

# Hybrid Coding for Selective Search

*in the VOC detection challenge*

Koen van de Sande

Jasper Uijlings

Cees Snoek

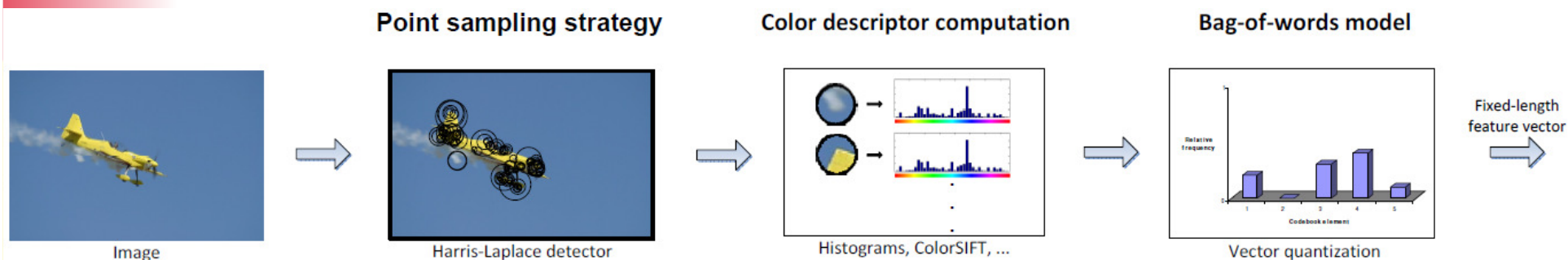
Arnold Smeulders

***University of Amsterdam***

PASCAL VOC Workshop 2012

October 12<sup>th</sup> 2012

# Lessons from Pascal VOC Classification



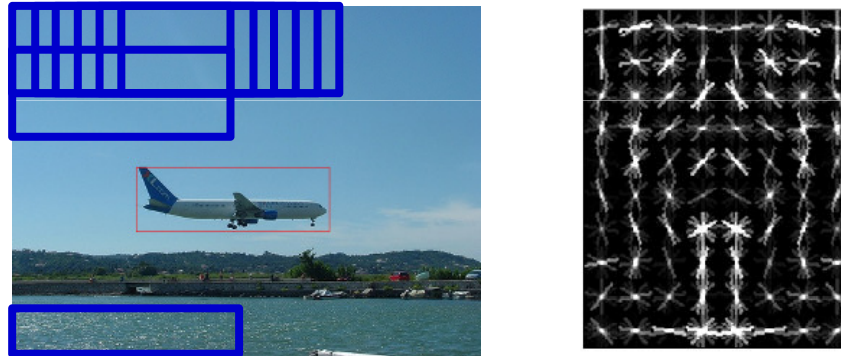
What works? [Zhang IJCV 2007, Song CVPR 2011]

- Ultra-dense sampling [Jurie ICCV 2005]
- Color descriptors [van de Sande TPAMI 2010]
- Soft, sparse and difference coding [Chatfield BMVC 2011]
- Efficient SVM solvers [Perronnin CVPR 2012]

Bag-of-words proven effective for classification

# Lessons from Pascal VOC Detection

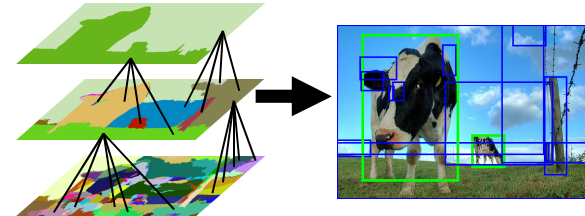
- Exhaustive search is state-of-the-art
  - Part-based [Felzenszwalb TPAMI 2010]
  - Improved by many [Zhang CVPR 2011] [Zhu TPAMI 2012]
  - Cheap features mandatory



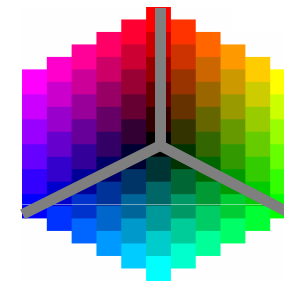
- Constrained search facilitates expensive features
  - Efficient subwindow search [Lampert TPAMI 2010]
  - Jumping Windows [Vedaldi TPAMI 2009]

