PASCAL VOC 2010 ECCV 2010/Crete/Greece CASIA


# Object Detection by Context and Boosted HOG-LBP

Yinan Yu, Junge Zhang, Yongzhen Huang, Shuai Zheng, Weiqiang Ren, Chong Wang

Advisors: Kaiqi Huang, Tieniu Tan

National Laboratory of Pattern Recognition
Institute of Automation, CAS
Beijing, P. R. China

# Background

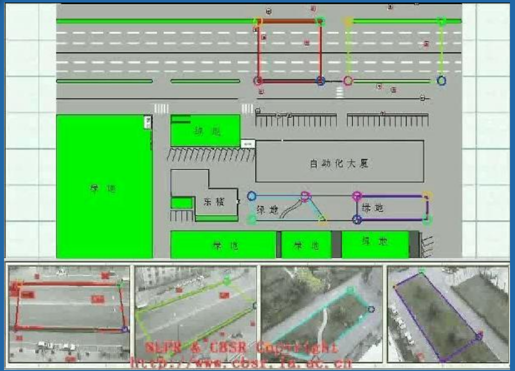


- *Object Detection is a fundamental and crucial module in Intelligent Visual Surveillance.*

*Anomaly Detection*

*Object Detection*


*Object Classification in Traffic Scenes*




*Intelligent Visual Surveillance System*

# Background

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

