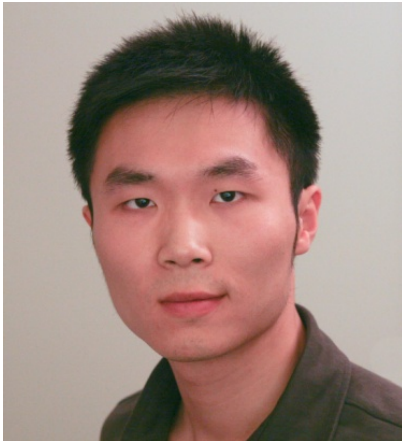




Large Scale Visual Recognition Challenge 2012



Jia Deng

Stanford

Alex Berg

Stony Brook

Sanjeev Satheesh

Stanford

Hao Su

Stanford

Aditya Khosla

MIT

Fei-Fei Li

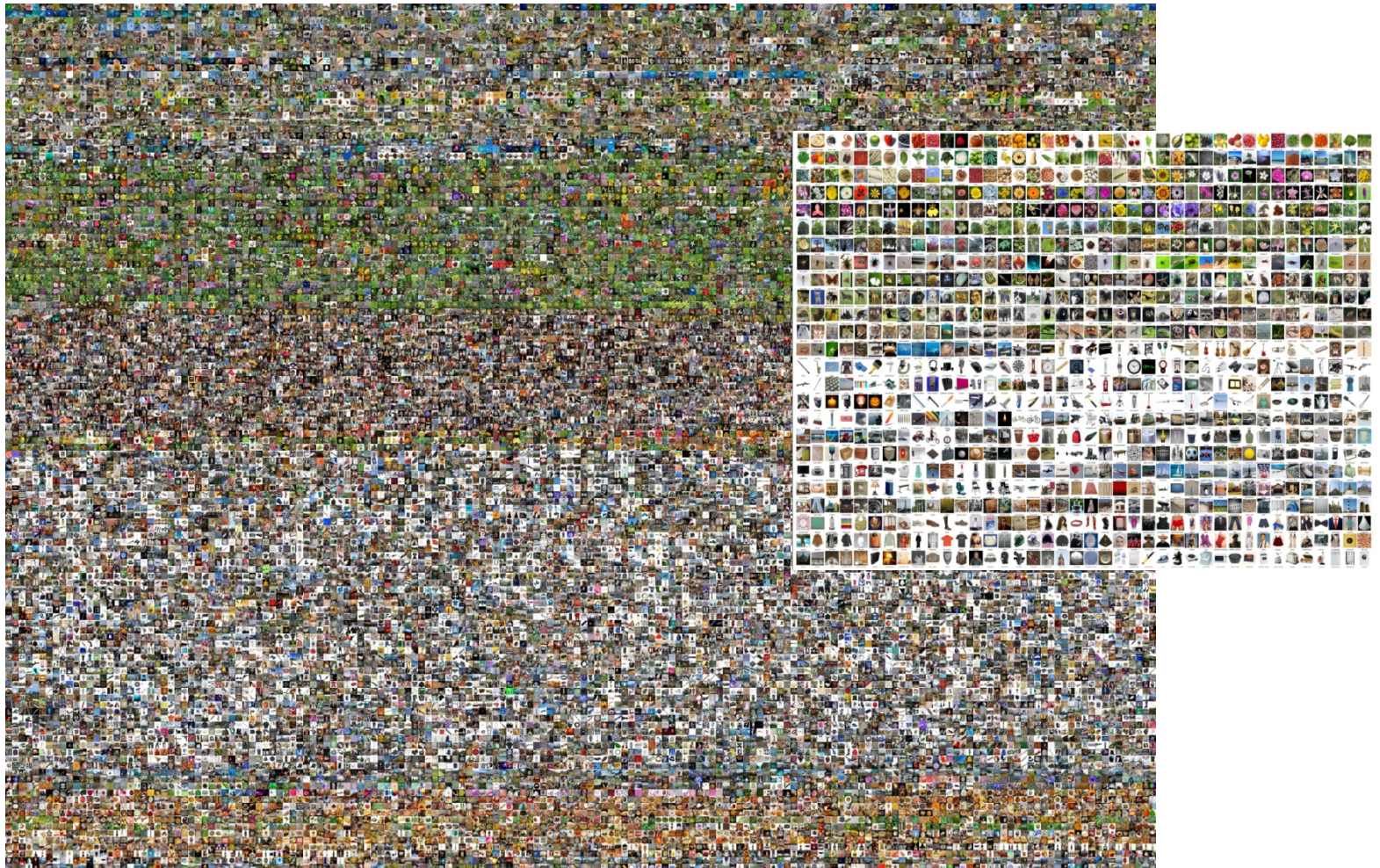
Stanford

agenda

- 15:30 - 16:00. Introduction and overview of results. Fei-Fei Li.
- 16:00 - 16:25. Invited talk. OXFORD_VGG team
- 16:25 - 16:40. Break
- 16:40 - 17:05. Invited Talk. ISI team
- 17:05 - 17:30. Invited Talk. SuperVision team
- 17:30 - 18:00. Discussion.

