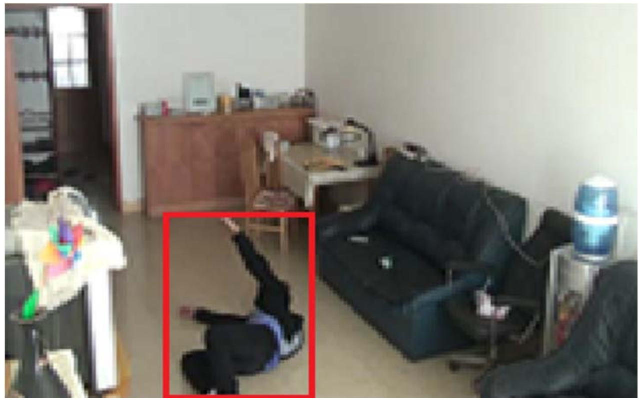
Object detection based on data decomposition, spatial mixture modeling and context

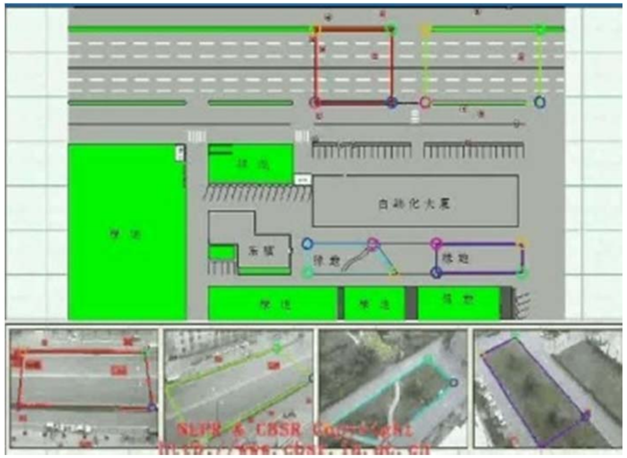
Junge Zhang, Yinan Yu, Yongzhen Huang, Chong Wang, Weiqiang Ren, Jinchen Wu
Advisors: Kaiqi Huang and Tieniu Tan

Intelligent Recognition & Digital Security Group
National Laboratory of Pattern Recognition,
Institute of Automation, CAS

Background

- Object detection is a fundamental function for visual surveillance and video analysis





Background

- *Video surveillance, Human-Machine Interface, Multimedia analysis…*

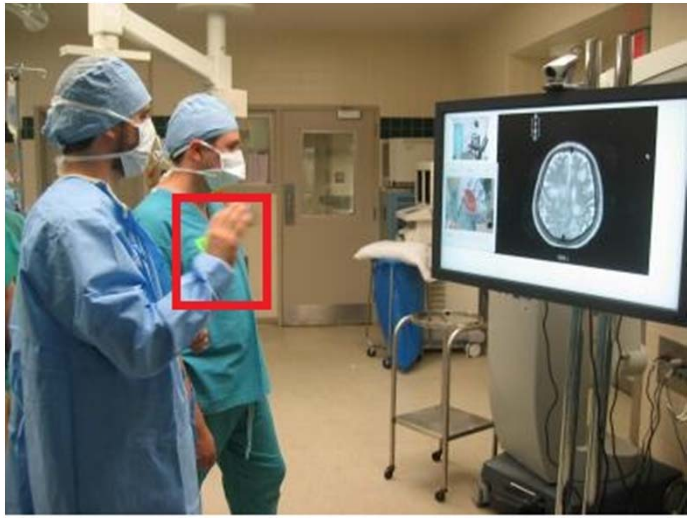
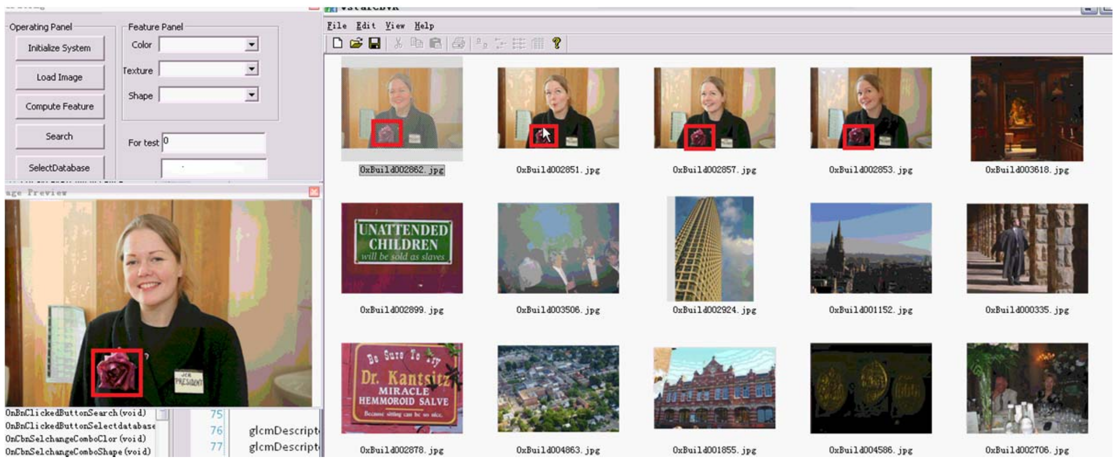


