Segmentation as Selective Search for Object Recognition in ILSVRC2011

Koen van de Sande
Jasper Uijlings
Arnold Smeulders
Theo Gevers
Nicu Sebe
Cees Snoek

University of Amsterdam, University of Trento

ILSVRC2011 - November 7th 2011

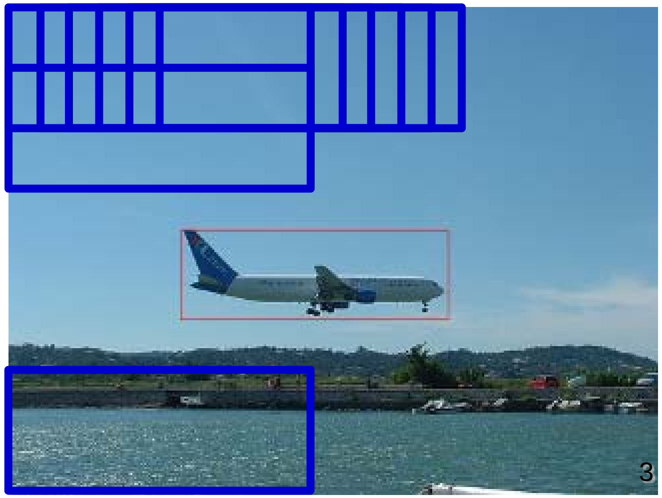
Object Recognition

- What is it?
  ➔ Classification (task #1)
- What is it + Where is it?
  ➔ Classification with localisation (task #2)

Viola IJCV 2004
Dalal CVPR 2005
Felzenszwalb TPAMI 2010
Vedaldi ICCV 2009

Exhaustive Search

- Logical first step, visits every location
- Imposes computational constraints on
  - #possible locations considered (coarse grid/fixed aspect ratio)
  - Evaluation cost per location (weak features/classifiers)
- Impressive results
- To go beyond them, need to be more sophisticated



Carreira CVPR 2010
Endres ECCV 2010
Uijlings CVPR 2009

Selective Search

- Selective search has been exploited convincingly for segmentation
- For segmentation, the emphasis is on finding few but good segments
  - 10-100 per image
  - Accurate delineation ➔ contour detector
- For recognition
  - Once discarded, an object will never be found again
  - Appearance including nearby context more effective
- Use segmentation with different goal

Selective Search for Recognition

- Design criteria
  - $\boxed{\text{High recall}}$
  - Coarse locations are sufficient
    - ⇨ Bounding boxes
  - Fast to compute
    - ⇨ Efficient low-level features
    - ⇨ <10s per image