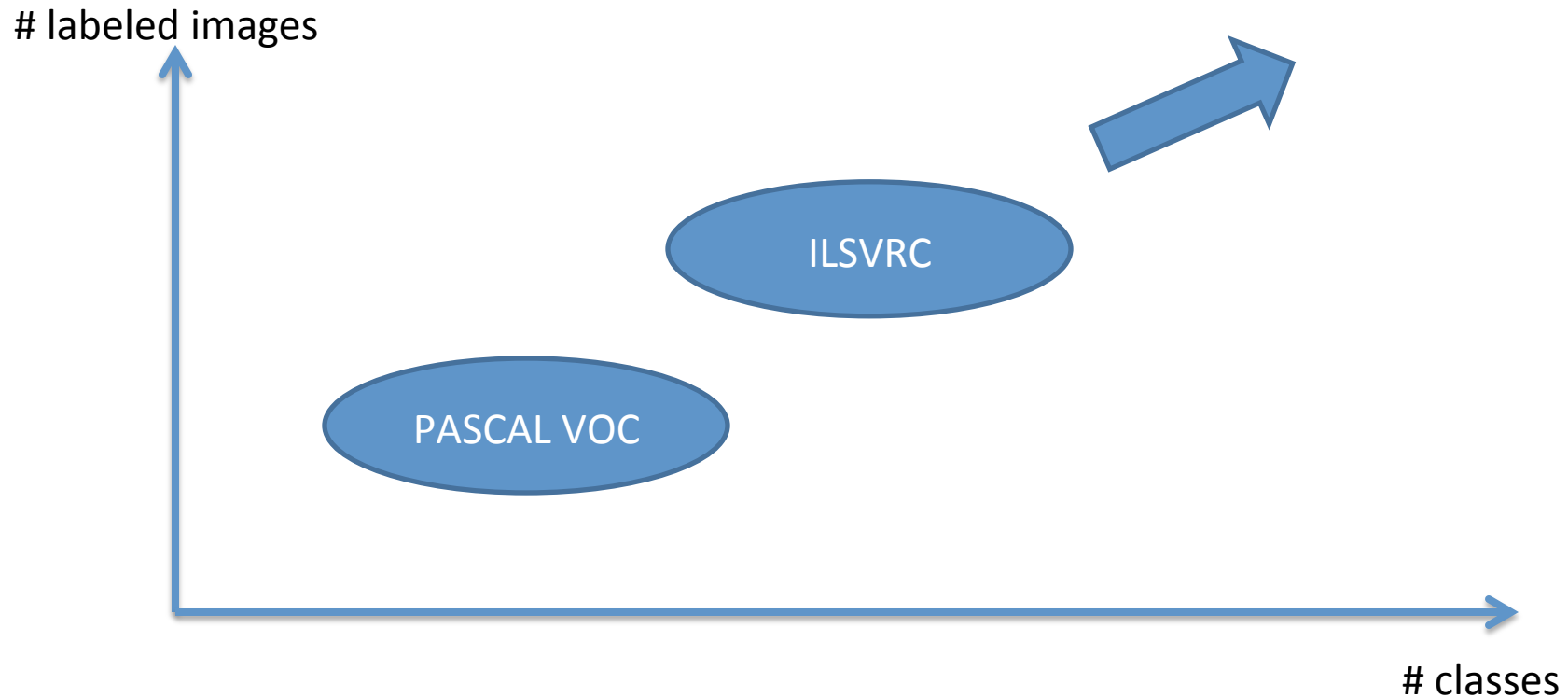
Holy Grail of Recognition

- Recognizing everything in the visual world



Why Large Scale

- Two trends in recognition
 - Use of massive, labeled datasets
 - A large number of fine-grained classes



Goals of ImageNet Challenges

- Foster research on large scale recognition
 - Complement PASCAL VOC
 - Provide a benchmark dataset
 - Let promising techniques emerge via competition

ImageNet Challenge 2012

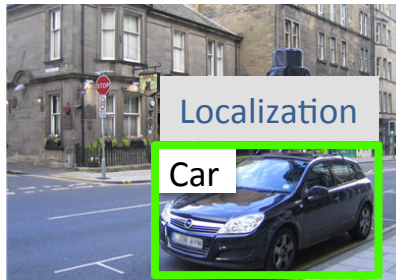
Task 1: Classification



Car

- Predict a class label
- 5 predictions / image
- 1000 classes
- 1,200 images per class for training
- Bounding boxes for 50% of training.

Task 2: Detection (Classification + Localization)



classification

Car

- Predict a class label and a bounding box
- 5 predictions / image
- 1000 classes
- 1,200 images per class for training
- Bounding boxes for 40% of training.

Task 3: Fine-grained classification



classification

Walker hound

- Predict a class label given a bounding box in test
- 1 prediction / image
- 120 dog classes (subset)
- ~200 images per class for training (subset)
- Bounding boxes for 100% of training

What's New in 2012

	ILSVRC 2010	ILSVRC 2011	ILSVRC 2012
Classification task	Y	Y	Y
Localization task	N	Y	Y
Fine-grained task	N	N	Y
# training images with bounding boxes	0%	25%	42%
# test images with bounding boxes	0%	100%	100%

Source for categories and training data



- ImageNet
 - 14,192,122 images, 21841 categories
 - Image found via web searches for WordNet noun synsets
 - Hand verified using Mechanical Turk
 - Bounding boxes for query object labeled
 - New data for validation and testing each year
- WordNet
 - Source of the labels
 - Semantic hierarchy
 - Contains large fraction of English nouns
 - Also used to collect other datasets like tiny images (Torralba et al)
 - Note that categorization is not the end/only goal, so idiosyncrasies of WordNet may be less critical

<http://www.image-net.org>

IMAGENET

14,197,122 images, 21841 synsets indexed

[Explore](#) ^{New!} [Download](#) ^{New!} [Challenge](#) [People](#) [Publication](#)
[About](#)

Not logged in. [Login](#) | [Signup](#)

ImageNet is an image database organized according to the **WordNet** hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures.