# Our approach

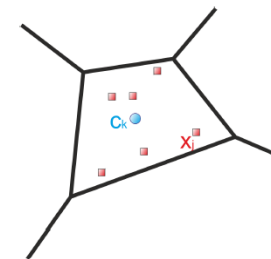
- Selective search



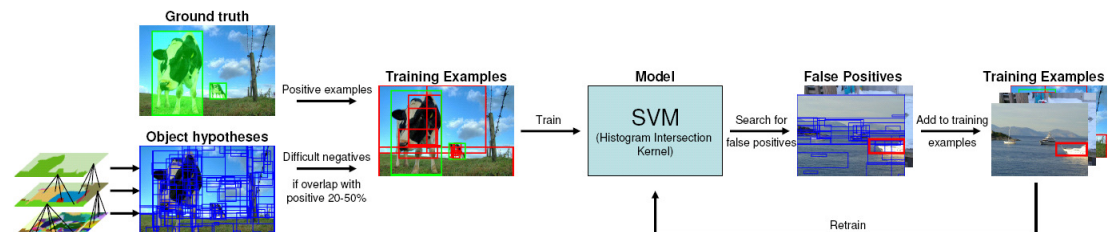
- New color descriptors



- Hybrid coding

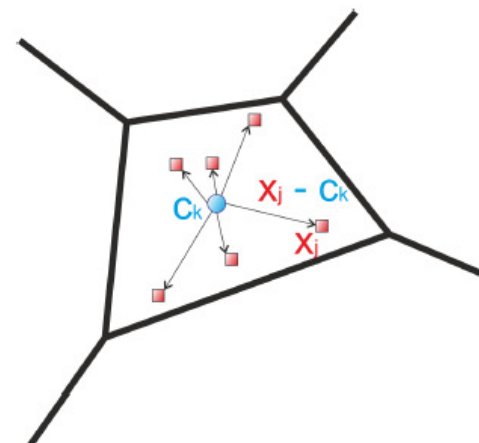
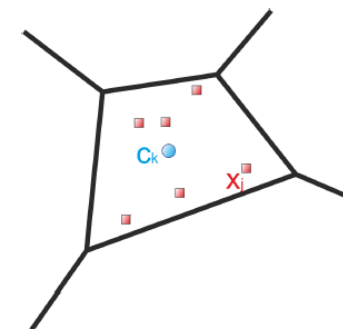


- Retraining



# Motivation Hybrid Coding

- Classical average coding
  - Small cells capture image details
  - Finesse within cell is lost
- Difference coding [Perronnin CVPR 2007]
  - Better maintain finesse within cell
  - Large cells compare wide range of image details
- Hybrid coding
  - Best of both encodings



# Hybrid Coding Details

## Classical Average Coding

PCA-reduce descriptor to 80D  
Find closest codeword  
Hard assignment  
L1 norm  
Dimensionality:  $k=4096$   
Spatial pyramid 1x1-2x2-3x3-4x4  
Histogram Intersection Kernel  
Fast Intersection SVM

## Difference Coding [Jegou TPAMI 2012]

PCA-reduce descriptor to 80D  
Find 4 closest codewords  
Weighed  $\{1, 1/2, 1/4, 1/8\}$  difference  
Power norm and L2 norm  
Dimensionality:  $k*d=20,480$   
Spatial pyramid 1x1-2x2-3x3-4x4  
Linear Kernel  
Optimized Cutting Plane SVM Solver

Product Quantization to handle longer vectors during training [Jegou TPAMI 2011]



