Semi-supervised Object Recognition Using Structure Kernel

Botao Wang, Hongkai Xiong
Institute of Image Communication and Information Processing (IICIP)
Department of Electronic Engineering
Shanghai Jiao Tong University

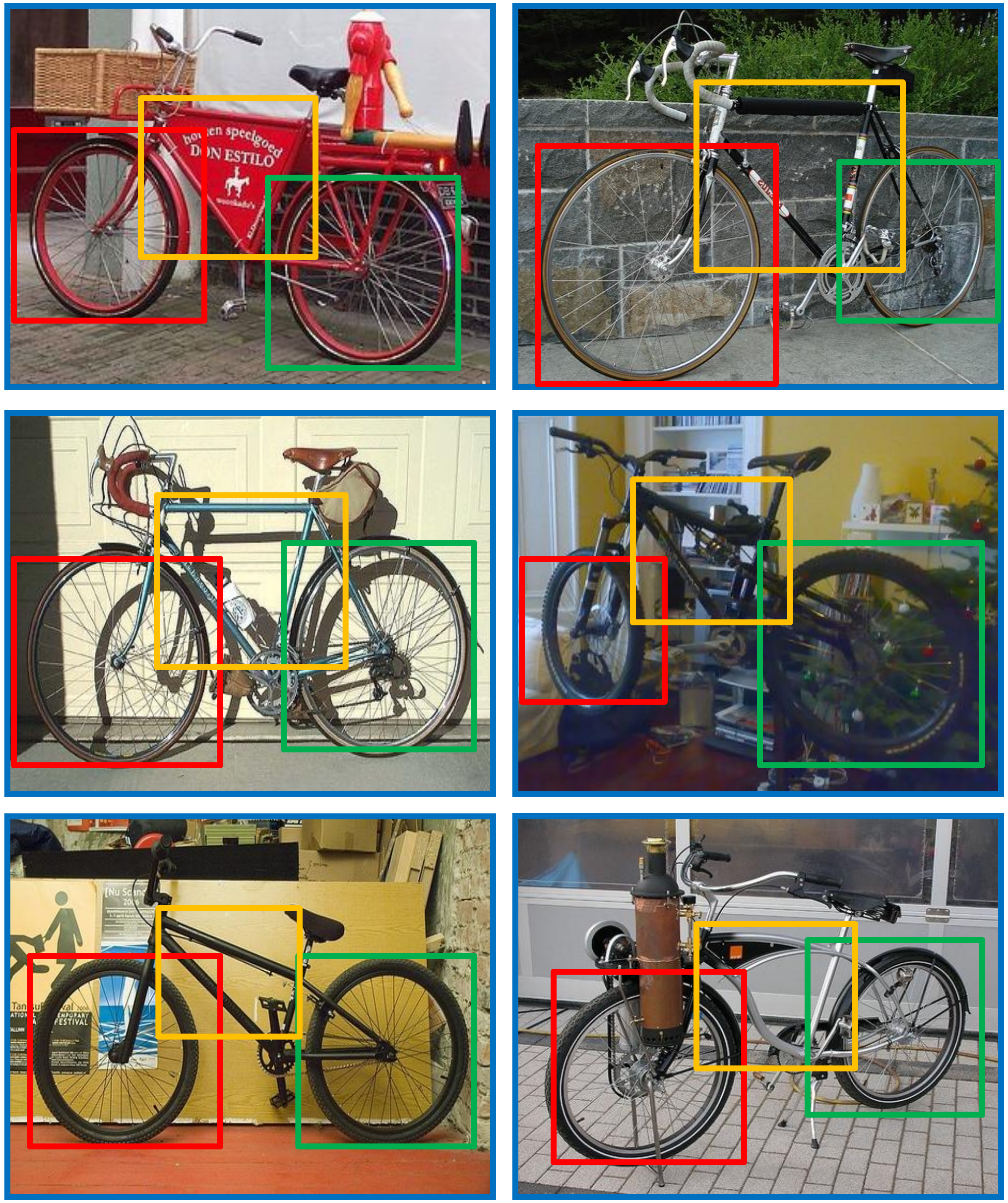
Xiaoqian Jiang
Division of Biomedical Informatics
University of California, San Diego

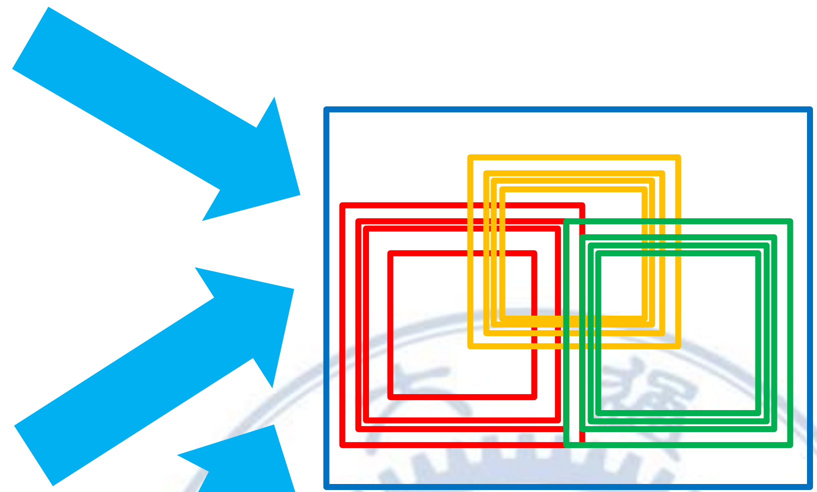
Fan Ling
Wireless Comm. Tech. Limited
QUALCOMM