Photo: Courtesy of Ben-Gurion
University of the Negev, Israel



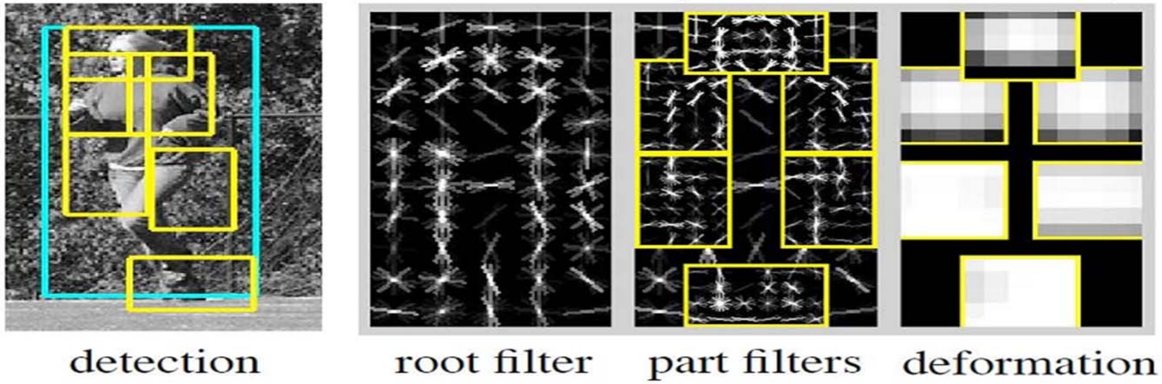
Object retrieval

Methods

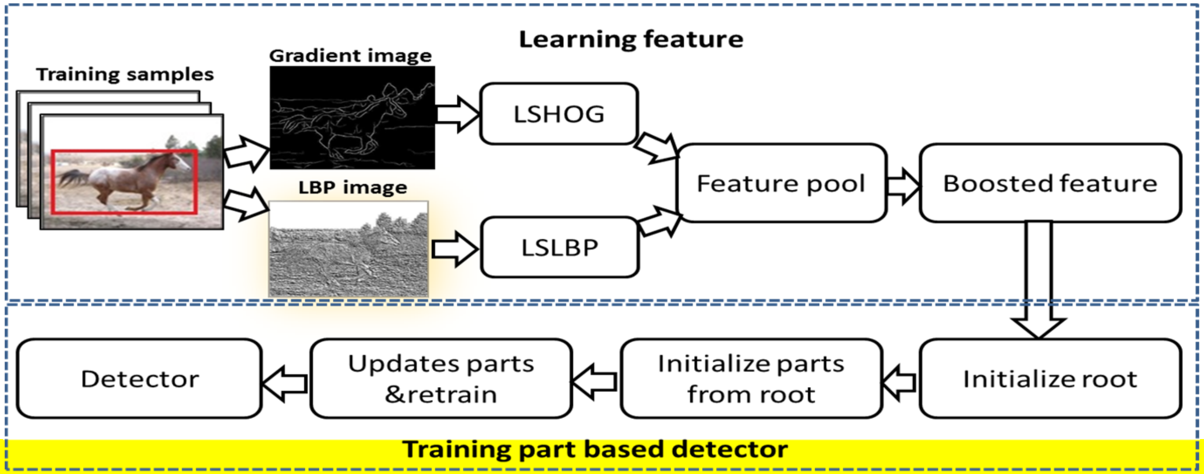
- Data Decomposition (DD) for part based model
- Spatial Mixture Modeling (SMM)
- Context Learning (NLPR_CLS)

Data decomposition for part based model

- Previous work
  - Felzenszwalb et al's deformable part based model (DPBM)[1]



  - Zhang et al's boosted LSHOG-LBP for DPBM[3]



Data decomposition for part based model

- Limits
  - The computational complexity is very large, especially when it is extended to enhanced models via multiple features, more flexible components or part models.
    - Moreover, it brings the risk of over-fitting when the length of those models becomes longer and longer
  - The original part based model is not “deformable” enough.
  - How to tackle these problems?