Selective Search: High Recall

- Image is intrinsically hierarchical

- Segmentation at a single scale won't find all objects

Selective Search: Approach

- Hypotheses based on hierarchical grouping

Selective Search: Approach

- Hypotheses based on hierarchical grouping

Ground truth

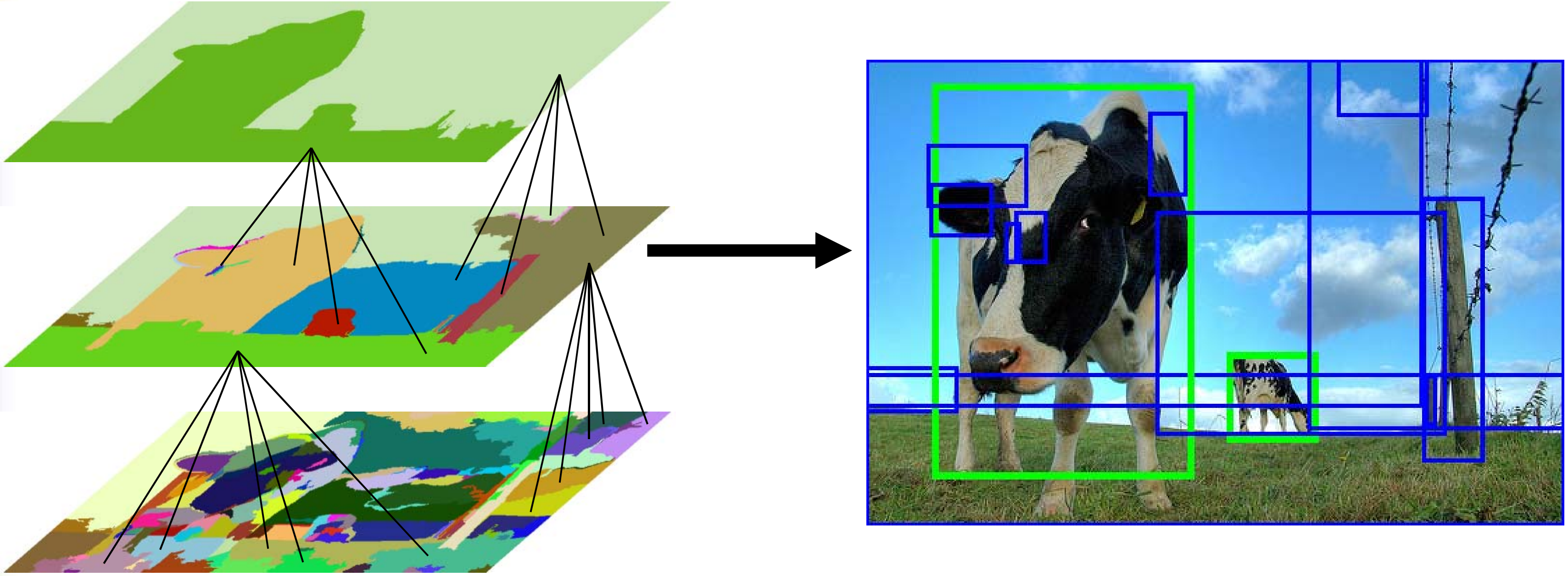
Selective Search: Approach

- Hypotheses based on hierarchical grouping

Initial segments from oversegmentation [Felzenszwalb2004]

Selective Search: Approach

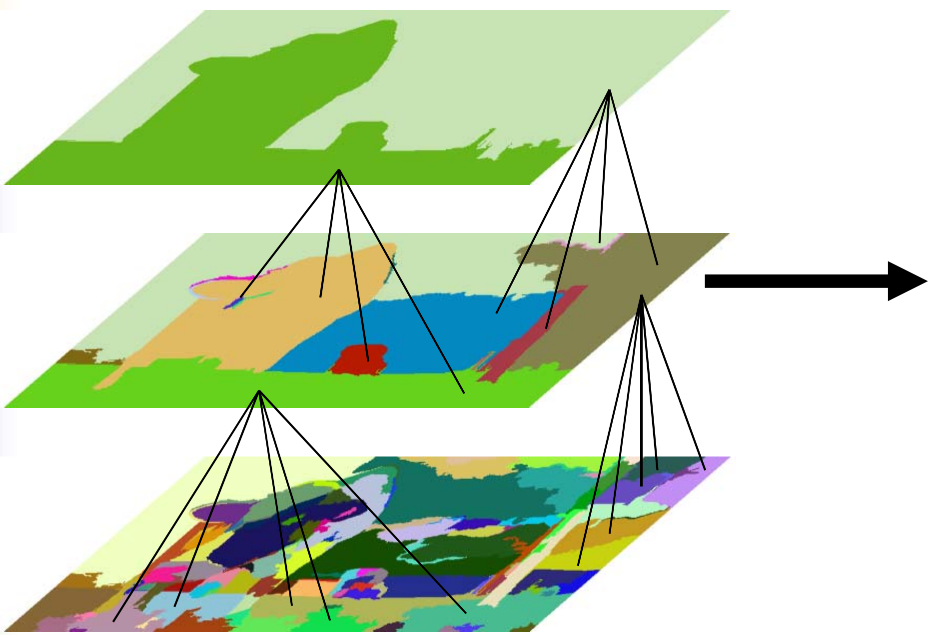
- Hypotheses based on hierarchical grouping

