LEARNING SUBCATEGORY RELEVANCES FOR CATEGORY RECOGNITION

Sinisa Todorovic and Narendra Ahuja
{sintod, n-ahuja}@uiuc.edu

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

BECKMAN INSTITUTE
AT THE UNIVERSITY OF ILLINOIS

CVPR 2008

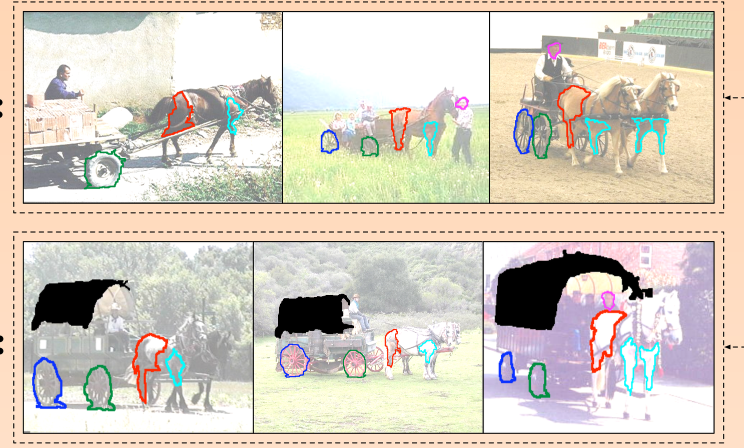
MOTIVATION

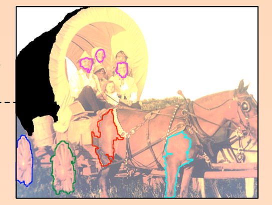
Does the new image belong to category c or c' ?

training images belonging to two categories

new image

Categorization by identifying common parts with training images





parts occurring in both categories

Relevance of image parts for categorization:

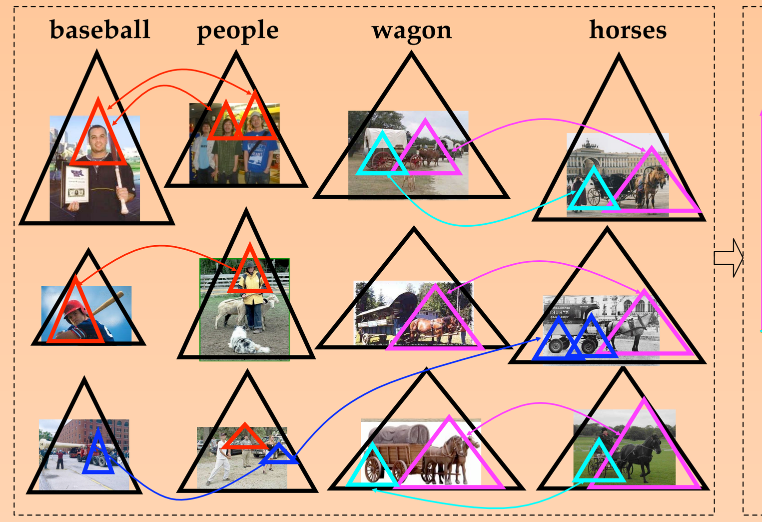
1) Similar regions recurring in training images are relevant = subcategories [1, 2]
3) Subcategories may be shared by many categories, or may be unique for a category
4) Evidence for categorization provided by a subcategory present in the new image:
   - If shared (e.g., wheels, faces) $\Rightarrow$ Poor
   - If unshared (e.g., wagon top) $\Rightarrow$ Strong
5) Relevance of a subcategory:
   - Varies for different image categories
   - Proportional to the relative degree of sharing