# Selective Search

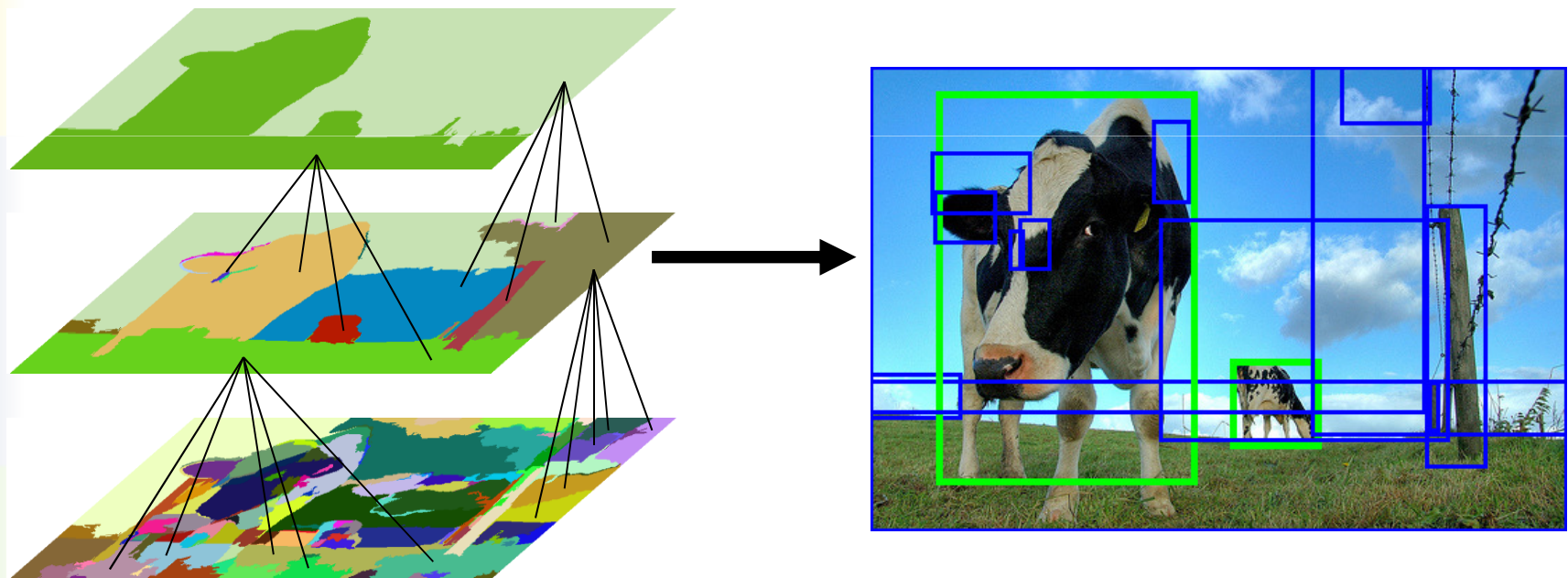
- Once discarded, an object will never be found again



- Image is intrinsically hierarchical
- Segmentation at a single scale won't find all objects

