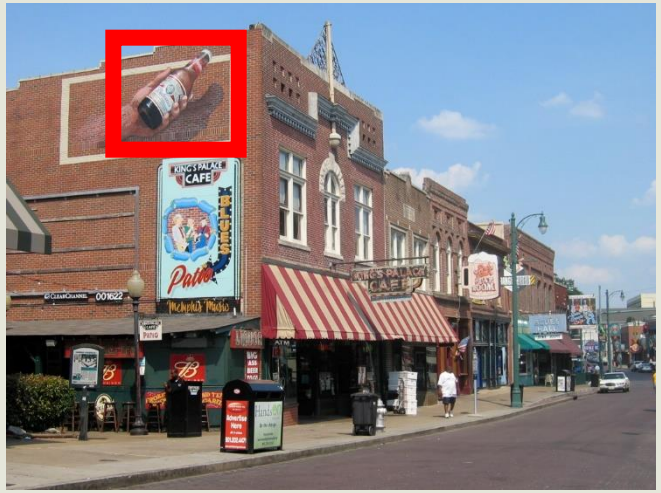
Sharing visual features for multiclass and multiview object detection

Antonio Torralba, Kevin P. Murphy, William T. Freeman

The problem

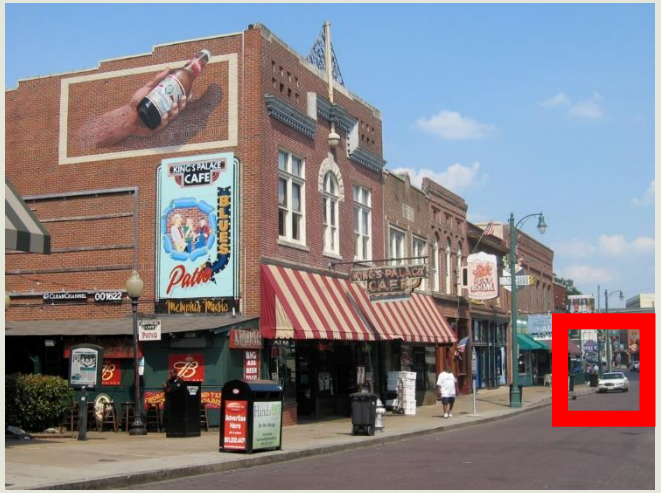
- In the previous papers we learned how to detect and classify images with 1 class.
- Our Recognition Problem is to recognize many object category with few images per category.
- How to detect many categories?

Find each category



...

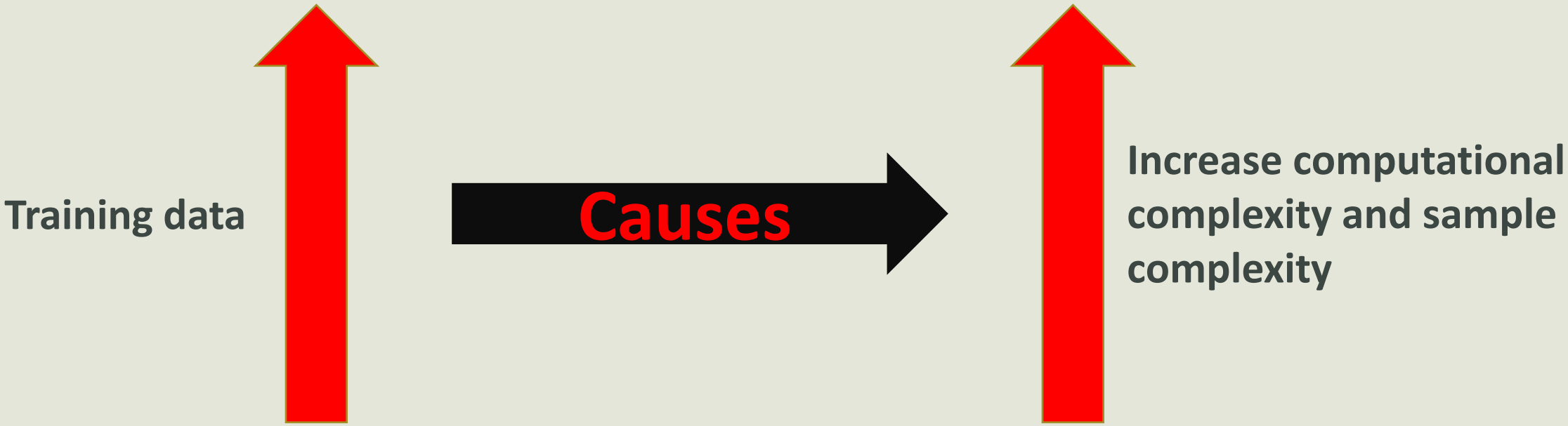
...




Patio
Memphis Music
CLEAR CHANNEL
001622
PATIO
BIG ASS BEER
ATM

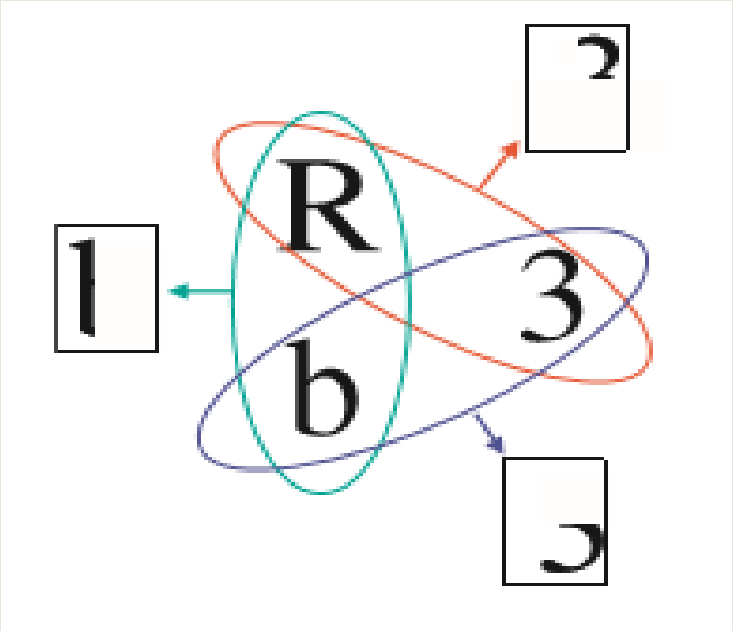
Why not?

- Require a lot of training data since each classifier require many different images.

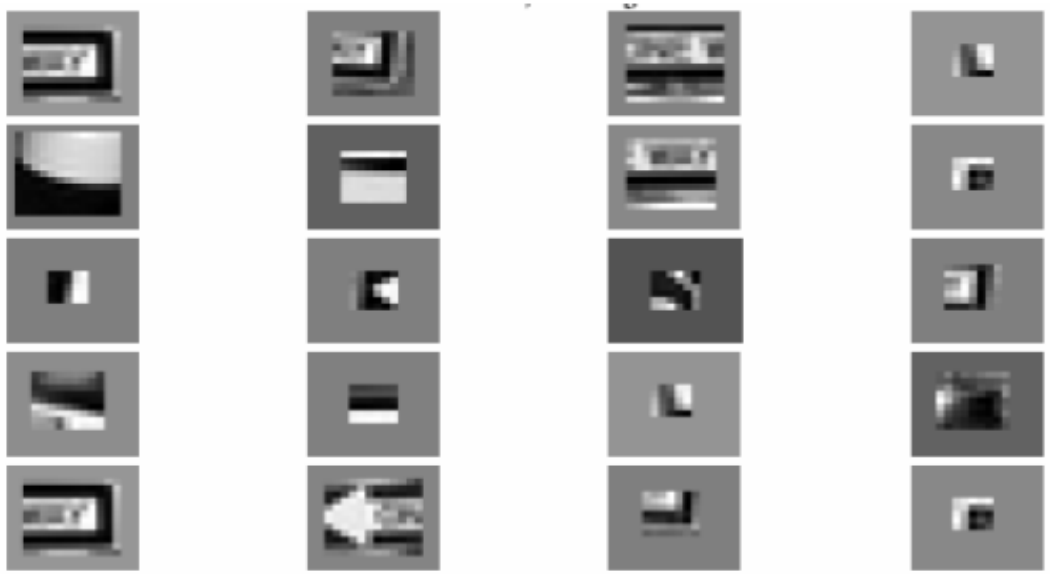


The solution - Feature sharing

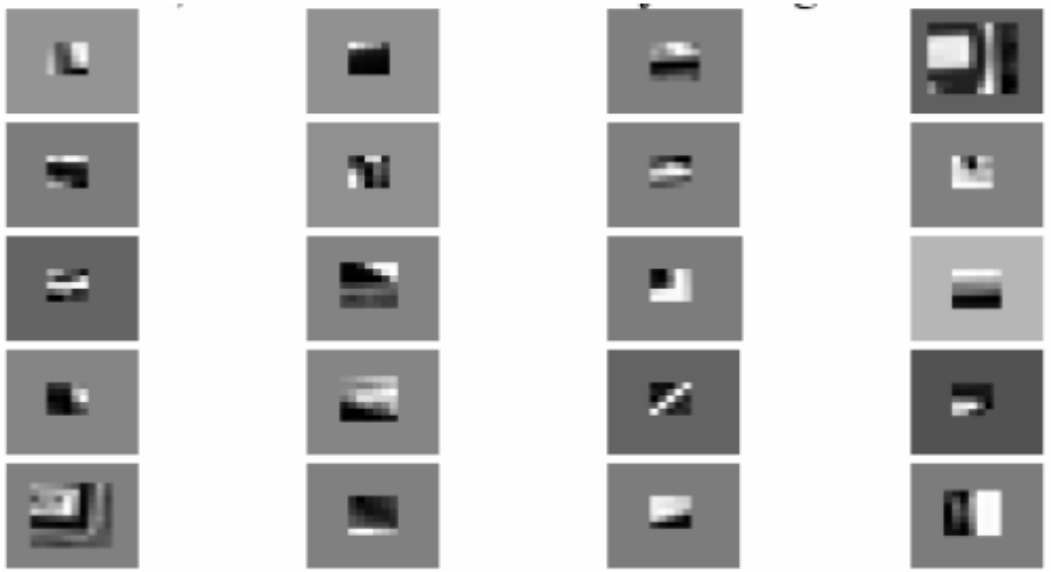
- Find common features that distinguish a subset of classes against the rest.