PROBLEM STATEMENT

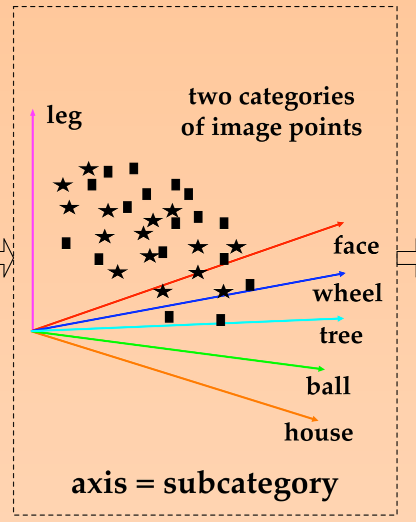
GIVEN a set of images, labeled by a visual category each image belongs to,
DISCOVER all subcategories occurring in the training set,
LEARN a definition of each label category in terms of the subcategories,
LEARN a region-based model of each subcategory that encodes:
1) Likelihood that the image contains the subcategory
2) Prior that the subcategory occurs
3) Relevances of the subcategory to the recognition of each image category

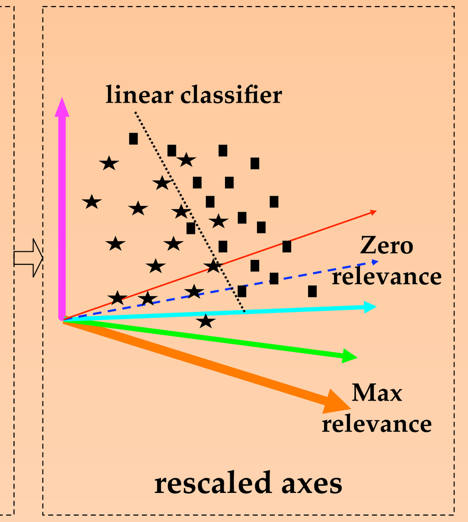
In a new image,
DETECT all occurrences of the subcategories using the learned models.
CATEGORIZE the new image
by accounting for the relevances of subcategory detections for each category

OVERVIEW OF OUR APPROACH



discovery of subcategories in segmentation trees

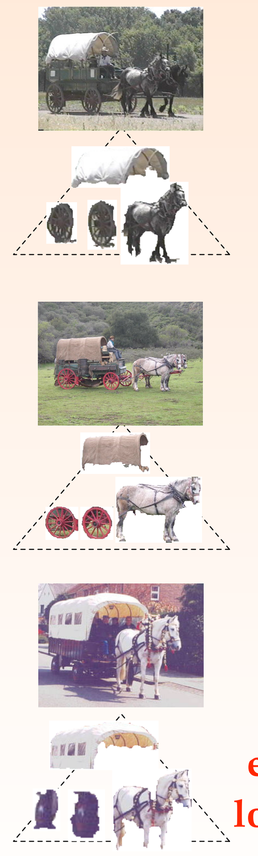




trees = points in the feature space of subcategories

1) Images = Segmentation trees $\Rightarrow$ Similar 2D objects = Similar subtrees [1, 2]
2) Similarity defined in terms of region properties:
   - Geometric (e.g., area, shape)
   - Photometric (e.g., intensity contrast with the surround)
   - Structural -- embedding of subregions within regions
3) Find similar subtrees via tree matching and cluster them $\Rightarrow$ Cluster = Discovered subcategory
4) Learn likelihoods and priors of occurrence of the subcategories from cluster properties
5) Images = Points in the feature space spanned by the posteriors of each subcategory
6) Learn the subcategory relevances by rescaling the feature space so that distance between in-class points < distance between out-of-class points
7) Categorize a new image by using a linear classifier that combines: likelihoods, priors, and relevances of the subcategories detected in the new image

RELATIONSHIP TO PRIOR WORK ON IMAGE CATEGORIZATION



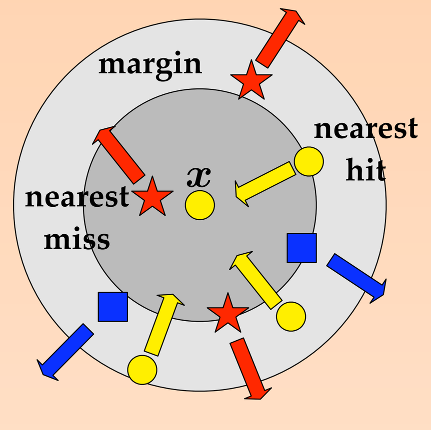
Scene-based (prior work):	Object-based (proposed):
histograms of keypoints	rich region-based image representation
no spatial information	allows object segmentation in addition to categorization
large training sets	small training sets
unclear robustness to occlusion and scale changes	robust to occlusion and scale changes
good performance mostly due to using powerful classifiers	good performance due to: efficient category modeling, local discriminative learning

