Attributes

Computer Vision

James Hays

Many slides from
Derek Hoiem

Recap: Human Computation

- Active Learning: Let the classifier tell you where more annotation is needed.
- Human-in-the-loop recognition: Have a human and computer cooperate to do recognition.
- Mechanical Turk is powerful but noisy
  - Determine which workers are trustworthy
  - Find consensus over multiple annotators
  - “Gamify” your task to the degree possible

Recap: Recognition Data Sets

- SUN Scene Database
  - *Not* Crowdsourced, 397 (or 720) scene categories
- PASCAL VOC
  - *Not* Crowdsourced, bounding boxes, 20 categories.
- LabelMe (Overlaps with SUN)
  - Sort of Crowdsourced, Segmentations, Open ended
- SUN *Attribute* database (Overlaps with SUN)
  - Crowdsourced, 102 attributes for every scene
- ImageNet
  - Large, Crowdsourced, Hierarchical, *Iconic* objects

– COCO
  - Large, Crowdsourced, 80 segmented object categories in complex scenes

Today – Crowd enabled recognition

- Recognizing Object Attributes
- Recognizing Scene Attributes

Describing Objects by their Attributes

Ali Farhadi, Ian Endres,
Derek Hoiem, David Forsyth

CVPR 2009


U.S. AIR FORCE

What do we want to know about this object?

What do we want to know about this object?

Object recognition expert:
“Dog”

What do we want to know about this object?

Object recognition expert:
“Dog”

Person in the Scene:
“Big pointy teeth”, “Can move fast”, “Looks angry”

Our Goal: Infer Object Properties

Can I **poke with it**?

Can I **put stuff in it**?

Is it **alive**?

What **shape** is it?

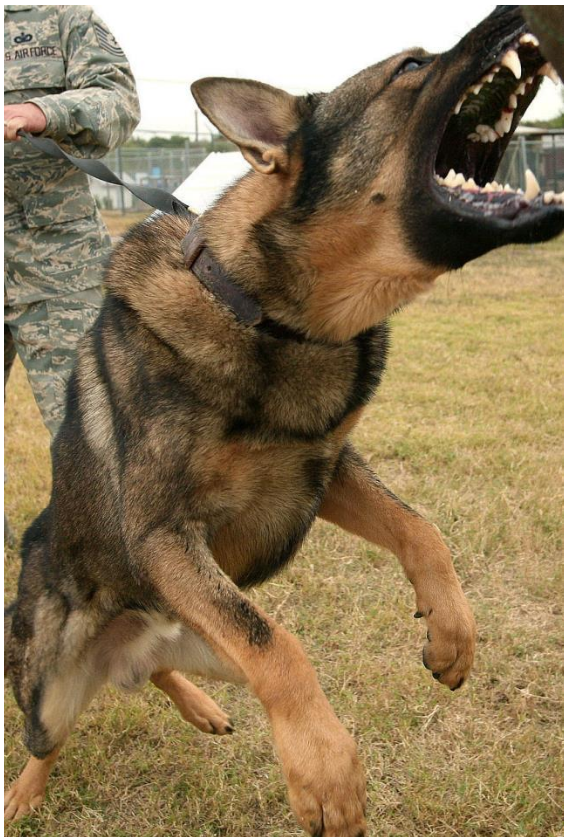
Is it **soft**?

Does it have a **tail**?

Will it **blend**?

Why Infer Properties

1. We want detailed information about objects



"Dog"
vs.
"Large, angry animal with pointy teeth"

Why Infer Properties

2. We want to be able to infer something about unfamiliar objects

Familiar Objects

New Object

Why Infer Properties

2. We want to be able to infer something about unfamiliar objects

If we can infer category names…

Familiar Objects

New Object

Cat

Horse

Dog

???

Why Infer Properties

2. We want to be able to infer something about unfamiliar objects

If we can infer properties…

Familiar Objects

Has Stripes
Has Ears
Has Eyes
….

Has Four Legs
Has Mane
Has Tail
Has Snout
….

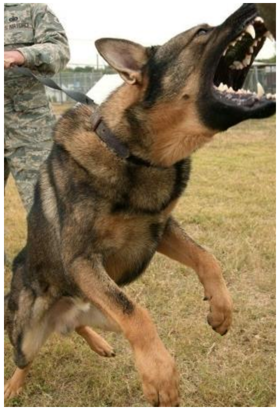
Brown
Muscular
Has Snout
….

New Object

Has Stripes (like cat)
Has Mane and Tail (like horse)
Has Snout (like horse and dog)

Why Infer Properties

3. We want to make comparisons between objects or categories



What is unusual about this dog?

What is the difference between horses and zebras?

Strategy 1: Category Recognition

Object Image

classifier

Category

"Car"

associated properties

Has Wheels
Used for Transport
Made of Metal
Has Windows
…

Category Recognition: PASCAL 2008
Category → Attributes: ??

Strategy 2: Exemplar Matching

Object Image

similarity function

Similar Image

associated properties

Has Wheels
Used for Transport
Made of Metal
Old
…

Malisiewicz Efros 2008
Hays Efros 2008
Efros et al. 2003

Strategy 3: Infer Properties Directly

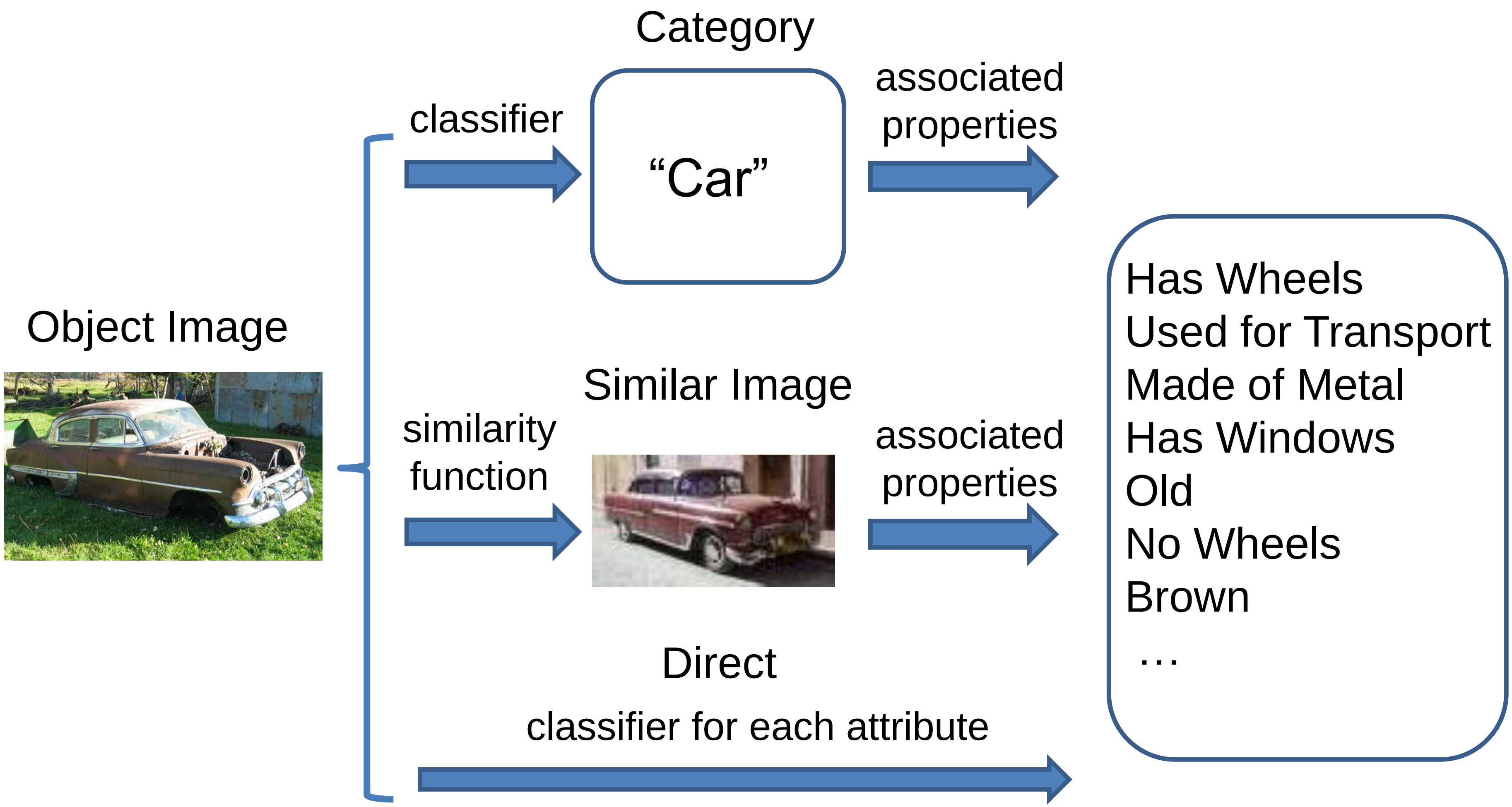
Object Image

classifier for each attribute

No Wheels
Old
Brown
Made of Metal
…

See also Lampert et al. 2009
Gibson's affordances

The Three Strategies



Our attributes

- Visible parts: “has wheels”, “has snout”, “has eyes”
- Visible materials or material properties: “made of metal”, “shiny”, “clear”, “made of plastic”
- Shape: “3D boxy”, “round”

Attribute Examples

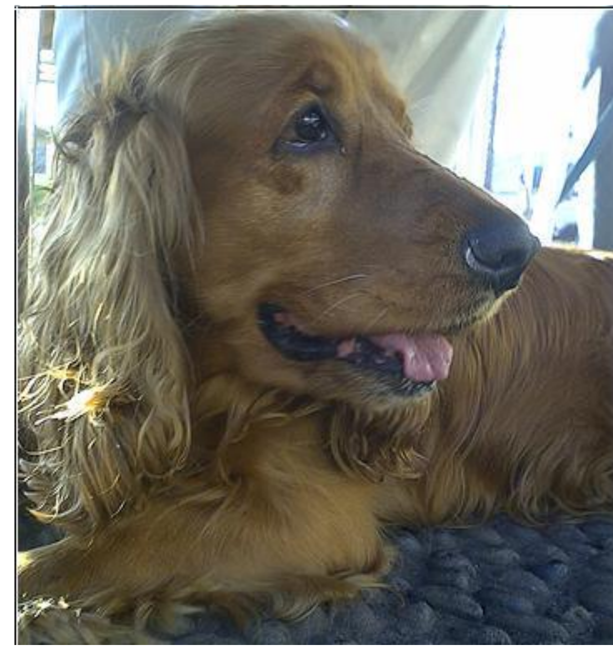
Shape: Horizontal Cylinder
Part: Wing, Propeller, Window, *Wheel*
Material: *Metal*, Glass

Shape:
Part: Window, *Wheel*, Door, Headlight, Side Mirror
Material: *Metal*, Shiny

Attribute Examples

Shape:
Part: Head, Ear, Nose, Mouth, Hair, Face, Torso, Hand, Arm
Material: Skin, Cloth