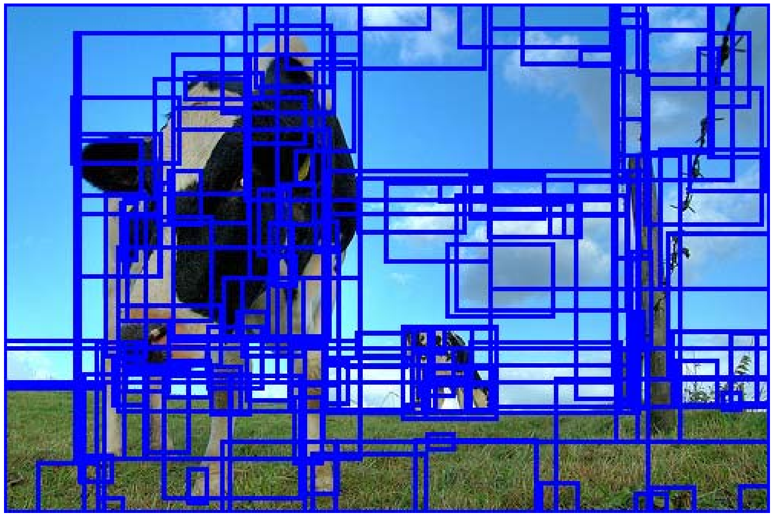


Group adjacent regions on color/texture cues

Selective Search: Approach

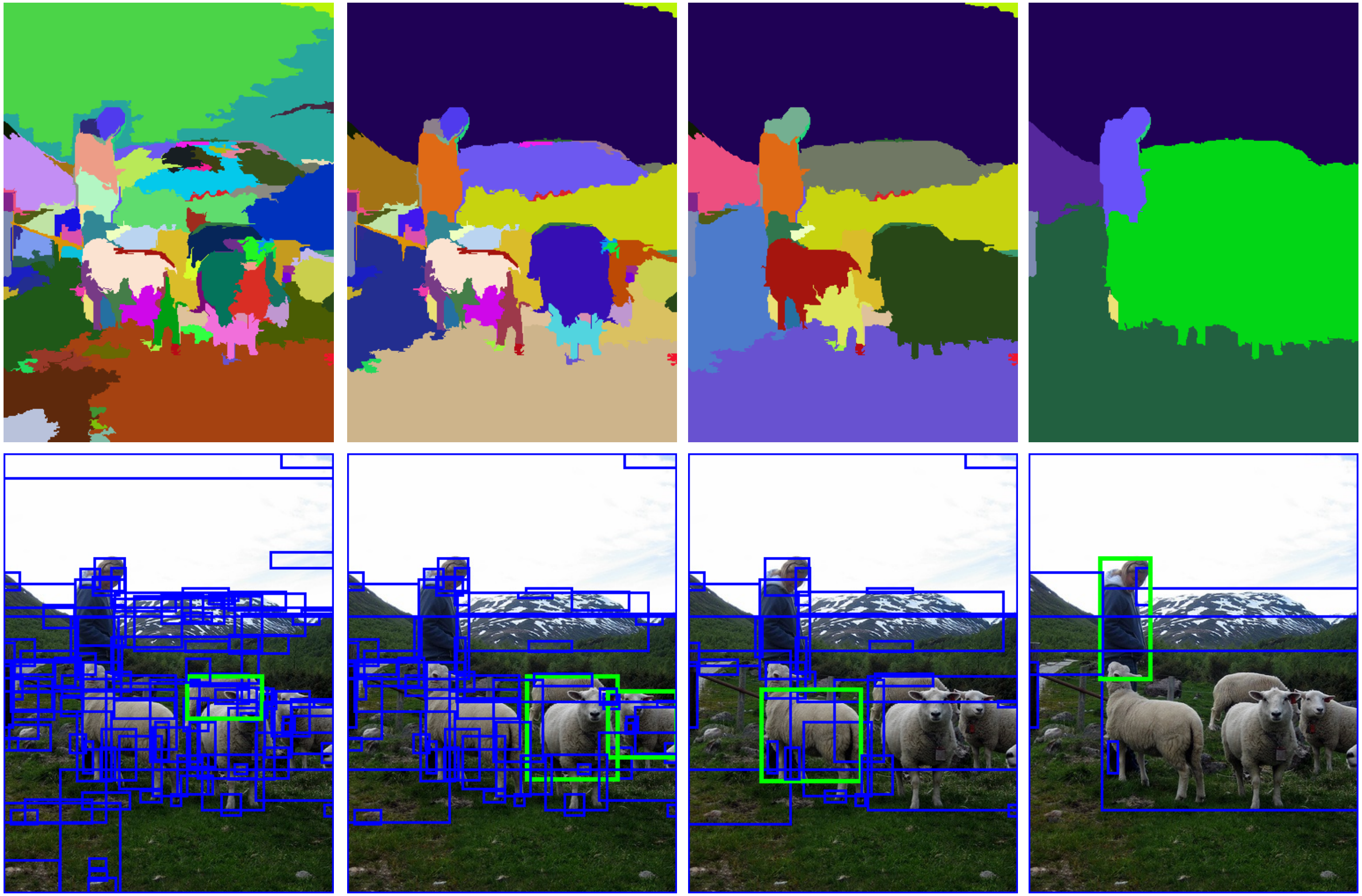
- Hypotheses based on hierarchical grouping





Object hypotheses from all hierarchy levels

Example 1



Example 2

Selective Search: High Recall

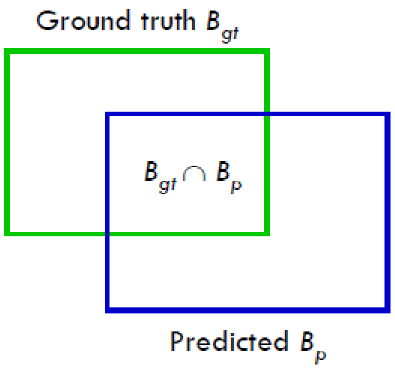
Color cues work best

Texture cues work, color fails

- No single segmentation strategy works everywhere
- **We need a set of complementary segmentation strategies**