a) Object



b) Selected features by a single detector



c) Selected features when trained jointly

Template-like features

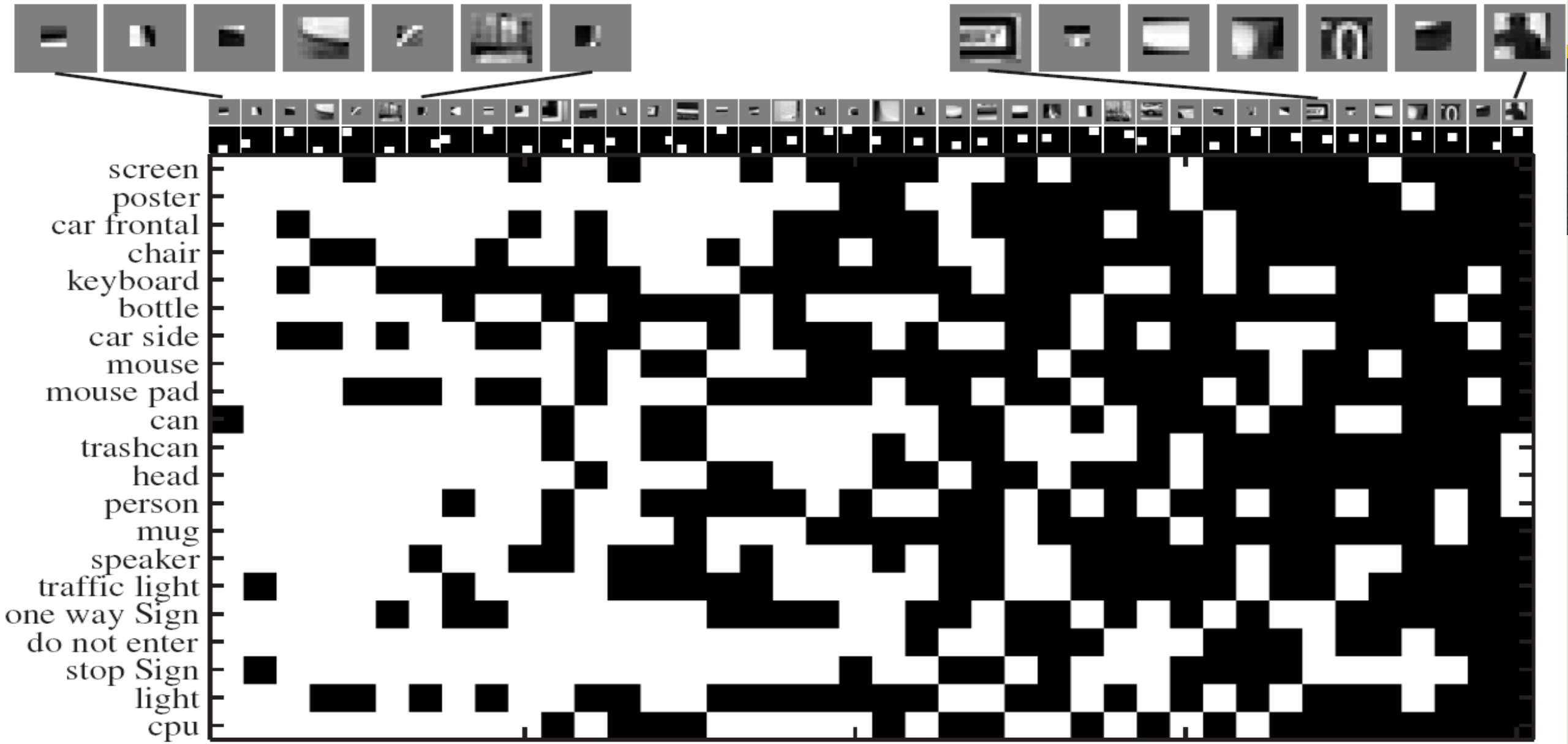
100% accuracy for a single object

But too specific.

Wavelet-like features,

weaker discriminative power

but shared in many classes.



Left part: common features which are being shared in most classes

Right part: specific features which are being shared by only a few classes

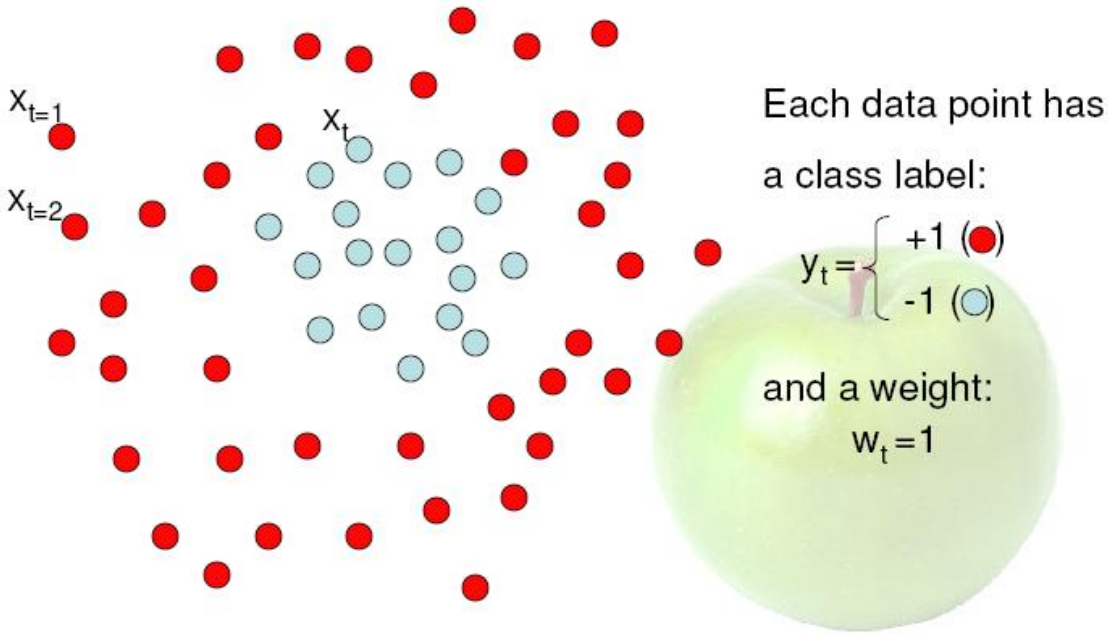
What we gain from sharing features?

- Essentially more positive samples
- Reuse the data

Example of the strong classifier

Boosting

- It is a sequential procedure:



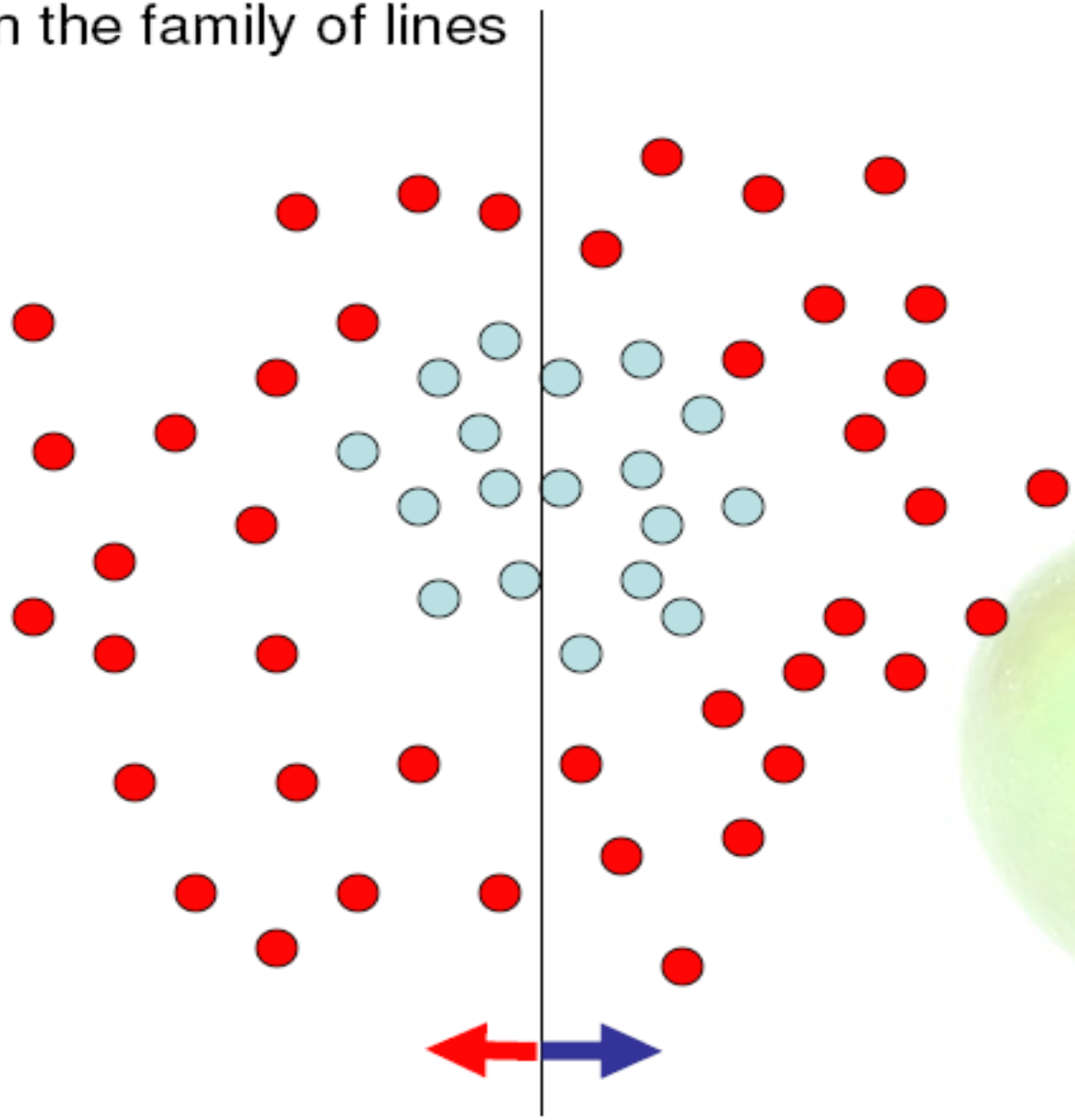
Each data point has a class label:

$$y_t = \begin{cases} +1 \ (\bullet) \\ -1 \ (\circ) \end{cases}$$

and a weight:

$w_t = 1$

Weak learners from the family of lines



Each data point has

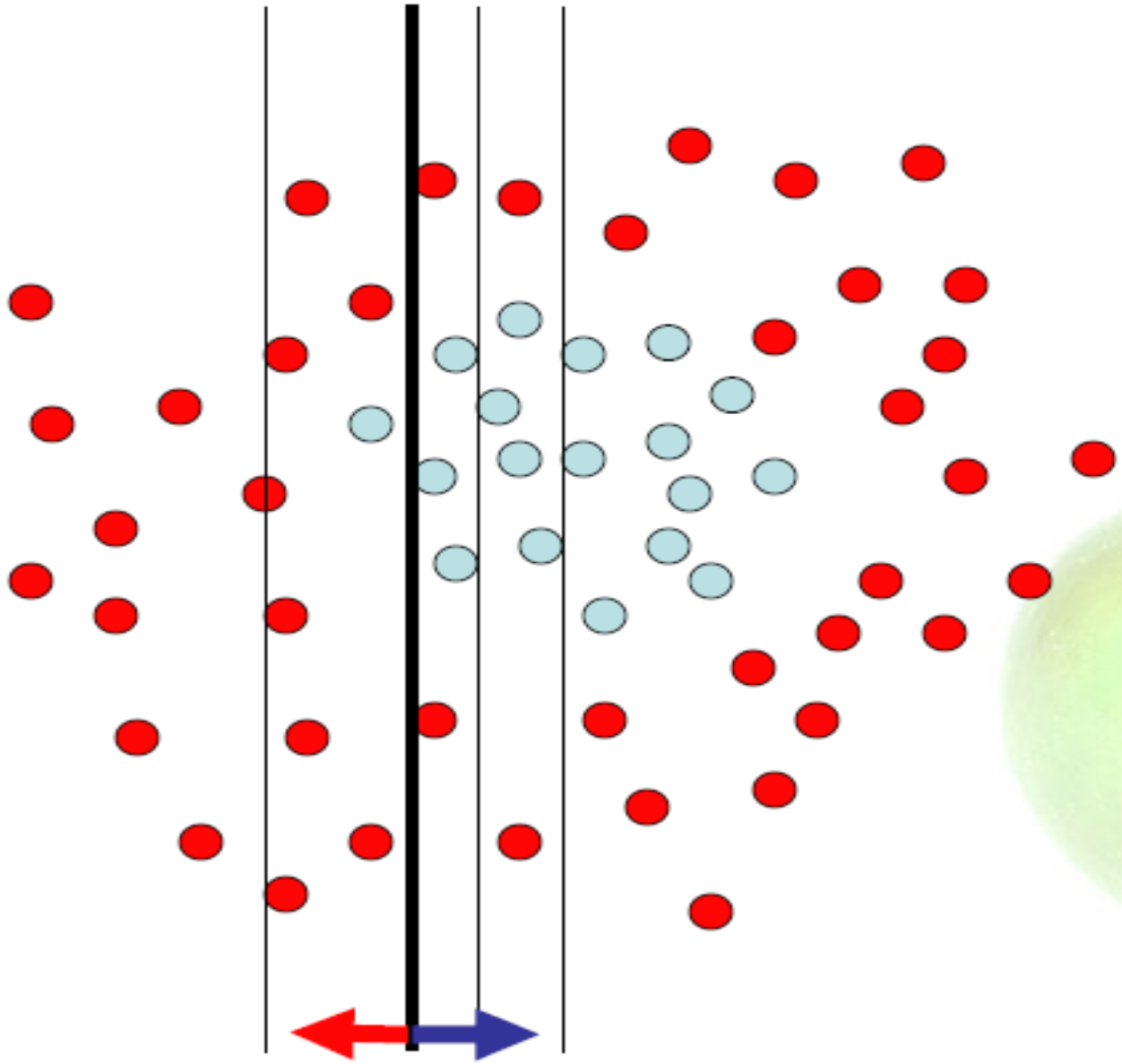
a class label:

$$y_t = \begin{cases} +1 \ (\bullet) \\ -1 \ (\circ) \end{cases}$$

and a weight:

$$w_t = 1$$

h => p(error) = 0.5 it is at chance



Each data point has a class label:

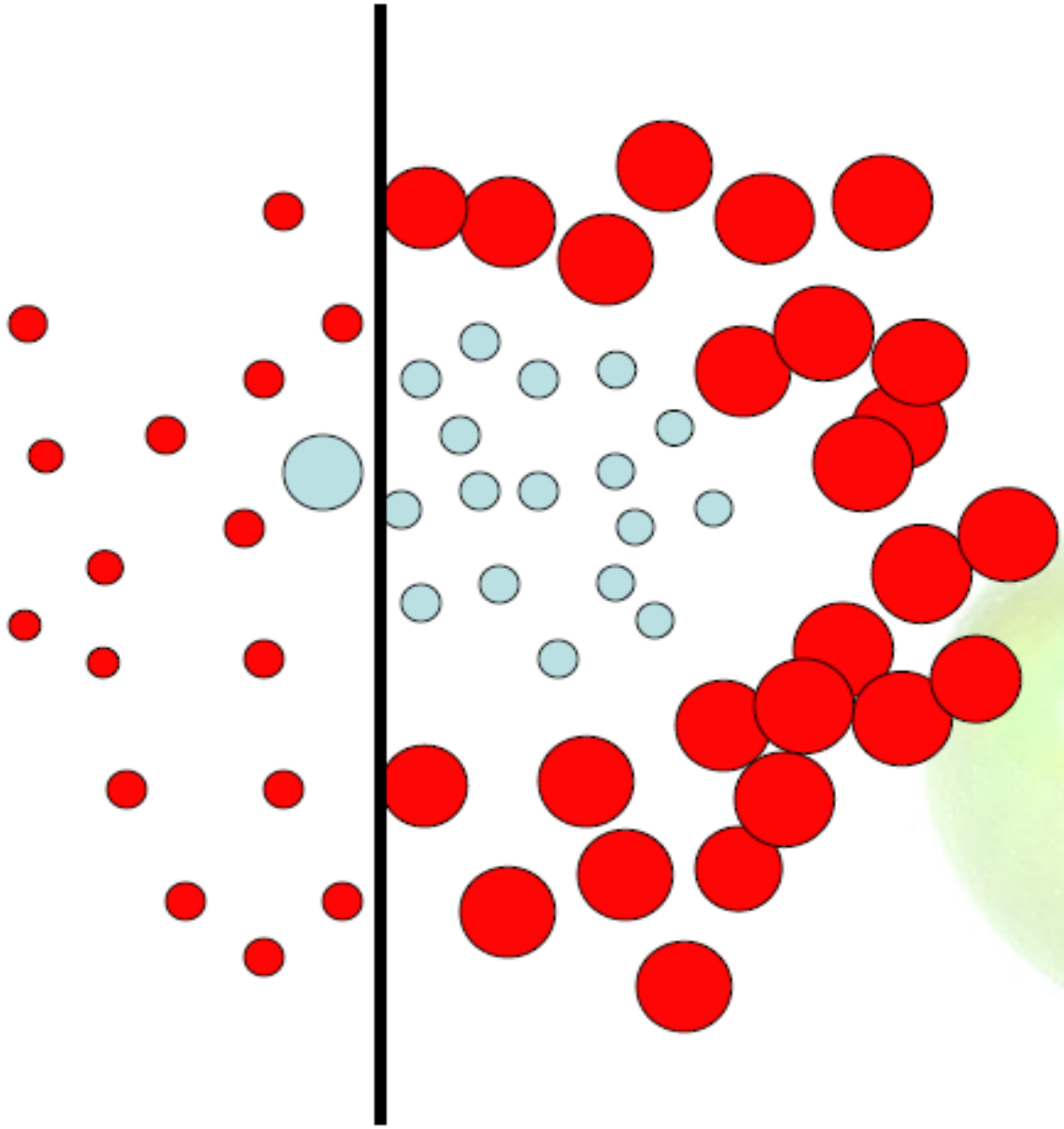
$$y_t = \begin{cases} +1 \ (\bullet) \\ -1 \ (\circ) \end{cases}$$

and a weight:

$$w_t = 1$$

This one seems to be the best

This is a '**weak classifier**': It performs slightly better than chance.



Each data point has

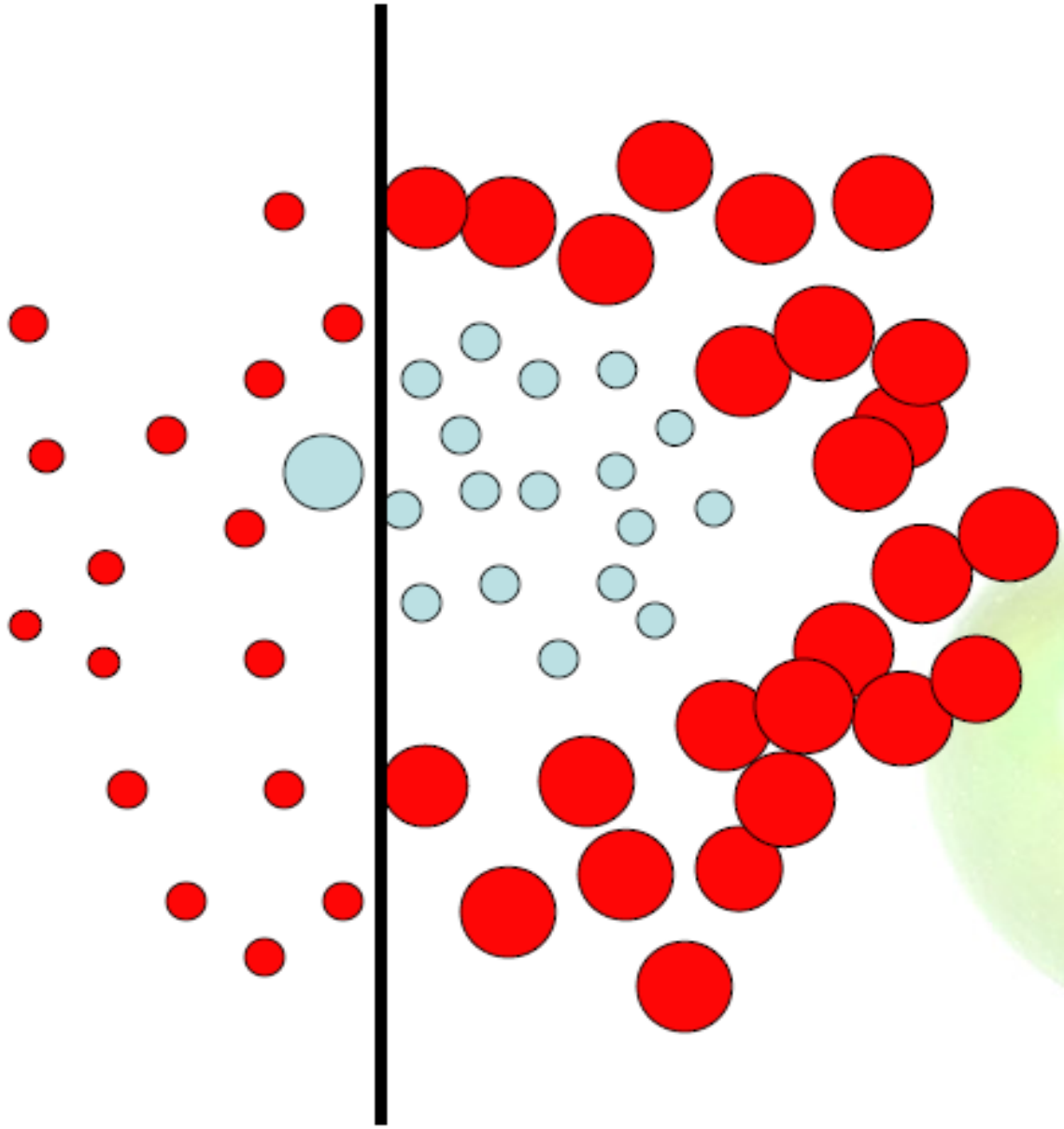
a class label:

$$y_t = \begin{cases} +1 \ (\bullet) \\ -1 \ (\circ) \end{cases}$$

We update the weights:

$$w_t \leftarrow w_t \exp\{-y_t H_t\}$$

We set a new problem for which the previous weak classifier performs at chance again



