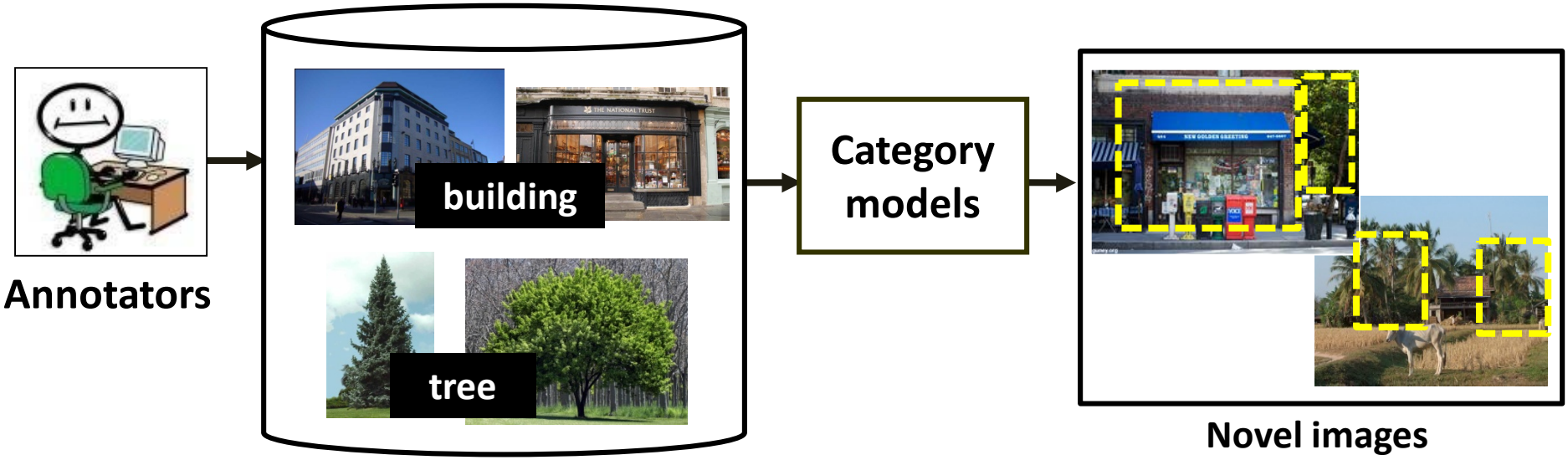
ECS 289G: Visual Recognition
Fall 2015

Yong Jae Lee
Department of Computer Science

Plan for today

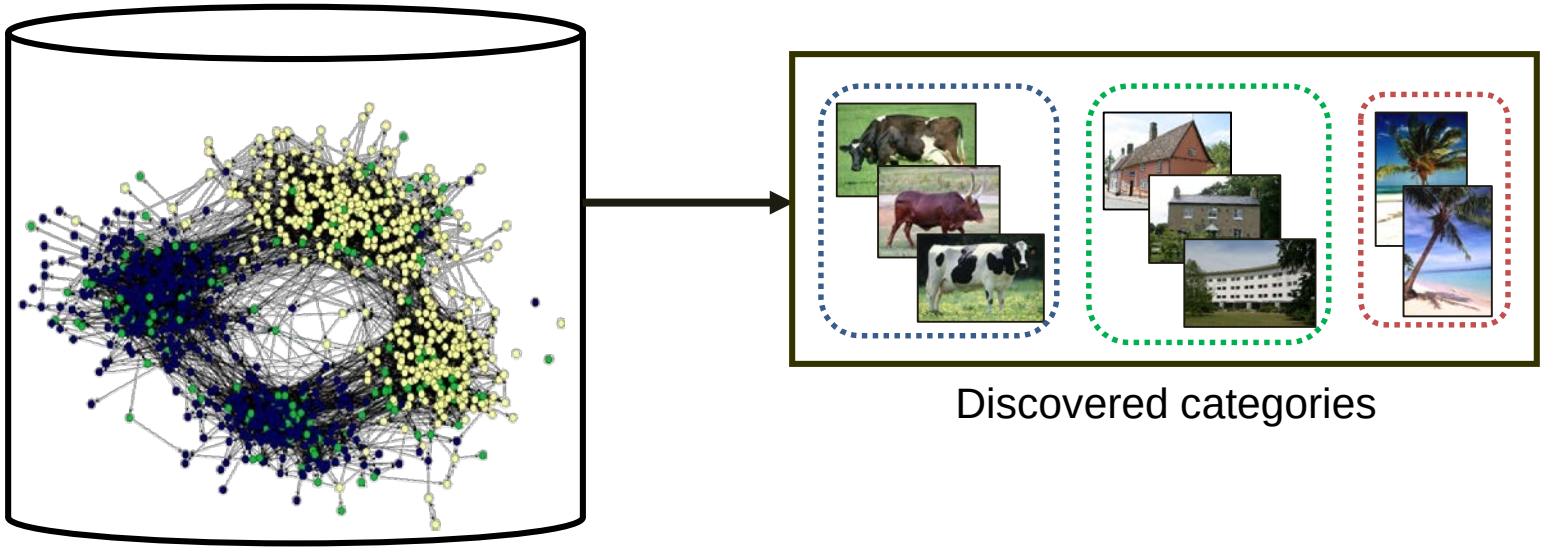
- Questions?
- Sign-up for paper
- Research overview

Standard supervised visual learning

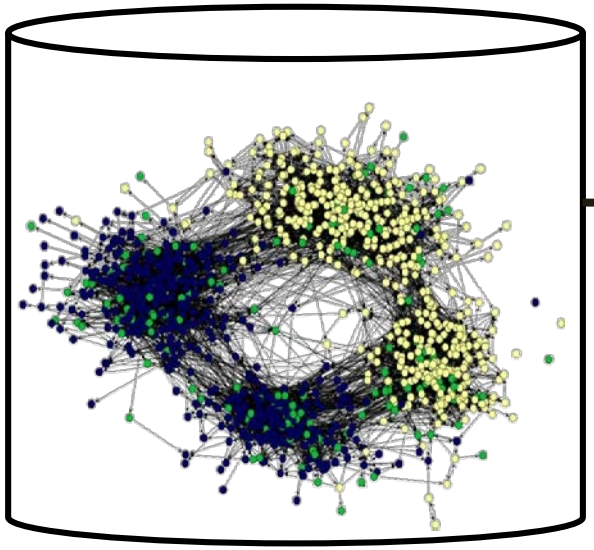


- Number of training images required can be costly
- Assumes closed-world setting where all categories are known

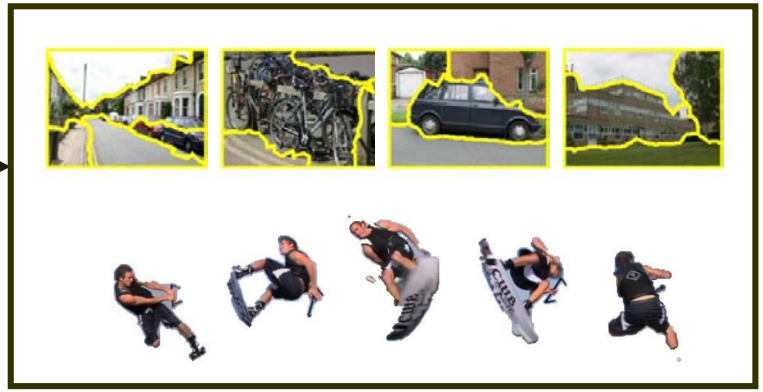
Unsupervised visual discovery



Unsupervised visual discovery

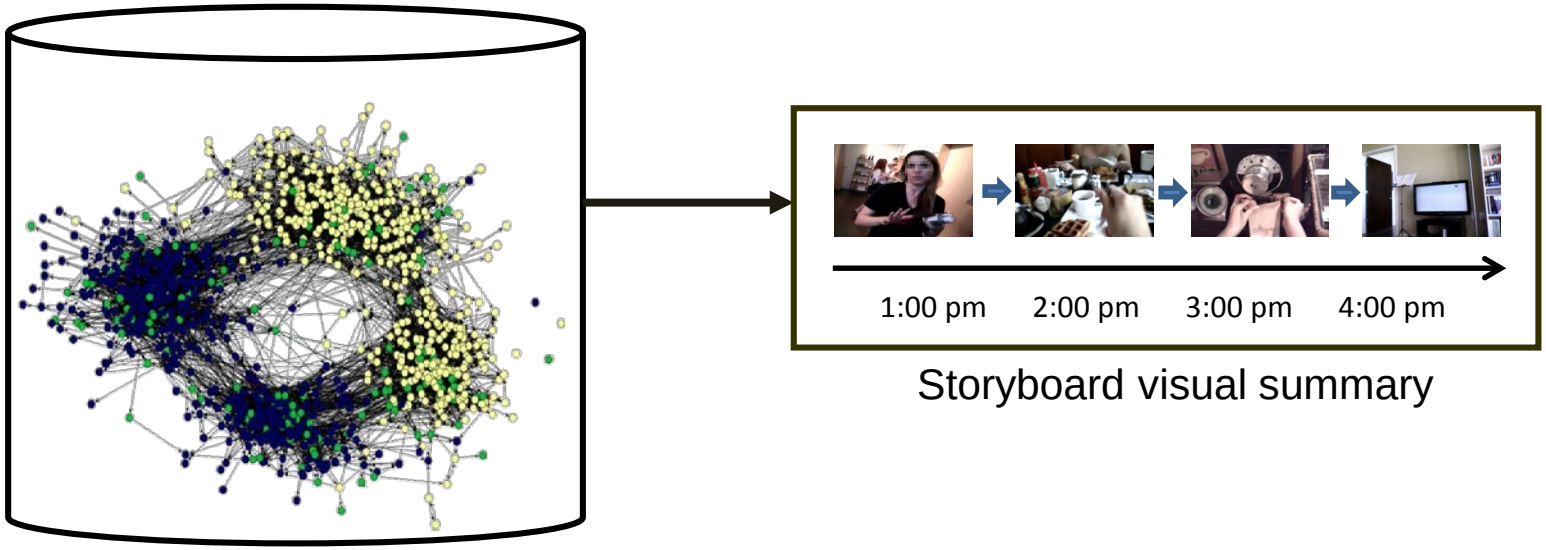


Visual world



Object segmentations
in images and video

Unsupervised visual discovery



Visual world

Storyboard visual summary

- No human to explicitly guide visual recognition process

Why visual discovery?

Exploring new environments

Why visual discovery?

Summarization

MSR Sensecam

Why visual discovery?

flickr

6 billion images

facebook

70 billion images

the simple image sharer imgur

1 billion images served daily

YouTube

100 hours uploaded per minute

photobucket

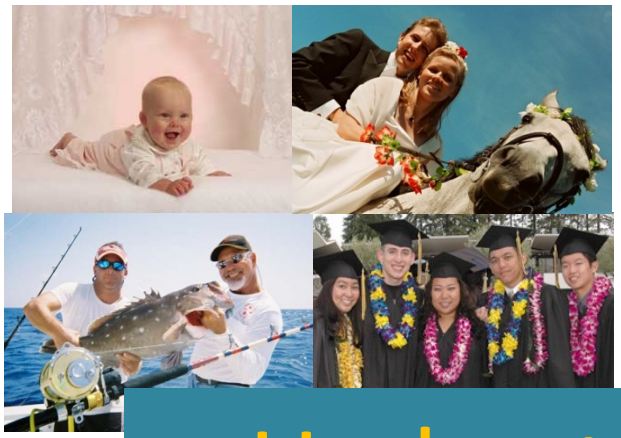
10 billion images

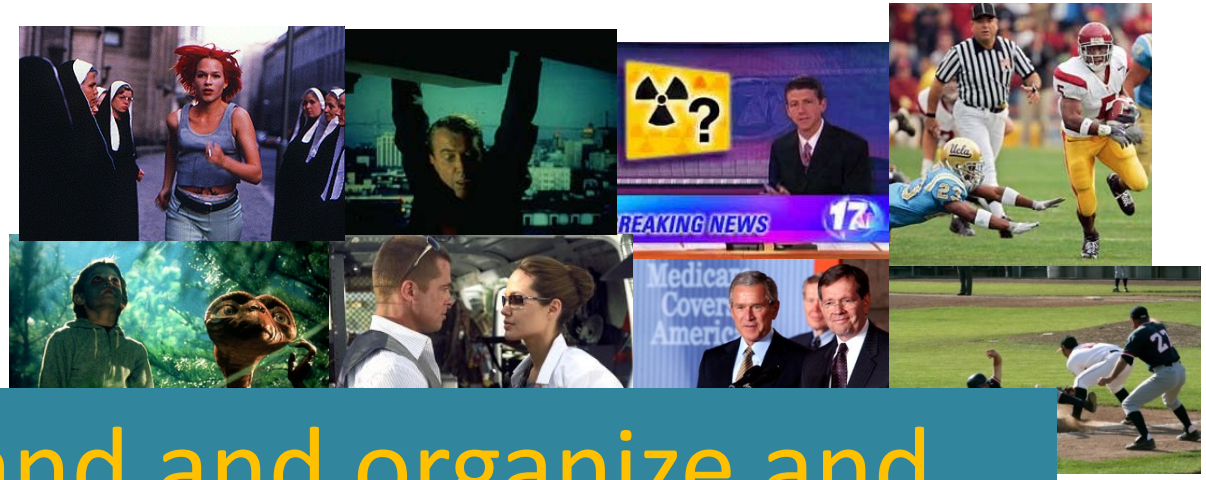
From cisco:

Almost 90% of web traffic is visual!

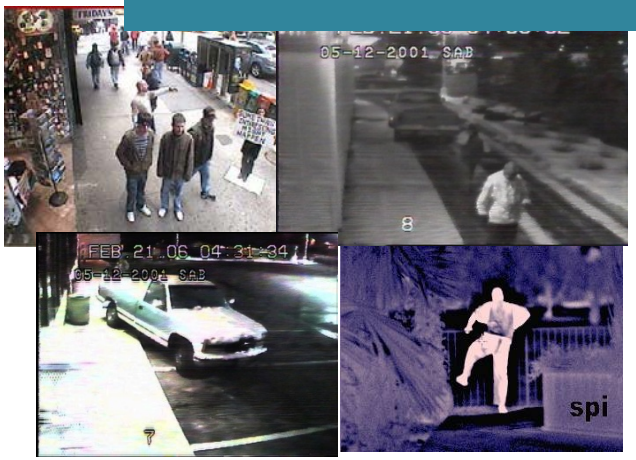
Most of it is unlabeled!!

