IMAGENET

22K categories and 15M images

- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
- Food
- Materials
- Structures
- Artifact
  - Tools
  - Appliances
  - Structures
- Person
- Scenes
  - Indoor
  - Geological Formations
- Sport Activity

www.image-net.org

Deng et al. 2009,
Russakovsky et al. 2015

Slide credit: Fei-Fei Li

What is WordNet?

Original paper by **[George Miller, et al 1990]** cited over 5,000 times

Organizes over 150,000 words into 117,000 categories called *synsets*.

Establishes ontological and lexical relationships in NLP and related tasks.

Slide credit: Fei-Fei Li and Jia Deng

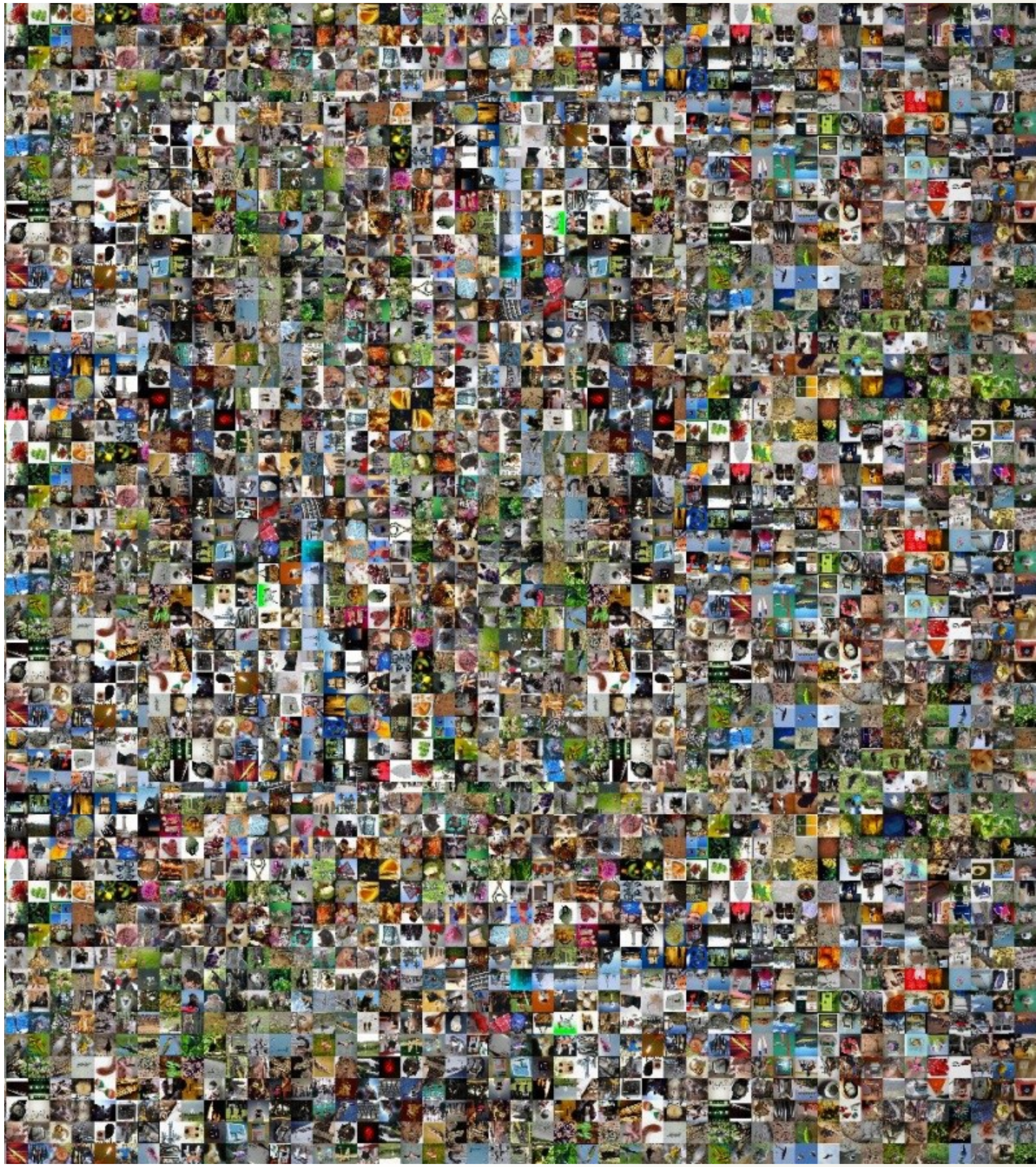
Individually Illustrated WordNet Nodes


jacket: a short coat

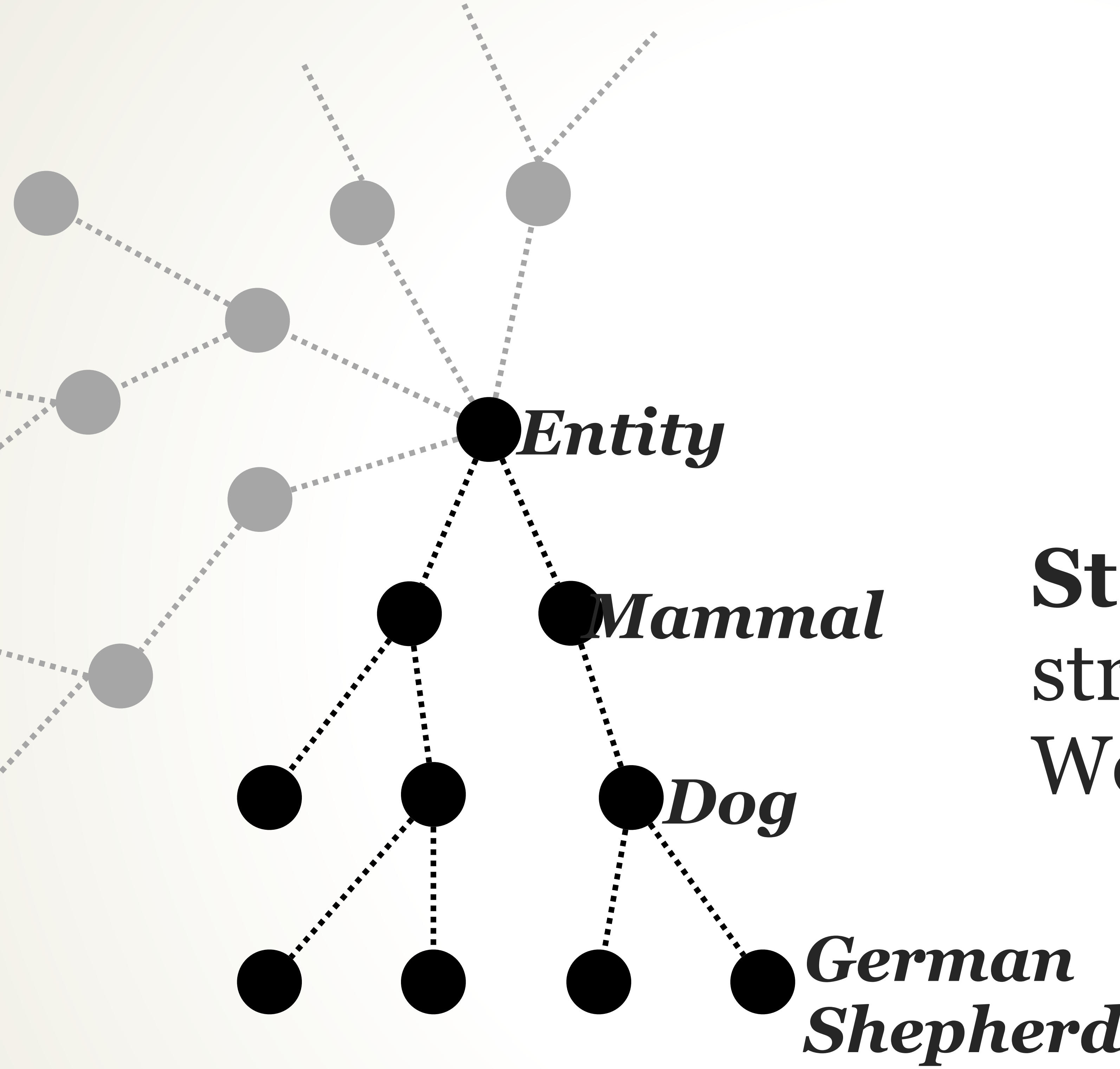

German shepherd: breed of large shepherd dogs used in police work and as a guide for the blind.


microwave: kitchen appliance that cooks food by passing an electromagnetic wave through it.


mountain: a land mass that projects well above its surroundings; higher than a hill.