Each data point has

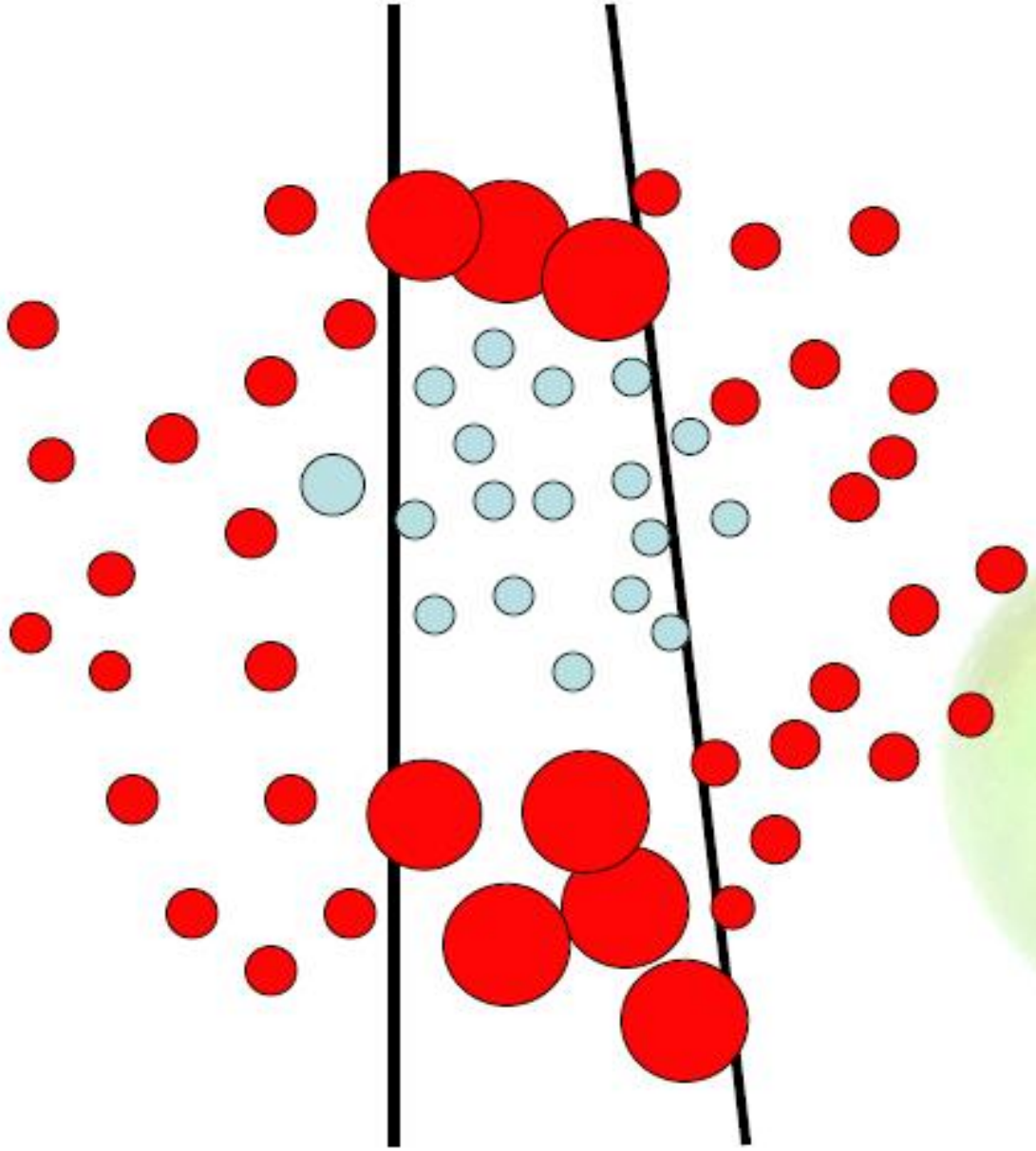
a class label:

$$y_t = \begin{cases} +1 \ (\bullet) \\ -1 \ (\circ) \end{cases}$$

We update the weights:

$$w_t \leftarrow w_t \exp\{-y_t H_t\}$$

We set a new problem for which the previous weak classifier performs at chance again



Each data point has

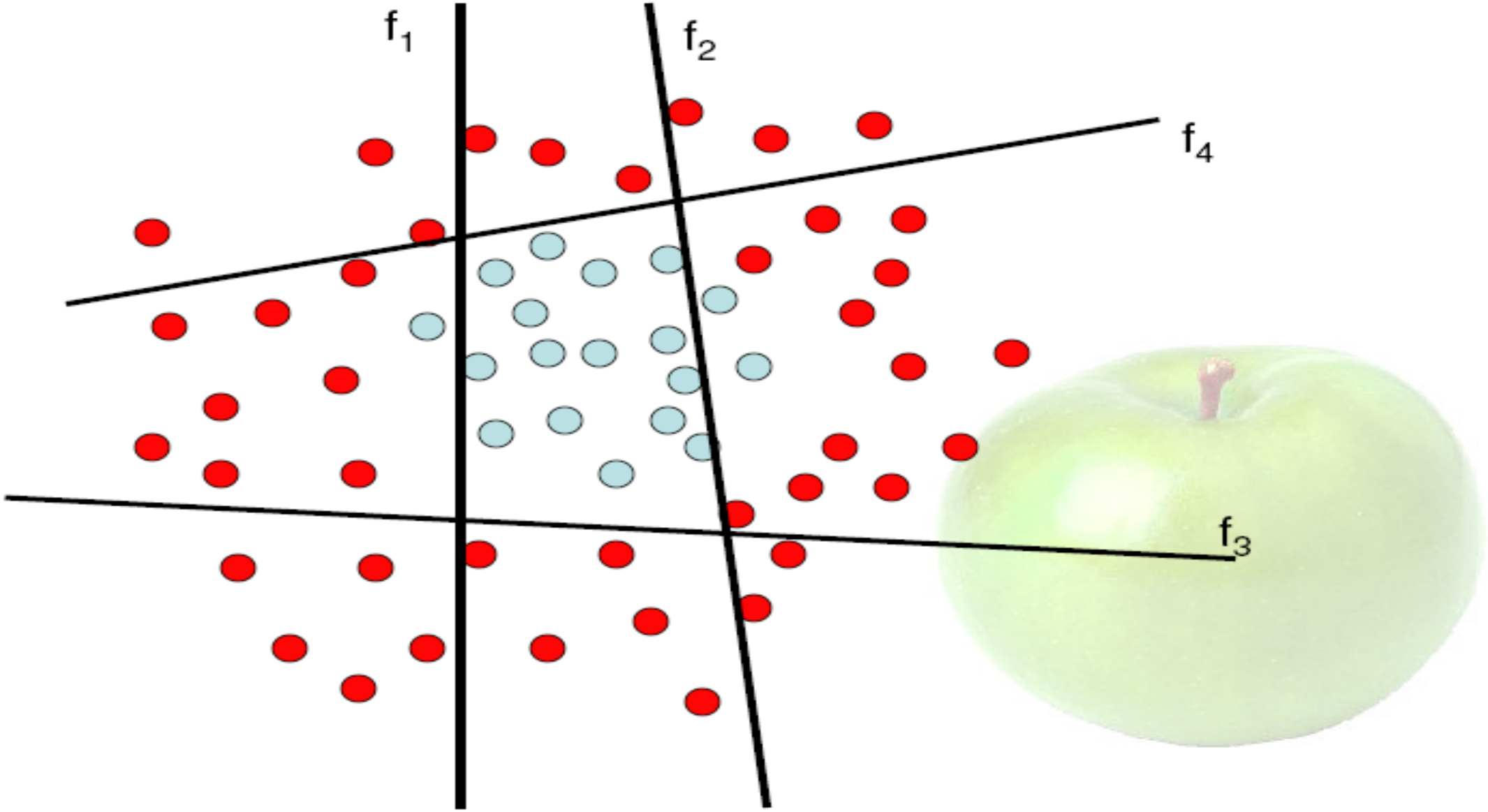
a class label:

$$y_t = \begin{cases} +1 \ (\bullet) \\ -1 \ (\circ) \end{cases}$$

We update the weights:

$$w_t \leftarrow w_t \exp\{-y_t H_t\}$$

We set a new problem for which the previous weak classifier performs at chance again



The strong (non- linear) classifier is built as the combination of all the weak (linear) classifiers.

Joint Boost - Multiclass

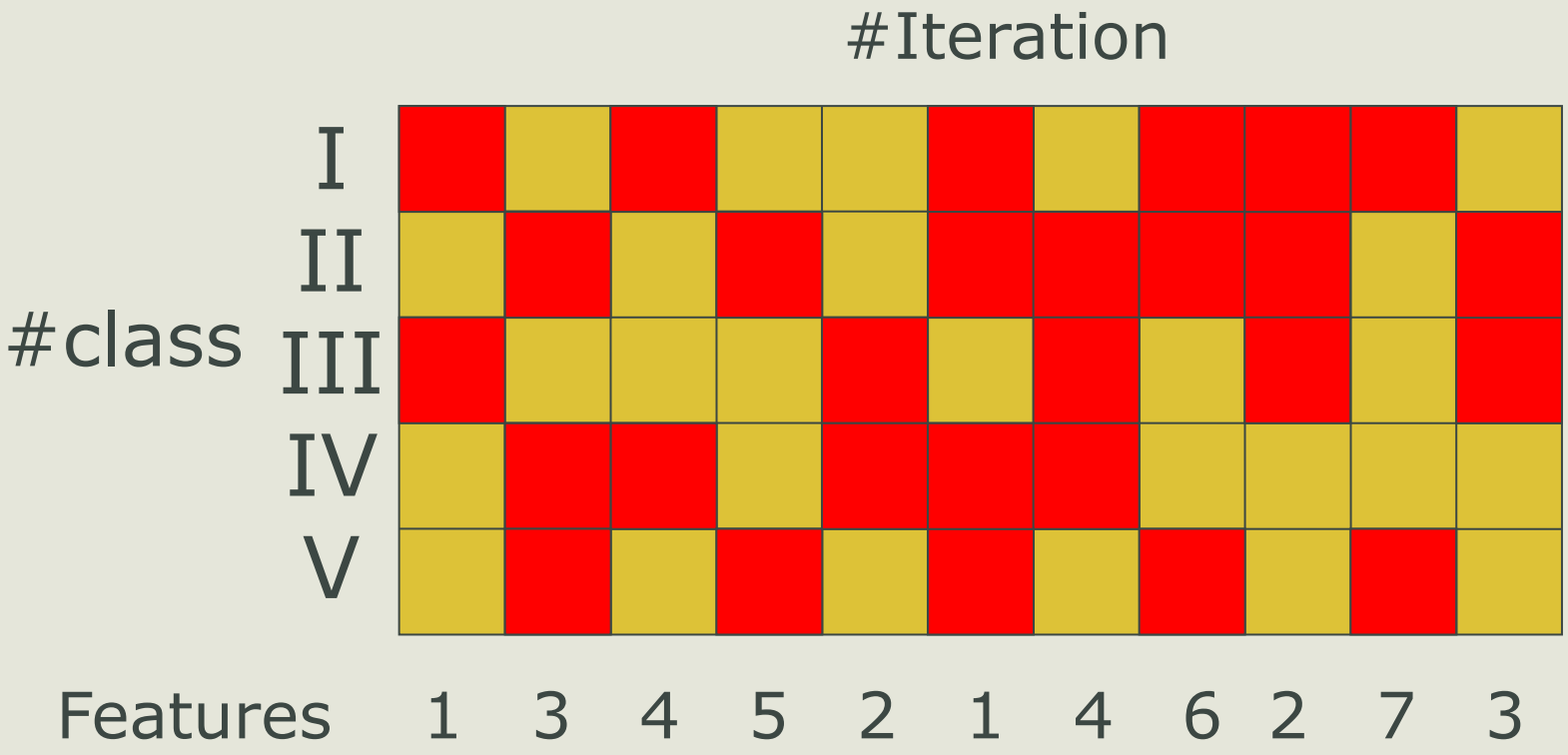
- We can minimize a similar function using one-vs-all strategy

$$H(v,c) = \sum_{m=1}^{M} h_m(v,c) \qquad J = \sum_{c=1}^{C} E\left[e^{-z^c H(v,c)}\right]$$