Outline

- **Multi-scale part-based object model**
- **Structure kernel**
- **Optimizing kernel parameters and training classifier**
- **Experimental results**

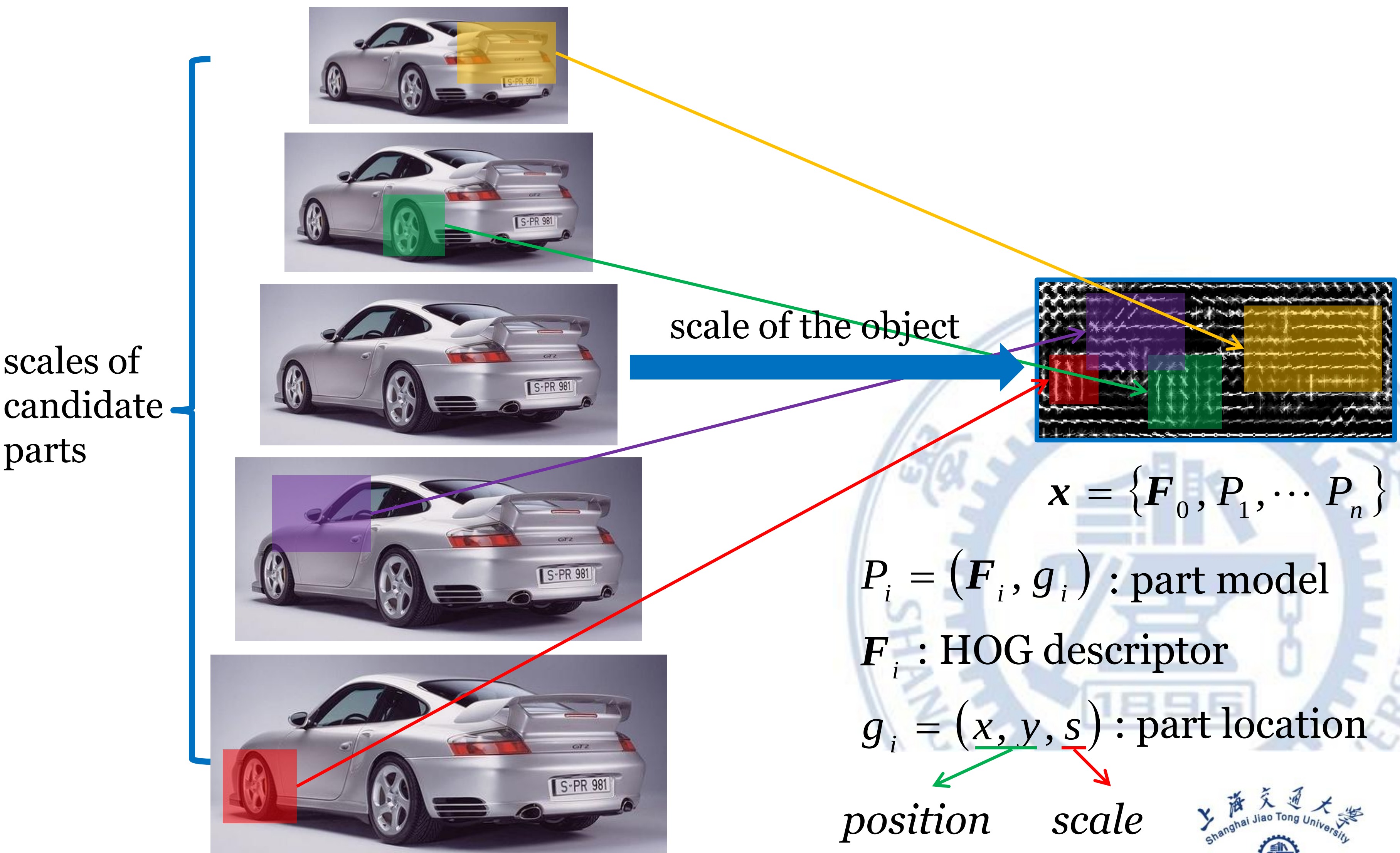
Part-based object representation





The part sizes of objects in real scenarios vary within certain scales because of intra-class variations, viewpoints, etc.