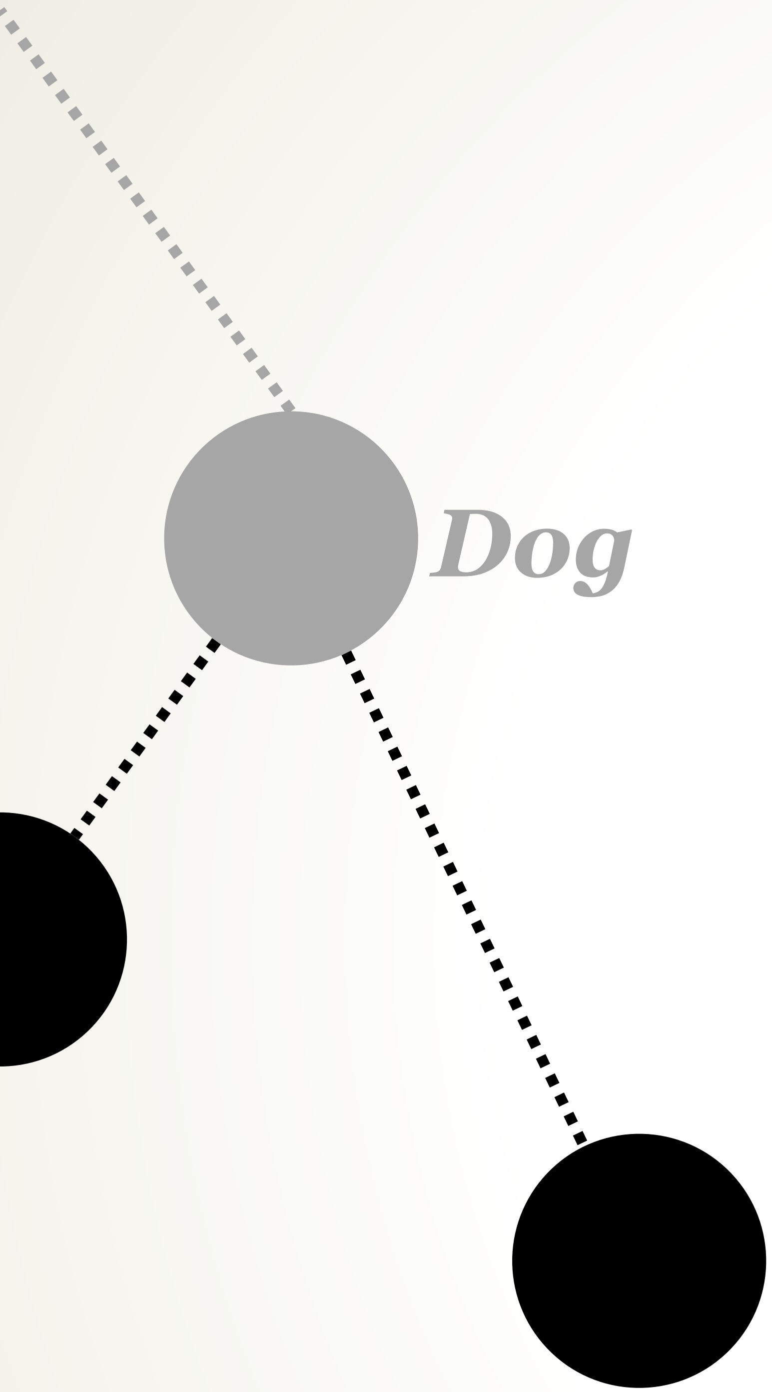


Slide credit: Fei-Fei Li and Jia Deng



Step 1: Ontological structure based on WordNet

Slide credit: Fei-Fei Li and Jia Deng

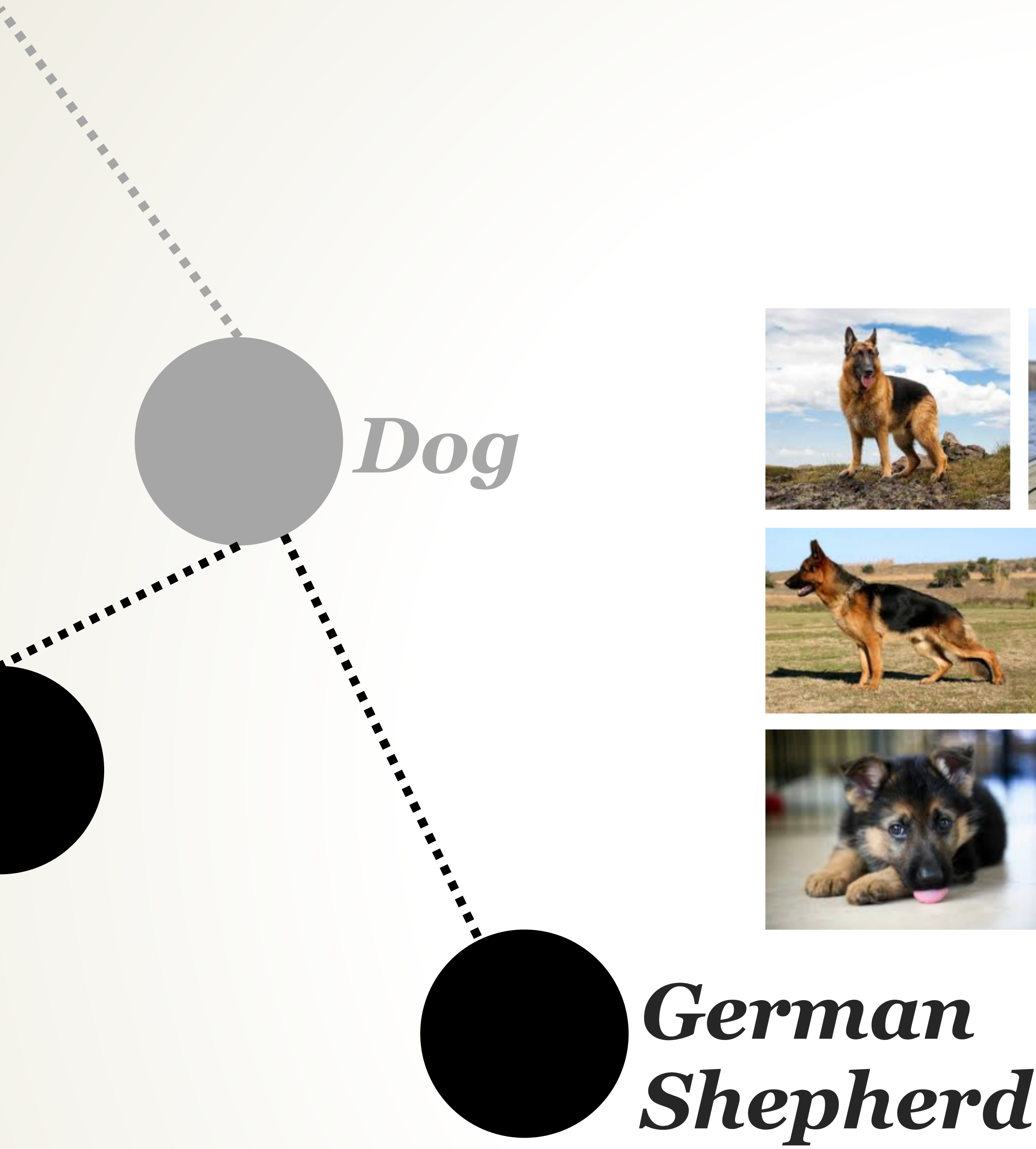


Dog

German Shepherd

Step 2: Populate categories with thousands of images from the Internet

Slide credit: Fei-Fei Li and Jia Deng



Step 3: Clean results by hand

Slide credit: Fei-Fei Li and Jia Deng

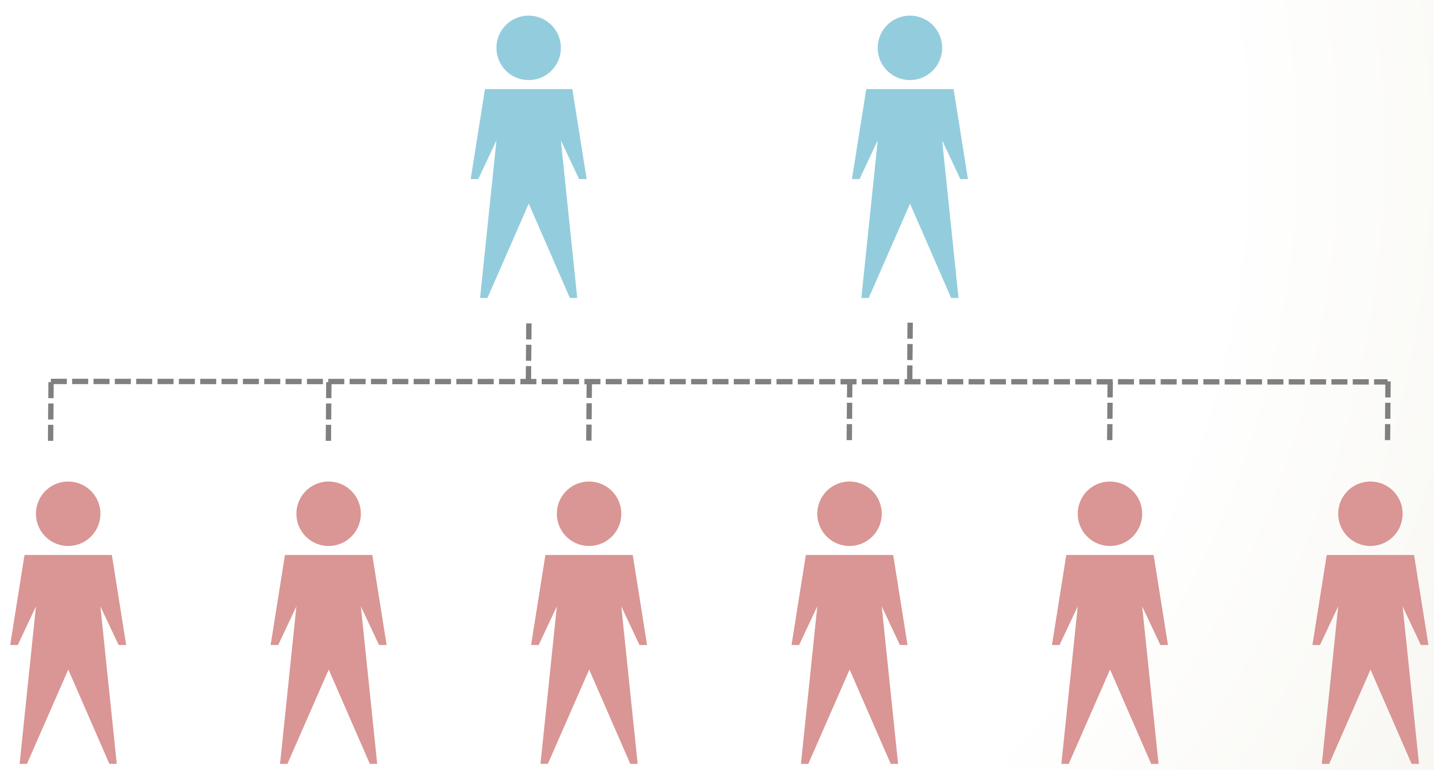
Three Attempts at Launching IMAGENET

Slide credit: Fei-Fei Li and Jia Deng

1st Attempt: The Psychophysics Experiment

ImageNet PhD Students

Miserable Undergrads



Slide credit: Fei-Fei Li and Jia Deng

1^{st} Attempt: The Psychophysics Experiment

- # of synsets: **40,000** (subject to: imageability analysis)
- # of candidate images to label per synset: **10,000**
- # of people needed to verify: **2-5**
- Speed of human labeling: **2 images/sec** (one fixation: ~200msec)
- **Massive parallelism (N ~ 10^2-3)**

40,000 × 10,000 × 3 / 2 = 6000,000,000 sec ≈ 19 years

Slide credit: Fei-Fei Li and Jia Deng

2nd Attempt: Human-in-the-Loop Solutions

Towards scalable dataset construction: An active learning approach

Brendan Collins, Jia Deng, Ka
{bmcollin, dengjia, li, feifei

Department of Computer Science, Princeton

Abstract. As computer vision research co
and greater variation within object categor
more exhaustive datasets are necessary. Ho
ing such datasets is laborious and monoto
in which many images have been automa
category (typically by automatic internet s
relevant images from noise. We present a d
which employs active, online learning to
with minimal user input. The principle adv
vious endeavors is its scalability. We demon
superior to the state-of-the-art, with scala
work.

1 Introduction

Though it is difficult to foresee the future of co
that its trajectory will include examining a gr
(such as objects or scenes), that the complexity