- This doesn't work very well, since it is separable in c.
- Put constraints. -> **shared features!**

Join Boost

- In each iteration, choose
  - One common feature
  - A subset of classes that use this feature
- So that the objective decreases most

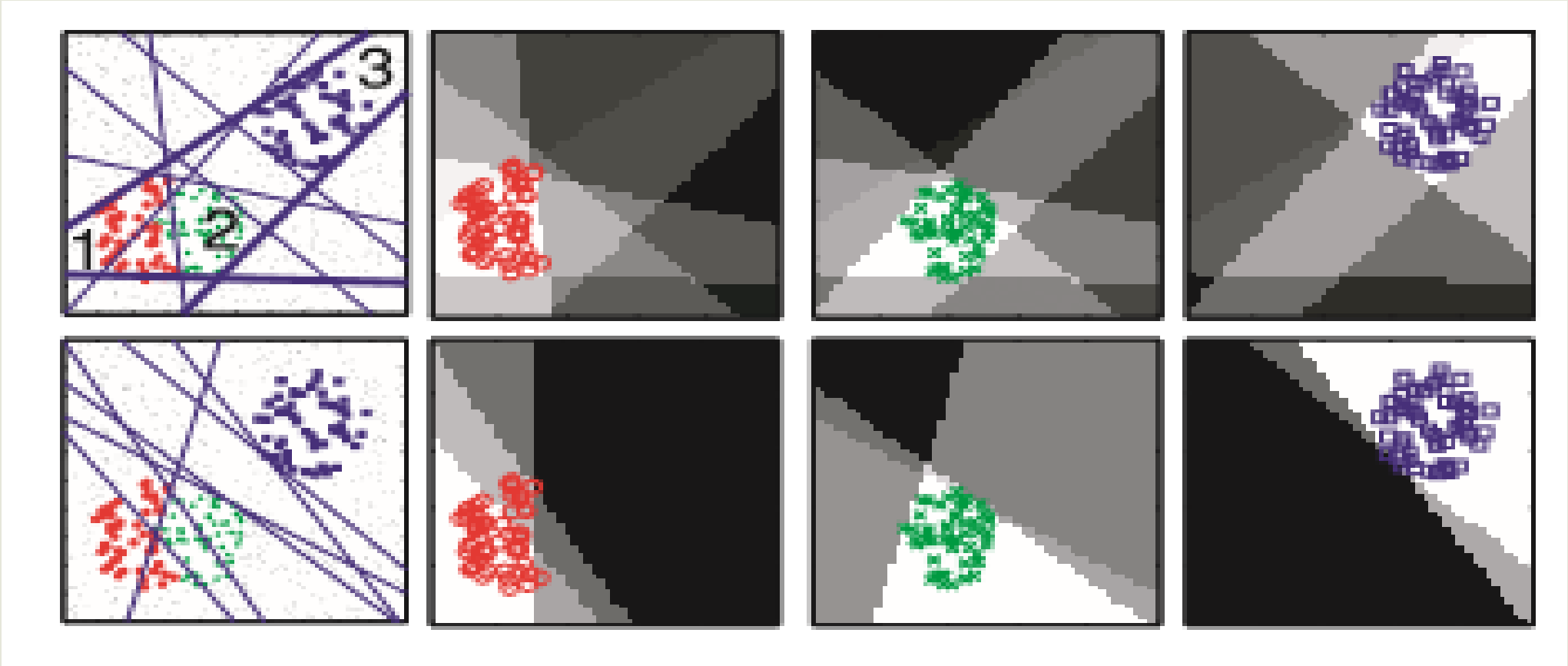

#Iteration
#class
I
II
III
IV
V
Features
1 3 4 5 2 1 4 6 2 7 3

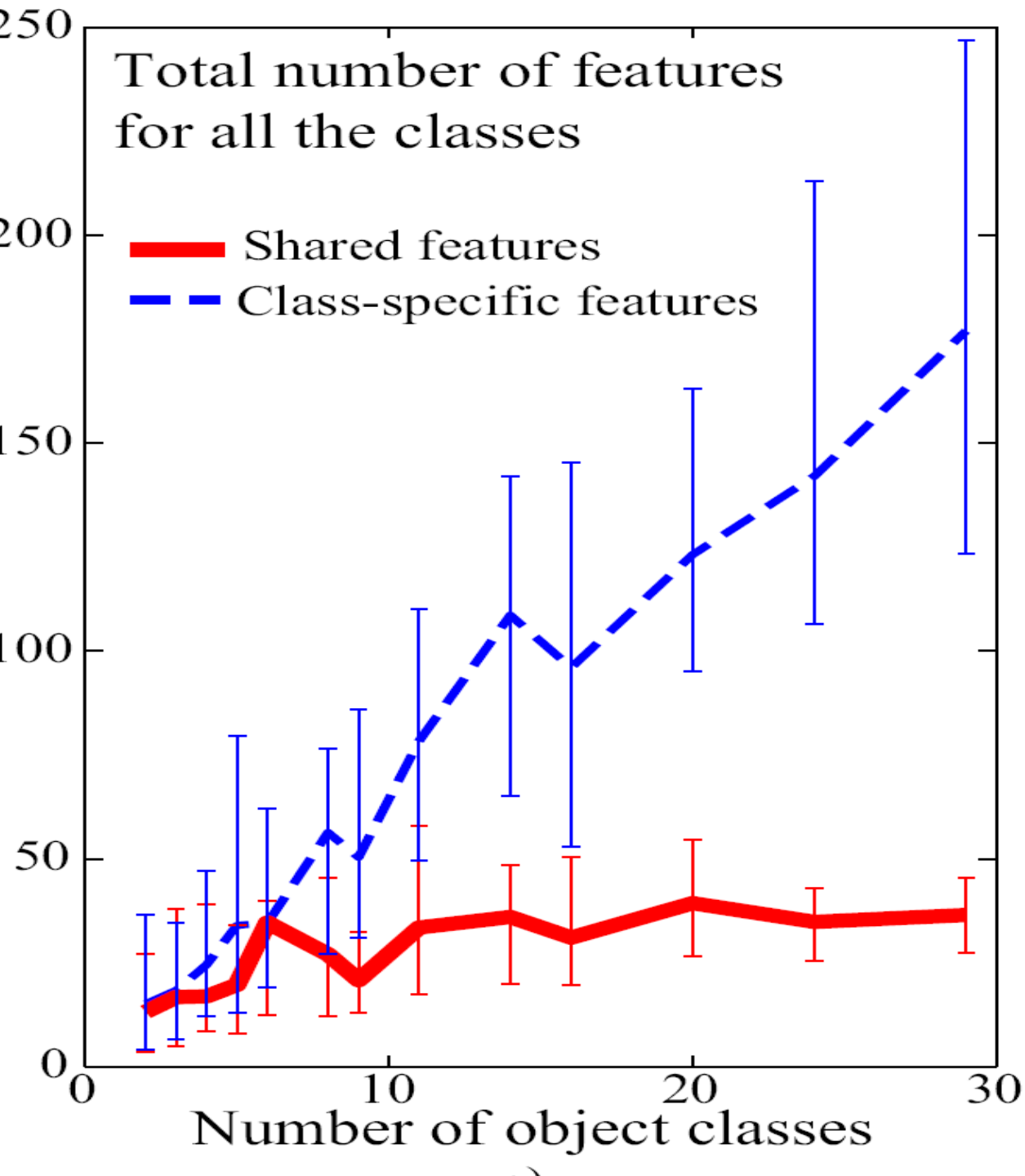
- Each class may have its own favorite feature
- a common feature may not be any of them, however it **simultaneously** decreases errors of many classes.

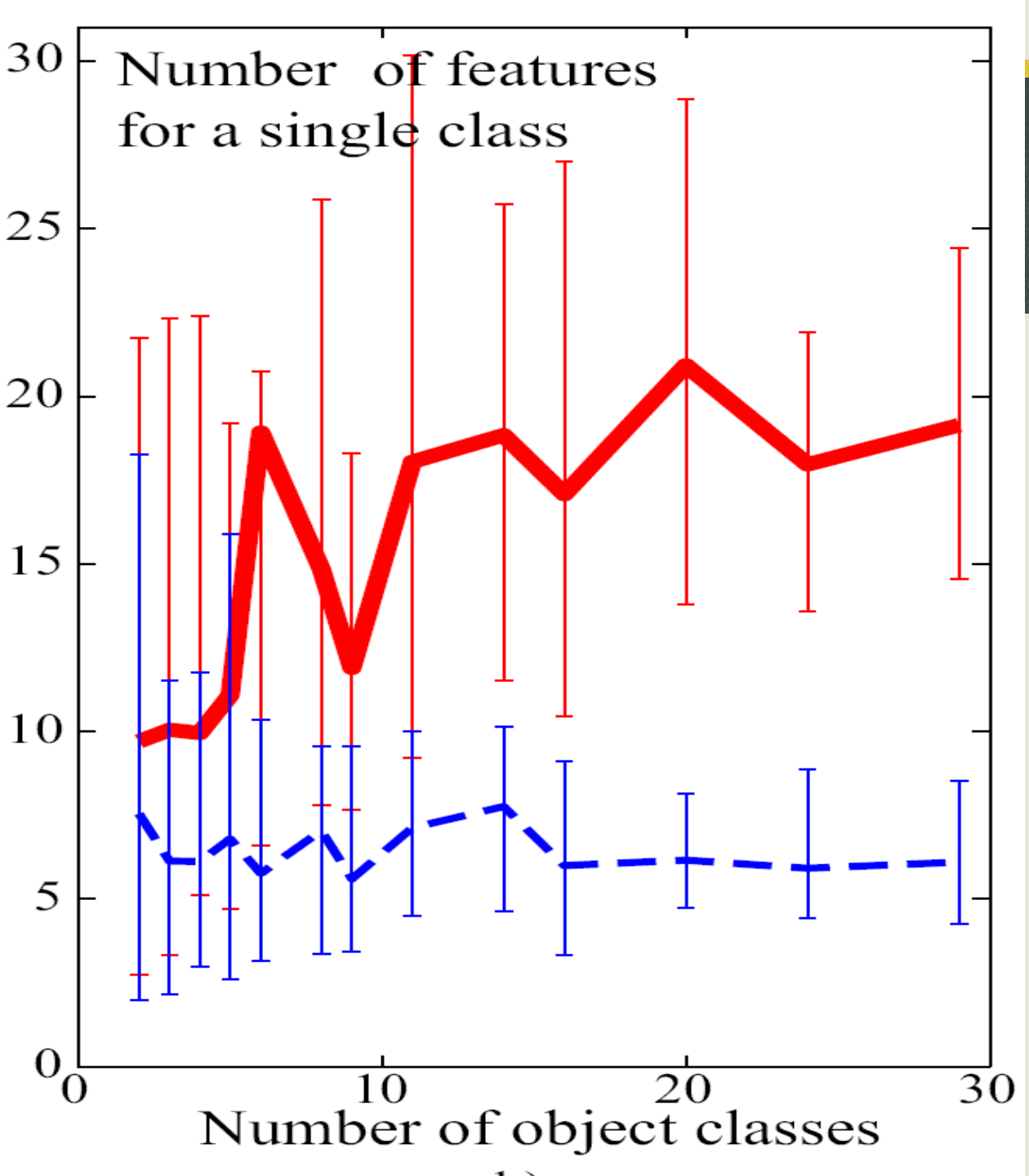
Computational issue

- Choose the best subset is prohibitive
- Use greedy approach
  - Choose one class and one feature so that the objective decreases the most
  - Iteratively add more classes until the objective increases again
    - Note the common feature may change