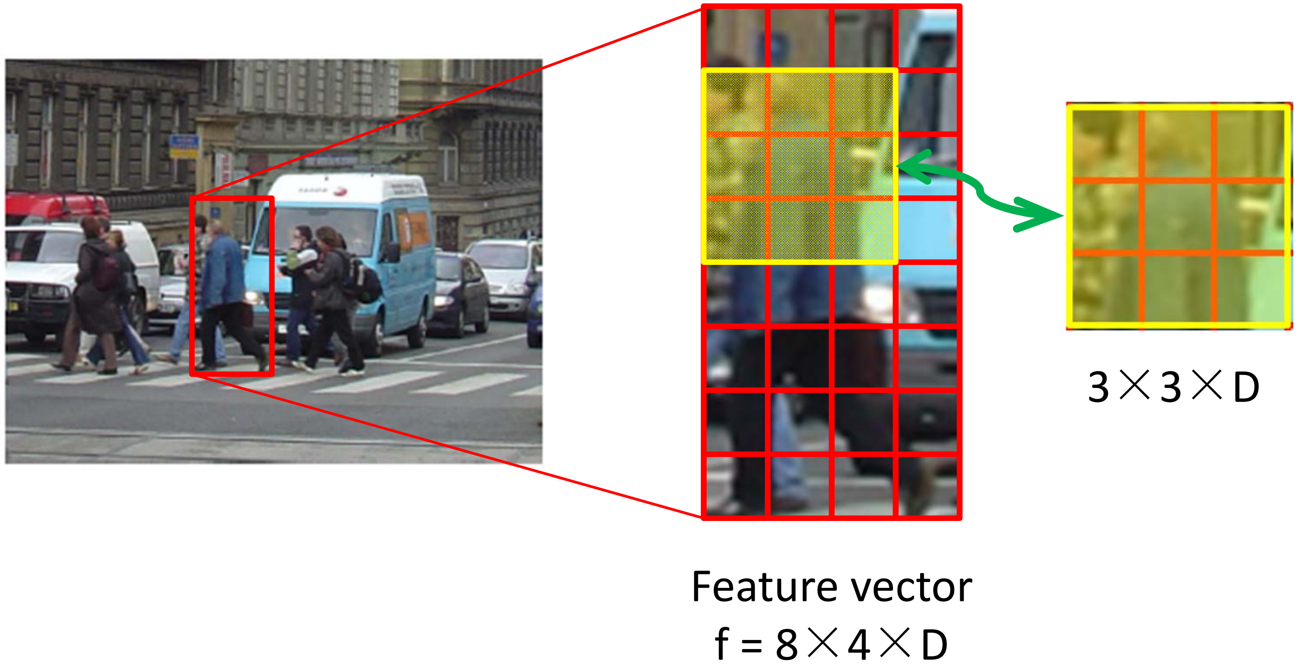
Data decomposition for part based model

- Pedersoli et al's work[2] indicates the inference cost is
- $$O\left(P\frac{L}{\delta^2}\left(D+\frac{L}{\delta^2 c}\right)\right)$$
- The dimension of filters and search space dominates the computation cost.
- Coarse-to-fine inference: reduce the search space
- Our goal and solution
  - Reduce the dimension of filters.
  - First, we'd like to maintain the high accuracy of those enhanced models via multiple features, more flexible components or parts. On the other hand, the time and memory cost should be reduced.

Data decomposition for part based model

- Decompose original filter space into lower dimensional space.
- Reconstruct the original filter with low cost

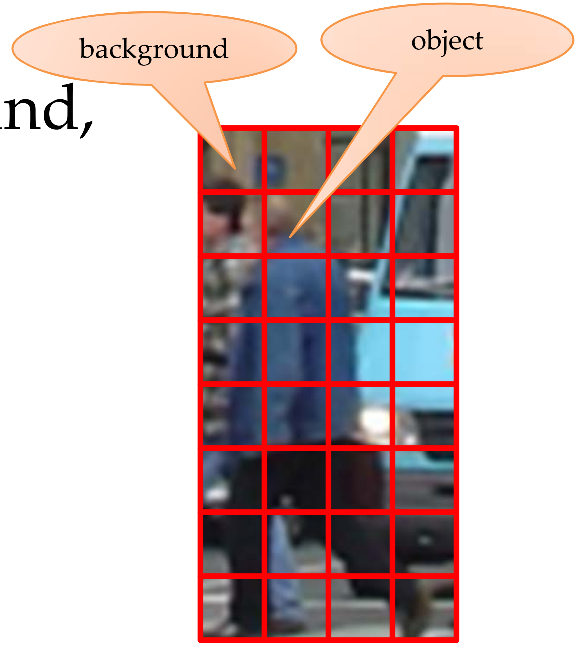
Data decomposition for part based model



- Data decomposition on root filter level or cell level?
- Cell level

Data decomposition for part based model

- Cell level data decomposition
- But, some cells correspond to background, others to objects
- Therefore, the decomposition method should be unsupervised, label independent, efficient
- PCA is one of those choices.

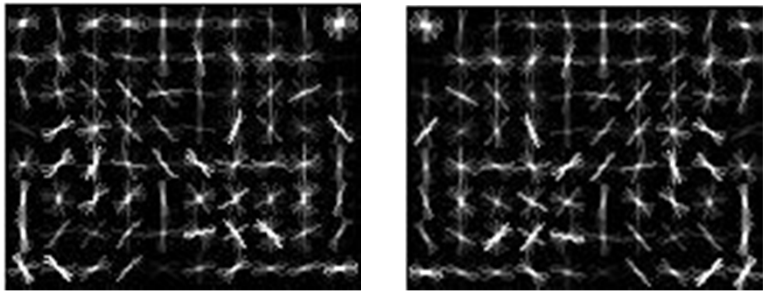


Data decomposition for part based model

- With PCA, we decompose each cell filter into a lower dimensional space.
- As known that, PCA is sensitive to the scaling of variables coming from different sources of data. (In our system, HOG and LBP are utilized)
- Data preprocessing

Data decomposition for part based model

- Similar to original DPBM, models are trained for horizontal symmetry to avoid over-fitting.



- How to find the symmetric model in the decomposition based framework?

Data decomposition for part based model

- Suppose the original filter extracted from left side images is f_l
- Its corresponding factorized feature and symmetric filter is m_l and f_r, respectively.
- We get