LEARNING SUBCATEGORY RELEVANCES

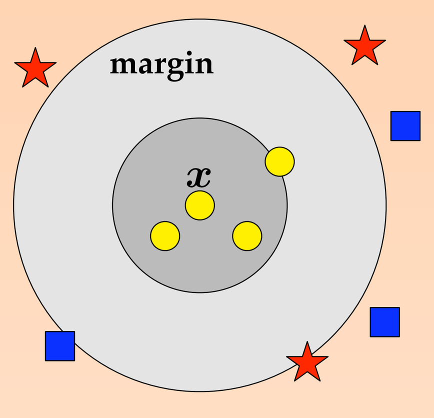
Goal: Maximize the discriminative power of 1-NN classifier over all image categories

Unknown about image categories: $\Longrightarrow$	1-NN classifier is suitable because:
1) Underlying distributions	1) Allows local learning of decision boundaries (efficient)
2) Decision boundaries between them	2) Discriminative power: hypothesis margin $\leq$ sample margin
	3) Maximizing the hypothesis margin $\Rightarrow$ Small generalization error
	4) Probability of error $\leq 2 \cdot$ Bayes probability of error

BEFORE

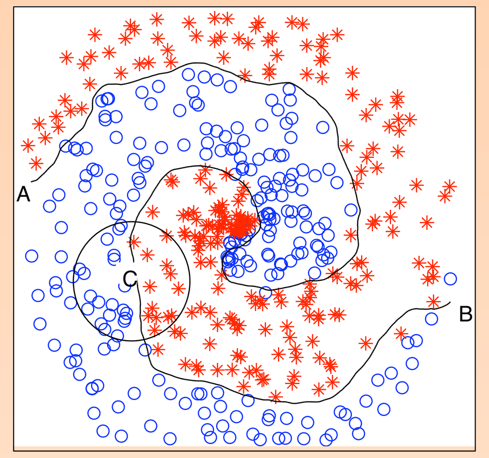


AFTER



images with the same label as x

images with different labels from that of x



Decision boundary A-C-B is complex, but locally linear in a neighborhood of C

Subcategory relevances for category c are learned by maximizing the hypothesis margin of 1-NN classifier

$$w(c) = \max_{w\in\mathbb{W}} \sum_{\substack{x\in\mathcal{X}\\ y=c}} \underbrace{[d_w(x,m(x)) - d_w(x,h(x))]}_{\text{hypothesis margin}} \quad \Longrightarrow \quad w(c) = \max_{w\in\mathbb{W}} w^{\mathrm{T}} \sum_{\substack{x\in\mathcal{X}\\ y=c}} [|x-m(x)| - |x-h(x)|]$$

$$d_w(x,x') = w^{\mathrm{T}}|x-x'|$$

Notation and Definitions:

image - $x = [x_1, \dots, x_i, \dots, x_n]^{\mathrm{T}} \in \mathcal{X}$

image label - $y \in \mathcal{Y} = \{1, \dots, c, \dots, C\}$

region properties - ψ

subcategory - i

posterior - $x_i = P(\psi|i)P(i)$

relevance weights - $w = [w_1, \dots, w_i, \dots, w_n]^{\mathrm{T}}$

simplex - $\mathbb{W} = \{w : w\in\mathbb{R}^n, \|w\|=1, w\geq 0\}$

distance - $d_w(x,x') = w^{\mathrm{T}}|x-x'|$

hits - $\mathbb{H}(x) = \{x' : x'\in\mathcal{X}, y'\in\mathcal{Y}, x'\neq x, y'=y\}$

misses - $\mathbb{M}(x) = \{x' : x'\in\mathcal{X}, y'\in\mathcal{Y}, y'\neq y\}$

nearest hit - $h(x) = \arg\min_{x'\in\mathbb{H}(x)} d_w(x,x')$

nearest miss - $m(x) = \arg\min_{x'\in\mathbb{M}(x)} d_w(x,x')$

Problem:

How to address the uncertainty about nearest hits and misses?