Shape:
Part: Head, Ear, Snout, Eye
Material: Furry

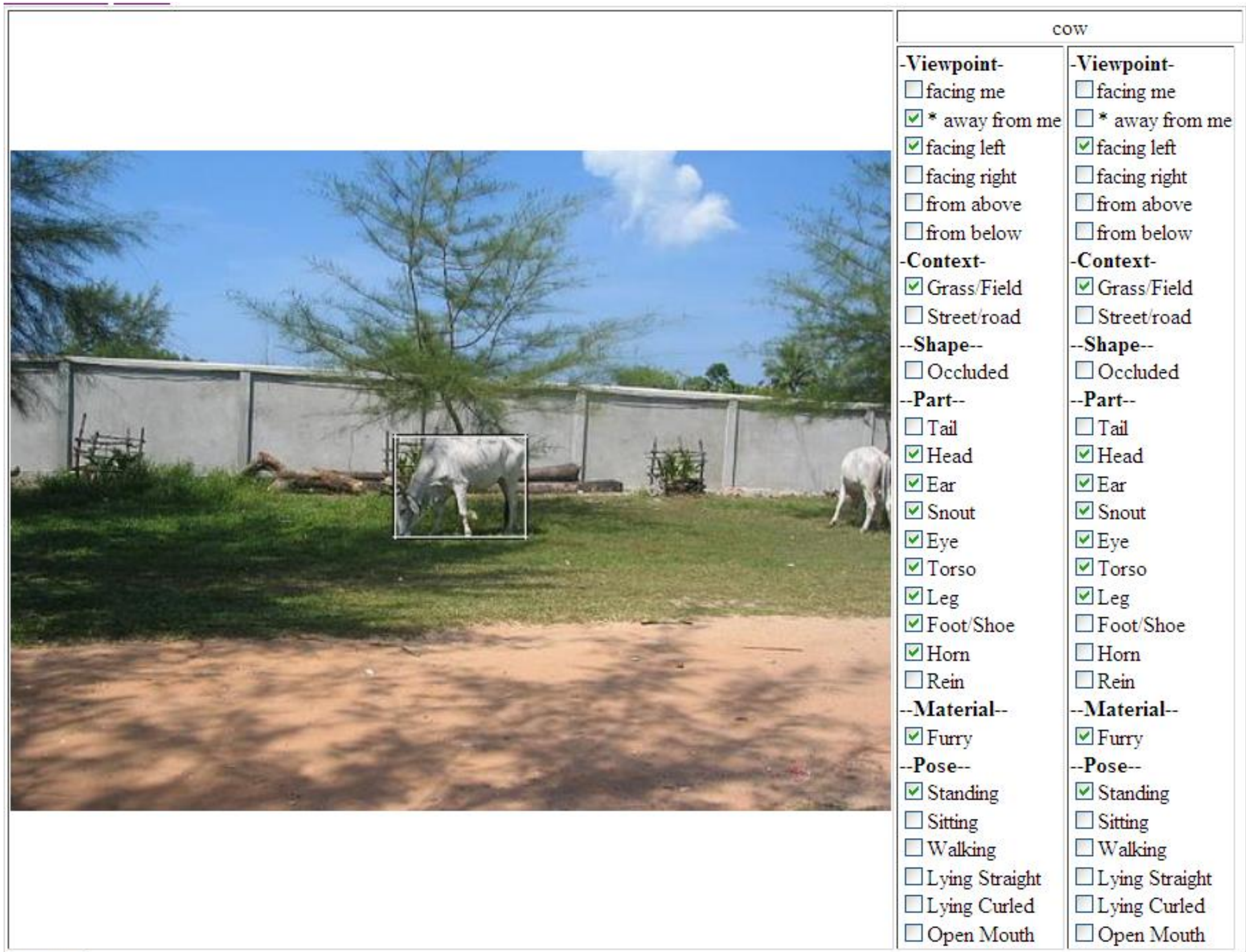


Shape:
Part: Head, Ear, Snout, Eye, Torso, Leg
Material: Furry

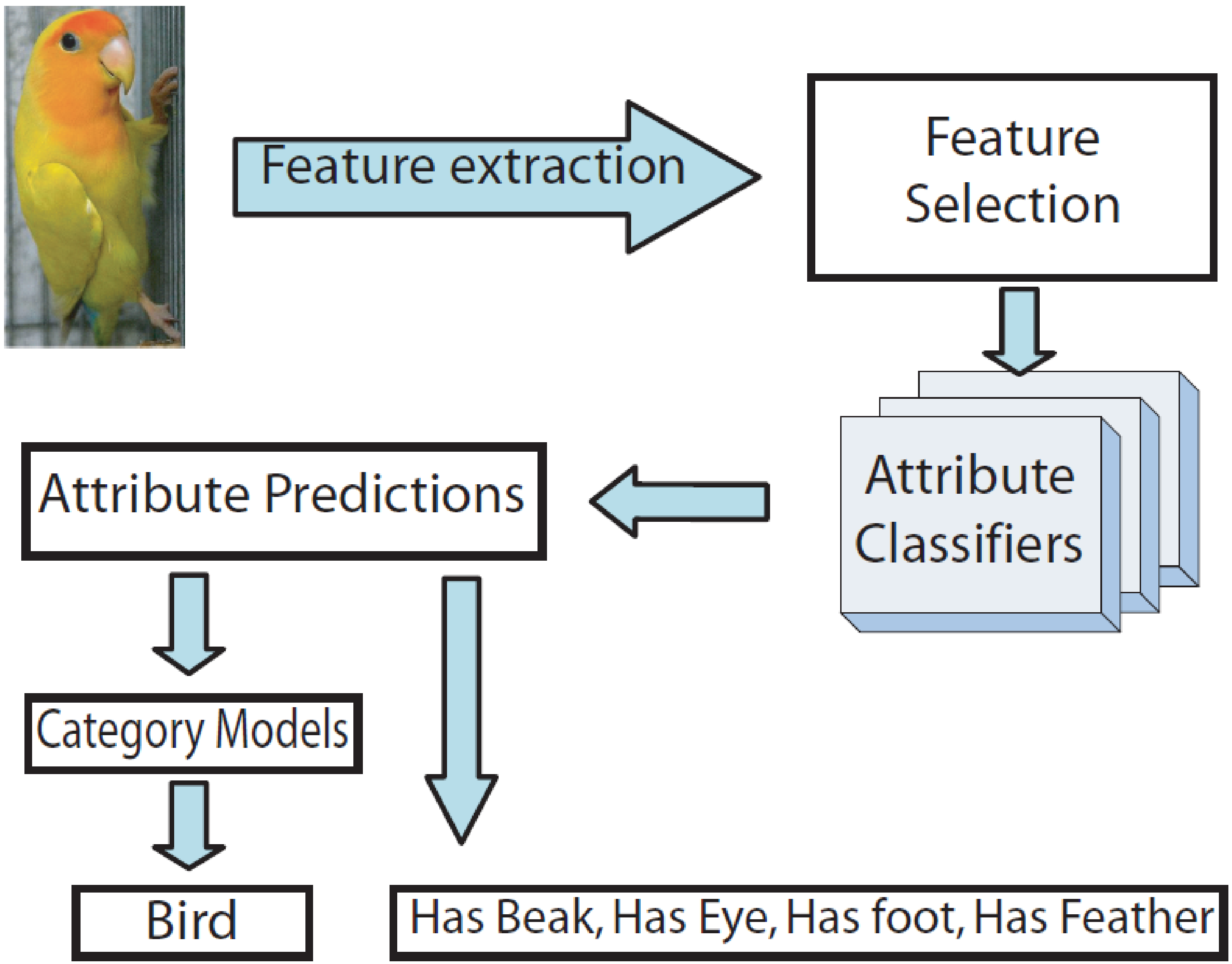
Datasets

- a-Pascal
  - 20 categories from PASCAL 2008 trainval dataset (10K object images)
    - airplane, bicycle, bird, boat, bottle, bus, car, cat, chair, cow, dining table, dog, horse, motorbike, person, potted plant, sheep, sofa, train, tv monitor
  - Ground truth for 64 attributes
  - Annotation via Amazon's Mechanical Turk
- a-Yahoo
  - 12 new categories from Yahoo image search
    - bag, building, carriage, centaur, donkey, goat, jet ski, mug, monkey, statue of person, wolf, zebra
  - Categories chosen to share attributes with those in Pascal
- Attribute labels are somewhat ambiguous
  - Agreement among "experts" 84.3
  - Between experts and Turk labelers 81.4
  - Among Turk labelers 84.1

Annotation on Amazon Turk



Our approach

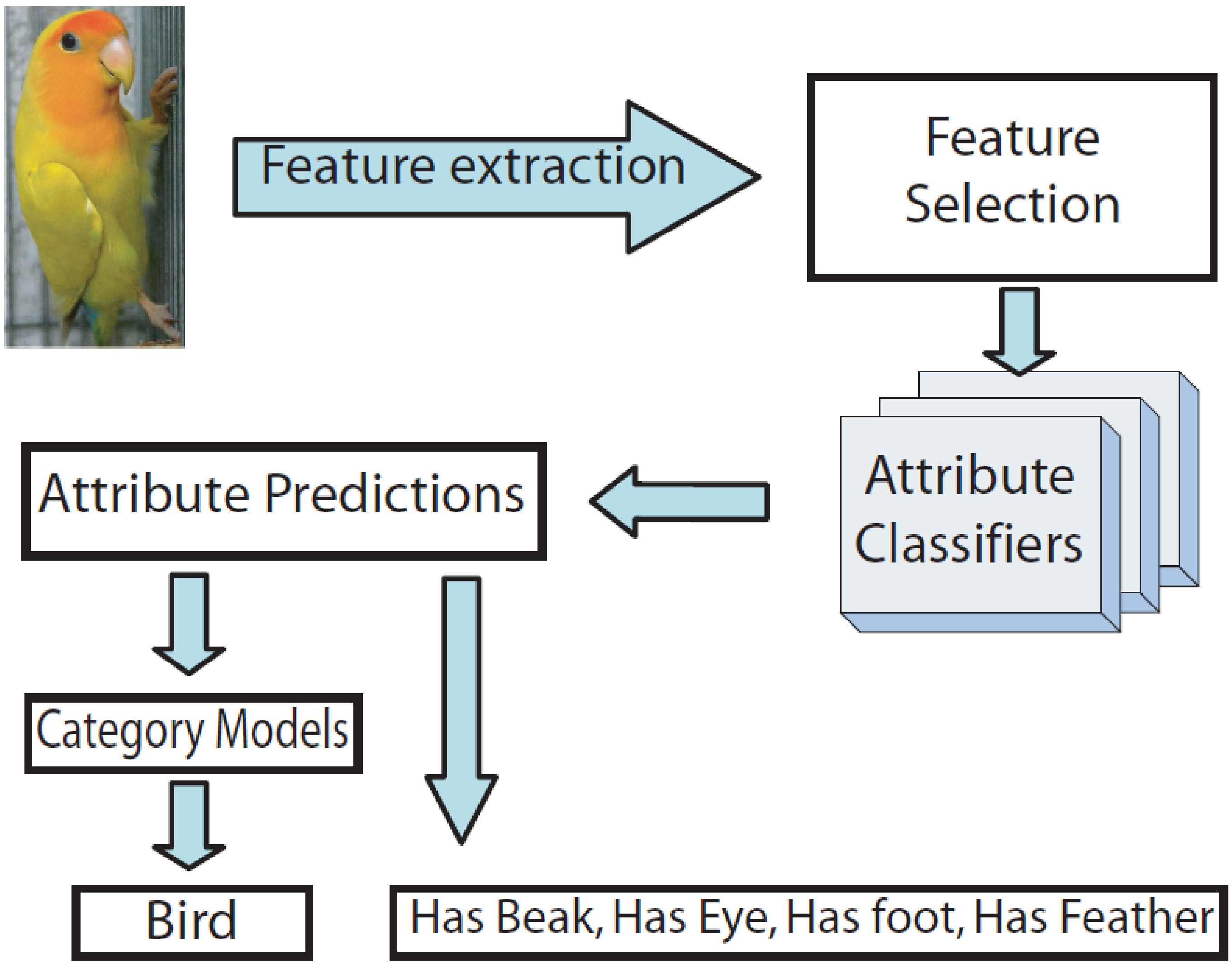


Features

Strategy: cover our bases

- Spatial pyramid histograms of quantized
  - Color and texture for **materials**
  - Histograms of gradients (HOG) for **parts**
  - Canny edges for **shape**

Our approach



Learning Attributes

- Learn to distinguish between things that have an attribute and things that do not
- Train one classifier (linear SVM) per attribute

Experiments

- Predict attributes for unfamiliar objects

- Identify what is unusual about an object

Describing Objects by their Attributes

No examples from these object categories were seen during training

Describing Objects by their Attributes

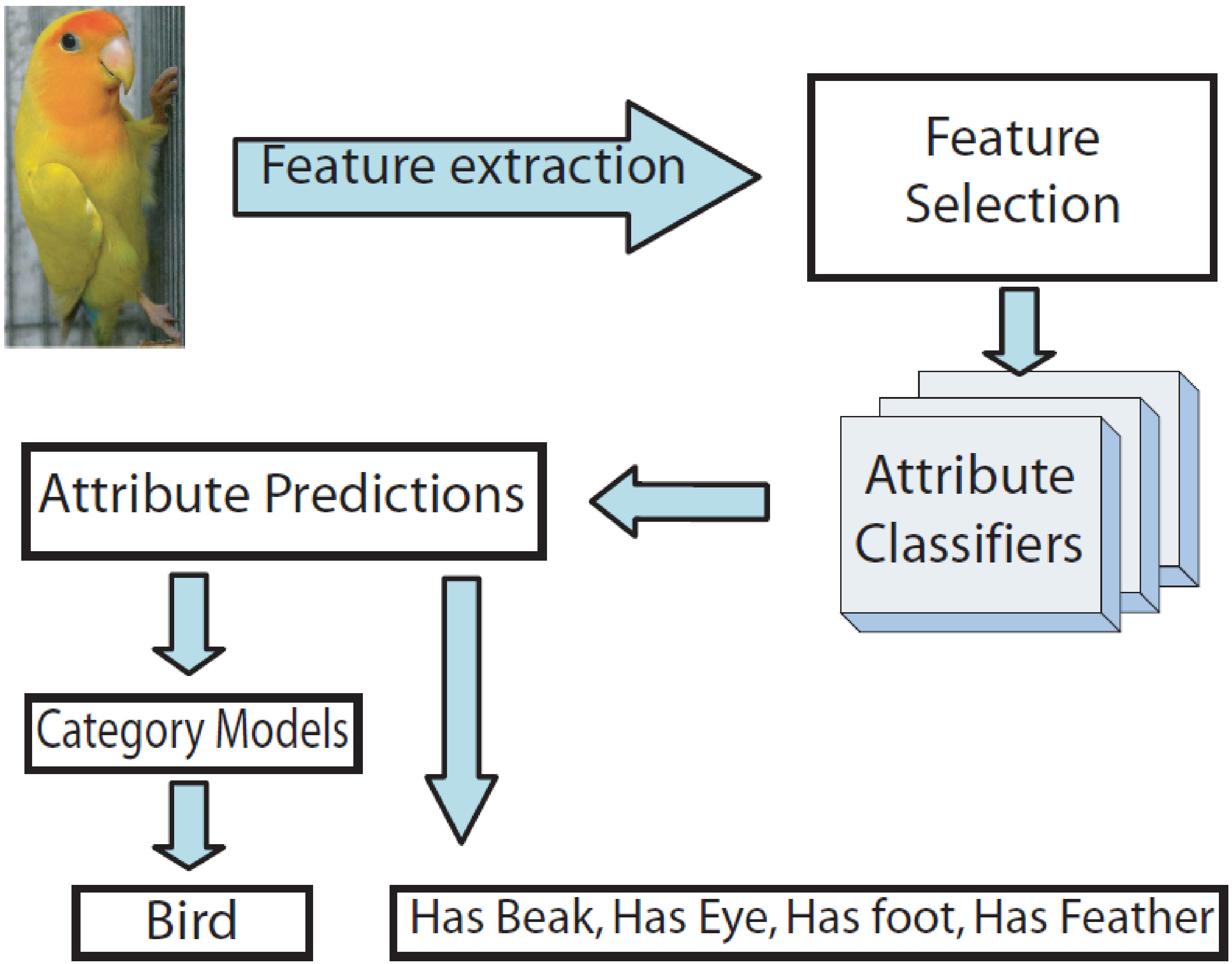
No examples from these object categories were seen during training