Multi-scale part-based object model



$$\boldsymbol{x} = \{\boldsymbol{F}_0, P_1, \cdots P_n\}$$

$P_i = (\boldsymbol{F}_i, g_i)$: part model

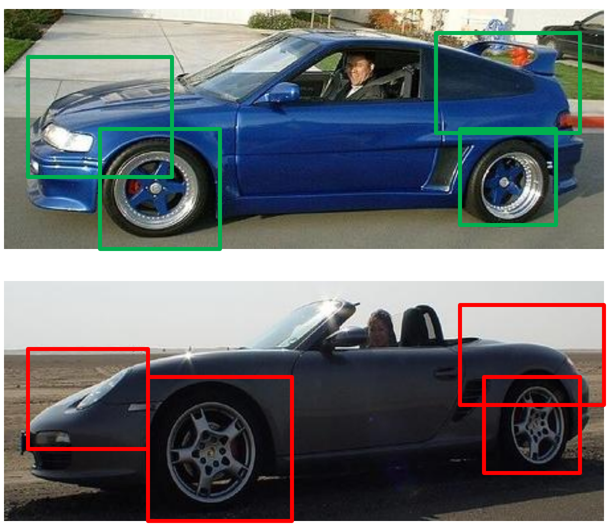
$\boldsymbol{F}_i$: HOG descriptor

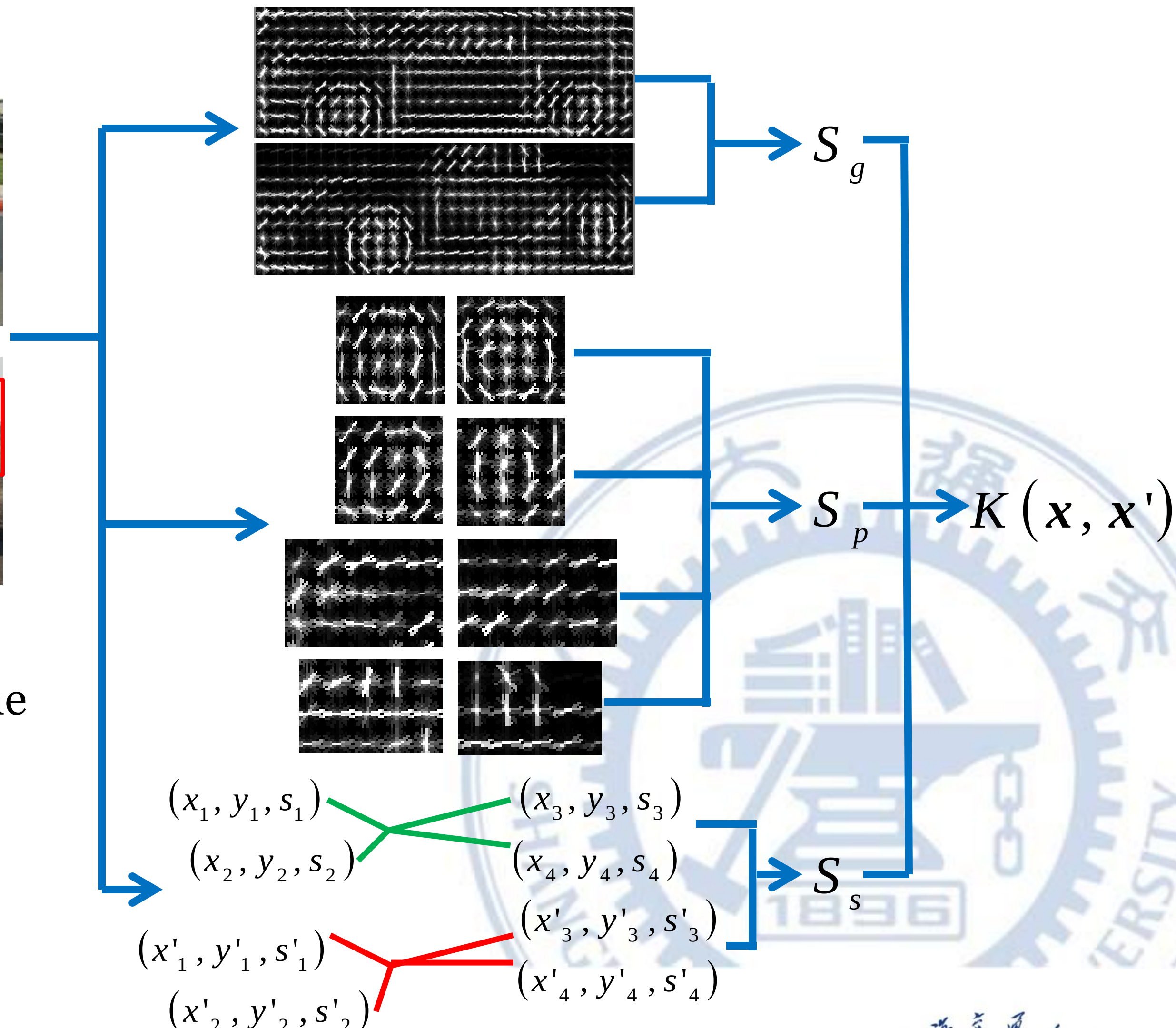
$g_i = (\underbrace{x, y}_{\textit{position}}, \underbrace{s}_{\textit{scale}})$: part location

Outline

- Multi-scale part-based object model
- Structure kernel
- Optimizing kernel parameters and training classifier
- Experimental results

Structure kernel





Structure kernel measures the similarity of two part-based represented objects in

(1) Global visual similarity;

(2) Part visual similarity;

(3) Spatial similarity.

Structure kernel

Global visual similarity

$$S_g = k(\boldsymbol{F}_0, \boldsymbol{F}_0')$$

$\boldsymbol{F}_0, \boldsymbol{F}_0'$: global feature vectors

k : base kernel

Part visual similarity

$$S_p = \sum_{i=1}^{n} w_i k(\boldsymbol{F}_i, \boldsymbol{F}_i')$$

$\boldsymbol{F}_i, \boldsymbol{F}_i'$: part feature vectors

k : base kernel

w_i : part weight

Spatial similarity

$$S_s = \sum_{i=1}^{n} \exp\left\{-\gamma(g_i - g_i')^2\right\}$$

g_i, g_i' : part locations

γ : penalty factor

Structure kernel

The structure kernel K measures the similarity between two part-based represented objects $\boldsymbol{x}$ and $\boldsymbol{x}'$ as

$$K(\boldsymbol{x}, \boldsymbol{x}') = k(\boldsymbol{F}_0, \boldsymbol{F}_0') + \sum_{i=1}^{n} w_i k(\boldsymbol{F}_i, \boldsymbol{F}_i') + \lambda \sum_{i=1}^{n} \exp\left\{-\gamma(g_i - g_i')^2\right\}$$

λ : parameter that balances the visual similarity and spatial similarity.