Approach (EM algorithm):

Replace nearest hits and misses with their expected values

$$w(c) = \max_{w\in\mathbb{W}} w^{\mathrm{T}} \underbrace{\sum_{\substack{x\in\mathcal{X}\\ y=c}} [\overline{|x-m(x)|} - \overline{|x-h(x)|}]}_{z(c)}$$

where

$$\overline{|x-m(x)|} = \textstyle\sum_{x'\in\mathbb{M}(x)} |x-x'| P(x'=m(x))$$

$$\overline{|x-h(x)|} = \textstyle\sum_{x'\in\mathbb{H}(x)} |x-x'| P(x'=h(x))$$

$$\Downarrow$$

E-step:

Find $P^{(t)}(x'=m(x))$ and $P^{(t)}(x'=h(x))$ using $w^{(t)}$

M-step:

$$w^{(t+1)}(c) = \max_{w\in\mathbb{W}} w^{\mathrm{T}} z^{(t)}(c) = \frac{[z^{(t)}(c)]_+}{\|[z^{(t)}(c)]_+\|}$$

where $[a]_+ = \max(0,a)$

THEORETICAL RESULTS

Lemma:

EM-based estimation of the subcategory relevances has a closed-form solution

Theorem:

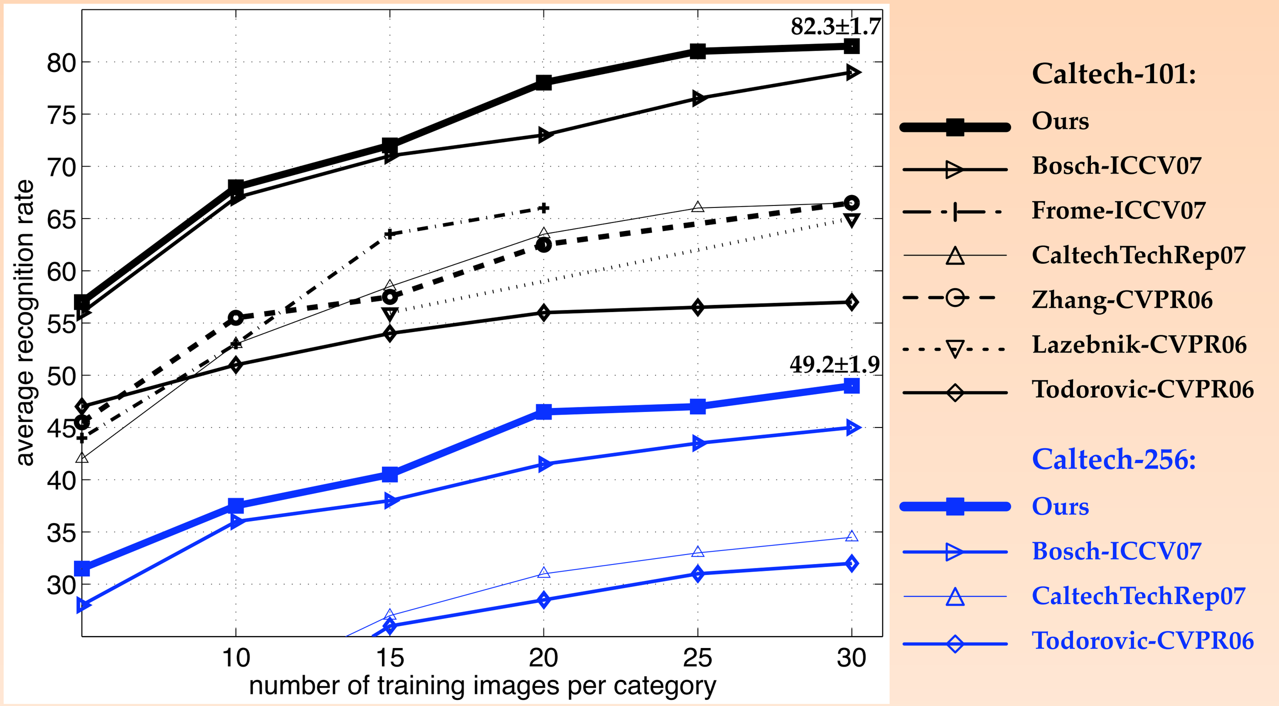
The learning algorithm for estimating the subcategory relevances converges to a unique, global solution regardless of the initialization point