[Click here](#) to learn more about ImageNet, [Click here](#) to join the ImageNet mailing list.

SEARCH

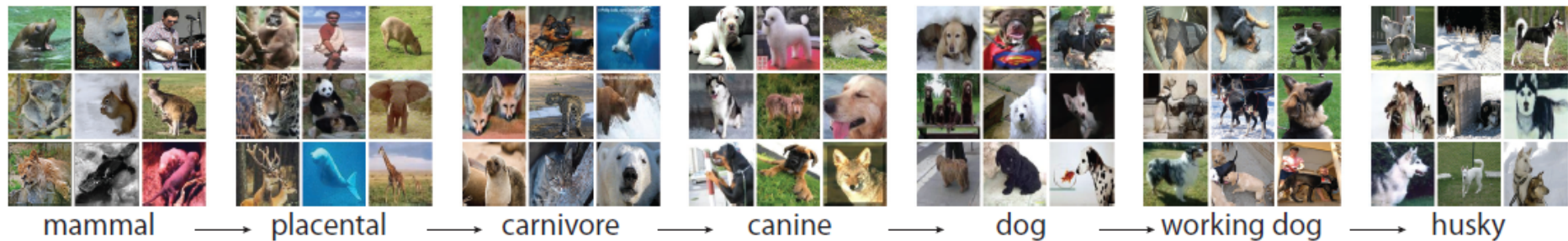


What do these images have in common? *Find out!*

Imagenet 2011 Fall Release is online now!

IMAGENET is a knowledge ontology

- Taxonomy



- [S: \(n\) Eskimo dog, husky](#) (breed of heavy-coated Arctic sled dog)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - [S: \(n\) working dog](#) (any of several breeds of usually large powerful dogs bred to work as draft animals and guard and guide dogs)
 - [S: \(n\) dog, domestic dog, Canis familiaris](#) (a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds) "*the dog barked all night*"
 - [S: \(n\) canine, canid](#) (any of various fissiped mammals with nonretractile claws and typically long muzzles)
 - [S: \(n\) carnivore](#) (a terrestrial or aquatic flesh-eating mammal) "*terrestrial carnivores have four or five clawed digits on each limb*"
 - [S: \(n\) placental, placental mammal, eutherian, eutherian mammal](#) (mammals having a placenta; all mammals except monotremes and marsupials)
 - [S: \(n\) mammal, mammalian](#) (any warm-blooded vertebrate having the skin more or less covered with hair; young are born alive except for the small subclass of monotremes and nourished with milk)
 - [S: \(n\) vertebrate, craniate](#) (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
 - [S: \(n\) chordate](#) (any animal of the phylum Chordata having a notochord or spinal column)
 - [S: \(n\) animal, animate being, beast, brute, creature, fauna](#) (a living organism characterized by voluntary movement)
 - [S: \(n\) organism, being](#) (a living thing that has (or can develop) the ability to act or function independently)
 - [S: \(n\) living thing, animate thing](#) (a living (or once living) entity)
 - [S: \(n\) whole, unit](#) (an assemblage of parts that is regarded as a single entity) "*how big is that part compared to the whole?*"; "*the team is a unit*"
 - [S: \(n\) object, physical object](#) (a tangible and visible entity; an entity that can cast a shadow) "*it was full of rackets, balls and other objects*"
 - [S: \(n\) physical entity](#) (an entity that has physical existence)
 - [S: \(n\) entity](#) (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

- Taxonomy



- [illegible]

ILSVRC 2012 Data

Training data

1,281,167 images in 1000 synsets

Min = 732 , max = 1300 (per synset)