# Selective Search: Approach

- Hypotheses based on hierarchical grouping

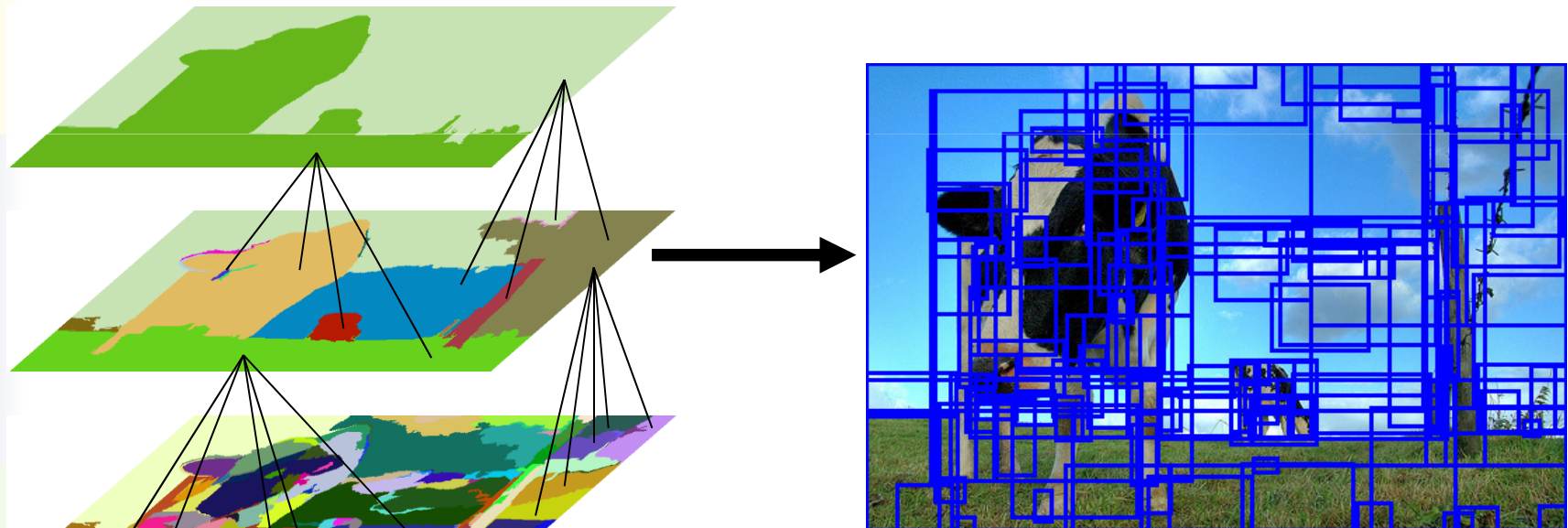


**Group adjacent regions on color/texture cues**



# Selective Search: Approach

- Hypotheses based on hierarchical grouping



Object hypotheses from all hierarchy levels

# Selective Search: Example



# Selective Search: High Recall



Color cues work best

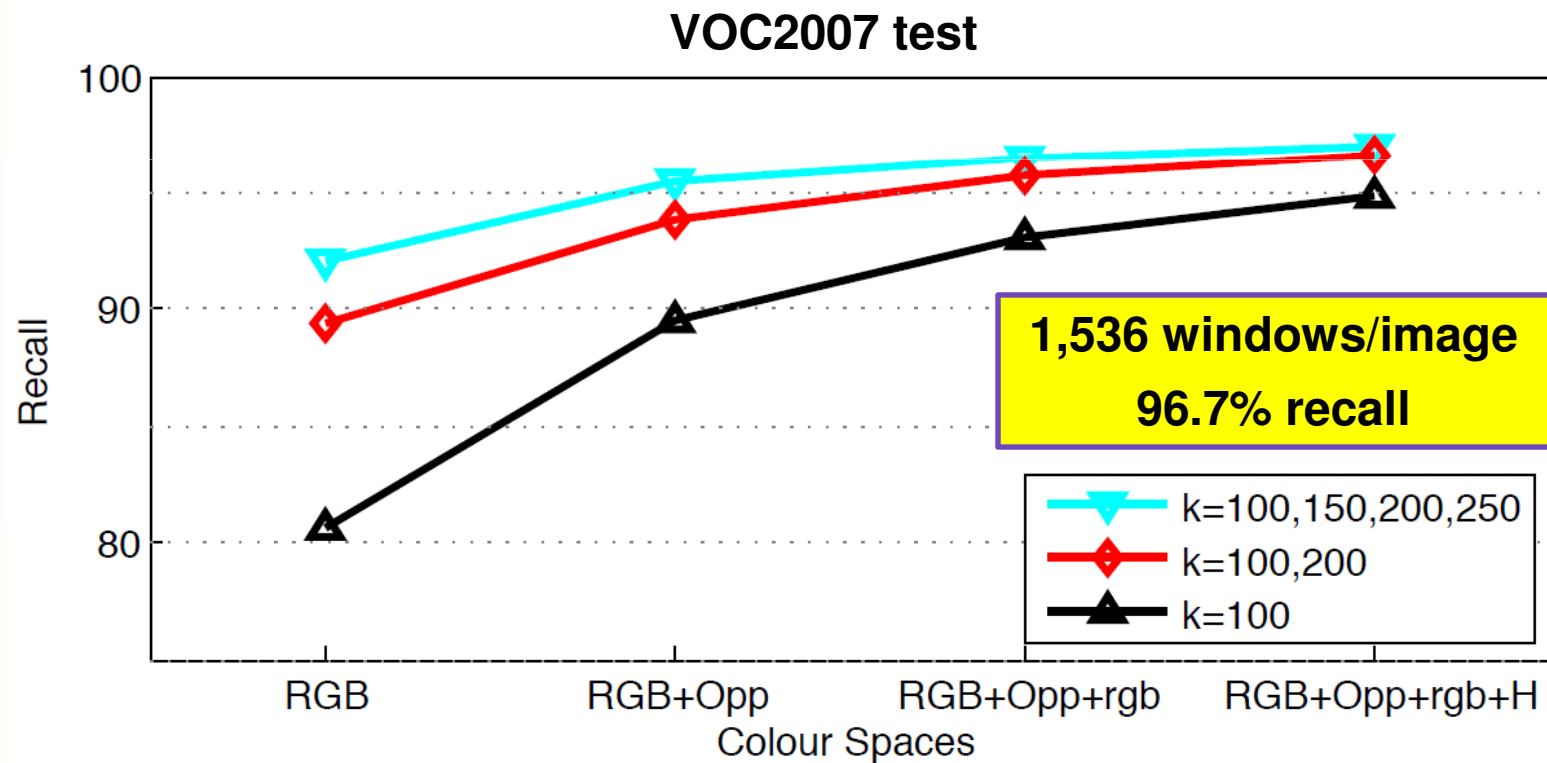


Texture cues work best

- No single segmentation strategy works everywhere
