# High Level Computer Vision

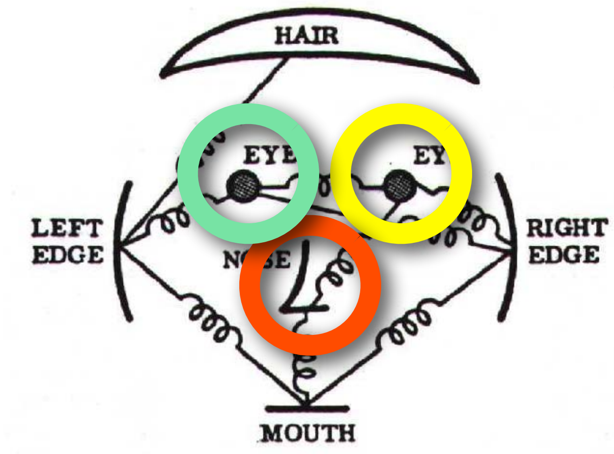
# Part-Based Models for Object Class Recognition Part 2

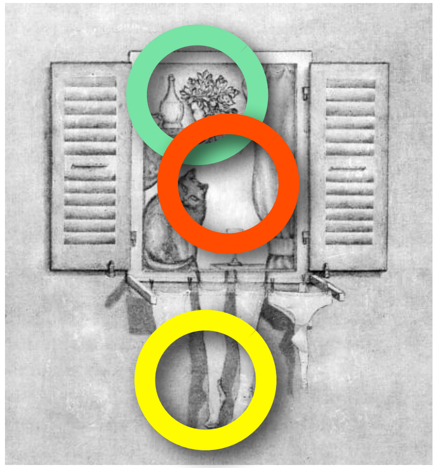
**Bernt Schiele - schiele@mpi-inf.mpg.de**
**Mario Fritz - mfritz@mpi-inf.mpg.de**

**https://www.mpi-inf.mpg.de/hlcv**

# Class of Object Models: Part-Based Models / Pictorial Structures

- Pictorial Structures [Fischler & Elschlager 1973]
  - Model has two components
    - **parts** (2D image fragments)
    - **structure** (configuration of parts)

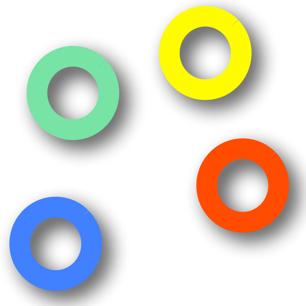


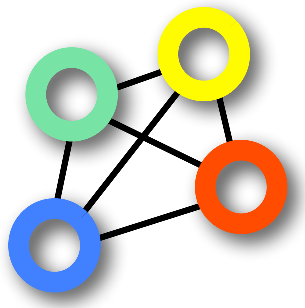
# “State-of-the-Art” in Object Class Representations

- Bag of Words Models (BoW)
  - object model = histogram of local features
  - e.g. local feature around interest points
- Global Object Models
  - object model = global feature object feature
  - e.g. HOG (Histogram of Oriented Gradients)
- Part-Based Object Models
  - object model = models of parts
    & spatial topology model
  - e.g. constellation model or
    ISM (Implicit Shape Model)
- **But: What is the Ideal Notion of Parts here?**
- **And: Should those Parts be Semantic?**

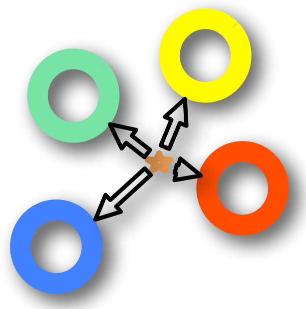
BoW: no spatial relationships



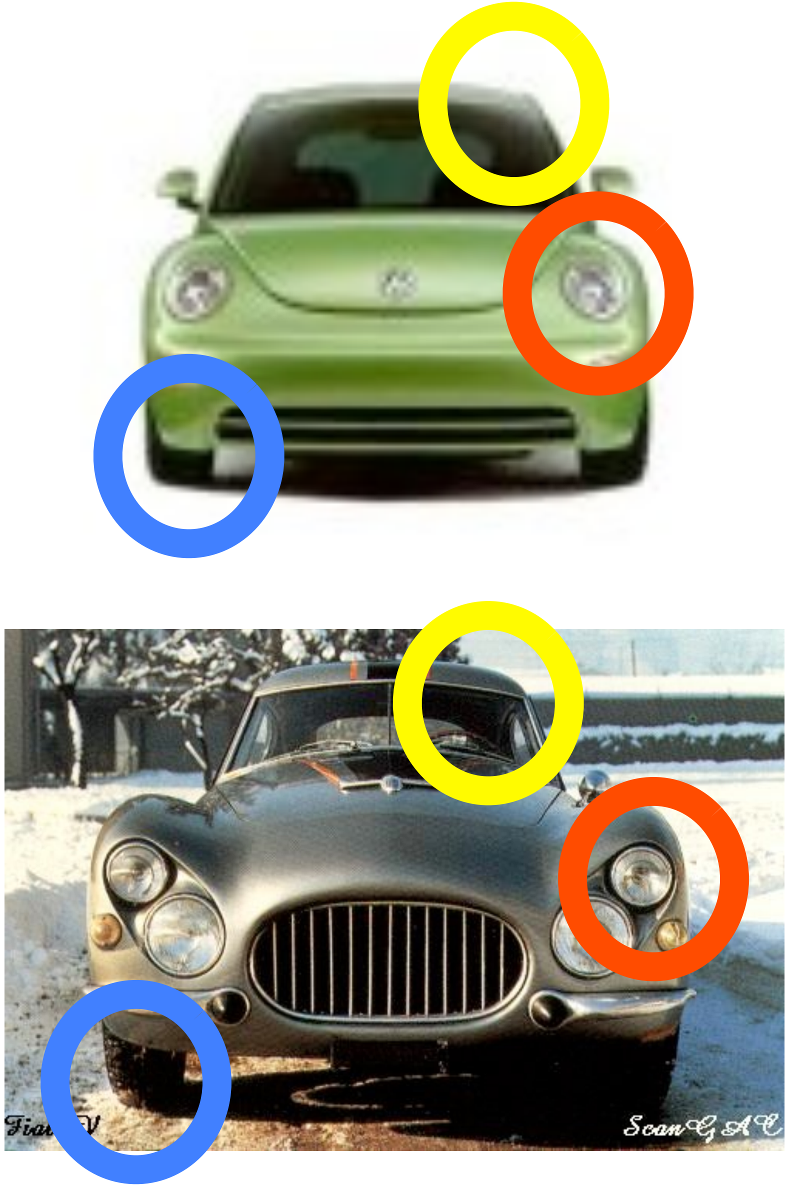
e.g. HOG: fixed spatial relationships



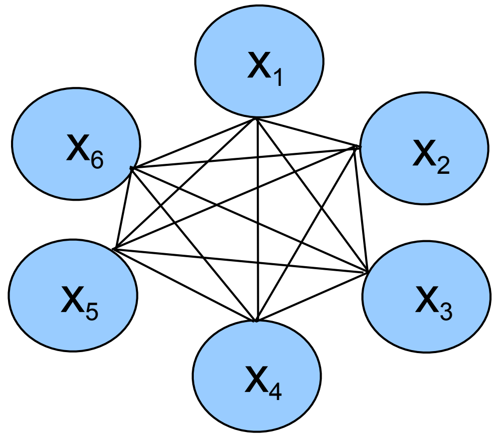
e.g. ISM: flexible spatial relationships



# Constellation of Parts



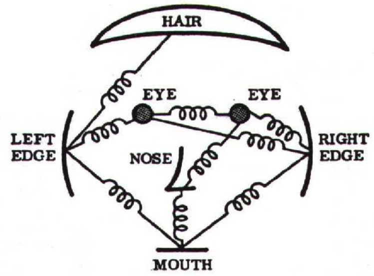

Fully connected shape model




Weber, Welling, Perona, ’00;
Fergus, Zisserman, Perona, 03

# Constellation Model

Weber, Welling, Perona, ’00;
Fergus, Zisserman, Perona, 03

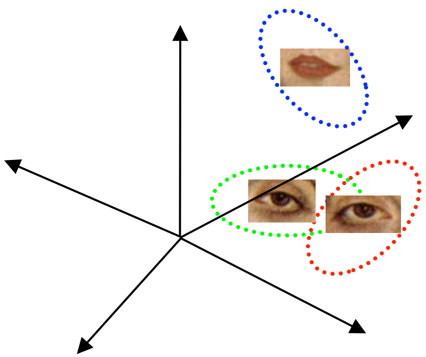



- **Joint model** for **appearance** and **structure** (=shape)
  - X: positions, A: part appearance, S: scale
  - h: Hypothesis = assignment of features (in the image) to parts (of the model)

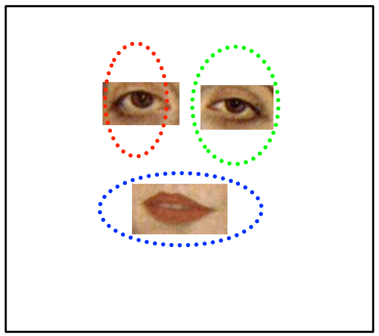
$$p(\mathbf{X},\mathbf{S},\mathbf{A}|\,\theta) = \sum_{\mathbf{h}\in H} p(\mathbf{X},\mathbf{S},\mathbf{A},\mathbf{h}|\,\theta)$$

$$= \sum_{\mathbf{h}\in H} \underbrace{p(\mathbf{A}|\mathbf{X},\mathbf{S},\mathbf{h},\theta)}_{Appearance}\,\underbrace{p(\mathbf{X}|\mathbf{S},\mathbf{h},\theta)}_{Shape}\,\underbrace{p(\mathbf{S}|\mathbf{h},\theta)}_{Rel.\ Scale}\,\underbrace{p(\mathbf{h}|\theta)}_{Other}$$

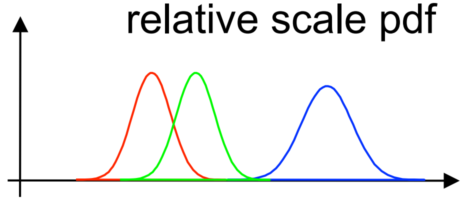
Gaussian part appearance pdf



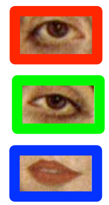
Gaussian shape pdf



Gaussian relative scale pdf

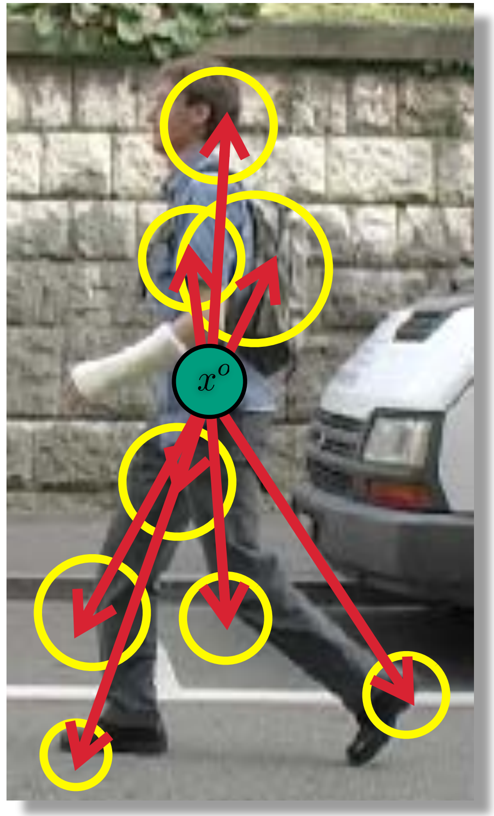



Prob. of detection



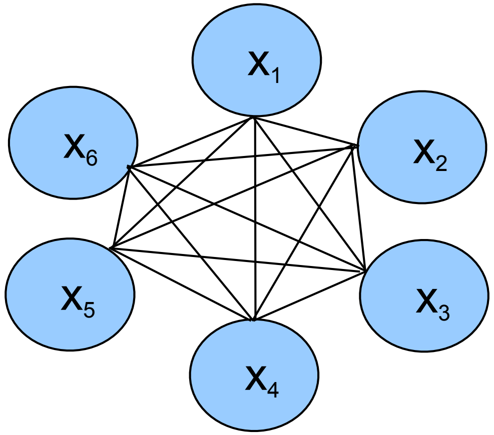
# Object Detection: ISM (Implicit Shape Model)

- Appearance of parts:
  Implicit Shape Model (ISM)
  [Leibe, Seemann & Schiele, CVPR 2005]



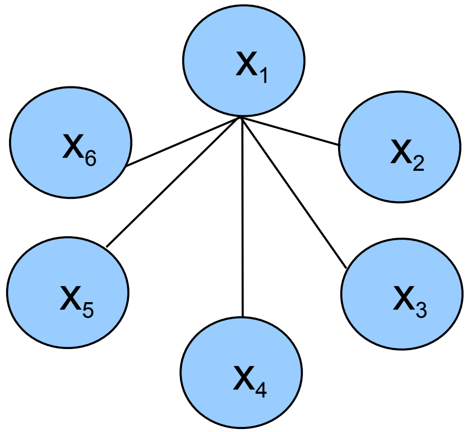
# Spatial Models for Categorization

Fully connected shape model




- e.g. Constellation Model
- Parts fully connected
- Recognition complexity: $O(N^P)$
- Method: Exhaustive search

"Star" shape model

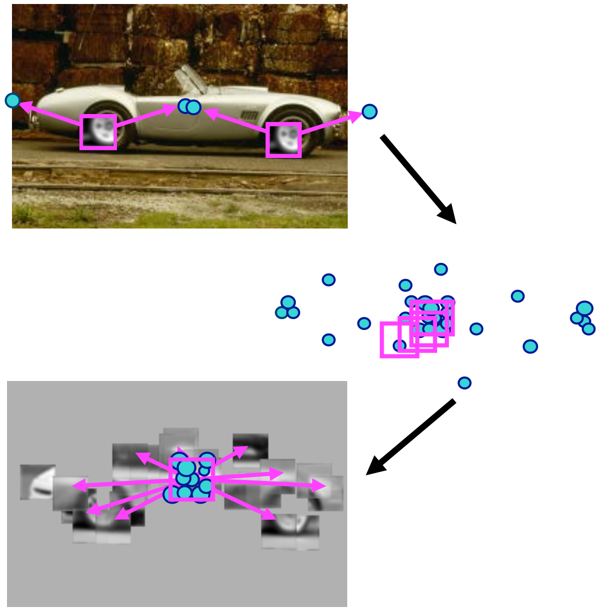



- e.g. ISM (Implicit Shape Model)
- Parts mutually independent
- Recognition complexity: O(NP)
- Method: Generalized Hough Transform

# Implicit Shape Model: What are Good Parts?

[Leibe,Schiele@bmvc03]
[Leibe,Leonardis,Schiele@ijcv08]

- Parts of the Implicit Shape Model
  - “parts” = feature clusters
  - lots of “parts” (in the order of 1’000 - 10’000 codebook entries) the more the better !
  - “parts” are mostly non-semantic



- “parts” = (mostly non semantic) feature clusters also true for
  - bag of words models
  - constellation model (much fewer “parts” - but still feature clusters)
  - ...

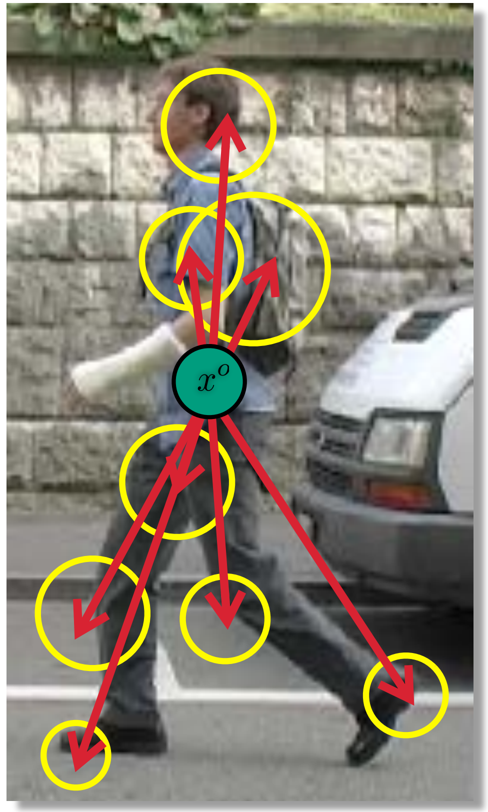
# Part-Based Models - Overview Today

- Last Week:
  - Part-Based based on Manual Labeling of Parts
    - Detection by Components, Multi-Scale Parts
  - The Constellation Model
    - automatic discovery of parts and part-structure
  - The Implicit Shape Model (ISM)
    - star-model of part configurations, parts obtained by clustering interest-points

- Today:
  - Pictorial Structures Model
  - Learning Object Model from CAD Data
  - Deformable Parts Model (DPM)
  - Discussion Semantic Parts vs. Discriminative Parts

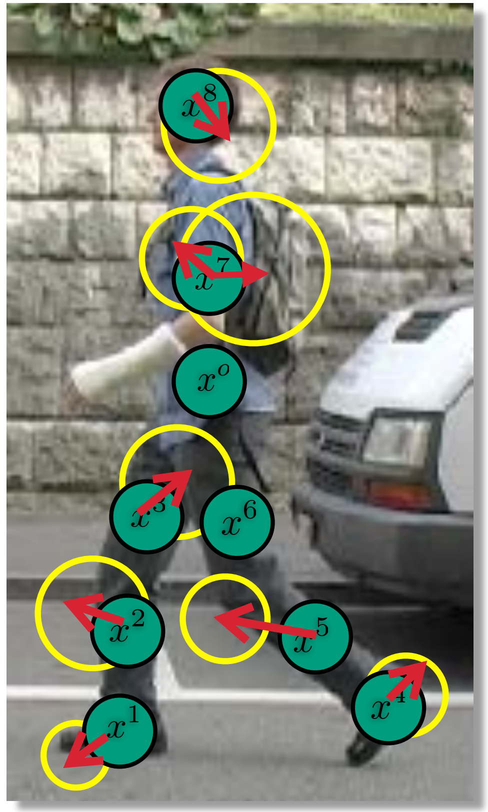
# People Detection: partISM

- Appearance of parts:
  Implicit Shape Model (ISM)
  [Leibe, Seemann & Schiele, CVPR 2005]



# People Detection: partISM

- Appearance of parts:
  Implicit Shape Model (ISM)
  [Leibe, Seemann & Schiele, CVPR 2005]
- Part decomposition and inference:
  Pictorial structures model
  [Felzenszwalb & Huttenlocher, IJCV 2005]




$$p(L|D) \propto p(D|L)p(L)$$

Body-part positions

Image evidence

# Pictorial Structures Model

$$p(L|D) = \frac{p(D|L)p(L)}{p(D)} \propto p(D|L)p(L)$$

Body-part positions    Image evidence

- Two Components
  - Prior (capturing possible part configurations): $p(L)$
  - Likelihood of Parts (capturing part appearance): $p(D|L)$

# Pictorial Structures: Model Components

[Andriluka,Roth,Schiele@cvpr09]

