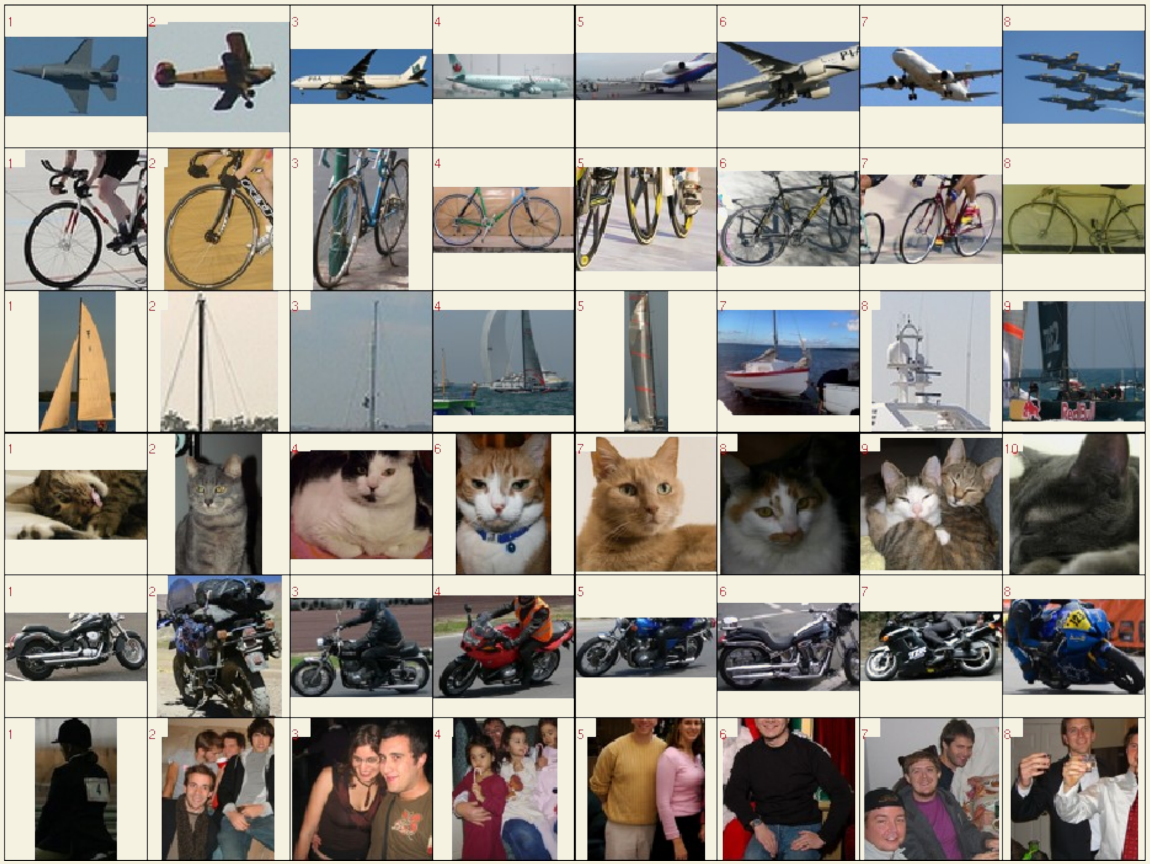
The Most Telling Window for Image Classification



Contributors:

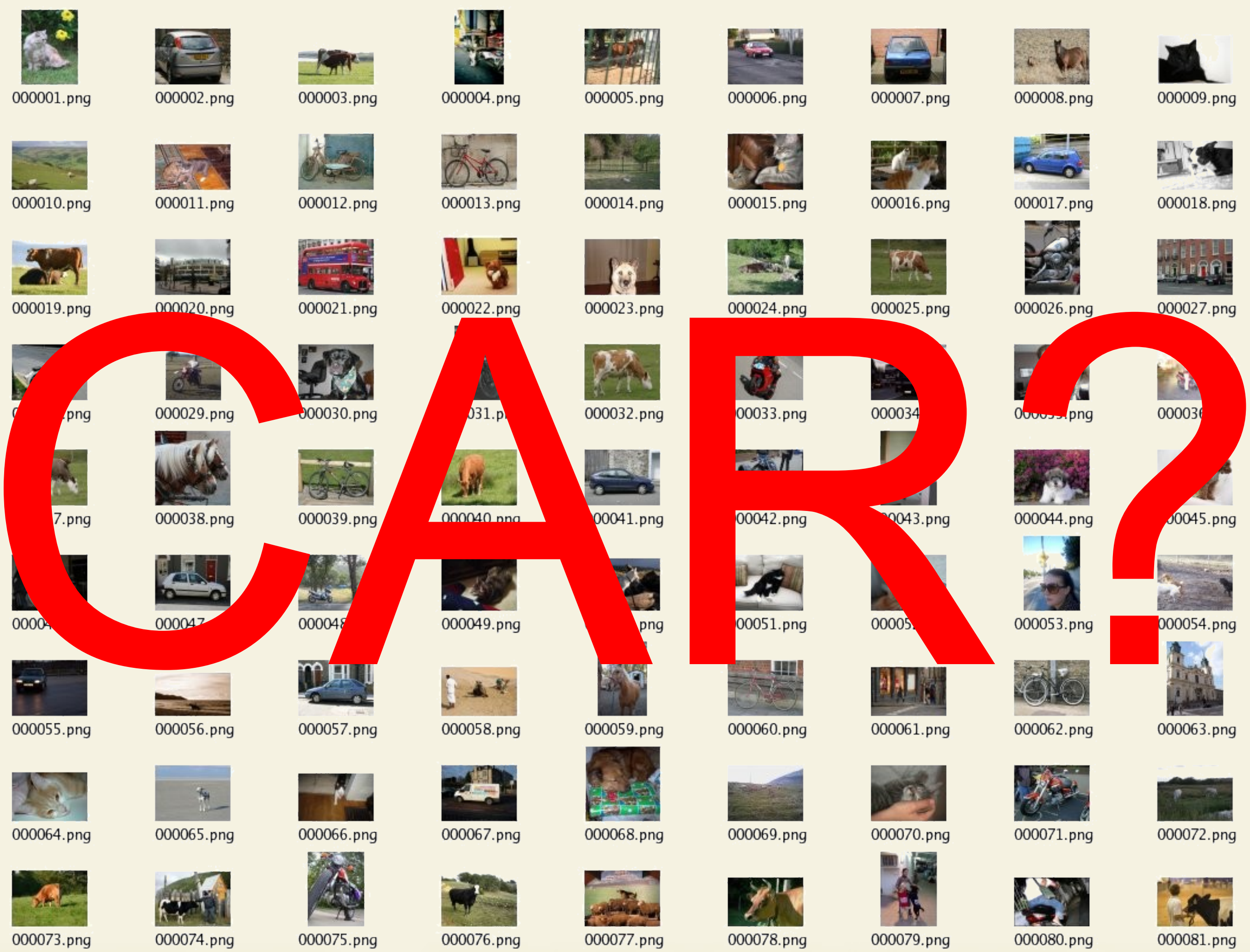
Jasper Uijlings
Koen van de Sande
Arnold Smeulders
Theo Gevers
Nicu Sebe
Cees Snoek

