



Large Scale Visual Recognition Challenge 2010

Alex Berg
Stony Brook

Jia Deng
Princeton &
Stanford

Fei-Fei Li
Stanford

Large Scale Recognition

- Millions to billions of images
- hundreds of thousands of labels
- Classifiers should provide information for indexing and retrieval
- Complement current Pascal VOC competitions
- This year's challenge:
 - Image categorization
 - 1000 categories
 - 1.2 million training images
 - 50 thousand new verification images
 - 150 thousand new test images
 - Success would provide techniques for image indexing and search
- Next year
 - Categorization and detection (now with bounding boxes!)
 - Object detection and image categorization but not full image parsing

Source for categories and training data

- ImageNet
 - 12 million images, 15 thousand categories
 - Image found via web searches for WordNet noun synsets
 - Hand verified using Mechanical Turk
 - All new data for validation and testing this 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 goal, but should provide information for other tasks, so idiosyncrasies of WordNet may be less critical



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

IM GENET

11,231,732 images, 15589 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.



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

ImageNet 2010 Spring Release is up! [Click here to check out what's new!](#)

Step 1: Collect Candidate Images from the Internet

- Query expansion
 - Synonyms: German shepherd, German police dog, German shepherd dog, Alsatian
 - Appending words from ancestors: sheepdog, dog
- Multiple languages
 - Italian, Dutch, Spanish, Chinese

e.g. ovejero alemán, pastore tedesco, 德国牧羊犬

- More engines



- Parallel downloading



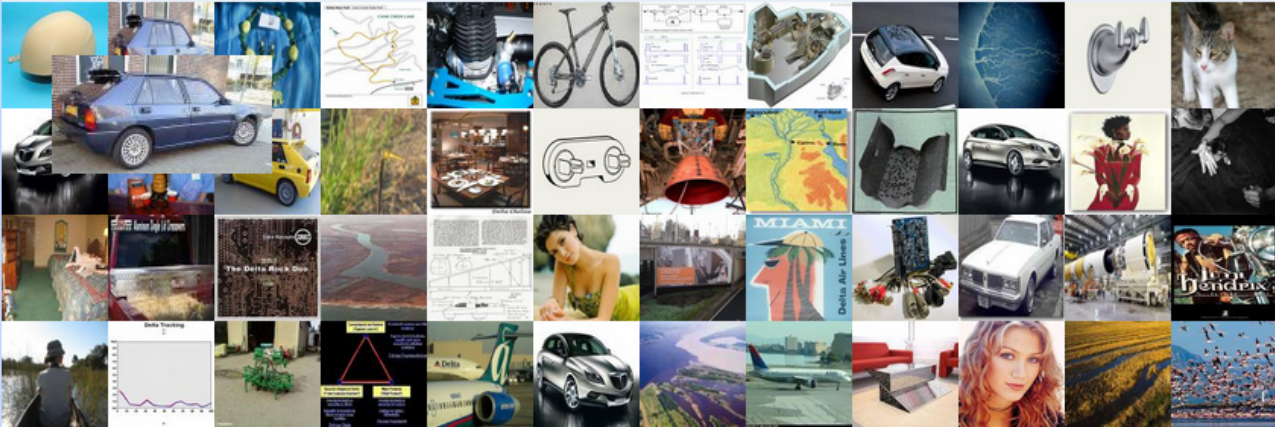
IMAGENET Basic User Interface

Click on the good images.

Main | Instructions | Unsure? Look up in Wikipedia | Google | [\[Additional input \] No good photos? Have expertise? comments? Click here!](#)

[First time workers please click here for instructions.](#)

Click on the photos that contain the object or depict the concept of : **delta**: a low triangular area of alluvial deposits where a river divides before entering a larger body of water; "the Mississippi River delta"; "the Nile delta" .(PLEASE READ DEFINITION CAREFULLY)
Pick as many as possible. *PHOTOS ONLY, NO PAINTINGS, DRAWINGS, etc.* It's OK to have other objects, multiple instances, occlusion or text in the image.
Do not use back or forward button of your browser. OCCASIONALLY THERE MIGHT BE ADULT OR DISTURBING CONTENT.