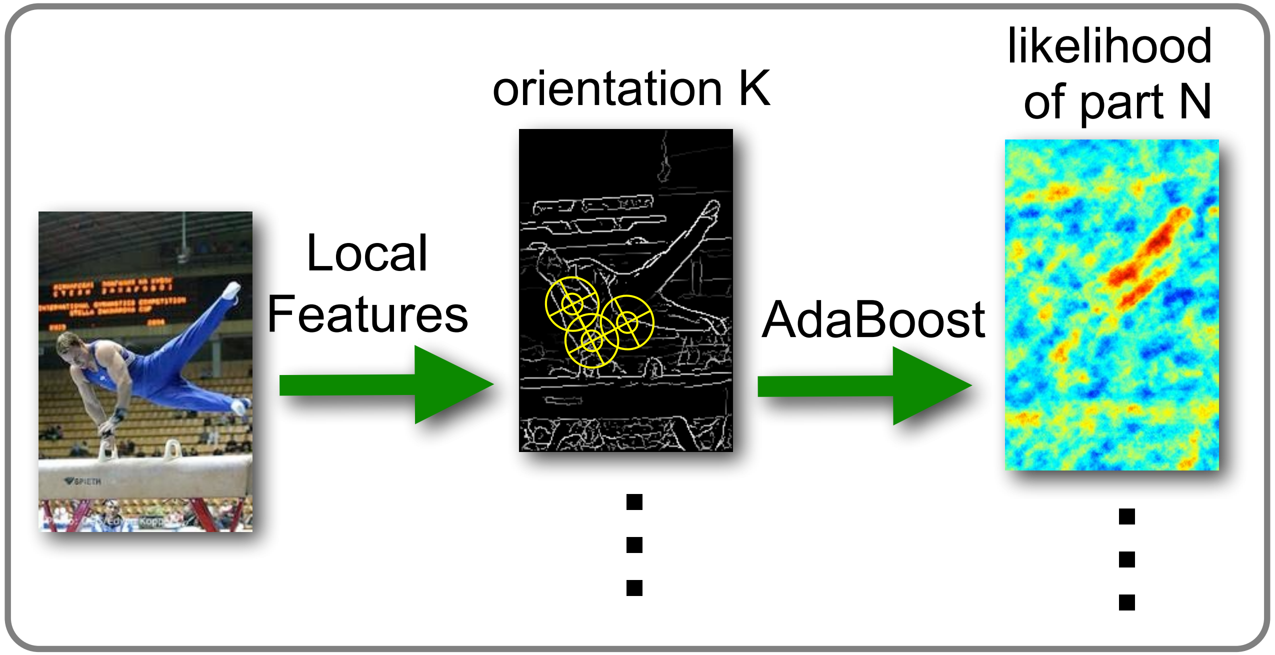
- Body is represented as flexible configuration of body parts

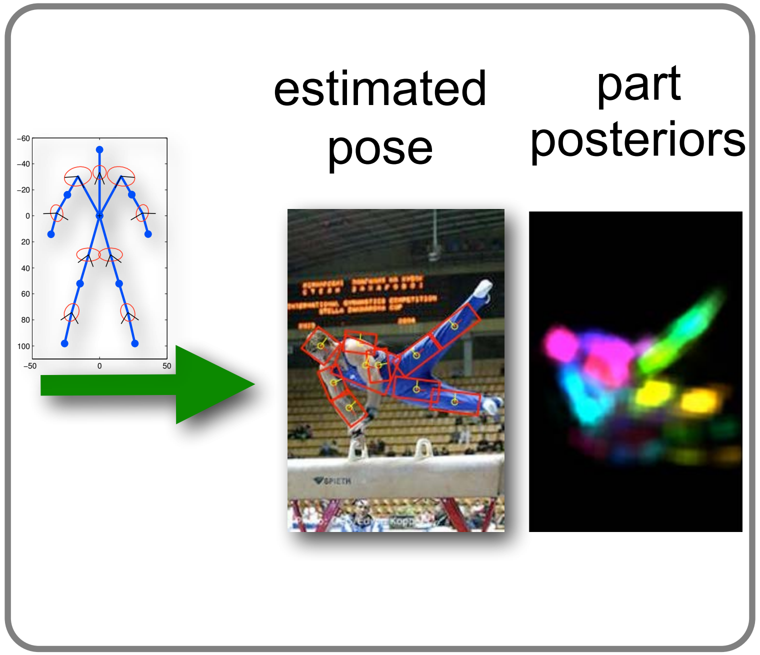
posterior over body poses

$$p(L|D) \propto p(D|L)p(L)$$

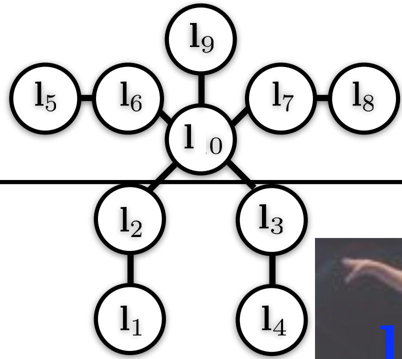
likelihood of observations

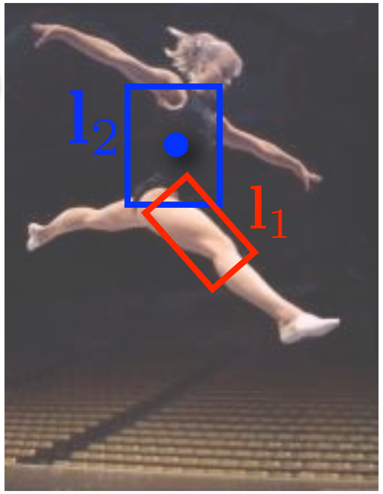
prior on body poses



# Kinematic Tree Prior
# (modeling the structure)

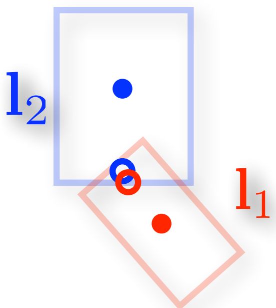


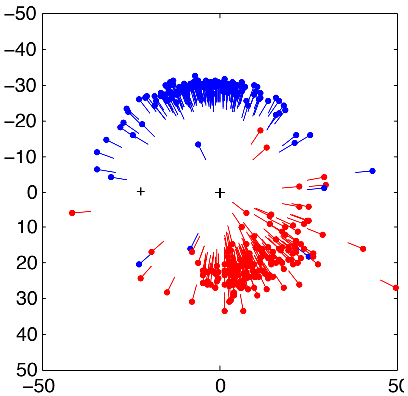

- Represent pairwise part relations [Felzenszwalb & Huttenlocher, IJCV'05]

$$p(L) = p(\mathbf{l}_0) \prod_{(i,j)\in E} p(\mathbf{l}_i|\mathbf{l}_j),$$

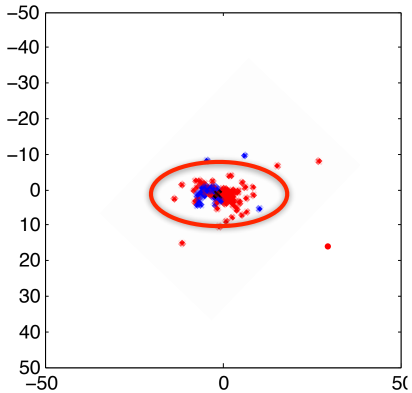
$$p(\mathbf{l}_2|\mathbf{l}_1) = \mathcal{N}(T_{12}(\mathbf{l}_2)|T_{21}(\mathbf{l}_1), \Sigma^{12})$$




part locations relative
to the joint



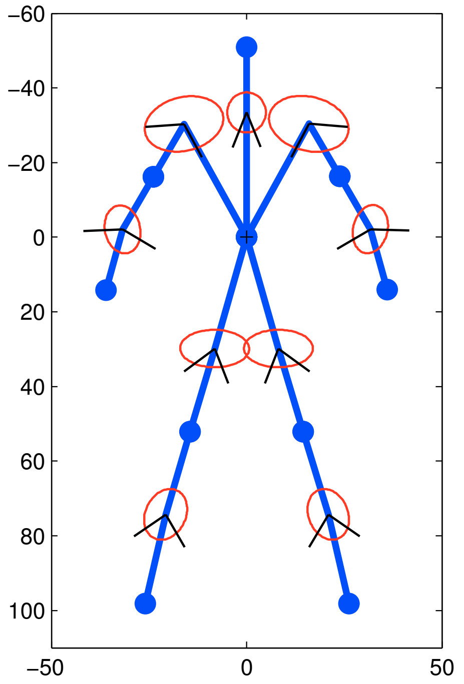

transformed
part locations



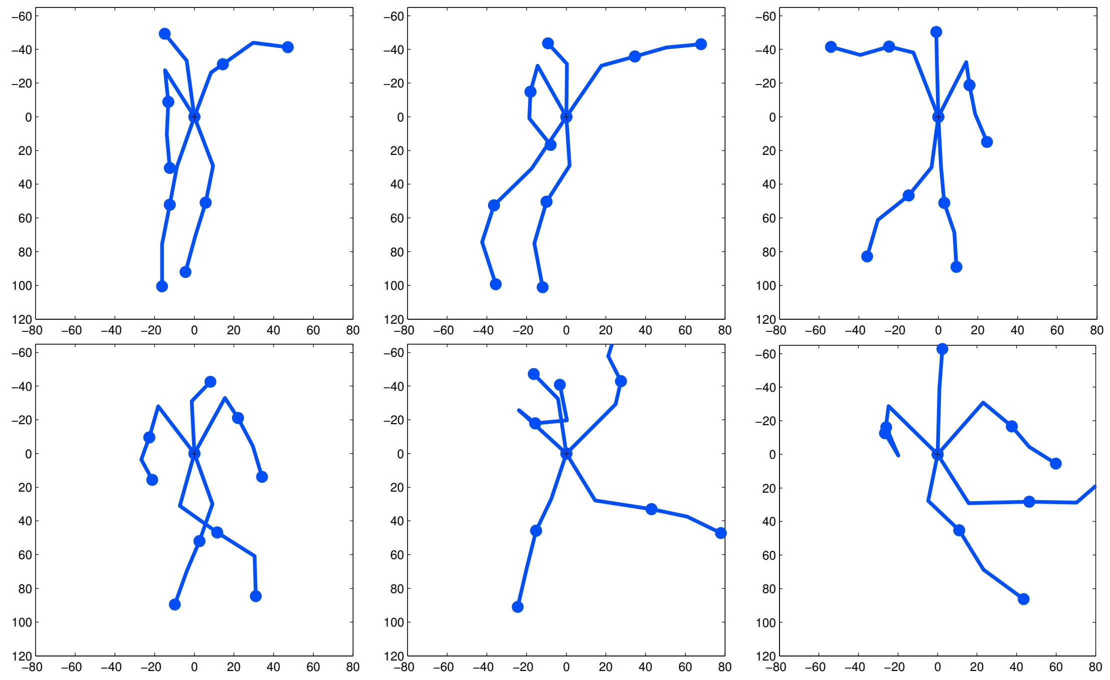
# Kinematic Tree Prior

- Prior parameters: $\{T_{ij}, \Sigma^{ij}\}$
- Parameters of the prior are estimated with maximum likelihood

mean pose




several independent samples



# Pictorial Structures: Model Components

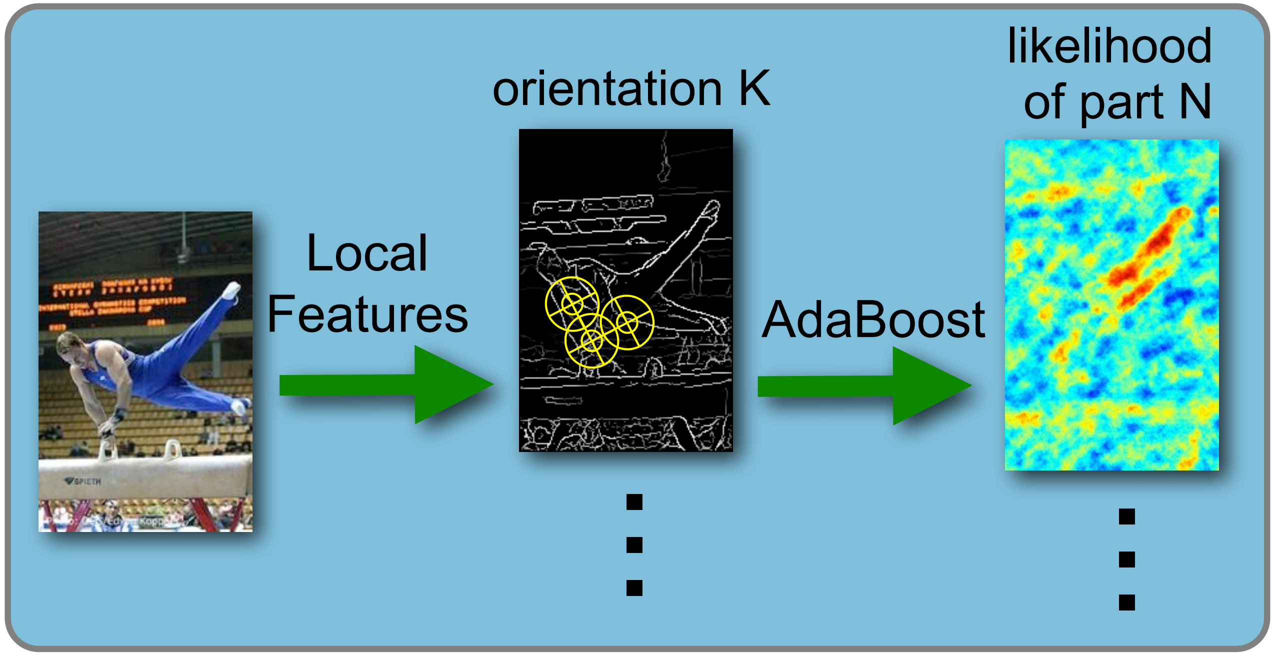
- Body is represented as flexible configuration of body parts

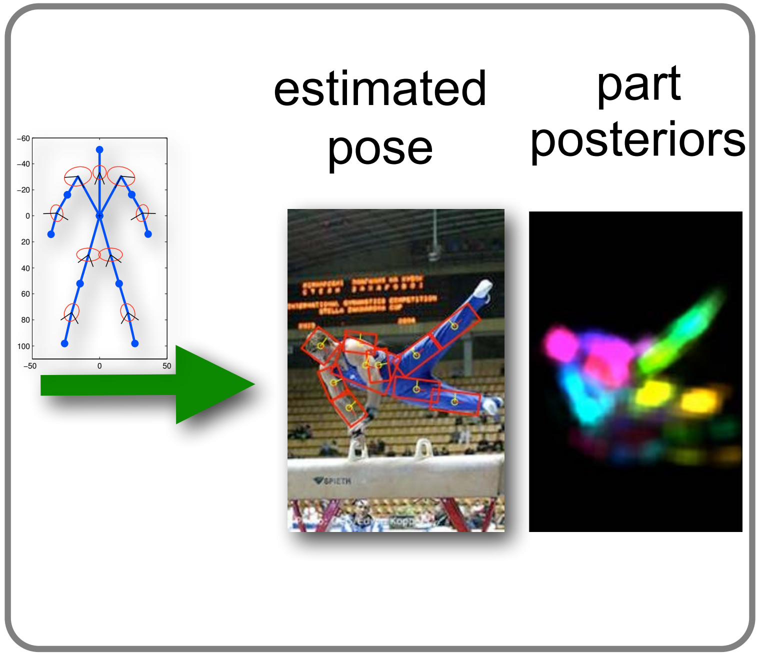
posterior over body poses

$$p(L|D) \propto p(D|L)p(L)$$

likelihood of observations

prior on body poses



# Likelihood Model

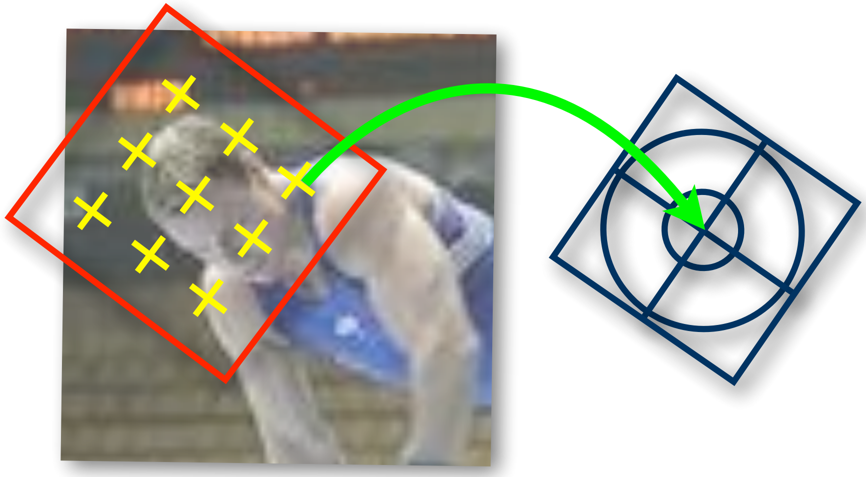
- Assumption:
  - evidence (image features) for each part independent of all other parts:

$$p(D|L) = \prod_{i=0}^{N} p(\mathbf{d}_i|\mathbf{l}_i)$$

- assumption clearly not correct, but
  - allows efficient computation
  - works rather well in practice
  - training data for different body parts should cover “all” appearances

# Likelihood Model

- Build on recent advances in object detection:
  - state-of-the-art image descriptor: Shape Context
    [Belongie et al., PAMI'02; Mikolajczyk&Schmid, PAMI'05]
  - dense representation
  - discriminative model: AdaBoost classifier for each body part



- Shape Context: 96 dimensions (4 angular, 3 radial, 8 gradient orientations)
- Feature Vector: concatenate the descriptors inside part bounding box
- head: 4032 dimensions
- torso: 8448 dimensions

# Likelihood Model

- Part likelihood derived from the boosting score:

decision stump weight

decision stump output

$$\tilde{p}(\mathbf{d}_i|\mathbf{l}_i) = \max\left(\frac{\sum_t \alpha_{i,t} h_t(\mathbf{x}(\mathbf{l}_i))}{\sum_t \alpha_{i,t}}, \varepsilon_0\right)$$

part location

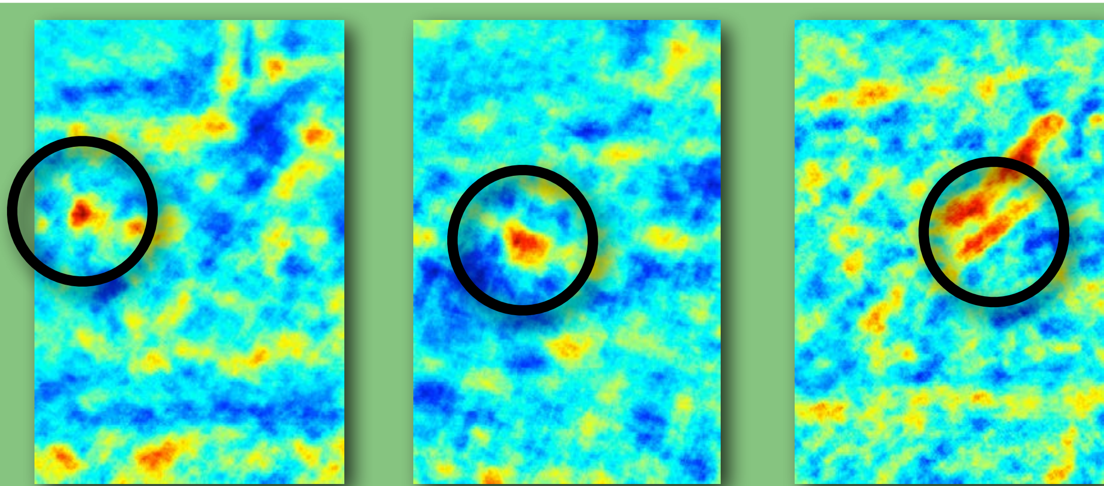
small constant to deal with part occlusions