- *Feature, Framework and Context*
- *The framework used in our method*

  *Deformable Part Model (UOCTTI)*

# Background

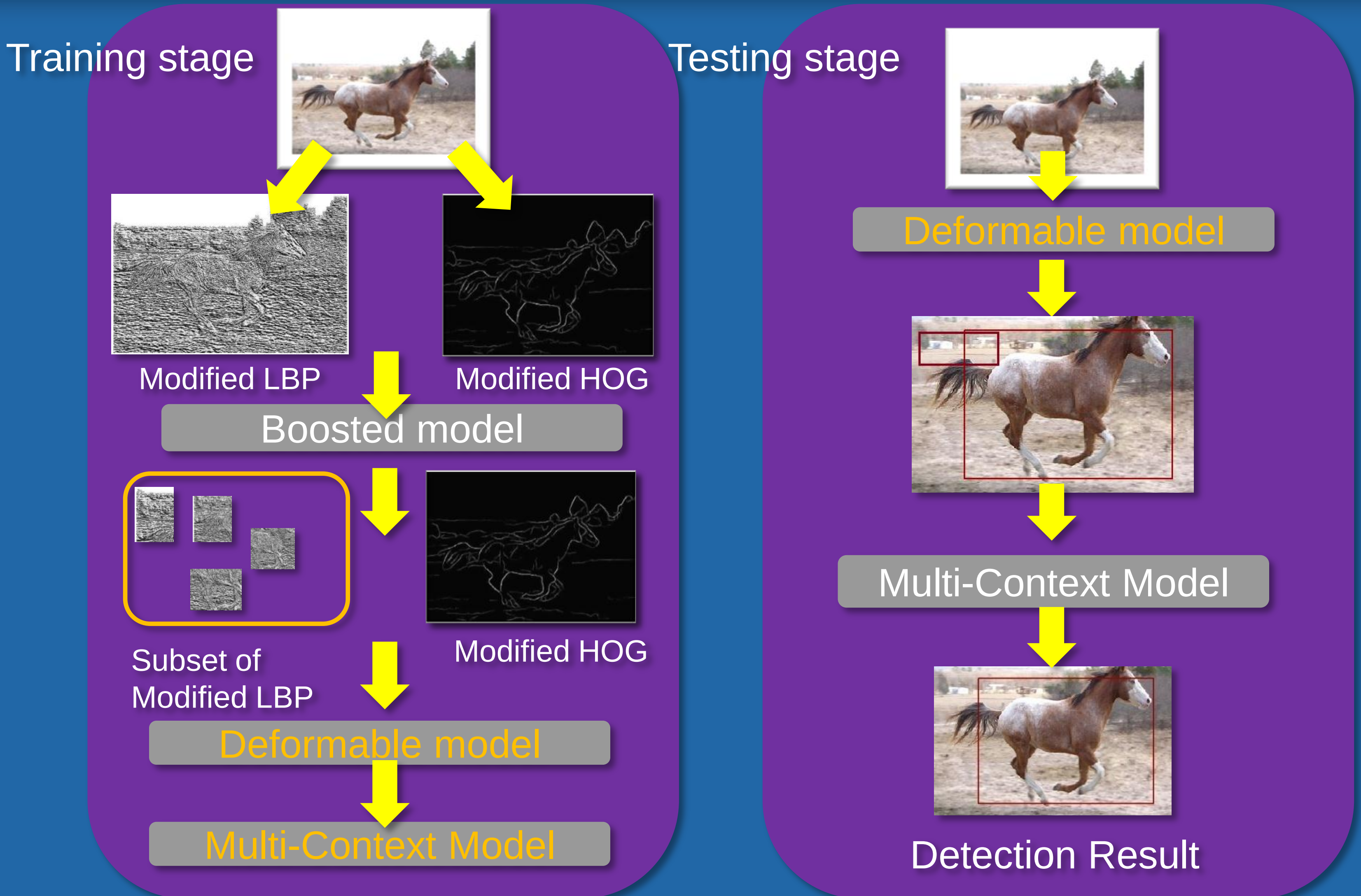


- *Object Detection*
  - *Video surveillance, Human-Machine Interface, Multimedia analysis…*

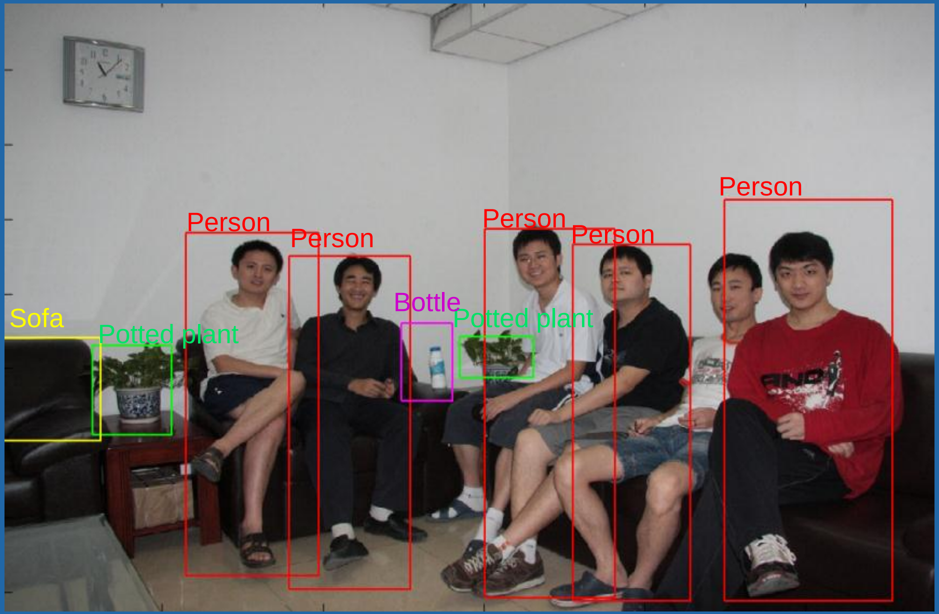
- *Feature, Framework and Context*
- *The framework used in our method*

  *Deformable Part Model (UOCTTI)*

# Overview of our method



# Overview of our method




Our team family

- **Boosted HOG-LBP**
- **Distinctive Multi-Context**

# Boosted HOG-LBP I : Multi-Features



Used features

- Shape feature

  Enhanced HOG, based on UOCTTI (CVPR 2008)