Below are the photos you have selected FROM THIS PAGE ONLY (they will be saved when you navigate to other pages). Click to deselect.

what's this? | select all | deselect all | < page 1 of 6 > | Submit | PREVIEW MODE. TO WORK ON THIS HIT, ACCEPT IT FIRST.

Definition quiz

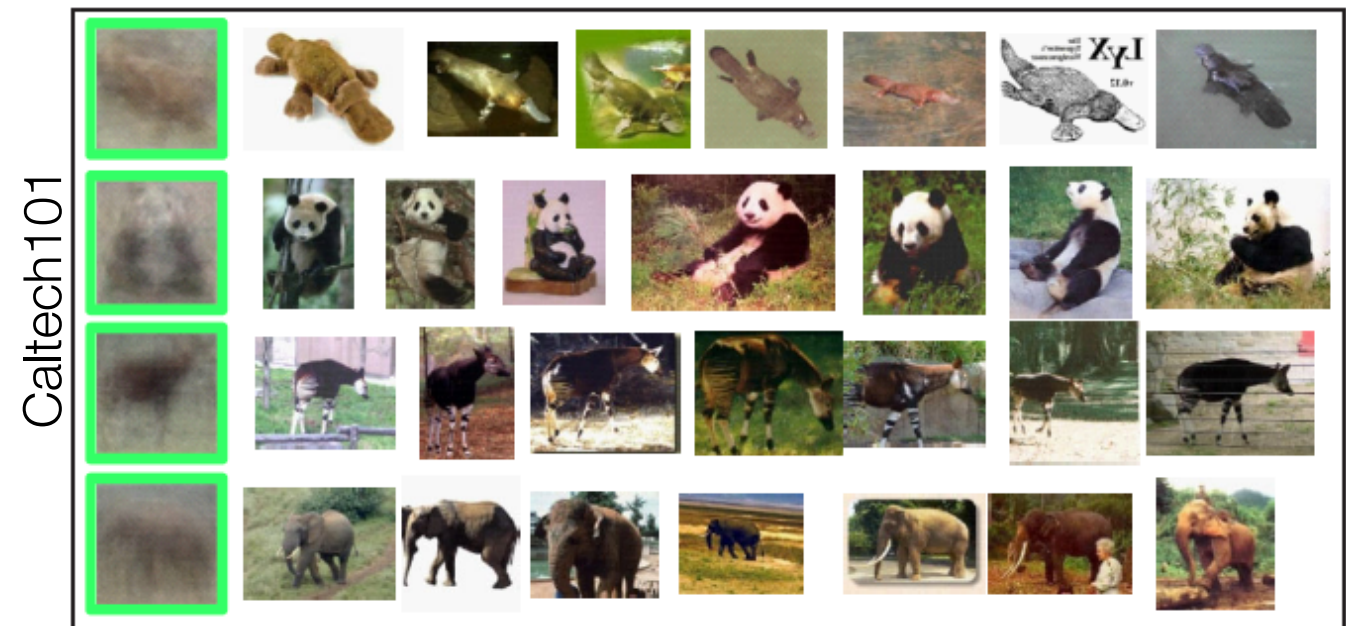
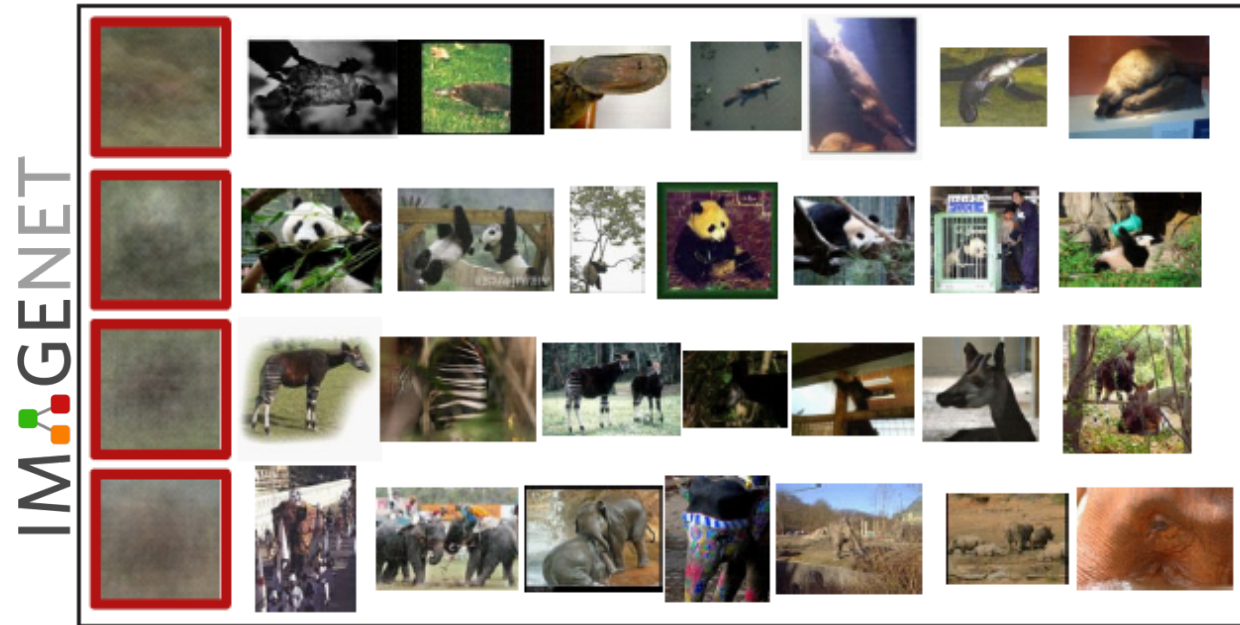
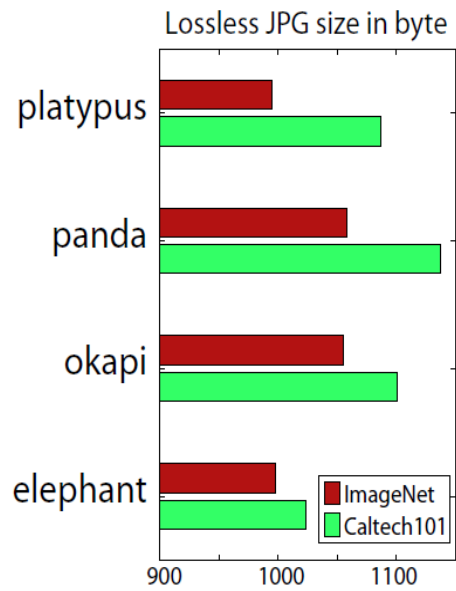
This HIT is about 'delta'.