544,546 images have bounding box annotations

Min = 250 / synset

619,207 bounding box annotations

Validation data

50 images / synset

76,750 bounding box annotations

Test data

100 images / synset

153,503 bounding box annotations

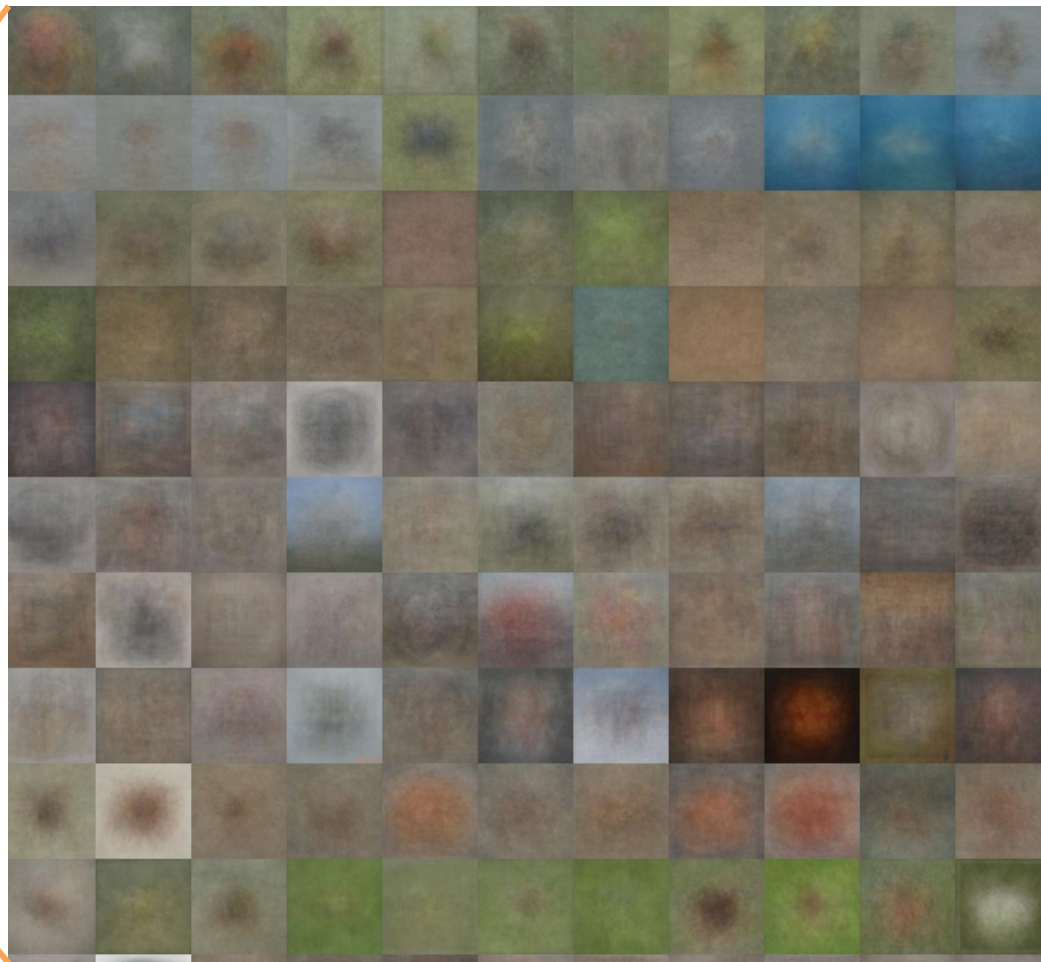
ILSVRC 2012 Data

Breakdown of categories (samples):

Subtree	# of leaf categories
Instrument	358
Canine	130
Covering	90
Vehicle	67
Invertebrate	61
Bird	59
Structure (construction)	58
Food	27
...	

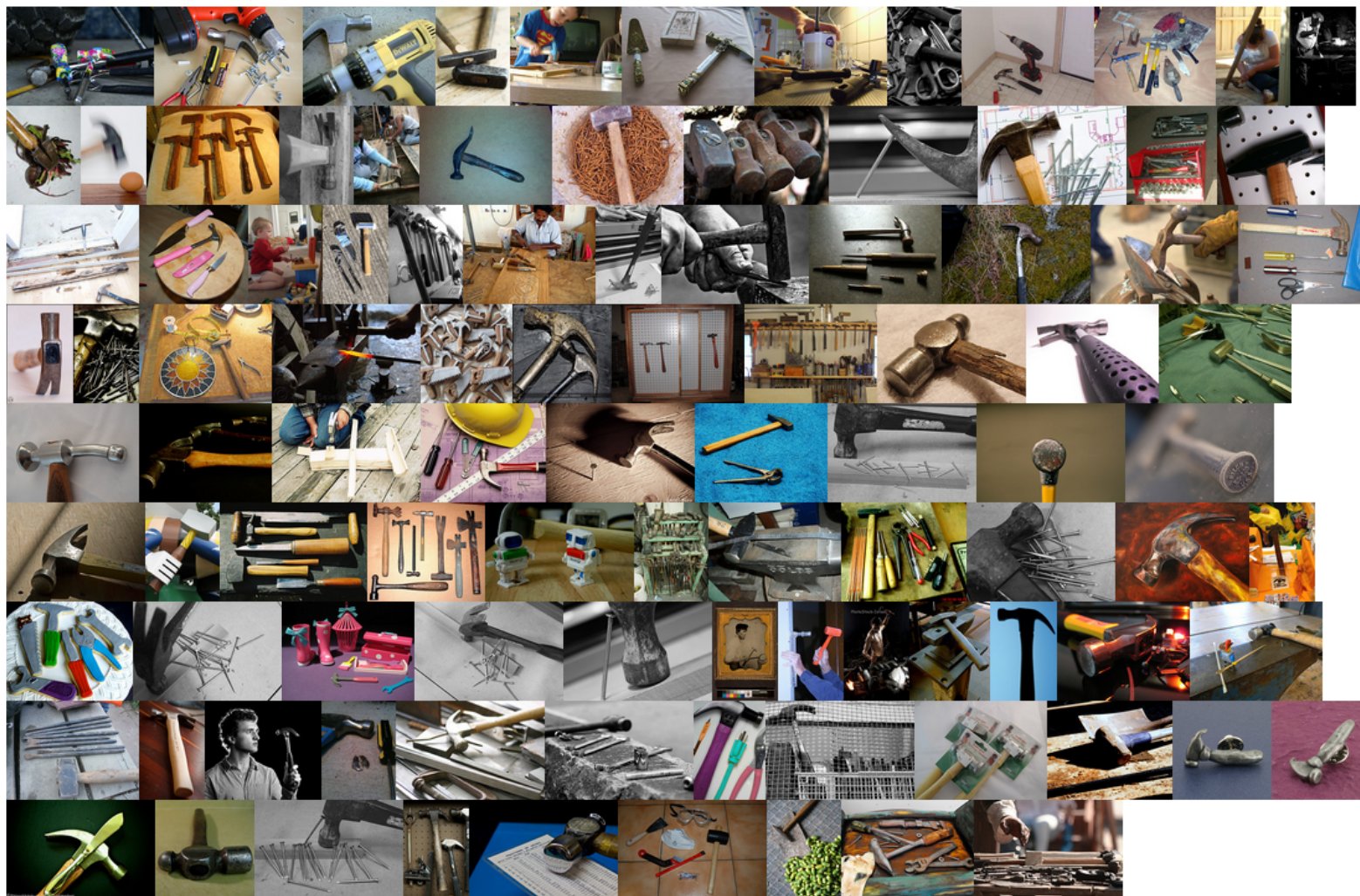
ILSVRC 2012 Data

Average Test images



ILSVRC 2012 Data

Test images for “Hammer”