Evaluation of object hypotheses

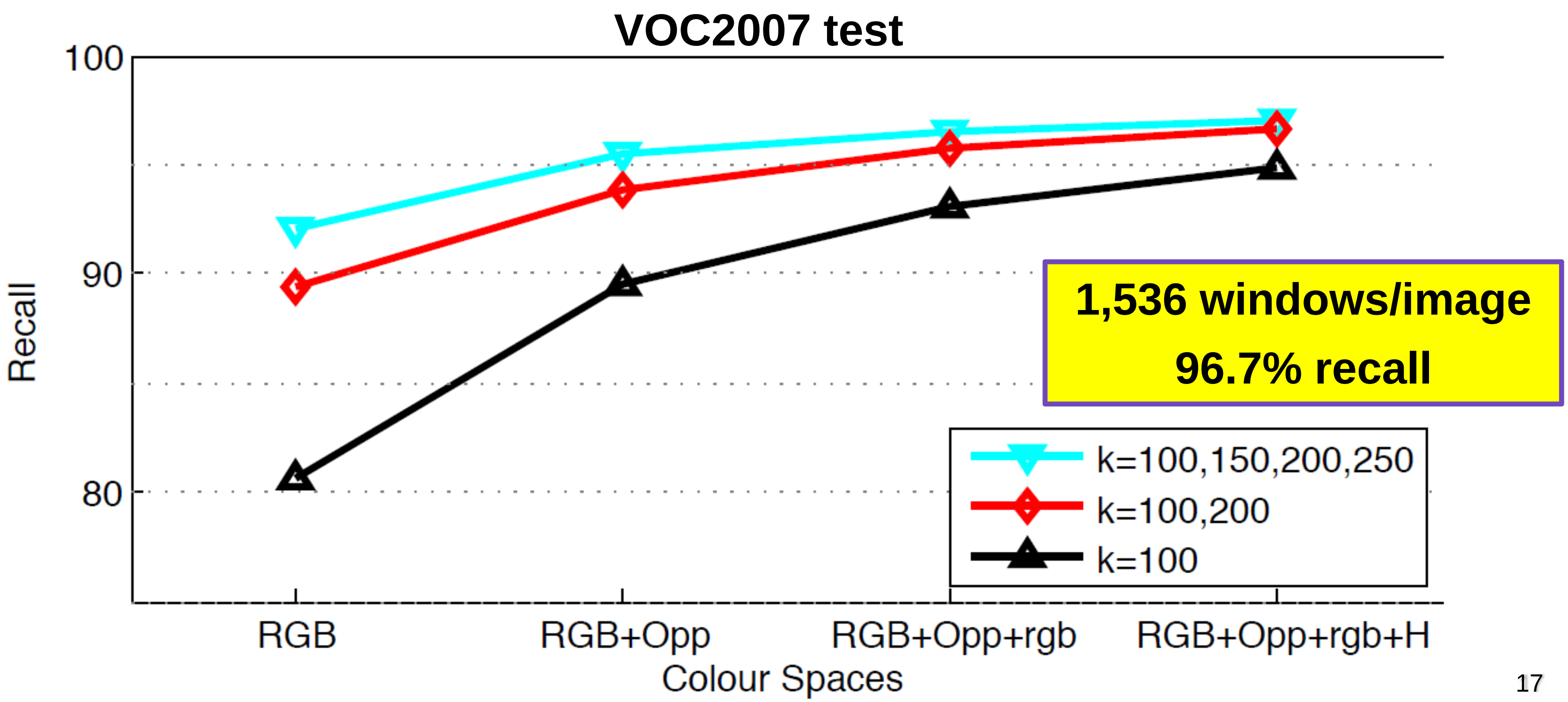
- Standard Pascal VOC overlap criterion
- Correct if overlap normalized for size >= 50%



- Recall is the % of objects for which there is a hypothesis with >=50% overlap

Multiple Complementary Color Spaces

- We diversify the set of segmentations: combine multiple initial segmentations and different color spaces
- Color spaces with complementary invariance properties: some include shadow/shading pixels in a segment, others do not
- Location hypotheses are class-independent