categories will increase, and that these catego
variation. It is unlikely that the researcher's
keep pace with the growing need for annotat
work aims to develop a system which can obta
ages with minimal supervision. The particula

OPTIMOL: automatic Online Picture collecTion via Incremental MOdel Learning

Li-Jia Li[1], Gang Wang[1] and Li Fei-Fei[2]

[1] Dept. of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, USA
[2] Dept. of Computer Science, Princeton University, USA
jiali3@uiuc.edu, gwang6@uiuc.edu, feifeili@cs.princeton.edu

Abstract

A well-built dataset is a necessary starting point for advanced computer vision research. It plays a crucial role in evaluation and provides a continuous challenge to state-of-the-art algorithms. Dataset collection is, however, a tedious and time-consuming task. This paper presents a novel automatic dataset collecting and model learning approach that uses object recognition techniques in an incremental method. The goal of this work is to use the tremendous resources of the web to learn robust object category models in order to detect and search for objects in real-world cluttered scenes. It mimics the human learning process of iteratively accumulating model knowledge and image examples. We adapt a non-parametric graphical model and propose an incremental learning framework. Our algorithm is capable of automatically collecting much larger object category

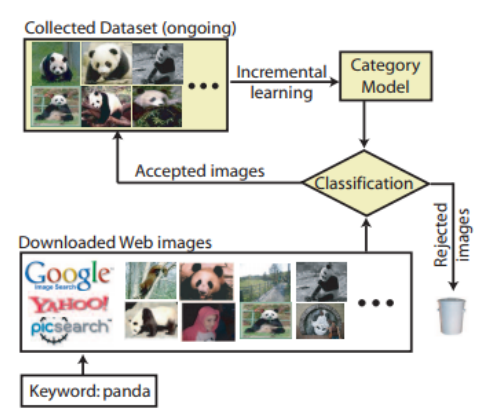


Figure 1. Illustration of the framework of the Online Picture collecTion via Incremental MOdel Learning (OPTIMOL) system. This framework works in an incremental way: Once a model is

Slide credit: Fei-Fei Li and Jia Deng

2nd Attempt: Human-in-the-Loop Solutions

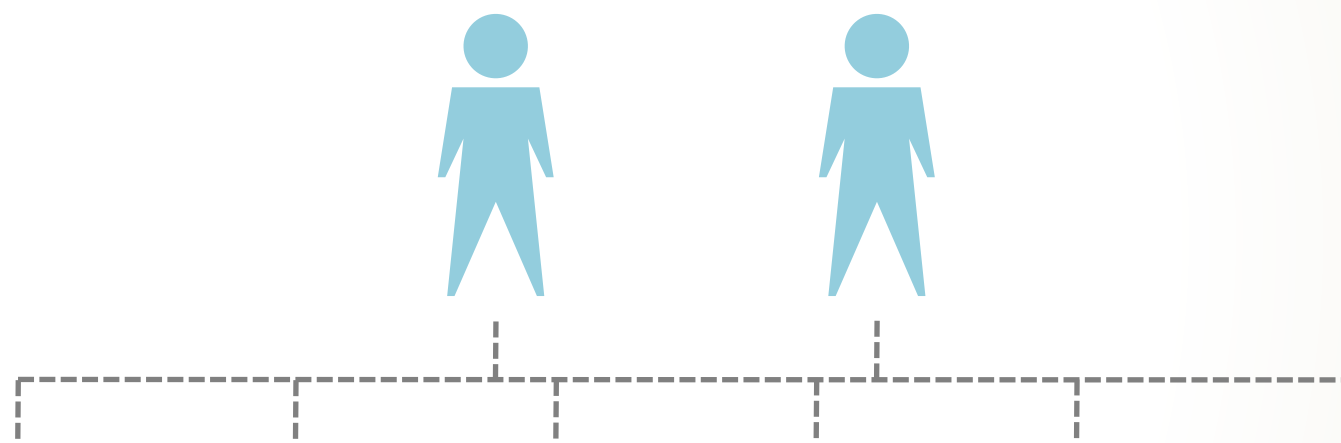
Machine-generated datasets can only match the best algorithms of the time.

Human-generated datasets transcend algorithmic limitations, leading to better machine perception.