Inputs today

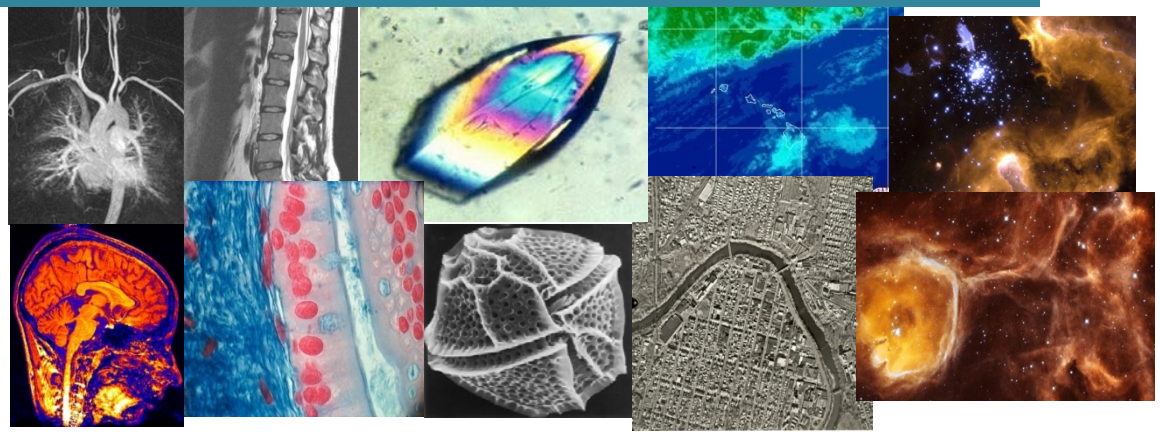




Understand and organize and index all this data!!



Surveillance and security



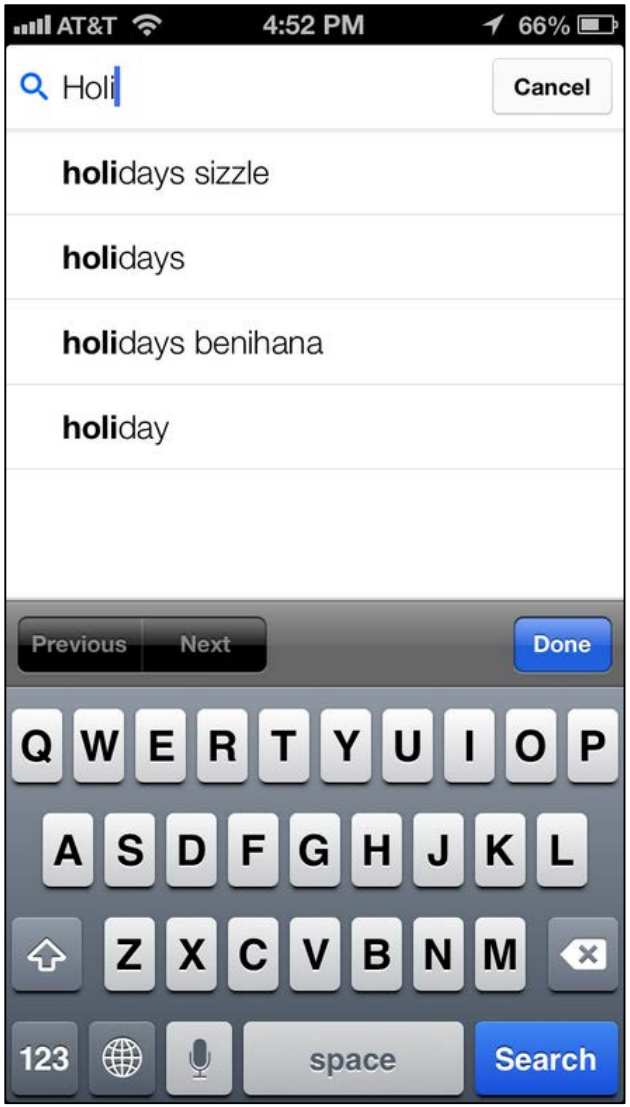
Medical and scientific images

Svetlana Lazebnik

Let’s first explore…

what we can do with big data!

Everyday use of big data: Predictive text



Big visual data

flickr

6 billion images

facebook

70 billion images

the simple image sharer imgur

1 billion images served daily

YouTube

100 hours uploaded per minute

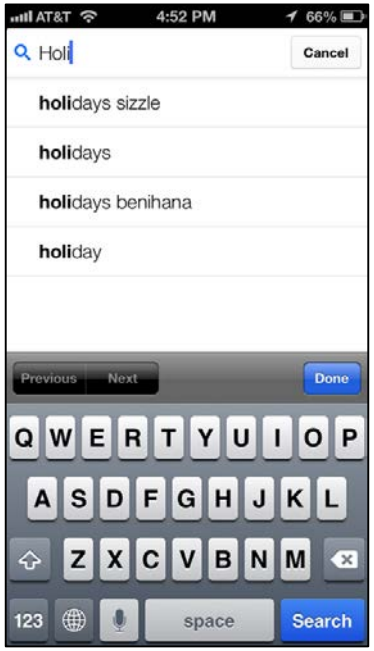
photobucket

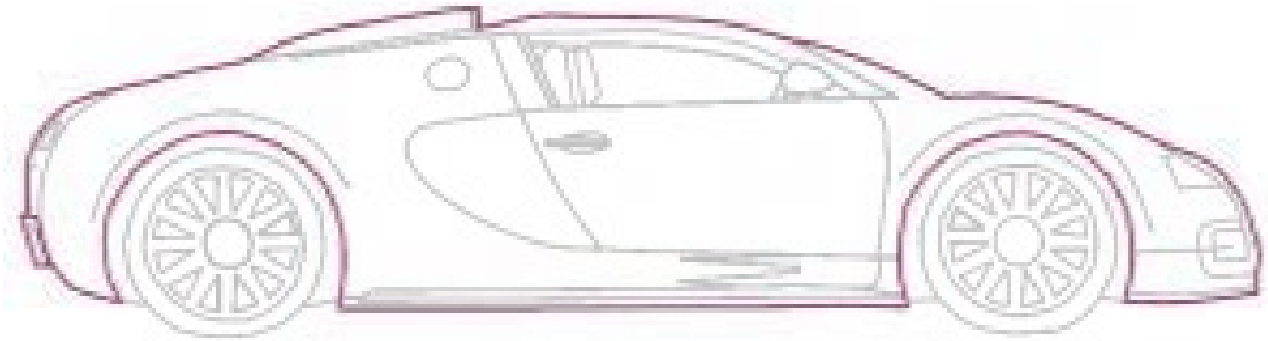
10 billion images

From CISCO:

Almost 90% of web traffic is visual!

Predictive drawing?





Drawing examples

Drawing examples

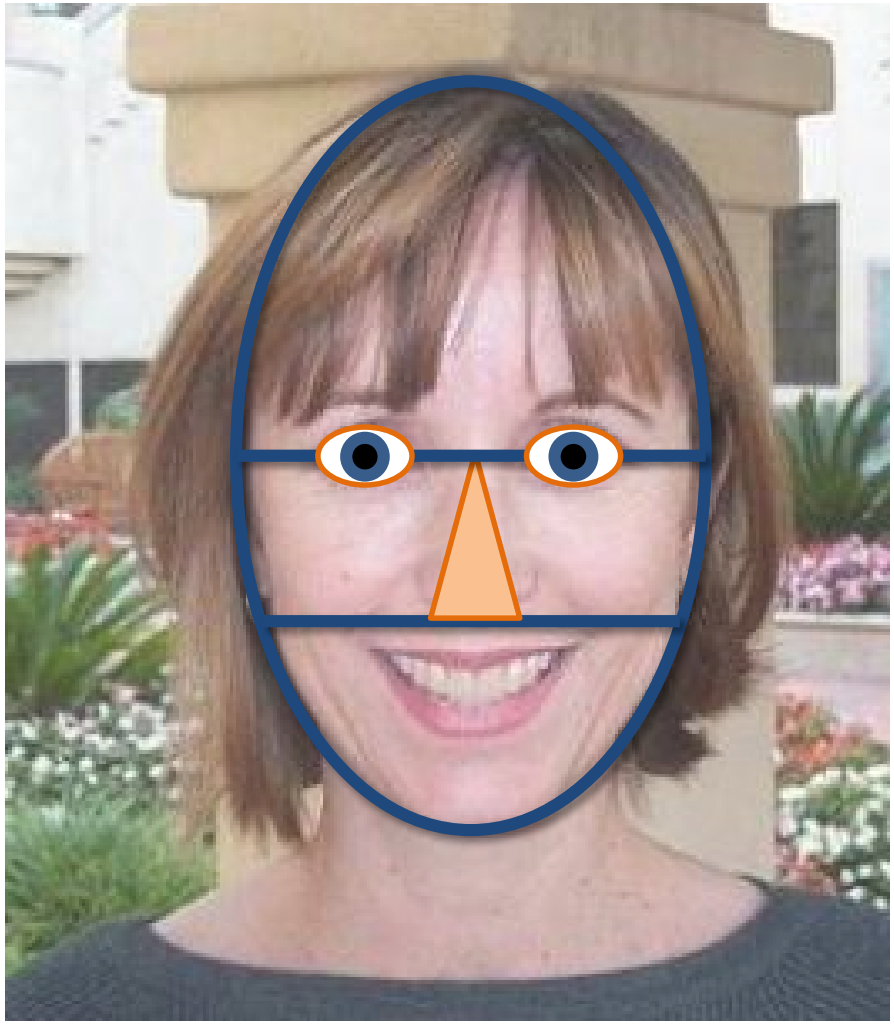
Subject 1

Subject 2

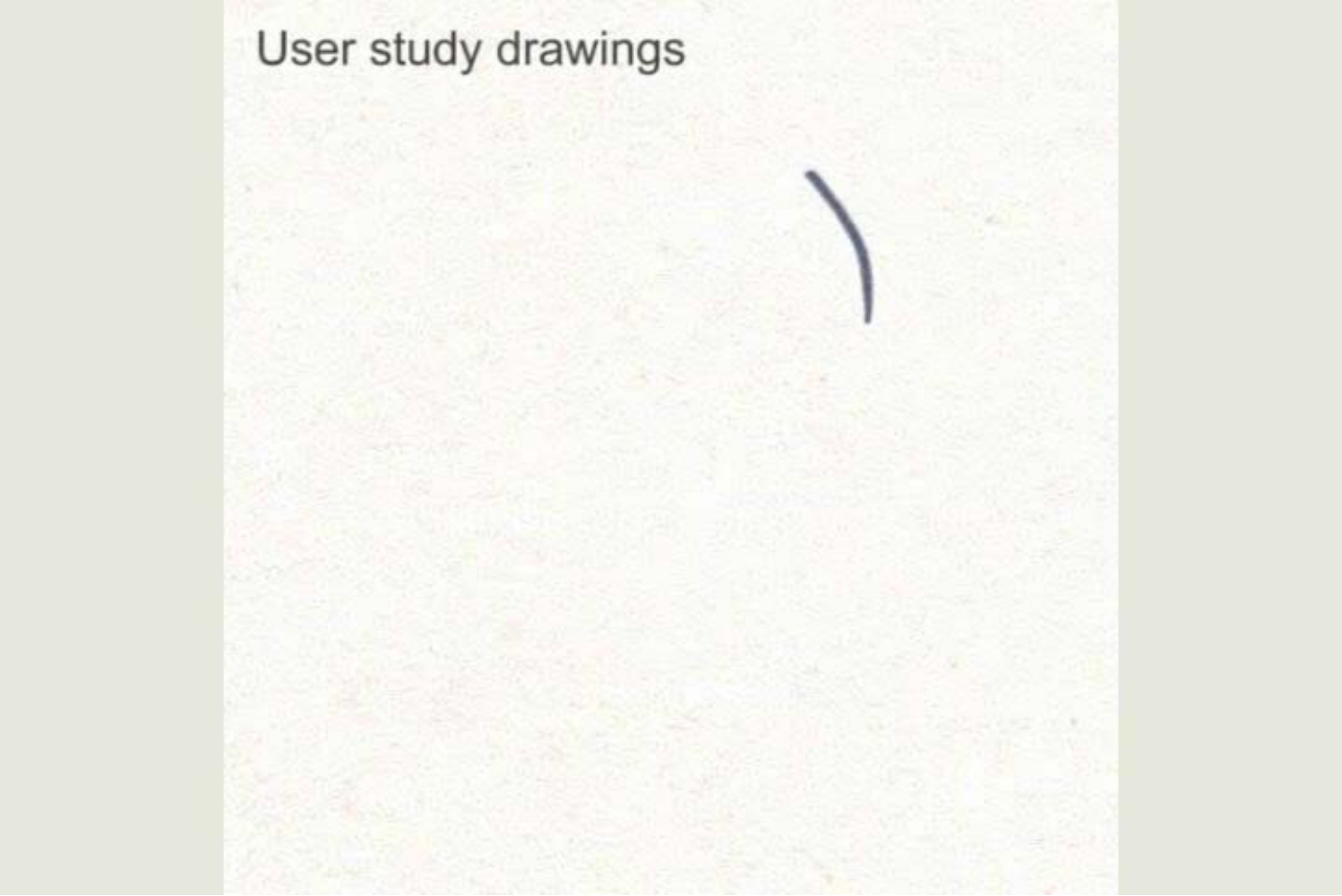
We need help..

Tracing

Drawing a face



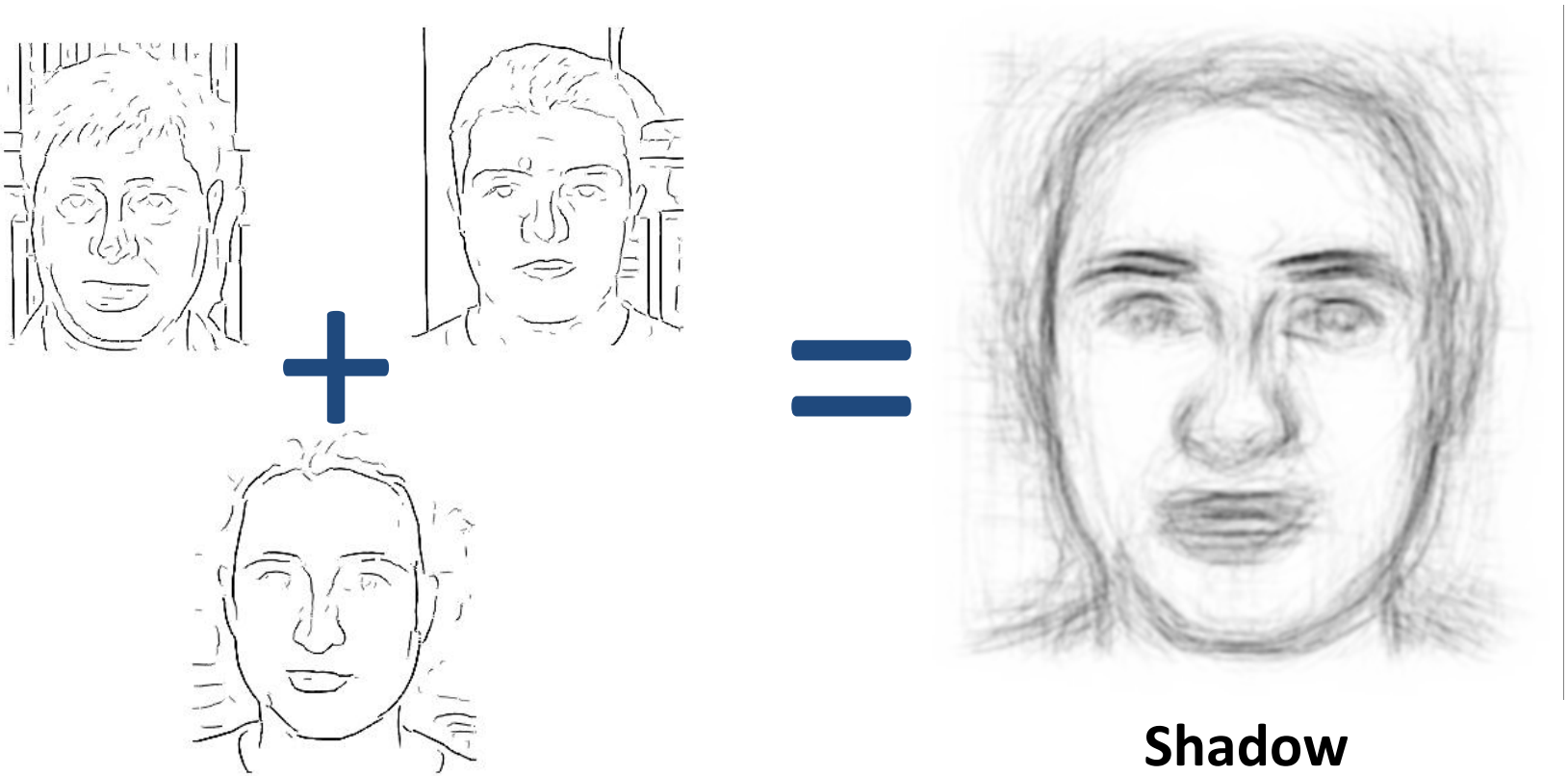
Our Idea: ShadowDraw



[Lee, Zitnick, Cohen, ShadowDraw, SIGGRAPH 2011]