- Texture feature

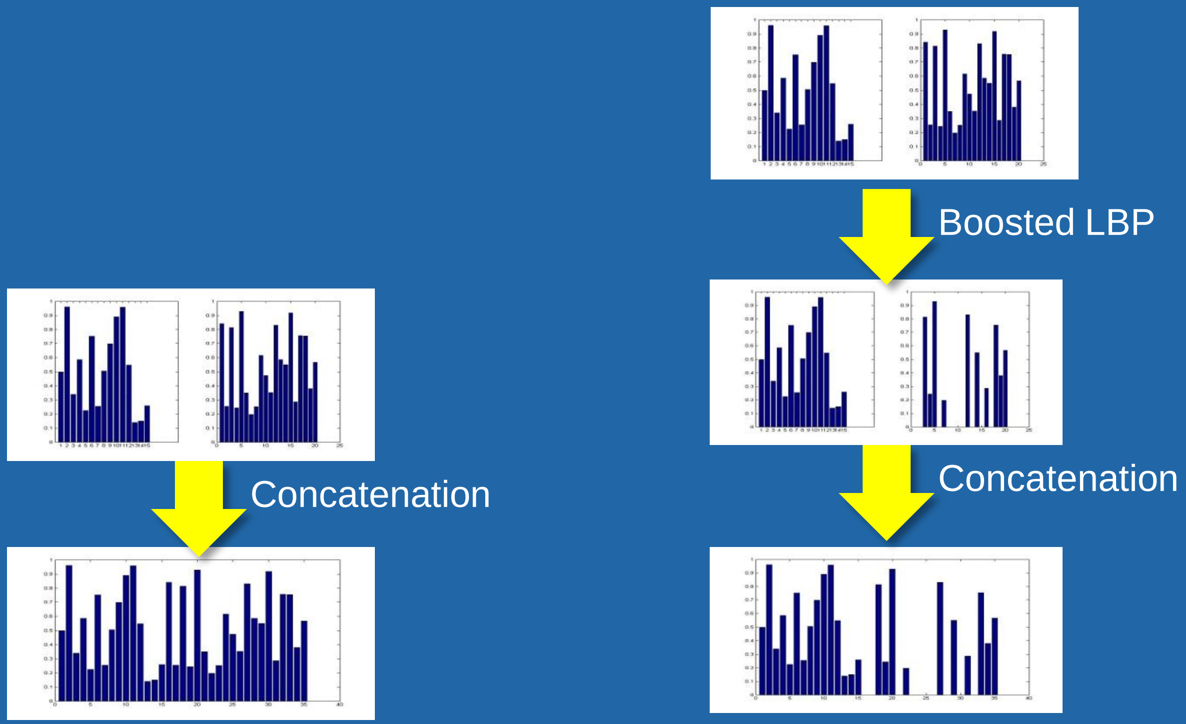
  Enhanced Uniform LBP, based on T. Oiala (PR 1998)

Shape information performs better than texture information for detection tasks, but they complement each other.

# Boosted HOG-LBP II : Combination

How to combine them?

- Concatenation of HOG and LBP
- Concatenation of HOG and a subset of LBP




Naïve Features Combination

Boosted Features Combination

# Boosted HOG-LBP II：Naïve Combination

## Some experiments on VOC2007

### Naïve combination

| Category | HOG | Naïve combination | Improvement |
|---|---|---|---|
| Person | 36.2% | 37.2% | **1.0%** |
| Chair | 16.5% | 15.2% | **-1.3%** |

- Not all LBP features improve the performance.
- See the chair class. It’s believed that some LBP features damage the detection performance

Baseline is the UOCTTI version 3.1

# Boosted HOG-LBP II：Boosted Combination

Gentle-Boost is used for feature selection

A subset of LBP features are selected, and most noisy features are removed

| Category | HOG(%) | HOG+BoostedLBP(%) | Improvement(%) |
|---|---|---|---|
| Person | 36.2 | 40.2 | **4.0** |
| Chair | 16.5 | 17.9 | **1.4** |
| Cat | 16.3 | 20.0 | **3.7** |
| Dog | 5.0 | 14.5 | **9.5** |

- The most effective LBP features are selected to enhance object description
- Object with rich textures benefits more from the Boosted LBP feature