# Likelihood Model

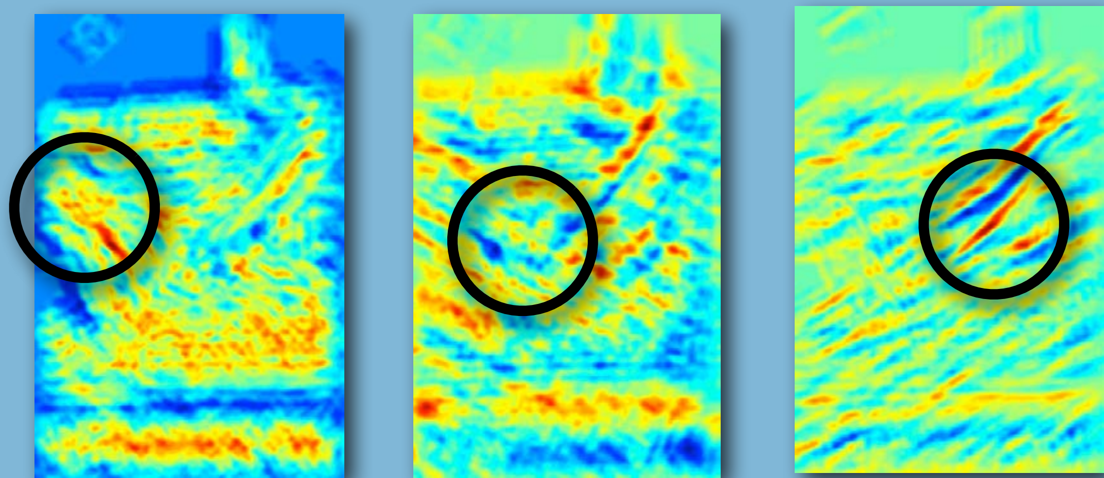
Input image

Head | Torso | Upper leg

**Our part likelihoods**

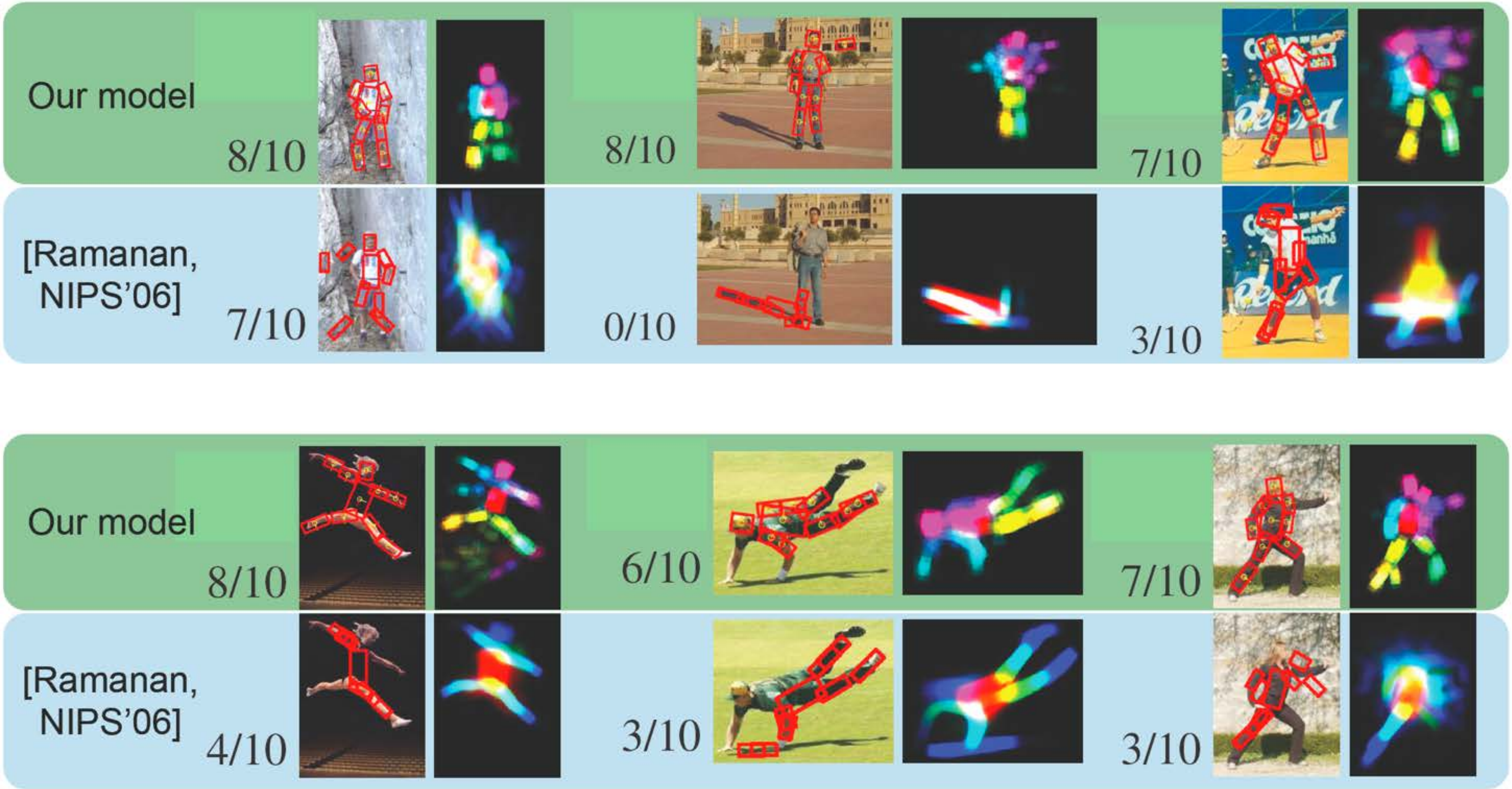


**[Ramanan, NIPS'06]**



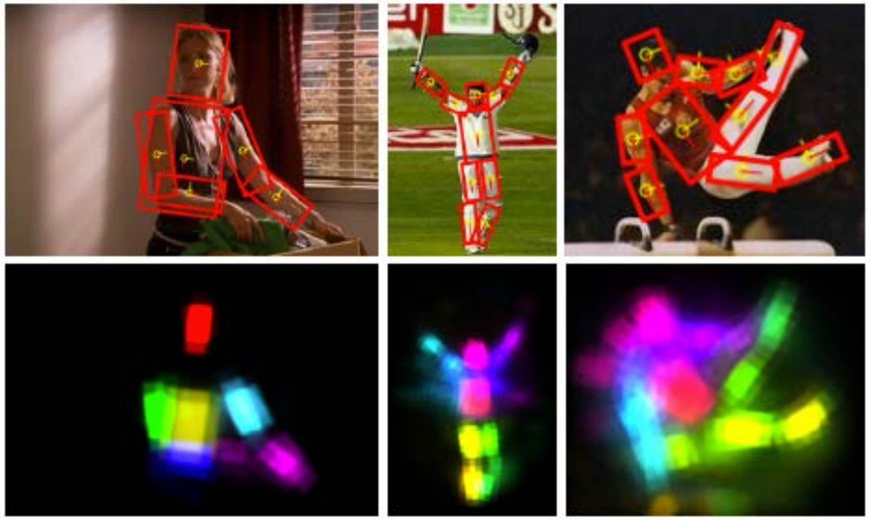
# Part-Based Model: 2D Human Pose Estimation

[Andriluka,Roth,Schiele@cvpr09]



# Pictorial Structures for Human Pose Estimation: What are Good Parts?

- Parts of the Pictorial Structures Model
  - “parts” = semantic body parts
  - pose estimation = estimation of body part configuration
  - semantic body parts allow to use motion capture data, etc. to improve kinematic tree prior



  - non-semantic parts (e.g. in the ISM-model) are more difficult to generalize across human body poses

# Part-Based Models - Overview Today

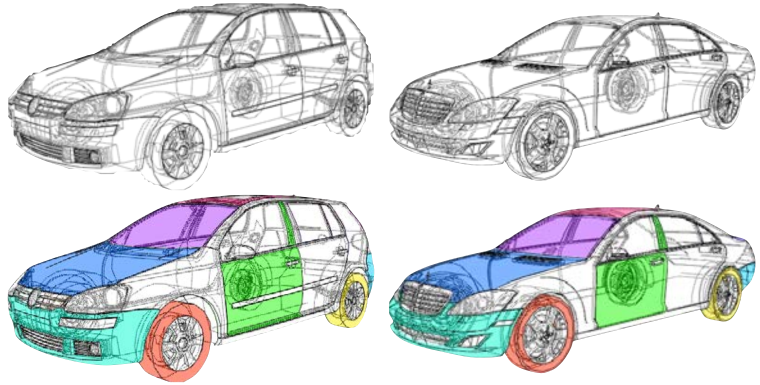
- Last Week:
  - Part-Based based on Manual Labeling of Parts
    - Detection by Components, Multi-Scale Parts
  - The Constellation Model
    - automatic discovery of parts and part-structure
  - The Implicit Shape Model (ISM)
    - star-model of part configurations, parts obtained by clustering interest-points

- Today:
  - Pictorial Structures Model
  - Learning Object Model from CAD Data
  - Deformable Parts Model (DPM)
  - Discussion Semantic Parts vs. Discriminative Parts

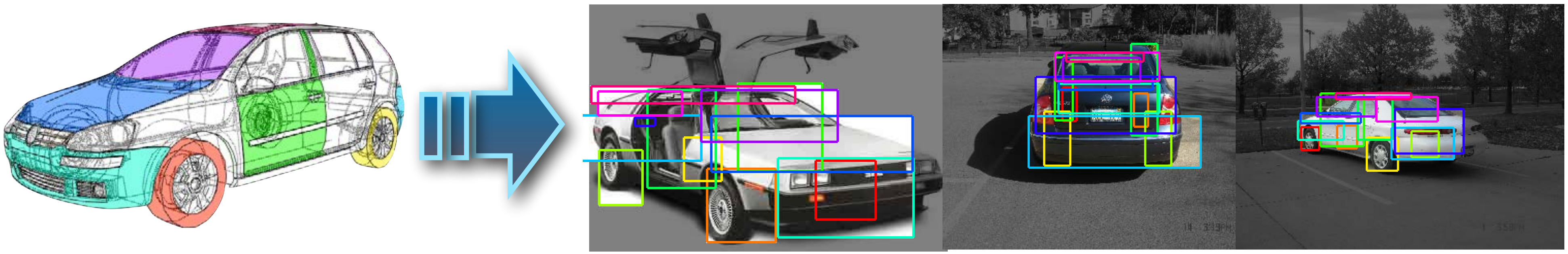
[Stark,Goesele,Schiele@bmvc10]

# Back to the Future: Learning Shape Models from 3D CAD Data

- 3D Computer Aided Design (CAD) Models
  - Computer graphics, game design
  - Polygonal meshes + texture descriptions
  - semantic part annotations (may) exist

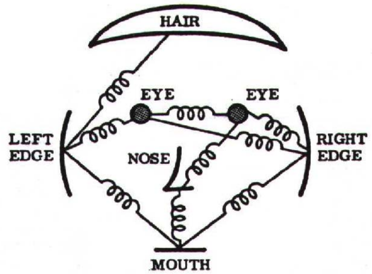


- Can we learn Object Class Models directly from 3D CAD data?
  - Issue: Transition between 3D CAD models and 2D real-world images



# Constellation Model

Weber, Welling, Perona, '00;
Fergus, Zisserman, Perona, 03

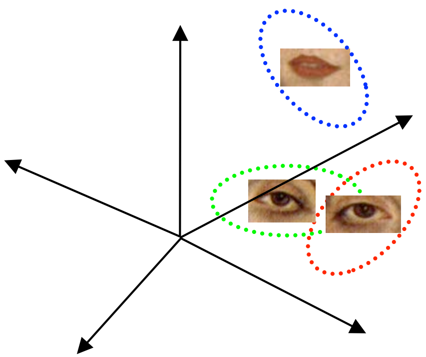



- **Joint model** for **appearance** and **structure** (=shape)
  - X: positions, A: part appearance, S: scale
  - h: Hypothesis = assignment of features (in the image) to parts (of the model)

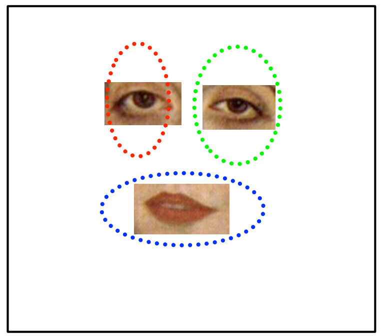
$$p(\mathbf{X},\mathbf{S},\mathbf{A}|\,\theta) = \sum_{\mathbf{h}\in H} p(\mathbf{X},\mathbf{S},\mathbf{A},\mathbf{h}|\,\theta)$$

$$= \sum_{\mathbf{h}\in H} \underbrace{p(\mathbf{A}|\mathbf{X},\mathbf{S},\mathbf{h},\theta)}_{Appearance} \underbrace{p(\mathbf{X}|\mathbf{S},\mathbf{h},\theta)}_{Shape} \underbrace{p(\mathbf{S}|\mathbf{h},\theta)}_{Rel.\ Scale} \underbrace{p(\mathbf{h}|\theta)}_{Other}$$

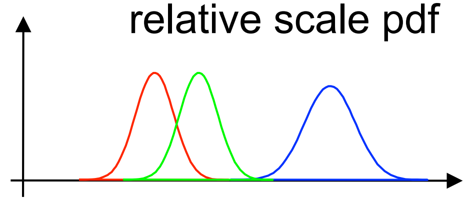
Gaussian part appearance pdf



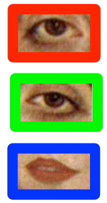
Gaussian shape pdf



Gaussian relative scale pdf

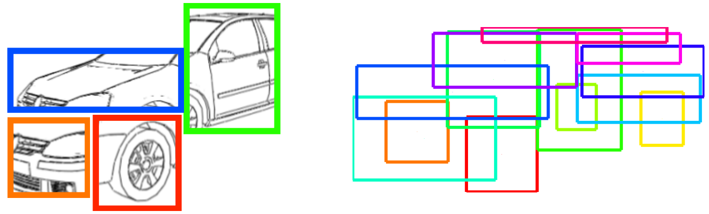



Prob. of detection



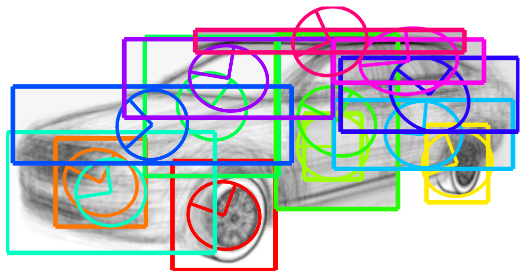
# Three Tools to Meet the Challenge

1. Shape-based appearance abstraction
   - Non-photorealistic rendering
   - Local shape + global geometry



2. Discriminative part detectors
   - Robust local shape features
   - AdaBoost classifiers

3. Powerful spatial model
   - Full covariance
   - Efficient DDMCMC inference

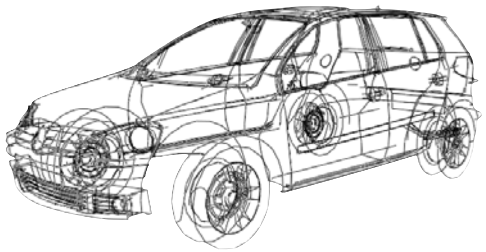


# Shape-based Appearance Abstraction ➻ Non-Photorealistic Rendering

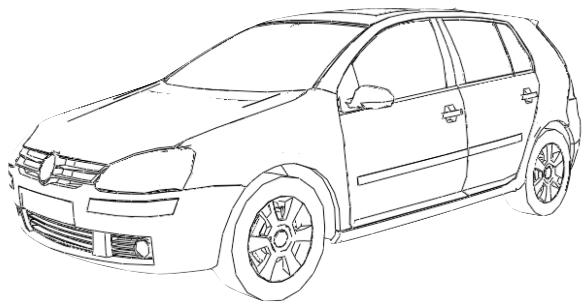
- Learn shape models from rendered images
  - we do NOT render photo-realistically / texture
  - But focus on 3D CAD model edges (mimic real-world image edges)

**Part boundaries**

**Mesh creases**



**Silhouette**

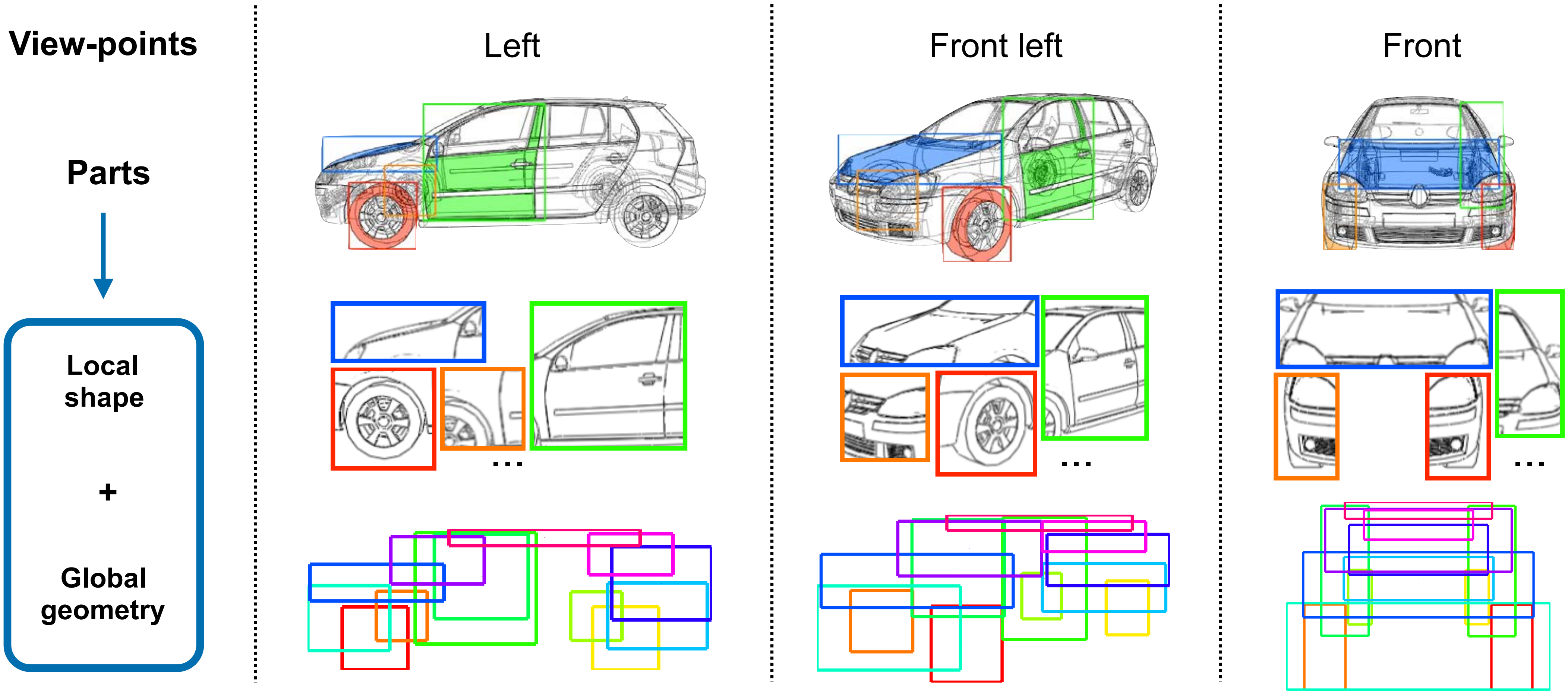


**Final edges**
(hidden edges removed)

[Stark,Goesele,Schiele@bmvc10]

# Shape - Local Shape + Global Geometry

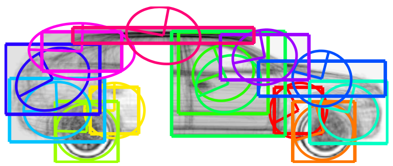
- Part-based object class representation
  - Semantic parts from 3D CAD models: *left front wheel*, *left front door*, etc.