ILSVRC 2012 Data

- Summary of Challenges
 - A large number of images in training
 - A large number of classes
 - Diversity of classes.
 - Diversity of images within classes

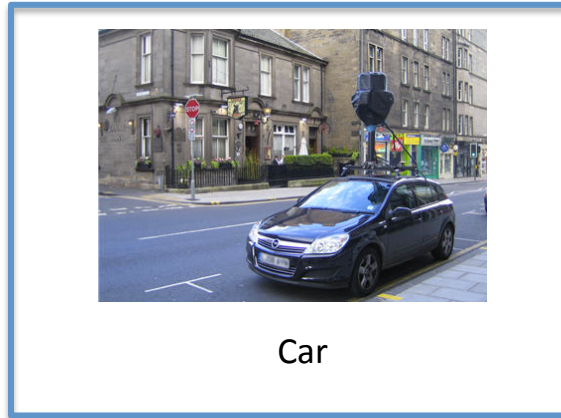
ILSVRC 2012 Results

Participation

107 registrations

30 submissions

Task 1: Classification

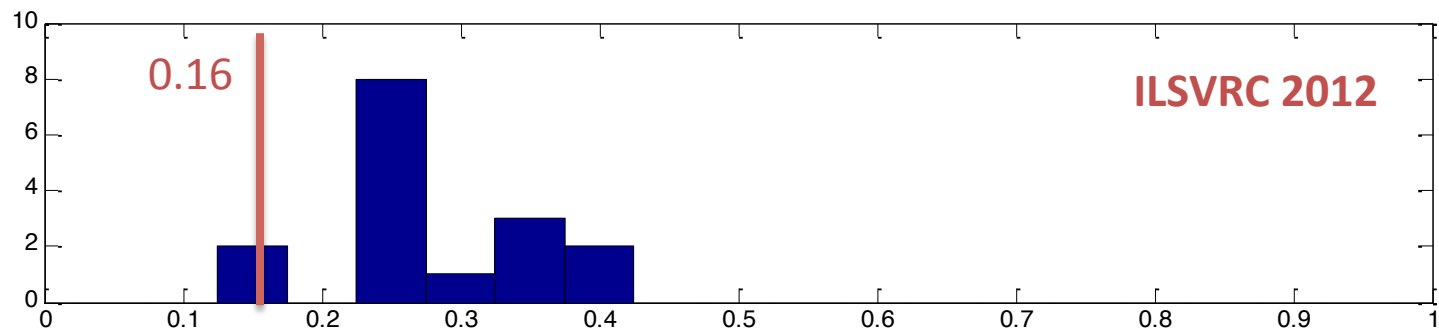
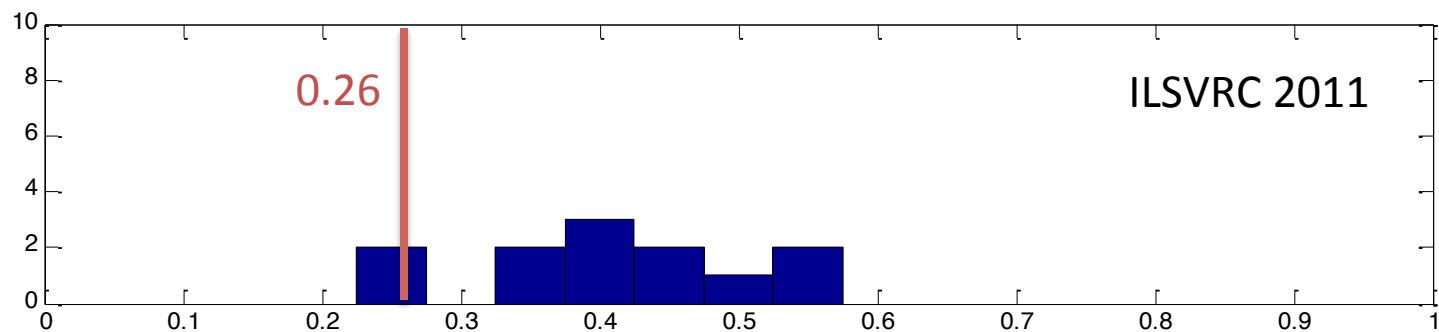
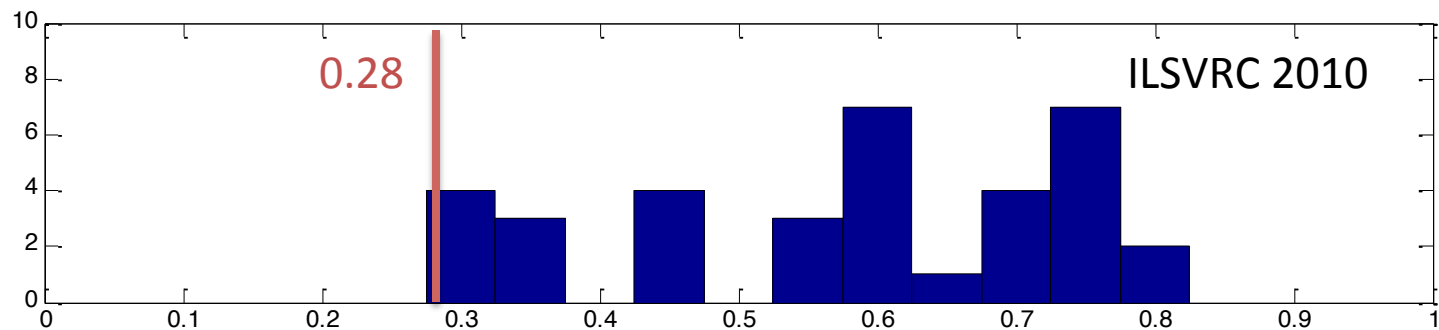