UNIVERSITY OF AMSTERDAM

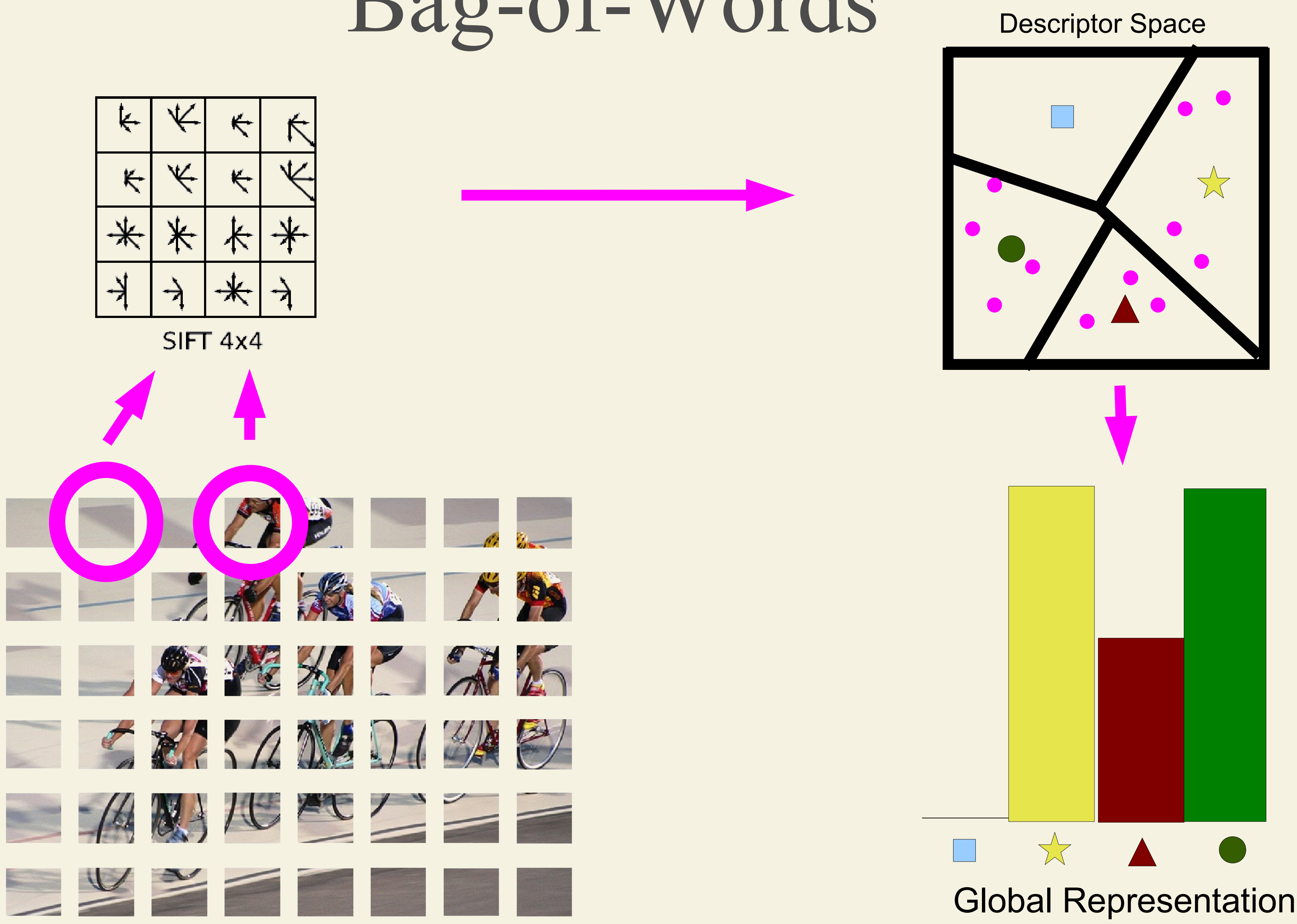
UNIVERSITY OF TRENTO - Italy

The Windows that Tell the Story of an Image, J.R.R. Uijlings and A.W.M. Smeulders. Under submission at TPAMI. Please contact jrr@disi.unitn.it before using this work.

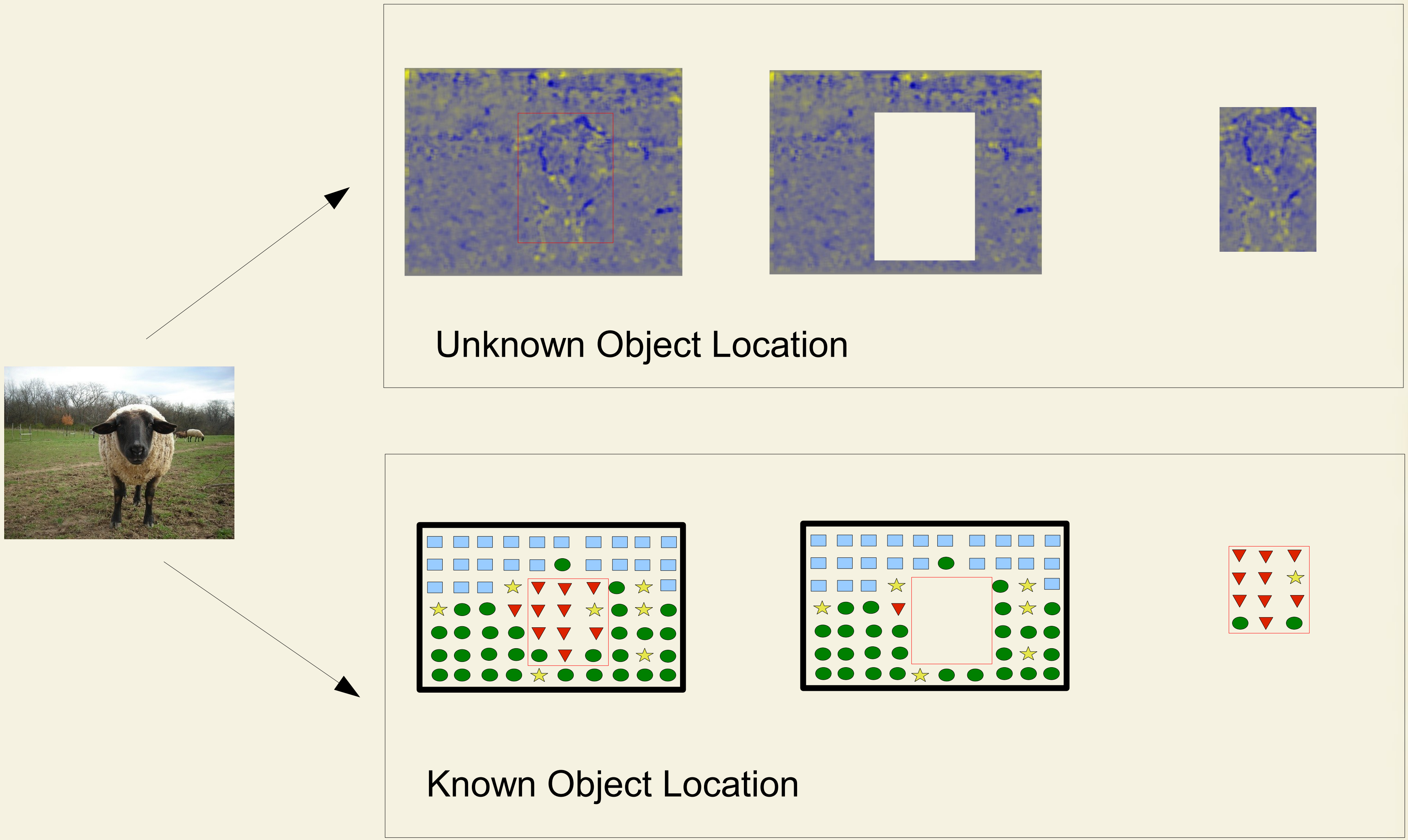
Image Classification



Bag-of-Words



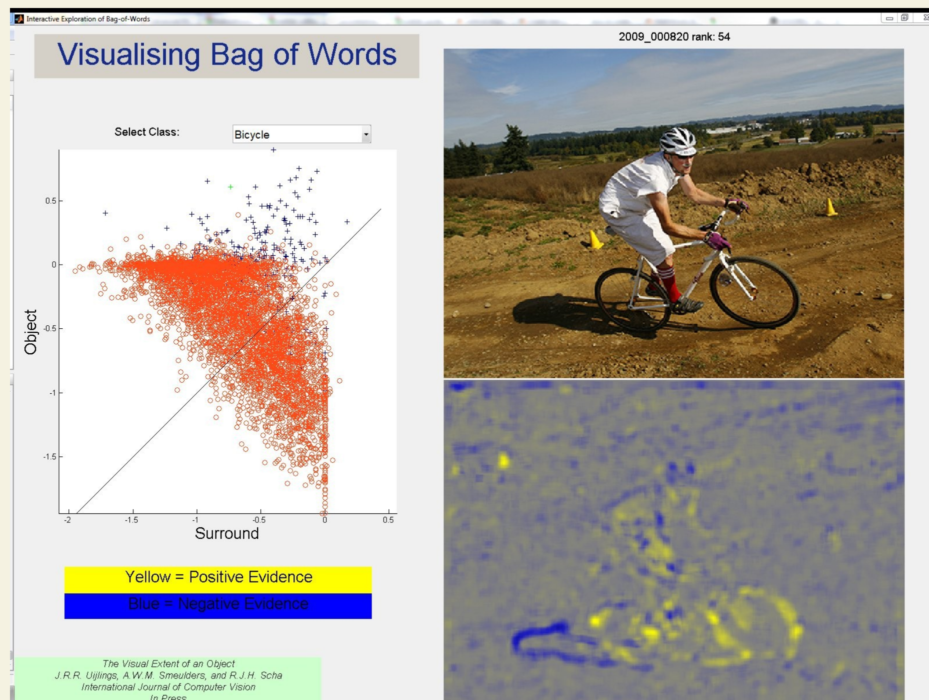
Is Global Optimal?



The Visual Extent of an Object. J.R.R. Uijlings, A.W.M. Smeulders and R.J.H. Scha, International Journal of Computer Vision, In press.

Is Global Optimal?

- Relevant conclusions:
  - The object alone yields significantly more accuracy than the whole image
  - Once the object location is known, context contributes very little.



Visualising Bag of Words
Demo @ ICCV
Tuesday 17:20 – 20:00

The Visual Extent of an Object. J.R.R. Uijlings, A.W.M. Smeulders and R.J.H. Scha, International Journal of Computer Vision, In press.

We need an explicit object location

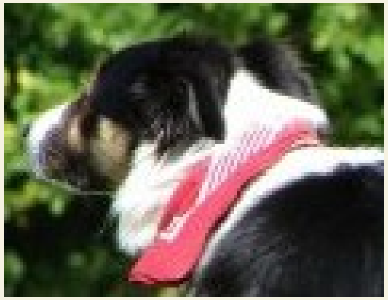
- It has been shown that object localisation can improve classification:

  "Combining Efficient Object Localization and Image Classification", H. Harzallah, F. Jurie, C. Schmid, CVPR 2009.

  Joint Winner Pascal 2008 Detection Challenge

Is Exact Localisation Optimal?

Is Exact Localisation Optimal?



Is Exact Localisation Optimal?

Parts were earlier used in "visual identification" to distinguish Bob's from Mary's Mercedes
Learning to Locate Informative Features for Visual Identification, IJCV 2008,
A. Ferencz, E. Learned-Miller, J. Malik

Is Exact Localisation Optimal?

Parts may be more discriminative because of pose change, often caused by interaction

Is Exact Localisation Optimal?