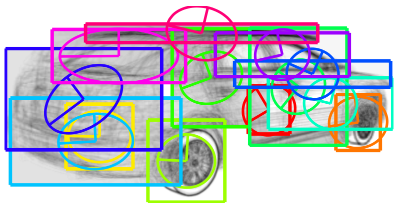
[Stark,Goesele,Schiele@bmvc10]

# Qualitative Results

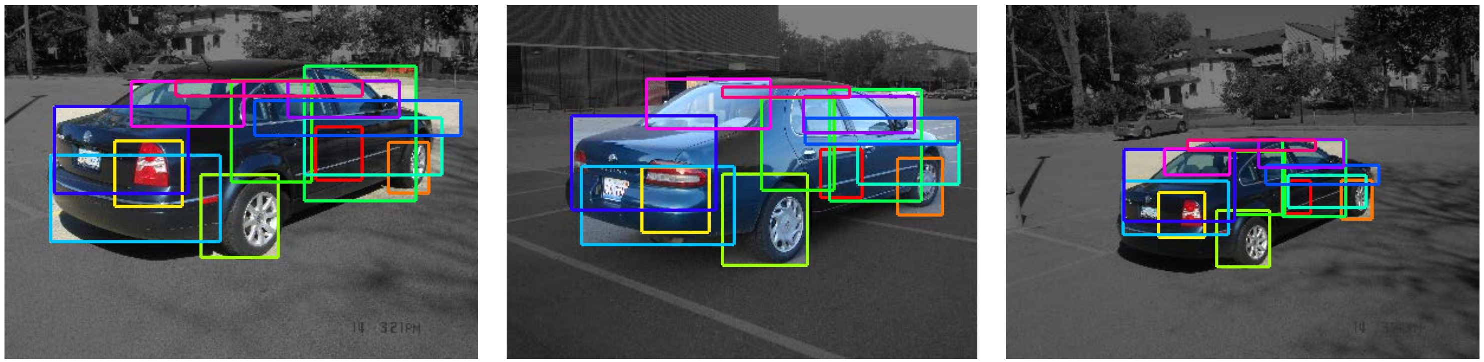
- Three strongest true positive detections per viewpoint model



Right



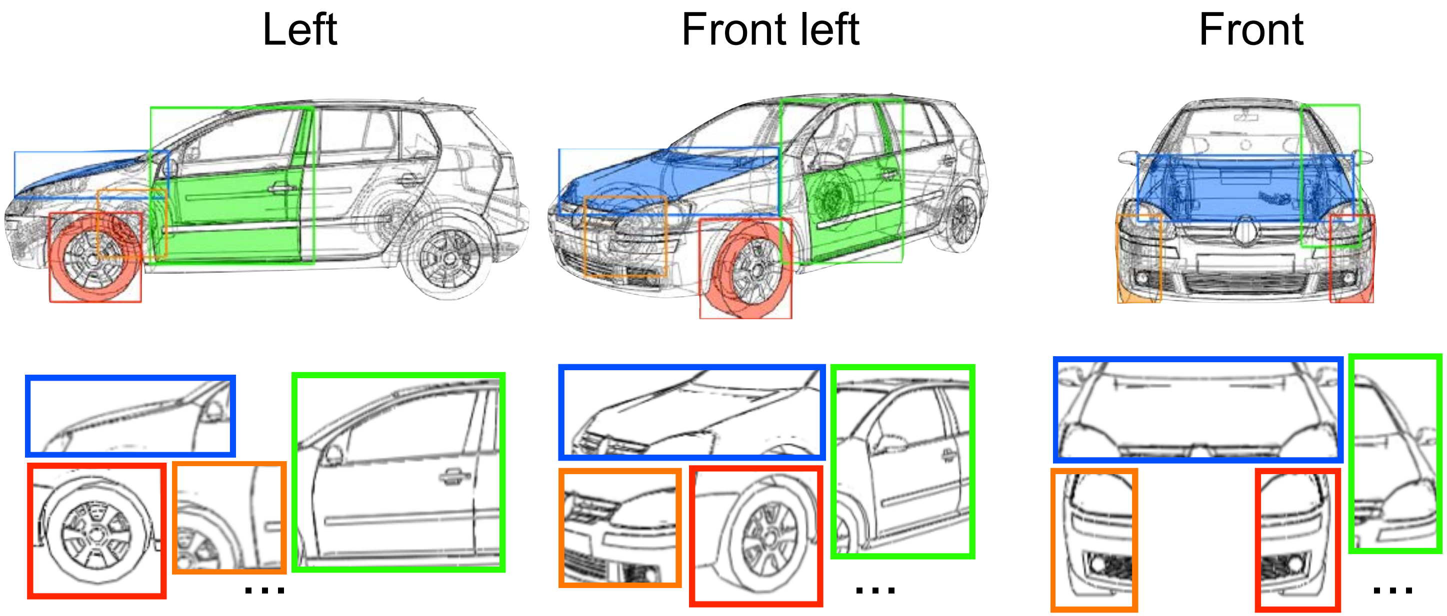
Back-right



- Observations
  - Accurate part localization
  - Predicted viewpoints do match

# Shape Model learned from 3D CAD-Data
# What are Good Parts?

- Parts of the Shape Model:
  - "parts" = just means to enable correspondence across 3D-models
  - semantics of parts:
    - in our case: yes - because of the employed 3D models
    - but: semantics neither necessary nor important



# Part-Based Models - Overview Today

- Last Week:
  - Part-Based based on Manual Labeling of Parts
    - Detection by Components, Multi-Scale Parts
  - The Constellation Model
    - automatic discovery of parts and part-structure
  - The Implicit Shape Model (ISM)
    - star-model of part configurations, parts obtained by clustering interest-points

- Today:
  - Pictorial Structures Model
  - Learning Object Model from CAD Data
  - Deformable Parts Model (DPM)
  - Discussion Semantic Parts vs. Discriminative Parts

# slide curtesy: Pedro Felzenszwalb

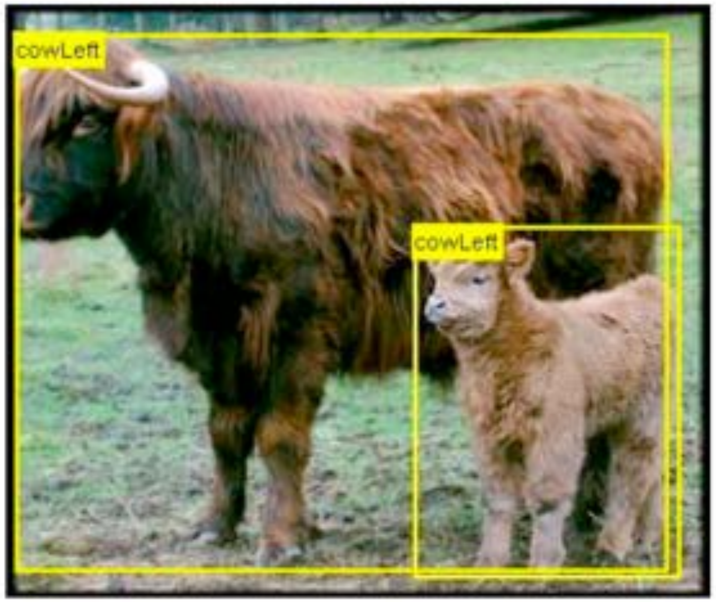
Object Detection with
Discriminatively Trained Part
Based Models

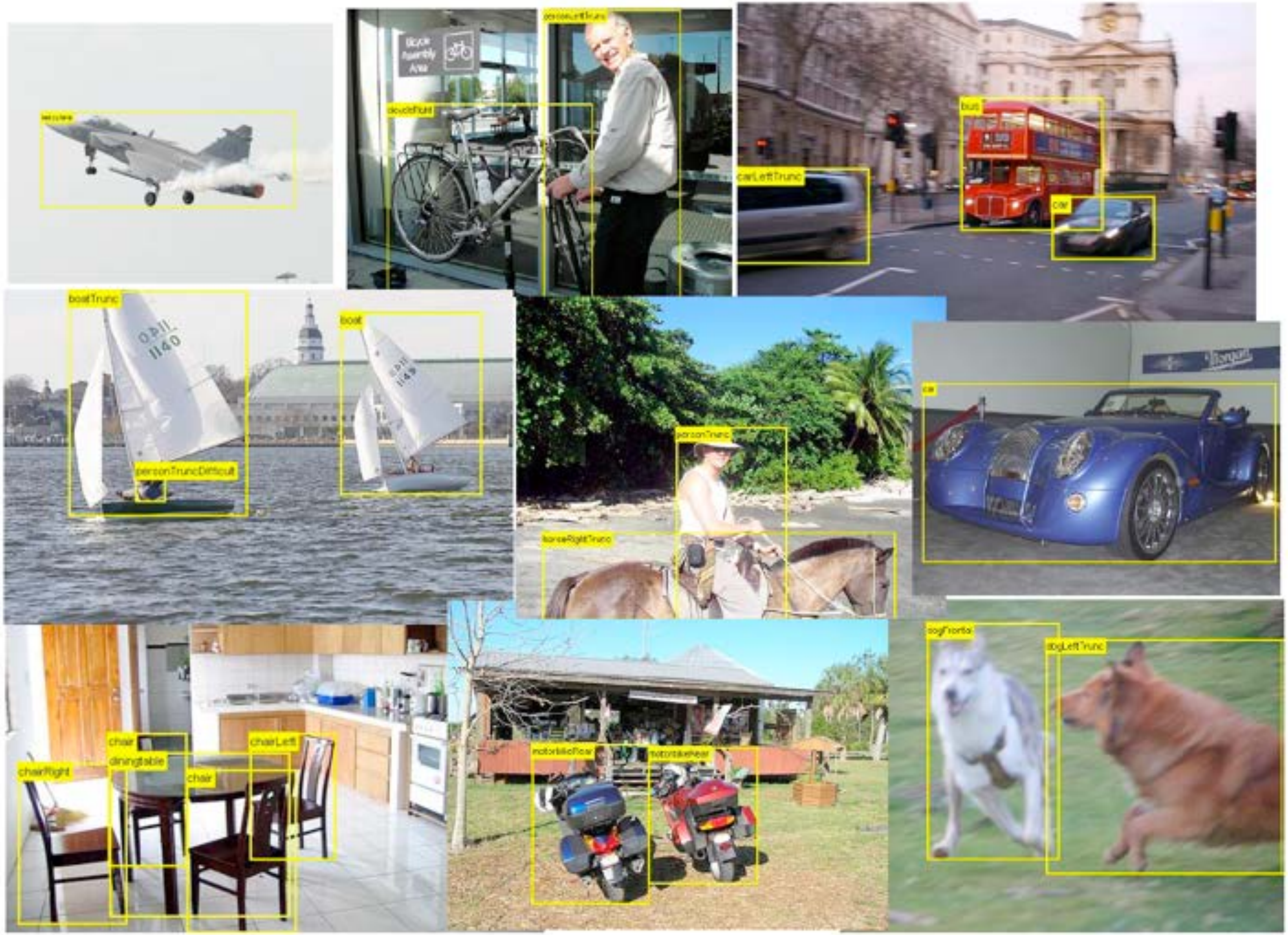
Pedro F. Felzenszwalb
Department of Computer Science
University of Chicago

Joint with David Mcallester, Deva Ramanan, Ross Girshick

# PASCAL Challenge

- ~10,000 images, with ~25,000 target objects
  - Objects from 20 categories (person, car, bicycle, cow, table...)
  - Objects are annotated with labeled bounding boxes

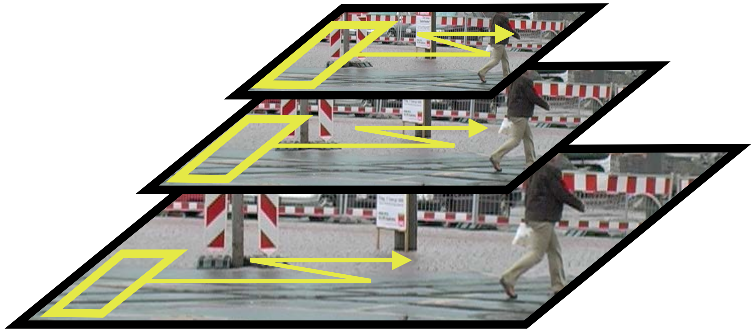




bus
carLeftTrunc
car
boatTrunc
boat
personTruncDifficult
personTrunc
horseRightTrunc
car
dogLeftTrunc
chair
diningtable
chair
chairLeft
chairRight

# Starting Point: Sliding Window Method

- Sliding Window Based People Detection:

Scan Image → Extract Feature Vector → Classify Feature Vector → Non-Maxima Suppression



'slide' detection window over all positions & scales

**Two Important Questions:**
**1) which feature vector**
**2) which classifier**

For example:

- HOG Pedestrian Detector
  - HOG descriptor
  - linear SVM

# Histogram of Oriented Gradients (HOG): Static Feature Extraction

Input Image

Detection Window

Normalise gamma

↓

Compute gradients

↓

Weighted vote in spatial & orientation cells

↓

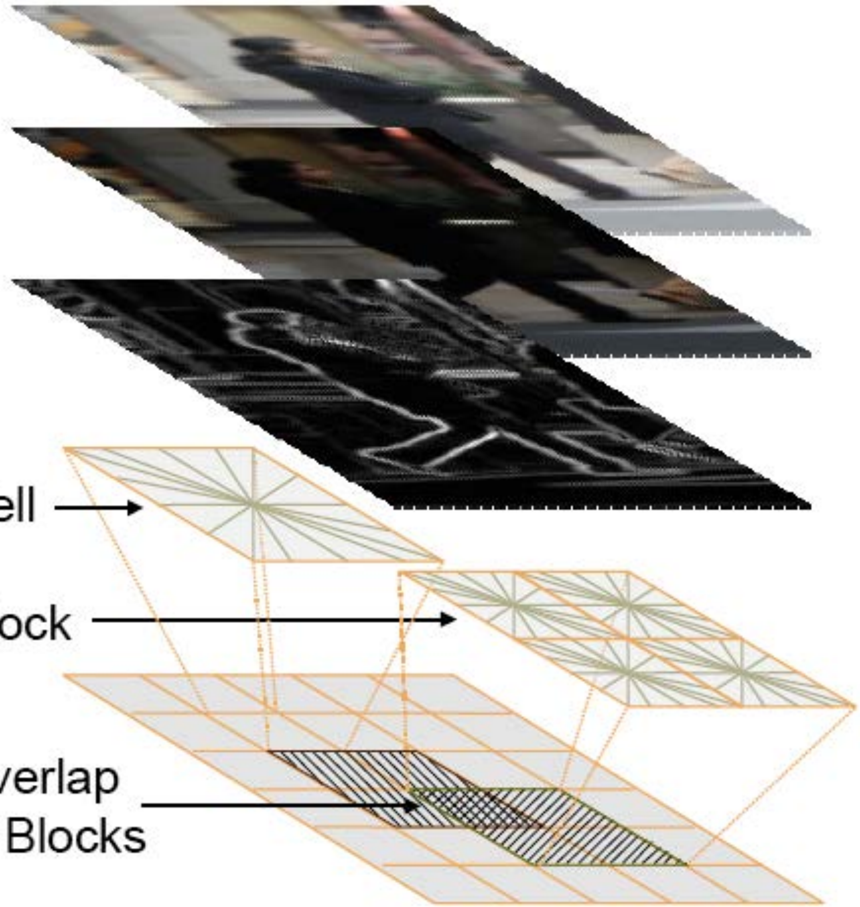
Contrast normalise over overlapping spatial cells

↓

Collect HOGs over detection window

↓

Linear SVM




Feature vector $f = [\ \ldots,\ \ldots,\ \ \ \ldots]$

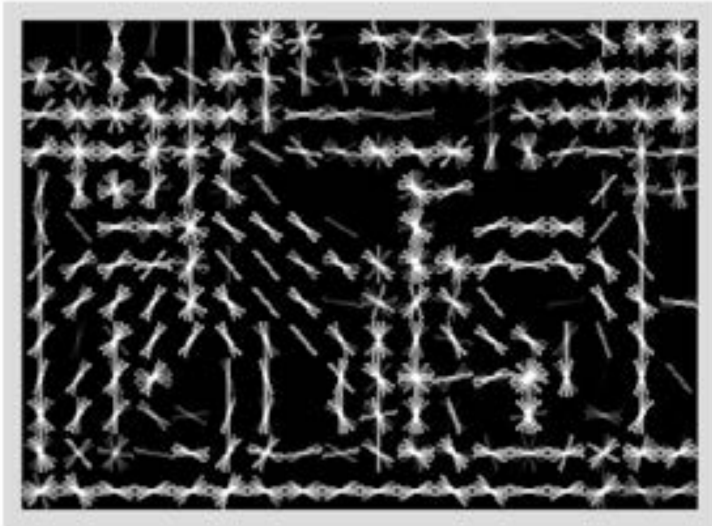
# Starting point: sliding window classifiers

Feature vector

$x = [\ldots, \ldots, \ldots, \ldots]$

- Detect objects by testing each subwindow
  - Reduces object detection to binary classification
  - Dalal & Triggs: HOG features + linear SVM classifier
  - Previous state of the art for detecting people

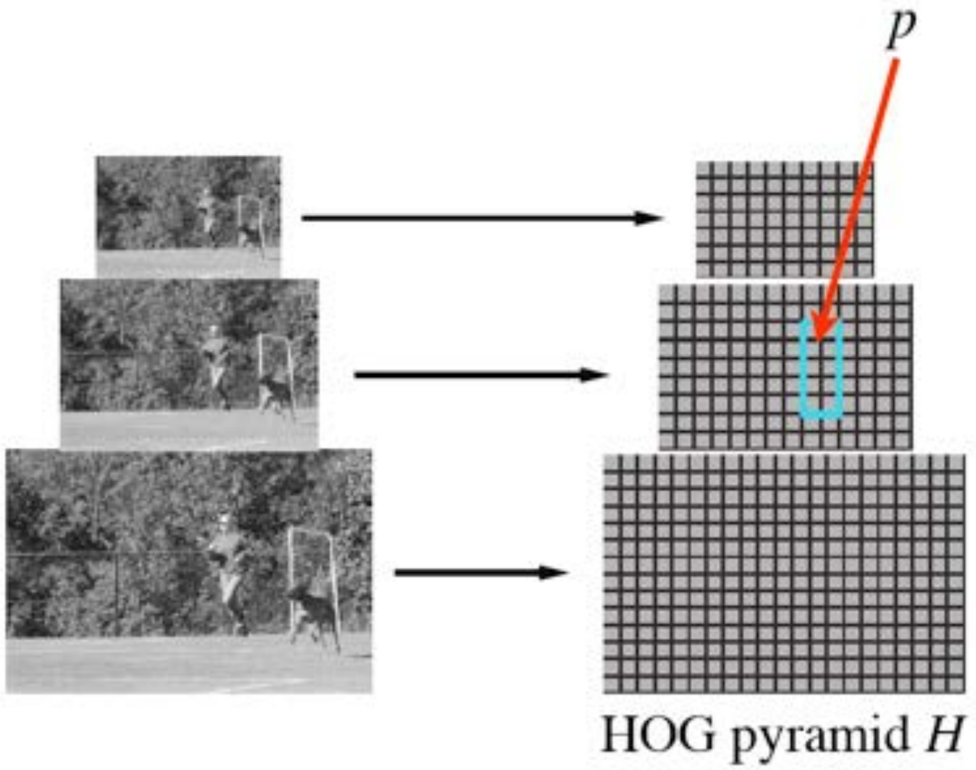
# Histogram of Gradient (HOG) features

- Image is partitioned into 8x8 pixel blocks
- In each block we compute a histogram of gradient orientations
  - Invariant to changes in lighting, small deformations, etc.
- Compute features at different resolutions (pyramid)

# HOG Filters

- Array of weights for features in subwindow of HOG pyramid
- Score is dot product of filter and feature vector