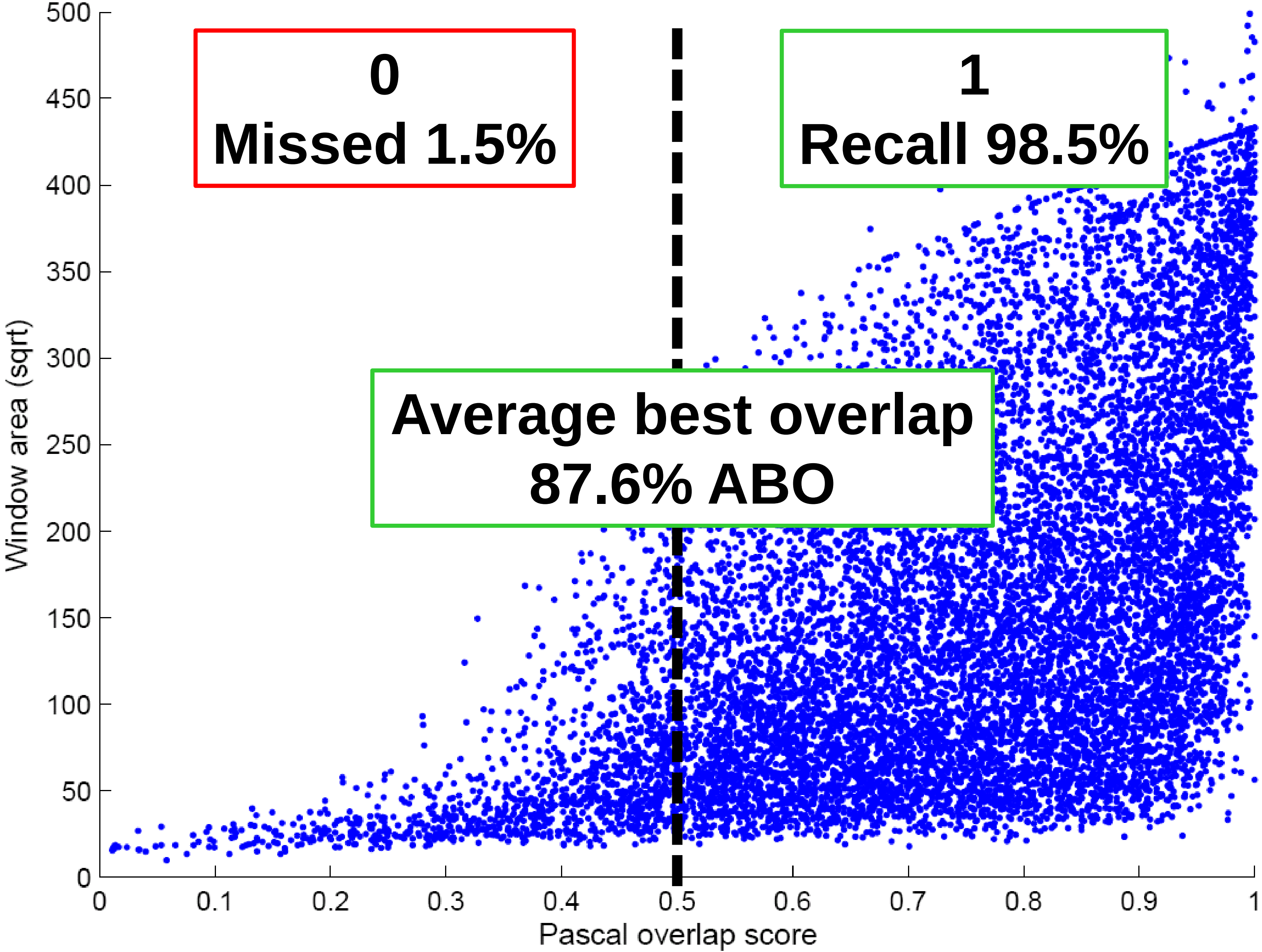
Selective Search on ILSVRC2011

- Apply to ILSVRC2011 train set
- Object hypotheses are class-independent

	ILSVRC2011 train
With bounding box annotations	315,525 images
Average #boxes/image	1,565
Average recall	98.5%

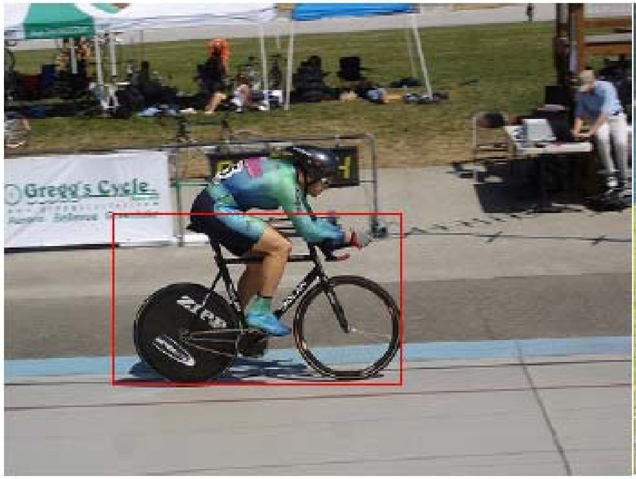
- Recall not informative anymore…

Is recall@50%-overlap still the right measure?

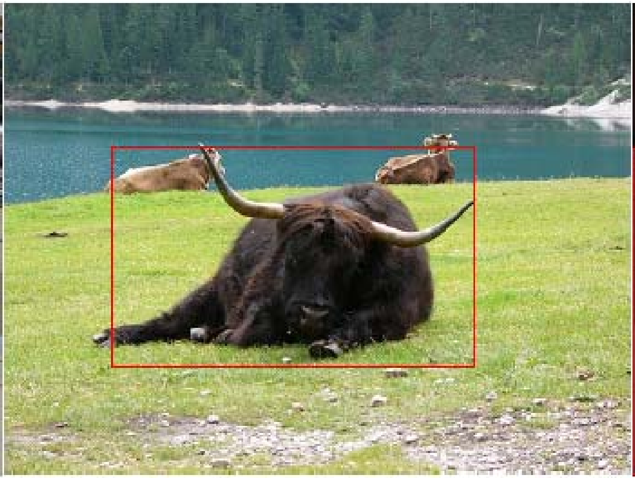


Average Best Overlap Example

- What does ~87.6% ABO look like?