Average ROC Area

Trained on a-PASCAL objects

Test Objects	Parts	Materials	Shape
a-PASCAL	0.794	0.739	0.739
a-Yahoo	0.726	0.645	0.677

Our approach



Category Recognition

- Semantic attributes not enough
  - 74% accuracy even with ground truth attributes
- Introduce discriminative attributes
  - Trained by selecting subset of classes and features
    - Dogs vs. sheep using color
    - Cars and buses vs. motorbikes and bicycles using edges
  - Train 10,000 and select 1,000 most reliable, according to a validation set

Attributes not big help when sufficient data

- Use attribute predictions as features
- Train linear SVM to categorize objects

PASCAL 2008	Base Features	Semantic Attributes	All Attributes
Classification Accuracy	58.5%	54.6%	**59.4%**
Class-normalized Accuracy	35.5%	28.4%	**37.7%**

Identifying Unusual Attributes

- Look at predicted attributes that are not expected given class label

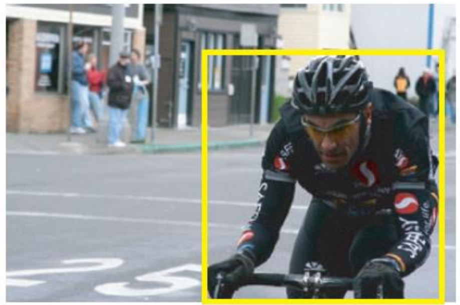
Absence of typical attributes

752 reports

68% are correct

Presence of atypical attributes

Motorbike
"cloth"



People
"label"

Bird
"Leaf"

Bus
"face"

Aeroplane
"beak"

Sofa
"wheel"

Bike
"Horn"

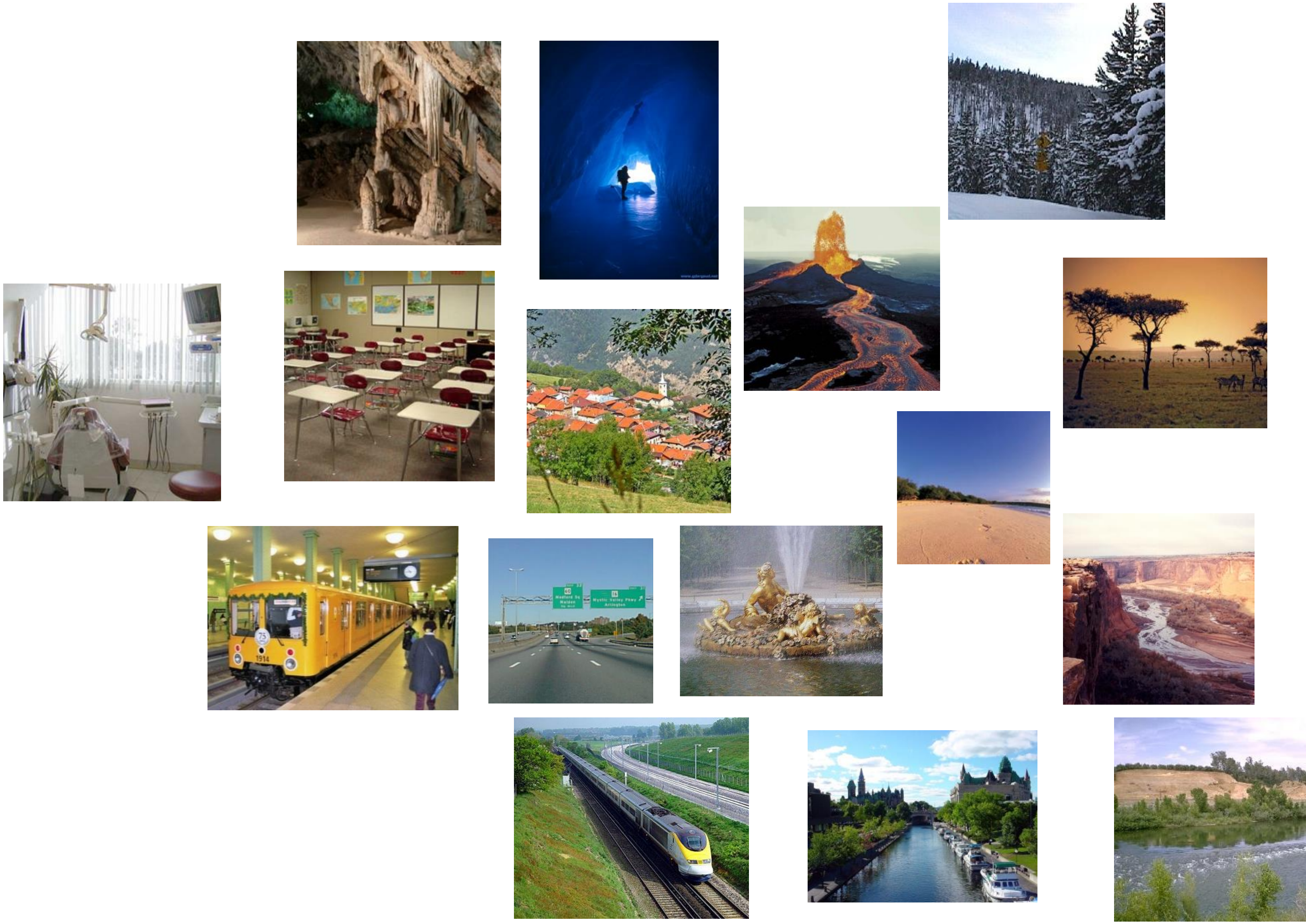
951 reports

47% are correct

Today – Crowd enabled recognition

- Recognizing Object Attributes
- Recognizing Scene Attributes

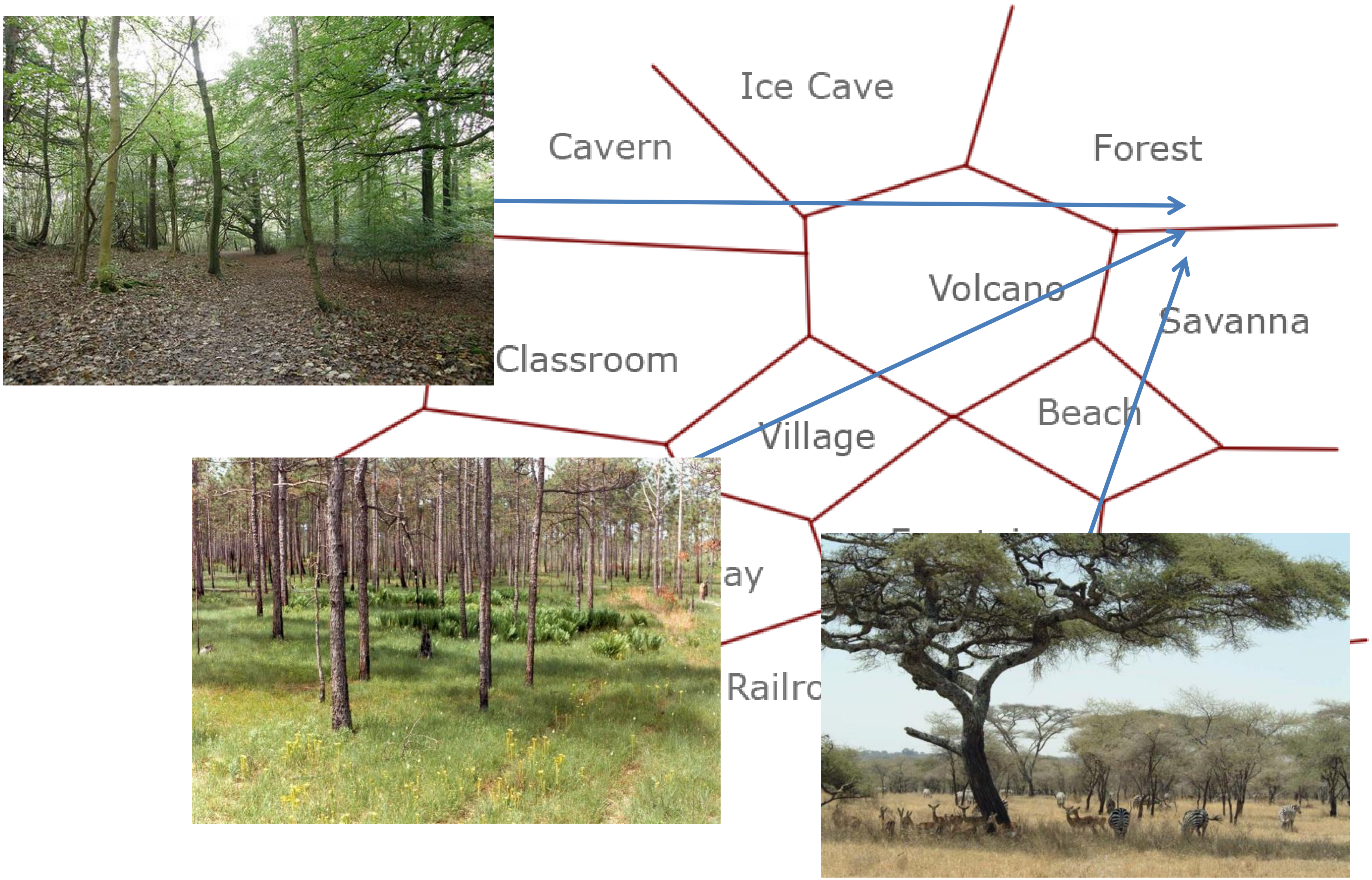
Space of Scenes



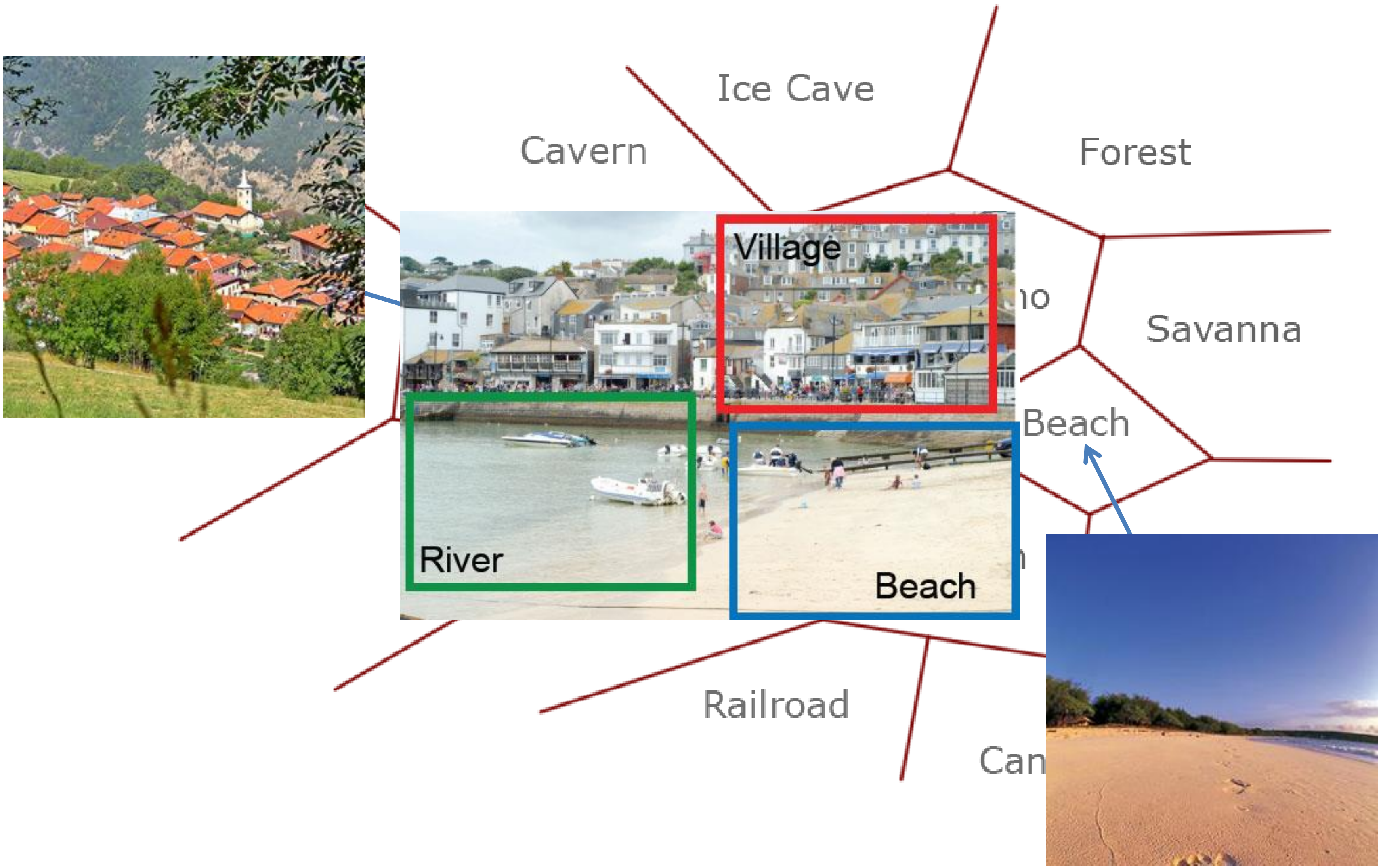
Genevieve Patterson and James Hays. CVPR 2012