Creating shadows

Collection of images…



- Shadow gives gist of many images *simultaneously*

Creating shadows

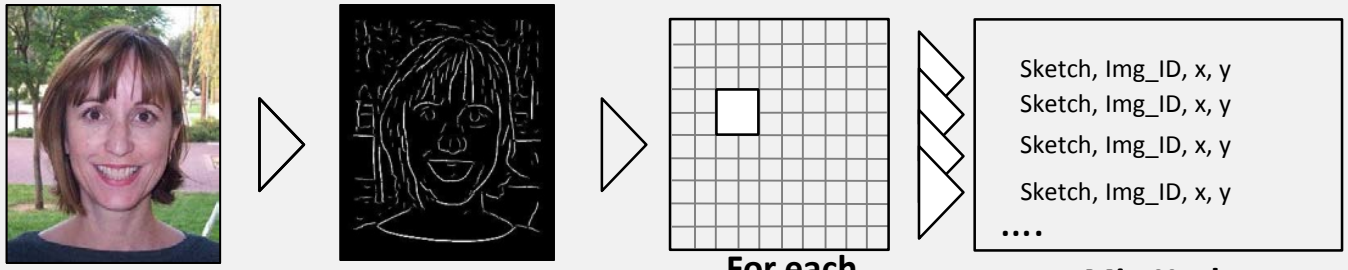
Drawing

Shadow

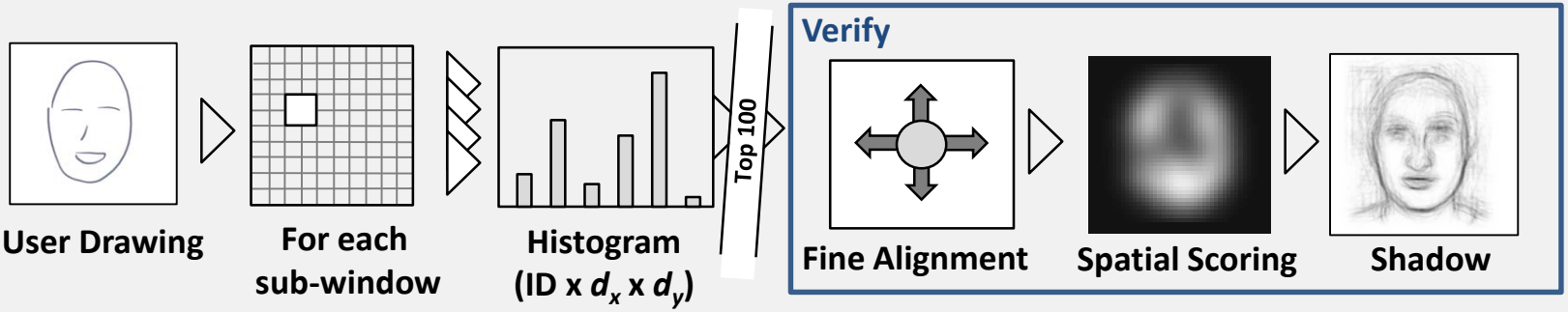
- Requires partial matching and dynamic updates

Approach overview

Database (offline)



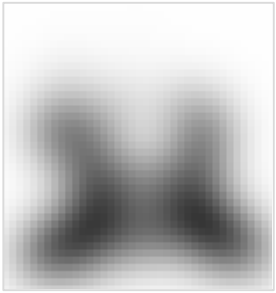
Query time (online)



Blending weights

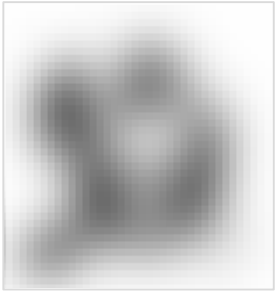
Pen Strokes

*



+

*

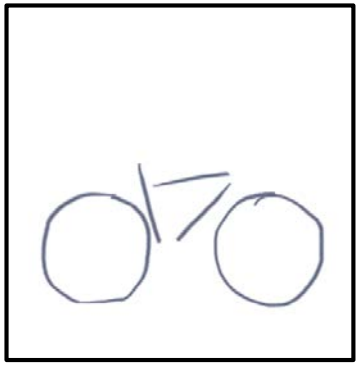


Top matches **Weights**

=

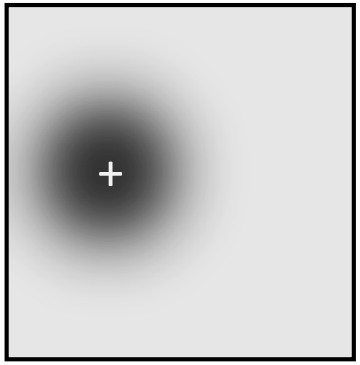
Shadows are the composite of many images

Rendering



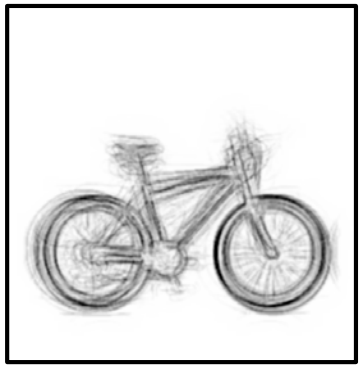
Pen Strokes

+



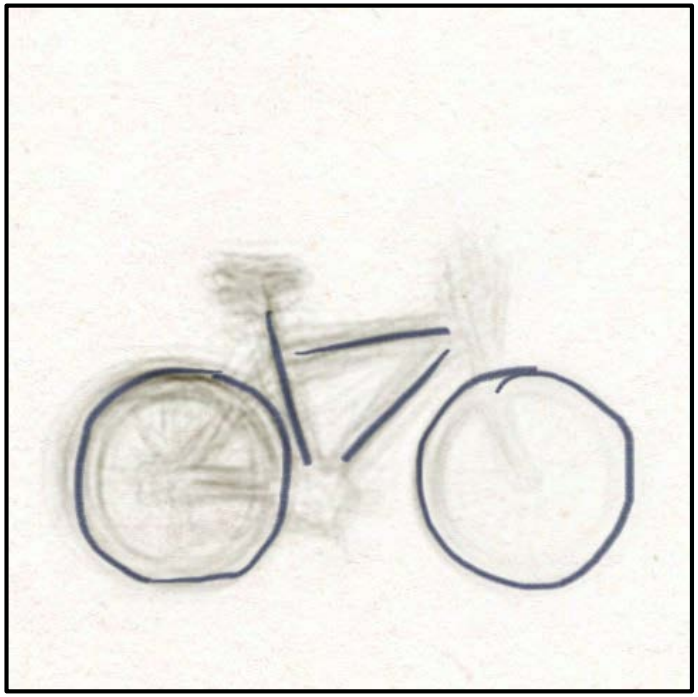
Pen Position

*



Shadow

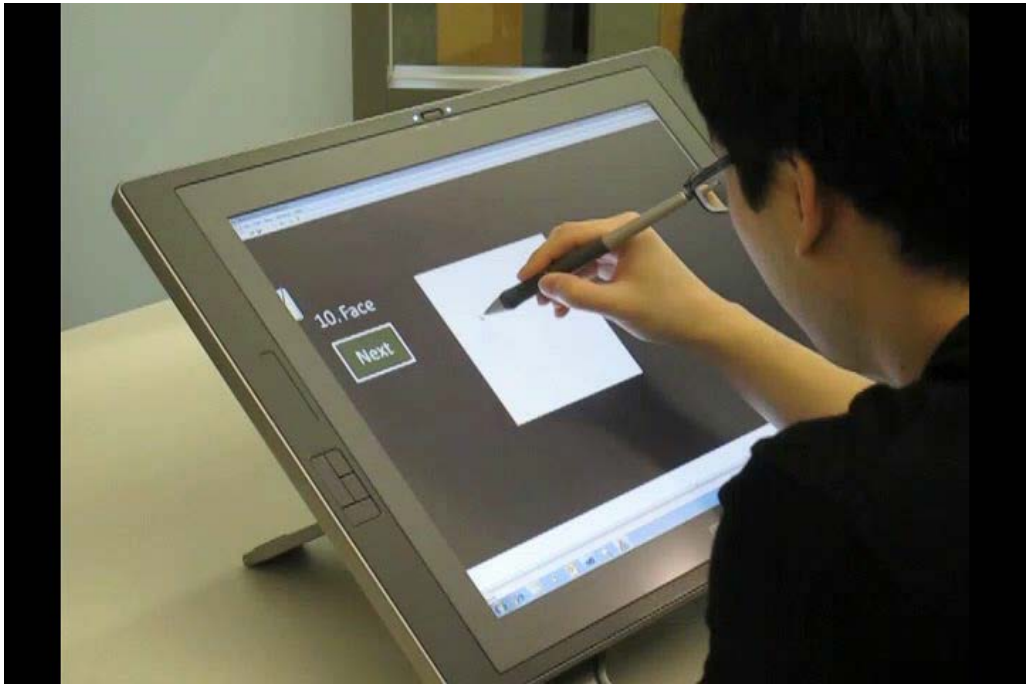
=



Output

User studies

- 30,000 images, 20 categories
- 16 drawers (8 men, 8 women), 8 evaluators
- 5 objects (shoe, face, bicycle, butterfly, rabbit)