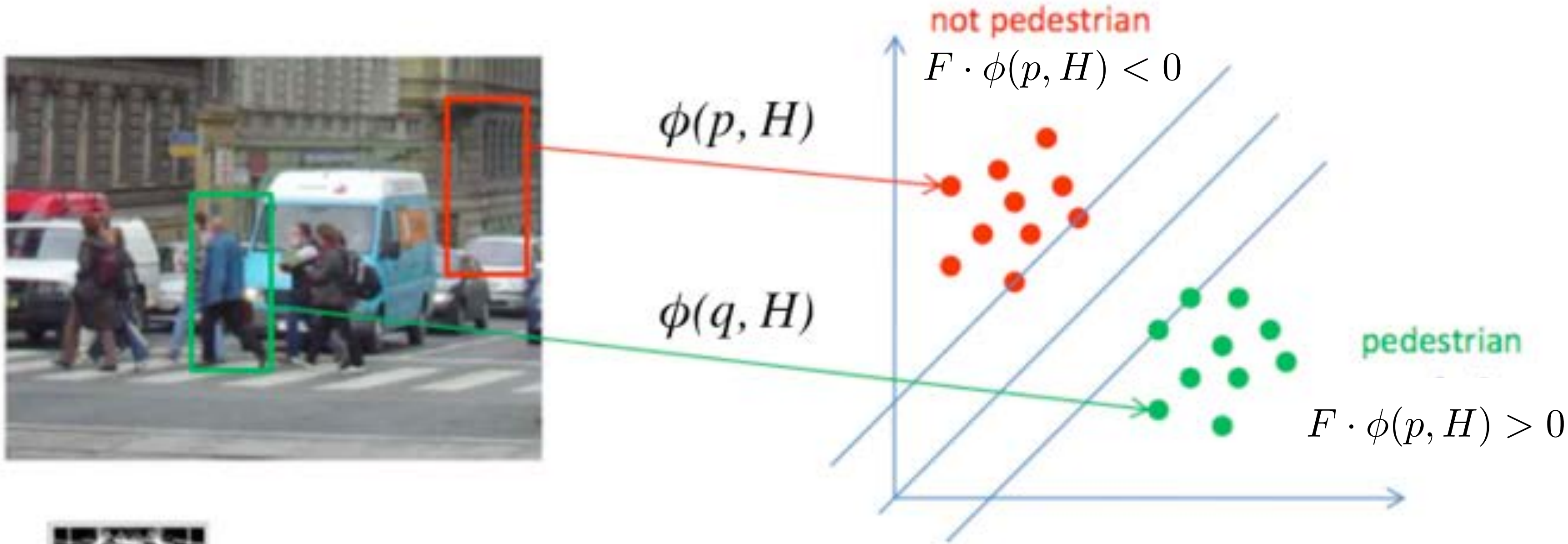



Filter $F$

Score of $F$ at position $p$ is
$F \cdot \phi(p, H)$

$\phi(p, H)$ = concatenation of HOG features from subwindow specified by $p$

# Dalal & Triggs: HOG + linear SVMs




Typical form of a model

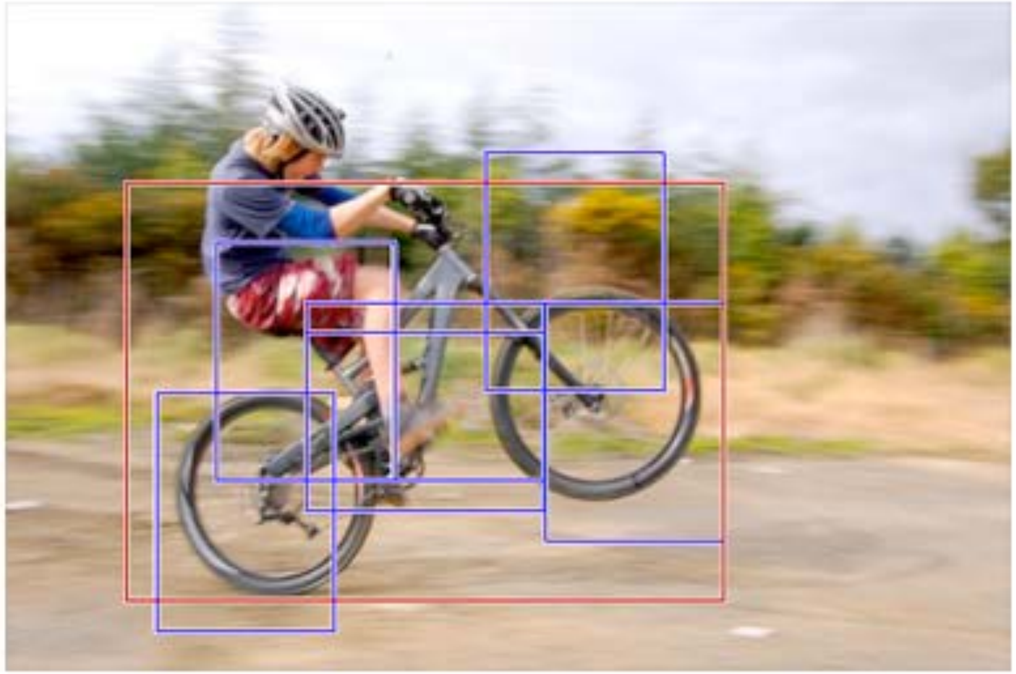
There is much more background than objects

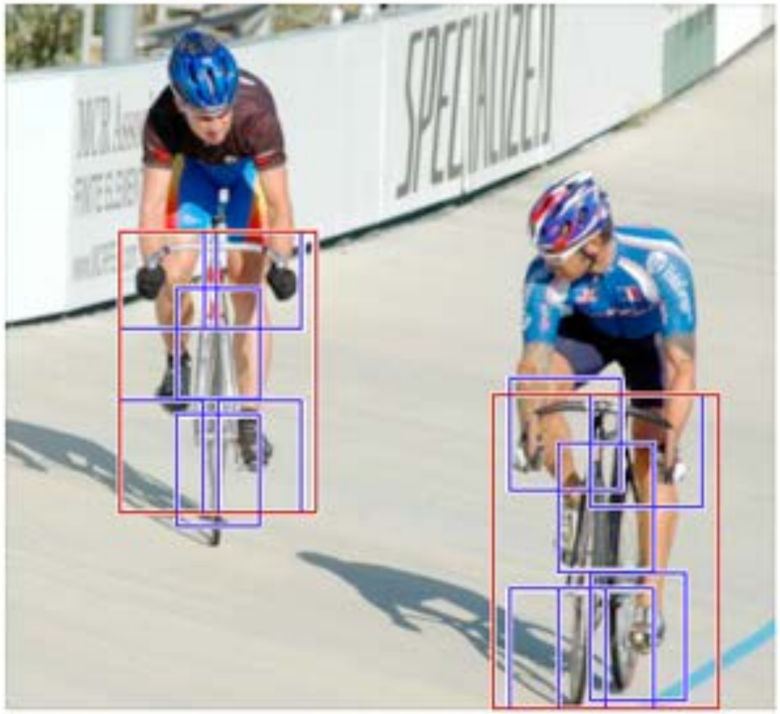
Start with random negatives and repeat:

1) Train a model

2) Harvest false positives to define “hard negatives”

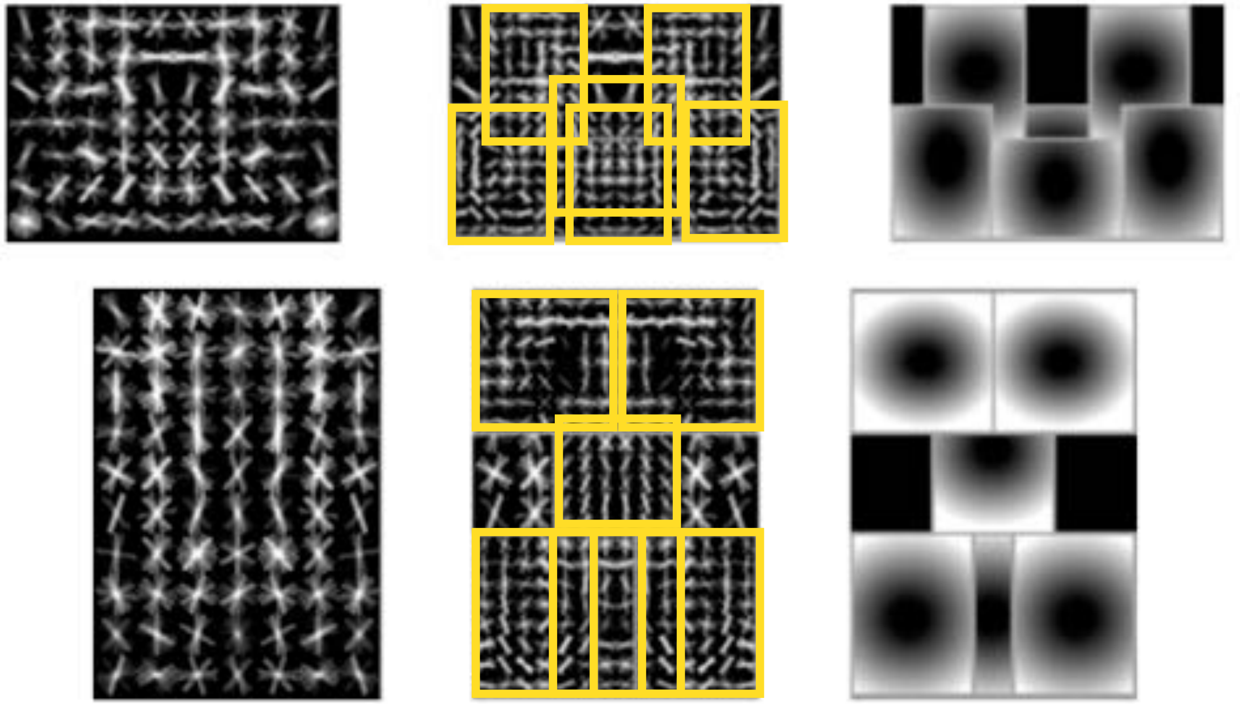
# Overview of our models






- Mixture of deformable part models
- Each component has global template + deformable parts
- Fully trained from bounding boxes alone

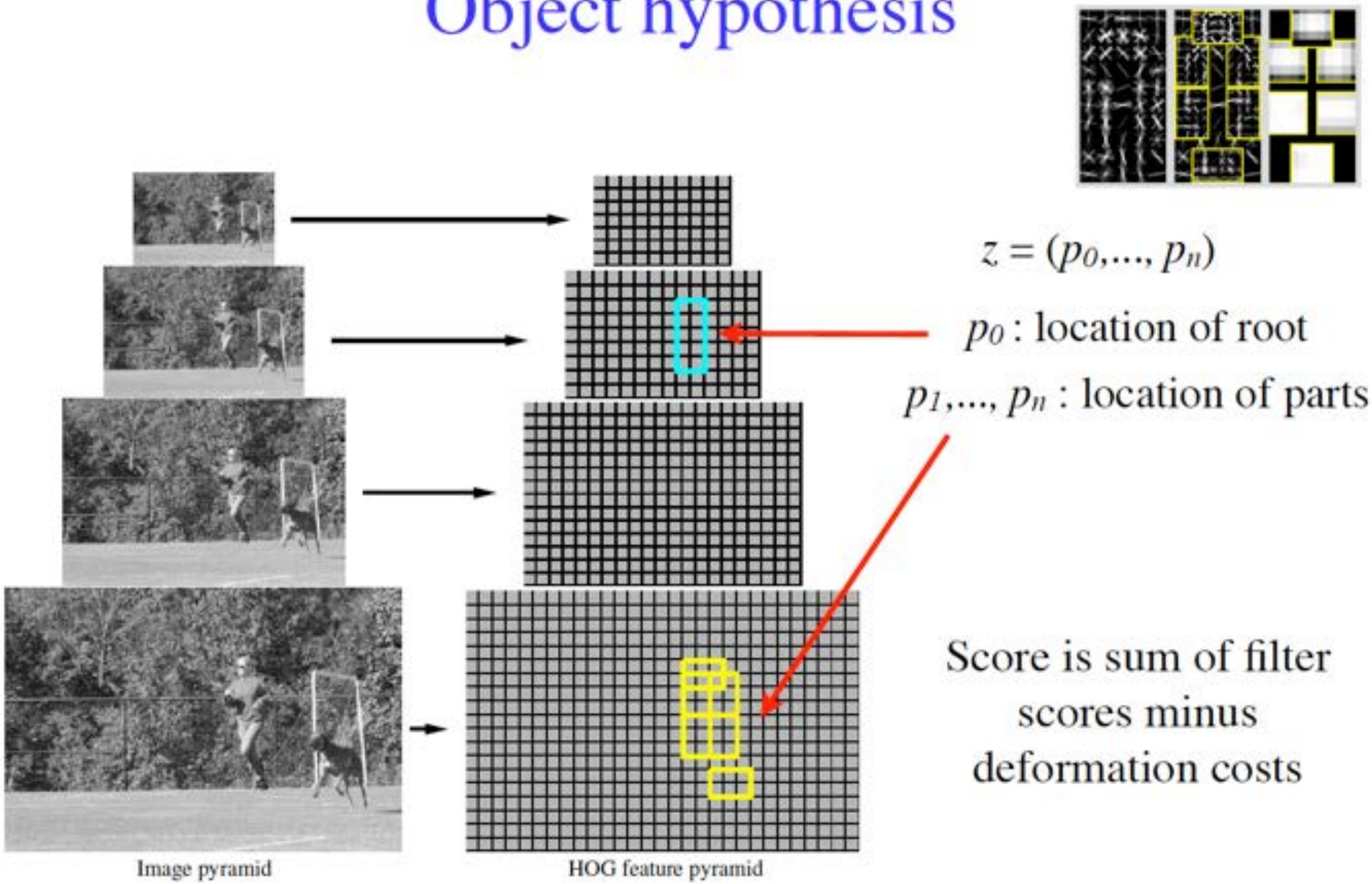
# 2 component bicycle model



root filters coarse resolution | part filters finer resolution | deformation models

Each component has a root filter $F_0$
and $n$ part models $(F_i, v_i, d_i)$

# Object hypothesis




$z = (p_0, \ldots, p_n)$

$p_0$ : location of root

$p_1, \ldots, p_n$ : location of parts

Score is sum of filter scores minus deformation costs

Multiscale model captures features at two-resolutions

# Score of a hypothesis

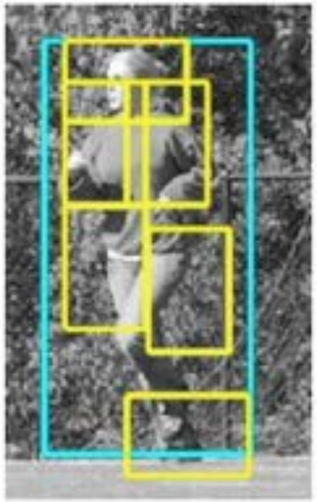
"data term" "spatial prior"

$$\text{score}(p_0,\ldots,p_n) = \boxed{\sum_{i=0}^{n} F_i \cdot \phi(H,p_i)} - \boxed{\sum_{i=1}^{n} d_i \cdot (dx_i^2, dy_i^2)}$$

filters

displacements

deformation parameters



$$\text{score}(z) = \beta \cdot \Psi(H,z)$$

concatenation filters and deformation parameters

concatenation of HOG features and part displacement features

# Matching

- Define an overall score for each root location
  - Based on best placement of parts

$$\text{score}(p_0) = \max_{p_1,\ldots,p_n} \text{score}(p_0, \ldots, p_n).$$

- High scoring root locations define detections
  - "sliding window approach"
- Efficient computation: dynamic programming + generalized distance transforms (max-convolution)

# Efficient Computation

- Overall score:

$$score(p_0, \ldots, p_n) = \sum_{i=0}^{n} F_i \cdot \phi(H, p_i) - \sum_{i=1}^{n} d_i \cdot (dx_i^2, dy_i^2)$$

- Maximization can be done separately:

$$score(p_0) = \max_{p_1, \ldots, p_n} score(p_0, \ldots, p_n)$$

$$= F_0 \cdot \phi(H, p_0) +$$

$$\max_{p_1} \left( F_1 \cdot \phi(H, p_1) - d_1 \cdot (dx_1^2, dy_1^2) \right) +$$

$$\cdots +$$

$$\max_{p_n} \left( F_n \cdot \phi(H, p_n) - d_n \cdot (dx_n^2, dy_n^2) \right)$$

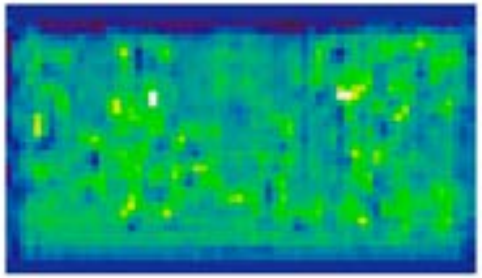
head filter

input image

Response of filter in l-th pyramid level

$$R_l(x, y) = F \cdot \phi(H, (x, y, l))$$

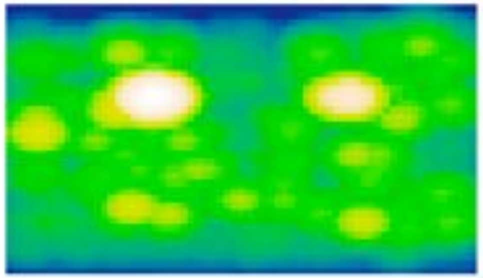
cross-correlation

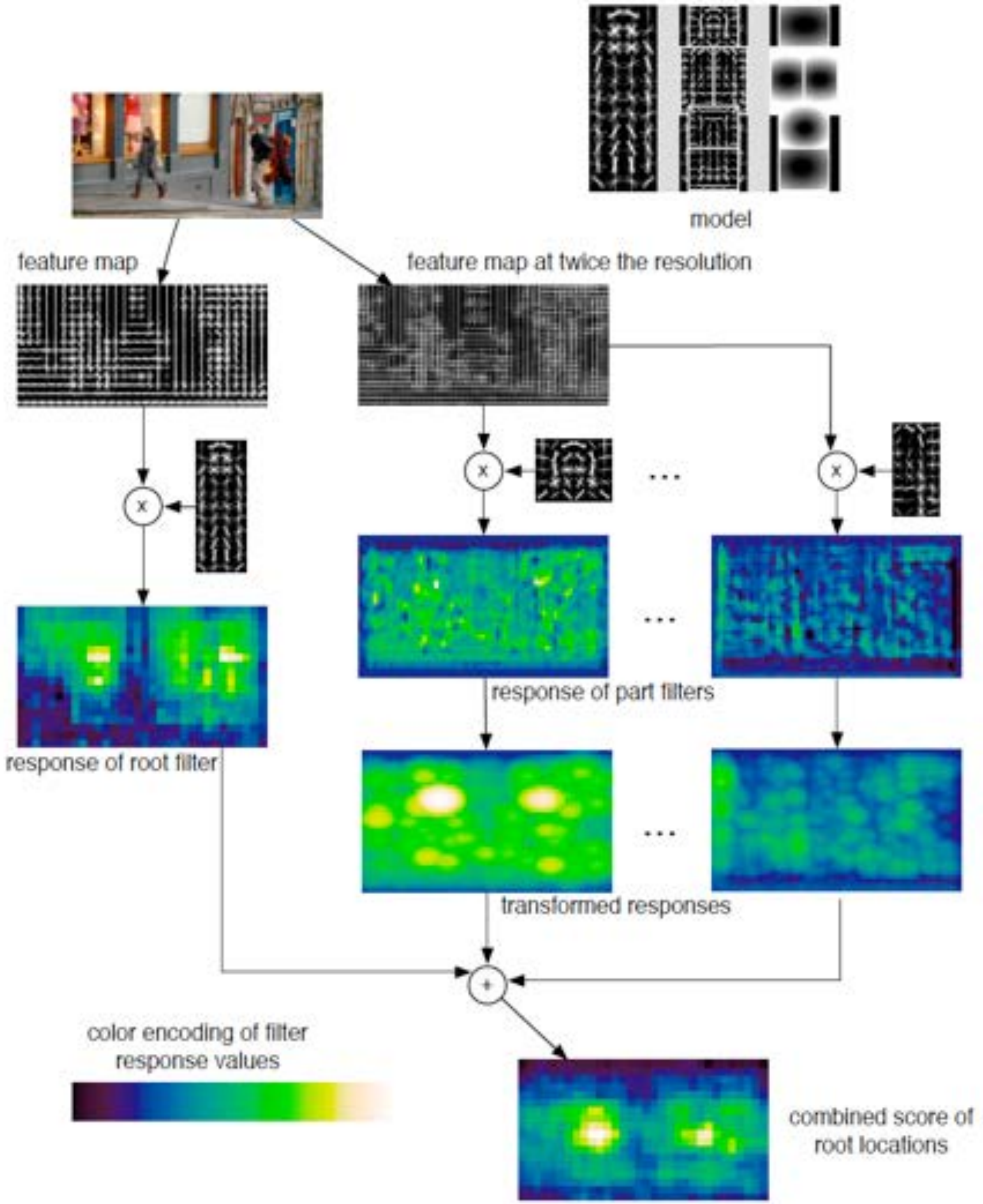


Transformed response

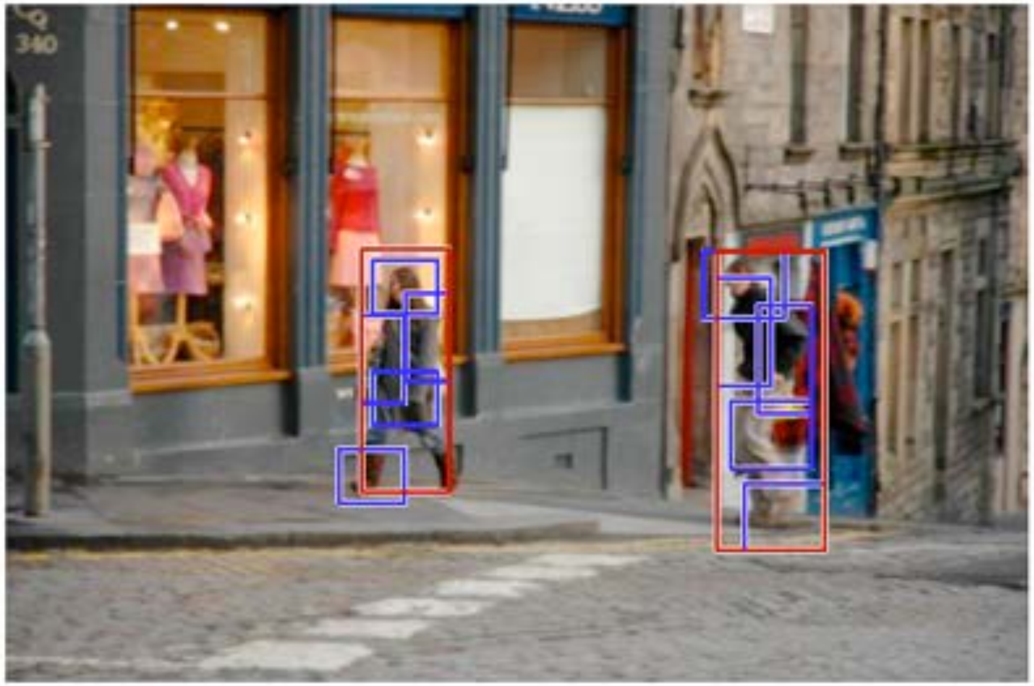
$$D_l(x, y) = \max_{dx,dy} \left(R_l(x + dx, y + dy) - d_i \cdot (dx^2, dy^2)\right)$$

