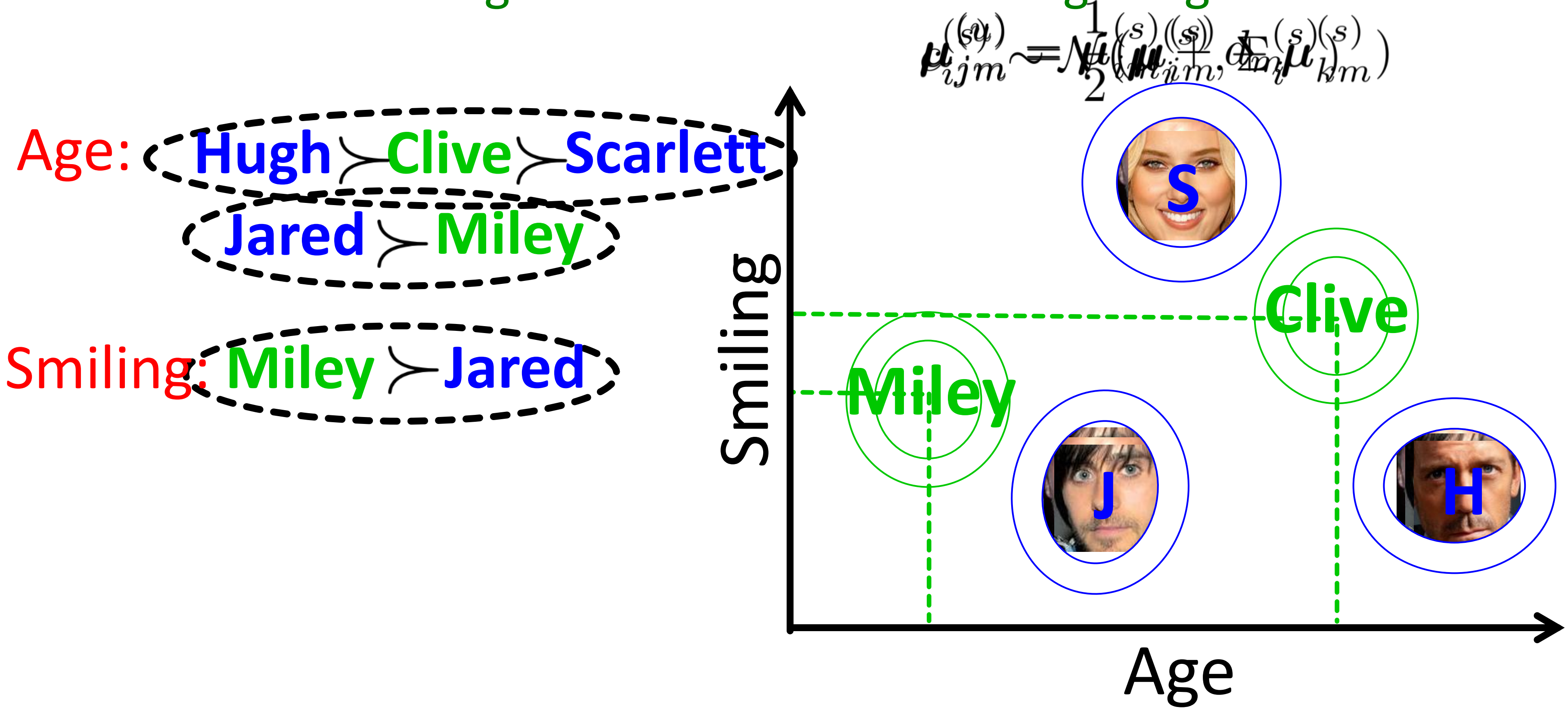
Test cases:
10 OSR, 20 PubFig



Open Discussion

Relative Zero-shot Learning

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



Infer image category using max-likelihood

Relative Zero-shot Learning

Training: Images from **S seen** categories and

Descriptions of **U unseen** categories

Age: **Hugh** $\succ$ **Clive** $\succ$ **Scarlett** **Jared** $\succ$ **Miley**

Smiling: **Miley** $\succ$ **Jared**

Need not use all attributes, or all seen categories

Testing: Categorize image into one of **S**+**U** categories

Method

- Model Structure

$$\begin{aligned} R &= \frac{p(\text{Object}|\mathbf{X},\mathbf{S},\mathbf{A})}{p(\text{No object}|\mathbf{X},\mathbf{S},\mathbf{A})} \\ &= \frac{p(\mathbf{X},\mathbf{S},\mathbf{A}|\text{Object})\, p(\text{Object})}{p(\mathbf{X},\mathbf{S},\mathbf{A}|\text{No object})\, p(\text{No object})} \\ &\approx \frac{p(\mathbf{X},\mathbf{S},\mathbf{A}|\,\theta)\, p(\text{Object})}{p(\mathbf{X},\mathbf{S},\mathbf{A}|\theta_{bg})\, p(\text{No object})} \end{aligned}$$

- Likelihood

$$p(\mathbf{X},\mathbf{S},\mathbf{A}|\,\theta) = \sum_{\mathbf{h}\in H} p(\mathbf{X},\mathbf{S},\mathbf{A},\mathbf{h}|\,\theta) =$$

$$\sum_{\mathbf{h}\in H} \underbrace{p(\mathbf{A}|\mathbf{X},\mathbf{S},\mathbf{h},\theta)}_{Appearance} \underbrace{p(\mathbf{X}|\mathbf{S},\mathbf{h},\theta)}_{Shape} \underbrace{p(\mathbf{S}|\mathbf{h},\theta)}_{Rel.\ Scale} \underbrace{p(\mathbf{h}|\theta)}_{Other}$$

Method contd.

- Appearance

$$\frac{p(\mathbf{A}|\mathbf{X},\mathbf{S},\mathbf{h},\theta)}{p(\mathbf{A}|\mathbf{X},\mathbf{S},\mathbf{h},\theta_{bg})} = \prod_{p=1}^{P}\left(\frac{\mathrm{G}(\mathbf{A}(h_p)|\mathbf{c}_p,V_p)}{\mathrm{G}(\mathbf{A}(h_p)|\mathbf{c}_{bg},V_{bg})}\right)^{d_p}$$

- Shape

$$\frac{p(\mathbf{X}|\mathbf{S},\mathbf{h},\theta)}{p(\mathbf{X}|\mathbf{S},\mathbf{h},\theta_{bg})} = \mathrm{G}(\mathbf{X}(\mathbf{h})|\boldsymbol{\mu},\Sigma)\,\alpha^f$$

- Relative Scale

$$\frac{p(\mathbf{S}|\mathbf{h},\theta)}{p(\mathbf{S}|\mathbf{h},\theta_{bg})} = \prod_{p=1}^{P}\mathrm{G}(\mathbf{S}(h_p)|t_p,U_p)^{d_p}\,r^f$$

- Occlusion

$$\frac{p(\mathbf{h}|\theta)}{p(\mathbf{h}|\theta_{bg})} = \frac{p_{Poiss}(n|M)}{p_{Poiss}(N|M)}\,\frac{1}{{}^nC_r(N,f)}\,p(\mathbf{d}|\theta)$$