Space of Scenes

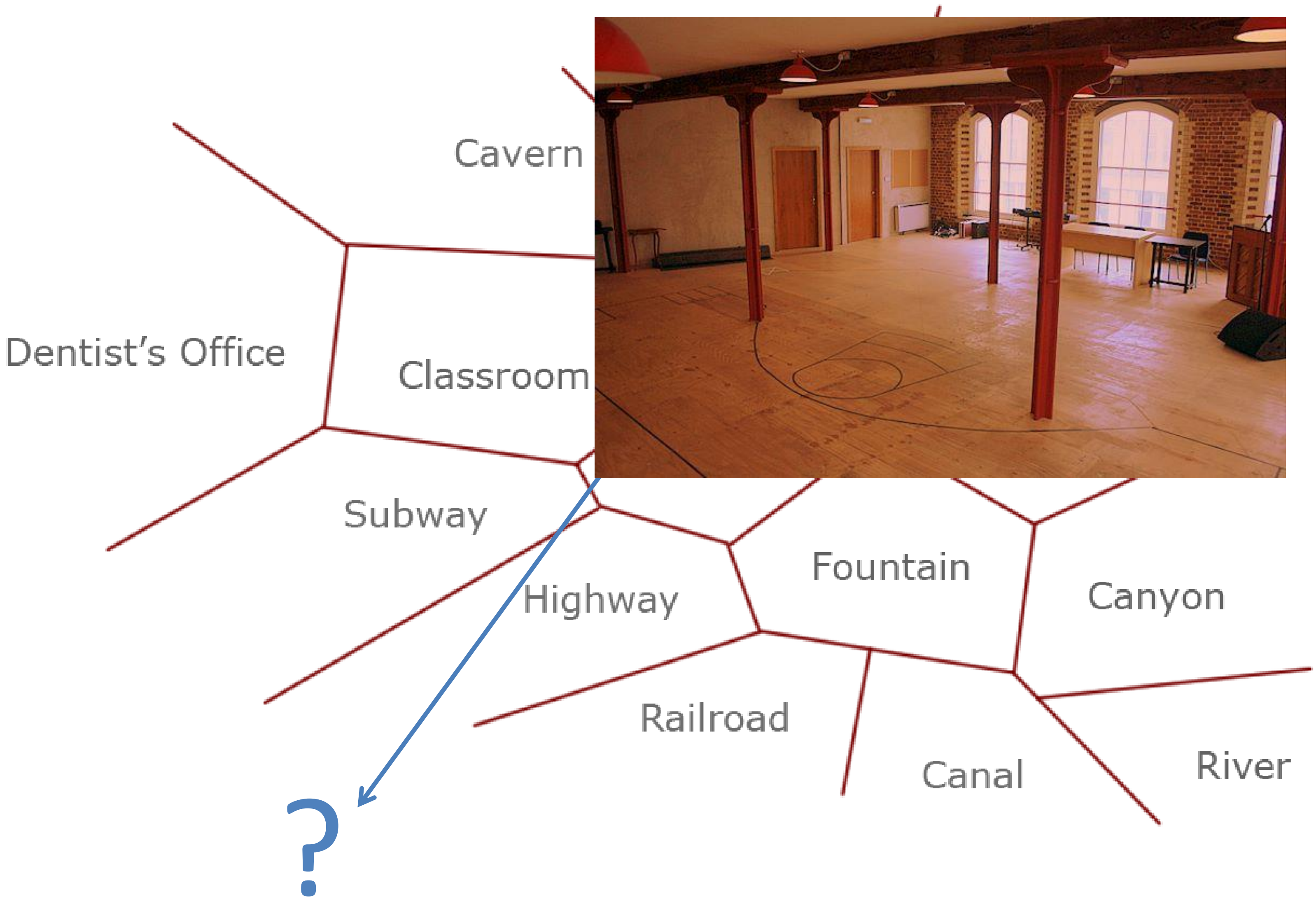
Space of Scenes



Space of Scenes



Space of Scenes



Space of Scenes

Big Picture

- Scenes don’t fit neatly into categories.
  - Objects often do!
- Categories aren’t expressive enough.

- We should reason about scene *attributes* instead of (or in addition to) scene categories.

Attribute-based Visual Understanding

Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer.

Lampert, Nickisch, and Harmeling. CVPR 2009.

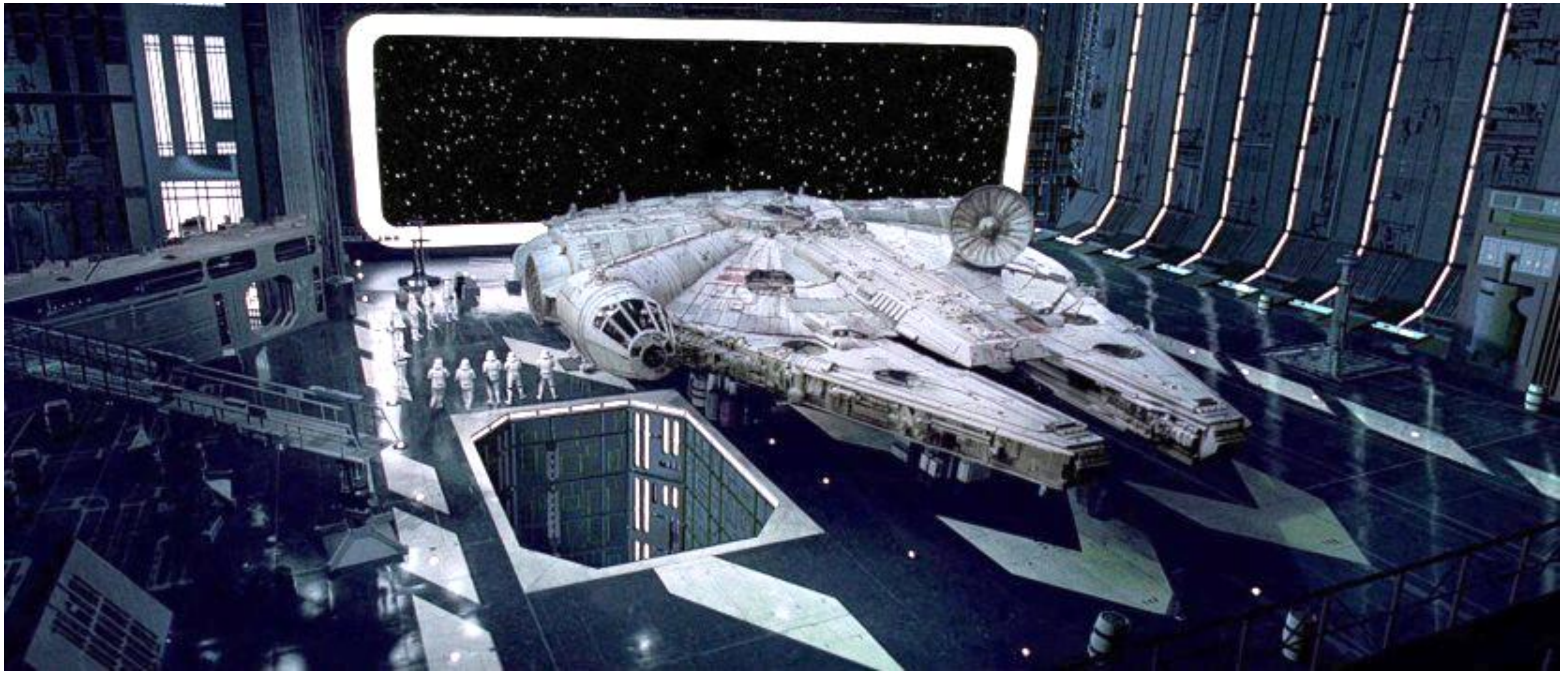
Describing Objects by their Attributes.

Farhadi, Endres, Hoiem, Forsyth. *CVPR* 2009.

Attribute and Simile Classifiers for Face Verification.

Kumar, Berg, Belhumeur, Nayar. ICCV 2009.

Numerous more recent works on **activity**, **texture**, **3d models**, etc.



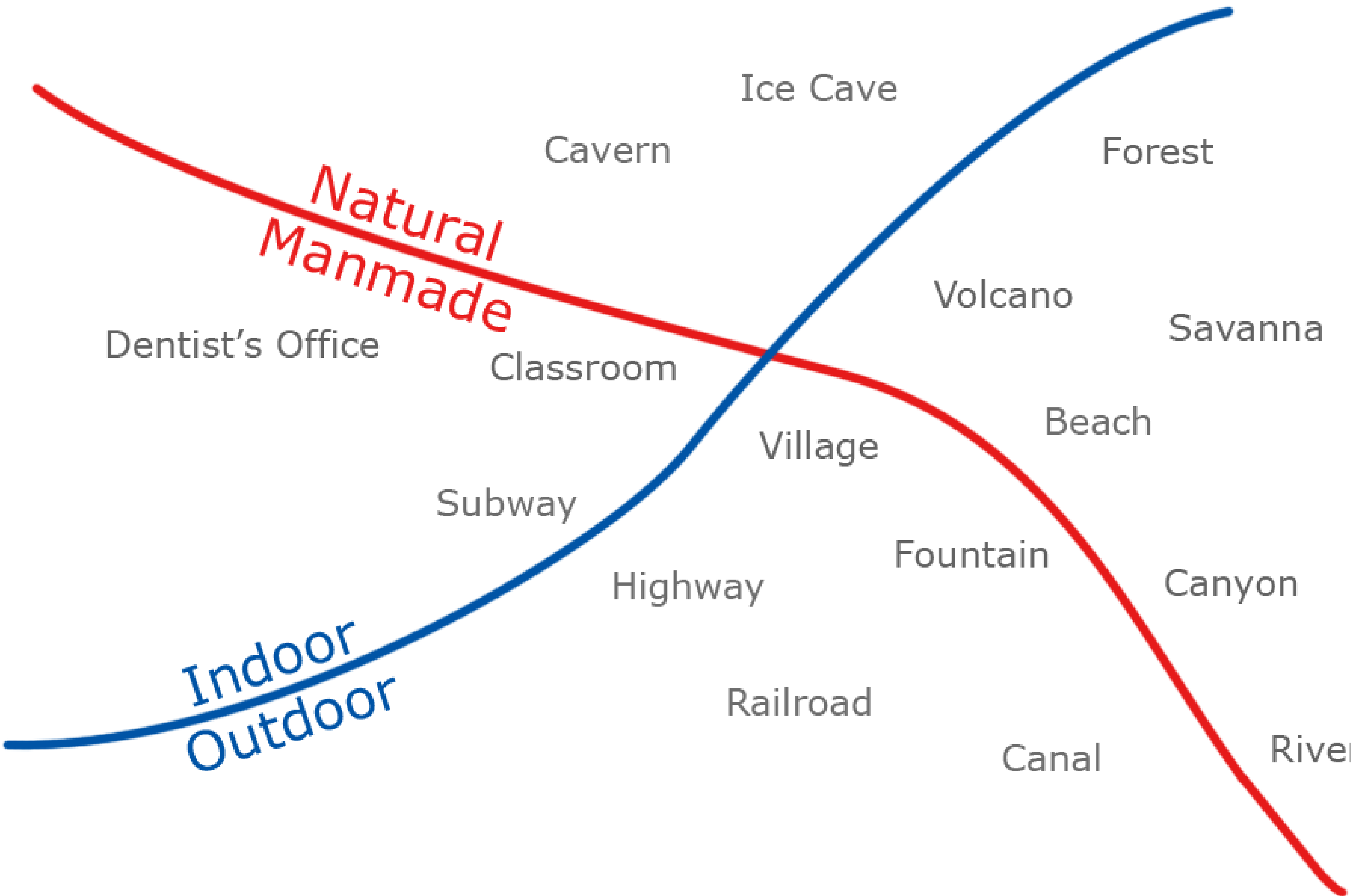
- Spatial layout: large, enclosed
- Affordances / functions: can fly, park, walk
- Materials: shiny, black, hard
- Object presence: has people, ships
- Simile: looks like Star Trek
- Emotion: scary, intimidating