- **Complementary segmentation strategies needed**

# Selective Search

- Multiple complementary invariant color spaces
- Location hypotheses are class-independent



*Uijlings, submitted to IJCV, code available*

# New Color Descriptors

## PASCAL VOC 2007

1. C-SIFT
2. OpponentSIFT
3. RGB-SIFT
4. SIFT

van de Sande TPAMI 2010

## PASCAL VOC

1. X-SIFT
2. C-SIFT
3. OpponentSIFT
4. RGB-SIFT

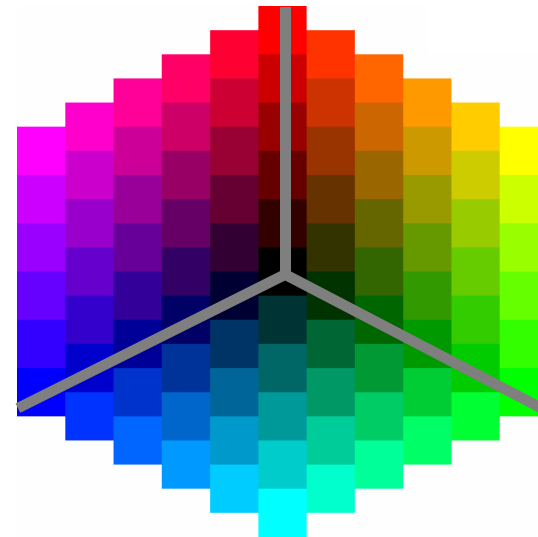
van de Sande, unpublished  
Technical Report in Nov 2012  
<http://koen.me/research/publications>