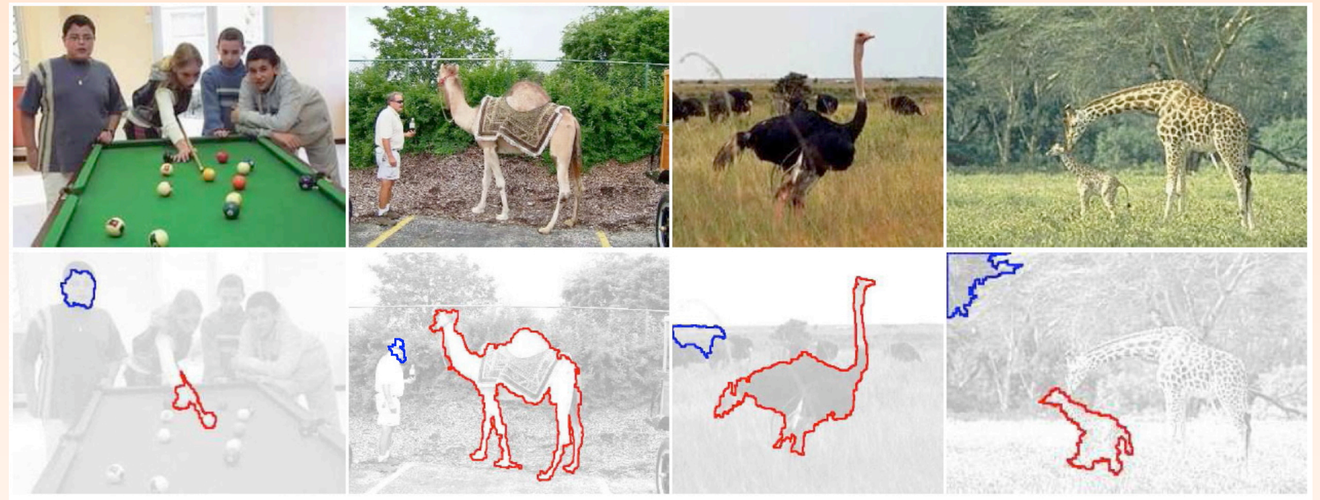
IMAGE CATEGORIZATION USING A LINEAR CLASSIFIER

$$\text{new image} = x \quad \Longrightarrow \quad y = \arg\max_{c\in\mathcal{Y}} w^{\mathrm{T}}(c)\cdot x$$

EXPERIMENTAL RESULTS

Best published categorizations on Caltech-101 and Caltech-256




Caltech-256 images from categories: billiards, camel, ostrich, and giraffe

Most relevant — Least relevant

Use of regions as image features $\Rightarrow$ Simultaneous categorization and segmentation

ACKNOWLEDGMENT

The support of the Office of Naval Research under grant N00014-06-1-0101 is gratefully acknowledged

REFERENCES

[1] Ahuja, Todorovic, "Learning the taxonomy and models of categories present in arbitrary images," in ICCV, 2007
[2] Todorovic, Ahuja, "Unsupervised category modeling, recognition and segmentation in images," in TPAMI, 2008