Outline

- Multi-scale part-based object model
- Structure kernel
- Optimizing kernel parameters and training classifier
- Experimental results

Training and optimizing parameters

1. Optimizing kernel parameters

$$K(\boldsymbol{x},\boldsymbol{x}') = k(\boldsymbol{F}_0,\boldsymbol{F}_0') + \sum_{i=1}^{n} \boxed{w_i}\, k(\boldsymbol{F}_i,\boldsymbol{F}_i') + \lambda \sum_{i=1}^{n} \exp\left\{-\gamma (g_i - g_i')^2\right\}$$

Inter-class objects

Intra-class objects

Low similarity

$$\textbf{min} \sum_{p,q} \sum_{i=1}^{n} w_i k\left(F_i^{\,p}, F_i^{\,q}\right) \qquad (1)$$

High similarity

$$\textbf{max} \sum_{p,q} \sum_{i=1}^{n} w_i k\left(F_i^{\,p}, F_i^{\,q}\right) \qquad (2)$$

$$\max_{\boldsymbol{w}} \sum_{p=1}^{N} \sum_{q=p}^{N} y_q \sum_{i=1}^{n} w_i k\left(\boldsymbol{F}_i^{\,p}, \boldsymbol{F}_i^{\,q}\right) \qquad (3)$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{n} w_i = 1 \\ w_i \geq 0, i = 1, \cdots n \\ y_p = +1 \end{cases}$$

上海交通大学 Shanghai Jiao Tong University

Normalized stochastic gradient ascent algorithm

$$\max_{\boldsymbol{w}} \; E(\boldsymbol{w}) = \sum_{p=1}^{N} \sum_{q=p}^{N} y_q \sum_{i=1}^{n} w_i k\left(\boldsymbol{F}_i^{\,p}, \boldsymbol{F}_i^{\,q}\right)$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{n} w_i = 1 & (1) \\ w_i \geq 0, i = 1, \cdots n & (2) \\ y_p = +1 \end{cases}$$

Sub-gradient of samples $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$: $\nabla = y_v \left[k\left(\boldsymbol{F}_1^{\,u}, \boldsymbol{F}_1^{\,v}\right), k\left(\boldsymbol{F}_2^{\,u}, \boldsymbol{F}_2^{\,v}\right), \cdots, k\left(\boldsymbol{F}_n^{\,u}, \boldsymbol{F}_n^{\,v}\right)\right]$

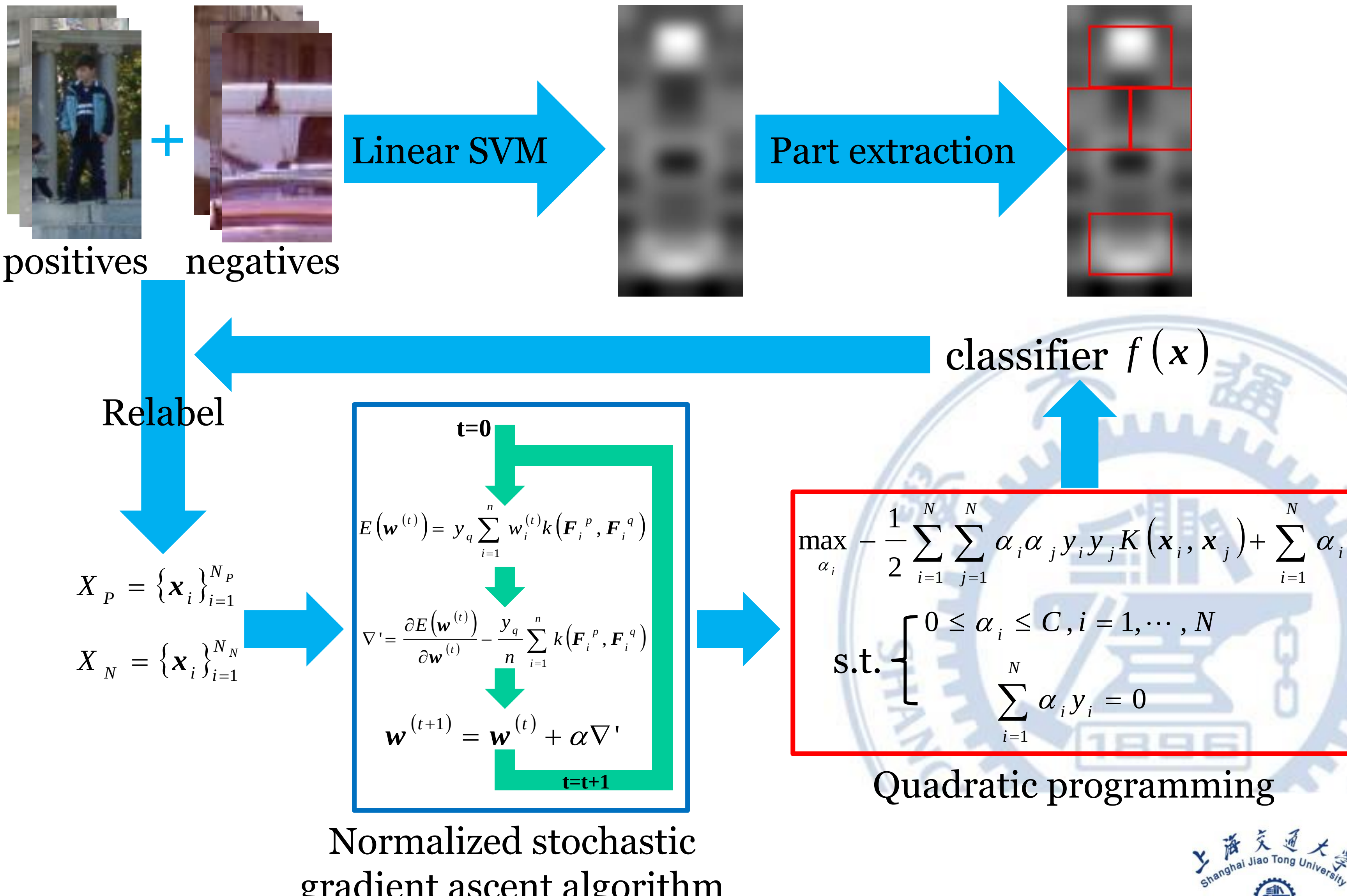
Normalized sub-gradient:

$$\nabla' = y_v \left[k\left(\boldsymbol{F}_1^{\,u}, \boldsymbol{F}_1^{\,v}\right) - \frac{1}{n}\sum_{i=1}^{n} k\left(\boldsymbol{F}_i^{\,u}, \boldsymbol{F}_i^{\,v}\right), \cdots, k\left(\boldsymbol{F}_n^{\,u}, \boldsymbol{F}_n^{\,v}\right) - \frac{1}{n}\sum_{i=1}^{n} k\left(\boldsymbol{F}_i^{\,u}, \boldsymbol{F}_i^{\,v}\right) \right]$$

If $\boldsymbol{w}^{(t)}$ satisfies (1) (2), $\boldsymbol{w}^{(t+1)} = \boldsymbol{w}^{(t)} + \alpha \nabla'$ also satisfies (1) (2) and

$$E\left(\boldsymbol{w}^{(t+1)}\right) \geq E\left(\boldsymbol{w}^{(t)}\right)$$

Training classifier with latent SVM



Outline

- Multi-scale part-based object model
- Structure kernel
- Optimizing kernel parameters and training classifier
- Experimental results

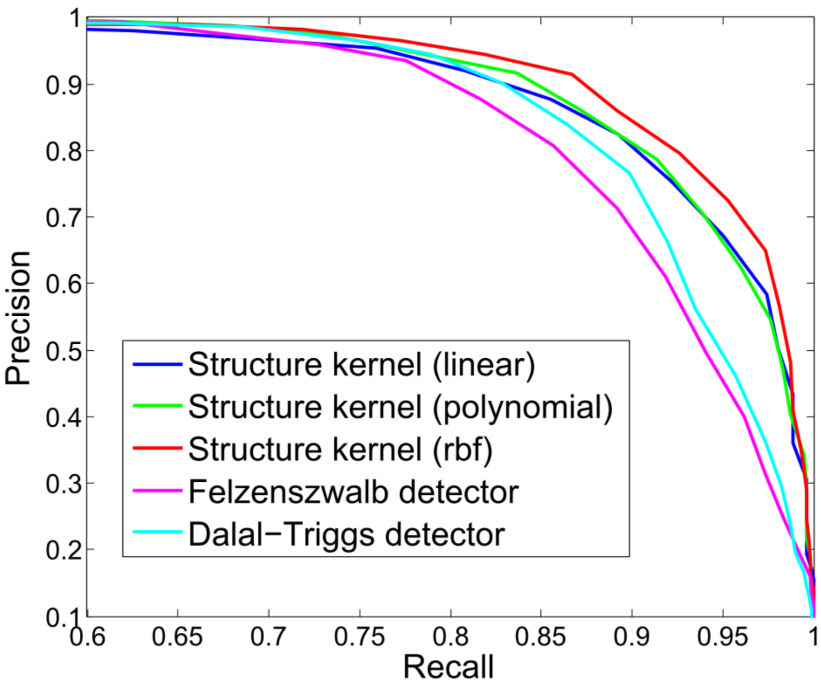
INRIA person dataset evaluation

1. **Different base kernels**

   a. linear kernel: $k(\boldsymbol{F}, \boldsymbol{F}') = \boldsymbol{F} \cdot \boldsymbol{F}'$

   b. polynomial kernel: $k(\boldsymbol{F}, \boldsymbol{F}') = (\boldsymbol{F} \cdot \boldsymbol{F}' + 1)^m$

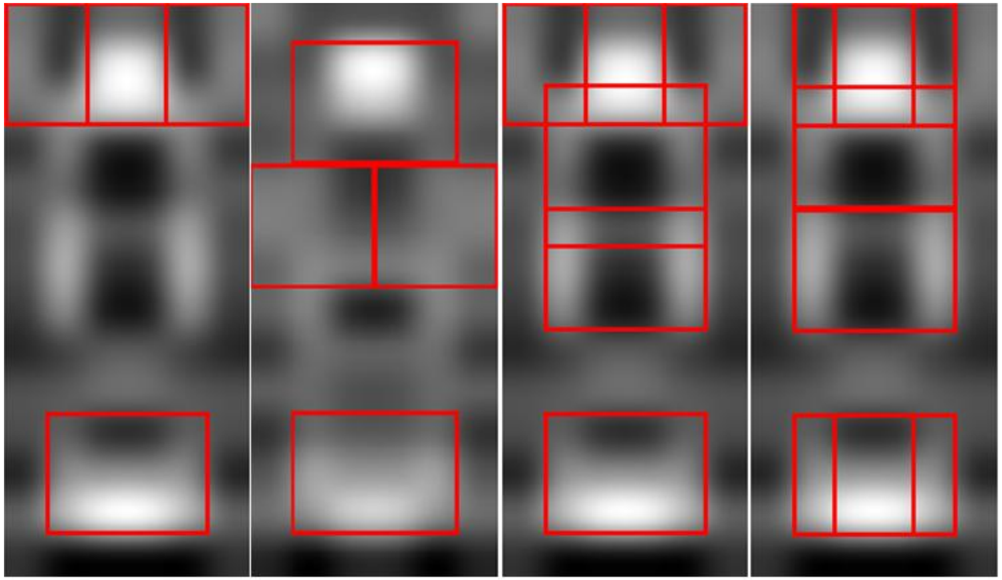
   c. RBF kernel: $k(\boldsymbol{F}, \boldsymbol{F}') = \exp\left\{-\delta \left\|\boldsymbol{F} - \boldsymbol{F}'\right\|^2\right\}$