User studies

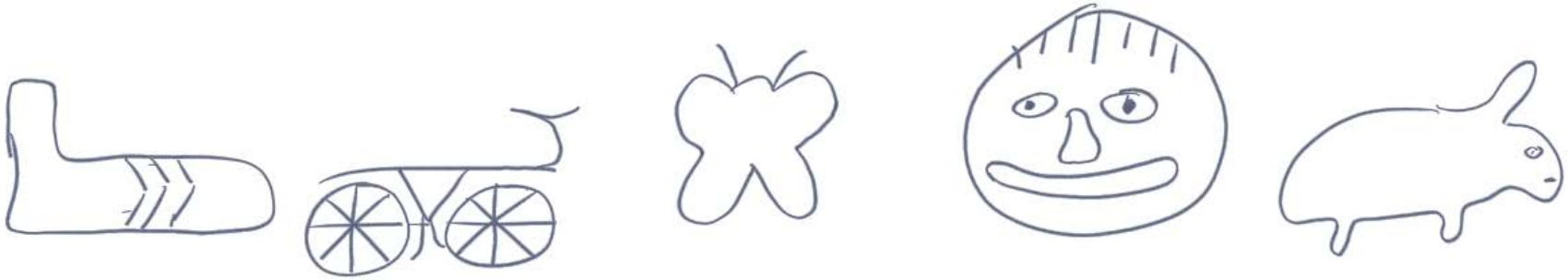
With ShadowDraw

Without ShadowDraw

Good drawers

User studies

With ShadowDraw



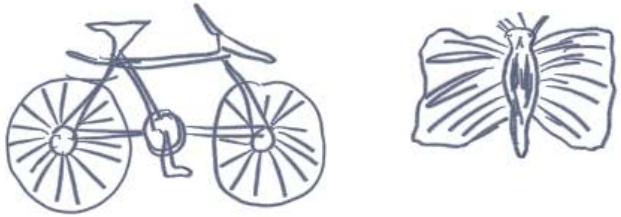
Without ShadowDraw

Bad drawers

User studies

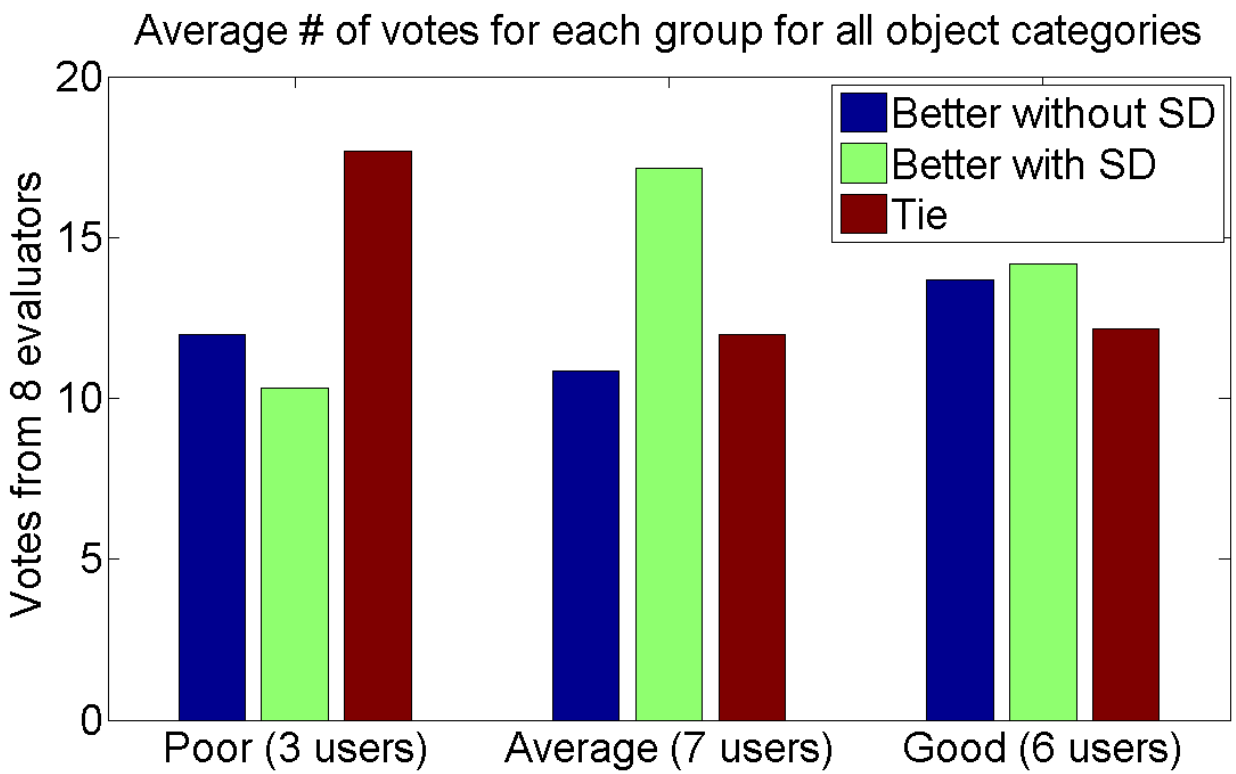
With ShadowDraw

Without ShadowDraw



Average drawers

User studies

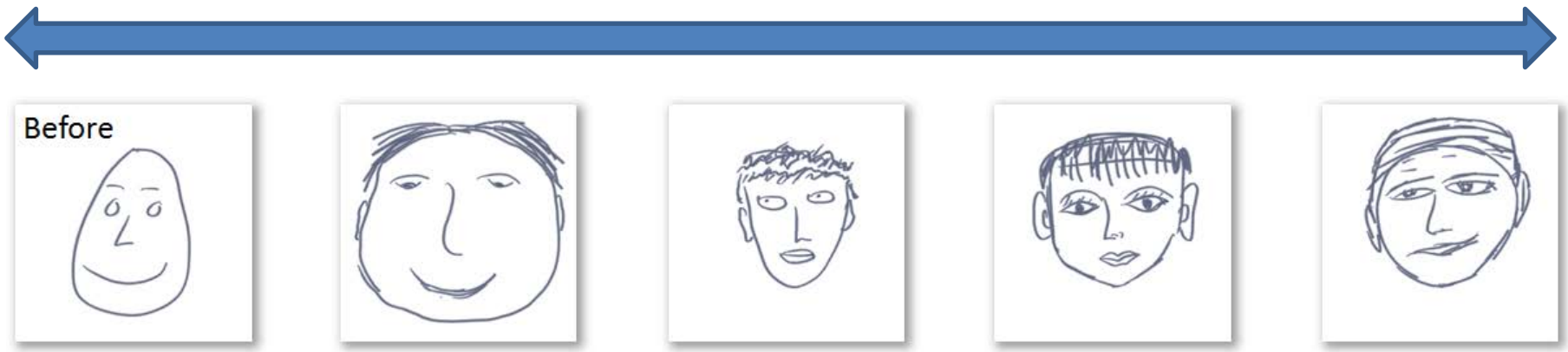


- Significant improvement for “Average” group

User studies

After training all users improved:

Poor skill *Good skill*



- Subject's *personal style* is maintained!

AverageExplorer: Interactive Exploration and Alignment of Visual Data Collections

Yong Jae Lee
University of California, Davis

SIGGRAPH 2014 work with Jun-Yan Zhu and Alexei A. Efros

Too Big for Humans

[Perona 2010]

Google wedding kiss Sign in

Web Images Videos Shopping News More Search tools SafeSearch


Romantic


First

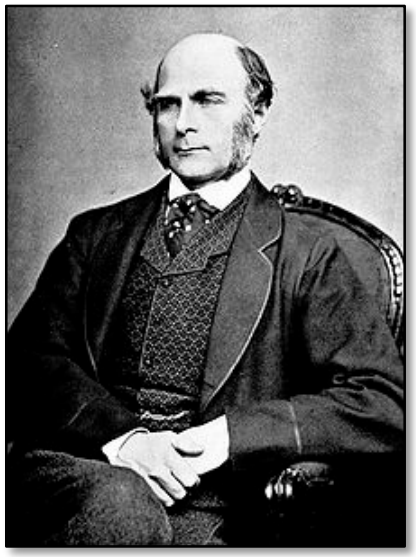

Church


Black and White


Beach

www.google.com/imgres?imgurl=http://www.decrocephotography.com/data/photos/5486_1wedding_kiss__sky.jpg&im...ml&h=840&w=1500&tbnid=WksN2Il4Pzfg3M:&zoom=1&docid=4ITYd9SxKc3GXM&hl=en&ei=E_HZU9WuENGAogT0moGYCg&tbm=isch

Image Averaging



Sir Francis Galton

1822-1911

Multiple Individuals

Composite

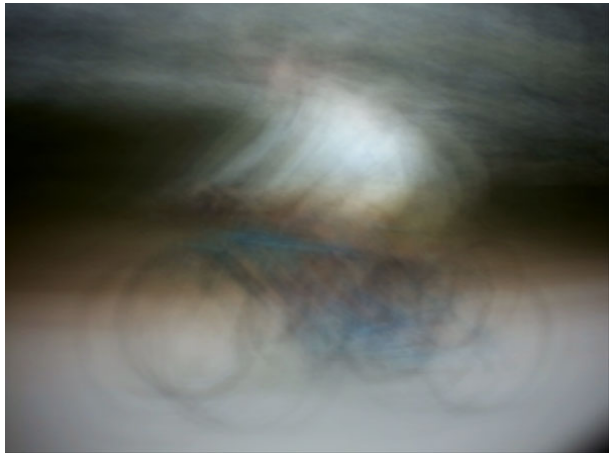
[Galton, “Composite Portraits”, Nature, 1878]

Average Images in Art

"60 passagers de 2e classe du metro, entre 9h et 11h"

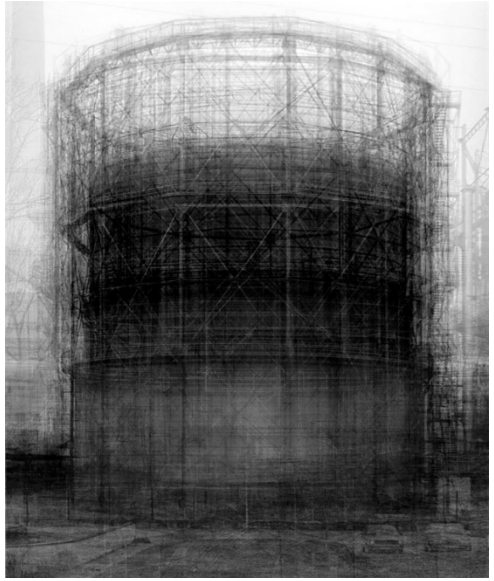
(1985)

Krzysztof Pruszkowski



"Dynamism of a cyclist"
(2001)

James Campbell



"Spherical type gasholders"
(2004)

Idris Khan

“100 Special Moments” (2004) by Jason Salavon

Newlyweds

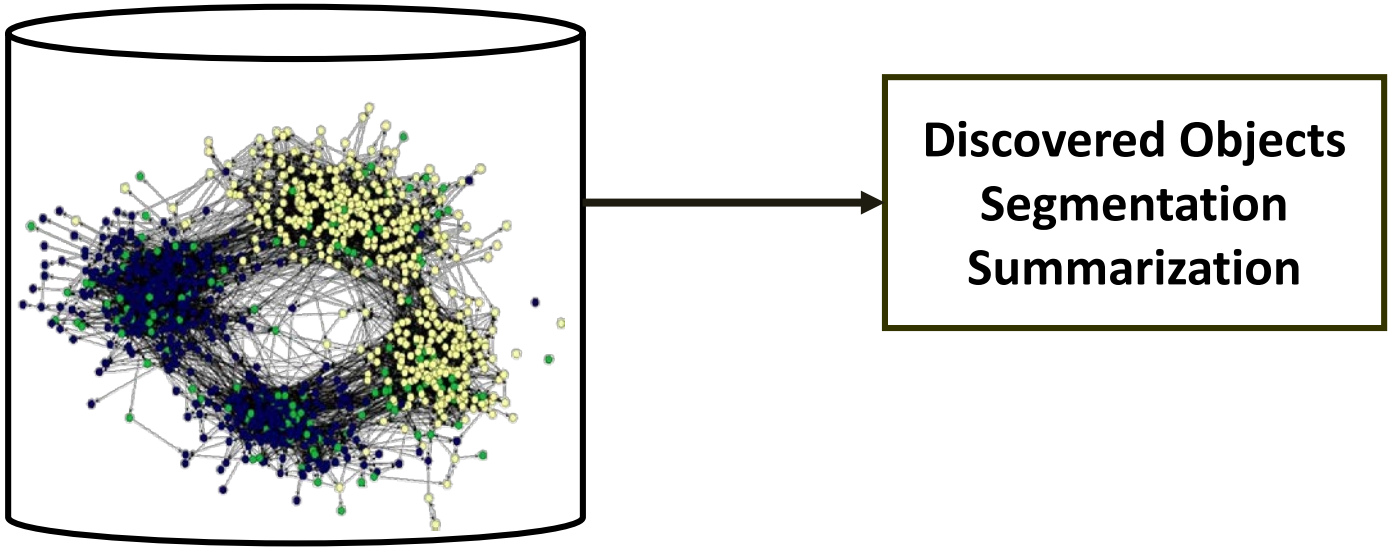
Little Leaguer

Kids with Santa

AverageExplorer

- video

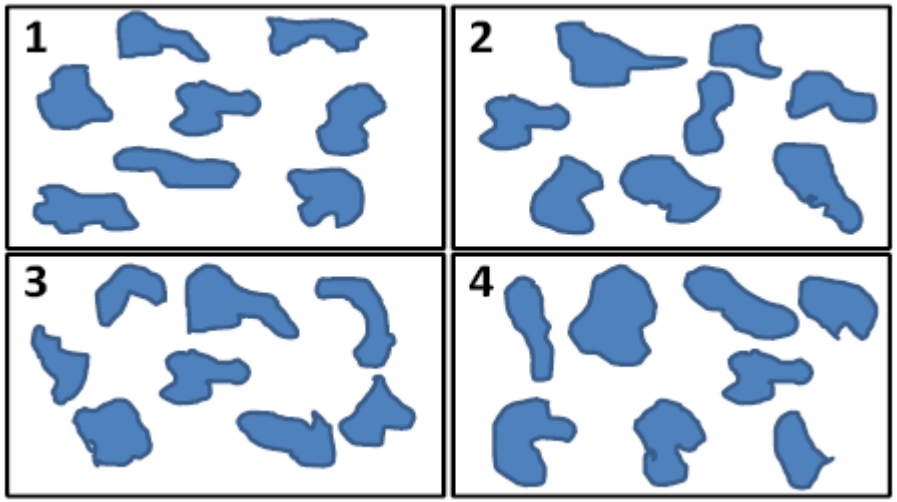
Research goal: Visual discovery



Visual world

- No human to explicitly guide visual recognition process

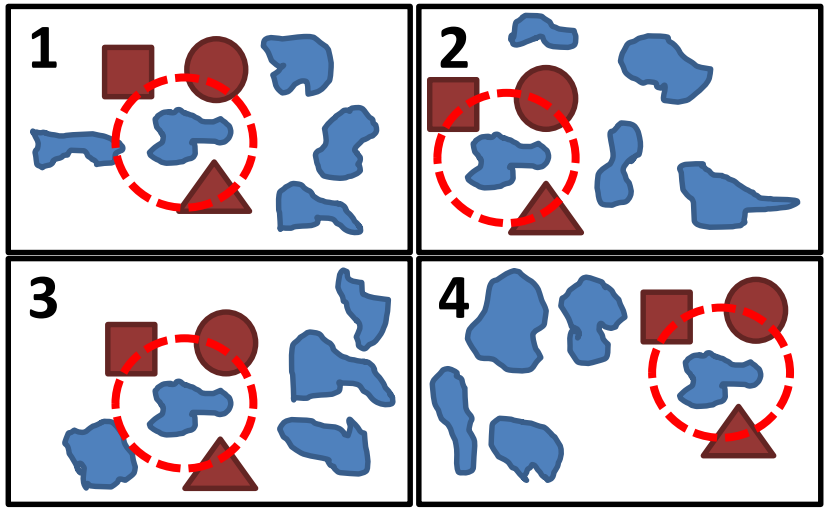
Intuition: context-aware discovery



Previous methods:

Cluster images/regions based on their appearance

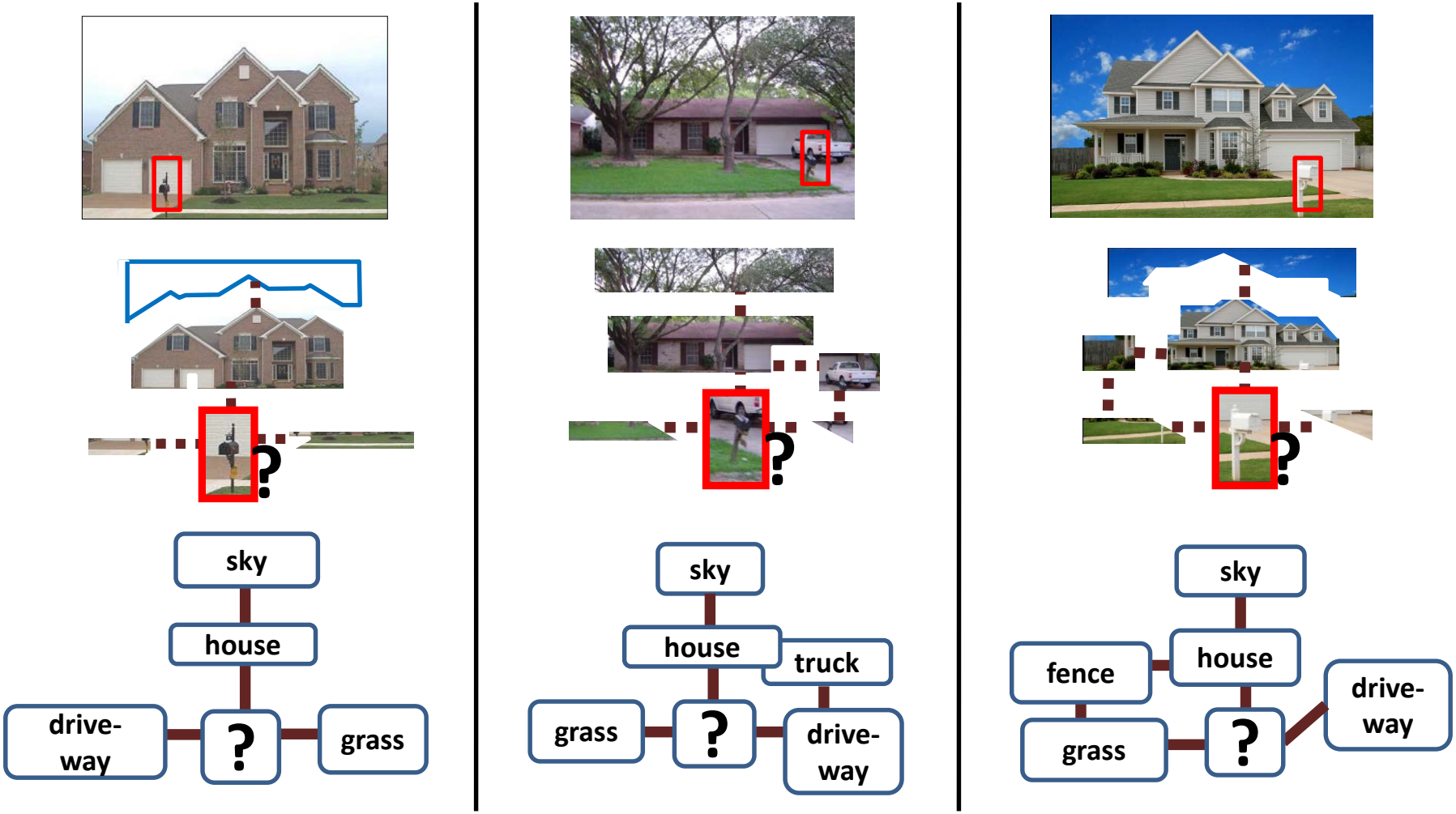
[Sivic et al. 2005, Russell et al. 2006, Liu et al. 2006, Lee et al. 2008, Kim et al. 2008,…]