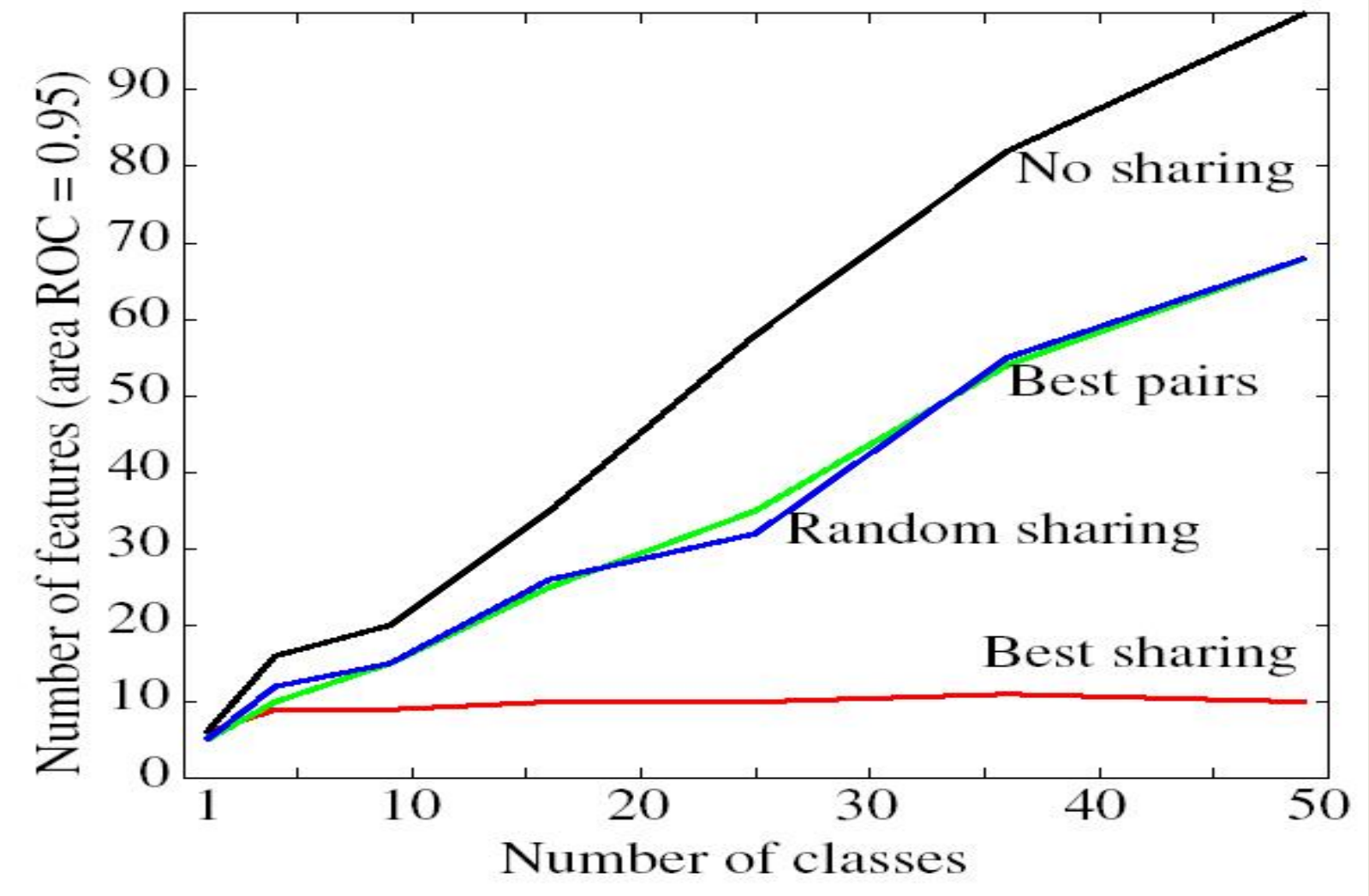
Joint boosting Vs. Independent




Total number of features
for all the classes
Shared features
Class-specific features
250
200
150
100
50
0
0
10
20
30
Number of object classes


Number of features
for a single class
30
25
20
15
10
5
0
0
10
20
30
Number of object classes

#features = O(log #class)

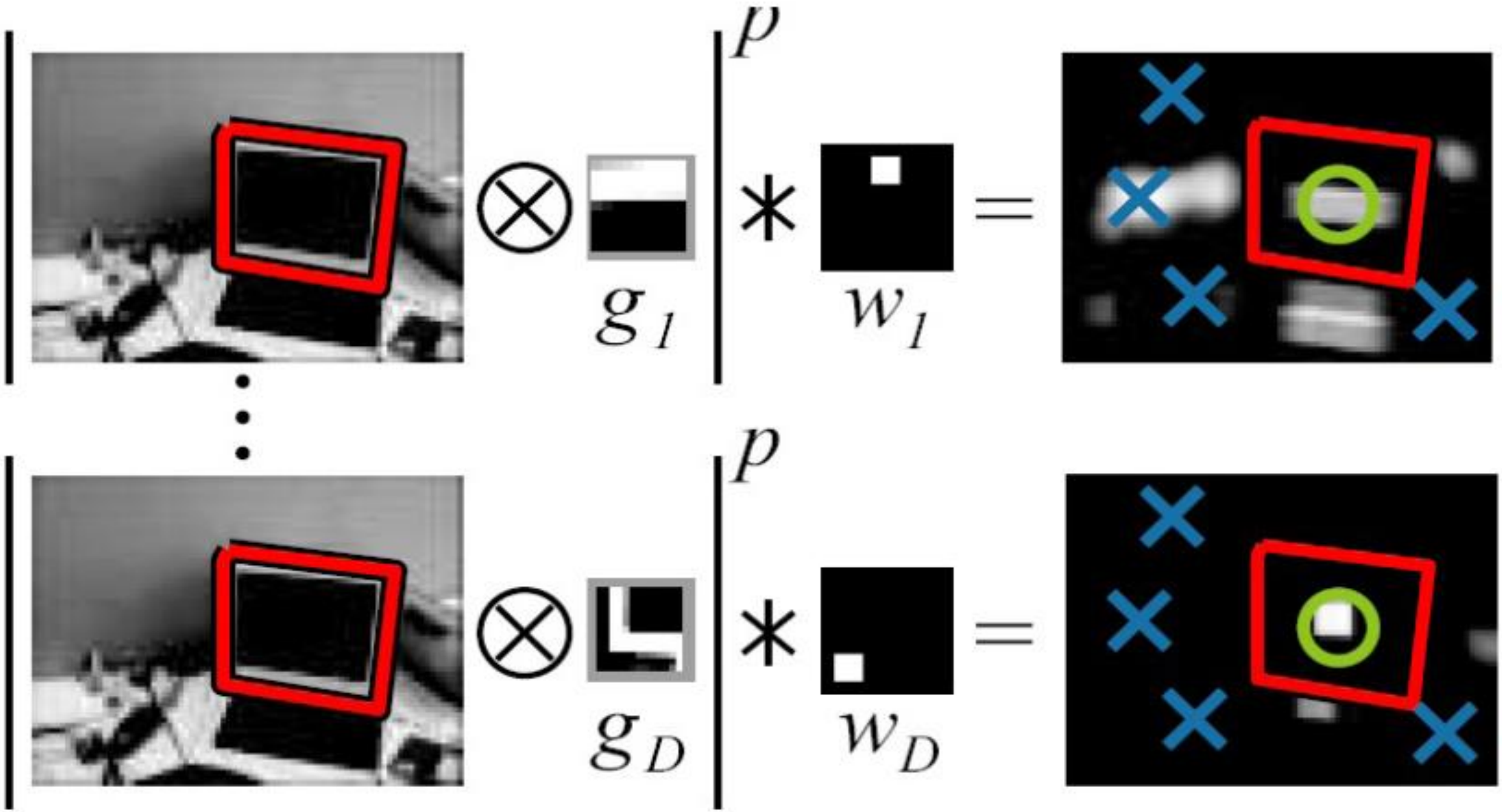


Feature they used in the paper

- Dictionary
  - 2000 random sampled patches
    - Of size from 4x4 to 14x14
    - no clustering
  - Each patch is associated with a spatial mask