## X-SIFT

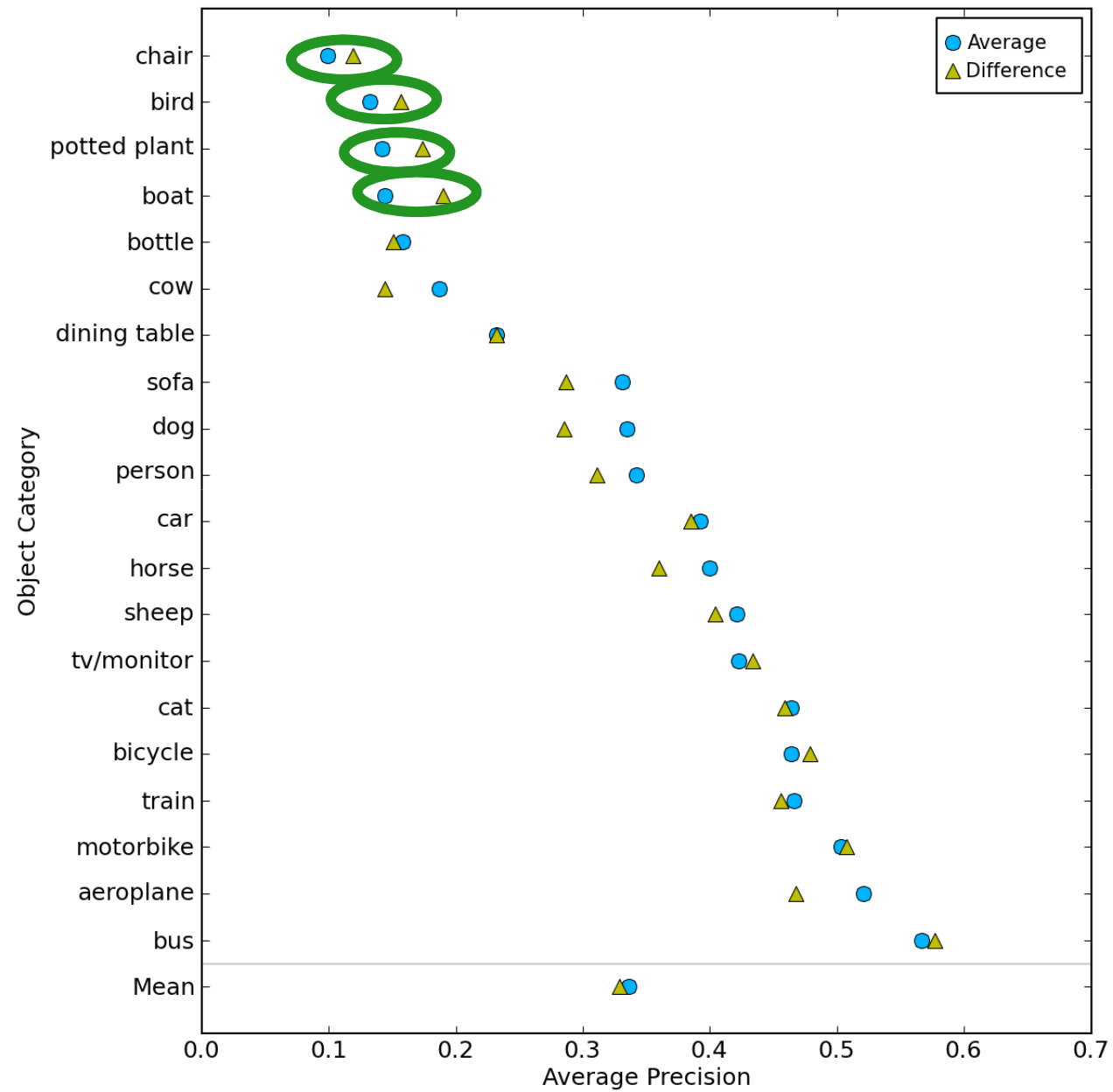
### ColorDescriptor 4.0

<http://www.colordescriptors.com>

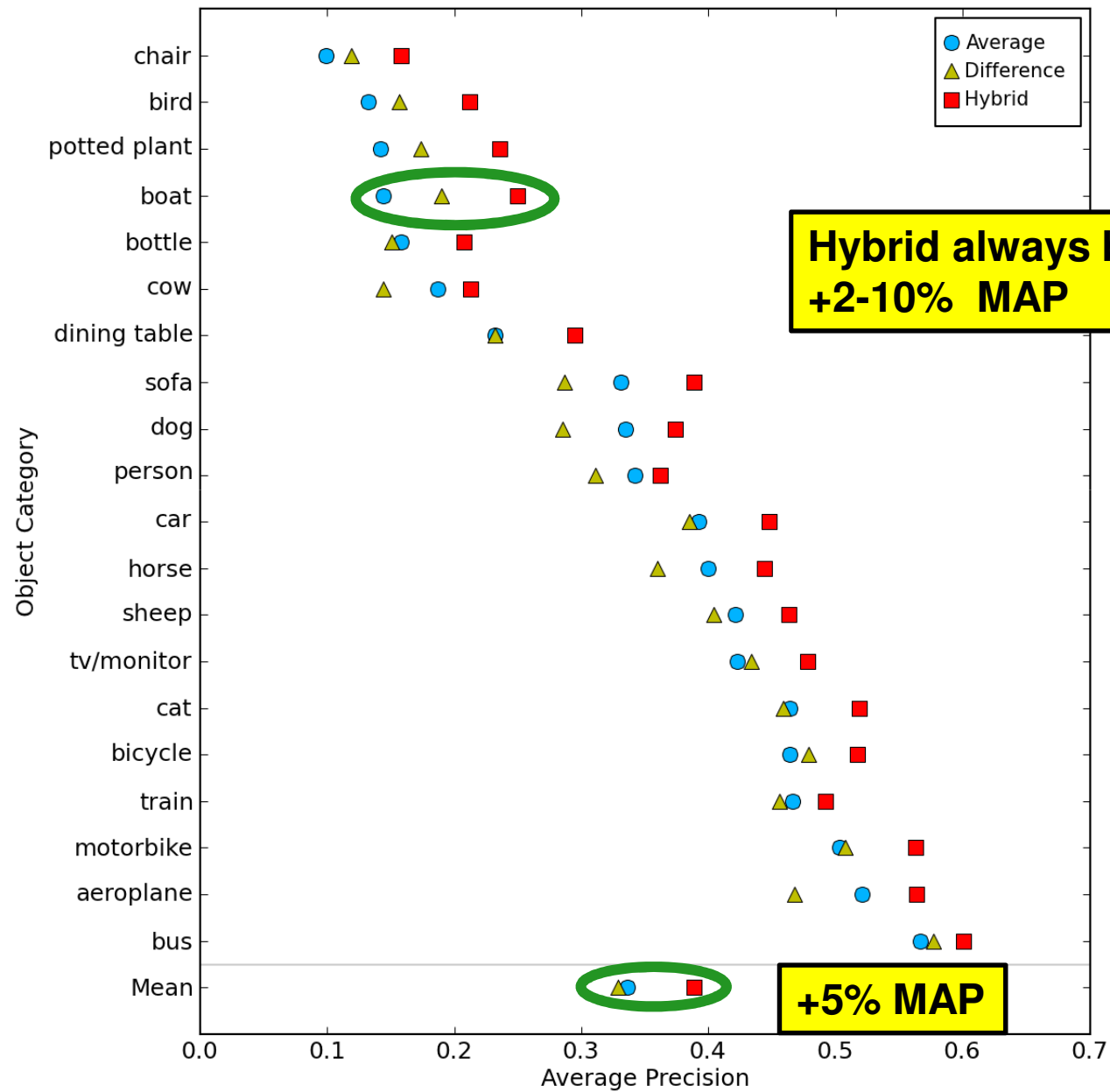
Includes GPU-accelerated CUDA version



# Quantitative Results on VOC2010val

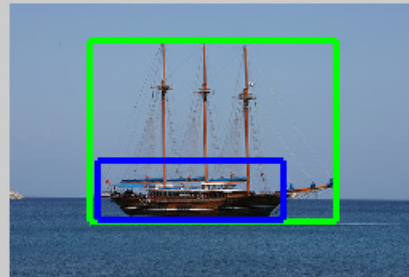
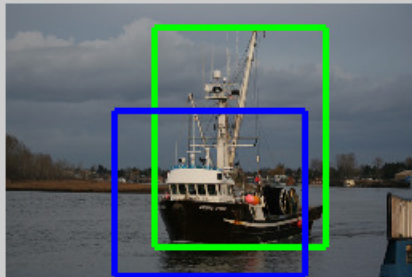
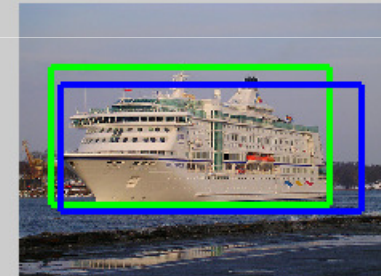