Space of Scenes

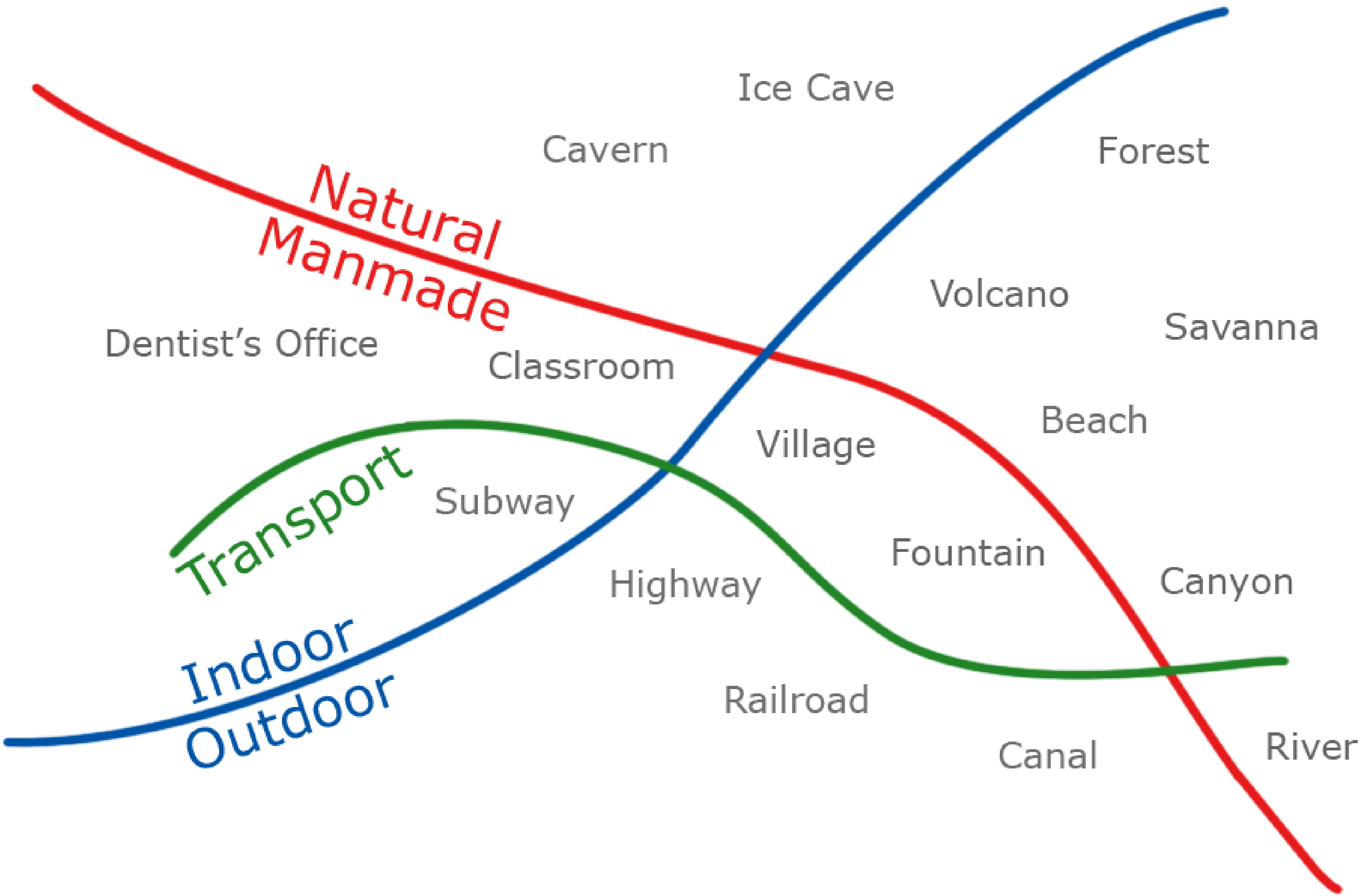
Ice Cave
Cavern
Forest
Volcano
Savanna
Dentist's Office
Classroom
Beach
Village
Subway
Fountain
Highway
Canyon
Railroad
Canal
River

Space of Scenes

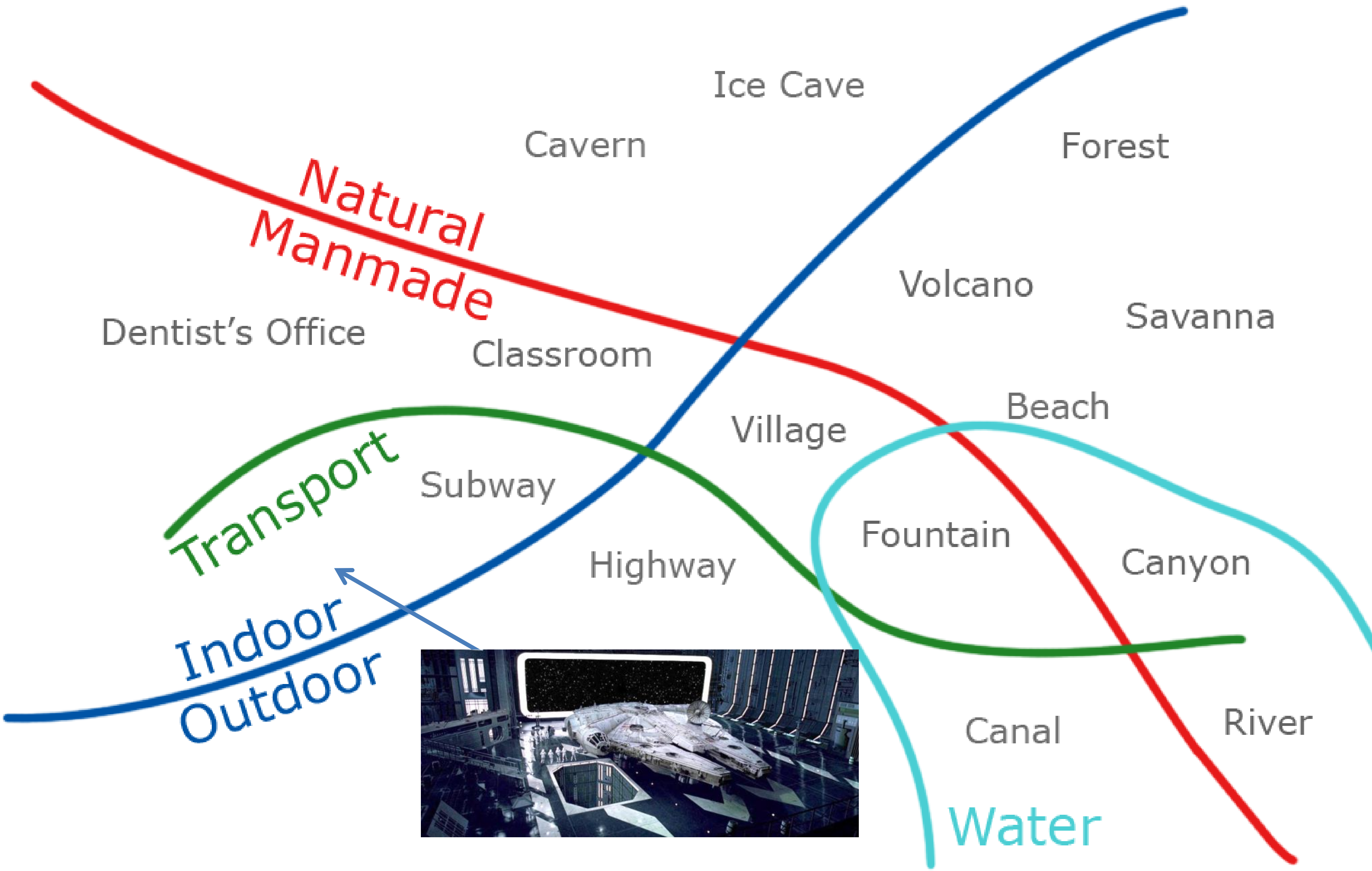
Space of Scenes



Space of Scenes

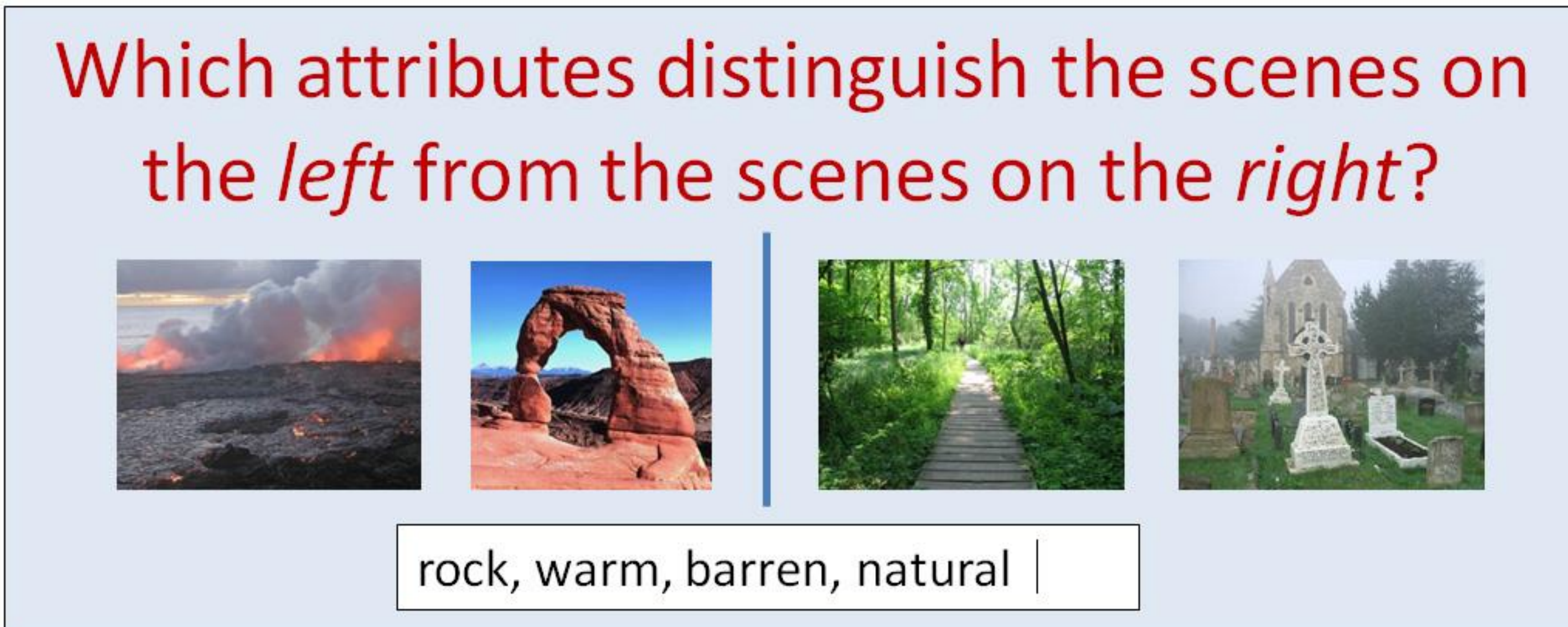


Space of Scenes



Which Scene Attributes are Relevant?

Inspired by the “splitting” task of Oliva and Torralba and “ESP game” by von Ahn and Blum.



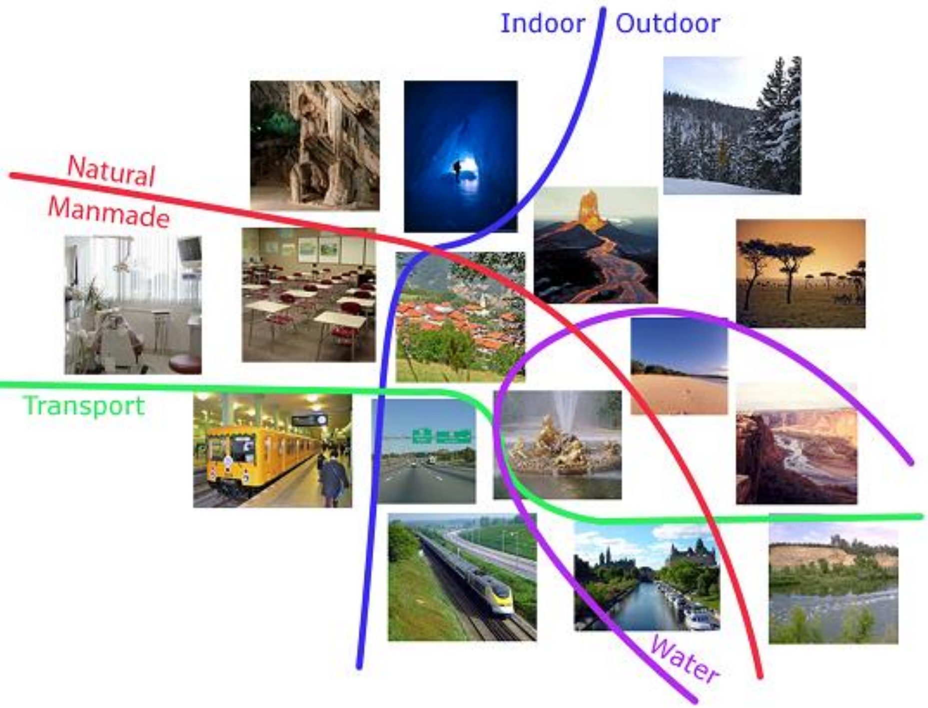
102 Scene Attributes

grass trees sand tiles aged waves dirt
biking marble leaves natural glass rusty matte snow
bathing railroad rock flowers working concrete cloth
exercise no-horizon glossy
competing gaming stressful cluttered cleaning rugged digging
vegetation climbing fencing ice using-tools pavement
symmetrical still-water studying scary constructing swimming
queuing electric-light medical-activity dry conducting-business
congregating vinyl sunbathing natural-light running-water hiking
diving far-away-horizon teaching wire transporting vacationing
spectating shrubbery soothing semi-enclosed sailing
damp socializing sterile praying camping shingles shopping
direct-sun farming open playing reading man-made cold
foliage carpet metal ocean eating research horizontal fire
driving paper wood smoke railing plastic asphalt
clouds warm brick
dirty sports vertical enclosed


Scene Attribute Labeling
Click on the scenes below that contain the following lighting or material:
camping: Either an actual camp site, or scene in wilderness suitable enough for humans to make a tent and/or sleep.
Example Scene
Example Scene
When you mouse over one of the images, a larger version of that image will appear in the box below.
These HITs are reviewed before being approved or rejected.
For futher instructions Click Here!
This task can be very subjective. If you are not sure about which images should be selected, please *SKIP THIS HIT* or email us to ask for clarification. There are more HITs with less subjective attributes.
Images continued down the page ...