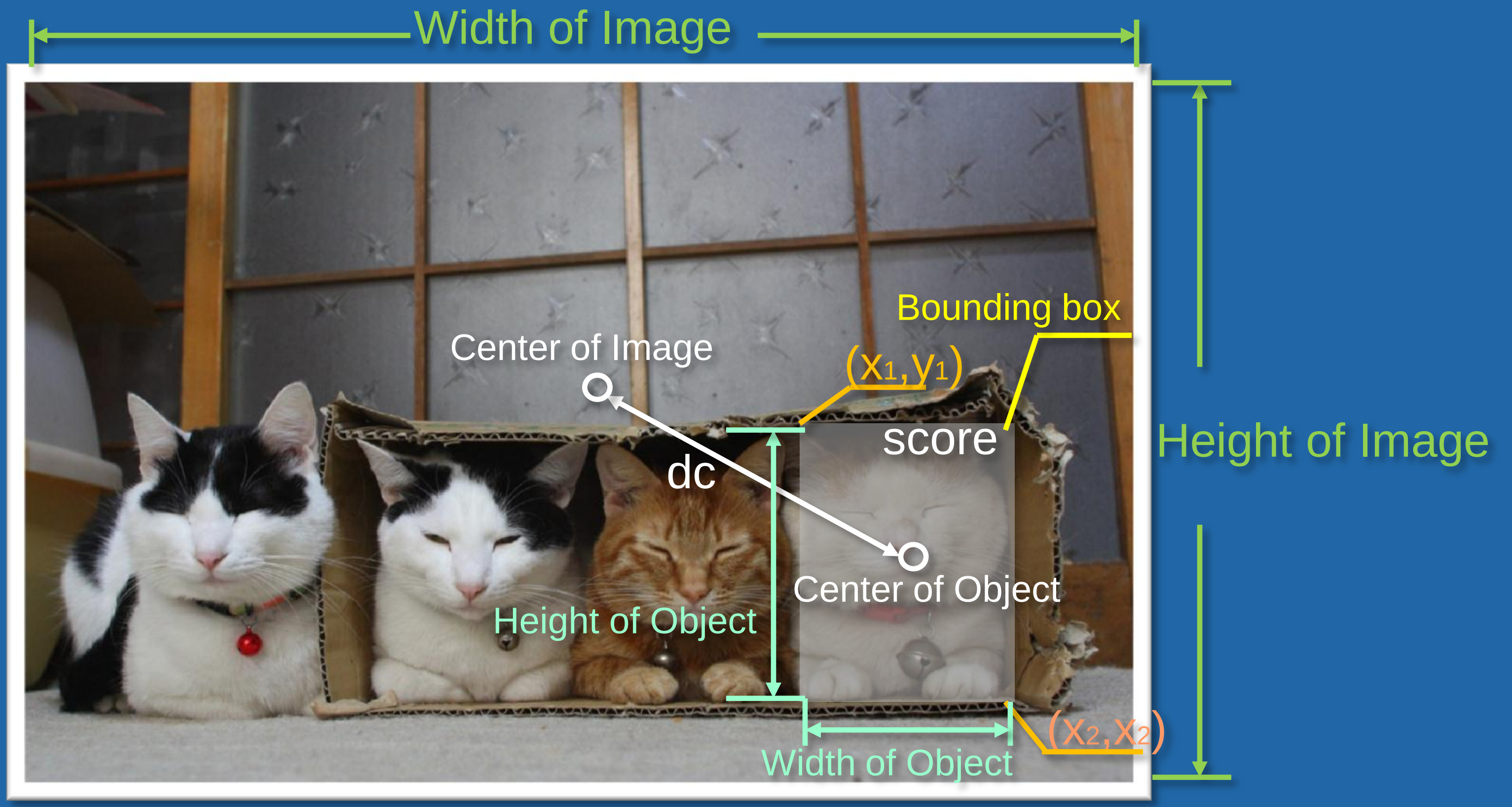
# Multi-Context

- Spatial Context
- Global Context
- Inter-Class Context

# Multi-Context : Spatial Context

For each detected bounding box (s, x1, x2, y1, y2)