Overlap 88.4%

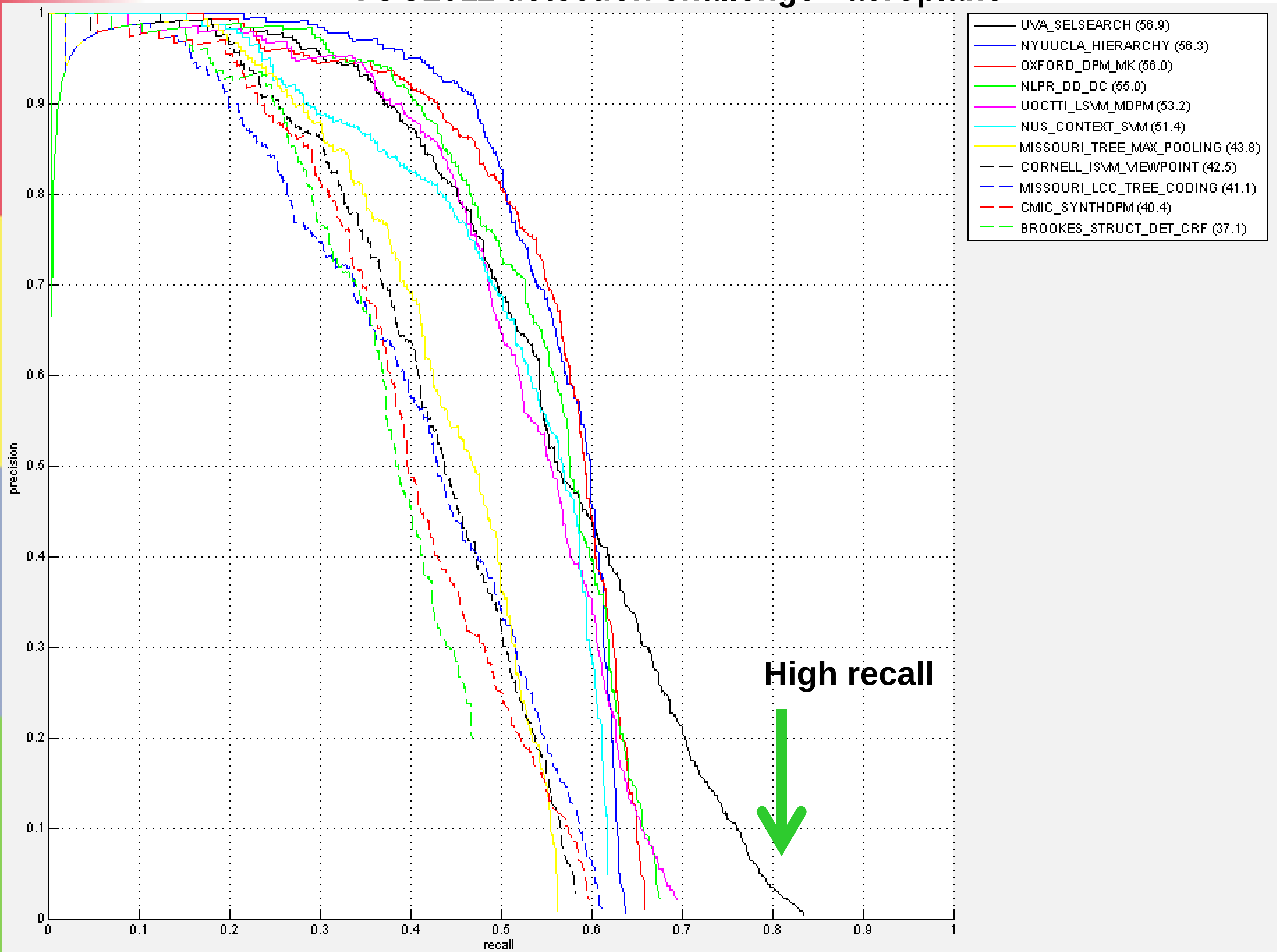


Overlap 87.9%