Winner

SuperVision

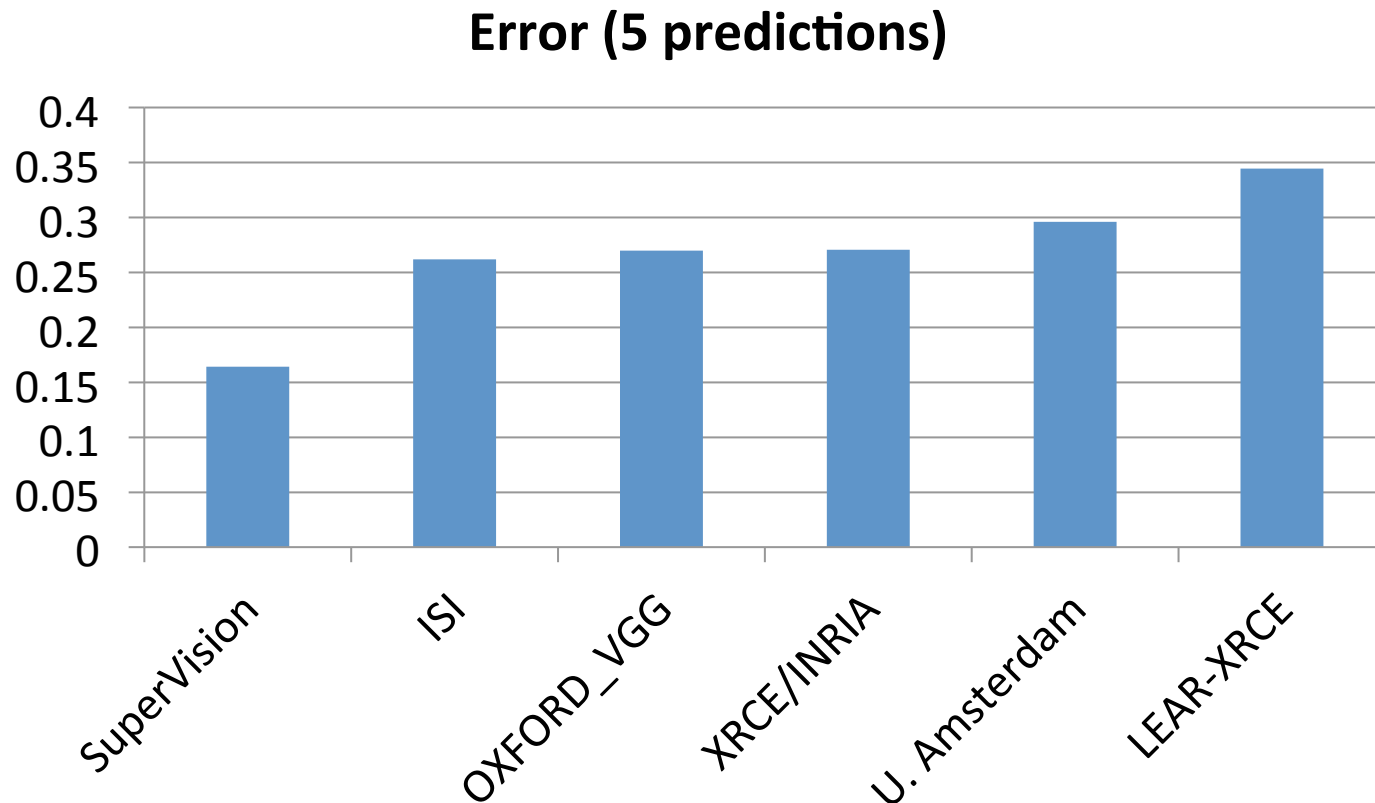
Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton
University of Toronto

Submissions



Error (5 predictions/image)

Ranking of the best results from each team



Easiest Categories

geyser 0.001667



odometer 0.011667

canoe 0.013333



yellow lady's slipper 0.015

web site, 0.015

gondola 0.016667



rapeseed 0.018333

flamingo 0.023333

electric locomotive 0.025



daisy 0.025

* Numbers indicate the mean error from the top 5 predictions from all submissions