Definition: a low triangular area of alluvial deposits where a river divides before entering a larger body of water; "the Mississippi River delta"; "the Nile delta"

Please read the above definition carefully. 'delta' might mean something different from what you think.

I HAVE READ IT

Diversity



This year's challenge

- Given an image predict categories of objects that may be present in the image
- 1000 “leaf” categories from ImageNet
- Two evaluation criteria based on cost averaged over test images
 - Flat cost – pay 0 for correct category, 1 otherwise
 - Hierarchical cost – pay 0 for correct category, height of least common ancestor in WordNet for any other category
- Allow a shortlist of up to 5 predictions
 - Use the lowest cost prediction each test image
 - Allows for incomplete labeling of all categories in an image

Participation

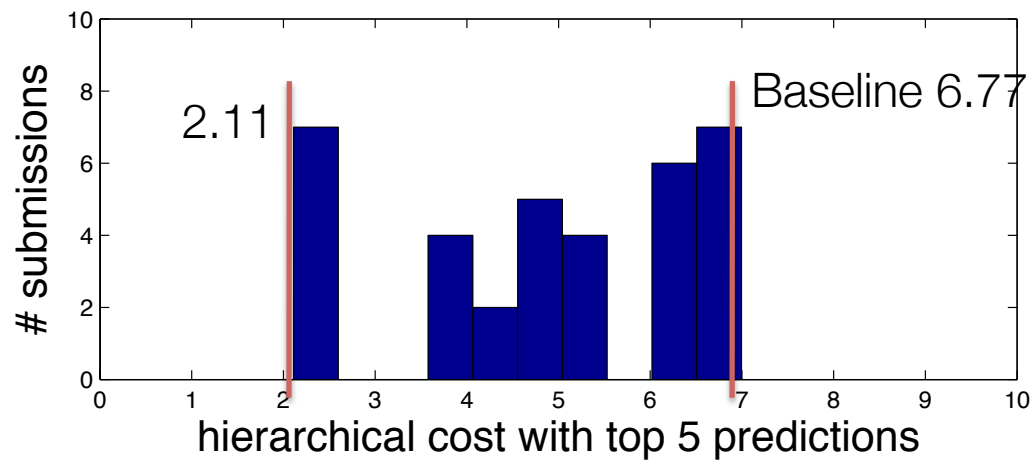
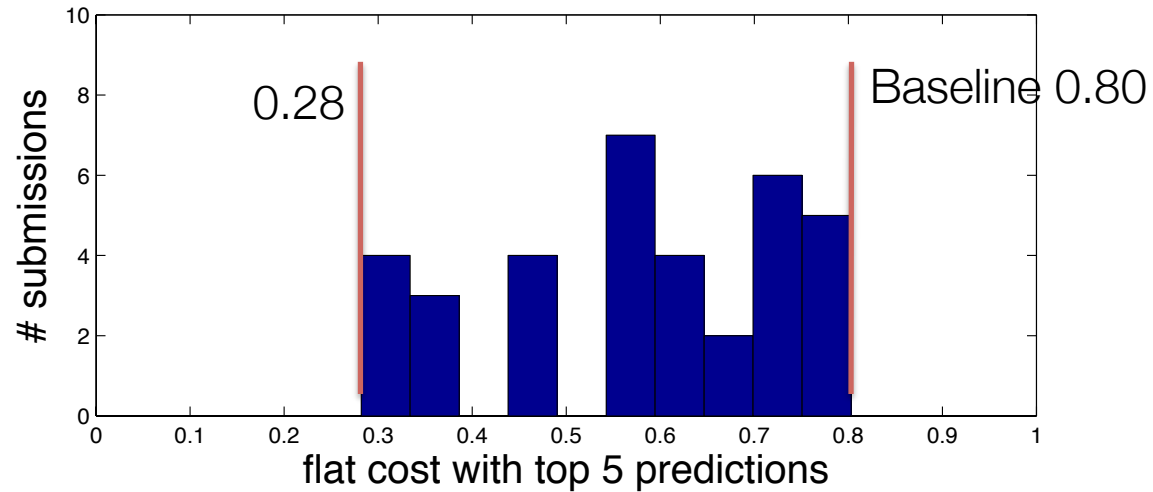
150+ registrations