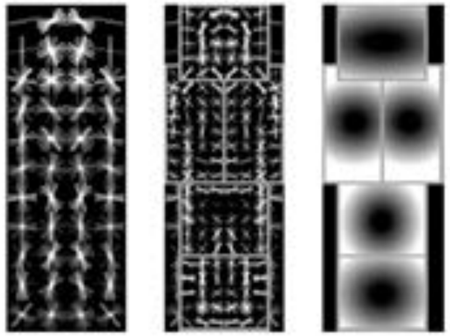
max-convolution, computed in linear time
(spreading, local max, etc)




model
feature map
feature map at twice the resolution
x
x
x
...
...
...
response of part filters
response of root filter
transformed responses
+
color encoding of filter response values
combined score of root locations

# Matching results



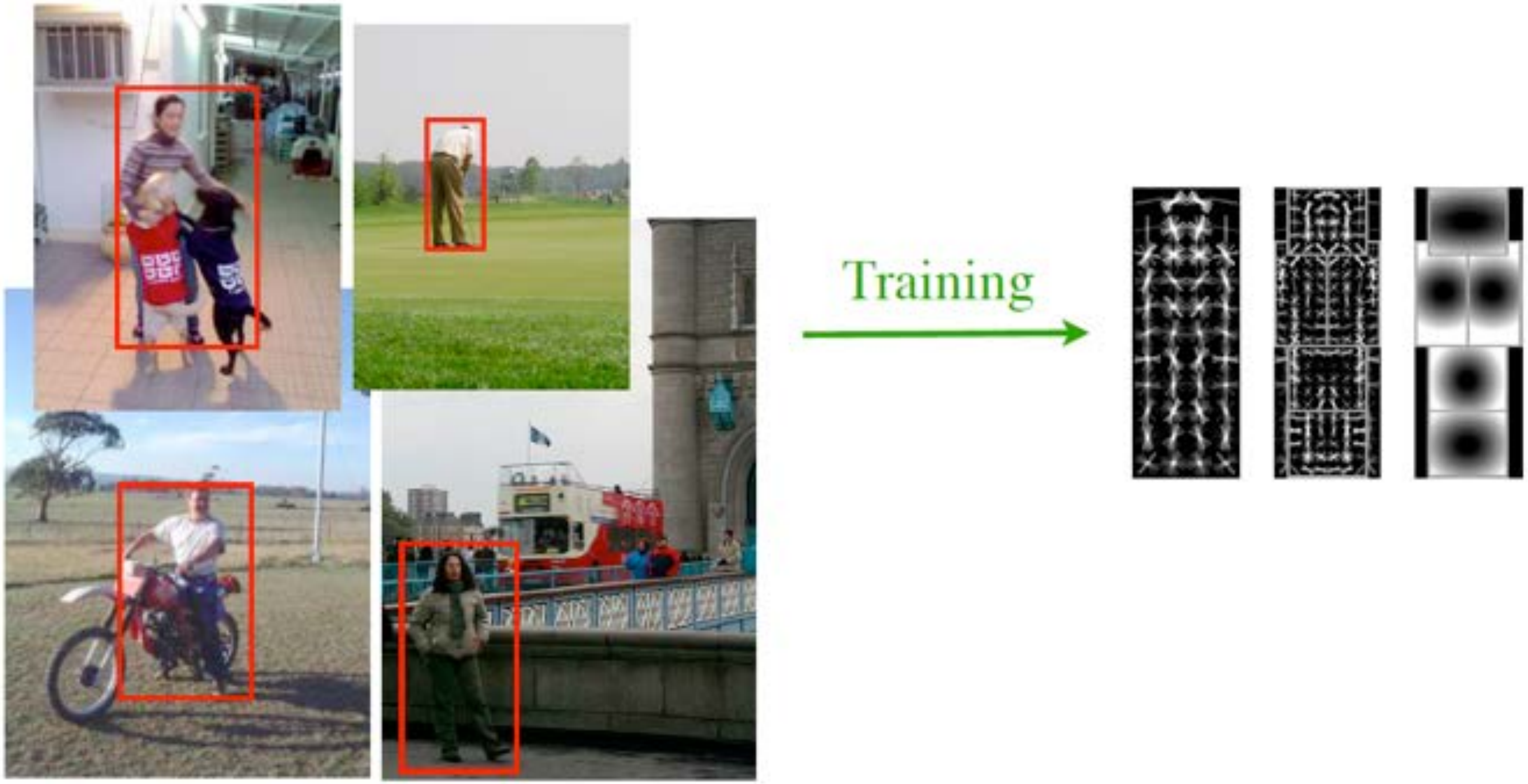



(after non-maximum suppression)

~1 second to search all scales

# Training

- Training data consists of images with labeled bounding boxes.
- Need to learn the model structure, filters and deformation costs.



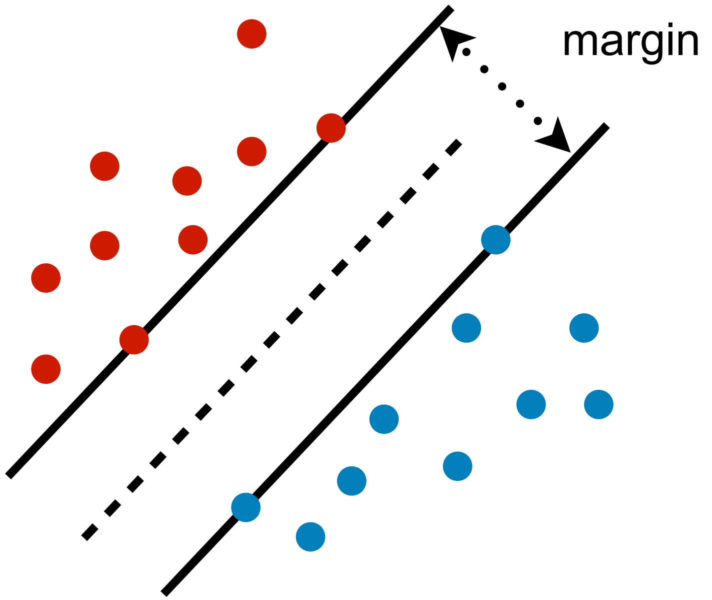
# SVM training

- Classifier scores and example x using:

$$f(x) = \beta \cdot \Phi(x)$$

  - model parameter: $\beta$
  - feature vector: $\Phi(x)$

- Linear SVM:
  - objective: maximize margin
    (for best generalization)



# SVM training

- Training data:

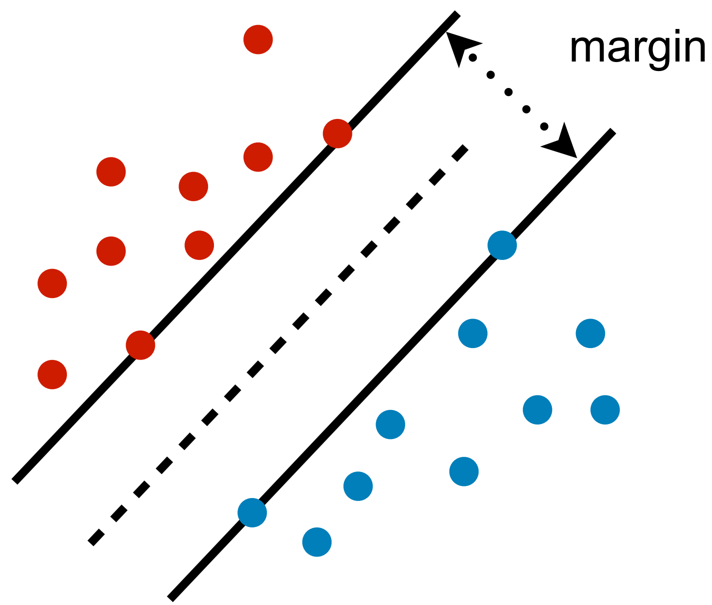
$$D = (\langle x_1, y_1 \rangle, \ldots, \langle x_n, y_n \rangle) \;\text{ with } y_i \in \{-1, 1\}$$

- Constraints:

$$f(x_i) \geq +1 \text{ for } y_i = +1$$
$$f(x_i) \leq -1 \text{ for } y_i = -1$$

$$\Rightarrow y_i f(x_i) \geq +1$$

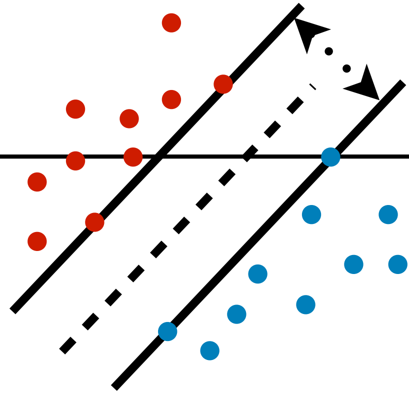
$$\Rightarrow 0 \geq 1 - y_i f(x_i)$$




- Training error:

$$\sum_{i=1}^{n} \max\left(0, 1 - y_i f(x_i)\right)$$

# SVM training

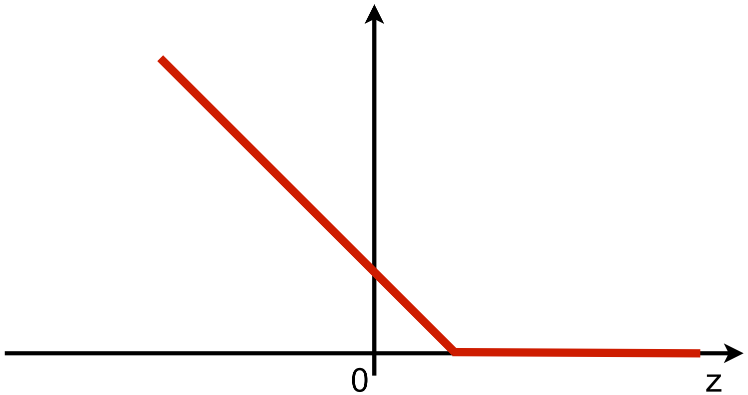


- Two objectives:
  - maximize margin: $$\min \frac{1}{2}||\beta||^2$$
  - minimize training error: $$\min \sum_{i=1}^{n} \max\left(0, 1 - y_i f(x_i)\right)$$
- Therefore minimize (primal formulation)

$$L(\beta) = \min_{\beta} \left( \frac{1}{2}||\beta||^2 + \sum_{i=1}^{n} \max\left(0, 1 - y_i f(x_i)\right) \right)$$

- Hinge loss:

$$H(z) = \max(0, 1 - z)$$



# Latent SVM (MI-SVM)

Classifiers that score an example $x$ using

$$f_\beta(x) = \max_{z \in Z(x)} \beta \cdot \Phi(x, z)$$

$\beta$ are model parameters
$z$ are latent values

Training data $D = (\langle x_1, y_1 \rangle, \ldots, \langle x_n, y_n \rangle) \qquad y_i \in \{-1, 1\}$

We would like to find $\beta$ such that: $y_i f_\beta(x_i) > 0$

Minimize

$$L_D(\beta) = \frac{1}{2}||\beta||^2 + C \sum_{i=1}^{n} \max(0, 1 - y_i f_\beta(x_i))$$

# Latent SVM training

$$L_D(\beta) = \frac{1}{2}||\beta||^2 + C\sum_{i=1}^{n} \max(0, 1 - y_i f_\beta(x_i))$$

- Convex if we fix $z$ for positive examples
- Optimization:
  - Initialize $\beta$ and iterate:
    - Pick best $z$ for each positive example
    - Optimize $\beta$ via gradient descent with data-mining

# Training algorithm, nested iterations

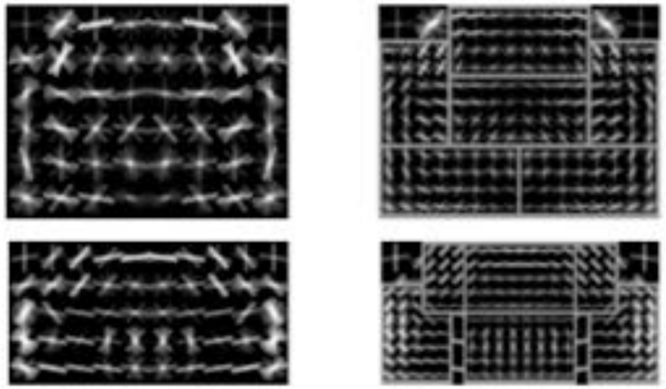
Fix "best" positive latent values for positives

Harvest high scoring (x,z) pairs from background images

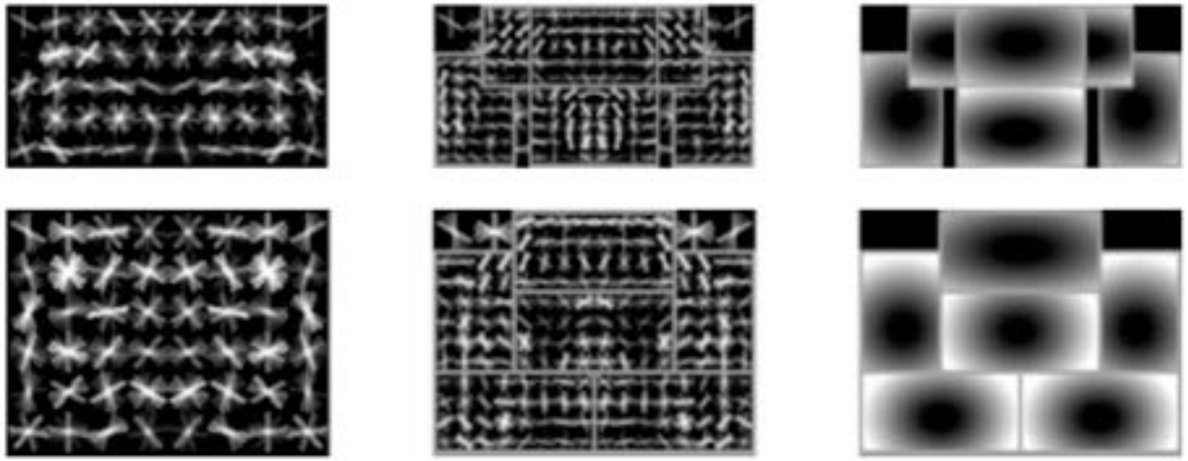
Update model using gradient descent

Trow away (x,z) pairs with low score

- Sequence of training rounds
  - Train root filters
  - Initialize parts from root
  - Train final model