(x1/width, x2/width, y1/height, y2/height, dc/diagonal, SizeO/SizeI)



# Multi-Context : Global Context

Global context:

- Scores from classification

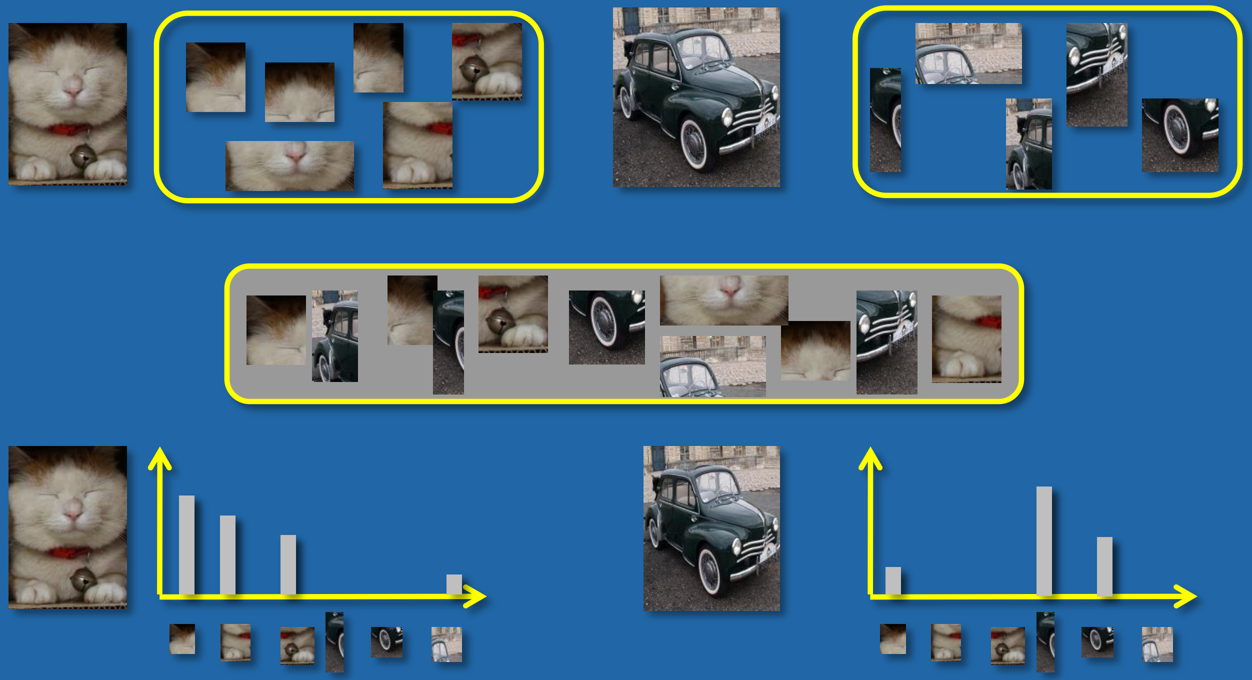
NLPR_VSTAR_CLS_DICTLEARN
4 No.1 positions in VOC 2010 Classification Challenge

- Maximum box scores (20 classes) (UOCTTI )