35 submissions from 11 teams

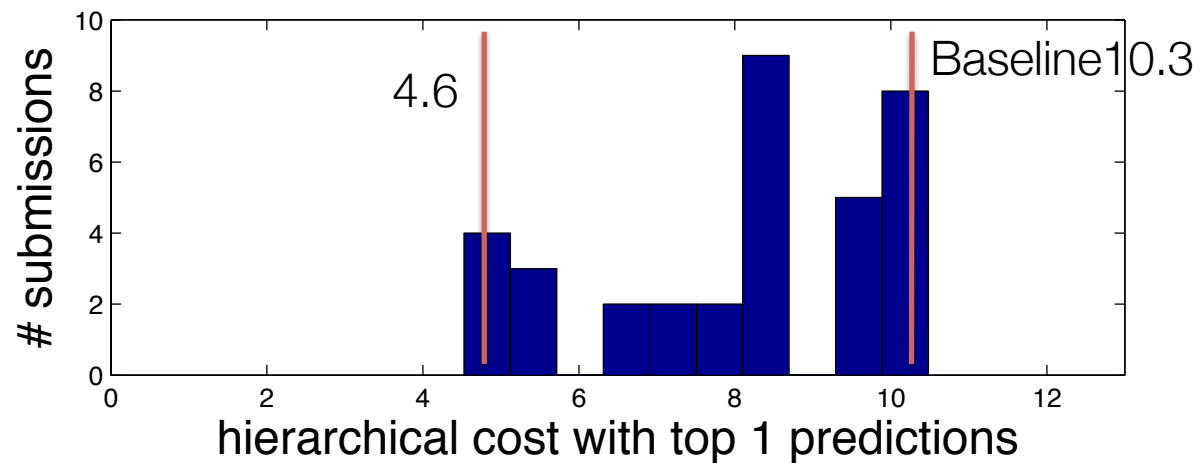
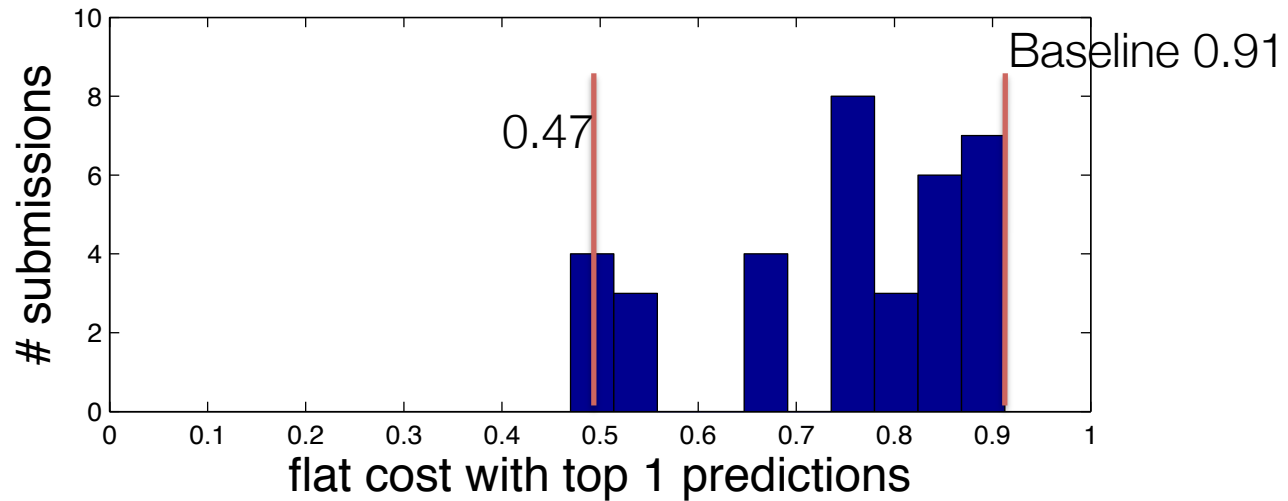
Teams from around the world

- NEC-UIUC, USA
- XRCE, France
- ISIL, University of Tokyo, Japan
- UC Irvine, USA
- MIT, USA
- Nanyang Technological University, Singapore
- LIG Grenoble, France
- IBM-ensemble, USA
- SRI International, USA
- National Institute of Informatics, Tokyo, Japan
- Harbin Institute of Technology, China

Results: 5 predictions



Results: 1 prediction



Flat versus hierarchical

Performance ordering was almost the same for flat and hierarchical evaluations.

Winners

- Winner: NEC-UIUC

Yuanqing Lin, Fengjun Lv, Shenghuo Zhu, Ming Yang, Timothee Cour, Kai Yu (NEC). LiangLiang Cao, Zhen Li, Min-Hsuan Tsai, Xi Zhou, Thomas Huang (UIUC). Tong Zhang (Rutgers).