# Car model



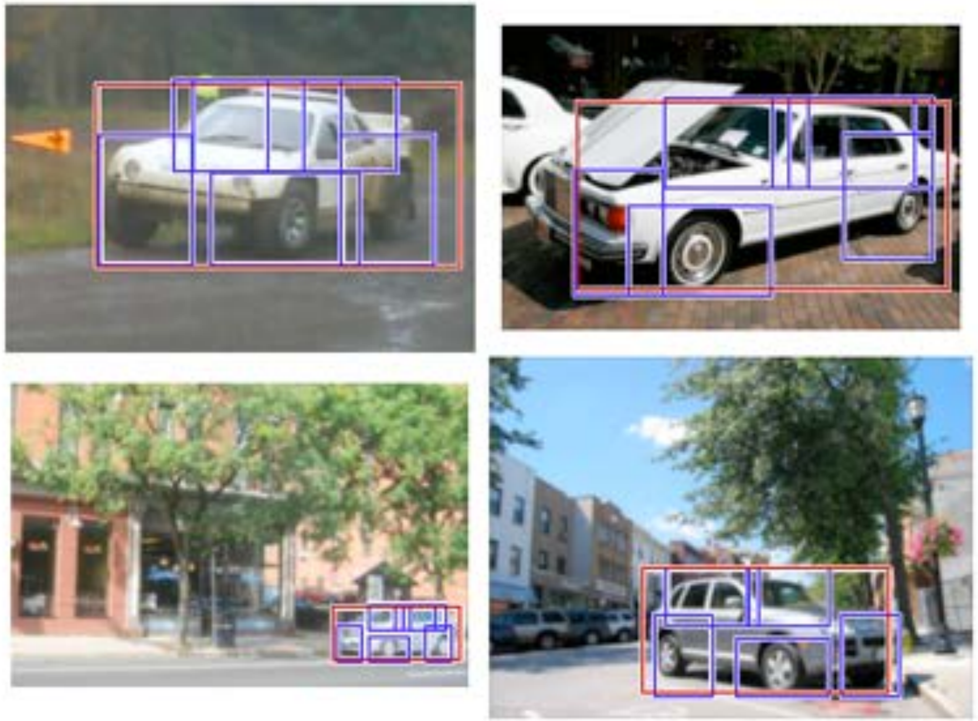
root filters
coarse resolution

part filters
finer resolution

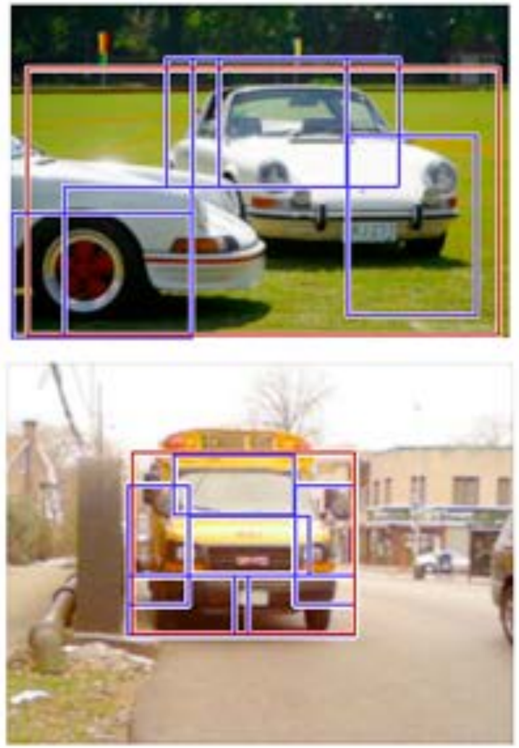
deformation
models

# Car detections

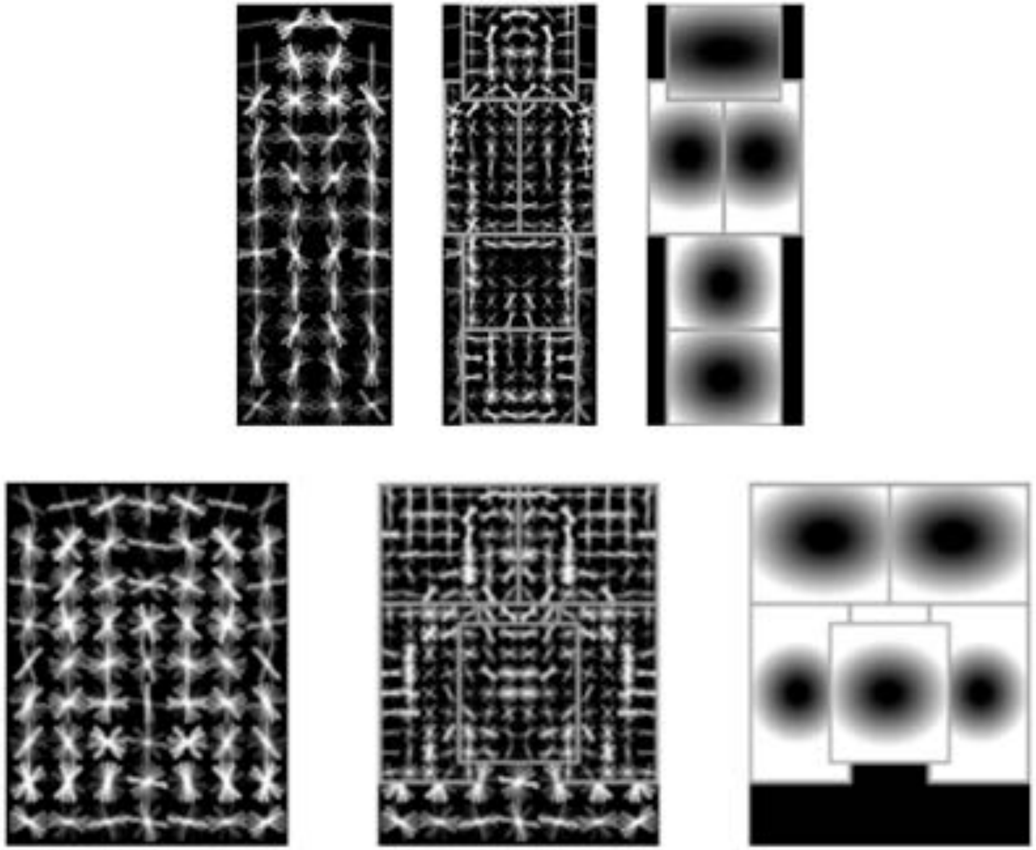
high scoring true positives



high scoring false positives



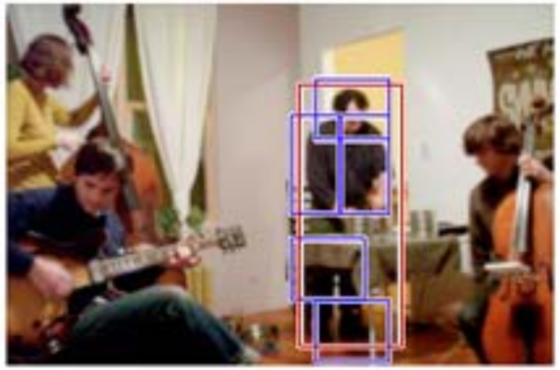
# Person model

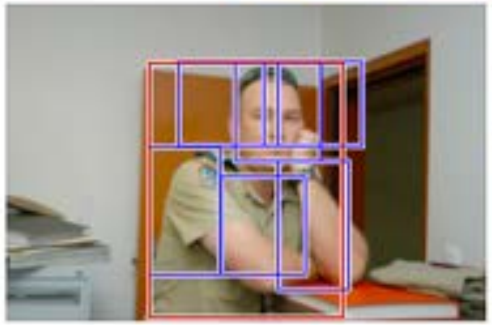


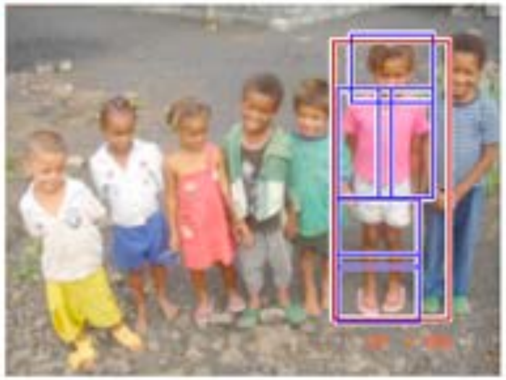
root filters coarse resolution | part filters finer resolution | deformation models

# Person detections

high scoring true positives





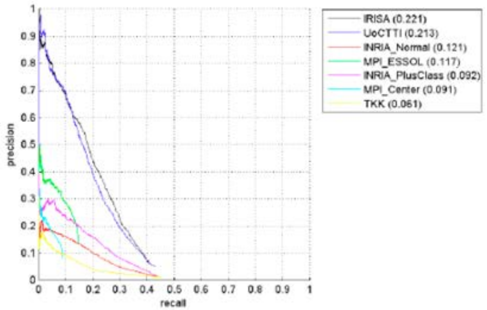


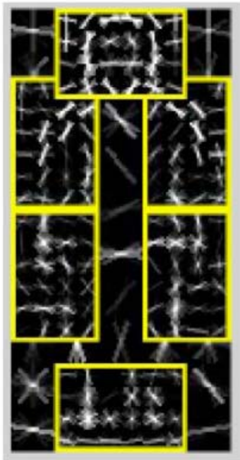
high scoring false positives
(not enough overlap)

# slides from Dan Huttenlocher

## PASCAL VOC 2007 Person Detection

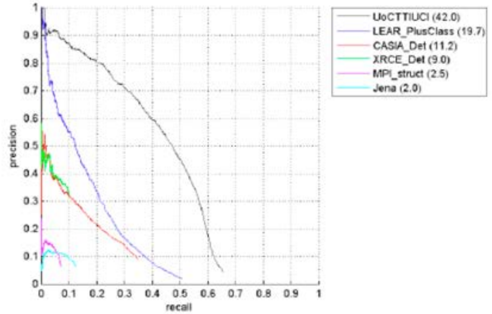
- Pictorial structure model
  - 45% precision at 20% recall

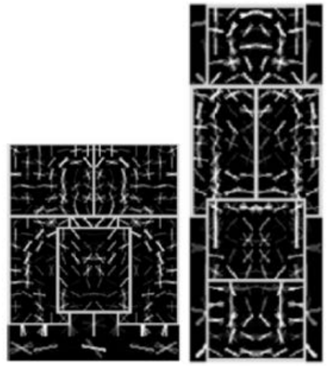


## PASCAL VOC 2008 Person Detection

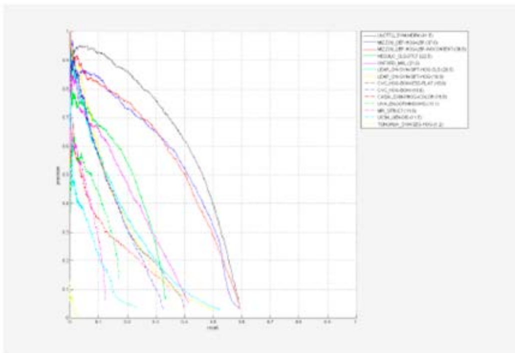
- Disjunction of two pictorial structures
  - 80% precision at 20% recall

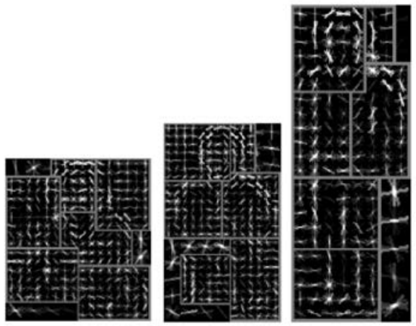


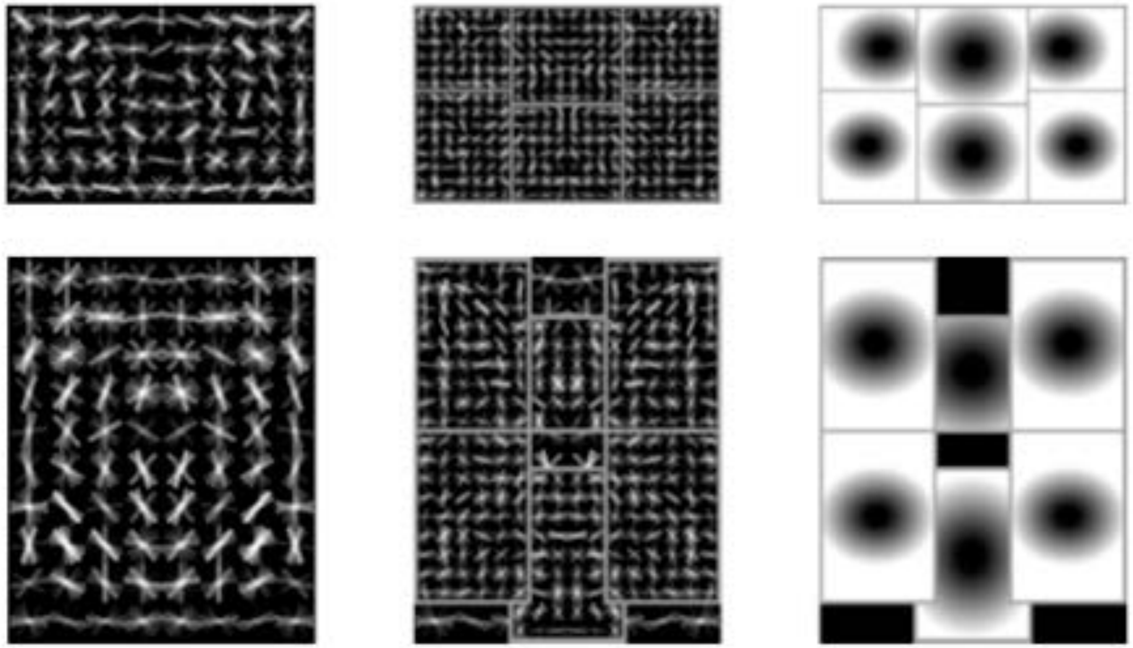
## PASCAL VOC 2009 Person Detection

- Disjunction of three pictorial structures
  - 85% precision at 20% recall





# Cat model



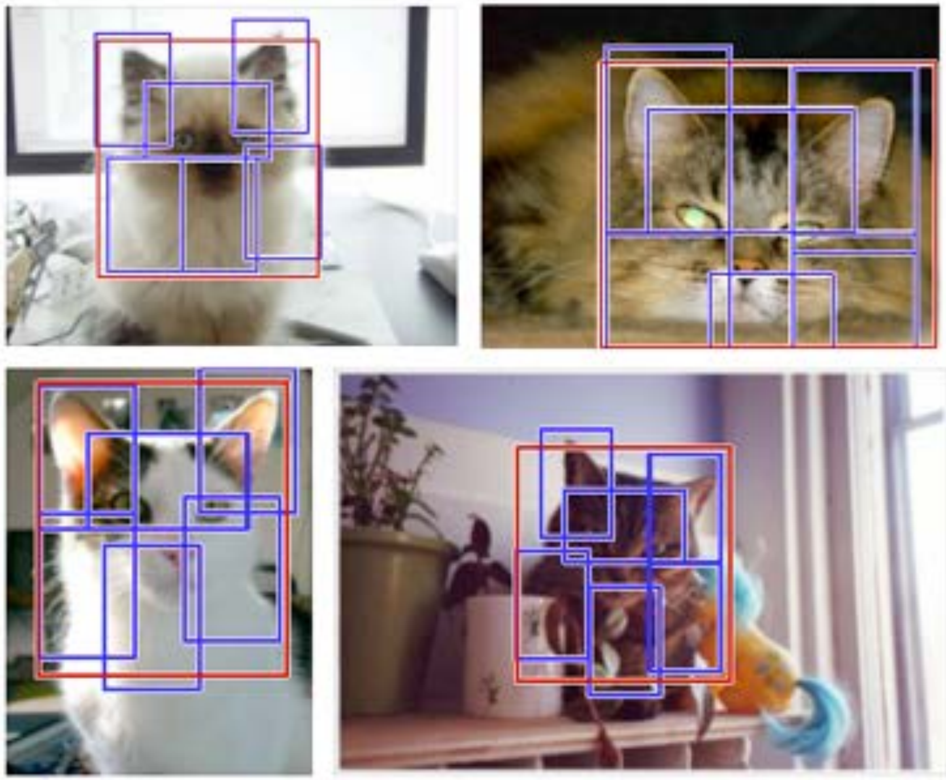
root filters
coarse resolution

part filters
finer resolution

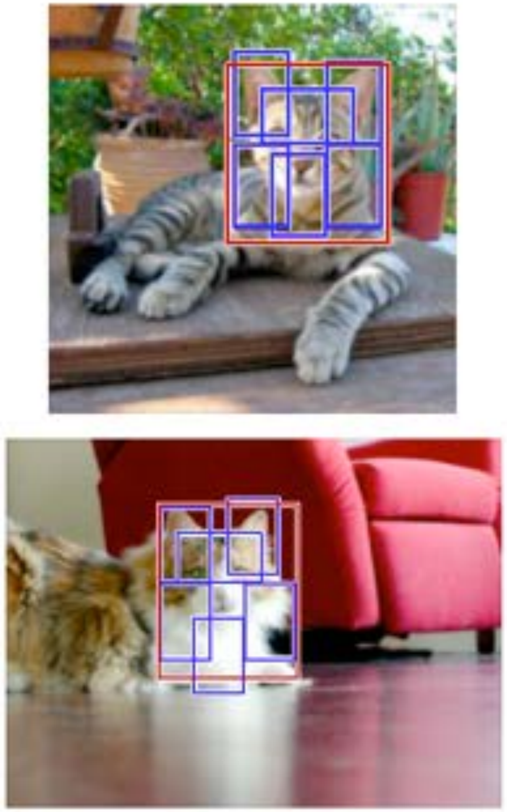
deformation
models

# Cat detections

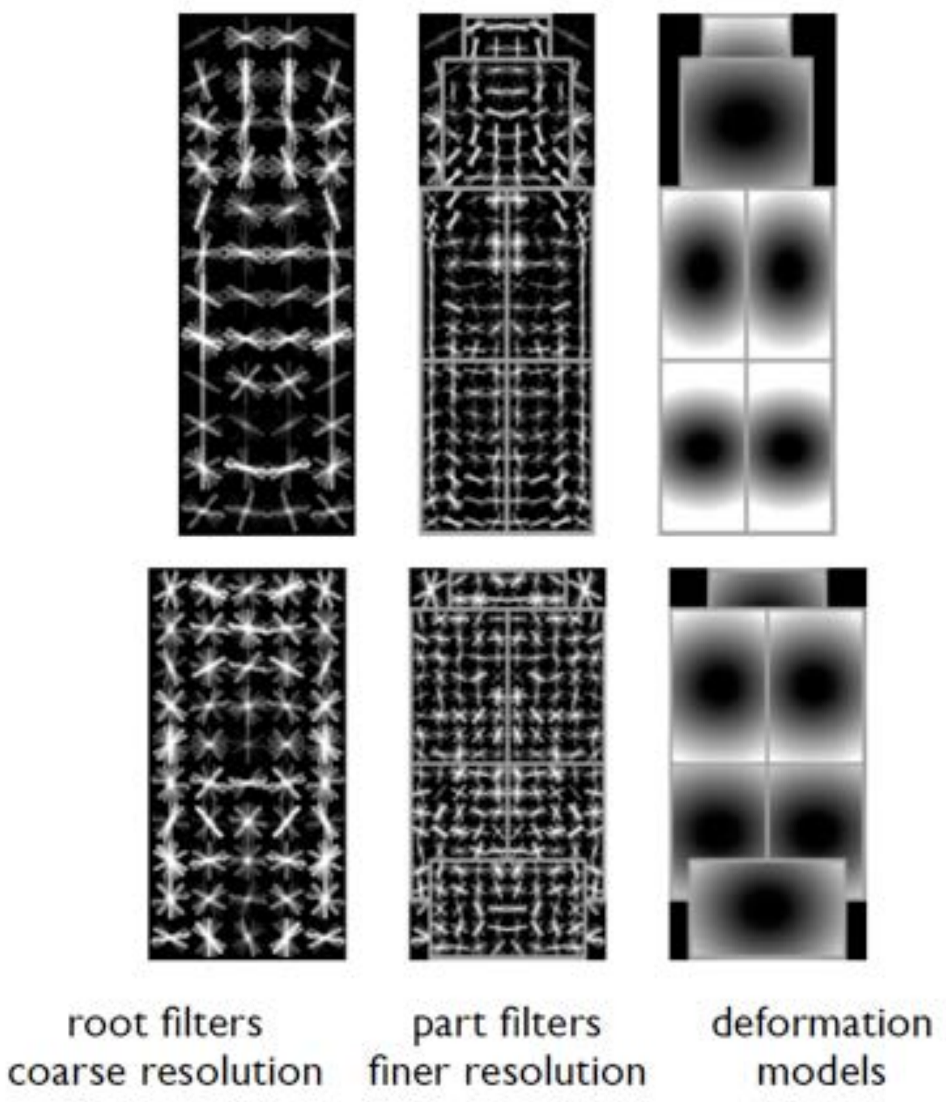
high scoring true positives



high scoring false positives
(not enough overlap)

