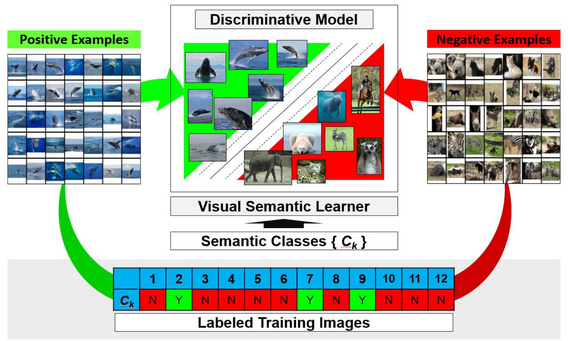
Semantics of Visual Discrimination:

Learning effective and useful visual semantic concepts using faceted hierarchical modeling

Introduction

- Data-driven learning is producing advances in image recognition
  - e.g., Convolutional Neural Nets (CNNs) performance on ImageNet
- Since image recognition is beginning to work at a scale, need more meaningful **visual discrimination** that reflects real life
- What this requires is:
  - Better representation of the visual semantic space
  - Support for multiple facets of visual content description
  - Facets for *people, objects, scenes, actions, activities, events*
- We point out critical semantic concept modeling issues and show how they impact visual discriminative learning
- We demonstrate how managing visual semantics using a faceted hierarchy improves discriminative learning and search in practice

Visual Semantic Learning Creates Discriminative Models – e.g., Support Vector Machines (SVMs), Ensemble Classifiers, Convolutional Neural Nets, etc.



What label?

Pick one:

a) Dog
b) Cat
c) Not Cat
d) Not Dog

Cat → Not Dog
Dog → Not Cat

Discriminative learning problem: more than simple photo bombing by cats

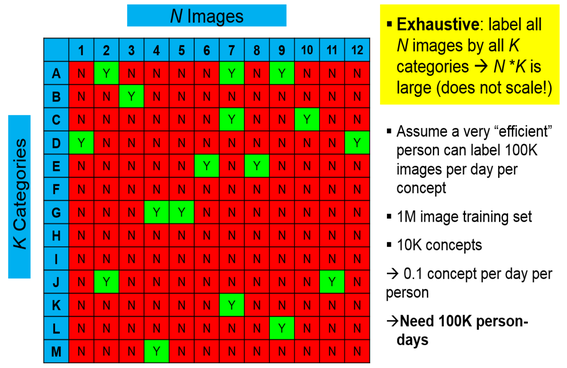
What label?

Pick one:

a) Zebra
b) Giraffe
c) Elephant
d) Deer

Zebra → Not Dog
Zebra → Not Giraffe
Zebra → Not Elephant
Zebra → Not Deer

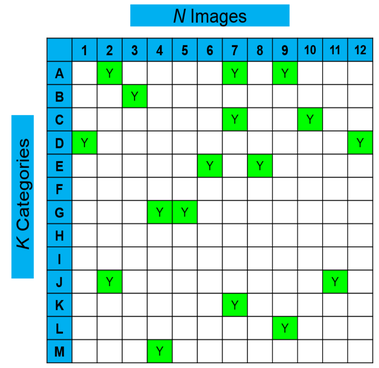
Paradox: images are mix-ins – need to support labeling of multiple semantics which is hard to do using discriminative models



- **Exhaustive**: label all *N* images by all *K* categories → *N*K* is large (does not scale!)
- Assume a very "efficient" person can label 100K images per day per concept
- 1M image training set
- 10K concepts