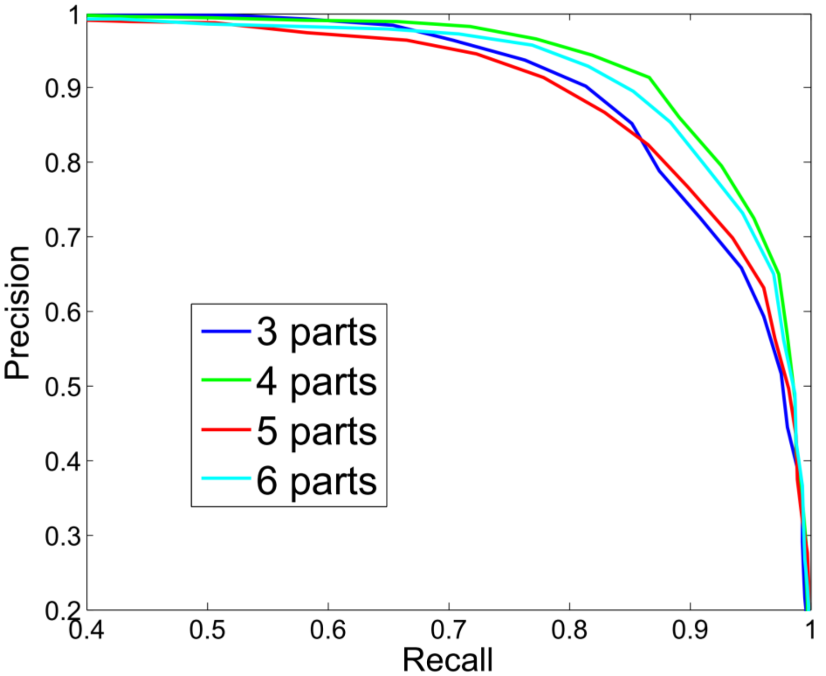
Base kernel	Average precision
SK (linear)	0.938
SK (polynomial)	0.946
SK (radial basis)	0.954
Felzenszwalb	0.917
Dalal-Triggs	0.927

INRIA person dataset evaluation

2. Different part numbers and configurations



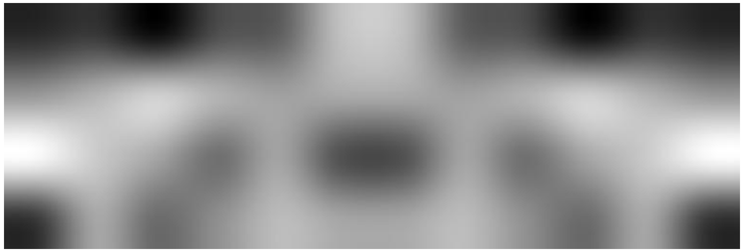
(a) 3 parts (b) 4 parts (c) 5 parts (d) 6 parts



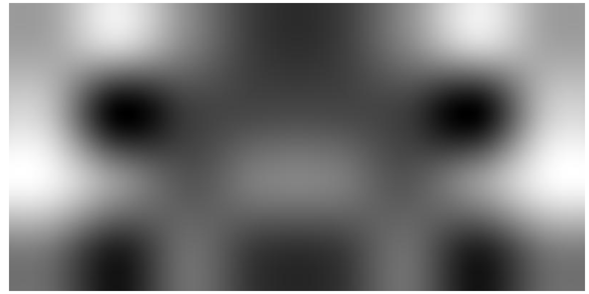
Part number	Average precision
3	0.934
4	0.954
5	0.930
6	0.945