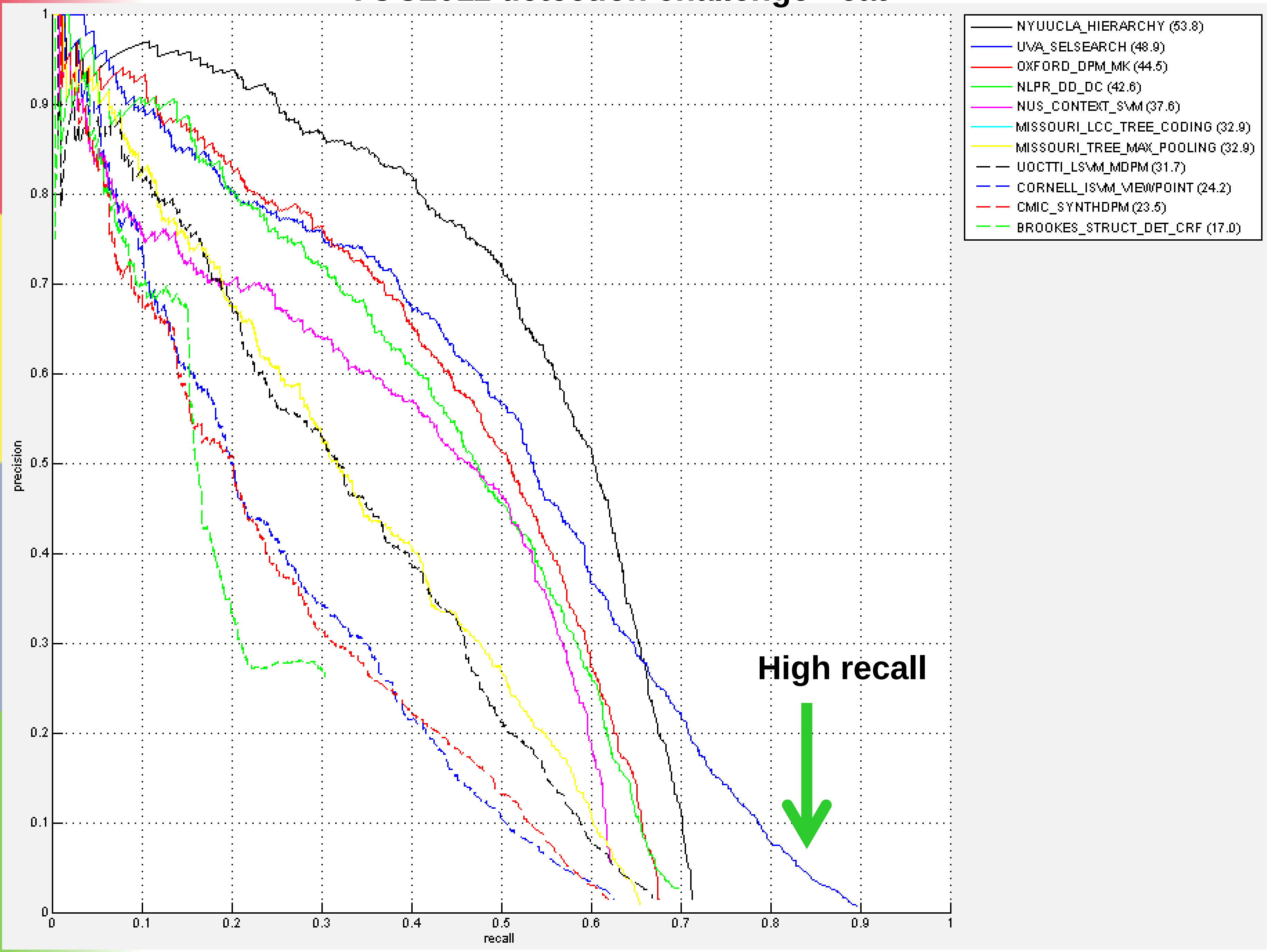
Overlap 87.4%

… is high recall visible in results?