SUN Attributes: A Large-Scale Database of Scene Attributes

http://www.cs.brown.edu/~gen/sunattributes.html



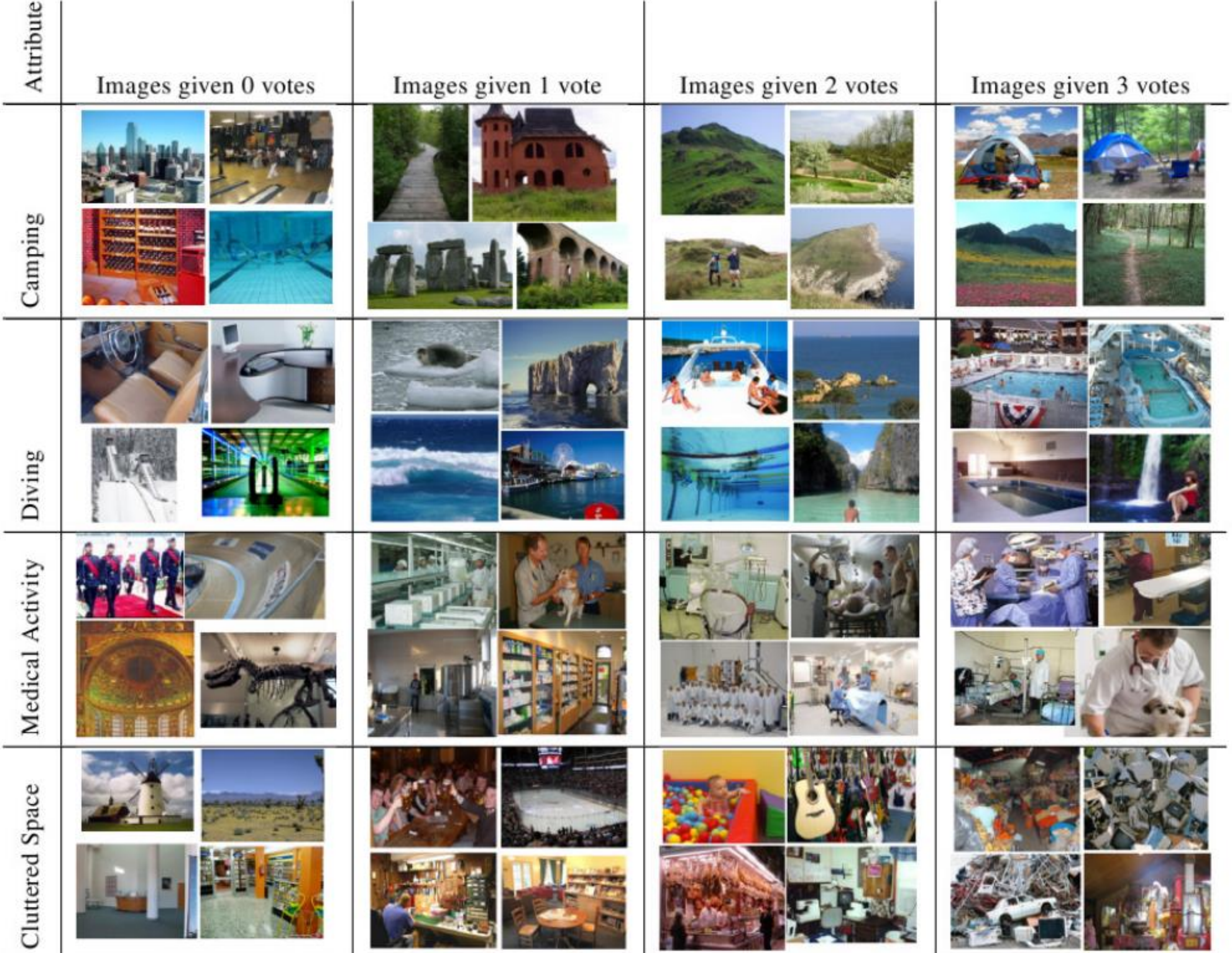
Space of Scenes
Organized by Attributes

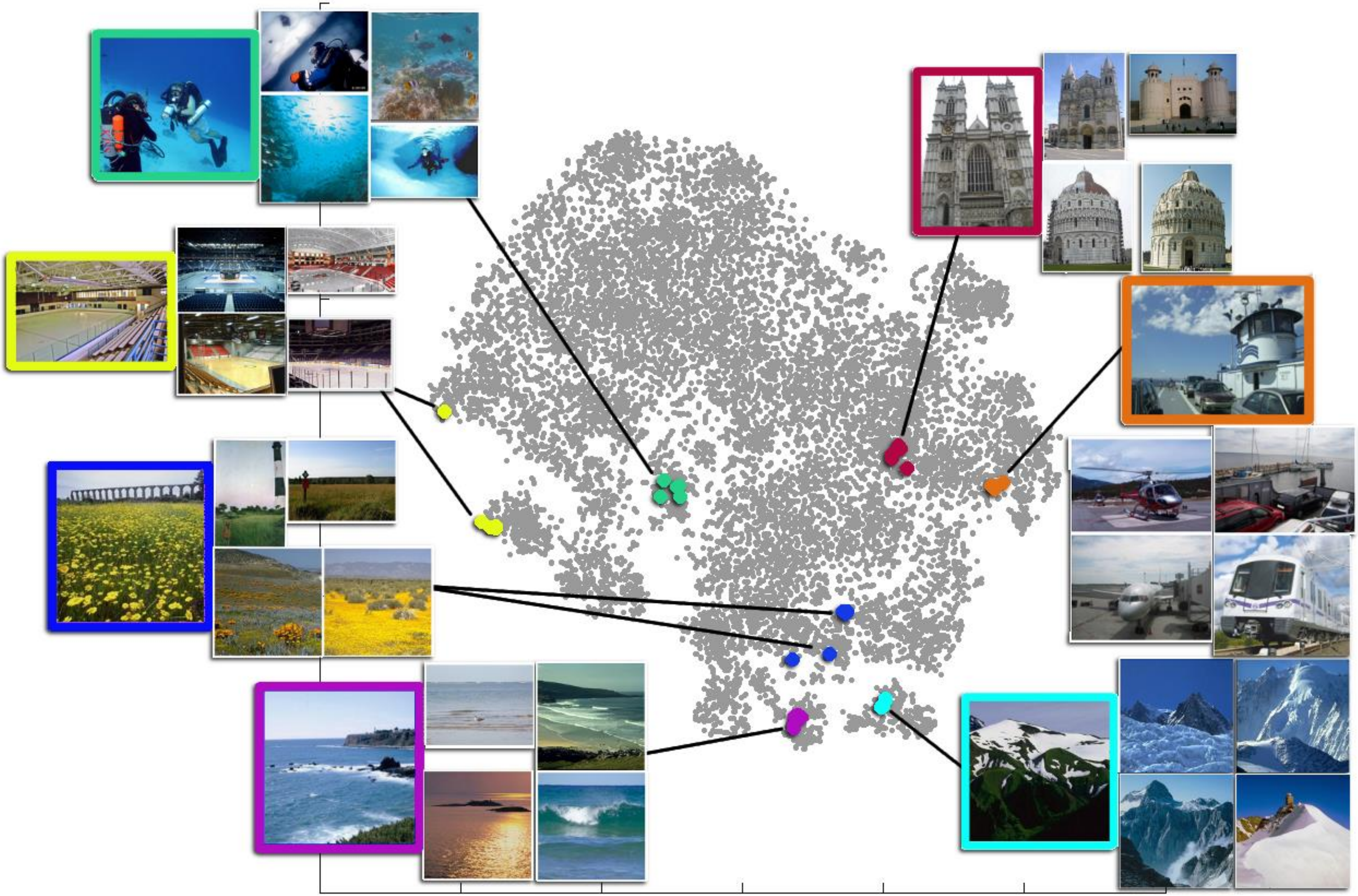
Global, binary attributes describing:

- Affordances / Functions (*e.g. farming, eating*)
- Materials (*e.g. carpet, running water*)
- Surface Properties (*e.g. aged, sterile*)
- Spatial Envelope (*e.g. enclosed, symmetrical*)

Statistics of database:

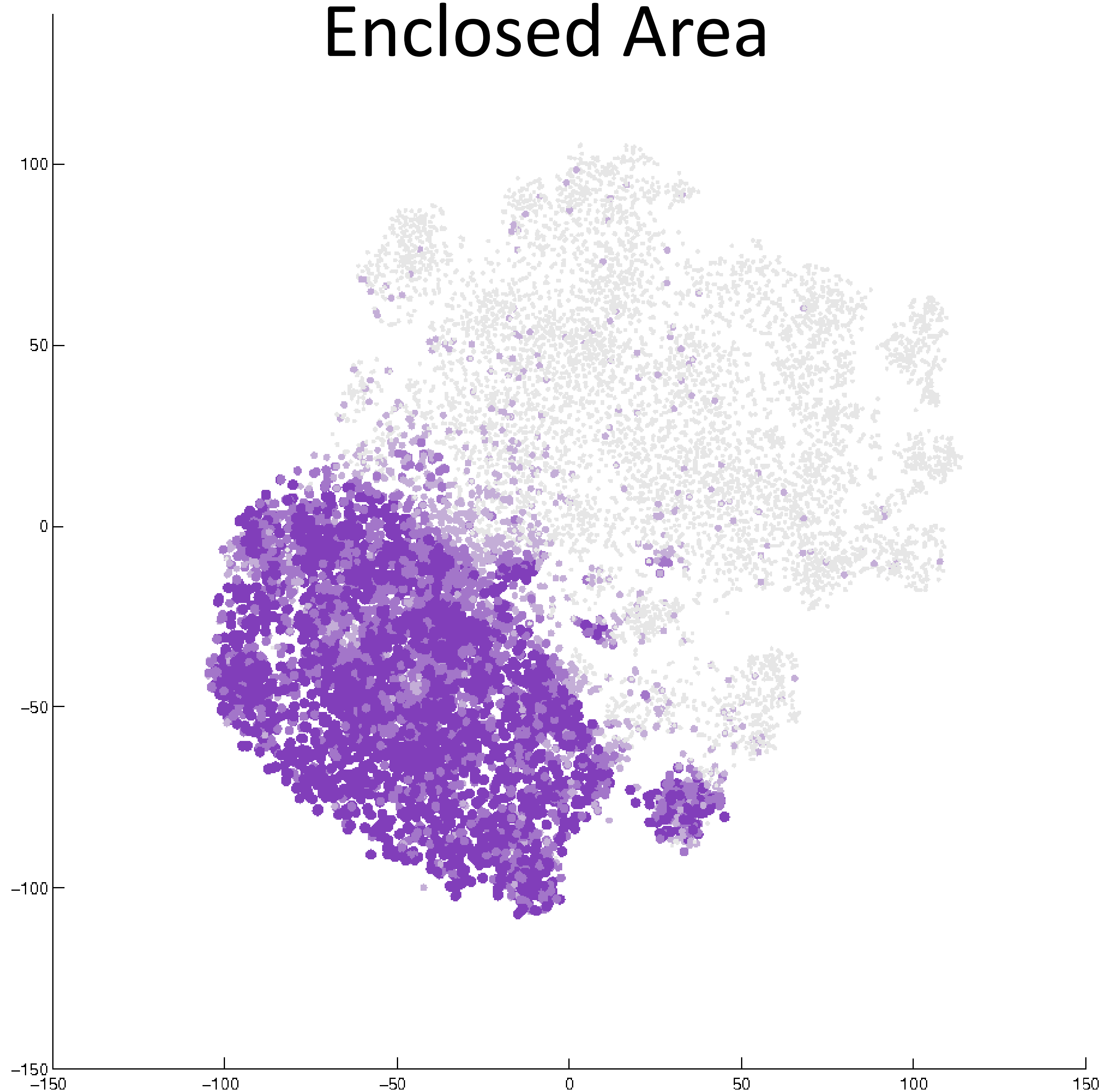
- 14,340 images from 717 scene categories
- 102 attributes
- 4 million+ labels
- good workers ~92% accurate
- pre-trained classifiers for download


Attribute
Images given 0 votes
Images given 1 vote
Images given 2 votes
Images given 3 votes
Camping
Diving
Medical Activity
Cluttered Space