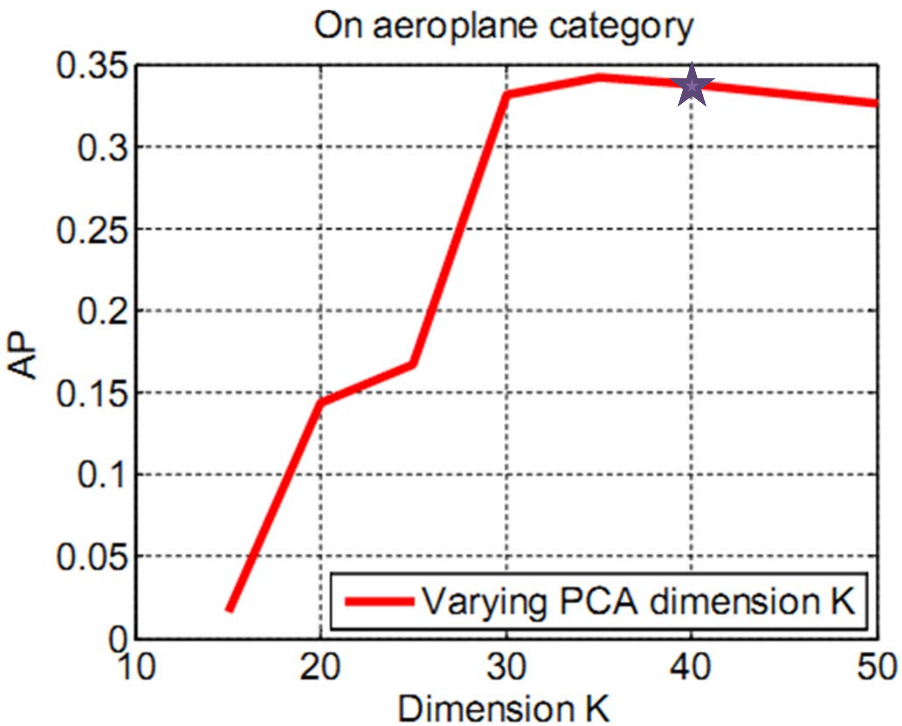
$$m_l = g\left(f_l\right)$$

$$m_r = g\left(f_r\right)$$

- $g(.)$ is the decomposition function. Then we can regress m_r on m_l. Direct back projection also works.

Empirical results

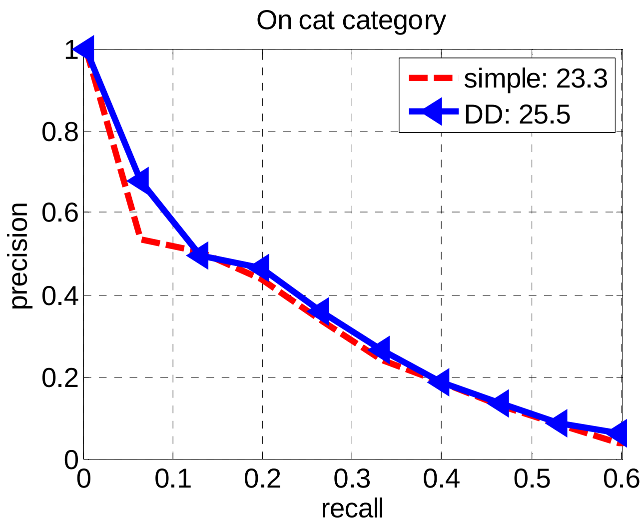
- Determine the number of PCs



Original: 31(HOG) + 59(LBP) = 90

Empirical results

- Performance



VOC2007	Methods	Training (hour)	Test (hour)	AP
cat	simple	24.1	3.5	23.3
	DD	18.2	1.8	25.5

On average it takes about 1.5s per image during testing

Simple: naïve combination; DD: data decomposition

10G vs 4G

Methods

- Data Decomposition (DD) for part based model
- Spatial Mixture Modeling (SMM)
- Context Learning (NLPR_CLS)

Spatial mixture modeling

Real World

Viewpoint

Same aspect ratio, different viewpoint

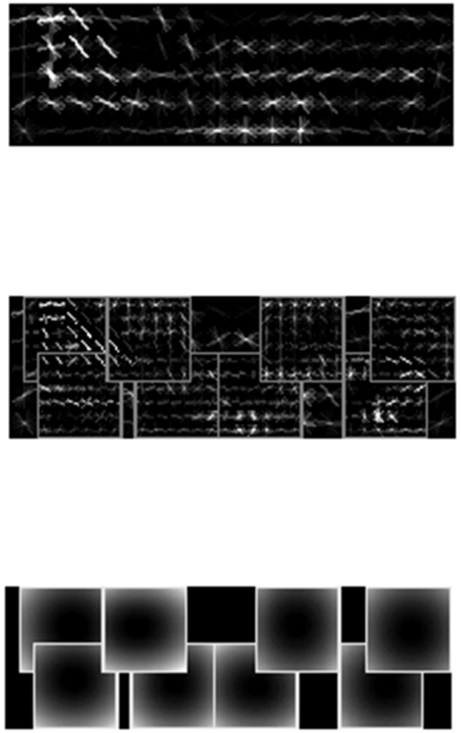
Pose

Different pose with the same aspect ratio should have different part location as well.

Part-Based Model

1. Components based on aspect ratio
2. Deformable parts whose location obeys single Gaussian distribution: one fixed optimal location.

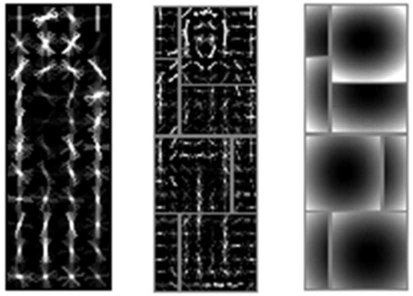
Different viewpoint but the same aspect ratio