Slide credit: Fei-Fei Li and Jia Deng

3^{rd} Attempt: Crowdsourcing

ImageNet PhD Students



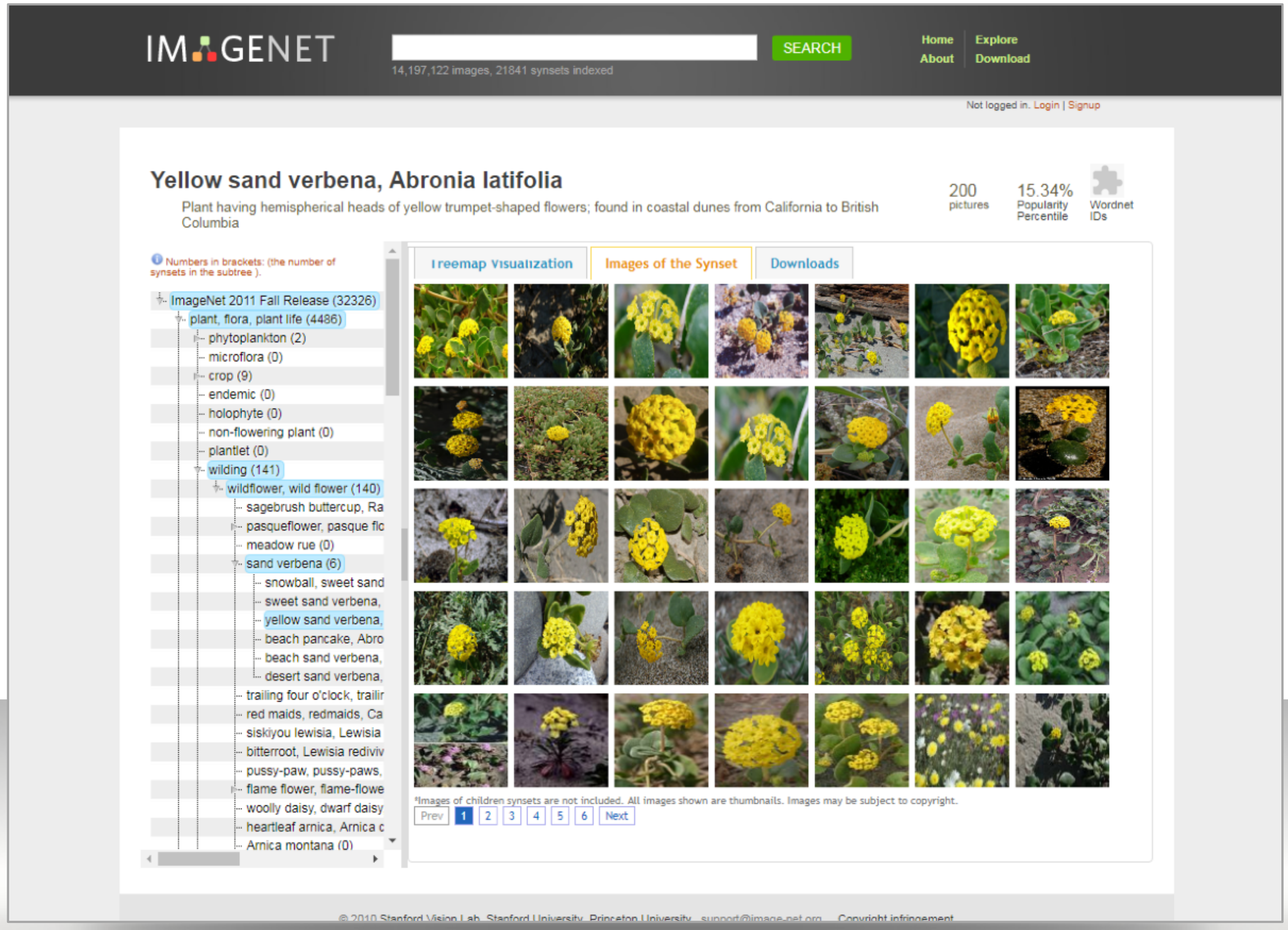
Crowdsourced Labor

amazon mechanical turk™
Artificial Artificial Intelligence

49k Workers *from* 167 Countries 2007-2010

Slide credit: Fei-Fei Li and Jia Deng

The Result: IMAGENET Goes Live in 2009



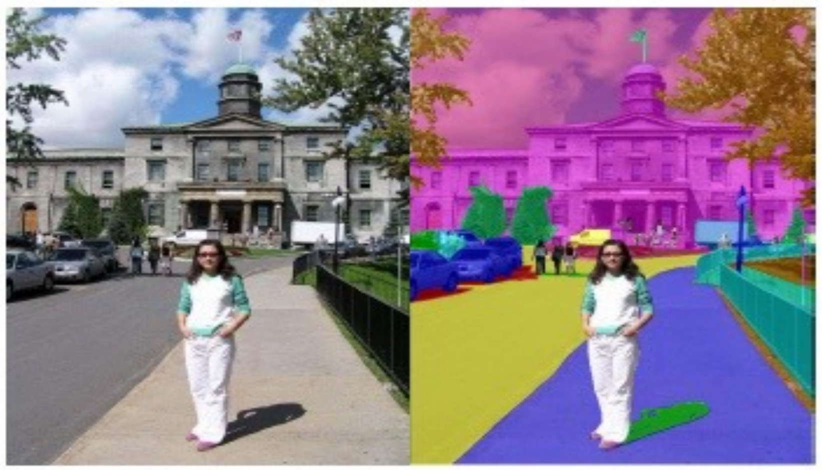
Slide credit: Fei-Fei Li and Jia Deng

Others Targeted Detail

LabelMe

Per-Object Regions and Labels

Russell et al, 2005



Lotus Hill

Hand-Traced Parse Trees

Yao et al, 2007

Slide credit: Fei-Fei Li and Jia Deng

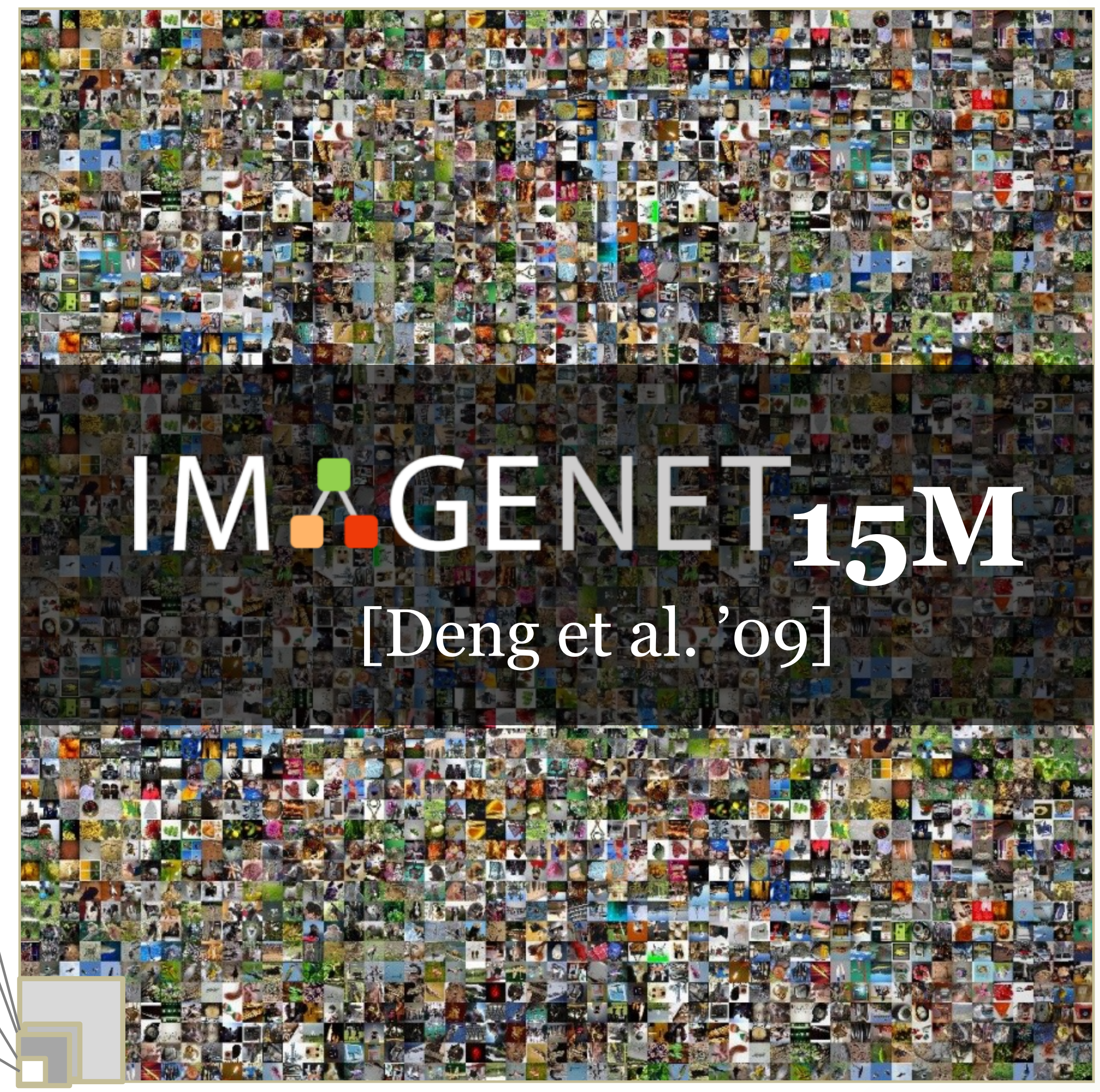
ImageNet Targeted Scale

SUN, 131K
[Xiao et al. ‘10]

LabelMe, 37K
[Russell et al. ’07]

PASCAL VOC, 30K
[Everingham et al. ’06-’12]

Caltech101, 9K
[Fei-Fei, Fergus, Perona, ‘03]



Slide credit: Fei-Fei Li and Jia Deng

Challenge procedure every year

1. Training data released: images and annotations
   - For classification, 1000 synsets with ~1k images/synset
2. Test data released: images only (annotations hidden)
   - For classification, ~ 100 images/synset
3. Participants train their models on train data
4. Submit text file with predictions on test images
5. Evaluate and release results, and run a workshop at ECCV/ICCV to discuss results

ILSVRC image classification task

Steel drum

Objects:	1000 classes
Training:	1.2M images
Validation:	50K images
Test:	100K images

ILSVRC image classification task

Steel drum

Output:
Scale
T-shirt
<u>Steel drum</u>
Drumstick
Mud turtle

✔

Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle

✗

ILSVRC image classification task

Steel drum

Output:
Scale
T-shirt
<u>Steel drum</u>
Drumstick
Mud turtle

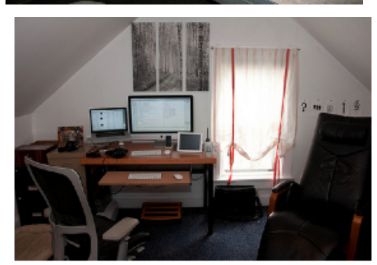
✔

Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle

✗

$$\text{Error} = \frac{1}{100{,}000} \sum_{\substack{100{,}000 \\ \text{images}}} 1[\text{incorrect on image i}]$$

Why not all objects?