- Honorable mention: XRCE

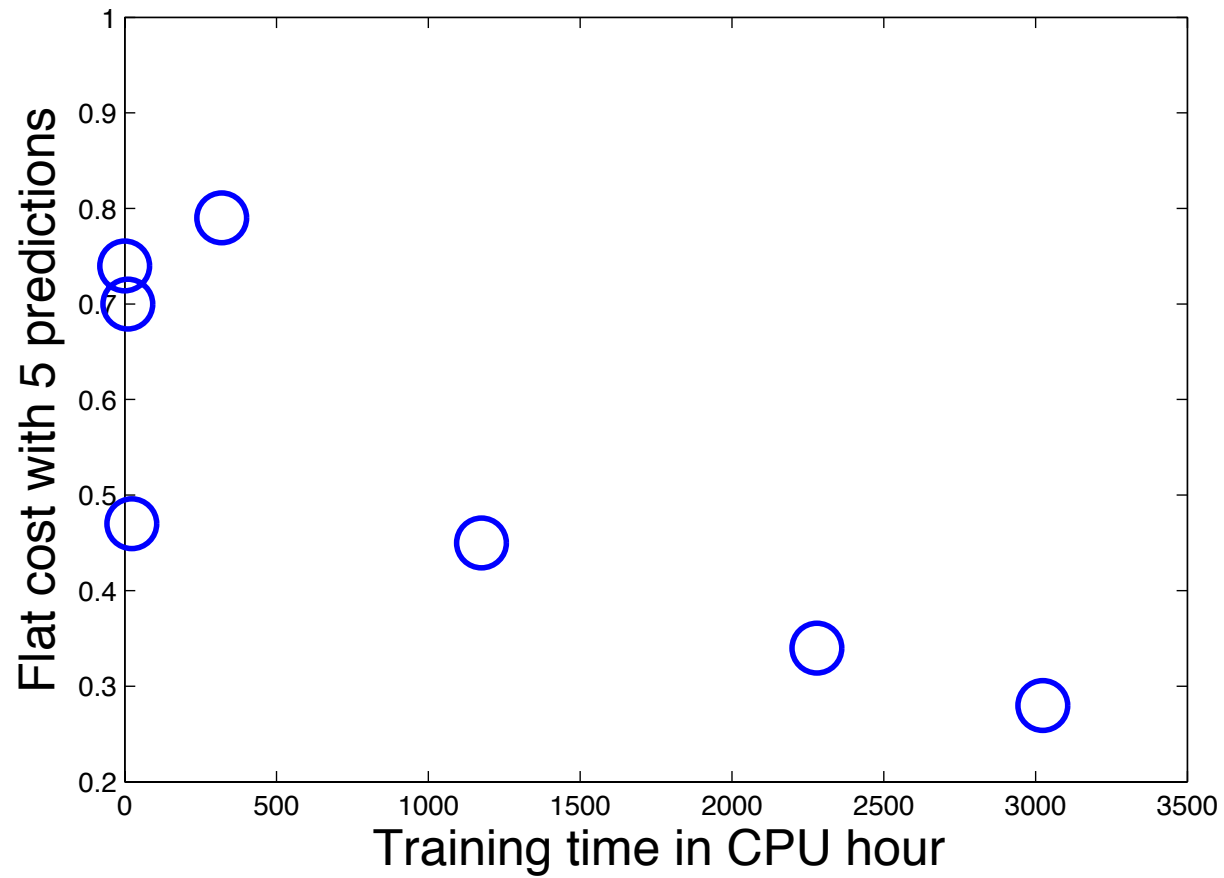
Jorge Sanchez, Florent Perronnin, Thomas Mensink (XRCE)