Our idea:

Let *familiar* objects serve as context for *unfamiliar* objects

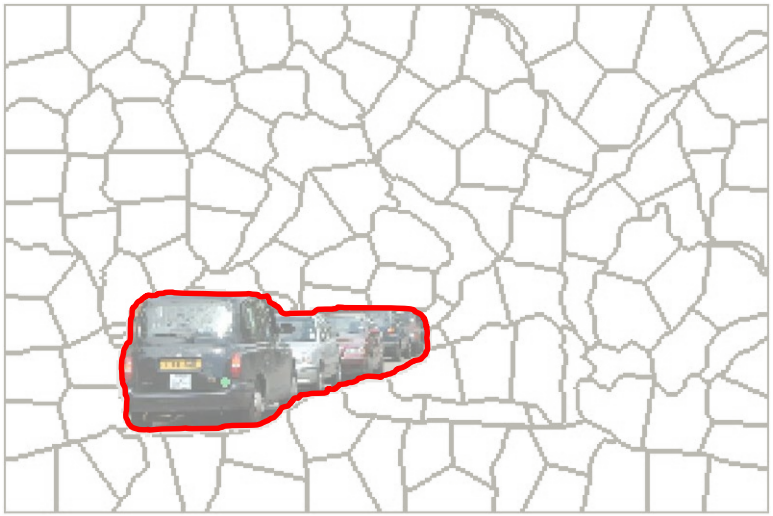
Intuition: context-aware discovery



[Lee & Grauman, Object-Graphs, CVPR 2010, TPAMI 2012]

Object-graphs

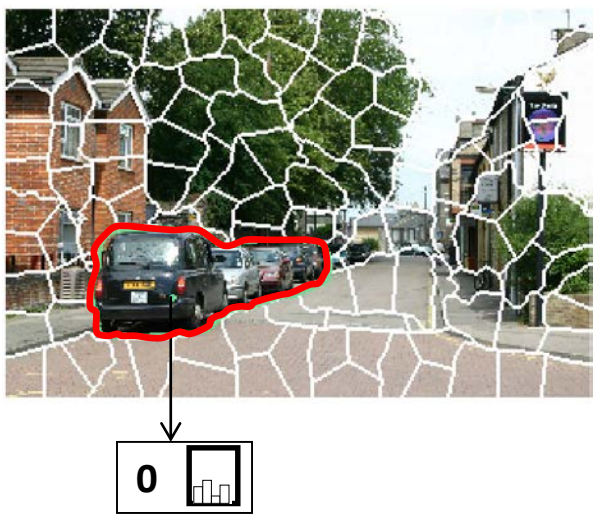
An unknown region within an image



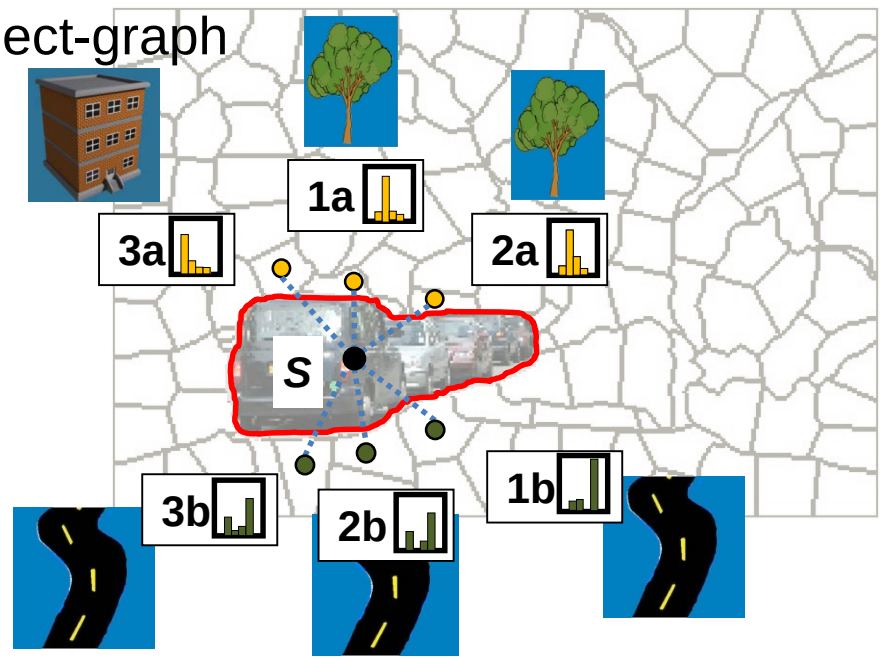
Model the topology of category predictions relative to the unknown (unfamiliar) region.

Object-graphs

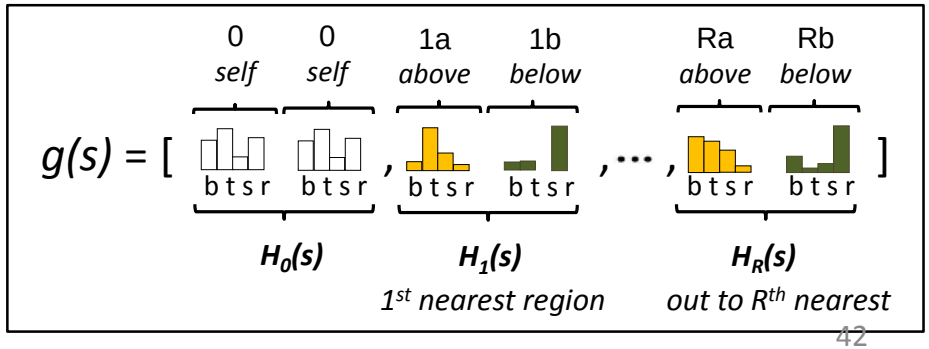
An unknown region within an image



Closest nodes in its object-graph



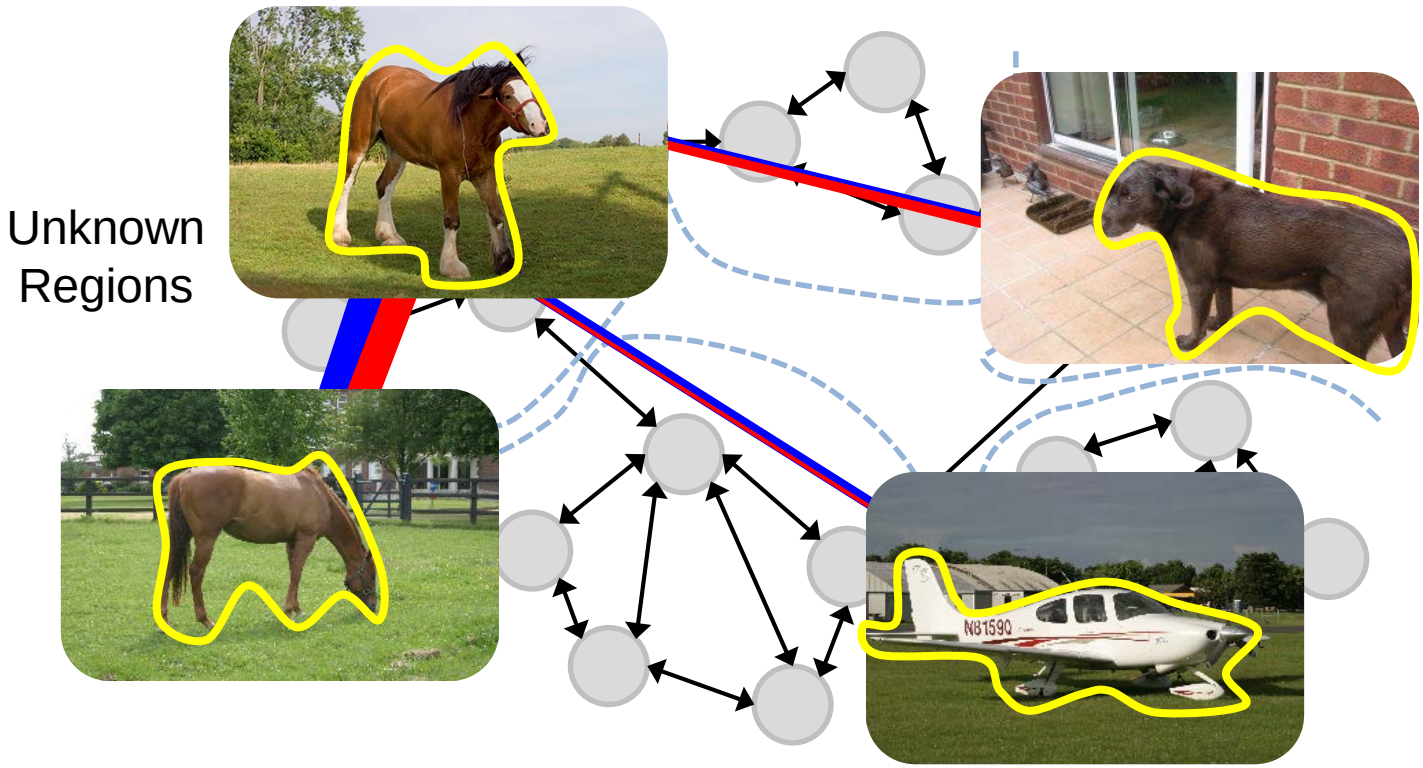
Consider spatially near regions *above* and *below*, record distributions for each known class.



Example object-graphs

- Colors indicate the predicted known category (max posterior)

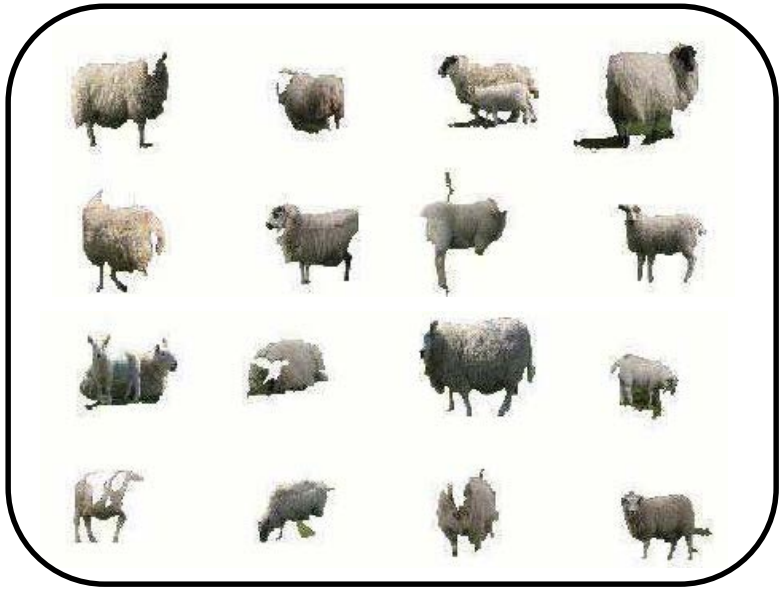
Clusters from region-region affinities

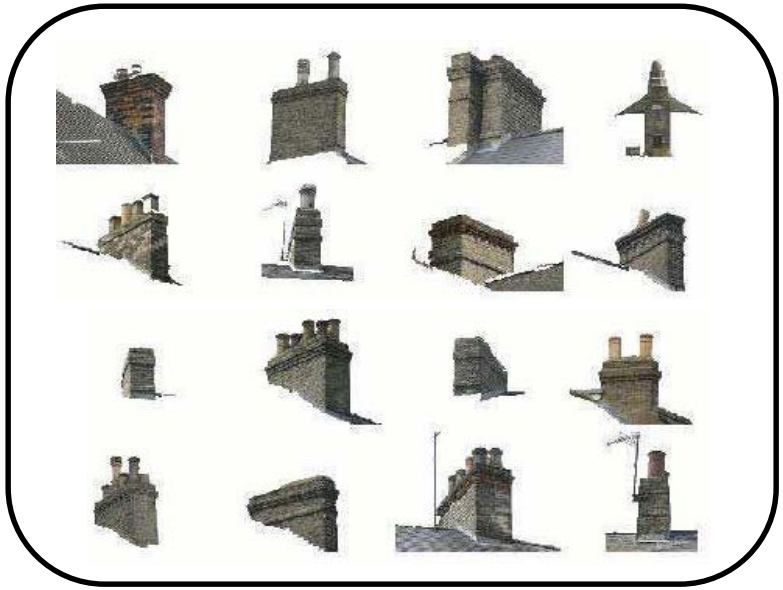


$$K(s_i, s_j) = K_{app}(s_i, s_j) + K_{obj-graph}(s_i, s_j)$$

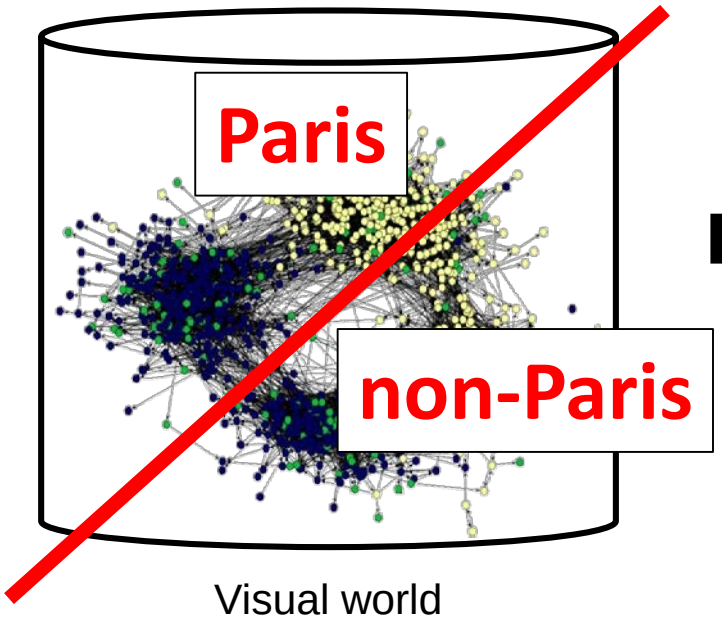
Object-level context provides more robust affinities