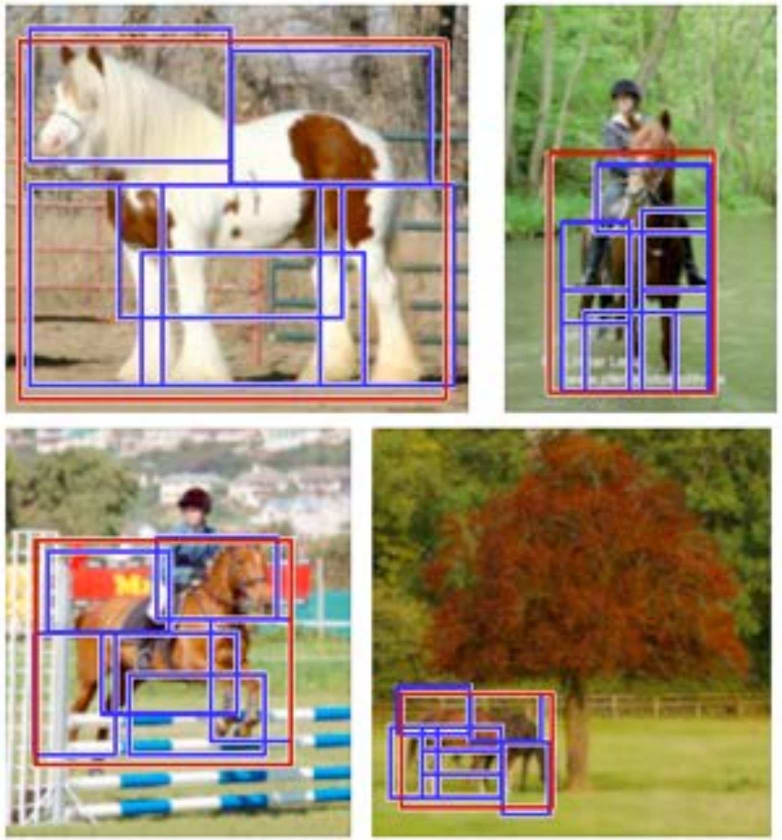


# Bottle model

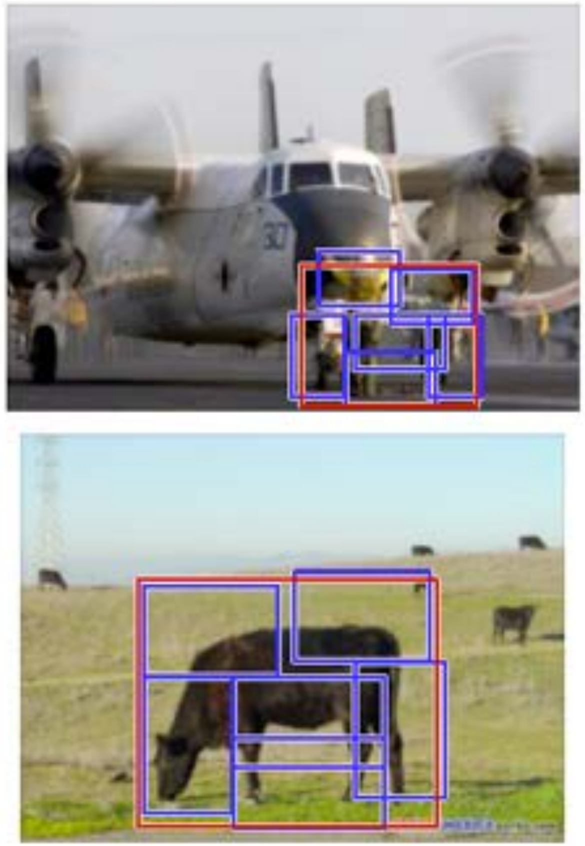


# Horse detections

high scoring true positives



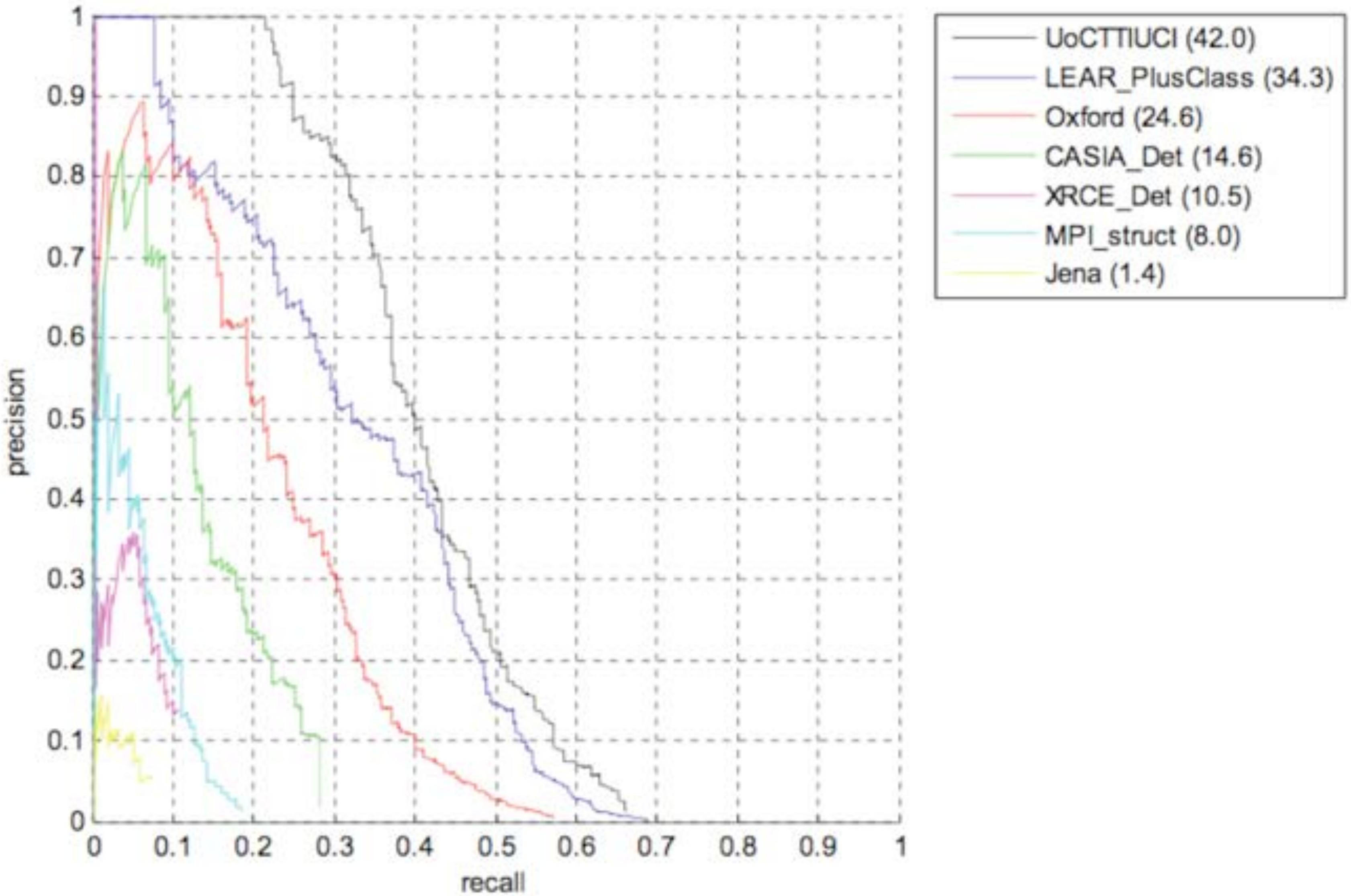
high scoring false positives



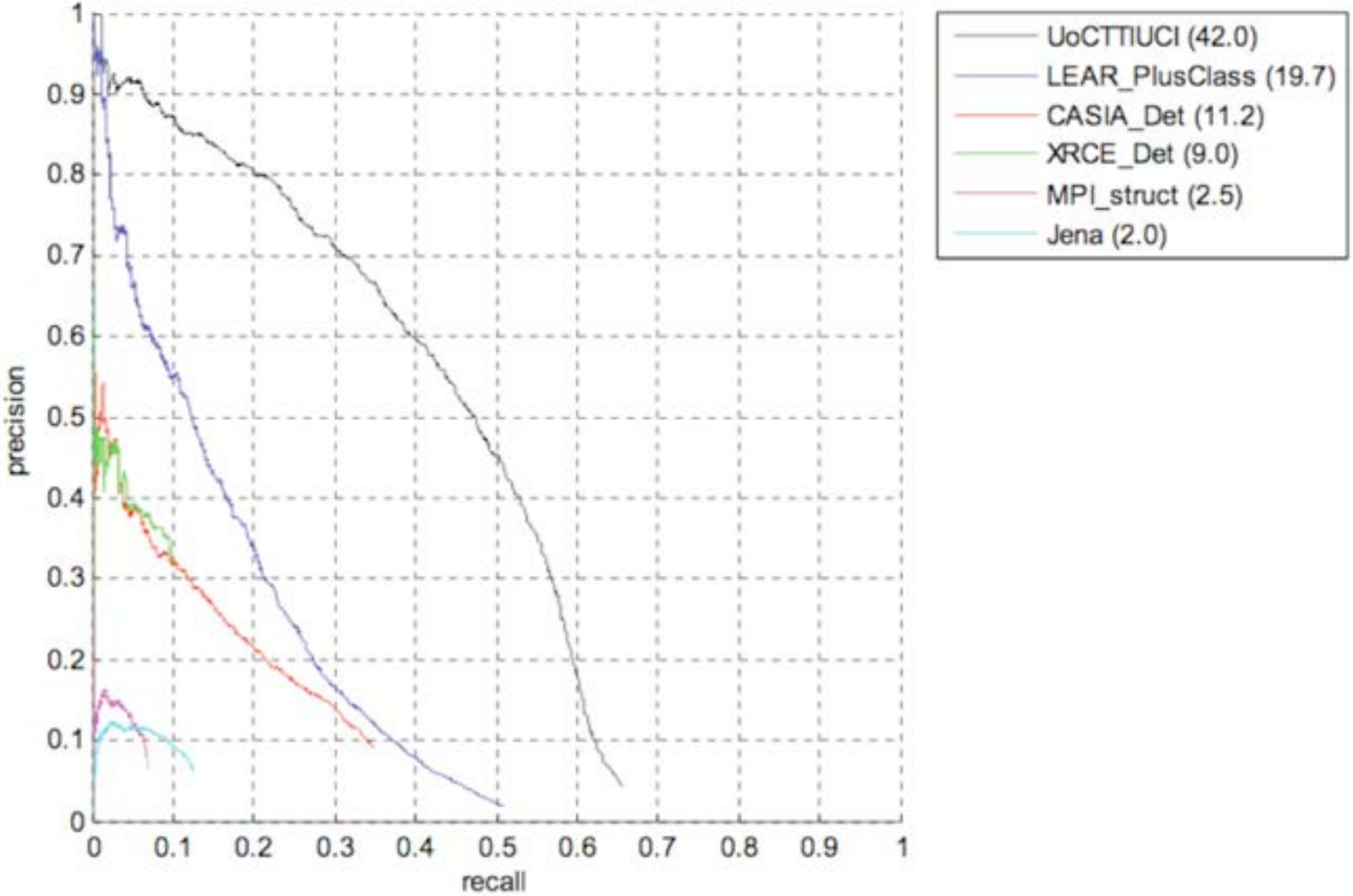
# Quantitative results

- 7 systems competed in the 2008 challenge
- Out of 20 classes we got:
  - First place in 7 classes
  - Second place in 8 classes
- Some statistics:
  - It takes ~2 seconds to evaluate a model in one image
  - It takes ~4 hours to train a model
  - MUCH faster than most systems.

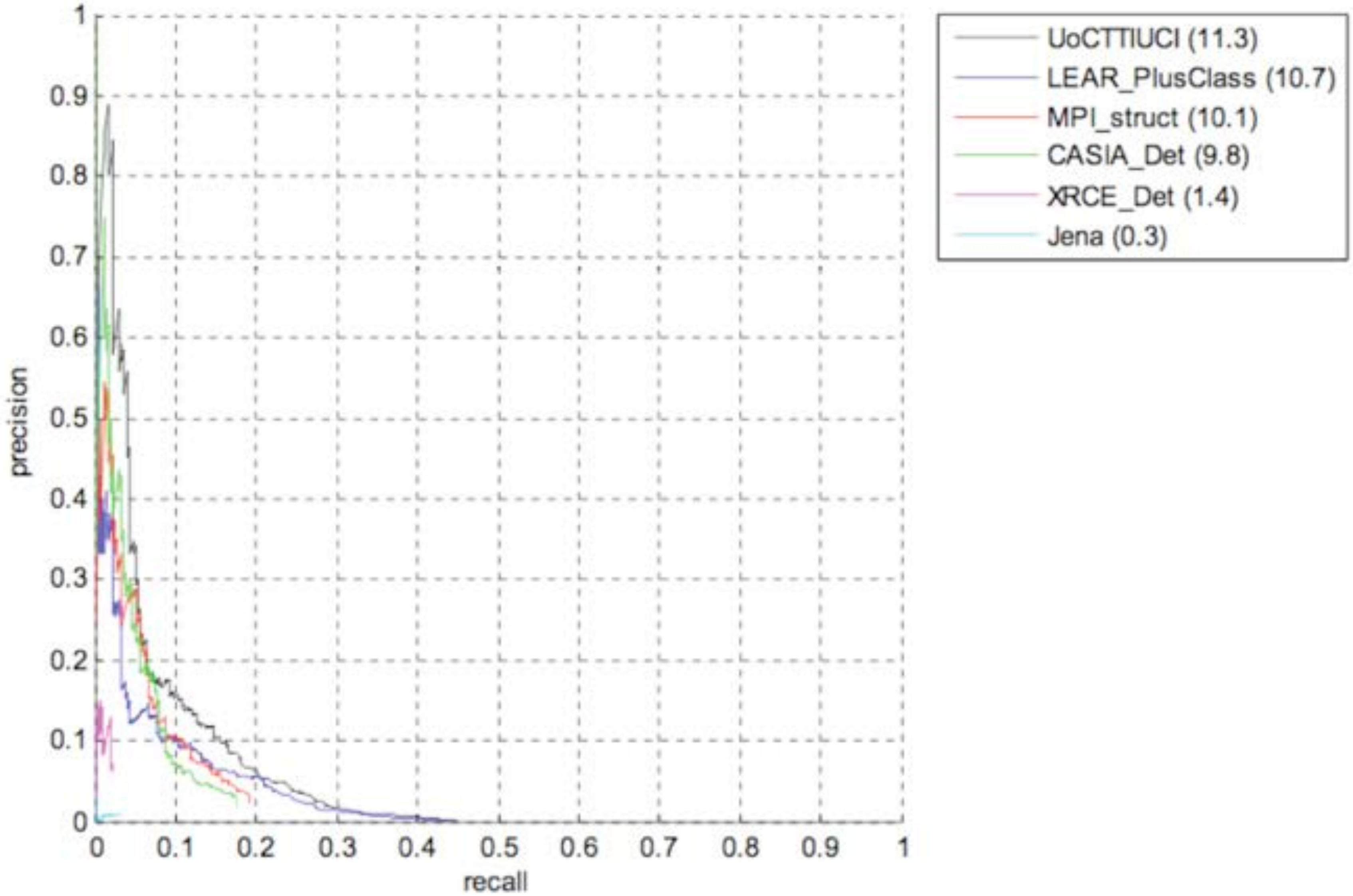
# Precision/Recall results on Bicycles 2008



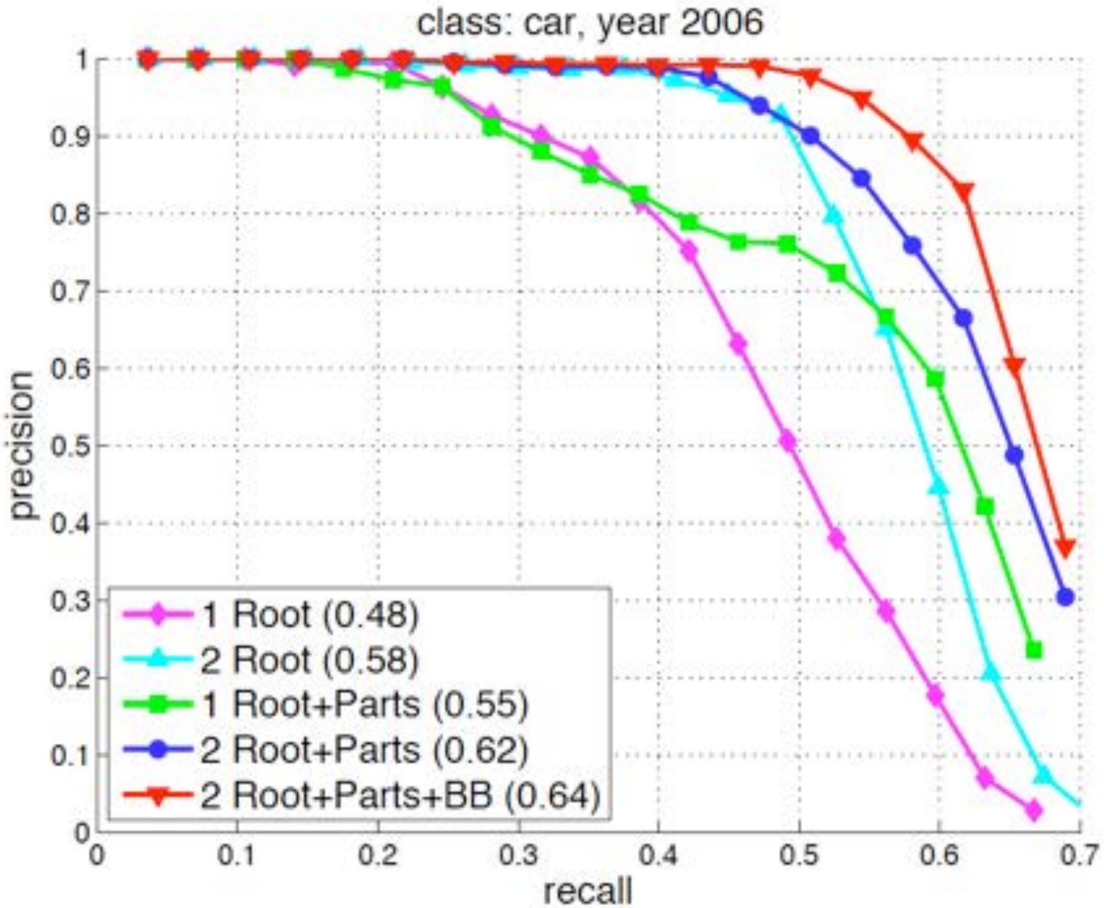
# Precision/Recall results on Person 2008



# Precision/Recall results on Bird 2008



# Comparison of Car models on 2006 data



# Summary

- Deformable models for object detection
  - Fast matching algorithms
  - Learning from weakly-labeled data
  - Leads to state-of-the-art results in PASCAL challenge
- Future work:
  - Hierarchical models
  - Visual grammars
  - AO* search (coarse-to-fine)

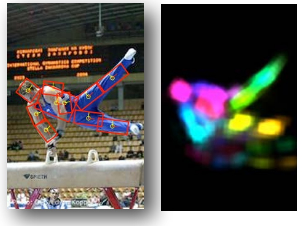
# Part-Based Models - Overview Today

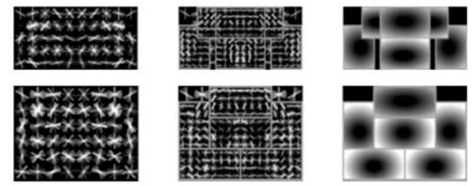
- Last Week:
  - Part-Based based on Manual Labeling of Parts
    - Detection by Components, Multi-Scale Parts
  - The Constellation Model
    - automatic discovery of parts and part-structure
  - The Implicit Shape Model (ISM)
    - star-model of part configurations, parts obtained by clustering interest-points

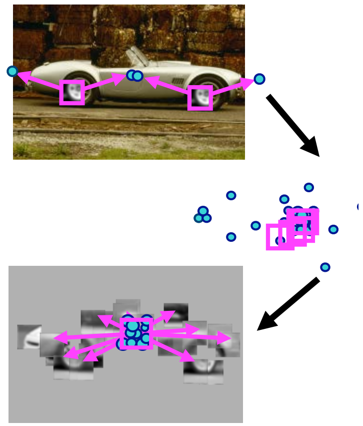
- Today:
  - Pictorial Structures Model
  - Learning Object Model from CAD Data
  - Deformable Parts Model (DPM)
  - Discussion Semantic Parts vs. Discriminative Parts

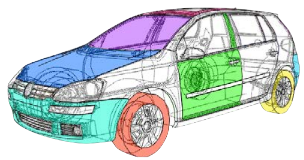
# What are Ideal Parts for Part-Based Object Models?

- Parts can/may
  - be semantic body parts - e.g. for articulated human body pose estimation
  - be feature clusters (typically many clusters) (e.g. ISM, constellation model, BoW)
  - support learnability of discriminant appearance (e.g. DPM model)
  - enable correspondence across 3D models
  - ...







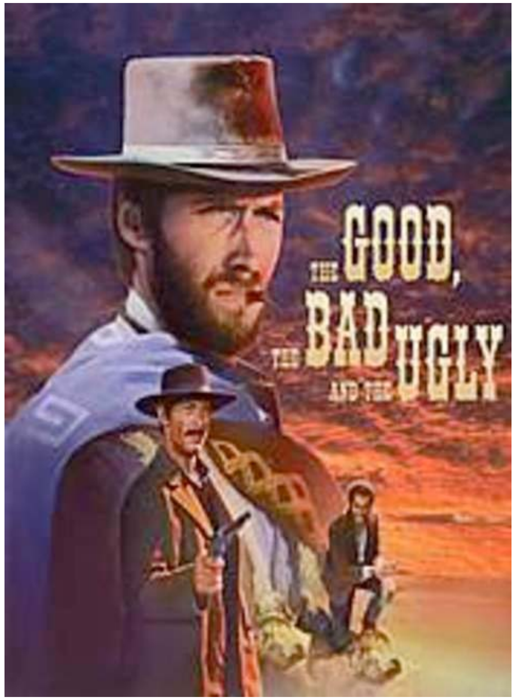
- in all those cases: the most important property is that "parts" facilitate correspondence across object instances

# What are Ideal Parts for Part-Based Object Models?

- Multiple motivations for part based models exist:
  - intuitiveness: semantic meaning of parts/attributes is attractive (e.g. enables use of language sources)
  - learnability: sharing of parts/attributes across instances/classes
  - scalability: transferability of parts/attributes across classes
  - ...
- in general, parts support learnability and scalability when they facilitate correspondence
  - across object instances
  - across object classes
  - across modalities (e.g. from language to visual appearance)
  - and semantics is only a secondary concern (for "intuitiveness")

max planck institut
informatik

UNIVERSITÄT
DES
SAARLANDES




# What are Ideal Parts for Part-Based Object Models?

# the Good, the Bad, and the Ugly

**thanks to:** Micha Andriluka, Bastian Leibe, Sandra Ebert, Mario Fritz, Diane Larlus, Marcus Rohrbach, Paul Schnitzspan, Stefan Roth, Michael Stark, Michael Goesele