Ranking by learning methods

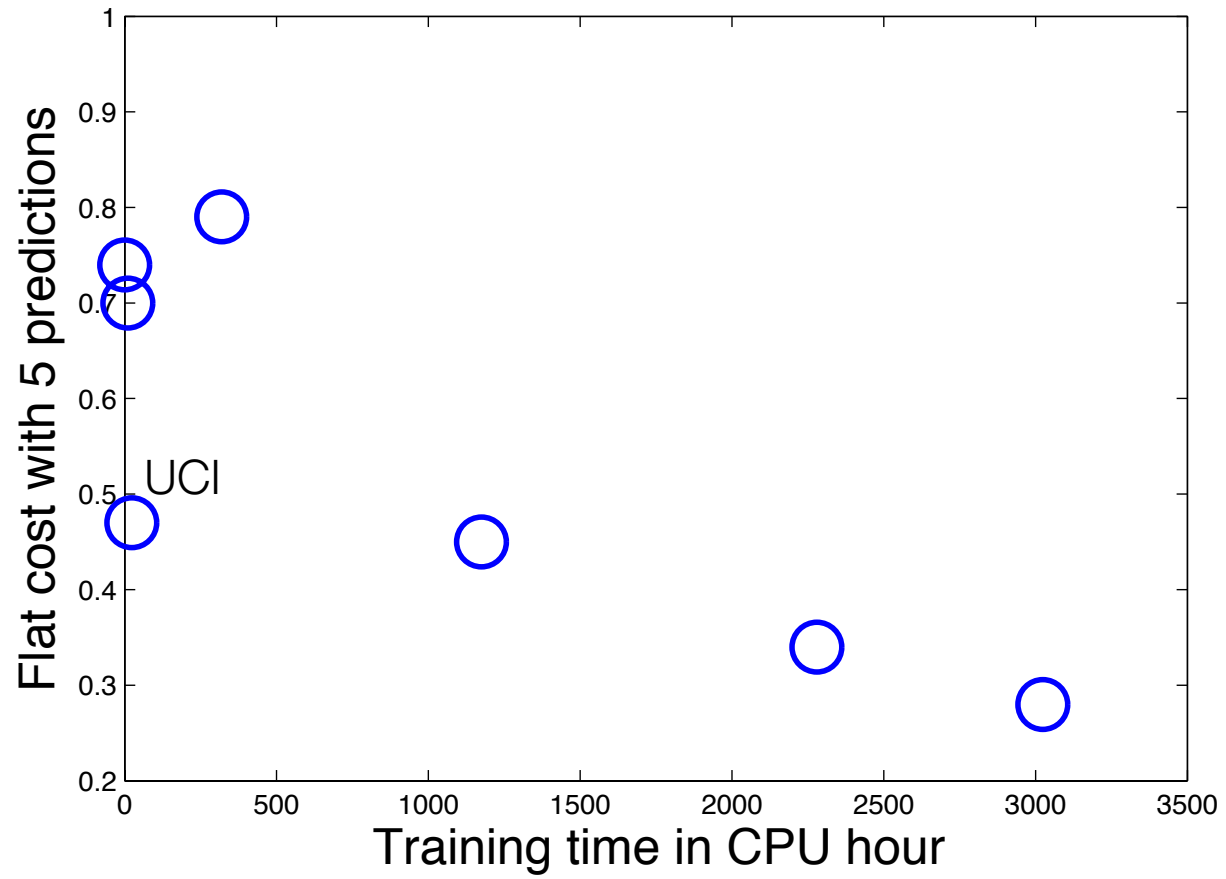
- Descriptor Coding + SVM, 0.28 --- NEC-UIUC
- Fisher kernel + SVM, 0.34 --- XRCE
- LI2C, 0.58 --- NTU_WZX
- KNN, 0.61 --- LIG
- Canonical Correlation Analysis, 0.79 -- NII

*number indicates flat cost with 5 predictions

Computation

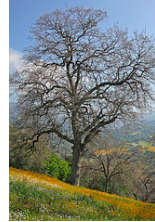


Computation



Hardest categories

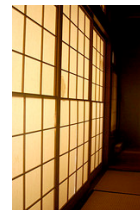
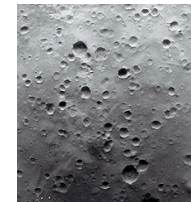
- Logwood, 0.92
- China tree, 0.92
- Red beech, 0.88
- Cap opener, 0.88
- Paintbrush, 0.87
- Needle, 0.87
- Weasel, 0.87
- Syringe, 0.87
- Alder tree, 0.87
- Sorrel tree, 0.86



*number indicates average flat cost with 5 predictions from all submissions

Easiest categories

- Odometer, 0.12
- Website, 0.14
- Lunar crater, 0.18
- Geyser, 0.20
- Monarch butterfly, 0.21
- Manhole cover, 0.22
- Rapeseed, 0.22
- Cliff dwelling, 0.24
- Sunflower, 0.25
- Shoji, 0.26



*number indicates average flat cost with 5 predictions from all submissions

Hardest subtrees

Subtree	# descendants	Average flat cost*
Implement (small tools)	50	0.68
Reptile	29	0.65
Fruit	55	0.64
Dog	31	0.62
Tree	68	0.62