Input \ Labels	Table	Chair	Bowl	Dog	Cat	…
	+	+	-	-	-	-
	+	-	+	-	+	-
	+	+	-	-	-	-
	-	-	-	+	-	-

(1000 objects)

100 million questions

(100K <u>test</u> images)

ILSVRC Task 2: Classification + Localization

Steel drum

Objects:	1000 classes	
Training:	1.2M images,	500K with bounding boxes
Validation:	50K images,	all 50K with bounding boxes
Test:	100K images,	all 100K with bounding boxes

Data annotation cost

Draw a tight bounding box around the moped

Data annotation cost

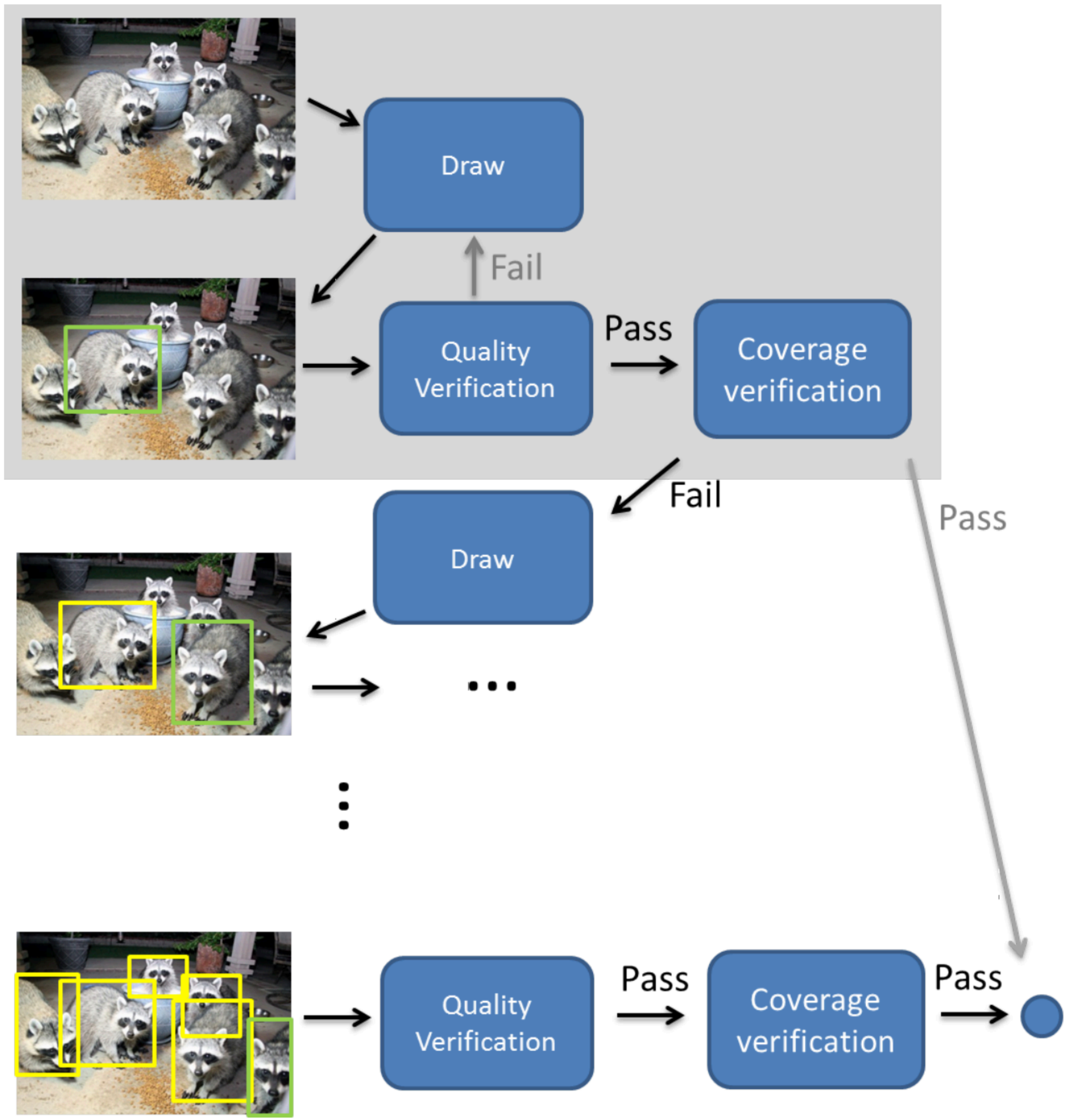
Draw a tight bounding box around the moped

Data annotation cost

Draw a tight bounding box around the moped

This took **14.5 seconds**

(**7 sec** [JaiGra ICCV'13],
10.2 sec [RusLiFei CVPR'15],
25.5 sec [SuDenFei AAAIW'12])

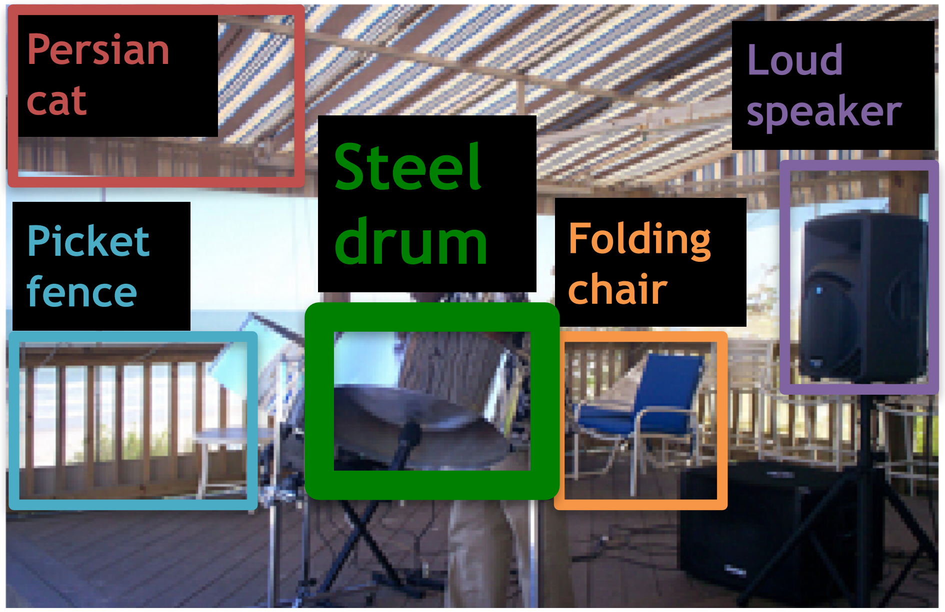


[Hao Su et al. AAAI 2010]

ILSVRC Task 2: Classification + Localization

Steel drum

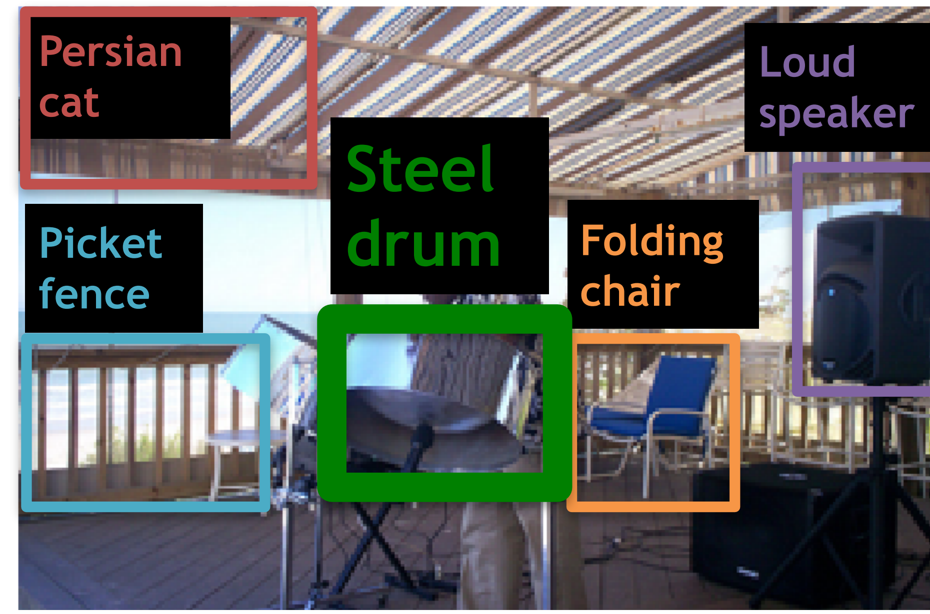
Output



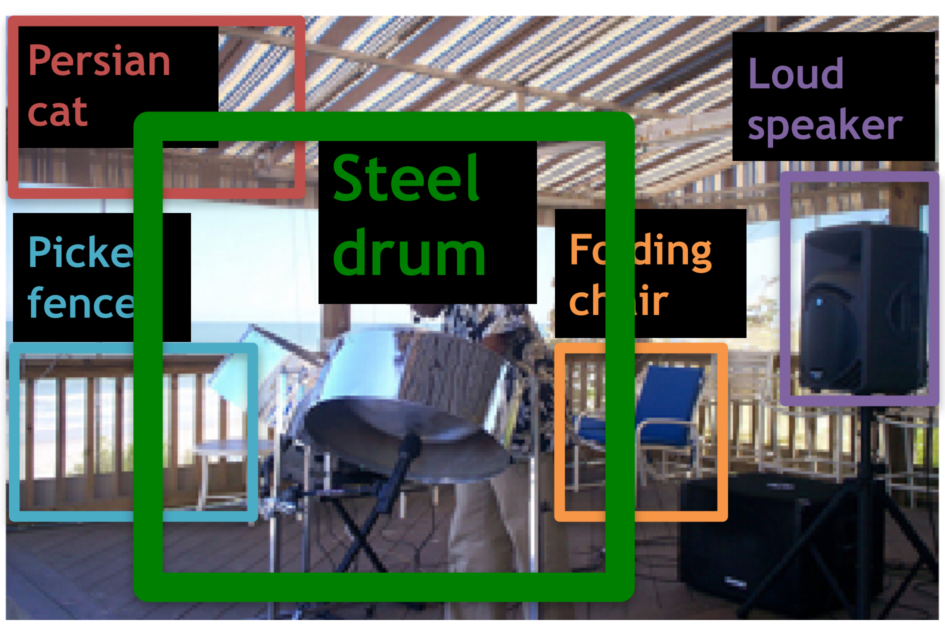
ILSVRC Task 2: Classification + Localization