For occluded objects only the non-occluded part is informative.

Is Exact Localisation Optimal?

In crowded scenes, compared to an individual object:
a collection is both more easy to find and may be more discriminative

Is Exact Localisation Optimal: NO

- Parts may be more discriminative for some classes.
- Interacting objects may change pose, retaining typical appearance only for object part.
- Occluded objects are hard to find when searching for complete objects.
- In crowded scenes groups are more easy to recognize.

The Windows that Tell the Story of an Image, J.R.R. Uijlings and A.W.M. Smeulders. Under submission at TPAMI. Please contact jrr@disi.unitn.it before using this work.

The Most Telling Window

- May focus on:
  - Object Parts
  - Complete Objects
  - Object Collections

The Windows that Tell the Story of an Image, J.R.R. Uijlings and A.W.M. Smeulders.
Under submission at TPAMI. Please contact jrr@disi.unitn.it before using this work.

Methodology Most Telling Window

- Object Location
- General framework training/classification

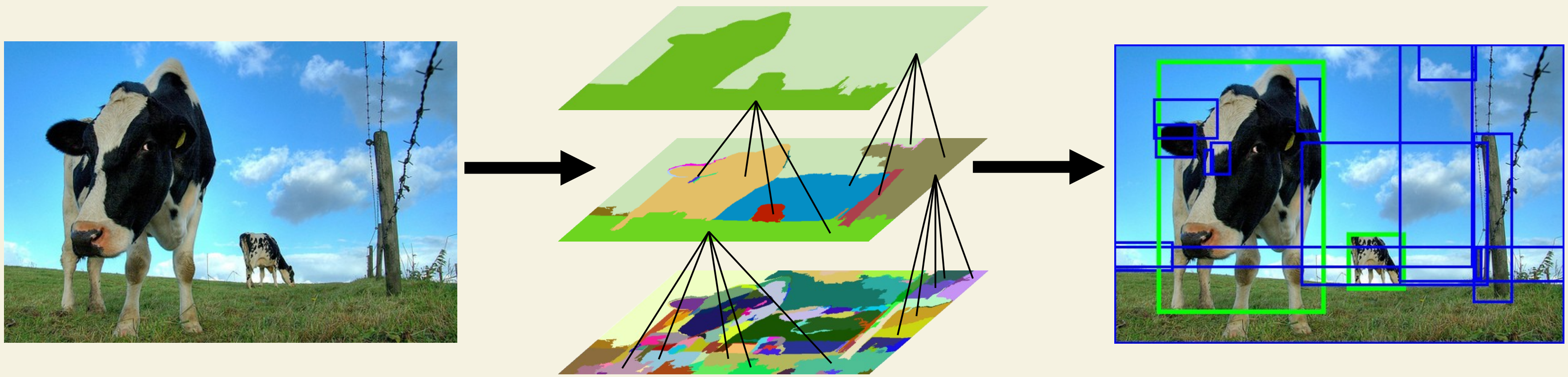
Methodology: Object Location

- Most Dominant: Sliding Windows.

- But yields 100.000 – 1.000.000 windows: infeasible for powerful Bag-of-Words implementation.
- Solution: Selective Search

Methodology: Object Location

- We introduce Selective Search



- Which uses multiple, complementary, hierarchical segmentations.

- More details in ILSVRC presentation

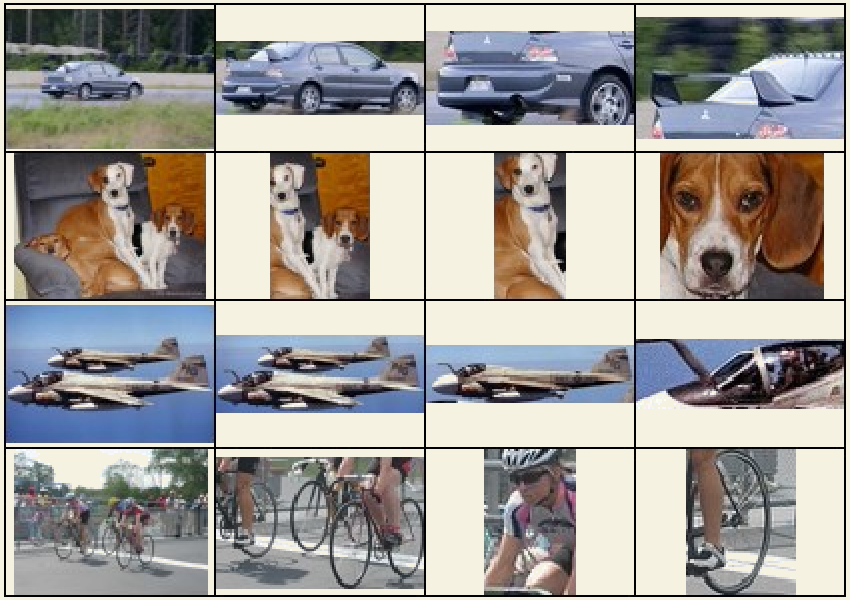
Segmentation as Selective Search for Object Recognition, ICCV 2011,
K.E.A. van de Sande, J.R.R. Uijlings, T. Gevers, and A.W.M. Smeulders,
Poster #42, Wednesday 17:20-20:00

Matlab pcode for selective search will be released soon.

Methodology: Object Location

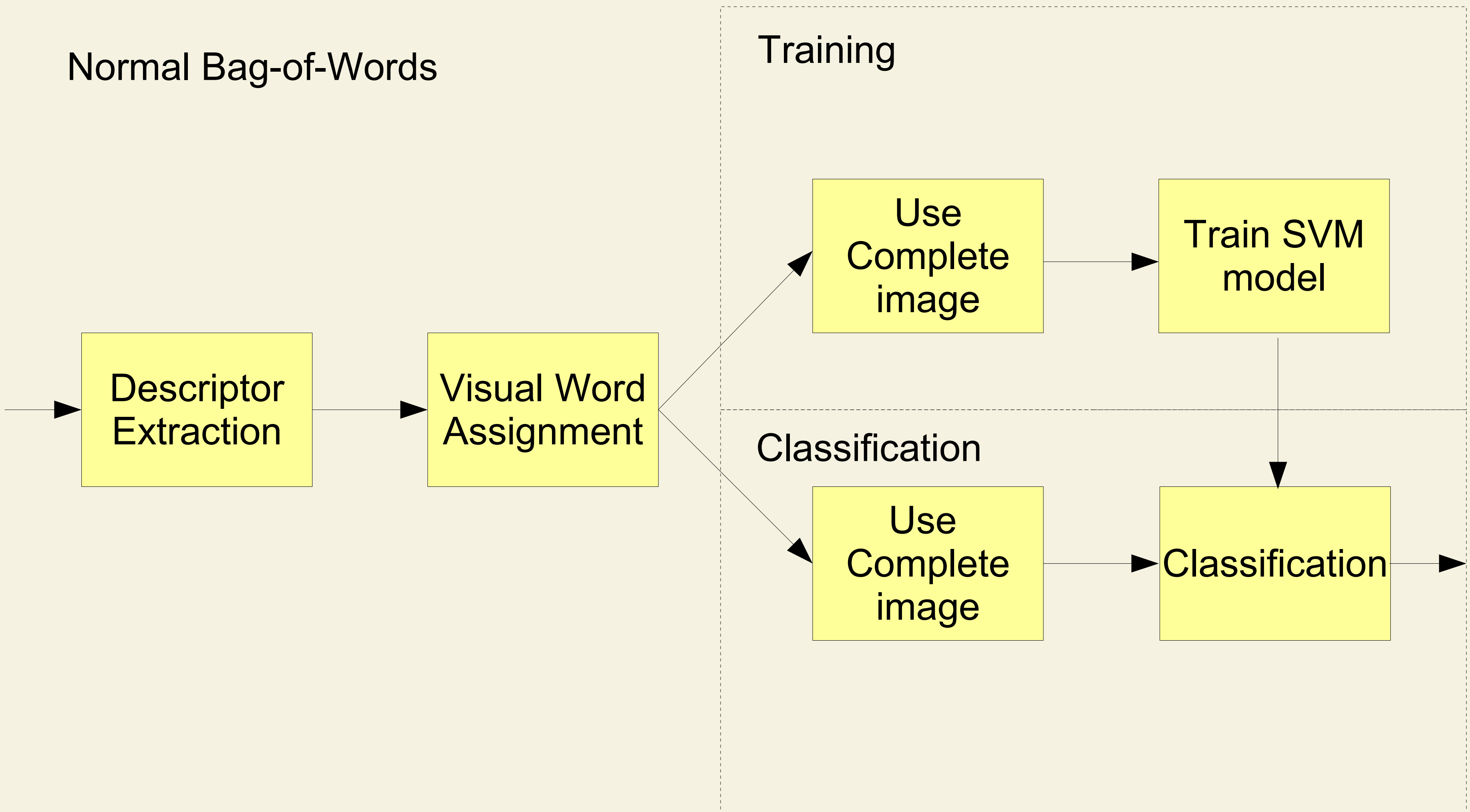
- Small set of class-independent locations
- Captures parts, objects, and collections

Example Windows generated by our method:

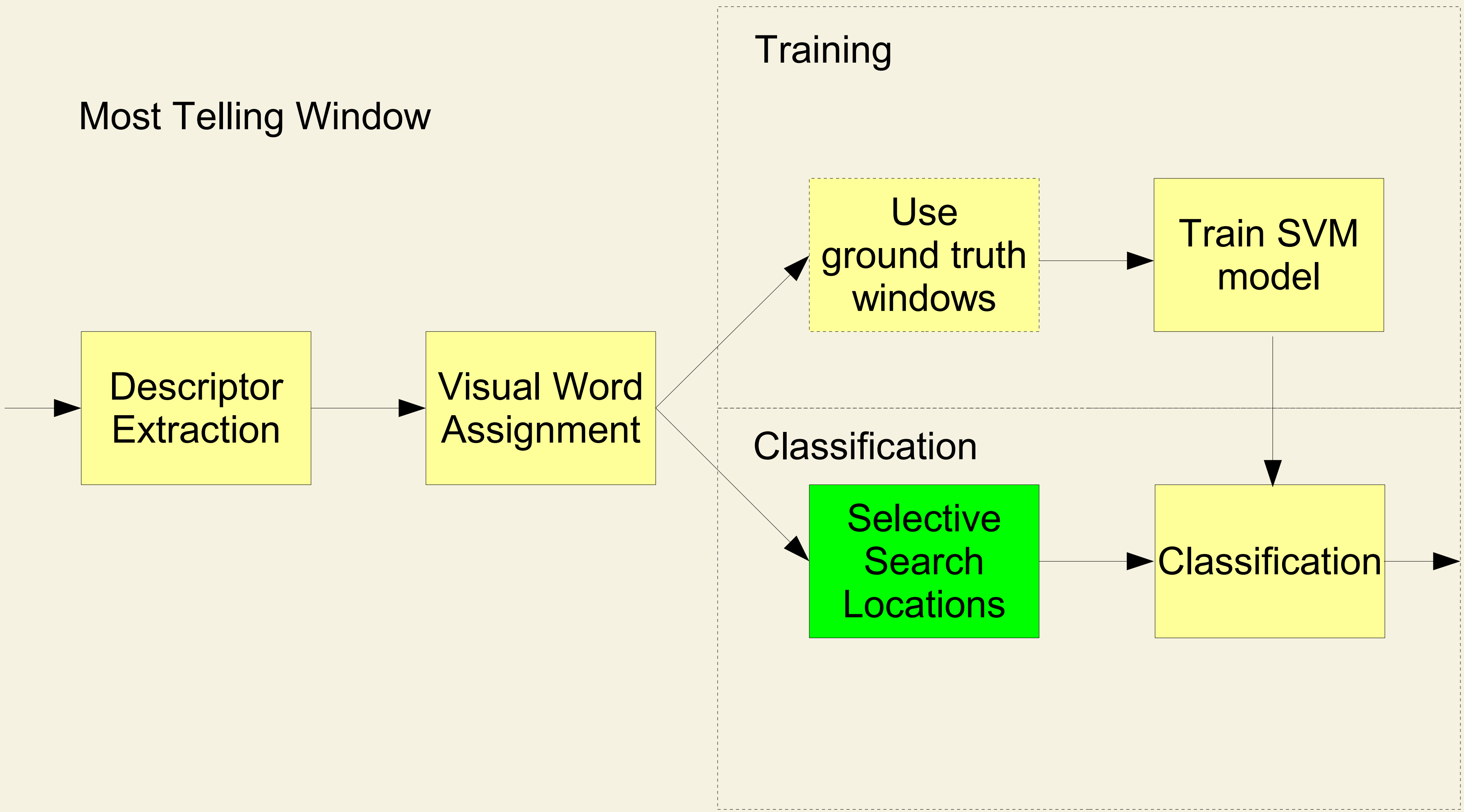


Segmentation as Selective Search for Object Recognition, ICCV 2011,
K.E.A. van de Sande, J.R.R. Uijlings, T. Gevers, and A.W.M. Smeulders,
Poster #42, Wednesday 17:20-20:00

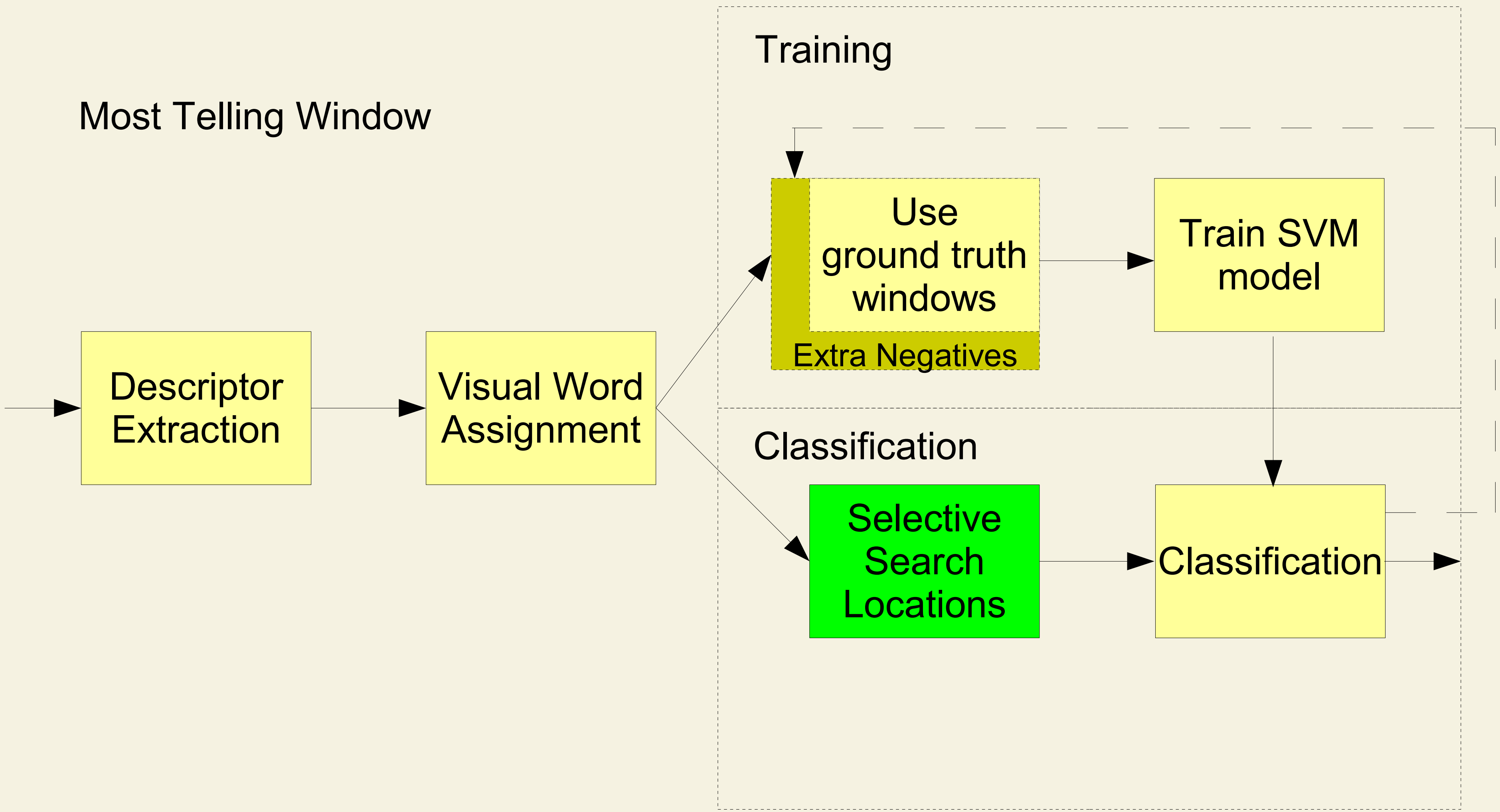
Methodology: Framework



Methodology: Framework



Methodology: Framework



Retraining: e.g. Laptev 2009, Felzenszwalb et al. 2010

Localisation vs Most Telling Window

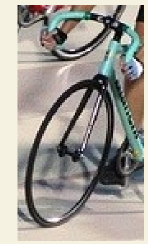
Localisation

Most Telling Window

No negative examples from positive images!

Localisation vs Most Telling Window

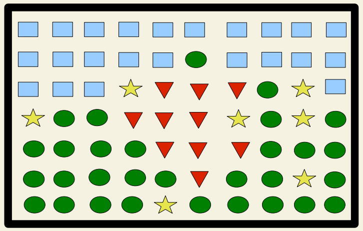
- Large difference in motivation:
  - Parts
  - Complete objects
  - Collections of objects



- Subtle difference in training windows
- Significant difference in final results
- (Of course, it would be better to also obtain new positive examples in retraining loop)