# Quantitative Results on VOC2010val



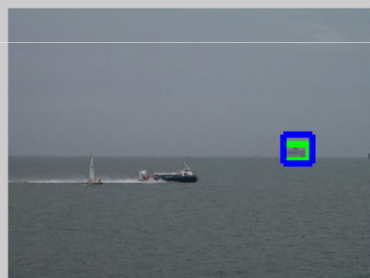


# Average coding

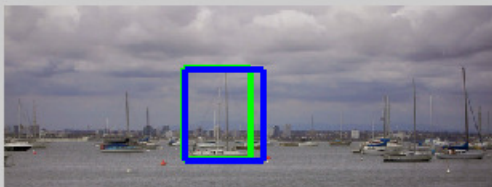




# Difference coding

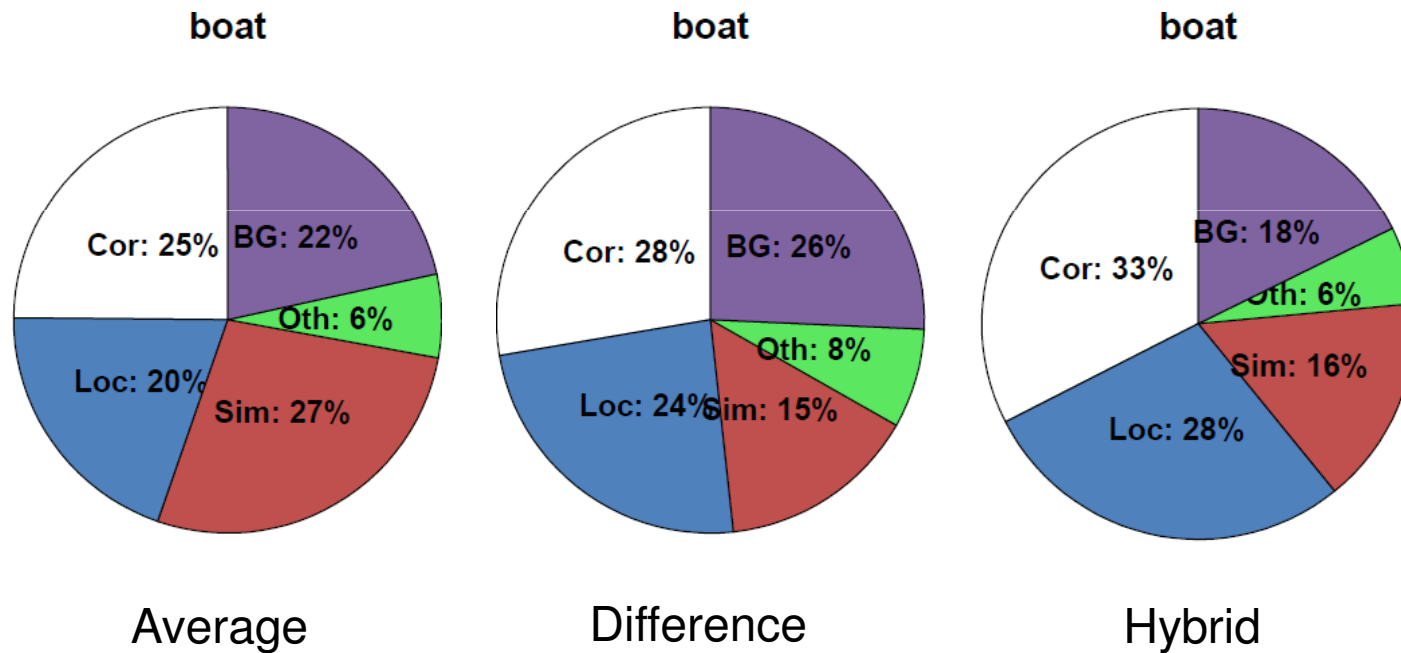


# Hybrid coding



# Boat Statistics

- Generated using [Hoiem ECCV 2012]



# Submission Details

- Adding classification helps [Harzallah ICCV 2009]
- Most Telling Window [Uijlings VOC 2011]
  - Detection system trained for classification

Scores added to Hybrid Coding output



# Conclusions

- Hybrid coding is effective for object recognition
- Selective search for few high quality object hypothesis

Thank you