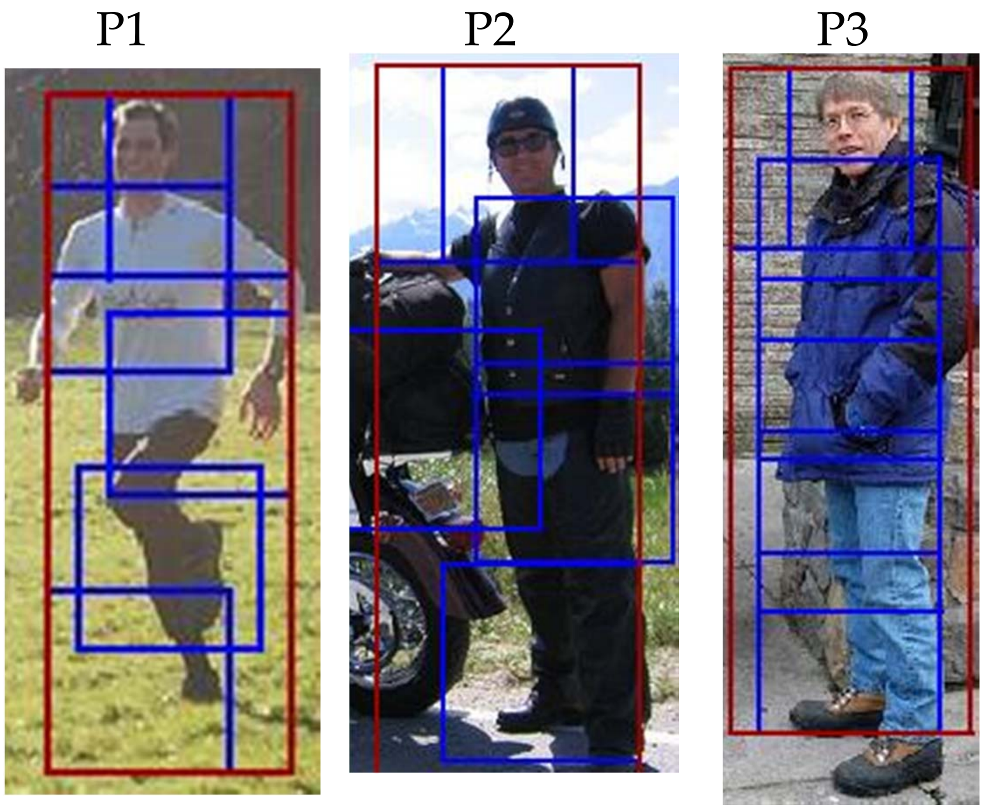


P1

P2

Different poses





The optimal part location in P1 should be different from P2 and P3. It's unfair to use the same penalty deformation distance.

Our solution

Original DPBM

↓

Detection — Perform detection on labeled positives to obtain latent information, namely, location of each part in each positive sample

↓

Clustering — Clustering of the locations of each part

↓

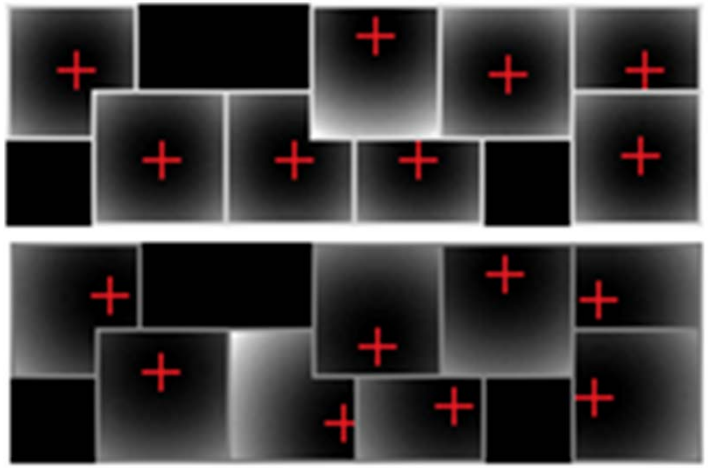
Updating — Update the model with clustering centers as the new mixture anchors of parts