Steel drum

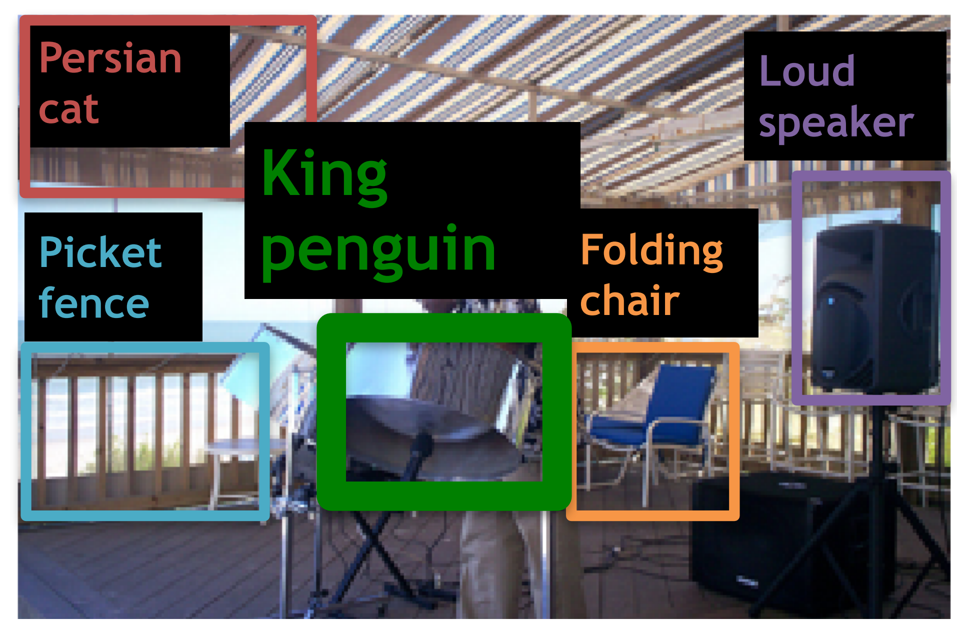
Output



Output (bad localization)



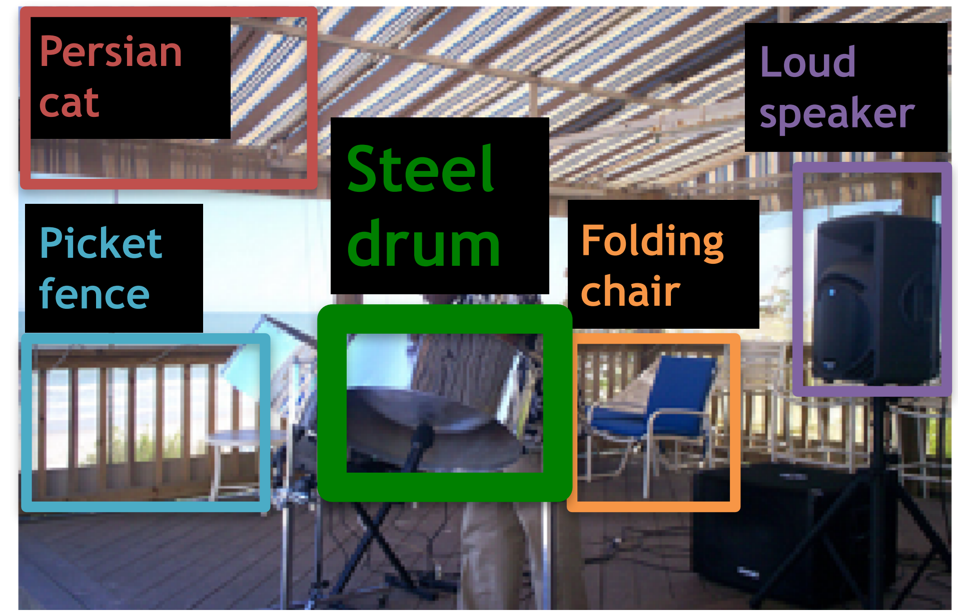
Output (bad classification)



ILSVRC Task 2: Classification + Localization

Steel drum

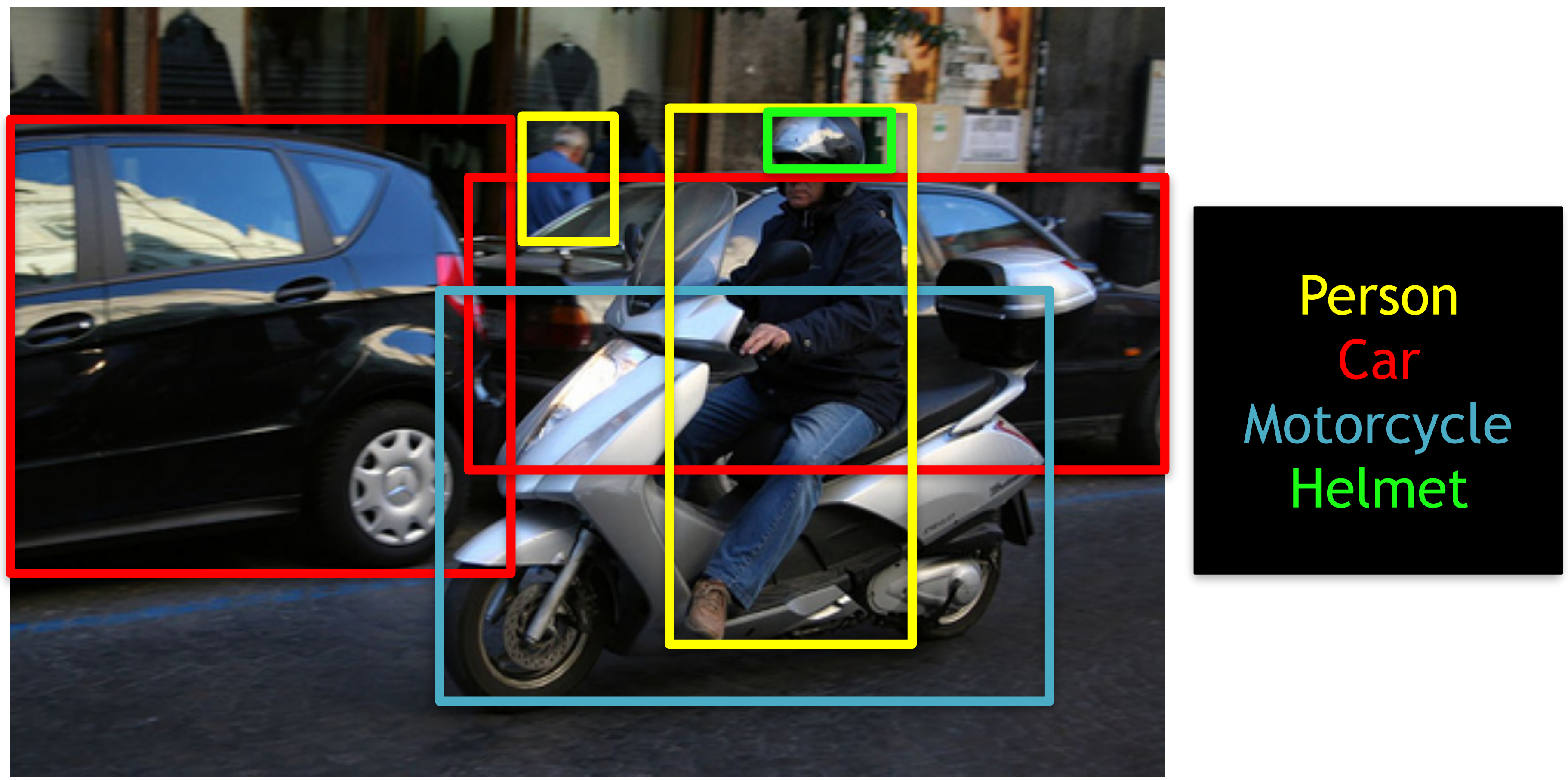
Output



$$\text{Error} = \frac{1}{100{,}000} \sum_{\substack{100{,}000 \\ \text{images}}} 1[\text{incorrect on image } i]$$

ILSVRC Task 3: Detection

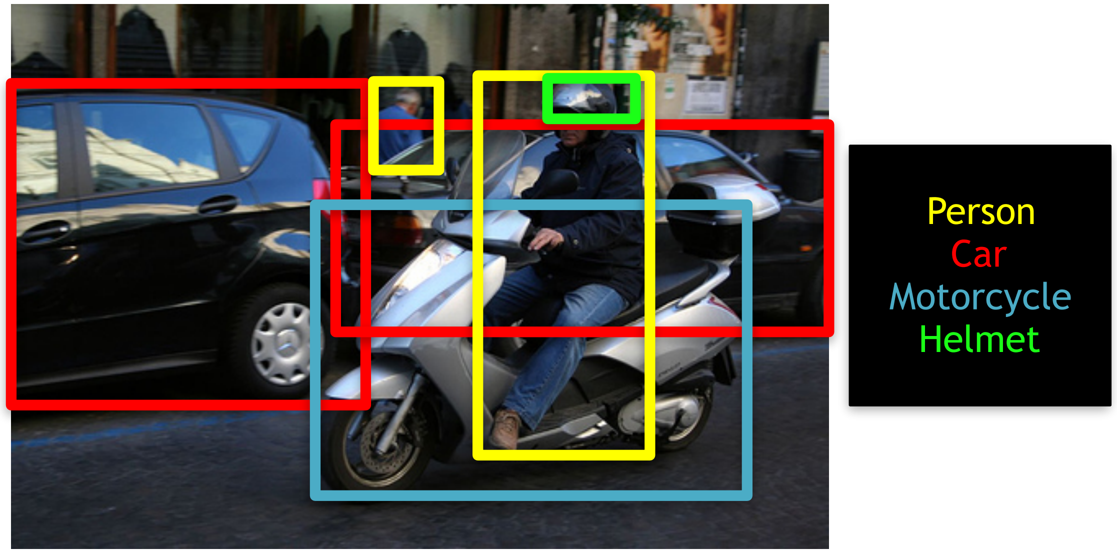
Allows evaluation of generic object detection
in cluttered scenes at scale