Toughest Categories

ladle 0.876667

hatchet 0.856667

spatula 0.833333

muzzle 0.831667

hook, claw 0.805

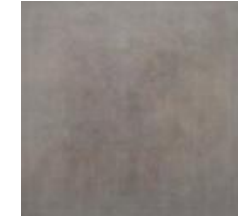
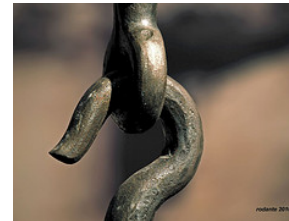
cleaver, 0.79

letter opener, paper knife 0.785

plunger, 0.758333

chime, bell, gong 0.75

power drill 0.745

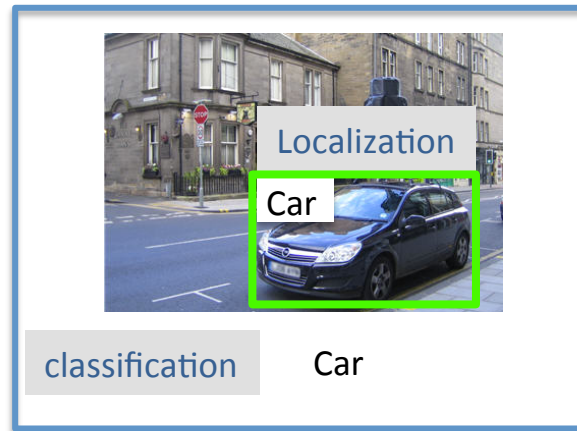


* Numbers indicate the mean error from the top 5 predictions from all submissions

A large collage of 30 small images arranged in a grid-like fashion, showcasing various people using different types of spoons and kitchen utensils. The images are diverse in setting and activity: some show people cooking in professional kitchens, others show them eating or playing with food. The utensils featured include large wooden spoons, metal ladles, slotted spoons, and even a baby in a bathtub with a spoon. The collage is a visual representation of the 'Spoon' theme, showing the many ways spoons are used in daily life.

[illegible]

Task 2: Detection (Classification + Localization)

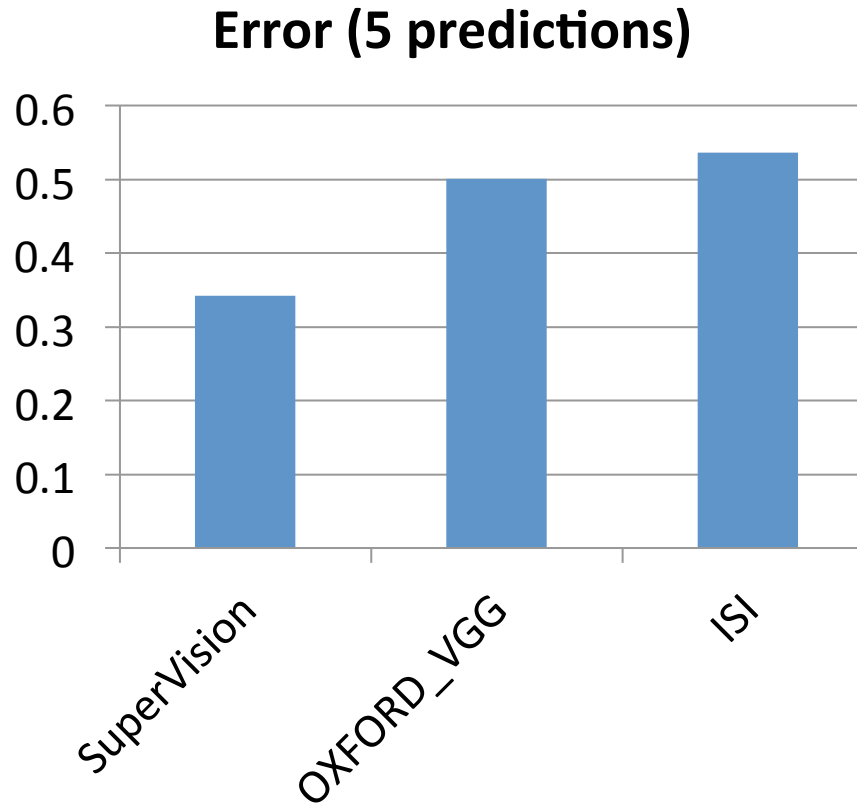


Winner

SuperVision

Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton
University of Toronto

Ranking of best results from each team



Easiest categories to localize given correct classification

prison, prison house	0.985965
confectionery, candy store	0.984379
toyshop	0.979347
pickup, pickup truck	0.978096
minibus	0.977033
ambulance	0.974638
convertible	0.970129
jersey, T-shirt, tee shirt	0.96158
web site, website, internet site, site	0.959527
Leonberg	0.958754



* Numbers indicate the percentage of correct localization given correct classification averaged over all submissions

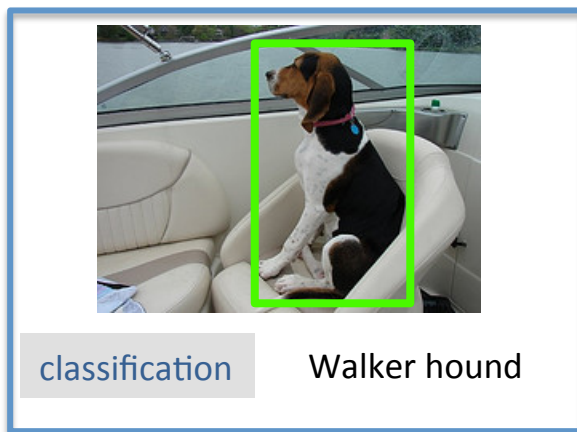
Hardest categories to localize given correct classification

ping-pong ball	0.123032
space bar	0.171802
basketball	0.173669
torch	0.19592
rugby ball	0.206638
puck, hockey puck	0.211481
water snake	0.212903
racket, racquet	0.214662
croquet ball	0.234641
ski	0.239927