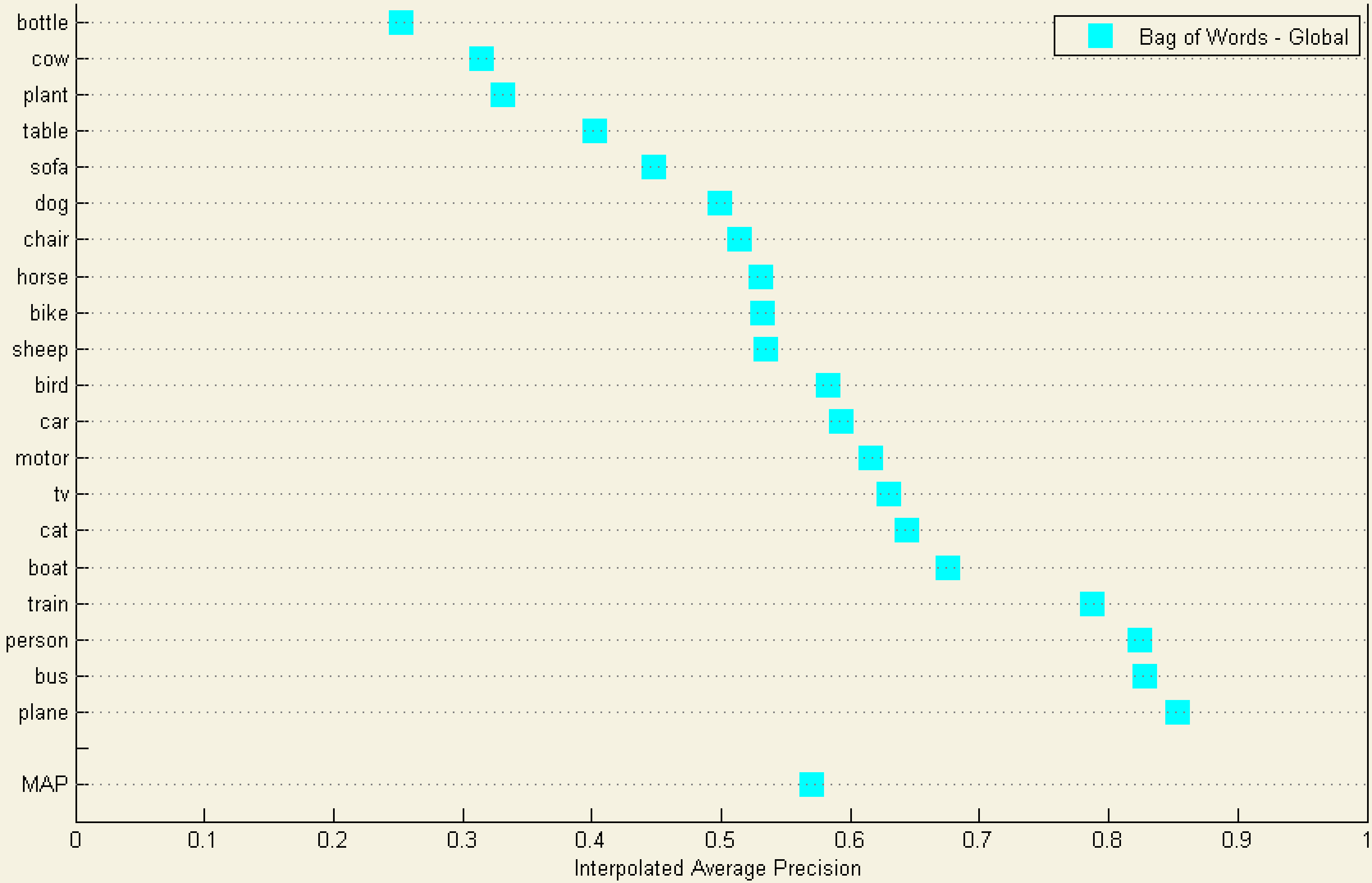
Implementation details

- Pixel-wise sampling
- (Colour) SIFT descriptors (Lowe04, Sande2010)
- K-means visual vocabulary
- Hard assignment.
- Store "Visual Word Images"



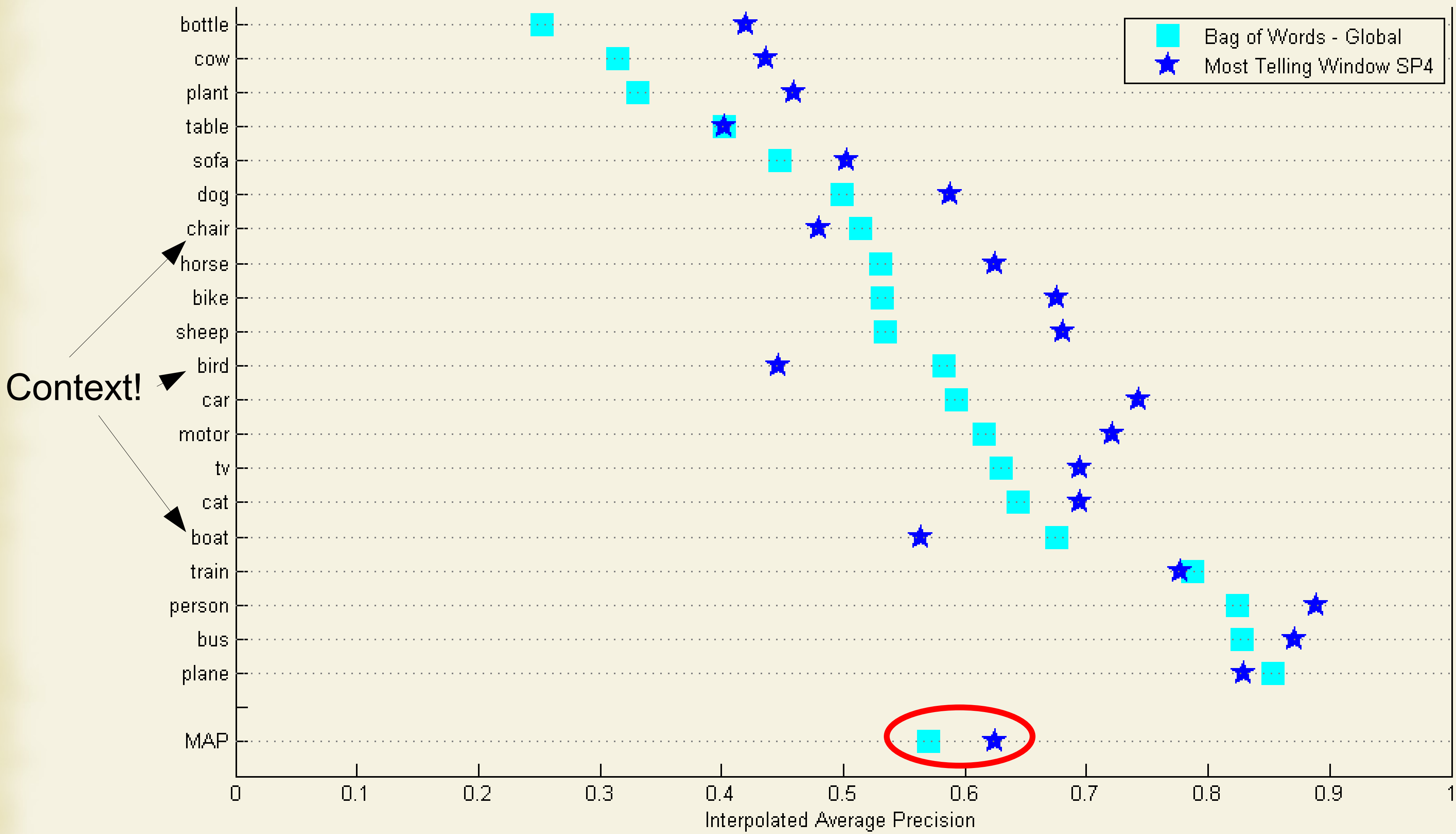
- Spatial Pyramid (Lazebnik06). BoW:1x1,2x2,1x3. MTW:2x2/4x4
- Bag-of-Words GPU acceleration (Sande2011)
- Selective Search (Sande 2011, Poster #42, Wednesday 17:20-20:00)
- Support Vector Machine with Histogram Intersection kernel. Fast additive classification (Maji 2009)

Results



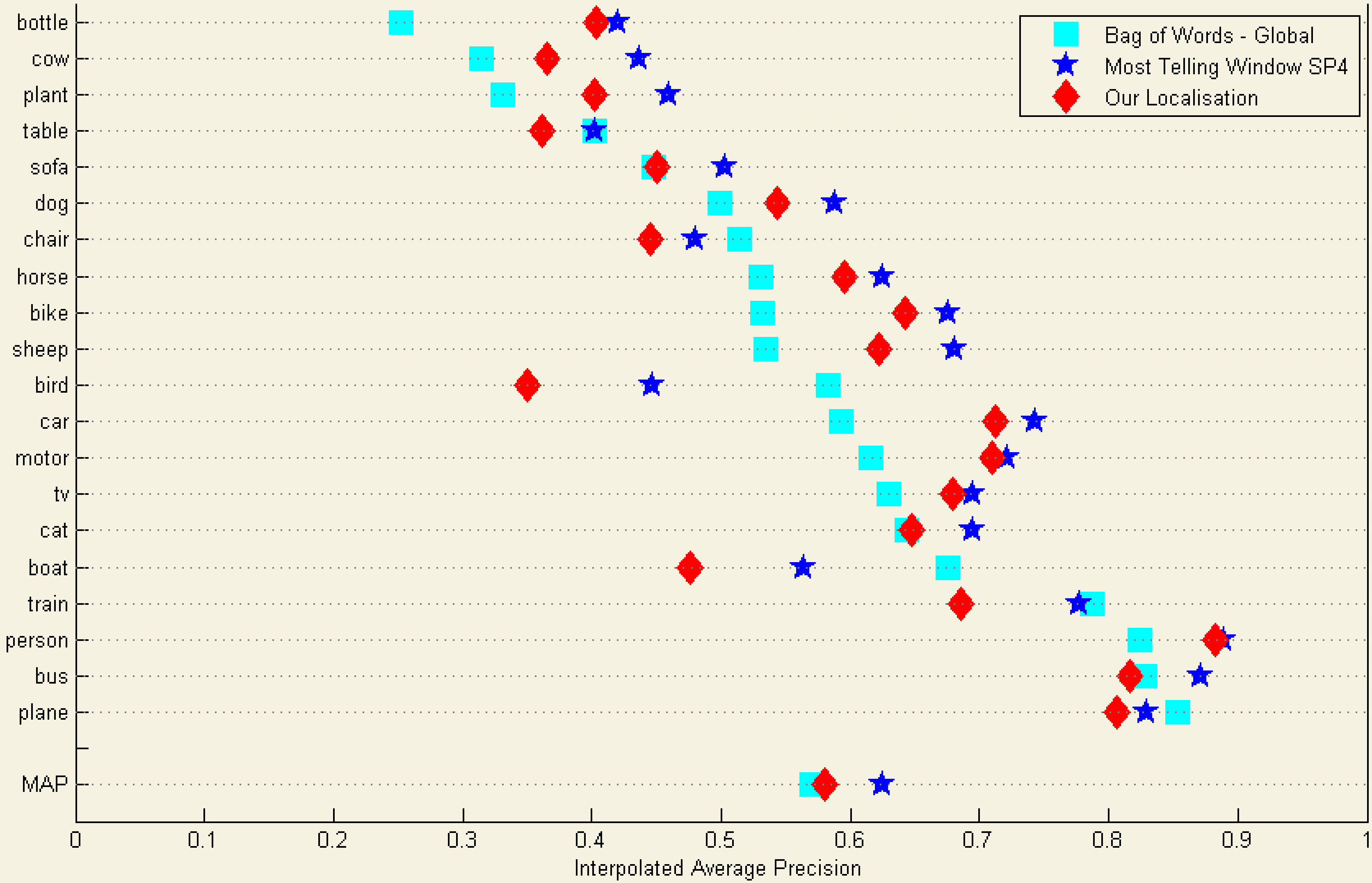
Comparable with top scores reported in e.g. Chatfield et al. BMVC 2011
- We: Pixel-wise sampling, 5 Colour SIFT (Sande 2010), kmeans vocabulary 4096
- Chatfield et al.: dense sampling, grey-SIFT only, Fisher/Sparse coding