Shanghai Jiao Tong University

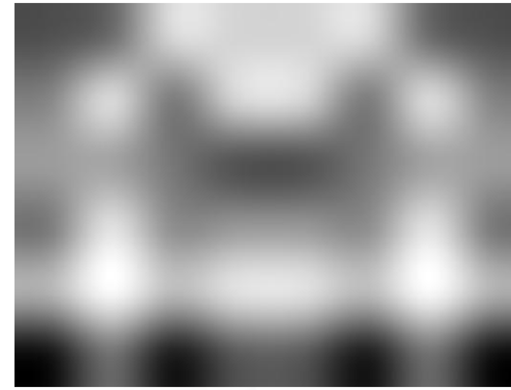
PASCAL VOC 2007 dataset evaluation



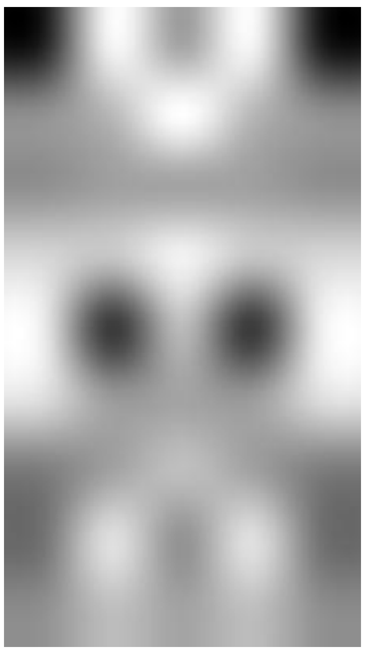
aeroplane



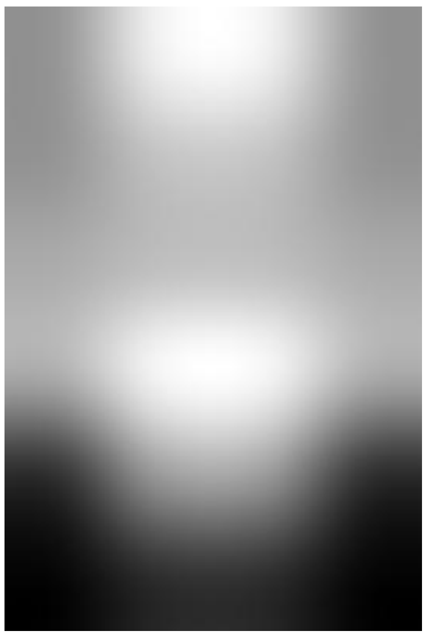
car



motorbike

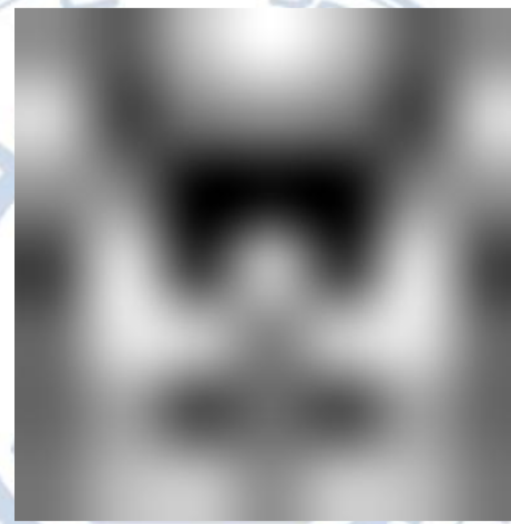


person



potted plant

dog

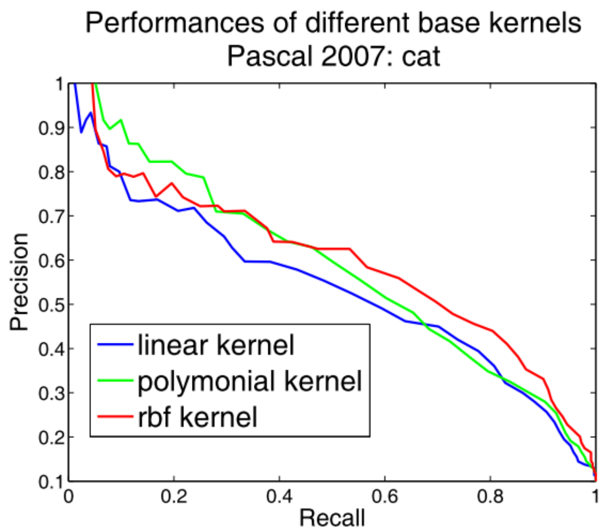


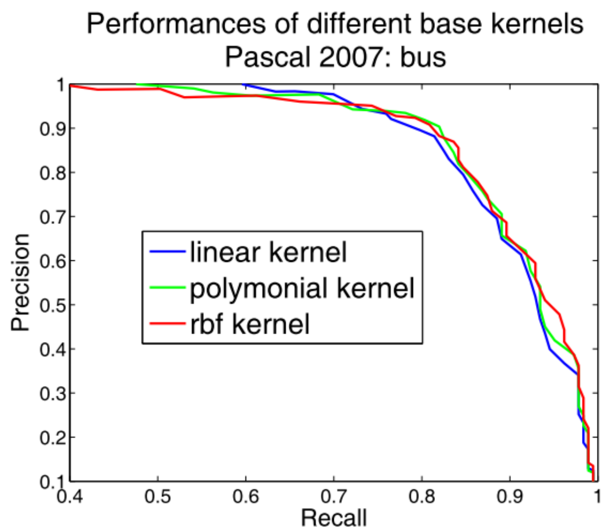
horse

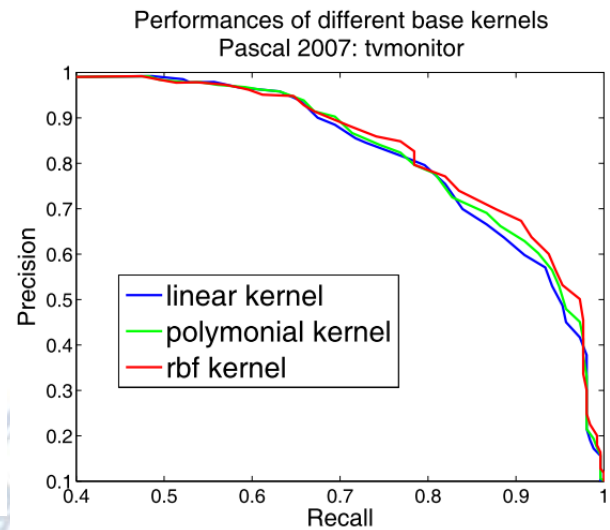
Global models of some object classes in PASCAL VOC 2007 dataset