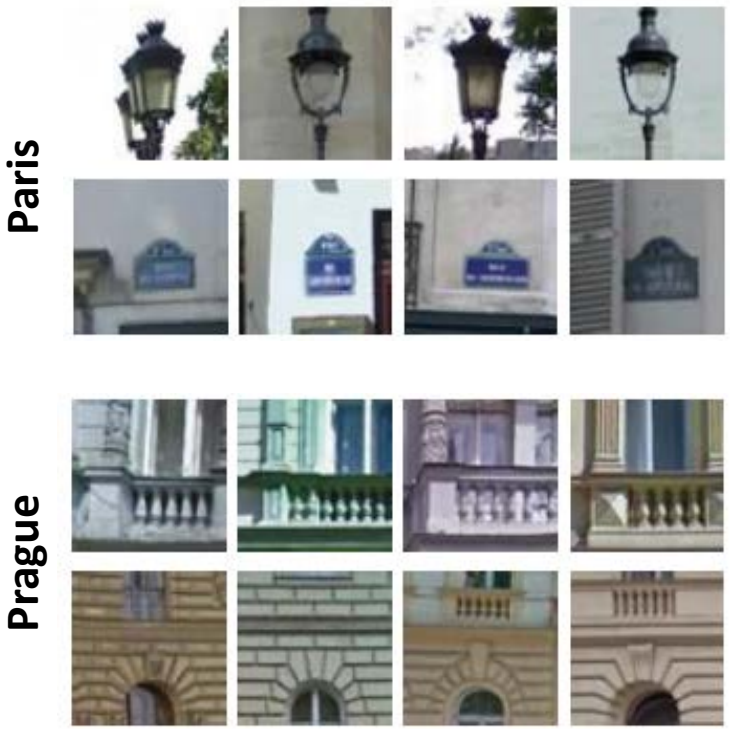
Example discoveries





Mining differences





Mid-level visual elements

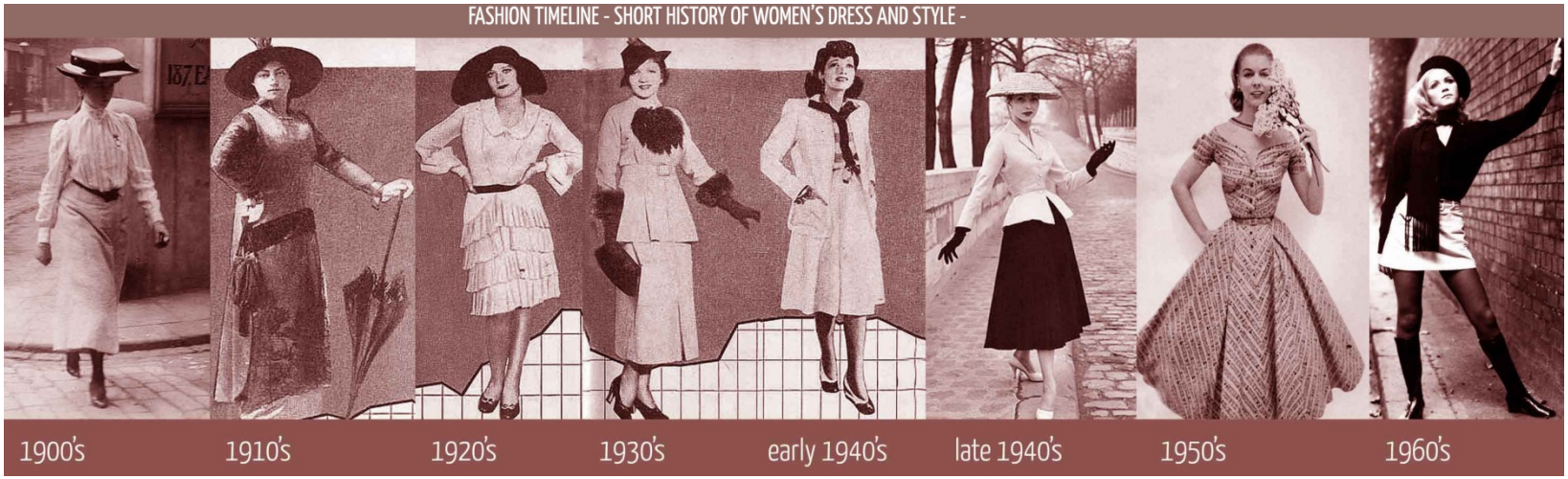
[Doersch et al. 2012, Endres et al. 2013, Juneja et al. 2013, Fouhey et al. 2013, Doersch et al. 2013]

- Recent methods mine *specific* visual patterns

Problem

- Much in our visual world undergoes a *gradual change*

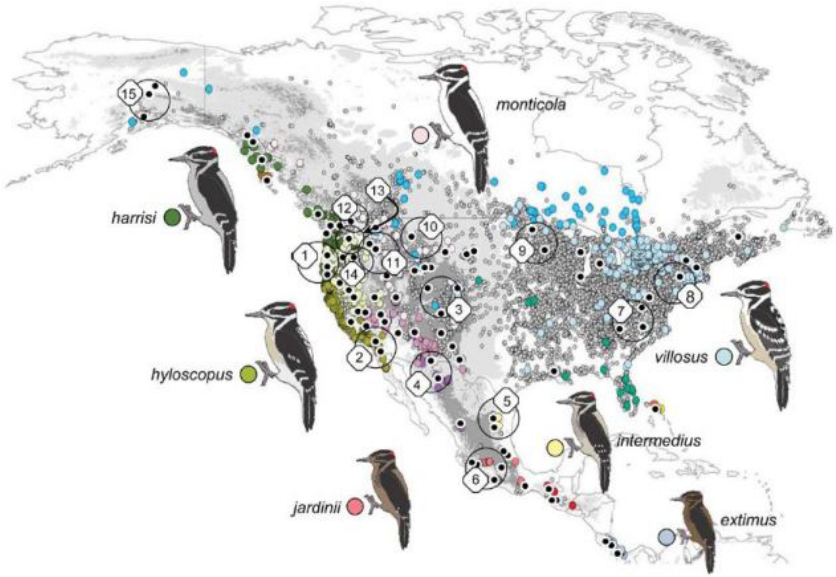
Temporal:

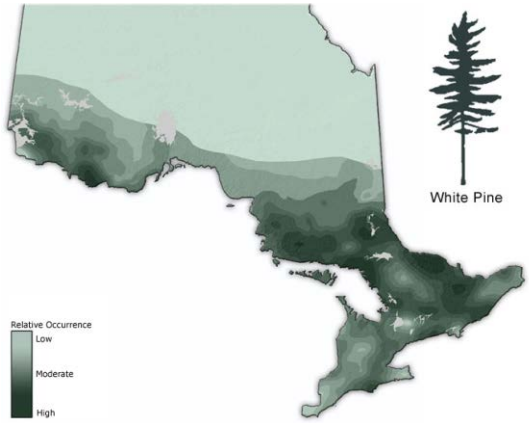


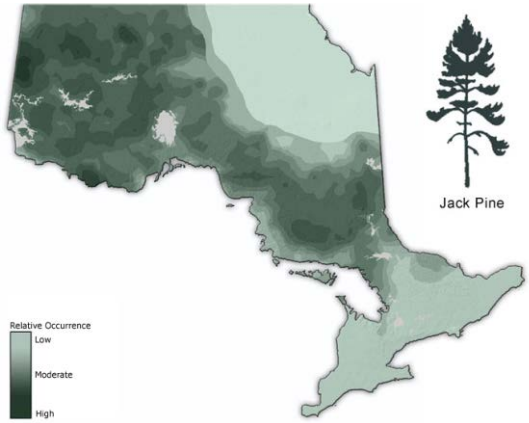
1887-1900 | 1900-1941 | 1941-1969 | 1958-1969 | 1969-1987

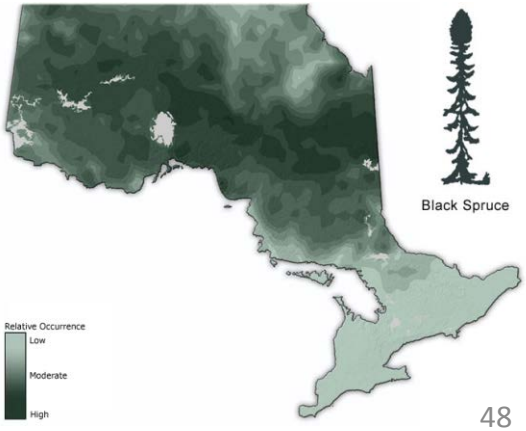
- Much in our visual world undergoes a *gradual change*

Spatial:










when?
1972
39·07·JJ

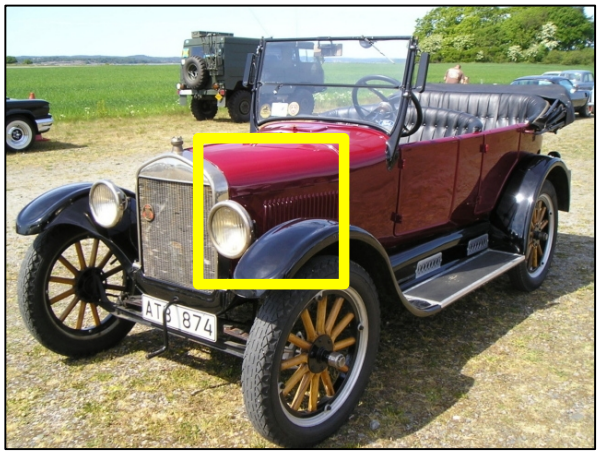
THE DISH
BIASED & BALANCED

"The View From Your Window" challenge

Key Idea

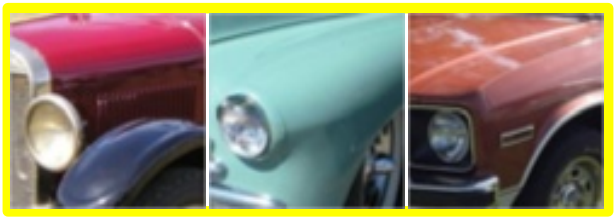
1) Establish connections

1926



1947

1975



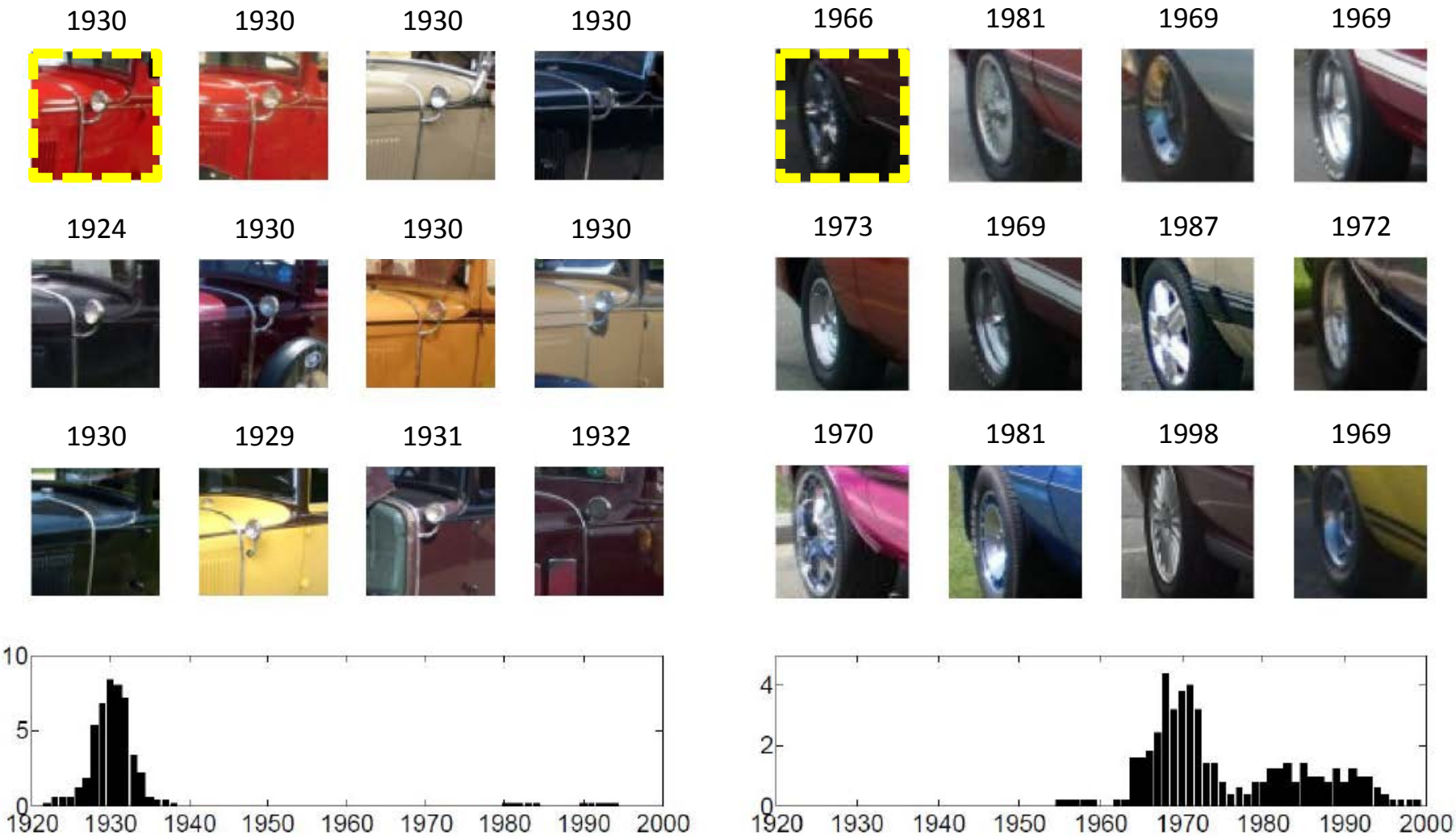
1926 1947 1975

"closed-world"