Objects:	200 classes
Training:	450K images, 470K bounding boxes
Validation:	20K images, all bounding boxes
Test:	40K images, all bounding boxes

ILSVRC Task 3: Detection

All instances of all target object classes expected to be localized on all test images

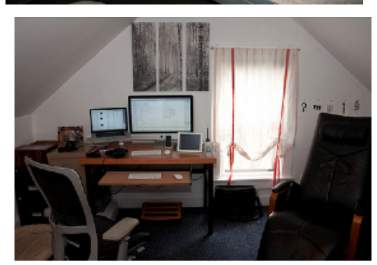


Evaluation modeled after PASCAL VOC:

- Algorithm outputs a list of bounding box detections with confidences
- A detection is considered correct if overlap with ground truth is big enough
- Evaluated by average precision per object class
- Winners of challenge is the team that wins the most object categories

Everingham, Van Gool, Williams, Winn and Zisserman. The PASCAL Visual Object Classes (VOC) Challenge. IJCV 2010.

Multi-label annotation

Input \ Labels	Table	Chair	Bowl	Dog	Cat	…
	+	+	-	-	-	-
	+	-	+	-	+	-
	+	+	-	-	-	-
	-	-	-	+	-	-

(200 objects)

(120K images)

24 million questions

[Deng et al. CHI'14]

Man-made objects

Furniture

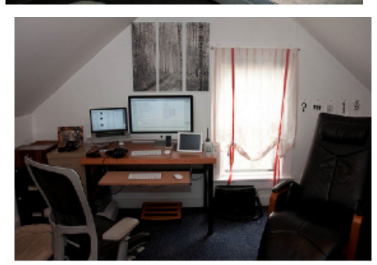
Animals

Semantic hierarchy

Labels

Input

(200 objects)

Input	Table	Chair	Bowl	Dog	Cat	…
	+	+	-	-	-	-
	+	-	+	-	+	-
	+	+	-	-	-	-
	-	-	-	+	-	-

(120K images)

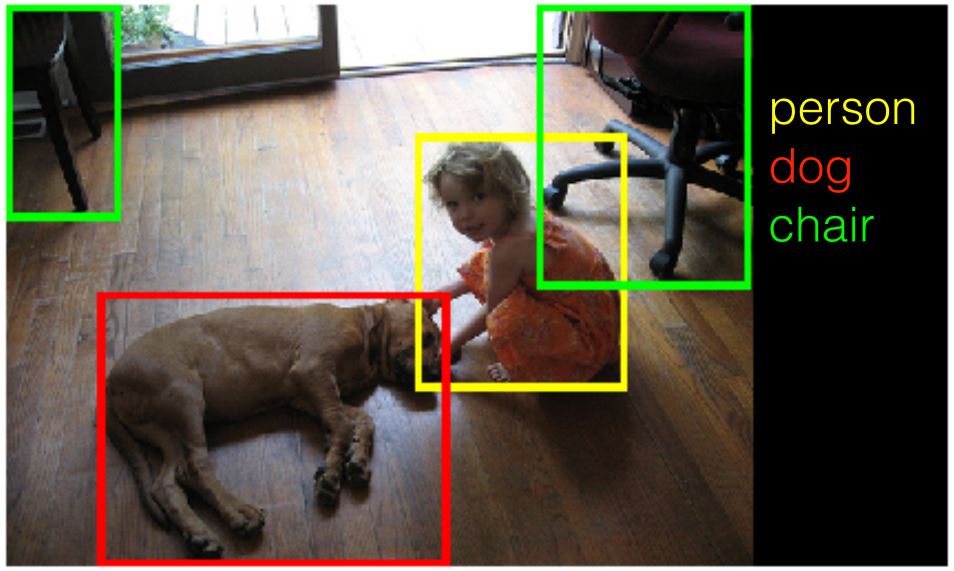
[Deng et al. CHI'14]

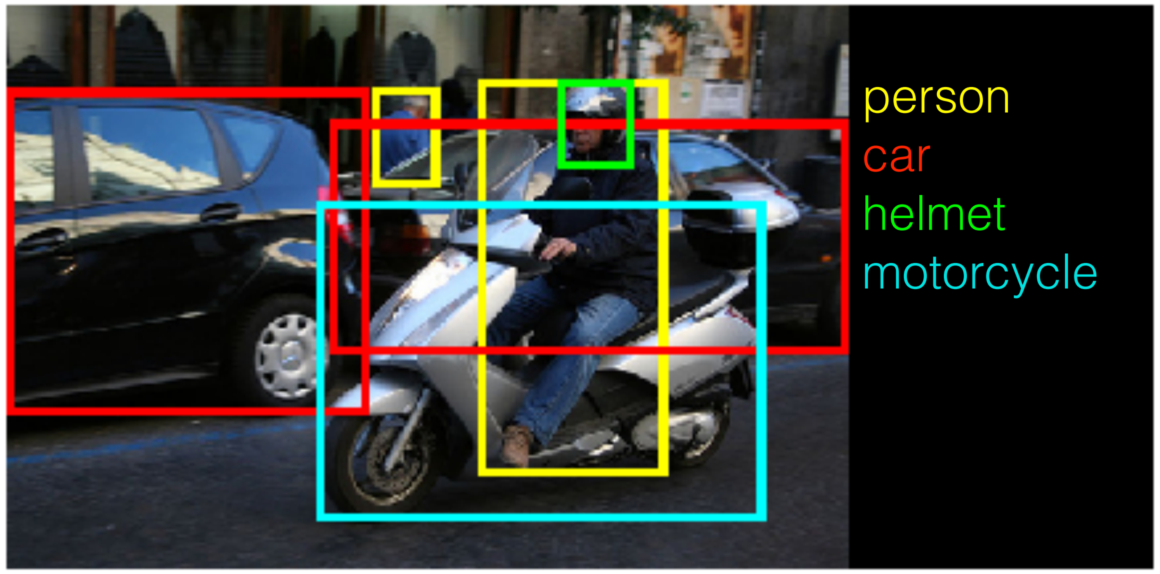
ImageNet object detection challenge

120,931 images 200 object classes

Compare to PASCAL VOC [EveVanWilWinZis '12]

22,591 images 20 object classes





Result:

6.2x savings in human cost

Large-scale **object detection** benchmark

[Deng et al. CHI'14] [Russakovsky et al. IJCV'15]

Annotation research

In-house annotation: Caltech 101, PASCAL
[FeiFerPer CVPR’04, EveVanWilWinZis IJCV’10]

Annotation research

In-house annotation: Caltech 101, PASCAL
[FeiFerPer CVPR'04, EveVanWilWinZis IJCV'10]

Decentralized annotation: LabelMe, SUN
[RusTorMurFre IJCV'07, XiaHayEhiOliTor CVPR'10]

AMT annotation: quality control; ImageNet
[SorFor CVPR'08, DenDonSocLiLiFei CVPR'09]

Annotation research

In-house annotation: Caltech 101, PASCAL
[FeiFerPer CVPR’04, EveVanWilWinZis IJCV’10]

Decentralized annotation: LabelMe, SUN
[RusTorMurFre IJCV’07, XiaHayEhiOliTor CVPR’10]

AMT annotation: quality control; ImageNet
[SorFor CVPR’08, DenDonSocLiLiFei CVPR’09]

Probabilistic models of annotators [WelBraBelPer NIPS’10]

Annotation research

In-house annotation: Caltech 101, PASCAL
[FeiFerPer CVPR'04, EveVanWilWinZis IJCV'10]

Decentralized annotation: LabelMe, SUN
[RusTorMurFre IJCV'07, XiaHayEhiOliTor CVPR'10]

AMT annotation: quality control; ImageNet
[SorFor CVPR'08, DenDonSocLiLiFei CVPR'09]

Probabilistic models of annotators [WelBraBelPer NIPS'10]

Iterative bounding box annotation [SuDenFei AAAIW'10]

Reconciling segmentations [VitHay BMVC'11]

Building an attribute vocabulary [ParGra CVPR'11]

Efficient video annotation: VATIC [VonPatRam IJCV12]

Annotation research

Computer vision community

In-house annotation: Caltech 101, PASCAL
[FeiFerPer CVPR'04, EveVanWilWinZis IJCV'10]

Decentralized annotation: LabelMe, SUN
[RusTorMurFre IJCV'07, XiaHayEhiOliTor CVPR'10]

AMT annotation: quality control; ImageNet
[SorFor CVPR'08, DenDonSocLiLiFei CVPR'09]

Probabilistic models of annotators [WelBraBelPer NIPS'10]