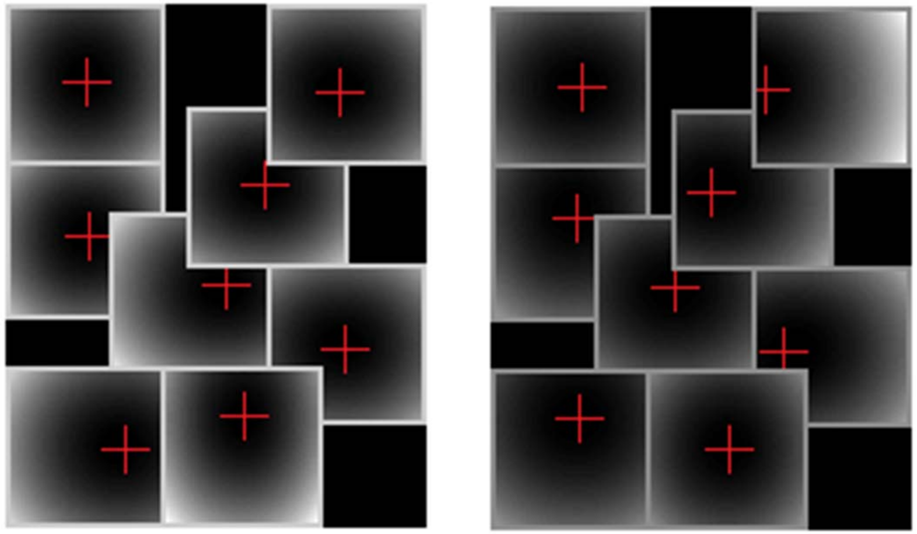
↓

Spatial mixture embedded Model

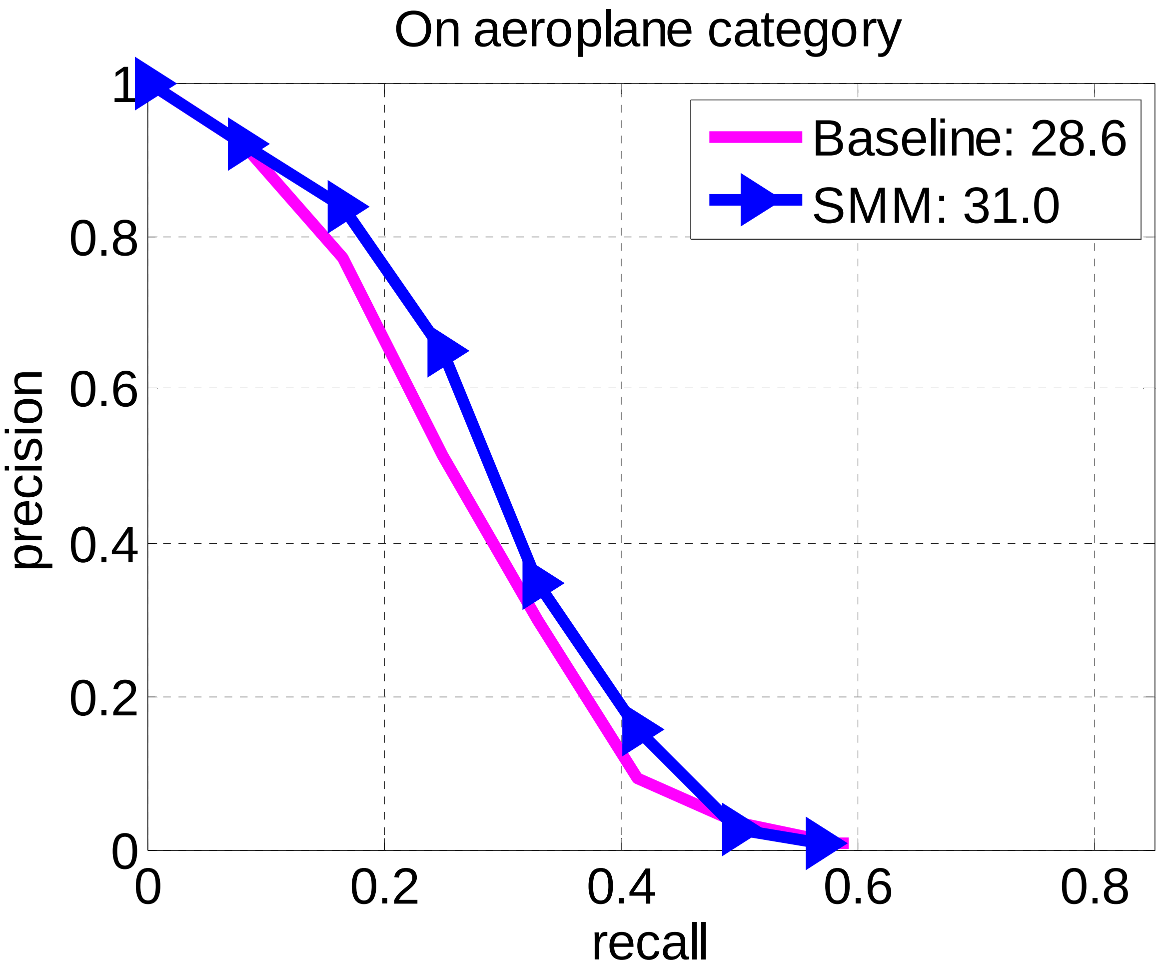
Our model

Examples: aeroplane and cat models wherein each part possesses 2 optimal locations