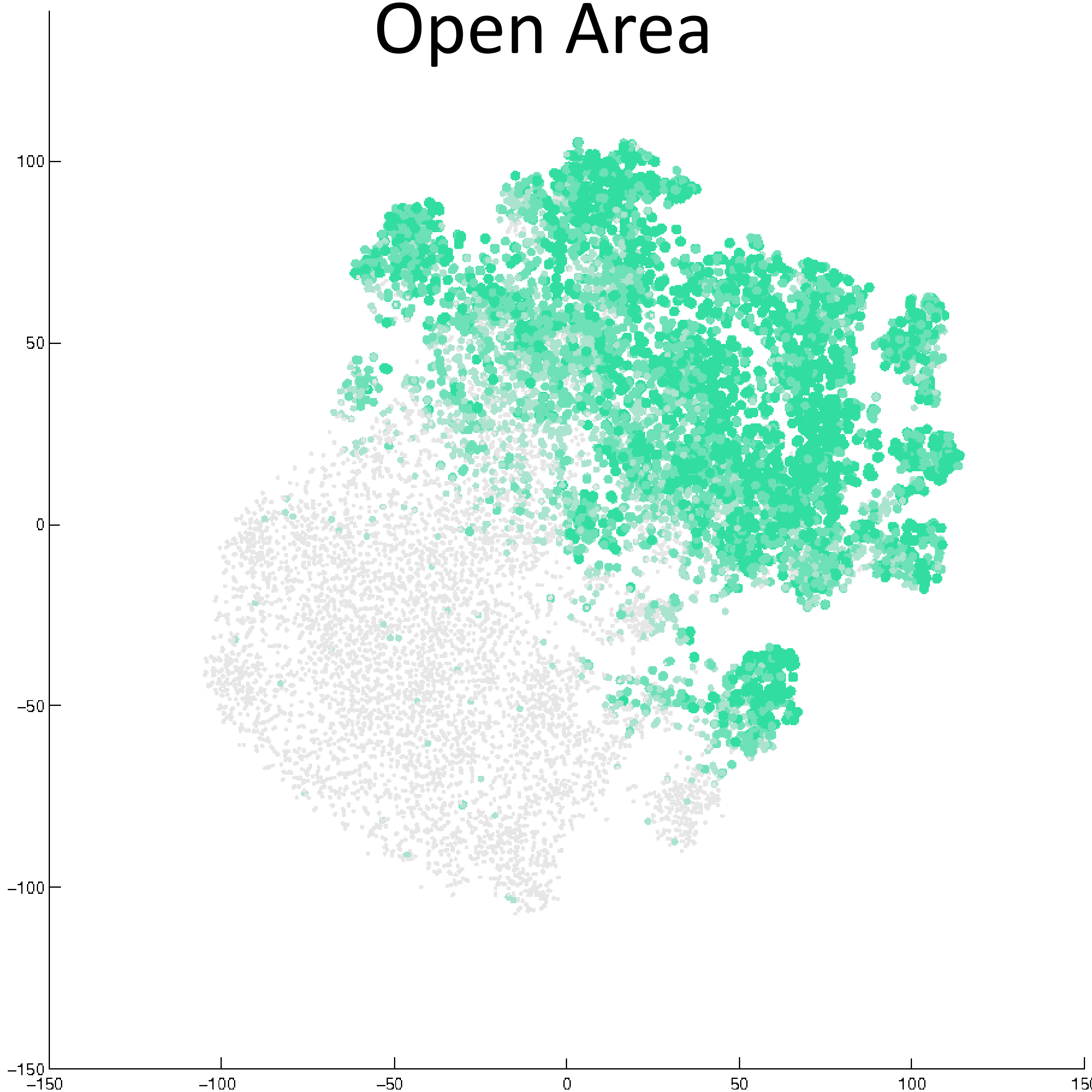


102 dimensional attribute space reduced to 2d with t-SNE

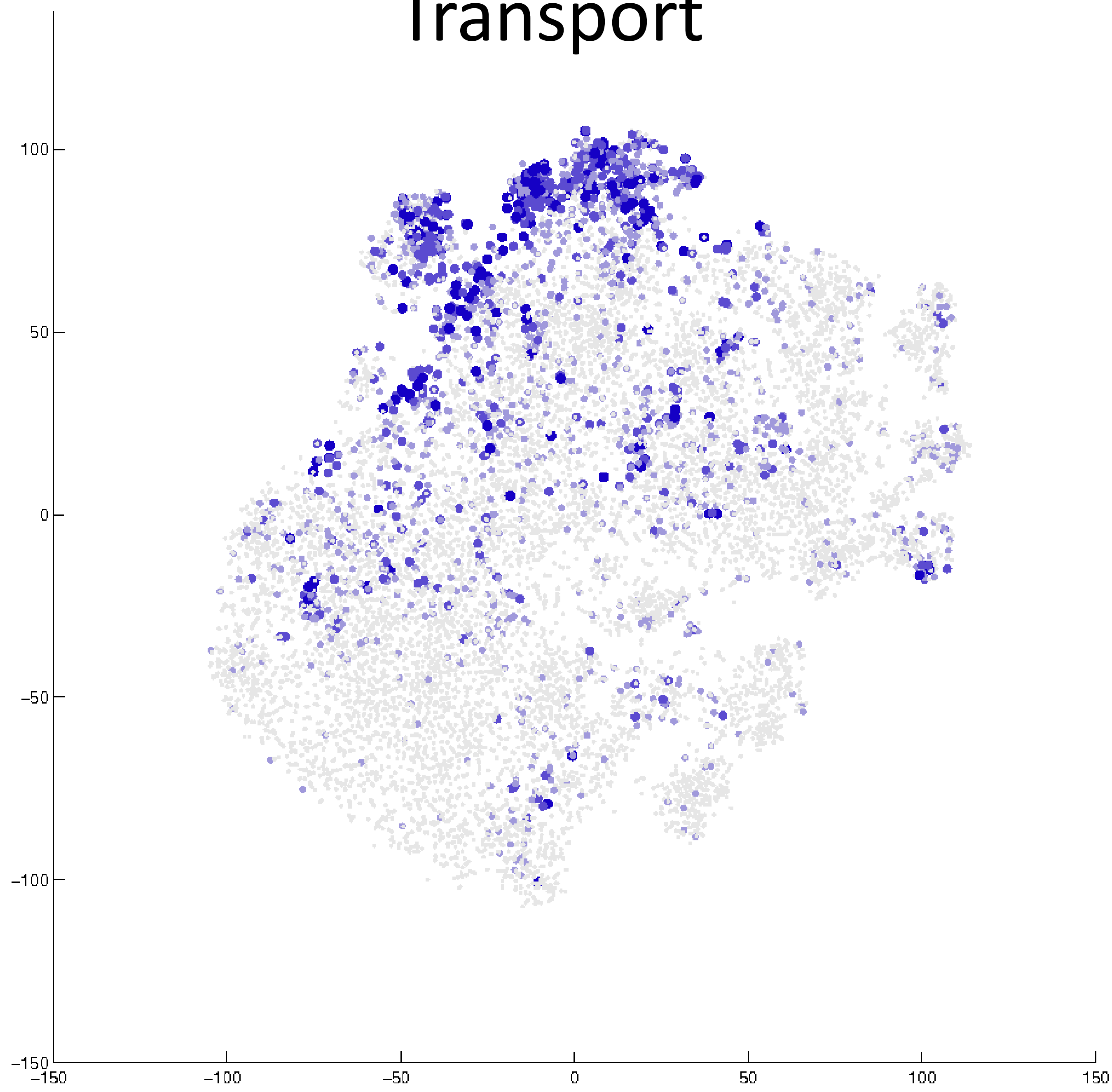
Enclosed Area



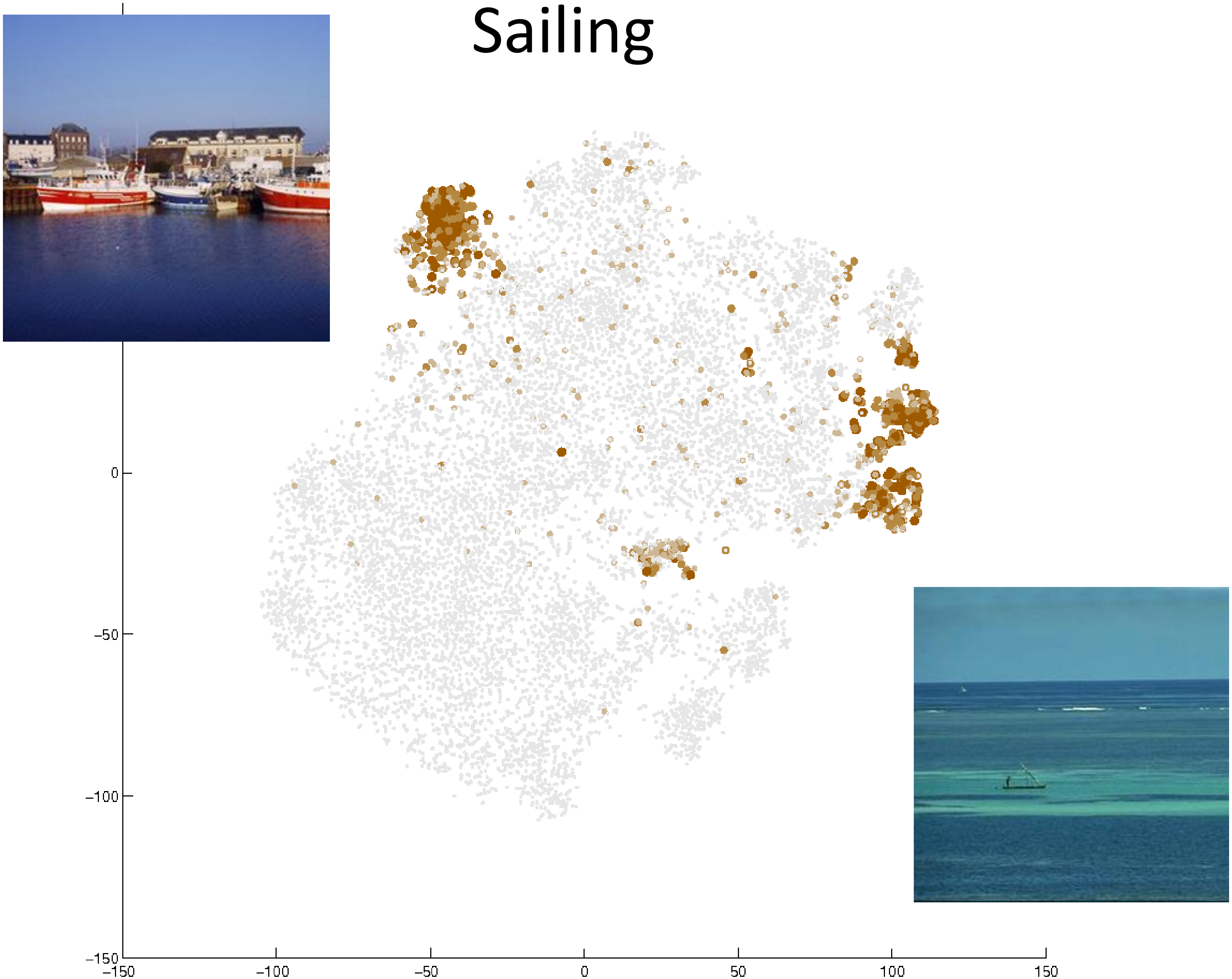
Open Area



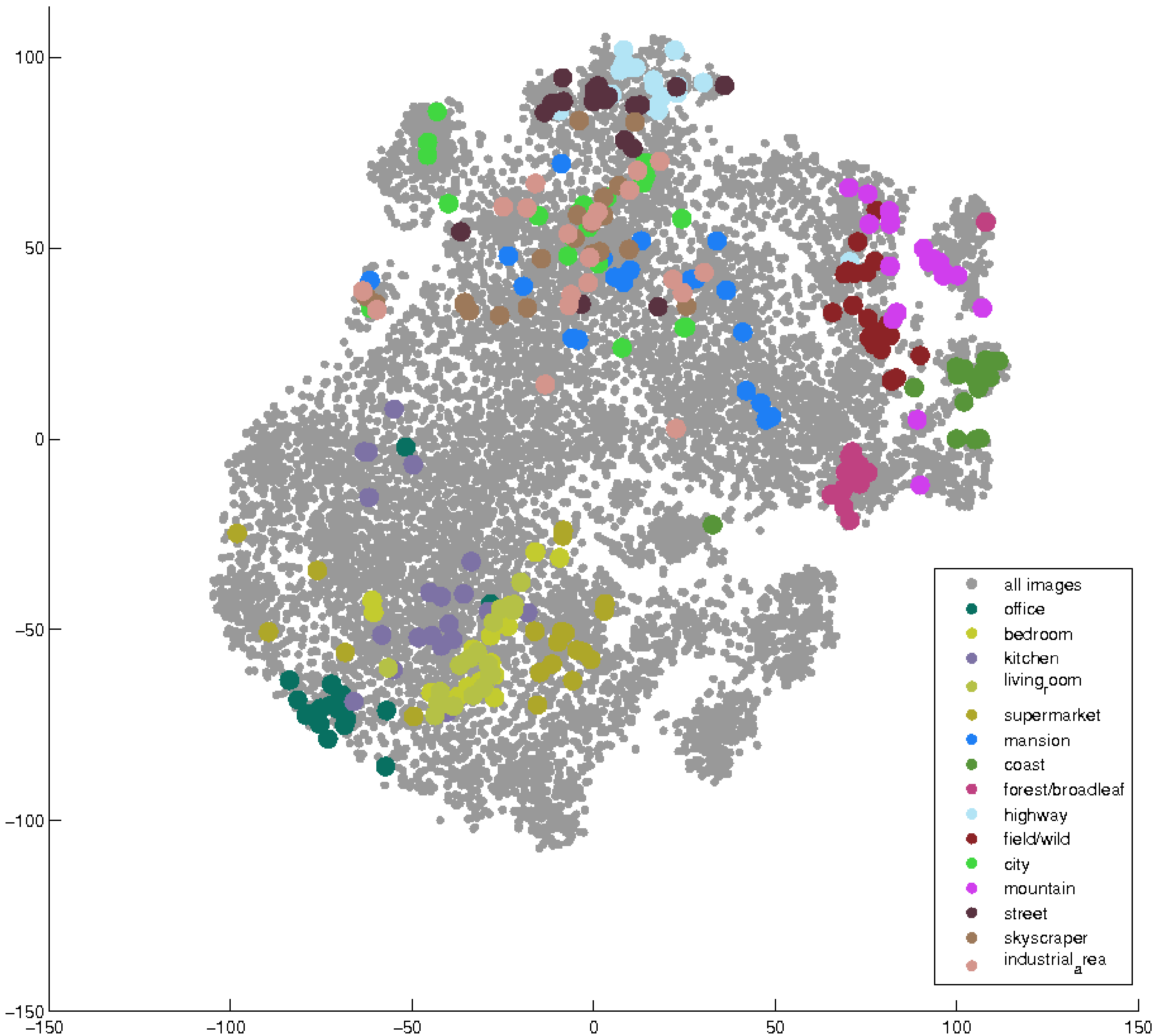
Transport



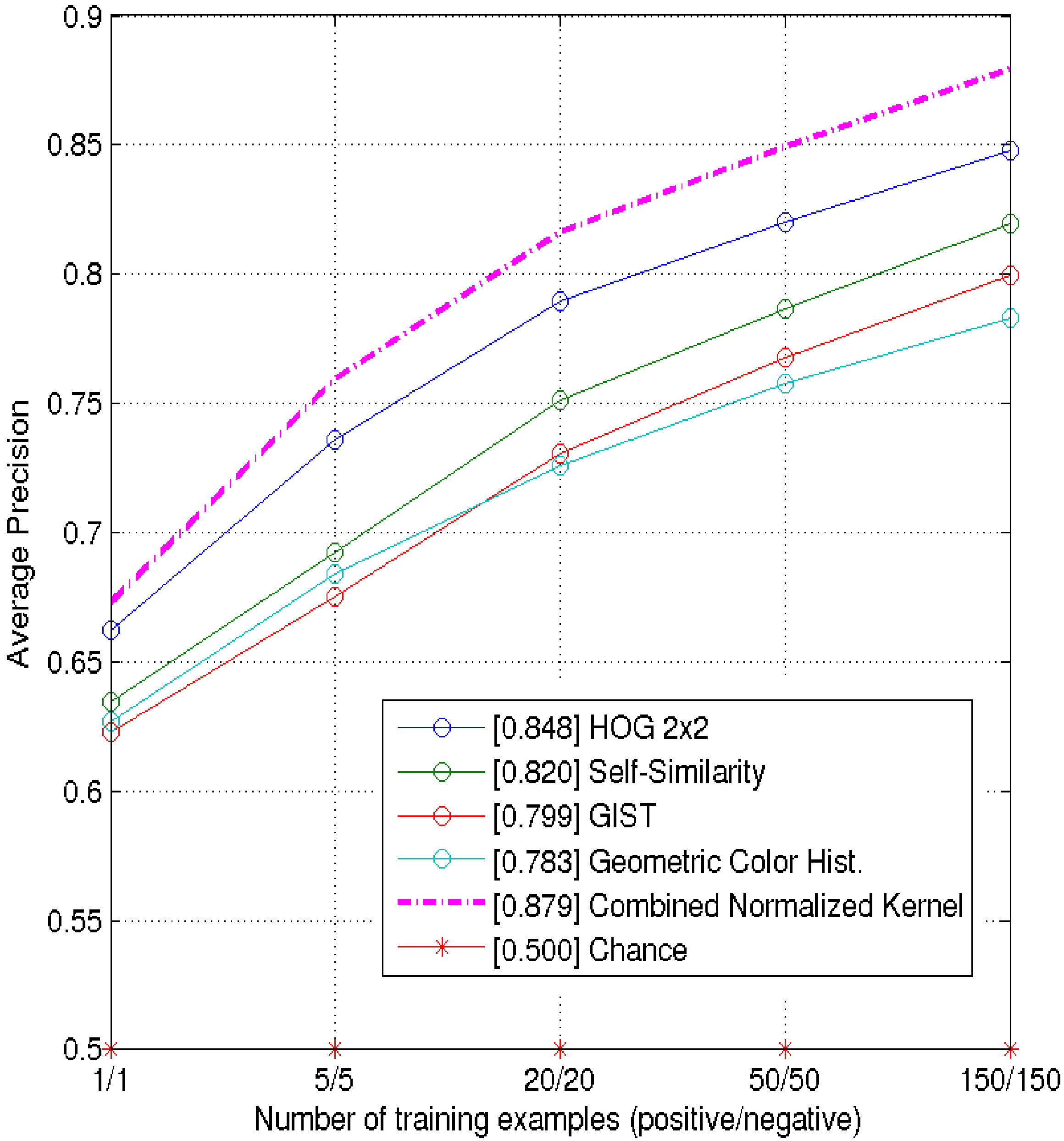
Sailing



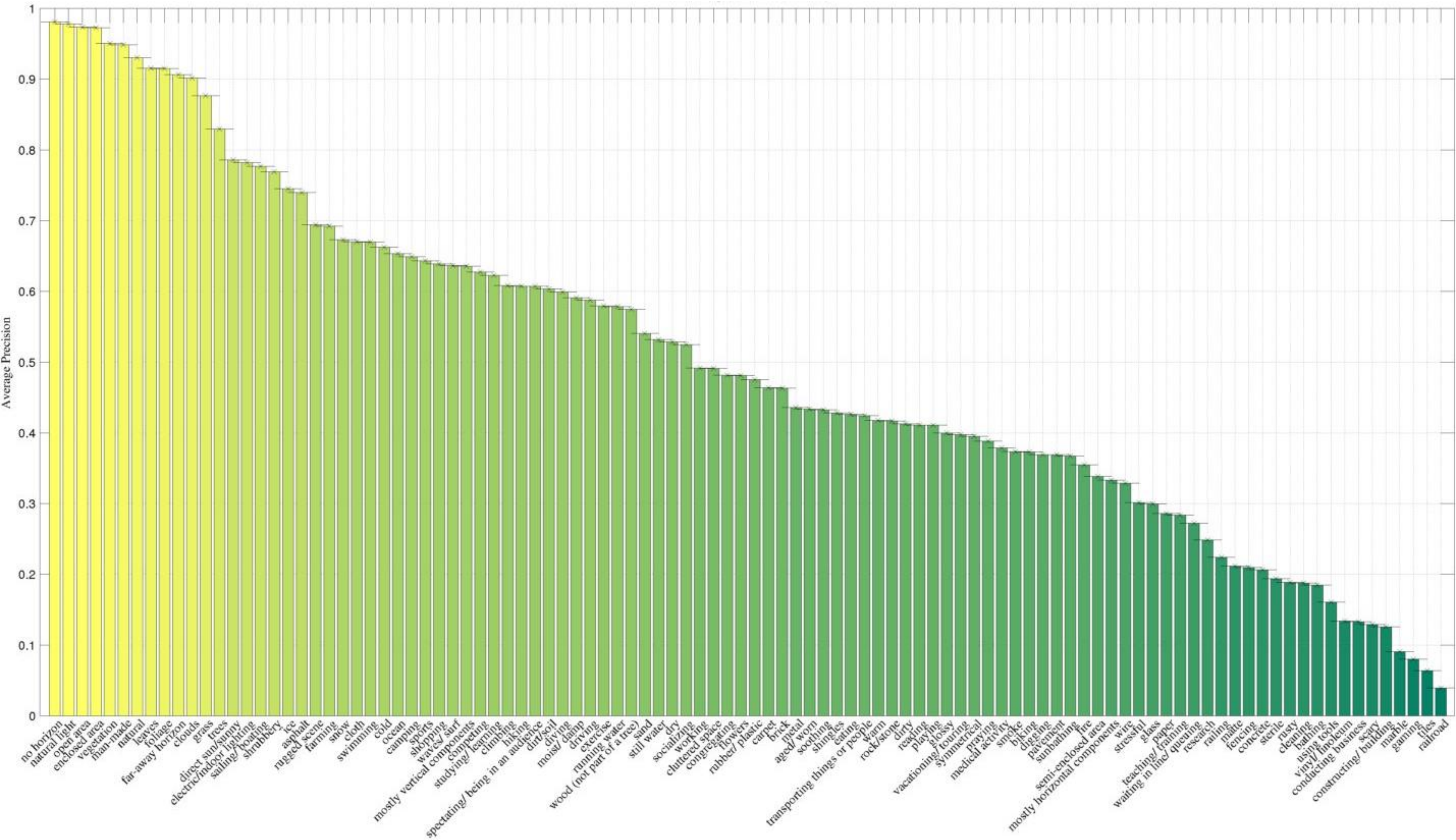
Instances of the “15 Scene” Categories



Average Precision of Attribute Classifiers



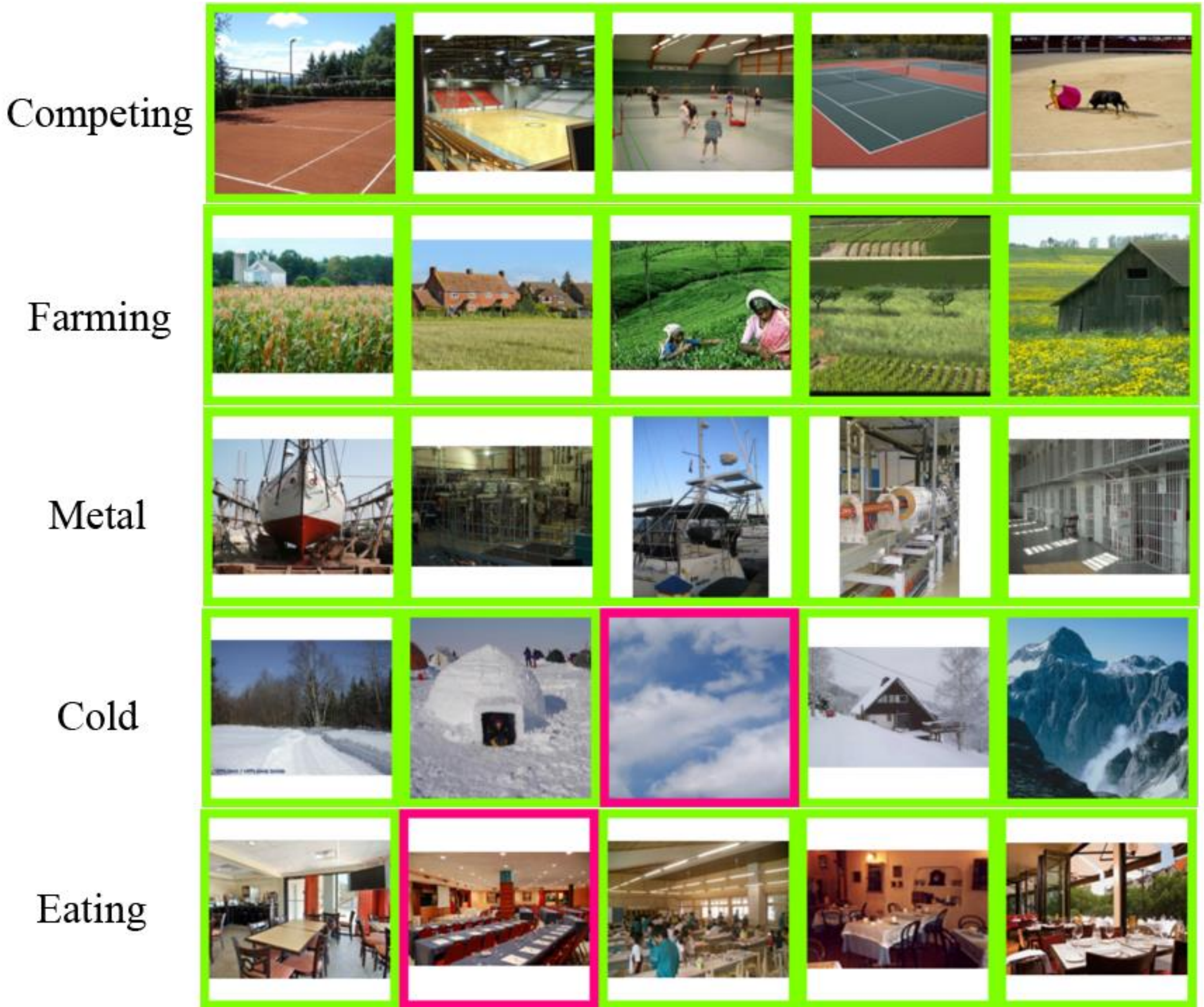
Average Precision of Attribute Classifiers



Attribute Recognition

Test Scene Images	*Highest Confidence* Attributes with Confidence Values	*Lowest Confidence* Attributes with Confidence Values
	0.74 vegetation	-1.33 studying
	0.63 open area	-1.36 gaming
	0.60 sunny	-1.38 fire
	0.57 sports	-1.42 carpet
	0.55 natural light	-1.60 tiles
	0.52 no horizon	-1.60 smoke
	0.51 foliage	-1.65 medical
	0.49 competing	-1.67 cleaning
	0.46 railing	-1.71 sterile
	0.46 natural	-1.74 marble
	0.91 eating	-1.07 gaming
	0.89 socializing	-1.11 running water
	0.70 waiting in line	-1.19 tiles
	0.51 cloth	-1.27 railroad
	0.42 shopping	-1.35 waves/ surf
	0.42 reading	-1.36 building
	0.39 stressful	-1.37 fire
	0.39 congregating	-1.40 bathing
	0.37 man-made	-1.50 ice
	0.31 plastic	-1.63 smoke

Most Confident Classifications



Most Confident Classifications

Recap: Attributes and Crowdsourcing

- If you can only get one label per instance, maybe a categorical label is the most informative.
- But now that crowdsourcing exists, we can get enough training data to simultaneously reason about a multitude of object / scene properties (e.g. attributes).
- In general, there is a broadening of interesting recognition tasks.
- Zero-shot learning: model category with an attribute distribution only.