2) Model style-specific differences

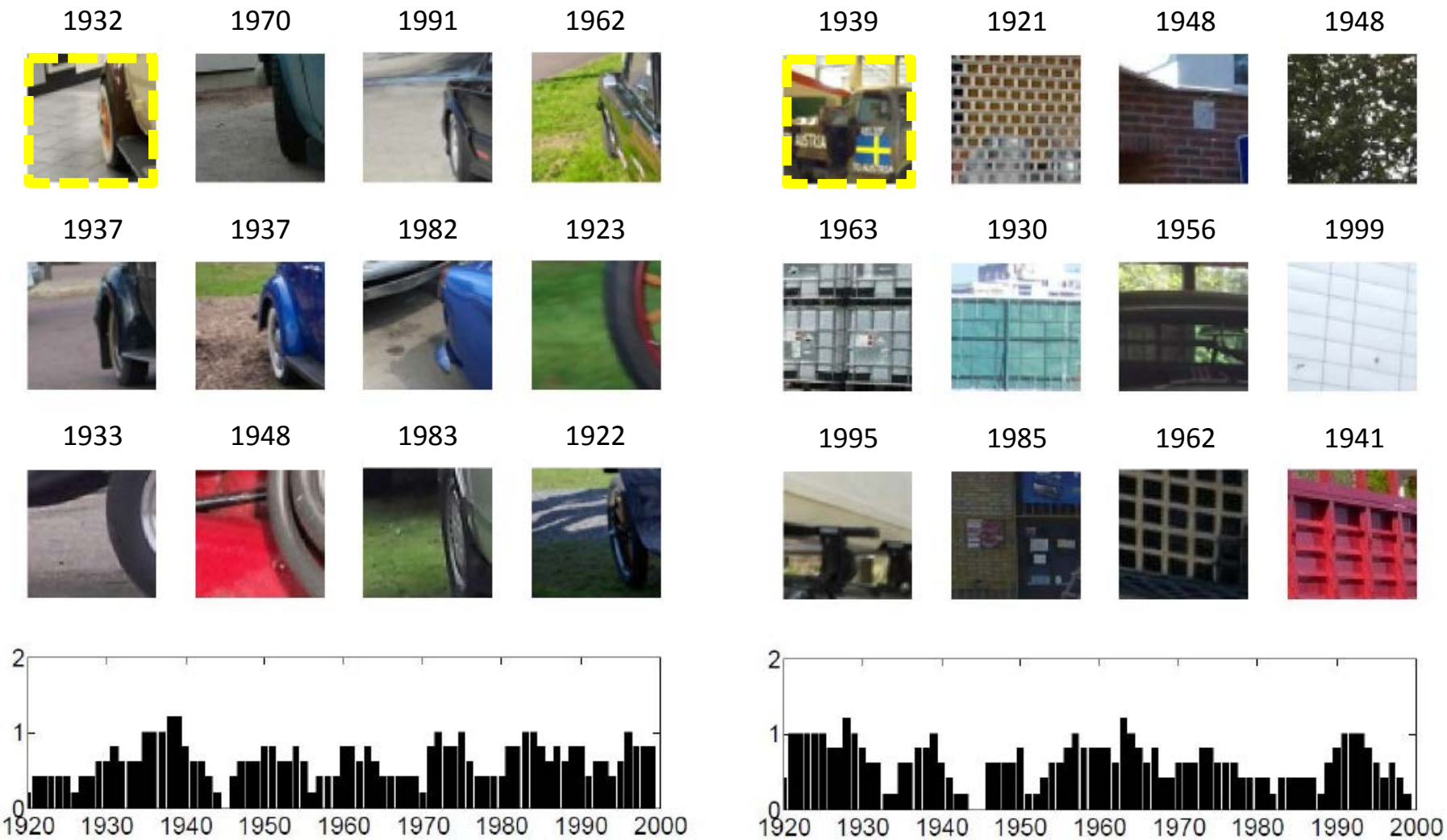
[Lee, Efros, Hebert, Style-aware mid-level representation, ICCV 2013]

Mining style-sensitive elements



(a) Peaky (low-entropy) clusters

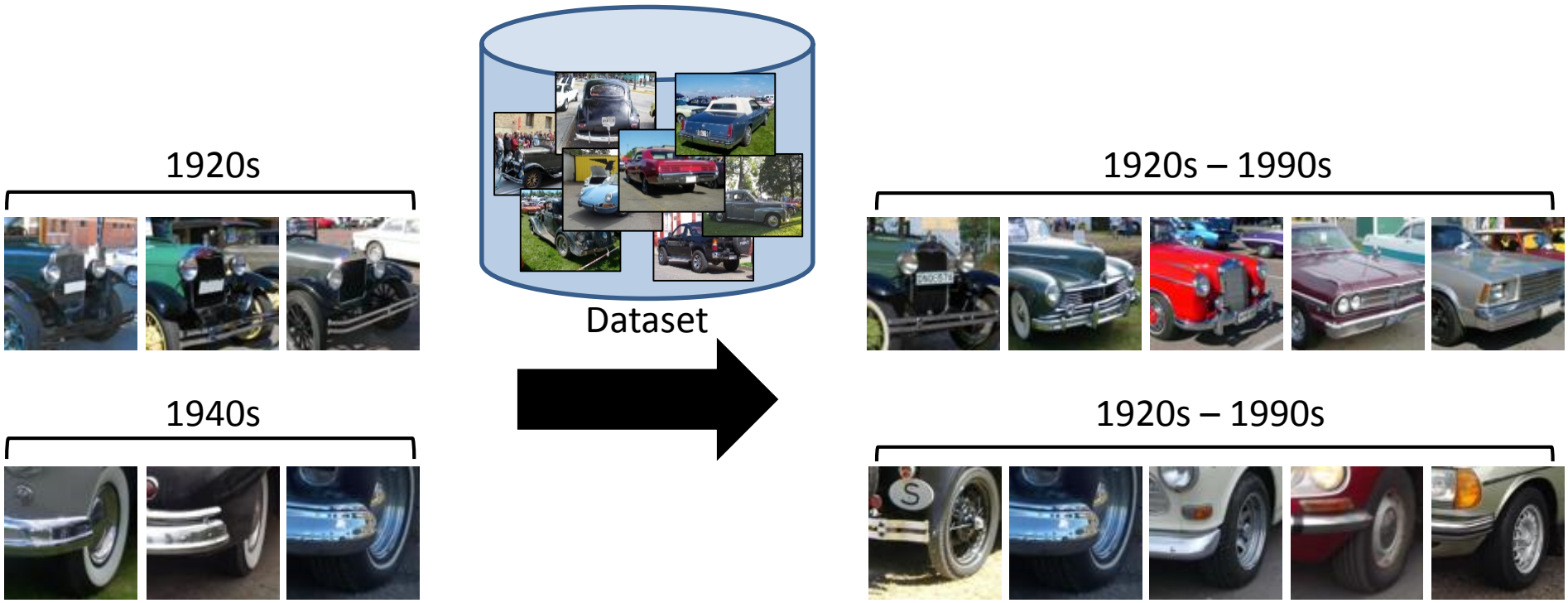
Mining style-sensitive elements



(b) Uniform (high-entropy) clusters

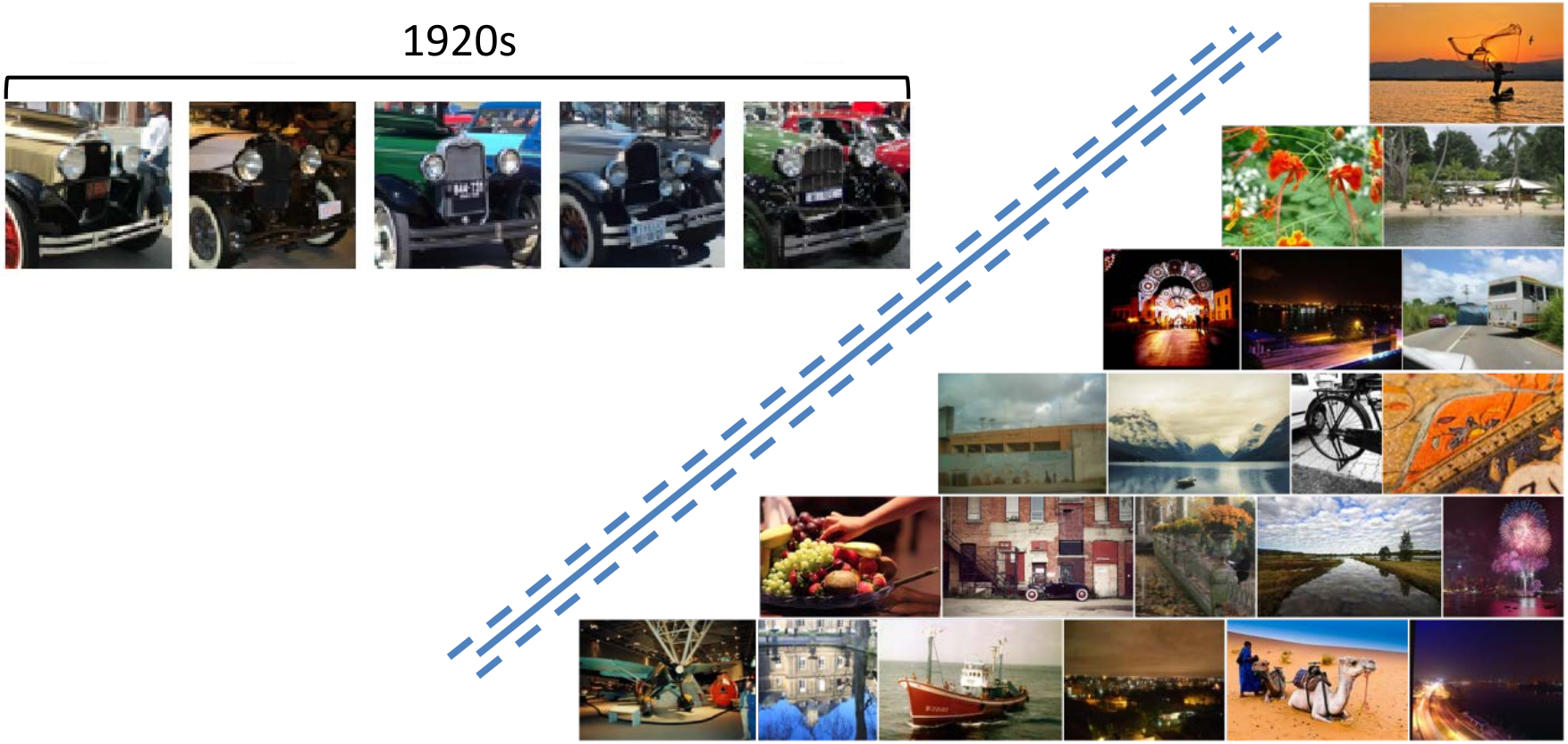
Making visual connections

- Take top-ranked clusters to build correspondences

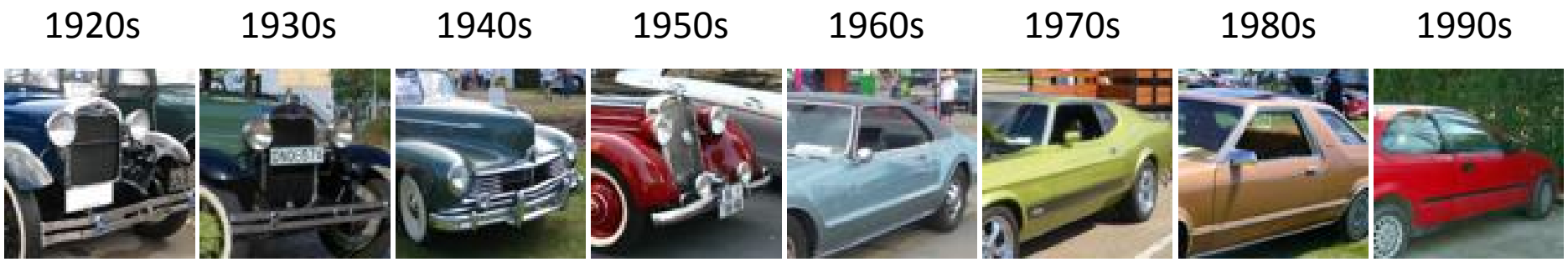


Making visual connections

- Train a detector (HoG + linear SVM) [Singh et al. 2012]



Making visual connections

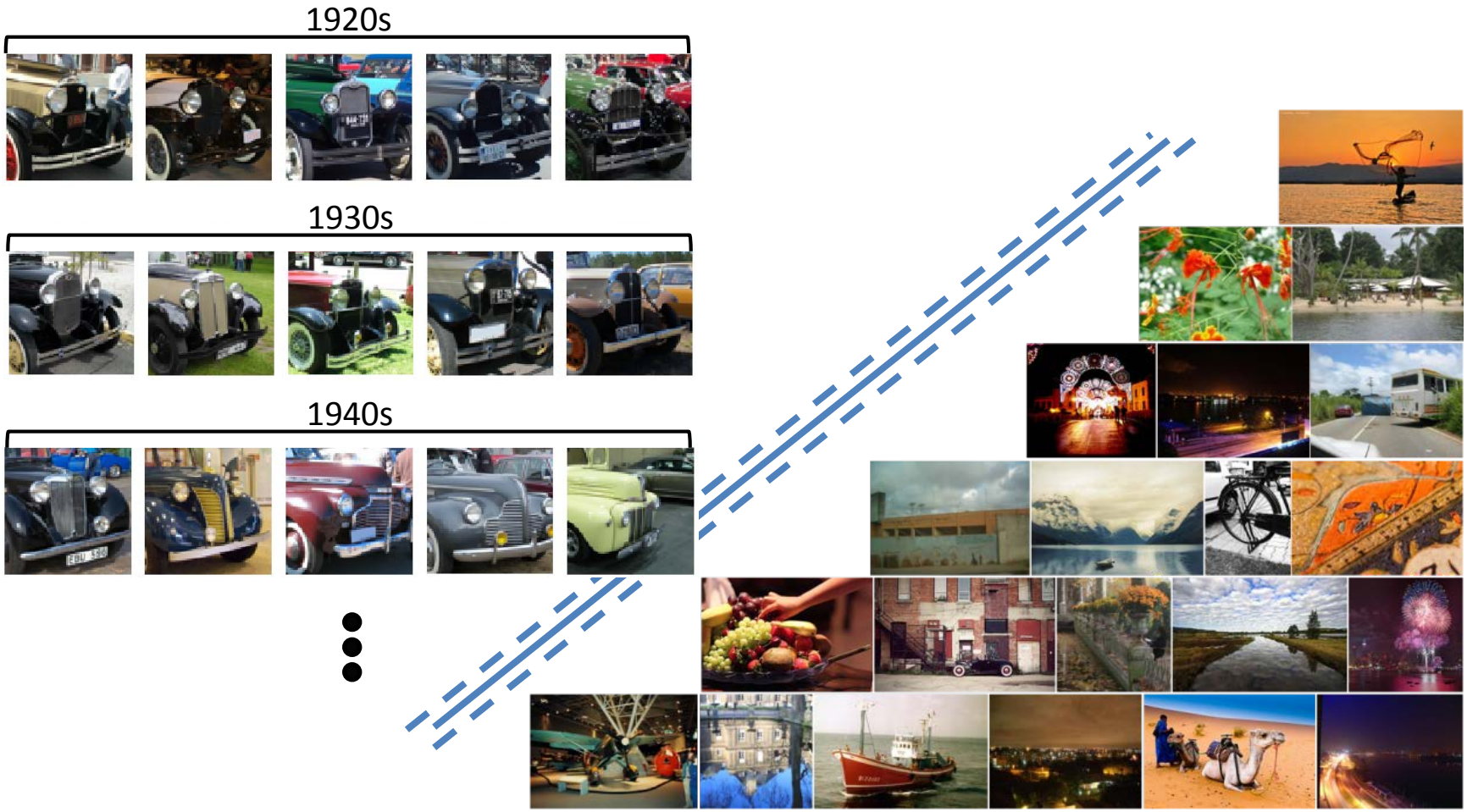


Top detection per decade

[Singh et al. 2012]