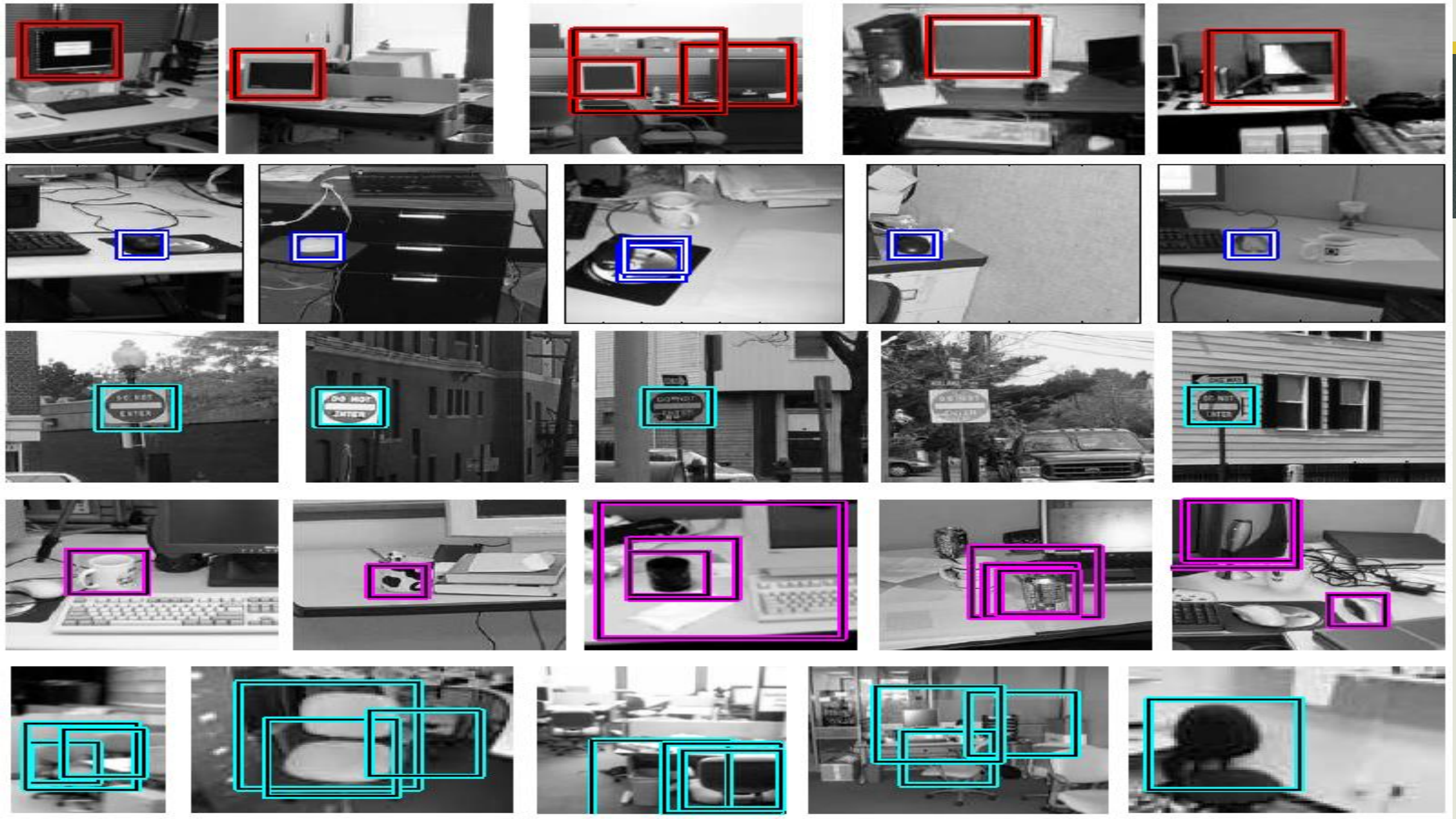
Features

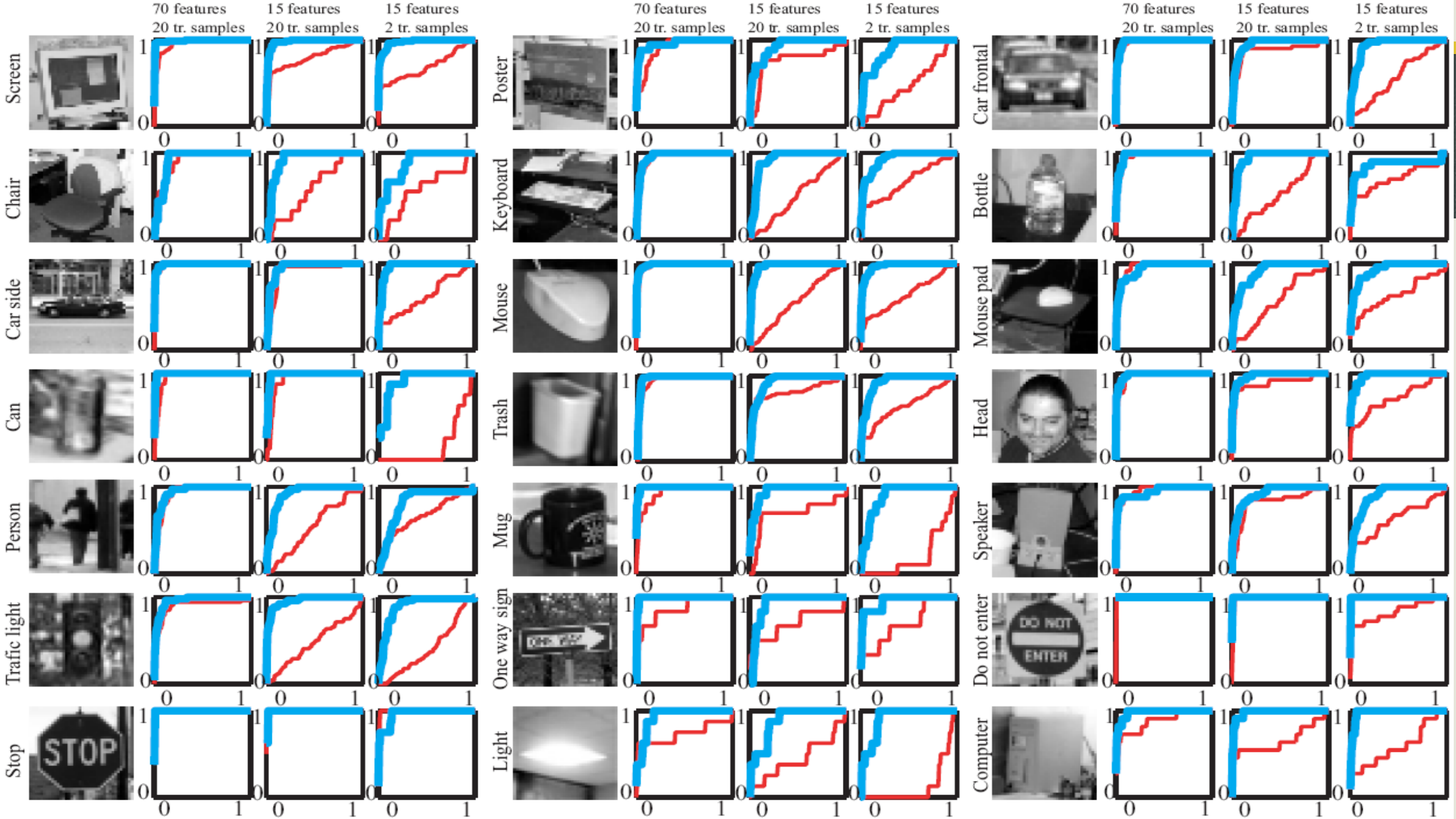
- Building feature vectors
  - Normalized correlation with each patch to get response
  - Raise the response to some power
    - Large value gets even larger and dominate the response (max operation)
  - Use spatial mask to align the response to the object center (voting)
  - Extract response vector at object center

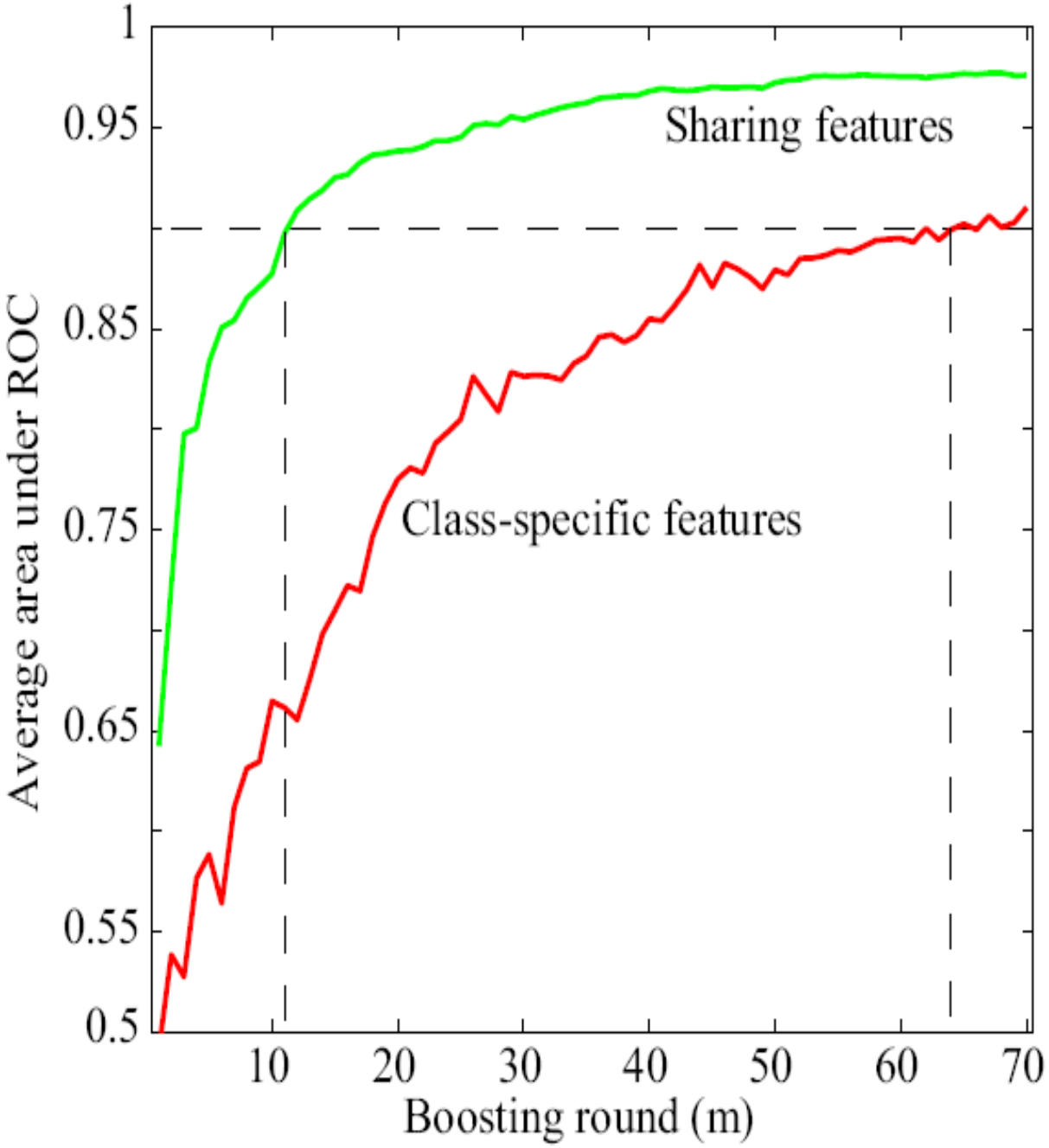

p
g 1
w 1
p
g D
w D

The Data

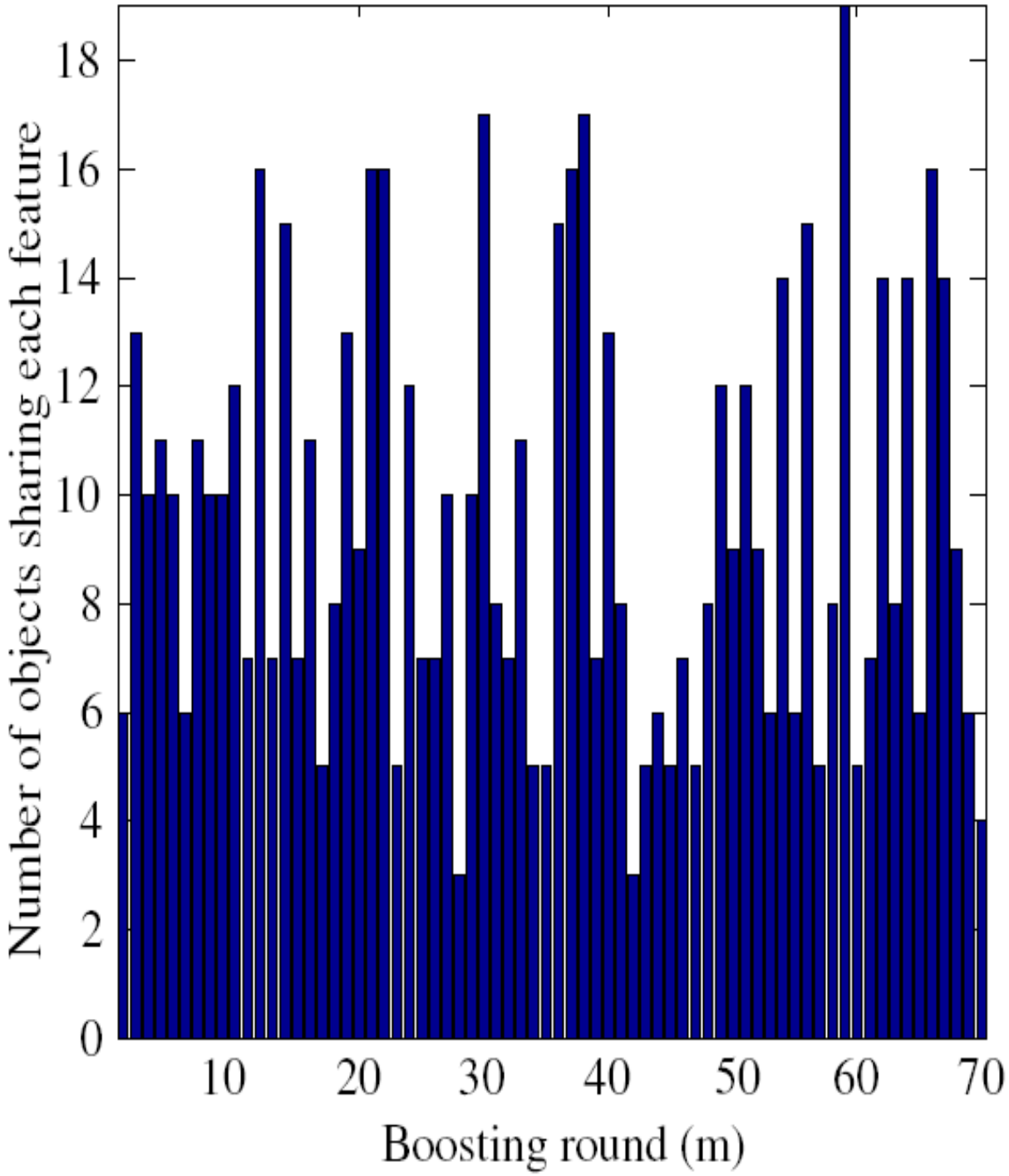
- Multiclass object recognition
  - Dataset: LabelMe
  - 21 objects, 50 samples per object
  - 500 rounds
- Multiview car recognition
  - Train on LabelMe, test on PASCAL
  - 12 views, 50 samples per view
  - 300 rounds