Iterative bounding box annotation [SuDenFei AAAIW'10]

Reconciling segmentations [VitHay BMVC'11]

Building an attribute vocabulary [ParGra CVPR'11]

Efficient video annotation: VATIC [VonPatRam IJCV12]

Annotation research

Computer vision community

In-house annotation: Caltech 101, PASCAL
[FeiFerPer CVPR'04, EveVanWilWinZis IJCV'10]

Decentralized annotation: LabelMe, SUN
[RusTorMurFre IJCV'07, XiaHayEhiOliTor CVPR'10]

AMT annotation: quality control; ImageNet
[SorFor CVPR'08, DenDonSocLiLiFei CVPR'09]

Probabilistic models of annotators [WelBraBelPer NIPS'10]

Iterative bounding box annotation [SuDenFei AAAIW'10]

Reconciling segmentations [VitHay BMVC'11]

Building an attribute vocabulary [ParGra CVPR'11]

Efficient video annotation: VATIC [VonPatRam IJCV12]

HCI community

ESP Game, Peekaboom: gamification of image labeling [AhnDab CHI'04, AhnLiuBlu CHI'06]

Estimating quality of crowd workers
[SheProIpe KDD'08]

Iterative workflow for handwriting recognition [DaiMauWel AAAI'10]

Clowder: optimizing/personalizing workflows [WelMauDai AAAI'11]

GalazyZoo: predictive models for consensus tasks [KamHacHor AAMAS'12]

Crowdsourcing taxonomy creation
[ChiLitEdgWelLan CHI'13, BraMauWel HCOMP'13]

Annotation research

Computer vision community

In-house annotation: Caltech 101, PASCAL
[FeiFerPer CVPR'04, EveVanWilWinZis IJCV'10]

Decentralized annotation: LabelMe, SUN
[RusTorMurFre IJCV'07, XiaHayEhiOliTor CVPR'10]

AMT annotation: quality control; ImageNet
[SorFor CVPR'08, DenDonSocLiLiFei CVPR'09]

Probabilistic models of annotators [WelBraBelPer NIPS'10]

Iterative bounding box annotation [SuDenFei AAAIW'10]

Reconciling segmentations [VitHay BMVC'11]

Building an attribute vocabulary [ParGra CVPR'11]

Efficient video annotation: VATIC [VonPatRam IJCV12]

HCI community

ESP Game, Peekaboom: gamification of image labeling [AhnDab CHI'04, AhnLiuBlu CHI'06]

Estimating quality of crowd workers
[SheProIpe KDD'08]

Iterative workflow for handwriting recognition [DaiMauWel AAAI'10]

Clowder: optimizing/personalizing workflows [WelMauDai AAAI'11]

GalazyZoo: predictive models for consensus tasks [KamHacHor AAMAS'12]

Crowdsourcing taxonomy creation
[ChiLitEdgWelLan CHI'13, BraMauWel HCOMP'13]

Annotation research

Computer vision community

In-house annotation: Caltech 101, PASCAL
[FeiFerPer CVPR'04, EveVanWilWinZis IJCV'10]

Decentralized annotation: LabelMe, SUN
[RusTorMurFre IJCV'07, XiaHayEhiOliTor CVPR'10]

AMT annotation: quality control; ImageNet
[SorFor CVPR'08, DenDonSocLiLiFei CVPR'09]

Probabilistic models of annotators [WelBraBelPer NIPS'10]

Iterative bounding box annotation [SuDenFei AAAIW'10]

Reconciling

Building an

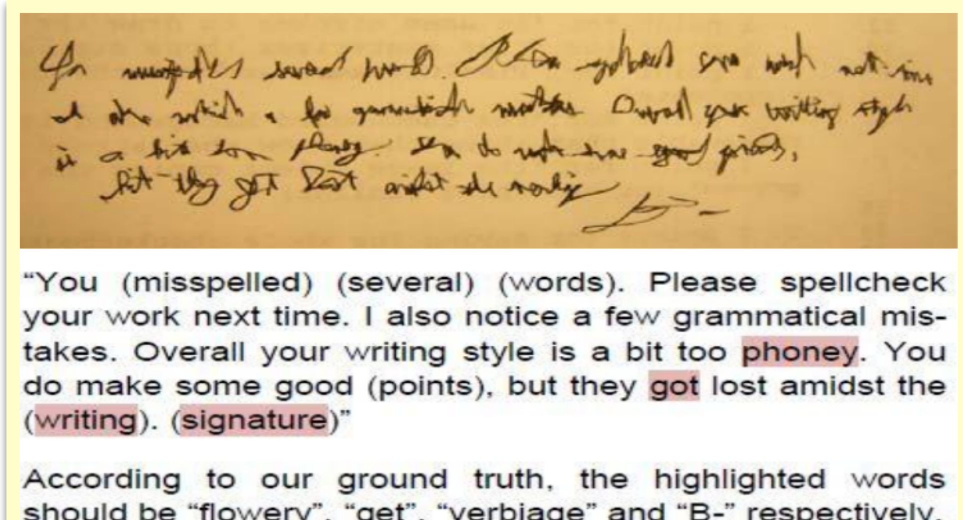
Efficient vid

HCI community

ESP Game, Peekaboom: gamification of image labeling [AhnDab CHI'04, AhnLiuBlu CHI'06]

Estimating quality of crowd workers
[SheProIpe KDD'08]

Iterative workflow for handwriting recognition [DaiMauWel AAAI'10]



Annotation research

Computer vision community

In-house annotation: Caltech 101, PASCAL
[FeiFerPer CVPR'04, EveVanWilWinZis IJCV'10]

Decentralized annotation: LabelMe, SUN
[RusTorMurFre IJCV'07, XiaHayEhiOliTor CVPR'10]

AMT annotation: quality control; ImageNet
[SorFor CVPR'08, DenDonSocLiLiFei CVPR'09]

Probabilistic models of annotators [WelBraBelPer NIPS'10]

Iterative bounding box annotation [SuDenFei AAAIW'10]

Reconciling segmentations [VitHay BMVC'11]

Building an attribute vocabulary [ParGra CVPR'11]

Efficient video annotation: VATIC [VonPatRam IJCV12]

HCI community

ESP Game, Peekaboom: gamification of image labeling [AhnDab CHI'04, AhnLiuBlu CHI'06]

Estimating quality of crowd workers
[SheProIpe KDD'08]

Iterative workflow for handwriting recognition [DaiMauWel AAAI'10]

Clowder: optimizing/personalizing workflows [WelMauDai AAAI'11]

GalazyZoo: predictive models for consensus tasks [KamHacHor AAMAS'12]

Crowdsourcing taxonomy creation
[ChiLitEdgWelLan CHI'13, BraMauWel HCOMP'13]

Sharing of insights

Computer vision community

In-house annotation: Caltech 101, PASCAL
[FeiFerPer CVPR'04, EveVanWilWinZis IJCV'10]

Decentralized annotation: LabelMe, SUN
[RusTorMurFre IJCV'07, XiaHayEhiOliTor CVPR'10]

AMT annotation: quality control; ImageNet
[SorFor CVPR'08, DenDonSocLiLiFei CVPR'09]

Probabilistic models of annotators [WelBraBelPer NIPS'10]

Iterative bounding box annotation [SuDenFei AAAIW'10]

Reconciling segmentations [VitHay BMVC'11]

Building an attribute vocabulary [ParGra CVPR'11]

Efficient video annotation: VATIC [VonPatRam IJCV12]

HCI community

ESP Game, Peekaboom: gamification of image labeling [AhnDab CHI'04, AhnLiuBlu CHI'06]

Estimating quality of crowd workers
[SheProIpe KDD'08]

Iterative workflow for handwriting recognition [DaiMauWel AAAI'10]

Clowder: optimizing/personalizing workflows [WelMauDai AAAI'11]

GalazyZoo: predictive models for consensus tasks [KamHacHor AAMAS'12]

Crowdsourcing taxonomy creation
[ChiLitEdgWelLan CHI'13, BraMauWel HCOMP'13]

Sharing of insights

Computer vision community

In-house annotation: Caltech 101, PASCAL
[FeiFerPer CVPR'04, EveVanWilWinZis IJCV'10]

Decentralized annotation: LabelMe, SUN
[RusTorMurFre IJCV'07, XiaHayEhiOliTor CVPR'10]

AMT annotation: quality control; ImageNet
[SorFor CVPR'08, DenDonSocLiLiFei CVPR'09]

Probabilistic models of annotators [WelBraBelPer NIPS'10]

Iterative bounding box annotation [SuDenFei AAAIW'10]

Reconciling segmentations [VitHay BMVC'11]

Building an attribute vocabulary [ParGra CVPR'11]

Efficient video annotation: VATIC [VonPatRam IJCV12]

HCI community

ESP Game, Peekaboom: gamification of image labeling [AhnDab CHI'04, AhnLiuBlu CHI'06]

Estimating quality of crowd workers
[SheProIpe KDD'08]

Iterative workflow for handwriting recognition [DaiMauWel AAAI'10]

Clowder: optimizing/personalizing workflows [WelMauDai AAAI'11]

GalazyZoo: predictive models for consensus tasks [KamHacHor AAMAS'12]

Crowdsourcing taxonomy creation
[ChiLitEdgWelLan CHI'13, BraMauWel HCOMP'13]

Scalable multi-label annotation

[RusDenSuKraSatEtal IJCV'15]

[DenRusKraBerBerFei CHI'14]