Results



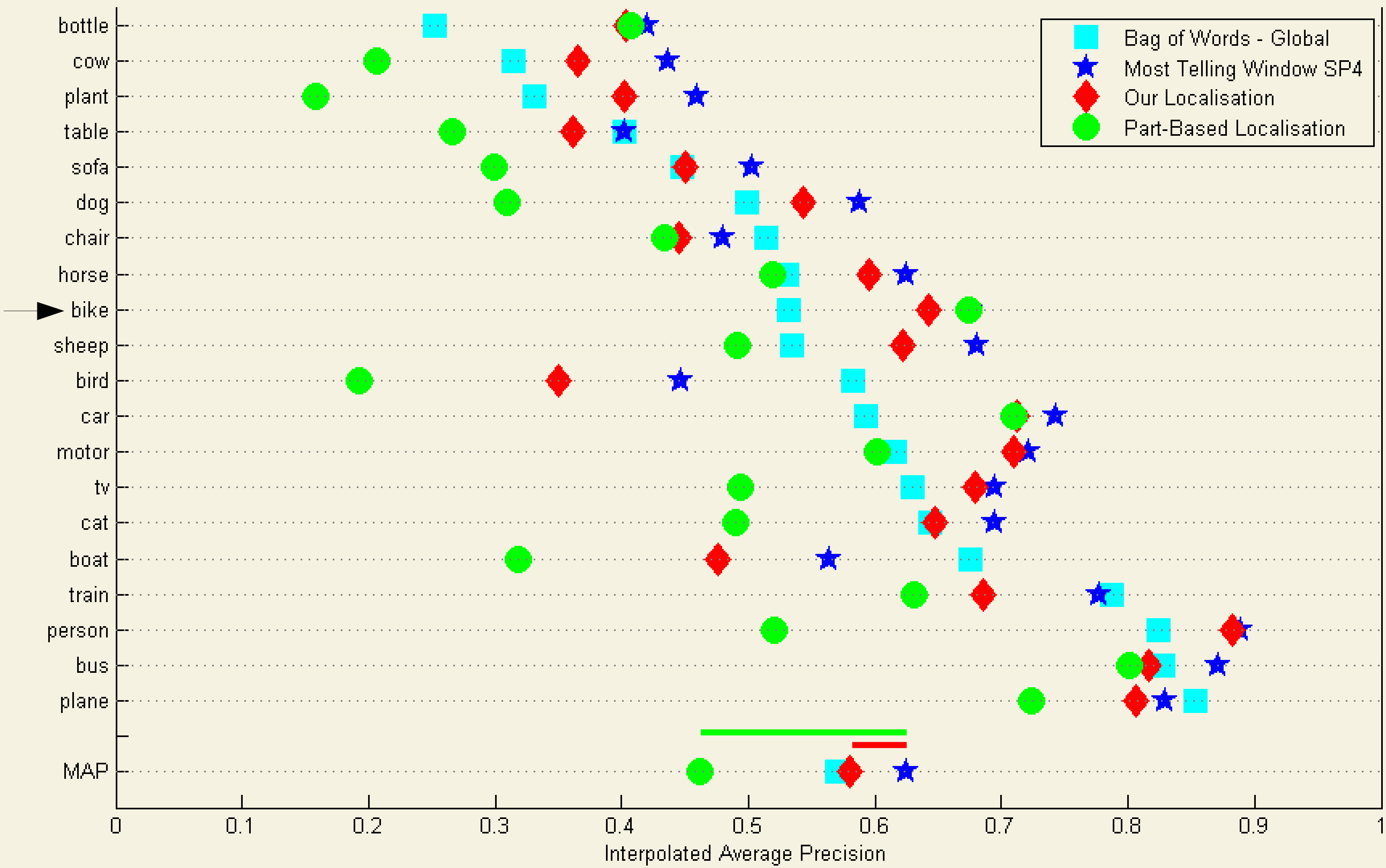
Significant improvement by using not the whole image but its Most Telling Window

Results



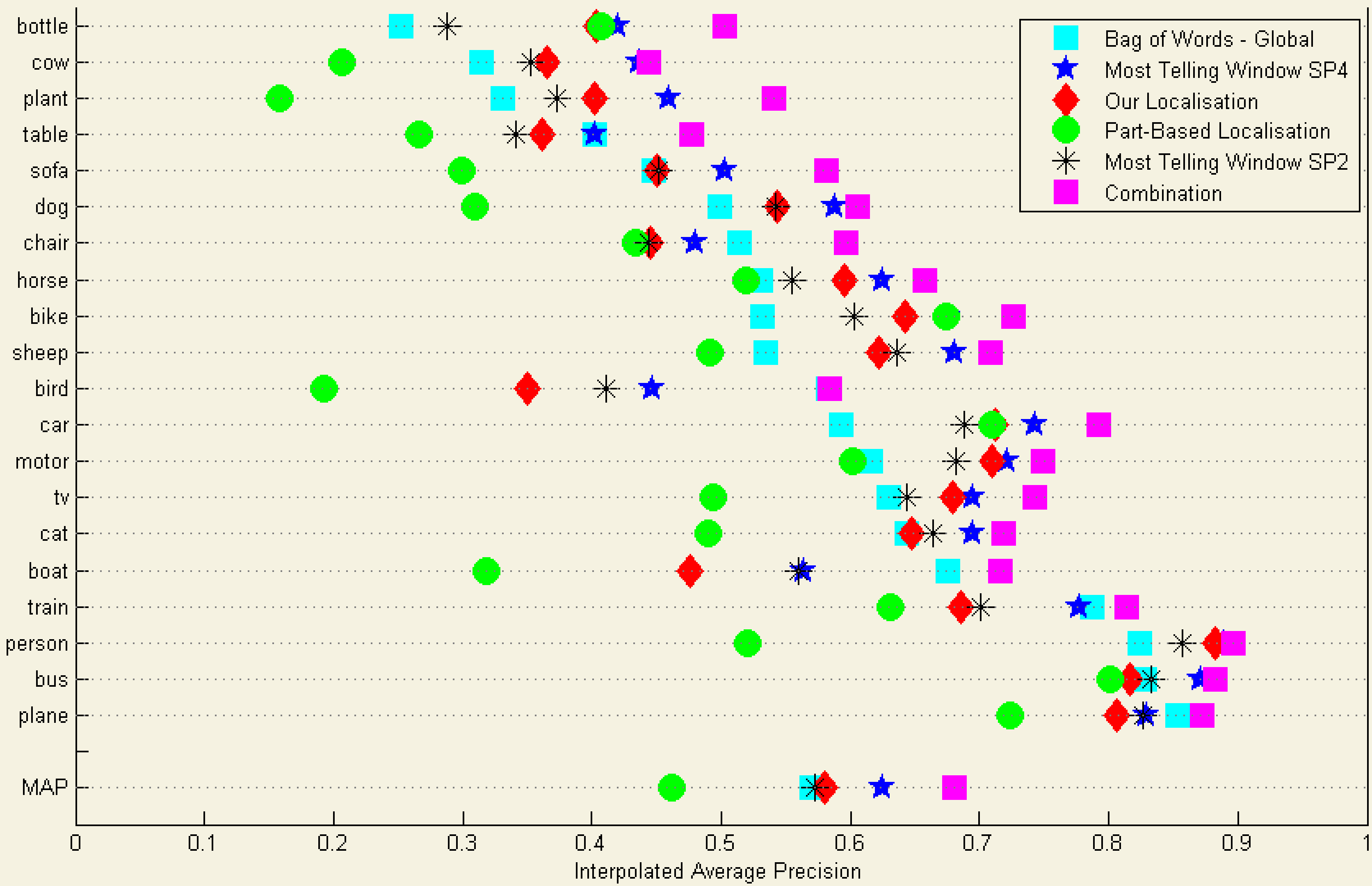
Most Telling Window consistently outperforms Exact Localisation (using same basic framework)

Results



Scores Detection Task: Felzenszwalb: 0.253 MTW: 0.317, Our localisation: 0.336, Discrepancy in results on detection and classification suggests that exact localisation tends to hallucinate objects that are not there while Most Telling Window finds object approximately.