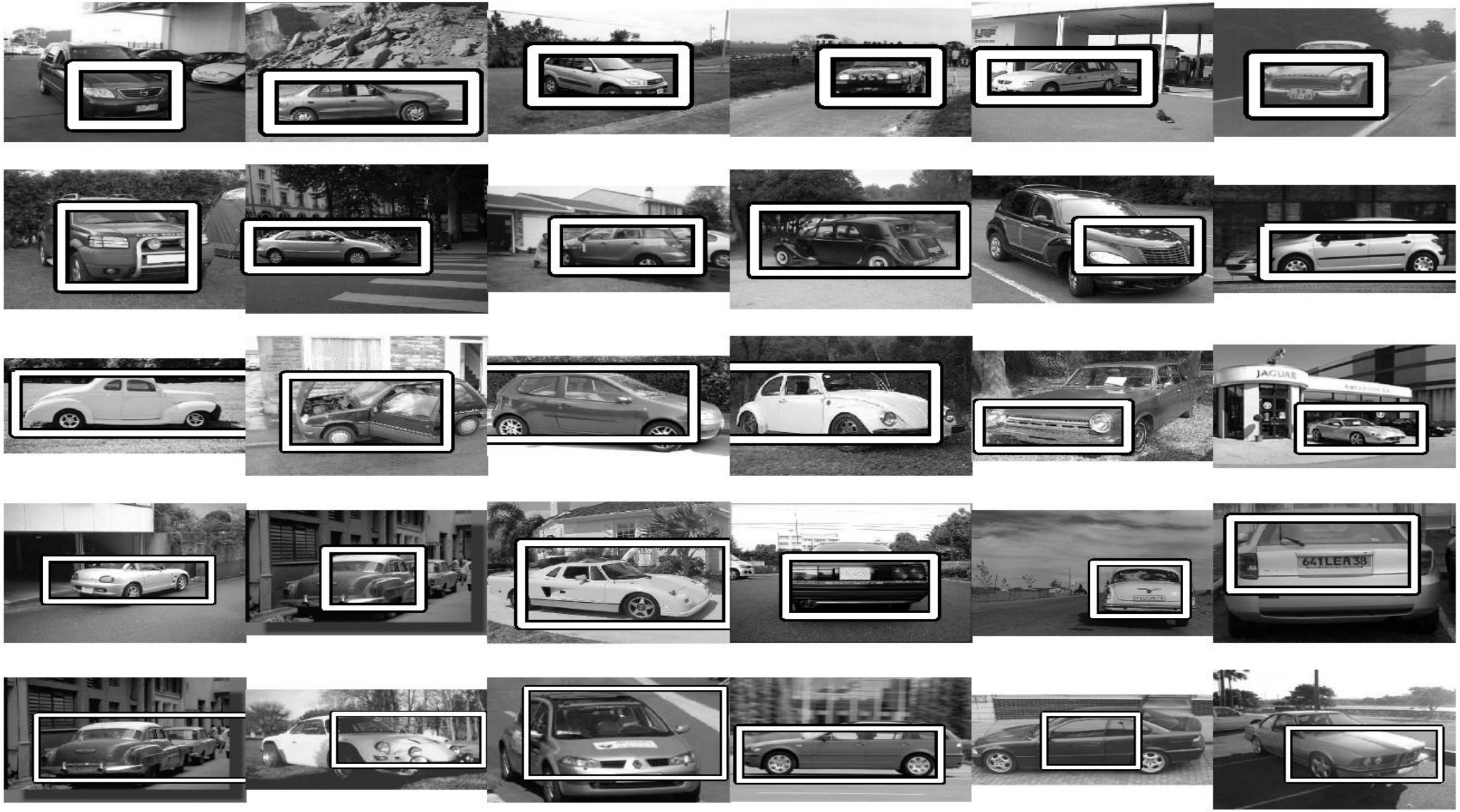


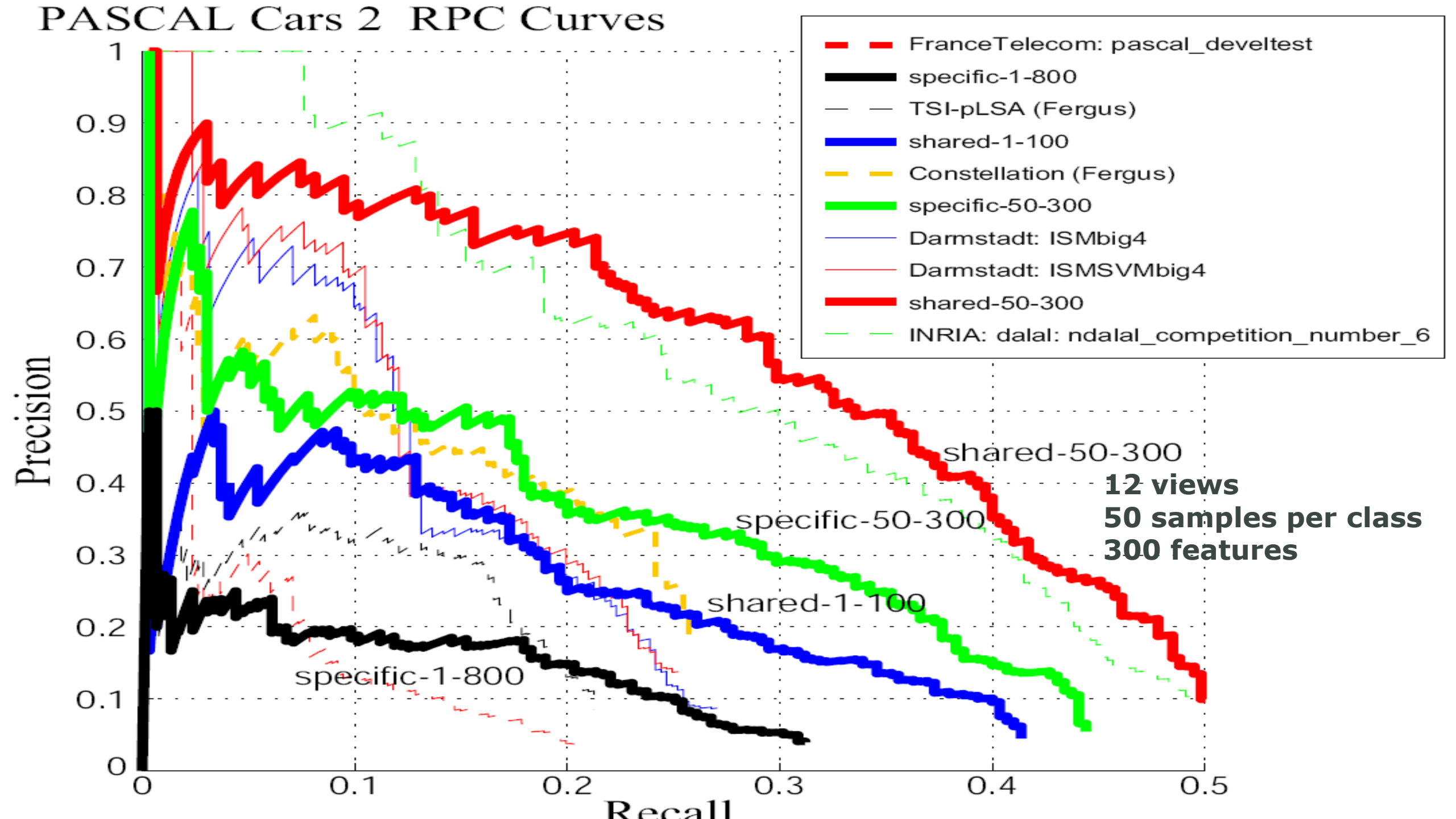

70 features 20 tr. samples
15 features 20 tr. samples
15 features 2 tr. samples
70 features 20 tr. samples
15 features 20 tr. samples
15 features 2 tr. samples
70 features 20 tr. samples
15 features 20 tr. samples
15 features 2 tr. samples
Screen
Chair
Car side
Can
Person
Trafic light
Stop
STOP
Poster
Keyboard
Mouse
Trash
Mug
One way sign
Light
Car frontal
Bottle
Mouse pad
Head
Speaker
Do not enter
DO NOT
ENTER
Computer
0
1


Average area under ROC
Boosting round (m)
Sharing features
Class-specific features
1
0.95
0.85
0.75
0.65
0.55
0.5
10
20
30
40
50
60
70

b)


Number of objects sharing each feature
Boosting round (m)
18
16
14
12
10
8
6
4
2
0
10
20
30
40
50
60
70


JAGUAR
641LEA38


PASCAL Cars 2 RPC Curves
FranceTelecom: pascal_develtest
specific-1-800
TSI-pLSA (Fergus)
shared-1-100
Constellation (Fergus)
specific-50-300
Darmstadt: ISMbig4
Darmstadt: ISMSVMbig4
shared-50-300
INRIA: dalal: ndalal_competition_number_6
shared-50-300
specific-50-300
shared-1-100
specific-1-800
12 views
50 samples per class
300 features
Precision
Recall
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Conclusion

- Joint Boosting indeed works
  - Especially when the number of images per class is not too small (otherwise NN)
- Better performance in the presence of
  - Many classes, each class has only a new samples
  - Introduce regularization that reduce over fitting
- Disadvantages
  - Train slowly, O(C^2).

The End