VOC2011 detection challenge - aeroplane



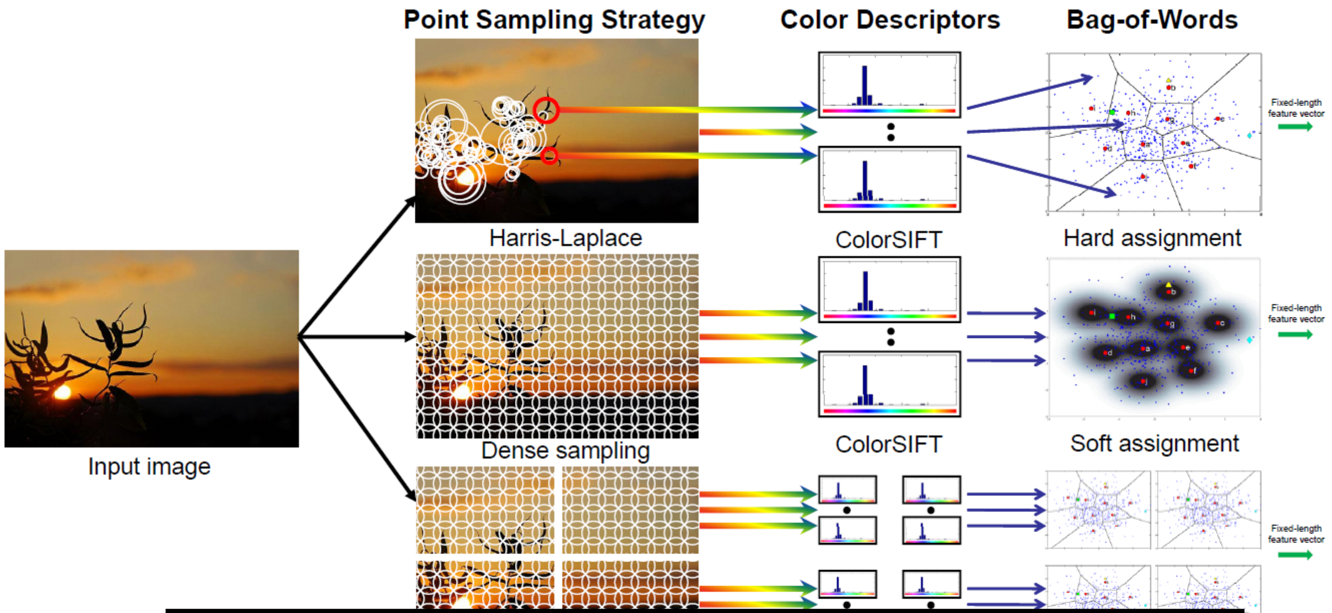
VOC2011 detection challenge - cat



Task #1 - Classification

Task #1: Classification

- Bag-of-words system with configuration similar to VOC2008 participation
  - Dense sampling and Harris-Laplace keypoints
  - SIFT, RGB-SIFT, “X-ColorSIFT” (under submission at IJCV)
  - Spatial pyramid 1x1 and 1x3
  - Vector Quantization with hard assignment
- We have 1.3 million training images and 1000 objects this time around…



Visit http://www.colordescriptors.com

Van de Sande TMM 2011
Maji CVPR 2009

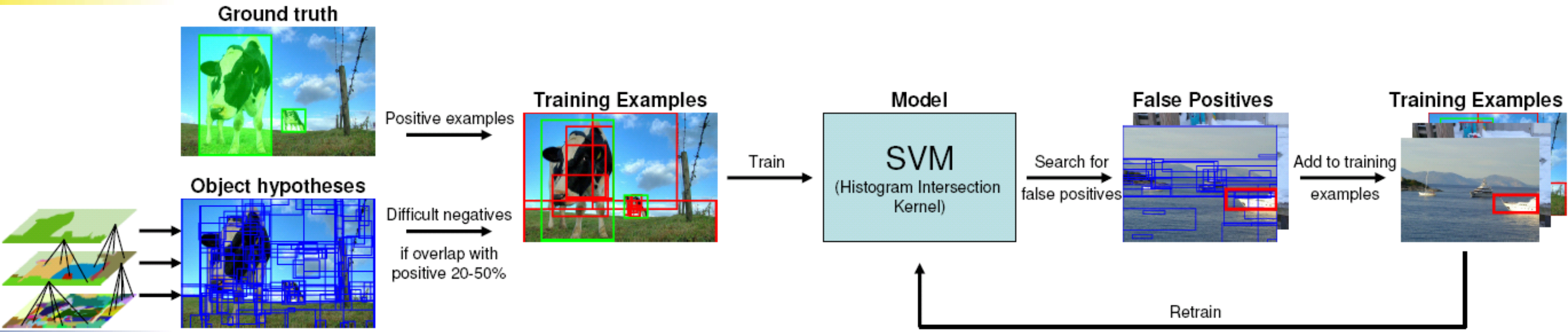
Training 1000 object classifiers

- Each object category has ~1,300 positive examples
- As negative examples: random 10% of the train set
- The same set of negatives for all categories
  ⇨ re-use big part of kernel matrix
- Compute kernel matrix on GPU: 10x faster than quad-core CPU (40x faster than single-core)
- Histogram Intersection Kernel with Fast Approximation
- Based on probabilistic output of LibSVM:
  - Select 5 objects with highest score per image
    ⇨ flat error 0.346

Task #2 – Classification with localisation

Localisation System Training

- Use positives and mirrored positives
- Use object hypotheses to create difficult initial negatives (at most 7,500)
- Add 2 iterations of false positives (from 4,000 images)



- Features: Bag-of-words, sample every pixel, SIFT, “X-ColorSIFT” and RGB-SIFT, pyramid up to level 3, codebook size 4096
- Histogram Intersection Kernel with Fast Approximation

Applying Object Localisation

- 100,000 test images × 1000 object detectors = 100M classifications
- Practical problem: deadline too close
- Each object occurs exactly 100 times
- Apply object detector only to:
  - Top-10 objects per image (in terms of classification likelihood)
  - Top-1000 images per object
  - Reduction to 1.24M classifications

ILSVRC 2011 Results

- Classification: in the 5 labels selected for an image, is the correct label in there?

Task 1	Flat error
Base	0.346
Reranked w/localisation	0.310

→ It is for 65.4% of the images

- Classification with localisation: in the 5 selected bounding boxes for an image, does it overlap >50% with ground truth box **and** is the box labeled correctly?

Task 2	Flat error
Localisation	0.425

→ **Just 8% error increase!**

Conclusions

- Adopted segmentation as selective search strategy for object localisation:
  - High recall: >96% with ~1,500 locations
  - Coarse locations are sufficient: bounding boxes
  - Fast to compute: <10s per image
  - Class-independent
  - Enables the use of bag-of-words features
- Doing localisation on top of classification increases error rate by just 8%

Come visit poster 42

Segmentation as Selective Search for Object Recognition, ICCV 2011,
K.E.A. van de Sande, J.R.R. Uijlings, T. Gevers, and A.W.M. Smeulders,
Poster #42, Wednesday 17:20-20:00