# Multi-Context : Global Context

NLPR_VSTAR_CLS_DICTLEARN

- Dense SIFT descriptor: grid = 4, #scale = 5
- Saliency Coding
- Linear SVM



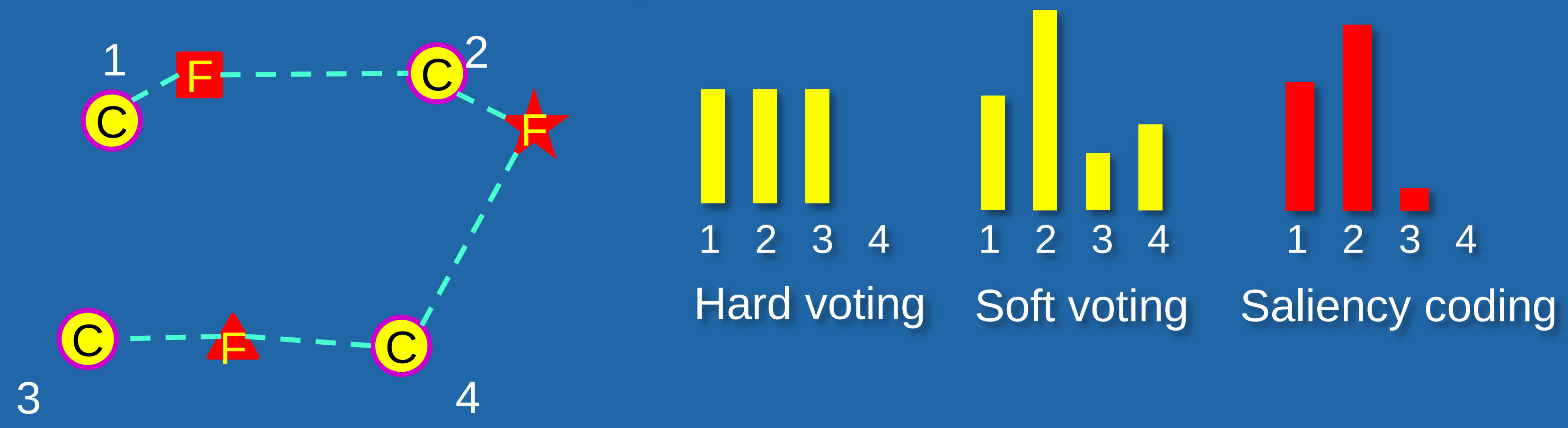
# Multi-Context : Global Context

Saliency Coding is proposed to find more salient weights

Our saliency definition is :

The closest distance << the second closest one

Our solution : 1-d1/d2 (the ratio between nearest distance and second nearest distance)



# Multi-Context : Global Context

## Dictionary Learning

Proximately solve derivative relation between AP and dictionaries

## Code Relation Modeling

Codes are not independent but restricted one another

# Multi-Context : Inter-Class Context

The inter-class context characterizes a weak spatial relationship between different classes objects

I. Classifiers of separate HOG and LBP are learned for each class (without parts)

II. Find maximum scores of the 20 HOG models and 20 LBP models respectively around the detected bounding boxes.

# Multi-Context : Inter Class Context


- Initial detected objects
- Searching region
- Window with maximum score

**Inter class learning results:**

**Bicycle often occurs near person**

# Multi-Context : Final Contexts

- The original score : 1
- Spatial Context : 6
- Global Context : 20+20
- Inter-Class Context : 20+20
- Total number of context features : 1+6+40+40=87