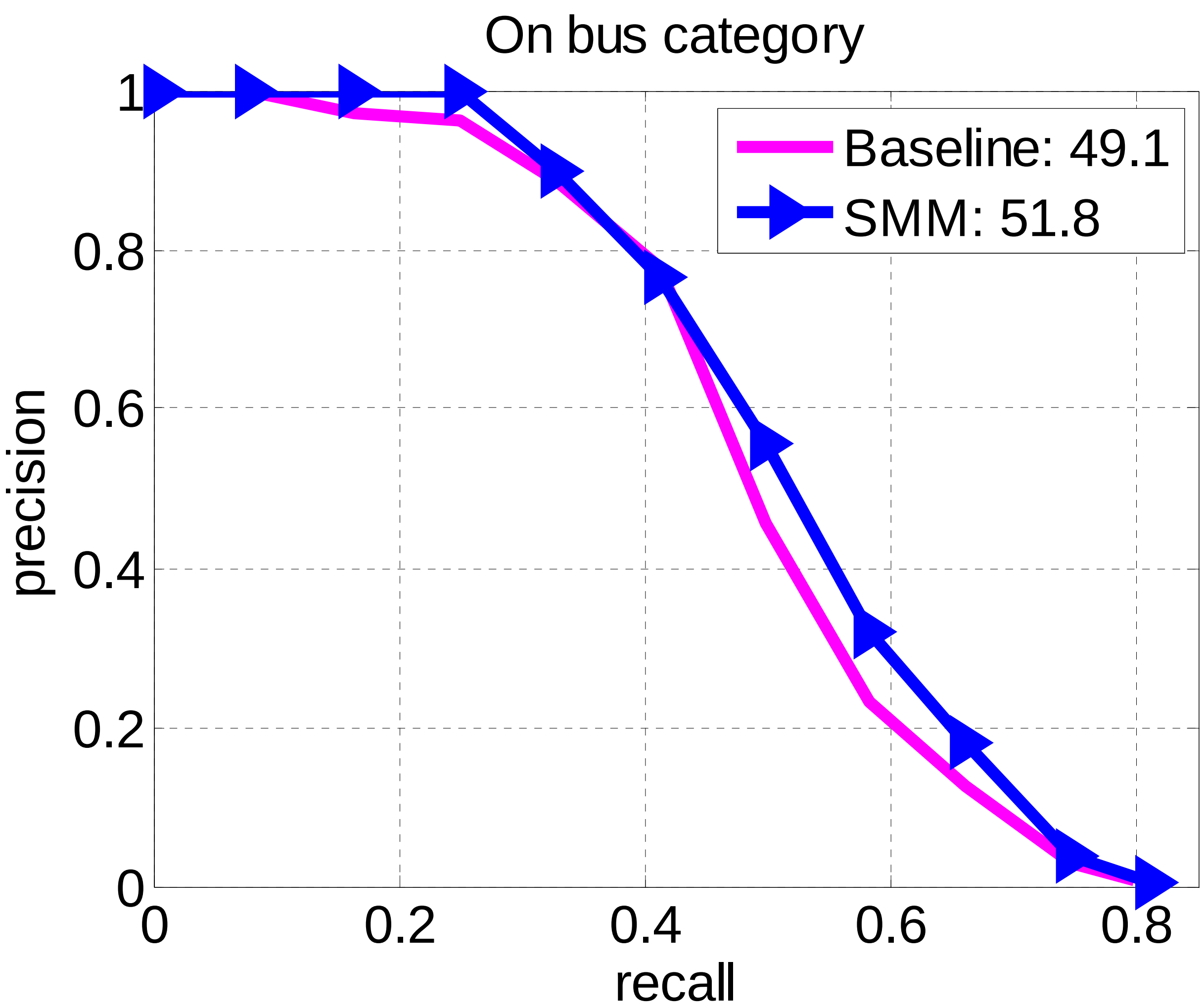


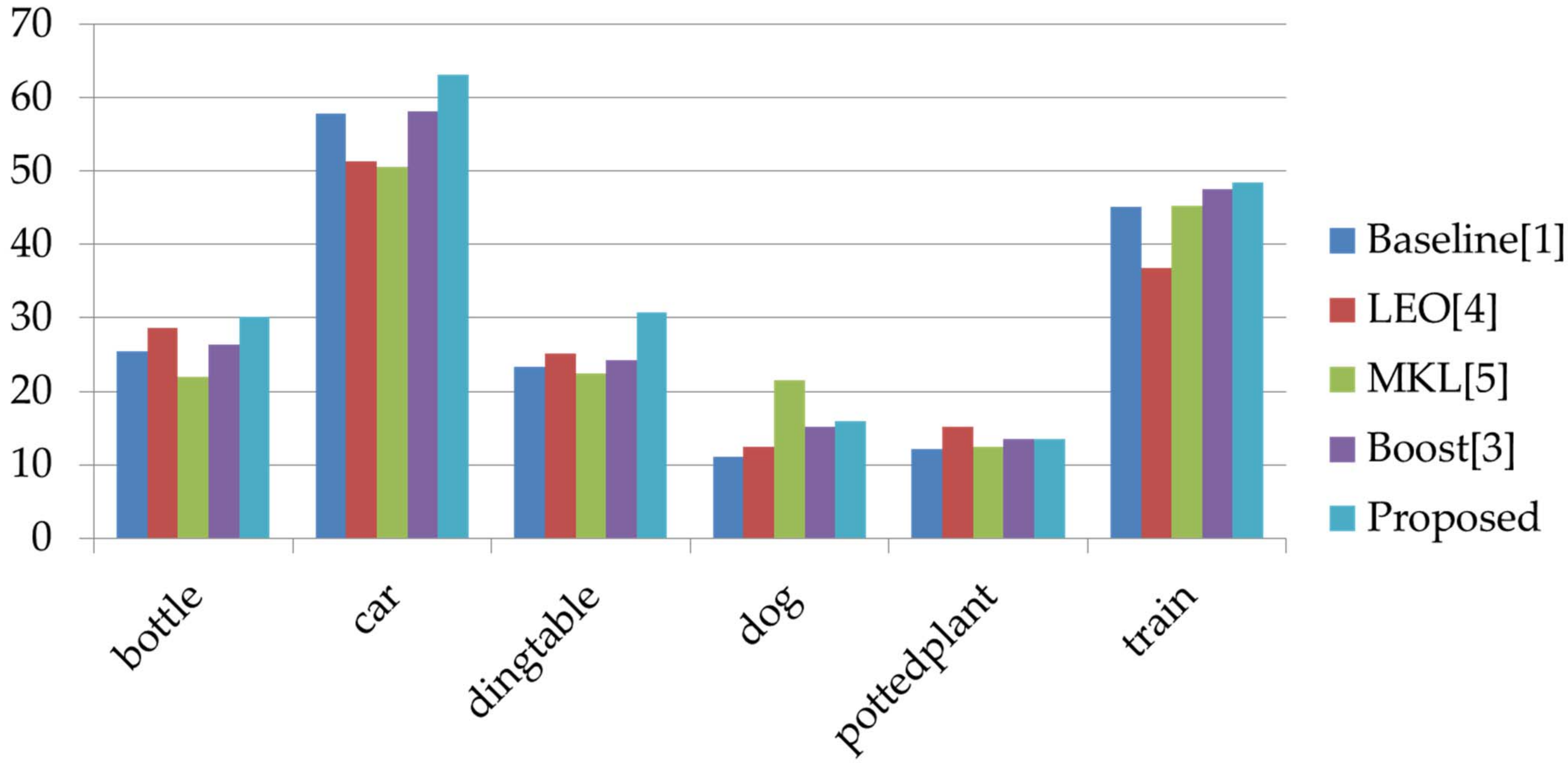


Empirical results



Empirical results




70
60
50
40
30
20
10
0
bottle
car
dingtable
dog
pottedplant
train
Baseline[1]
LEO[4]
MKL[5]
Boost[3]
Proposed

Methods

- Data Decomposition (DD) for part based model
- Spatial Mixture Modeling (SMM)
- Context Learning (NLPR_CLS)

Post-processing: Context Learning

- Scores from 20 categories classification results, from **NLPR_CLS@VOC'2011, 78.2% mAP, 8 first and 9 second places.**