* Numbers indicate the percentage of correct localization given correct classification averaged over all submissions

Task 3: Fine-Grained Classification

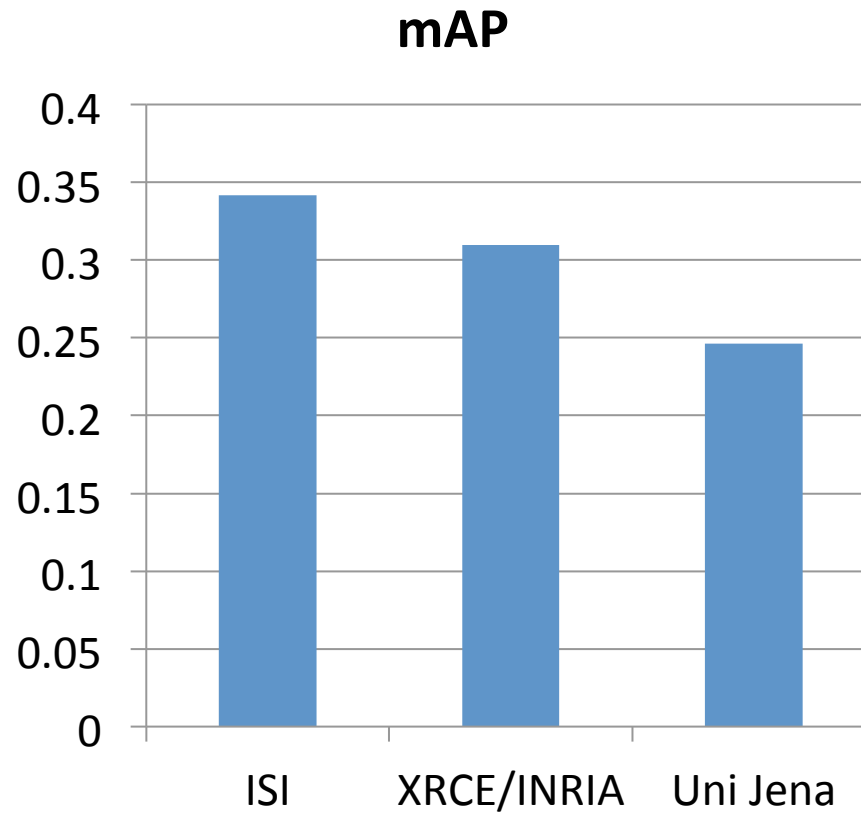


Winner

ISI

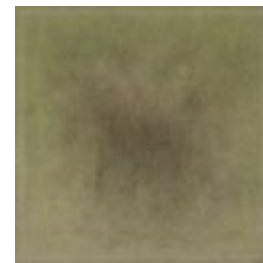
Naoyuki Gunji (the Univ. of Tokyo), Takayuki Higuchi (the Univ. of Tokyo), Koki Yasumoto (the Univ. of Tokyo), Hiroshi Muraoka (the Univ. of Tokyo), Yoshitaka Ushiku (the Univ. of Tokyo), Tatsuya Harada (the Univ. of Tokyo & JST PRESTO), Yasuo Kuniyoshi (the Univ. of Tokyo)

Results



Easiest Dogs

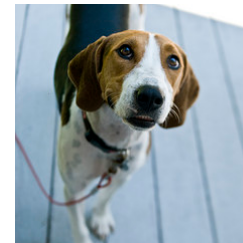
African hunting dog,	0.791974
Leonberg	0.790108
komondor	0.698301
Samoyed, Samoyede	0.605438
Tibetan mastiff	0.601007
keeshond	0.595873
dhole, Cuon alpinus	0.577688
Afghan hound, Afghan	0.523334
Sealyham terrier, Sealyham	0.520428
chow, chow chow	0.509193



* Numbers indicate the mean average precision from all submissions

Hardest Dogs

soft-coated wheaten terrier	0.034898
Great Dane	0.073262
American Staffordshire terrier	0.09573
miniature poodle	0.098169
standard schnauzer	0.100944
otterhound, otter hound	0.107735
Irish terrier	0.112036
Chihuahua	0.115103
cocker spaniel, English cocker spaniel, cocker	0.116995
Walker hound, Walker foxhound	0.122125



* Numbers indicate the mean average precision from all submissions

Future of ImageNet Challenges

What we have learned

- Promising techniques

Method	Year
Locality-constrained Linear Coding	ILSVRC 2010
Improved Fisher Vectors	ILSVRC 2011
Convolutional Neural Networks	ILSVRC 2012

- Also, 400+ citation of ImageNet
 - ECCV 2012 Best Paper award (“Segmentation Propagation in ImageNet”, Kuettel et al.)

Should we run ILSVRC 2013?

ILSVRC 2013?

- What tasks?
 - (1) Classification
 - (2) Classification with complete ground truth labels
 - (3) Fine-grained classification
 - (4) Classification + Localization
 - (5) PASCAL VOC style detection on 1K classes

ILSVRC 2013?

- Logistics?
 - Data collection
 - Budget

agenda

- 15:30 - 16:00. Introduction and overview of results. Fei-Fei Li.
- 16:00 - 16:25. Invited talk. OXFORD_VGG team
- 16:25 - 16:40. Break
- 16:40 - 17:05. Invited Talk. ISI team
- 17:05 - 17:30. Invited Talk. SuperVision team
- 17:30 - 18:00. Discussion.