Results



Final combination by cross-validation using weighted addition of classifier output:

- 2 parts Most Telling Window SP 4x4
- 1 part Most Telling Window SP 2x2
- 2 parts Localisation (Felzenszwalb 2010)
- 1 part global Bag-of-Words

3 variations of global Bag-of-Words and our exact localisation were discarded. Location is crucial!

Visualising the Most Telling Window of top-ranked images

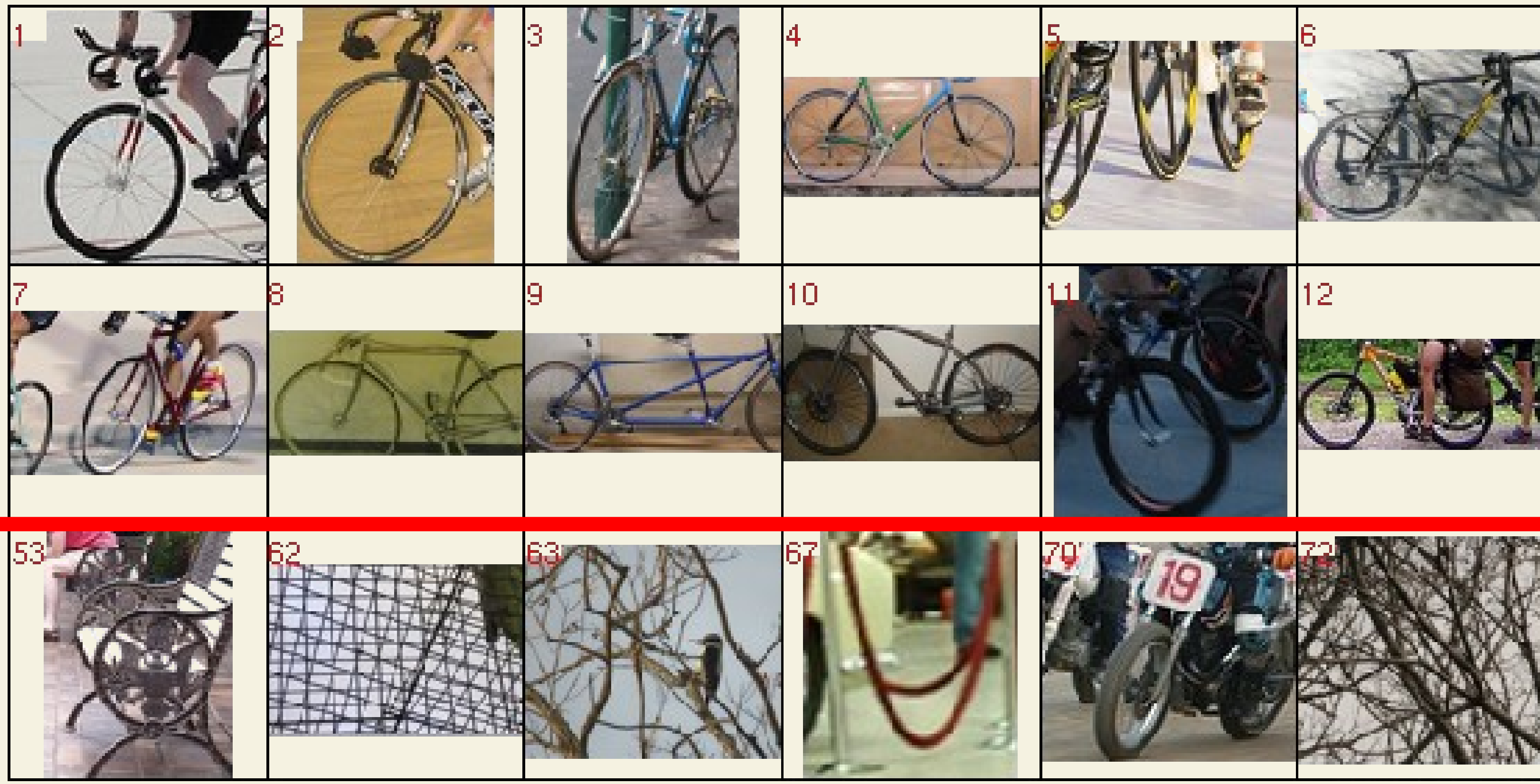
High-ranked Positives

High-ranked Negatives

Aeroplane

Visualising the Most Telling Window of top-ranked images

High-ranked
Positives



High-ranked
Negatives

Bicycle

Visualising the Most Telling Window of top-ranked images

High-ranked
Positives

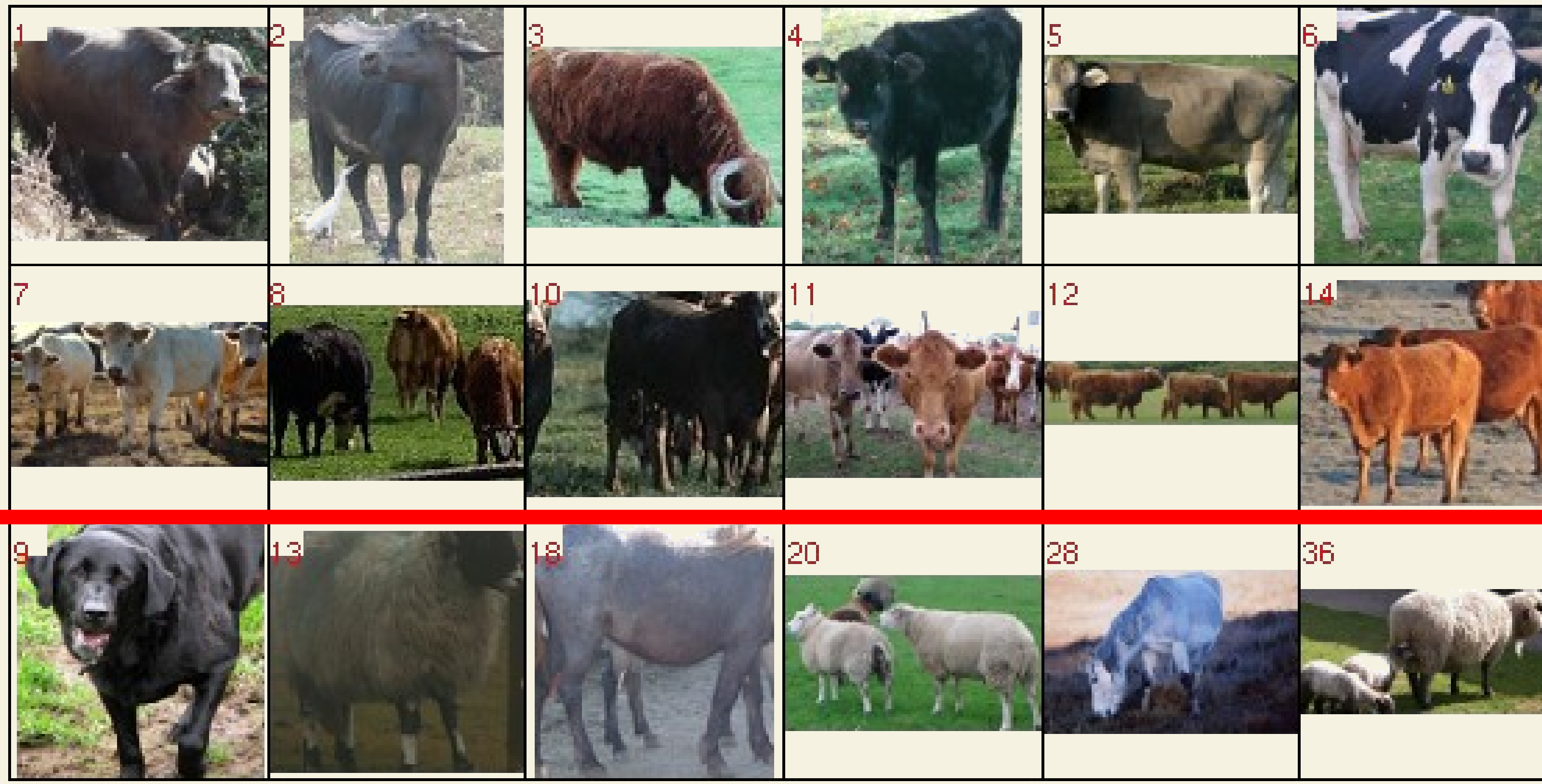
High-ranked
Negatives

Cat

Visualising the Most Telling Window of top-ranked images

High-ranked Positives

High-ranked Negatives



Cow

Visualising the Most Telling Window of top-ranked images

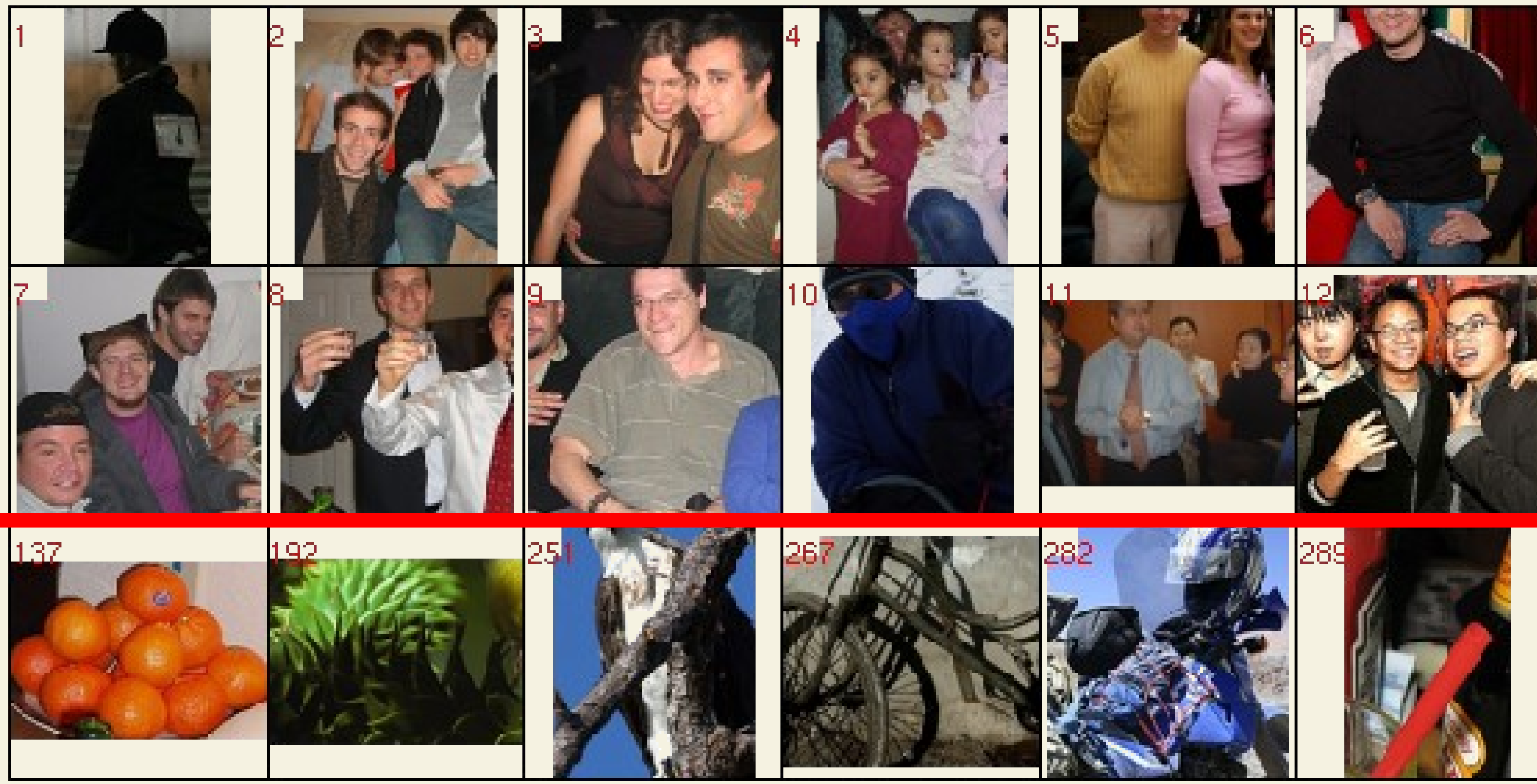
High-ranked Positives

High-ranked Negatives

Motorcycle

Visualising the Most Telling Window of top-ranked images

High-ranked
Positives



High-ranked
Negatives

Person

Pascal VOC 2011 Classification Challenge

	plane	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motor	person	plant	sheep	sofa	train	tv		MAP
NUSPSL_CTX_GPM	95.5	81.1	79.4	82.5	58.2	87.7	84.1	83.1	68.5	72.8	68.5	76.4	83.3	87.5	92.8	56.5	77.7	67	91.2	77.5		78.6
NLPR_SS_VW_PLS	94.5	82.6	79.4	80.7	57.8	87.8	85.5	83.9	66.6	74.2	69.4	75.2	83	88.1	93.5	56.2	75.5	64.1	90	76.6		78.2
NUSPSL_CTX_GPM_SVM	94.3	78.5	76.4	80	57	86.3	82.1	81.5	65.6	74.7	66.5	73.4	81.9	85.3	91.9	53.2	73.9	65.1	89.5	76		76.7
UVA_MOSTTELLING	90.1	74.1	66.5	76	57	85.6	81.2	74.5	63.5	62.7	64.5	66.6	76.5	81.2	90.8	58.7	69.3	66.3	84.7	77.2		73.4
MSRAUSTC_HIGH_ORDER_SVM	92.8	74.8	69.6	76.1	47.3	83.5	76.4	76.9	59.8	54.5	63.5	67	75.1	78.8	90.4	43.1	63.1	60.4	85.6	71.1		70.5
MSRAUSTC_PATCH	92.7	74.5	69.4	75.4	45.7	83.4	76.5	76.6	59.6	54.5	63.4	67.4	74.8	78.6	90.3	43	63.1	58.6	85.2	71.3		70.2
LIRIS_CLSDET	90	66.2	63.3	70.9	47	80.9	73.9	63.9	61.1	52.7	57.9	56.9	69.6	73.8	88.4	46.3	65.3	54.2	81.3	72.7		66.8
BPACAD_COMB_LF_AK_WK_NO	86.5	58.3	59.7	67.4	33.2	74.2	64	65.5	58.5	44.8	53.5	57	60.7	70.8	84.6	39.4	55.4	50.5	80.7	63.1		61.4
NLPR_SVM_BOWDET_CONV	83.8	69.8	47.8	60.5	45.4	80.5	74.6	60.4	54	51.3	45.3	51.5	64.5	72.6	87.7	35.9	57.7	39.8	75.8	62.7		61.1