Making visual connections

- We expect style to change gradually…



Natural world “background” dataset

Making visual connections

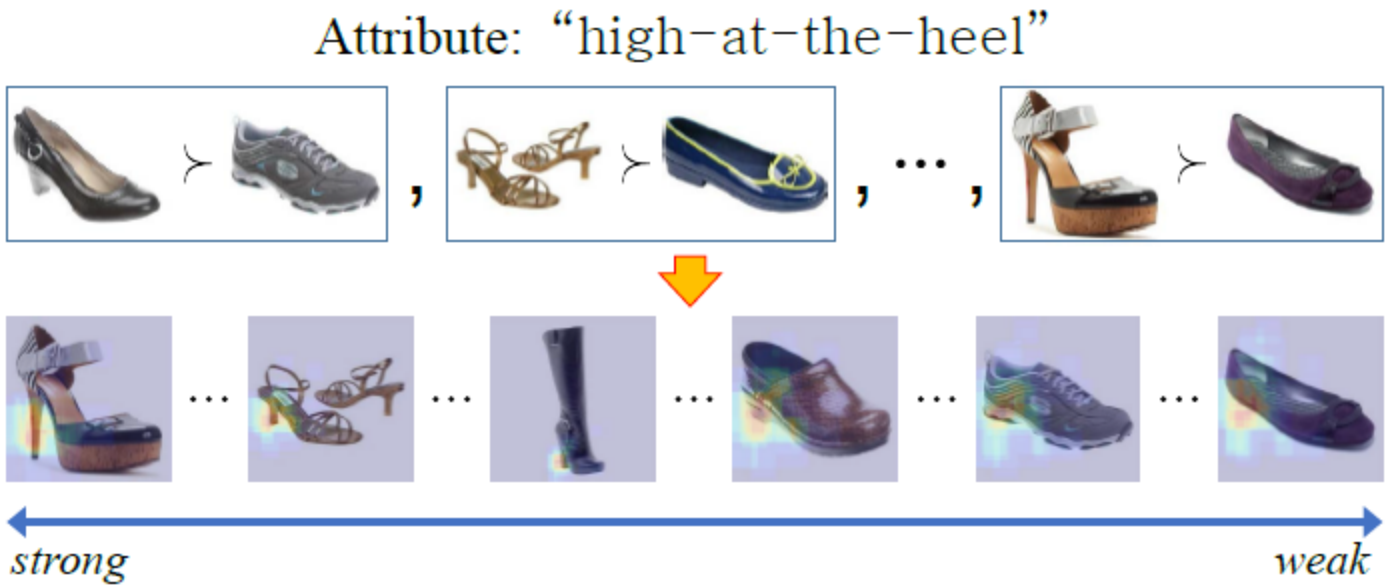
1920s 1930s 1940s 1950s 1960s 1970s 1980s 1990s

Top detection per decade

Results: Learned styles

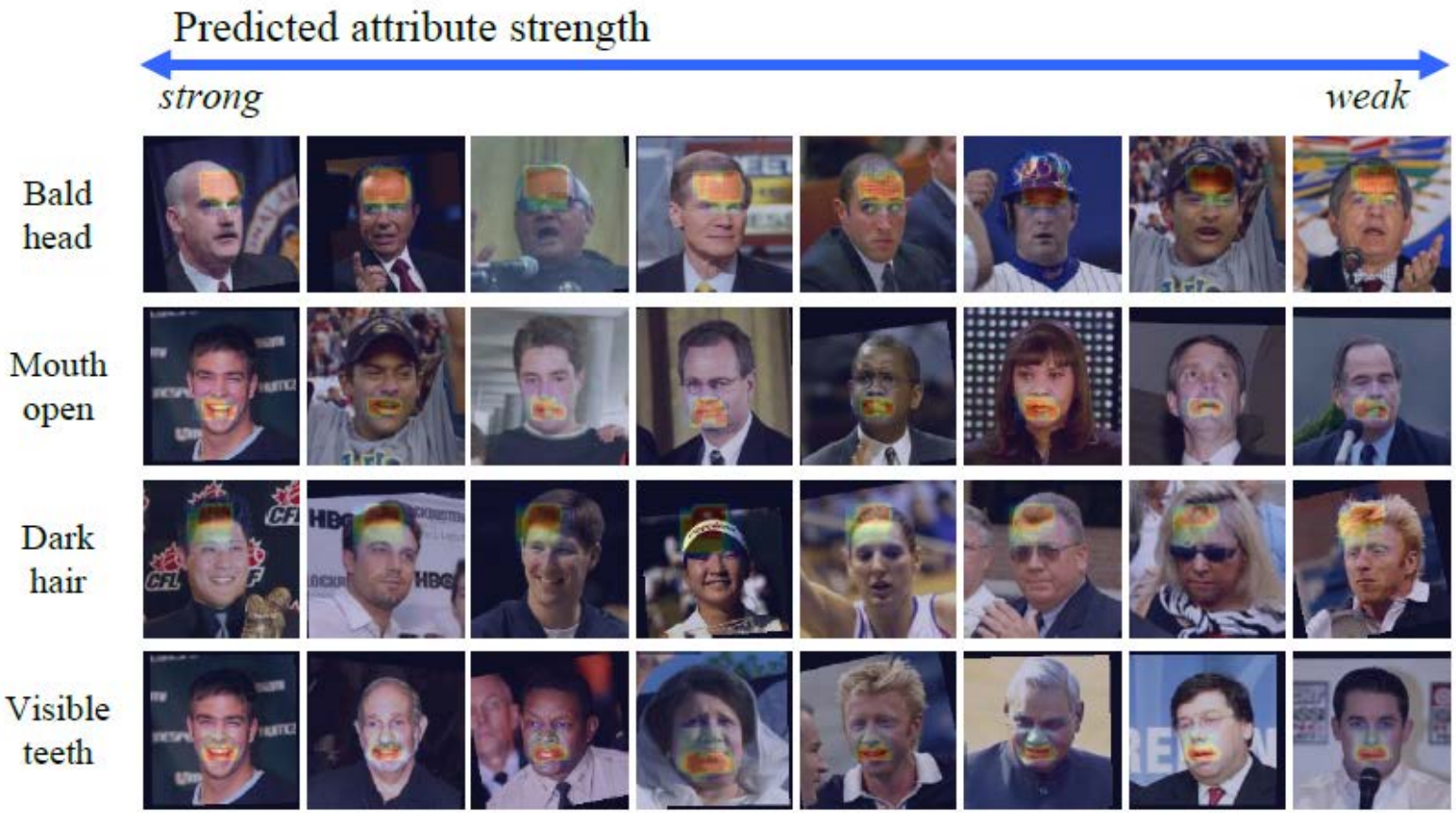
Average of top predictions per decade

Mining spatial extent of relative attributes



[Xiao & Lee, Spatial Extent of Relative Attributes, ICCV 2015]

Mining spatial extent of relative attributes

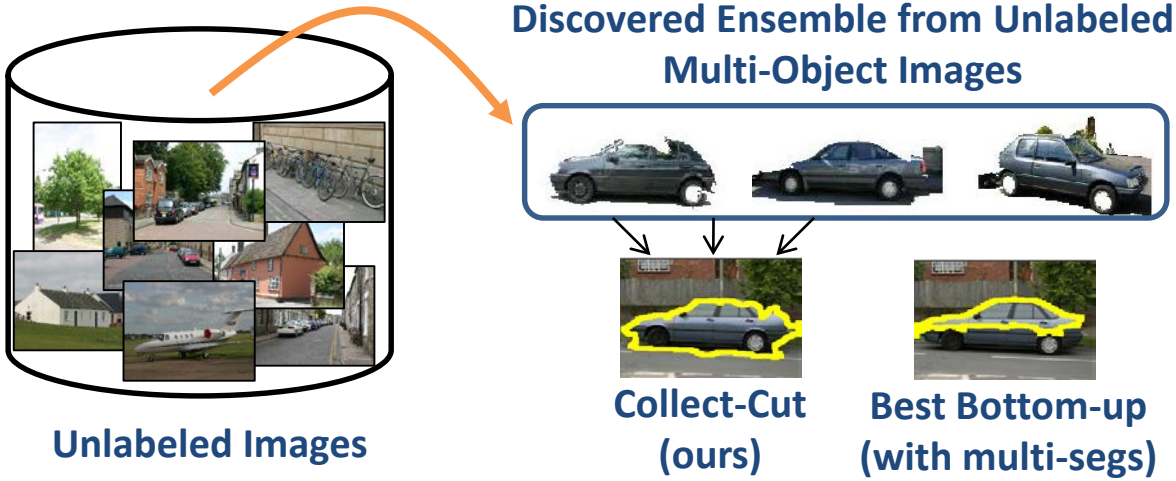


[Xiao & Lee, Spatial Extent of Relative Attributes, ICCV 2015]

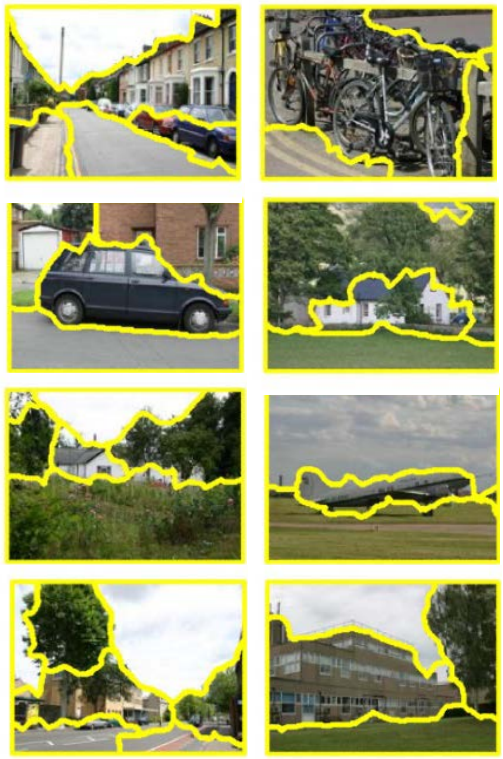
Mining spatial extent of relative attributes

[Xiao & Lee, Spatial Extent of Relative Attributes, ICCV 2015]

Collect-Cut



Unsupervised Segmentation Examples



[Lee & Grauman, Collect-Cut, CVPR 2010]

Problem: Video object segmentation

How to segment the foreground objects in video when

- background is moving and changing
- categories of foreground objects are unknown in advance

Input: Unannotated video

Desired output: Segmentation of high-ranking foreground object

- Existing methods group pixels using low-level features, which can result in an "over-segmentation." [Brendel & Todorovic 2009, Vazquez-Reina et al. 2010, Grundmann et al. 2010, Brox & Malik 2010]

Mining first-person camera data

GoPro

Google Glass

Looxcie

Tobii

SMI

Pivothead

Mining first-person camera data

90’s

Steve Mann life logger

Problem: Summarizing egocentric videos

Wearable camera

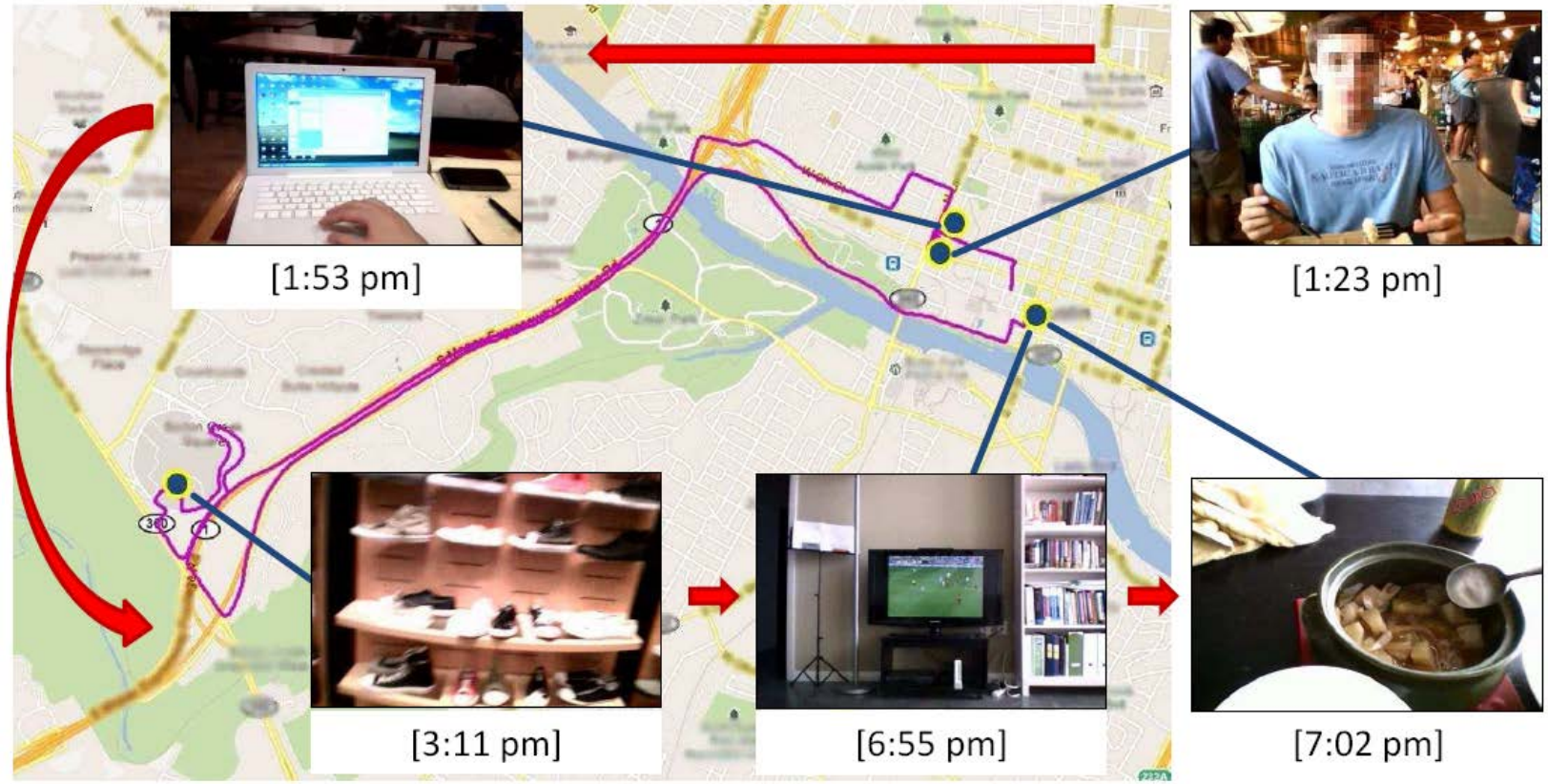
Input: Egocentric video of the camera wearer's day

9:00 am 10:00 am 11:00 am 12:00 pm 1:00 pm 2:00 pm

Output: Storyboard summary of discovered *important people and objects*

[Lee, Ghosh, Grauman, Egocentric video summarization, CVPR 2012]

Results: Egocentric video summarization



Coming up

- Sign-up for paper
- Next class
  - ImageNet classification with deep convolutional neural networks. A. Krizhevsky, I. Sutskever, and G. E. Hinton. NIPS 2012.
    Very Deep Convolutional Networks for Large-Scale Image Recognition. K. Simonyan and A. Zisserman. ICLR 2015.

- Read both papers
- Write a review for one of them