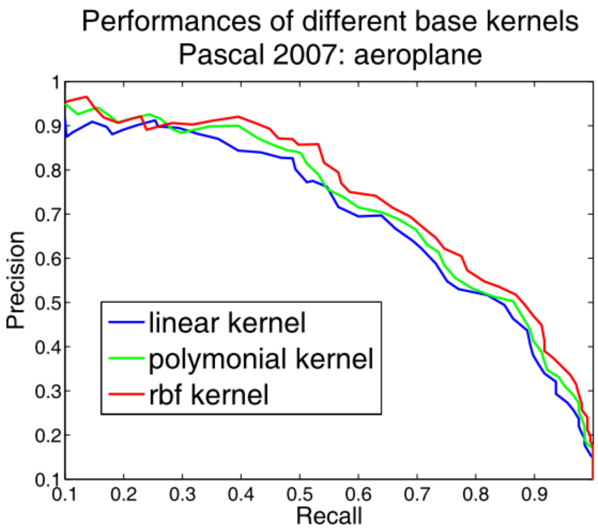
PASCAL VOC 2007 dataset evaluation

1. Different base kernels





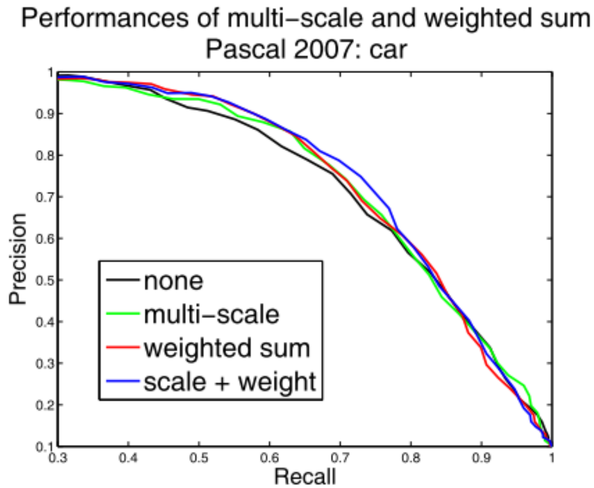


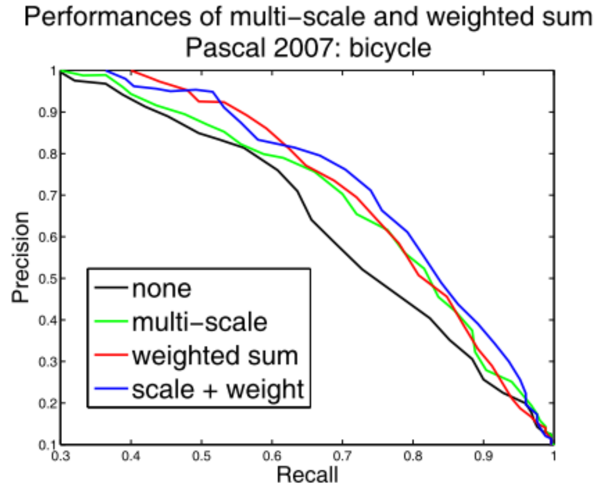


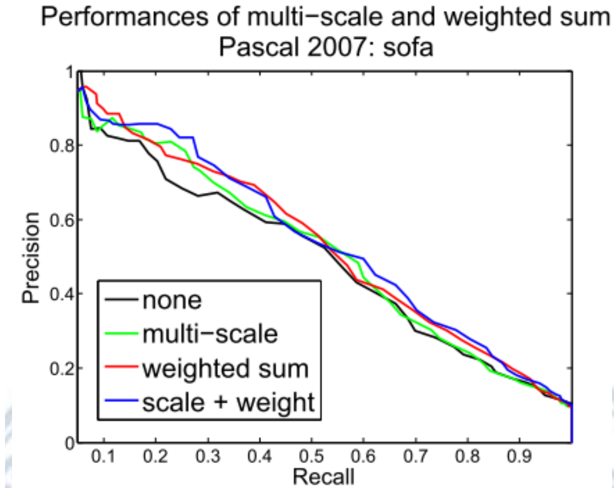
	cat	bus	aeroplane	TV/monitor
Linear kernel	0.538	0.910	0.717	0.887
Polynomial kernel	0.583	0.913	0.740	0.891
RBF kernel	0.596	0.913	0.764	0.897

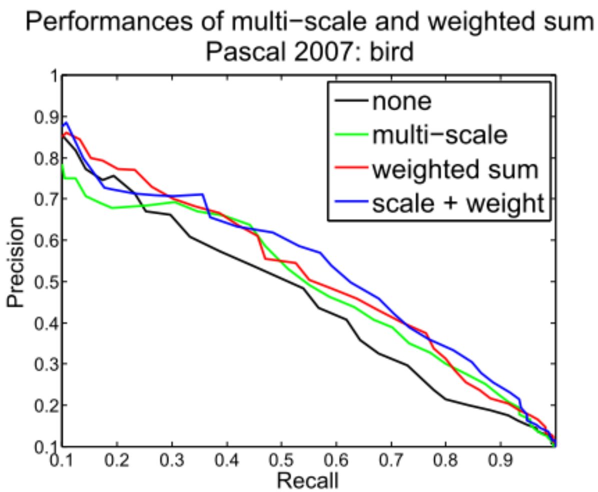
PASCAL VOC 2007 dataset evaluation

2. Multi-scale part representation and weighted combination of part similarities









	car	bicycle	sofa	bird
None	0.793	0.738	0.512	0.483
Multi-scale	0.799	0.776	0.526	0.526
Weighted-combination	0.802	0.792	0.547	0.541
Multi-scale + weighted combination	0.807	0.804	0.554	0.553

Conclusions and future work

- The multi-scale part-based model is more robust to intra-class variations and viewpoint changes than existing part-based models.
- The structure kernel measures the similarity of two part-based objects in both visual appearance and spatial layout in a joint manner.
- The normalized stochastic gradient ascent method makes the kernel parameters adaptive to the distribution of data.
- A more sophisticated cost function can be used to penalize the deformation of objects

Thanks!