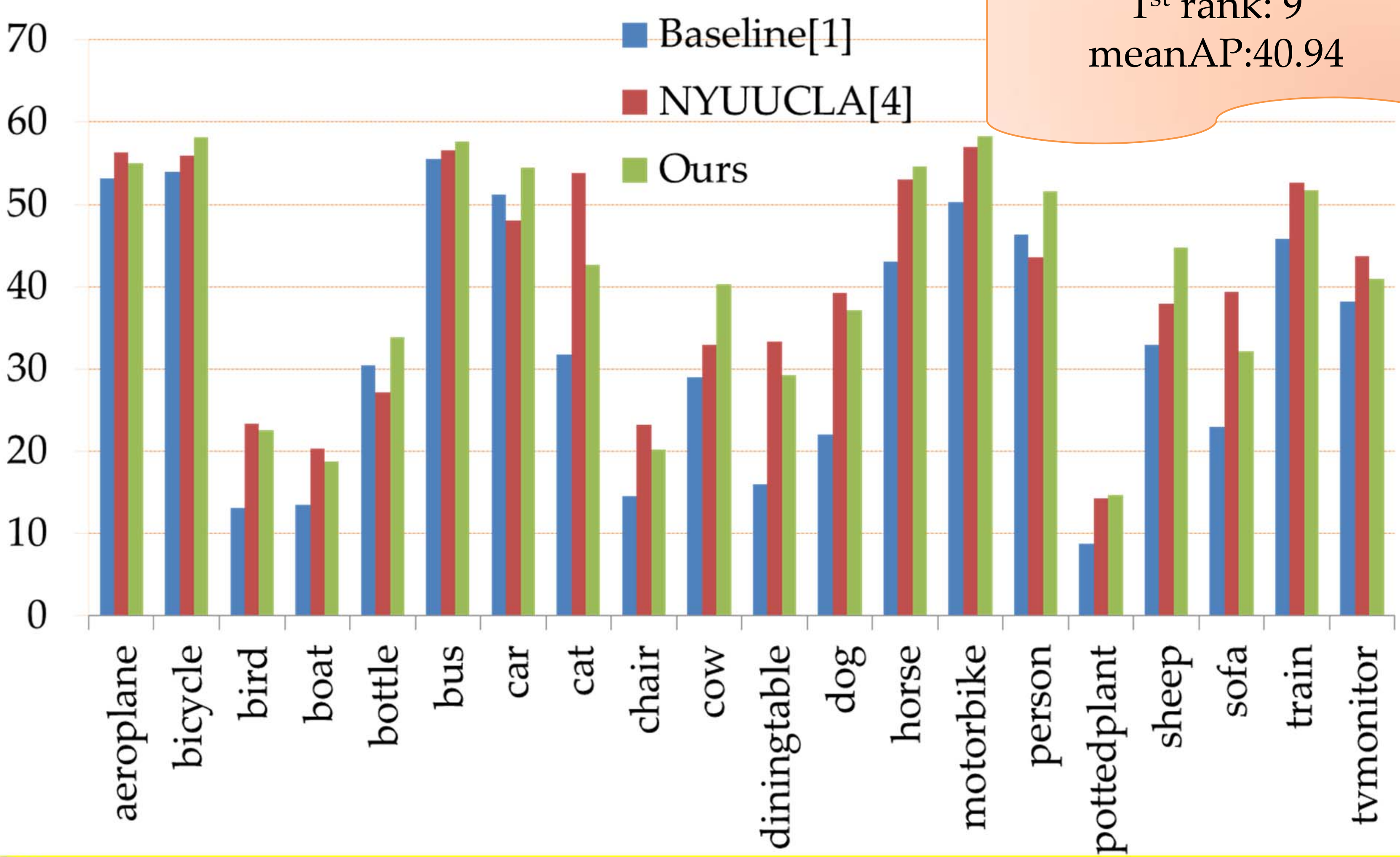
- Others:
  - Maximum scores from 20 categories detectors[1].
  - The overlap between candidate windows and supervised segmentation region[6].
  - Spatial prior[1]

Our Classification: NLPR_CLS

- Partial Least Squares (PLS)
  - Little memory, over 16000 compression ratio (reduce from 500k to 30), scalable to multiple features
  - Preserving discrimination, 0.5%~1% improvement usually
- Semi-Semantic Visual Words (SSVW)
  - Discriminative Visual Words from Deformable Part Based Model[1]
  - Semi-Semantic level image representation
  - Explicit code semantic relationship
  - Low-dimensional high-order code co-occurrence
- Efficient Multiple Linear Kernel Learning (EMLKL)
  - Find appropriate weights for multiple features

PASCAL VOC 2011

1st rank: 9
meanAP:40.94



Reference

- [1] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, Object Detection with Discriminatively Trained Part Based Models, TPMAI 2010.
- [2] Marco Pedersoli, Andrea Vedaldi, Jordi Gonzàlez, **A** Coarse-to-fine approach for fast deformable object detection, in CVPR, June, 2011 .
- [3] Junge Zhang, Kaiqi Huang, Yinan Yu, Tieniu Tan. Boosted Local Structured HOG-LBP for Object Localization, in CVPR, 2011.
- [4] Long Zhu, Yuanhao Chen, Alan Yuille, William Freeman "Latent Hierarchical Structural Learning for Object Detection". CVPR 2010.
- [5] A. Vedaldi, V. Gulshan, M. Varma, and A. Zisserman. "Multiple Kernels for Object Detection", ICCV, 2009.
- [6] Lubor Ladicky, Philip H.S. Torr. "Automatic Labelling Environment (Semantic Segmentation) ".
- [7] Yongzhen Huang, Kaiqi Huang, Tieniu Tan. Salient Coding for Image Classification. In CVPR,2011.

Thanks very much

- If you have any question, please drop email to: jgzhang@nlpr.ia.ac.cn