→ 0.1 concept per day per person

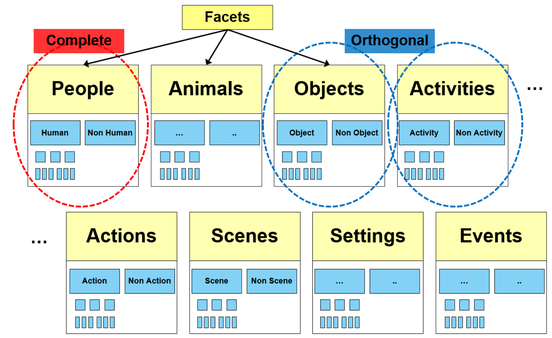
→**Need 100K person-days**



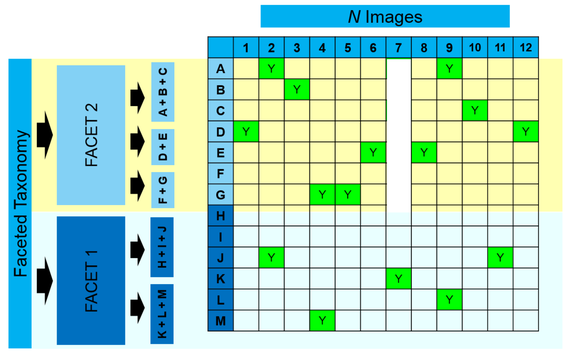
- **Exhaustive**: label all *N* images by all *K* categories → *N*K* is large (does not scale!)
- **Binary**: label *p* positives (*p*<<*M*) and *n* negatives (*n*<<*N*) for each of *K* categories → *(p+n)*K* << *N*K* (much better, still a lot)
- **Multi-Class**: label *p* positives for each of *K* categories and infer negatives → *p*K*

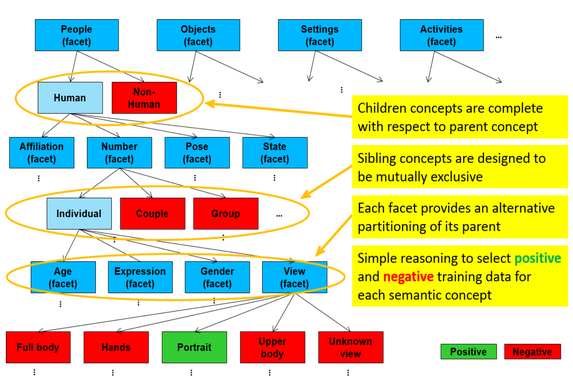
→ **Need 100 person-days**

Proposed Approach: Managing Visual Semantics using Faceted Taxonomy



Faceted Taxonomy Authoring and Maintenance (*p*K*$_i$ labels per FACET *i*):





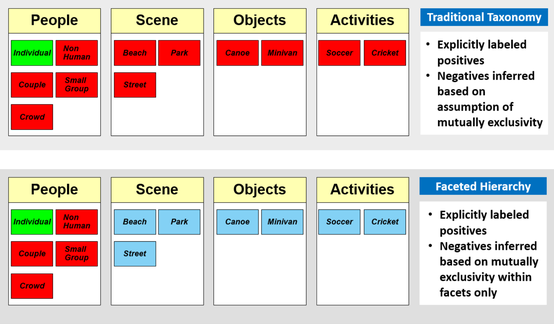
Children concepts are complete with respect to parent concept

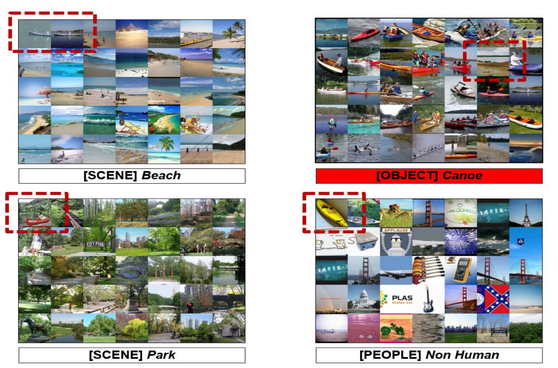
Sibling concepts are designed to be mutually exclusive