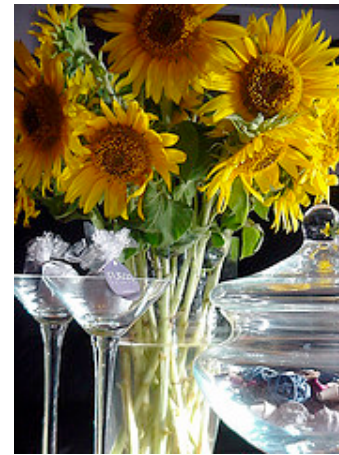
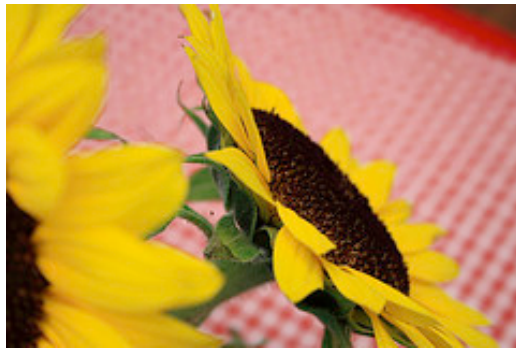
*number indicates average flat cost with 5 predictions from all submissions

Easiest subtrees

Subtree	# descendants	Average flat cost*
Geological formation	12	0.35
Vehicle	60	0.46
Flower	87	0.49
Structure	53	0.50
Equipment (e.g. home appliance)	45	0.52

*number indicates average flat cost with 5 predictions from all submissions

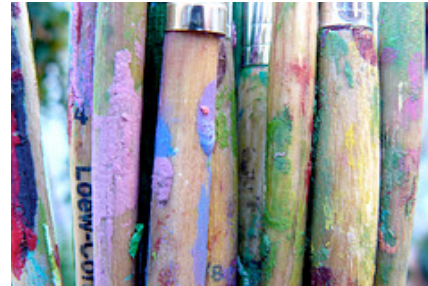
Most difficult sunflowers



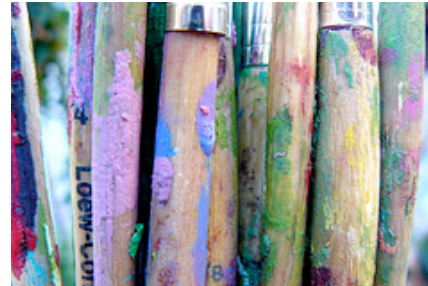
Easiest sunflowers



Most difficult?



Most difficult paintbrushes!

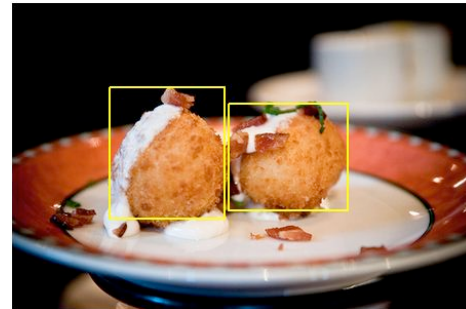
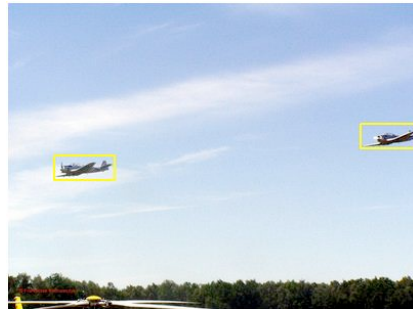
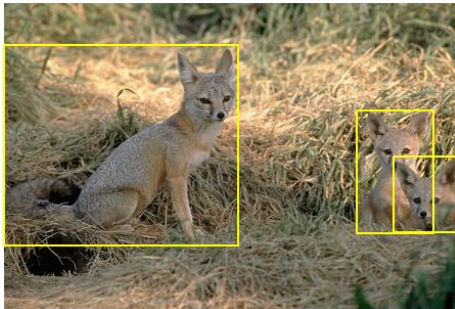


Easiest paintbrushes



2011 Large Scale Visual Recognition Challenge!

- Data:
 - Bounding boxes
 - Not yet full parsing of images
- Task:
 - Image categorization
 - Object detection/localization



Full 2010 results and team info online

<http://www.image-net.org/challenges/LSVRC/2010/results>