| | | | | |
|---|---|---|---|---|
| S | GC_DET | GC_CLS | ICC_HOG | ICC_LBP |

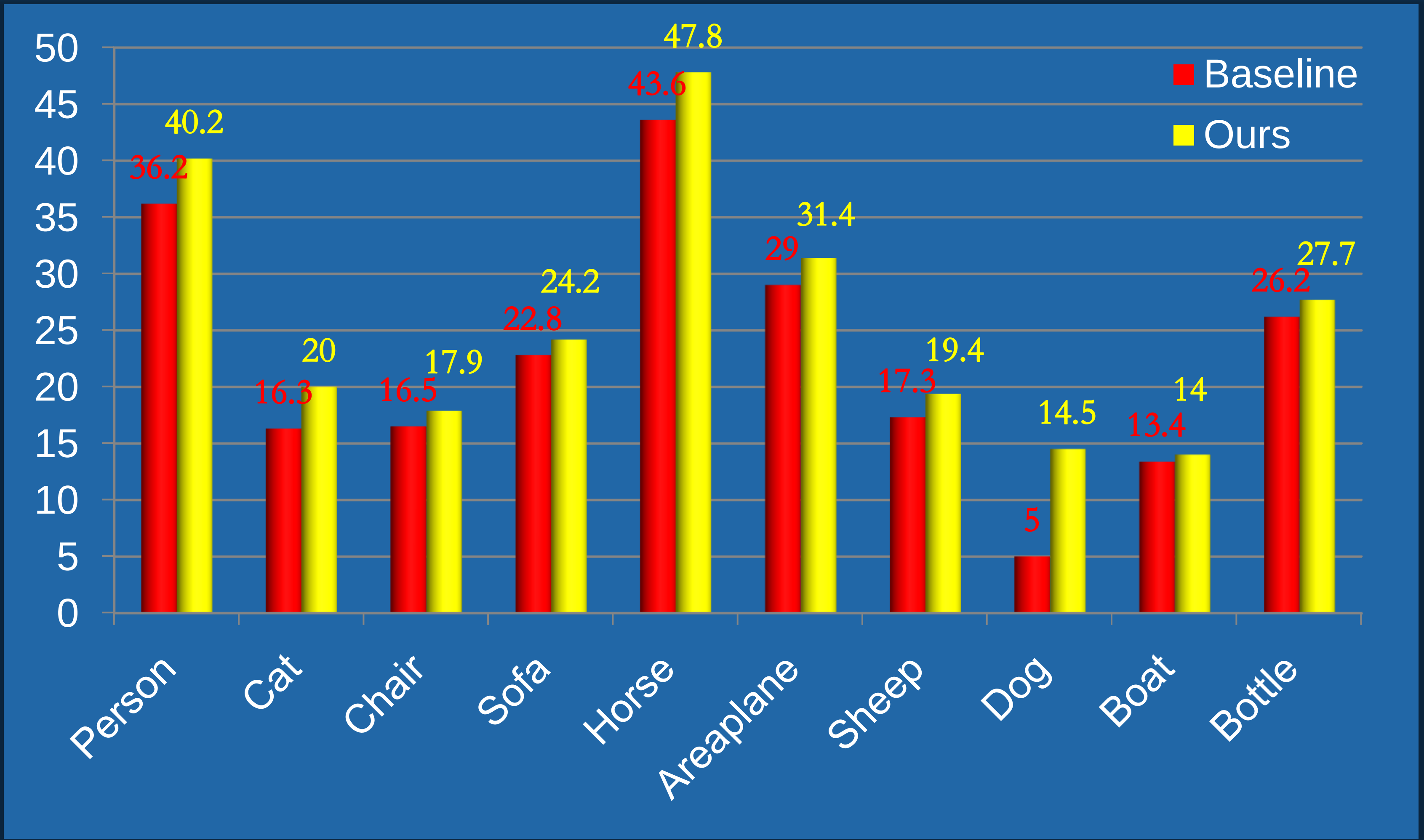
# Multi-Context : Context Learning

SVM is employed for context learning:

• Linear and Kernel SVMs are effective for learning contexts

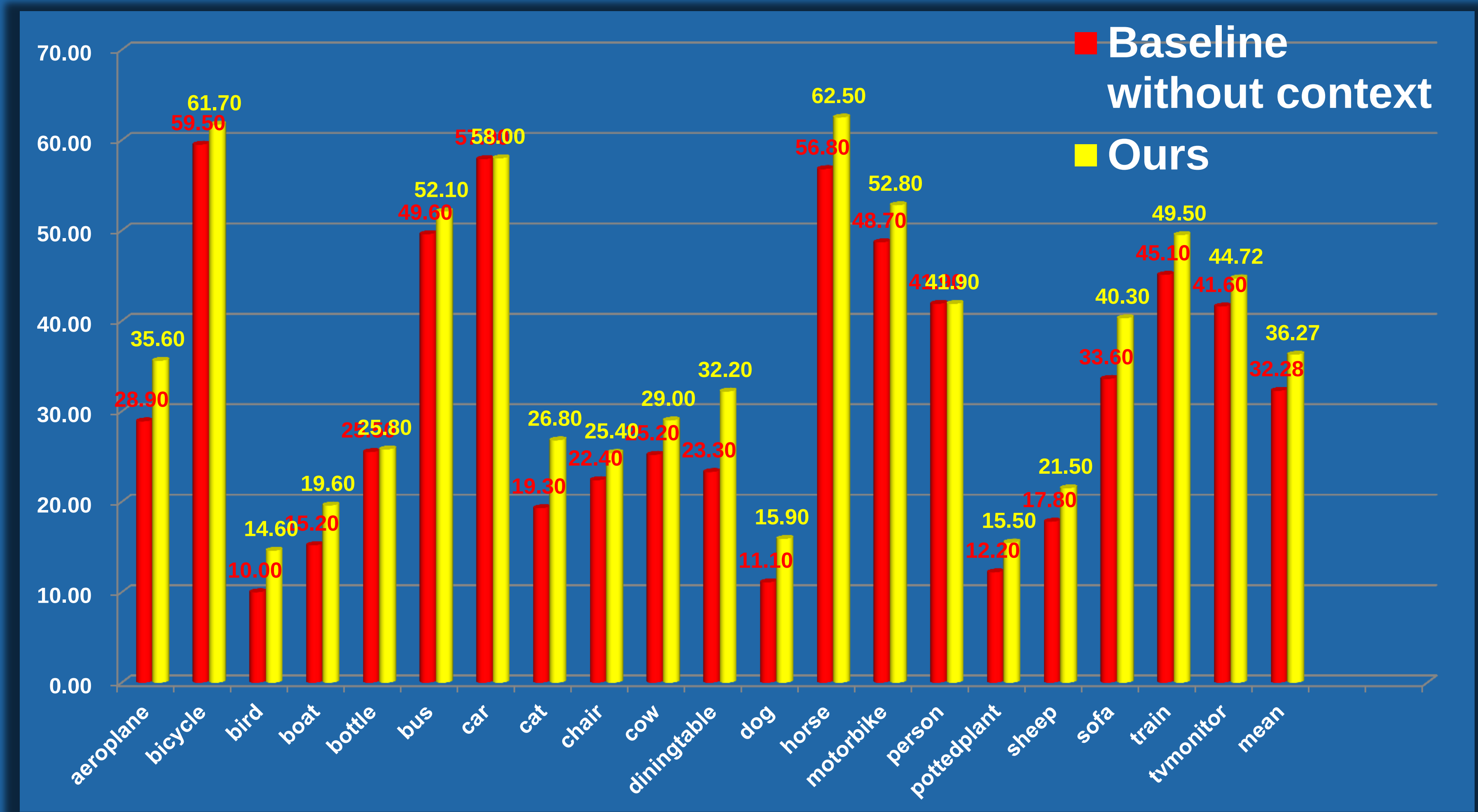
• **RBF-SVM performs best in most classes**

# Results : Boosted HOG-LBP VOC2007 (without contexts)




Baseline is the UOCTTI version 3.1 VOC2007

# Results : Only Contexts - Improvement : 4%




Baseline is the UOCTTI version 4, VOC2007

# Results : VOC2010

- **Six classes rank first**
- **Five classes rank second**
- **36.8% mean AP**

| categories | AP |
|---|---|
| aeroplane | 53.3 |
| bicycle | **55.3** |
| bird | **19.2** |
| boat | **21** |
| bottle | 30 |
| bus | **54.4** |
| car | **46.7** |
| cat | 41.2 |
| chair | **20** |
| cow | **31.5** |
| diningtable | 20.7 |
| dog | 30.3 |
| horse | **48.6** |
| motorbike | **55.3** |
| person | **46.5** |
| pottedplant | 10.2 |
| sheep | 34.4 |
| sofa | 26.5 |
| train | **50.3** |
| tvmonitor | 40.3 |
| mAP | **36.8** |

# Conclusion

- For object representation, boosted multiple features are more distinctive than naïve combination
- Utilizing multiple contexts performs well in depressing false positives
- The multi-context information is very effective for multi-classes object detection.


# Thank You

Contact: ynyu@nlpr.ia.ac.cn