Each facet provides an alternative partitioning of its parent

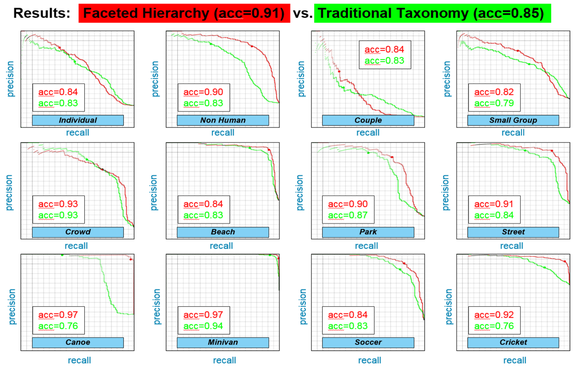
Simple reasoning to select positive and negative training data for each semantic concept

Experimental Setup: Faceted Hierarchy vs. Traditional Taxonomy

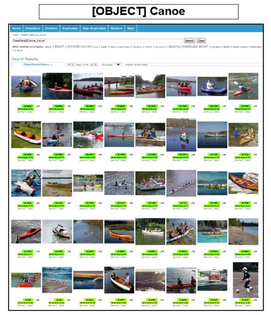




Results: Faceted Hierarchy (acc=0.91) vs. Traditional Taxonomy (acc=0.85)



[OBJECT] Canoe

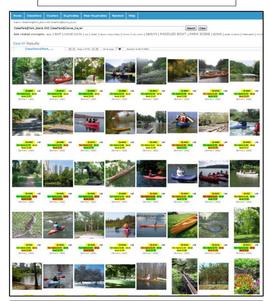


Scene for *canoe* can be a *beach* or *park*, and *canoe* can have *people* or *no people*
→ **not mutually exclusive across facets**

[ACTIVITY] Cricket

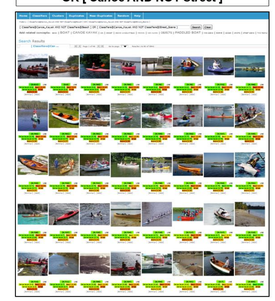
Scene for *cricket* can be a *park*, and *cricket* can have *people* or *no people*
→ **not mutually exclusive across facets**

Park* AND *Canoe

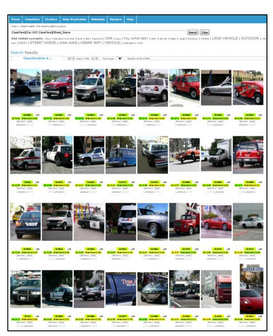


Park* AND *Canoe combines discriminative semantic concepts across facets

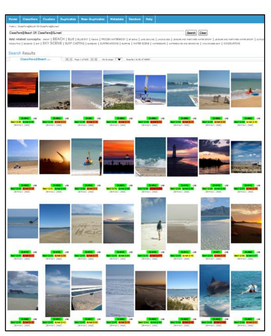
[*Canoe* AND NOT *Beach*] OR [*Canoe* AND NOT *Street*]



[***Canoe*** AND NOT ***Beach***] OR [***Canoe*** AND NOT ***Street***] → approx. ***Park*** AND ***Canoe***

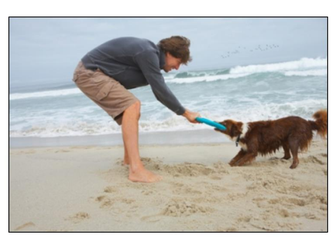


[OBJECT] *Car* AND [SCENE] *Street Scene*



[SCENE] *Beach* OR [SCENE] *Sunset*

What label?



Can multiple pick labels from multiple facets!

Pick one:

	FACETS:
a) Man	[Person]
b) Dog	[Animal]
c) Frisbee	[Object]
d) Beach	[Setting]
e) Playing	[Activity]

Man → Not Woman
Dog → Not Cat
Frisbee → Not Ball
Beach → Not Park
Playing → Not Working