LIRIS_CLS	88.3	56.2	59.3	68.6	33.2	76.6	62.2	64.5	55.3	42.6	55.1	56.2	61.9	70	82.5	37.3	56.4	48.3	79.6	64.7		60.9
SJT_SIFT_LLC_PCAPOOL_DET_SV	85.6	66.5	51.9	60.3	45.4	76.8	70.3	65.1	56.4	34.3	49.6	52.4	63.1	71.5	86.8	26.1	56.9	47.9	75.5	65.6		60.4
NLPR_SVM_BOWDET	82.9	69.4	45.4	60.1	46	80	75.1	59.9	54.9	50.7	43.3	49.9	63.4	72.2	88.1	36.1	57.1	37.7	75.2	58.5		60.3
BPACAD_CS_FISH256_1024_SVM	85	57	57.7	65.9	30.7	75	62.4	64.4	56.9	42.2	50.9	55.3	59.1	69.1	84.2	39.3	52.3	46.7	78.9	61.8		59.7
SJT_SIFT_LLC_PCAPOOL_SVM	83.2	52.5	49.3	59.6	26	73.5	58.2	64.4	52.1	36.6	44.9	52.1	57.8	63.8	78.1	19.1	52.8	44.1	72	57.4		54.9
JDL_K17_AVG_CLS	84.2	52	54.5	63.2	25.3	71.2	58	61.1	50.2	33.3	44.3	49.7	57.9	65.1	79.9	20.9	47.4	43	77.7	56.7		54.8
NANJING_DMC_HIK_SVM_SIFT	55.6	25.5	31	36.5	15.8	41.4	40	40.6	30	17.8	21.1	34	27	31	57.9	11.9	20.7	22.6	48.4	35.7		32.2
BUPT_NOPATCH	65.1	23.8	17.3	36	12.6	40.5	31.1	35.4	27.2	10.4	20.8	31.3	13.6	29.5	54.9	10.7	19.1	19.2	42.1	30.8		28.6
BUPT_ALL	61.5	11.9	12.4	29.7	8.7	30.6	18.4	23.6	21.6	5.8	14.8	18.5	7.1	12.3	47.7	7.2	15	9.8	18.8	19.2		19.7
NLPR_KF_SVM	10.5	9.1	10.7	6	6.5	7.2	13.3	12.2	11.5	9.5	5.6	16.7	8.6	6.6	38.9	5.3	15	5	8.3	5.4		10.6

Best score   Within 95% of best score

The top-3 each has a different focus for boosting classification performance:

1st NUSPSL: Focus on combination of exact localisation and classification
(Song et al. CVPR 2011)

2nd NLPR: Focus on vocabulary: Semi-semantic, Salient and Supervector coding.
(Huang et al. CVPR 2011)

3rd UVA/DISI: Focus on location: The Most Telling Window
(Uijlings and Smeulders, submitted to TPAMI, Sande et al. ICCV 2011)

Conclusions Most Telling Window

- The Most Telling Window is the window that is most discriminative for classifying the presence of an object. It can be an (1) Object Part. (2) Whole Object. (3) Object Collection.
- First time that window within the image yields better results by itself than whole image?
- The Most Telling Window works better than exact localisation.
- Suboptimal positive windows suggest room for improvement.
- Selective Search enables powerful, local Bag-of-Words

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

  - Class independent parts, wholes, and collections.

The Windows that Tell the Story of an Image, J.R.R. Uijlings and A.W.M. Smeulders. Under submission at TPAMI. Please contact jrr@disi.unitn.it before using this work.