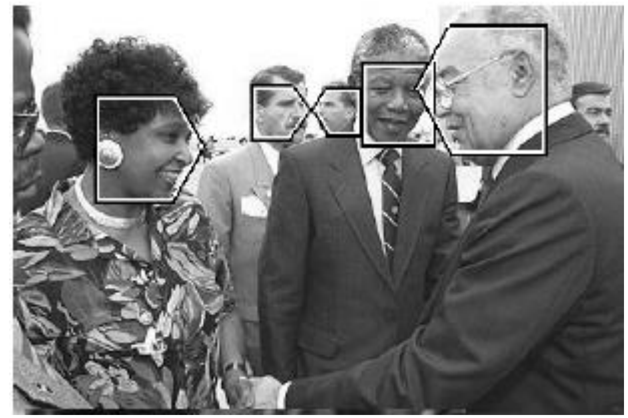
Object Category Detection: Parts-based Models

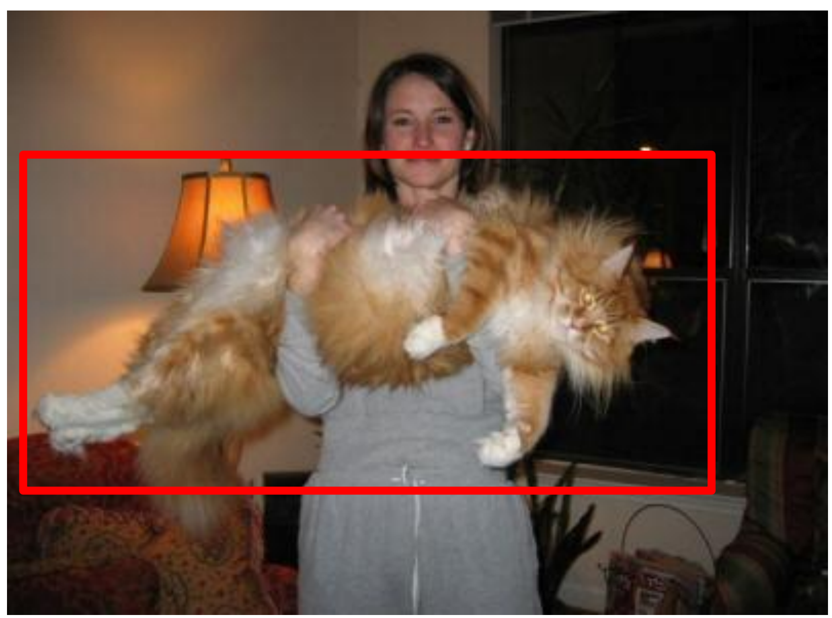
Slides borrowed from Derek Hoiem

Goal: Detect all instances of objects

Cars

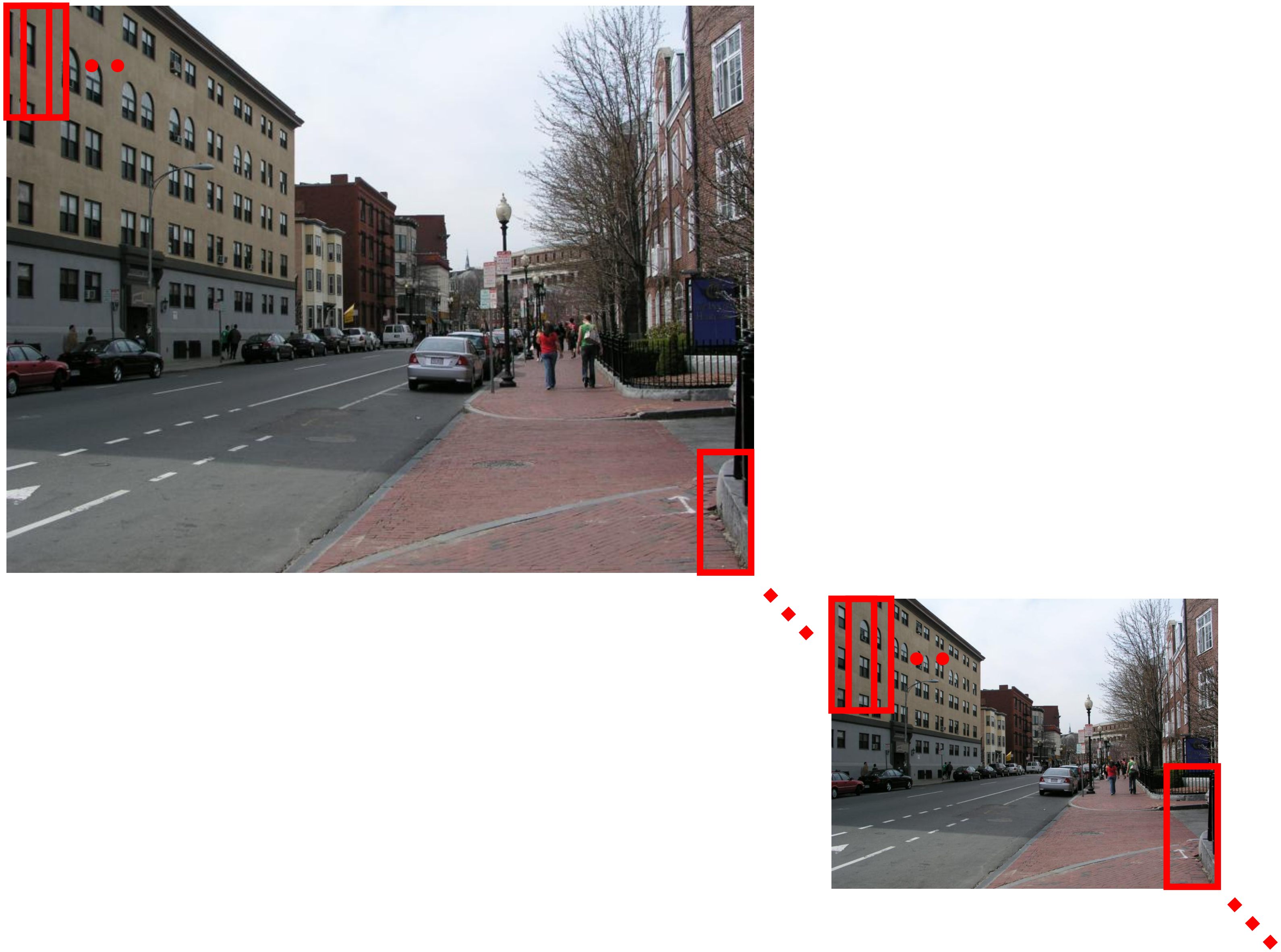
Faces





Cats

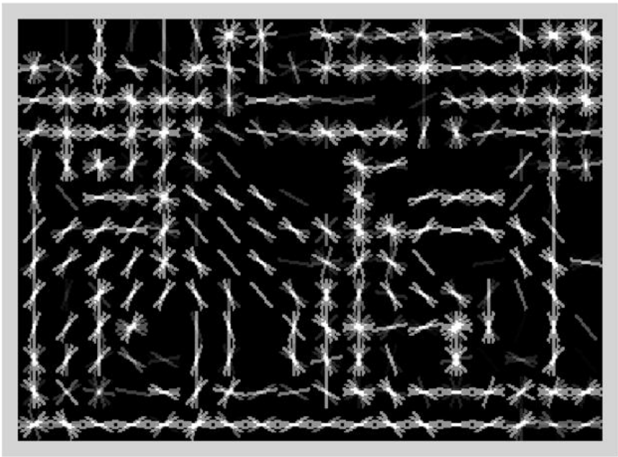
Last class: sliding window detection



Object model: last class

- Statistical Template in Bounding Box
  - Object is some (x,y,w,h) in image
  - Features defined wrt bounding box coordinates

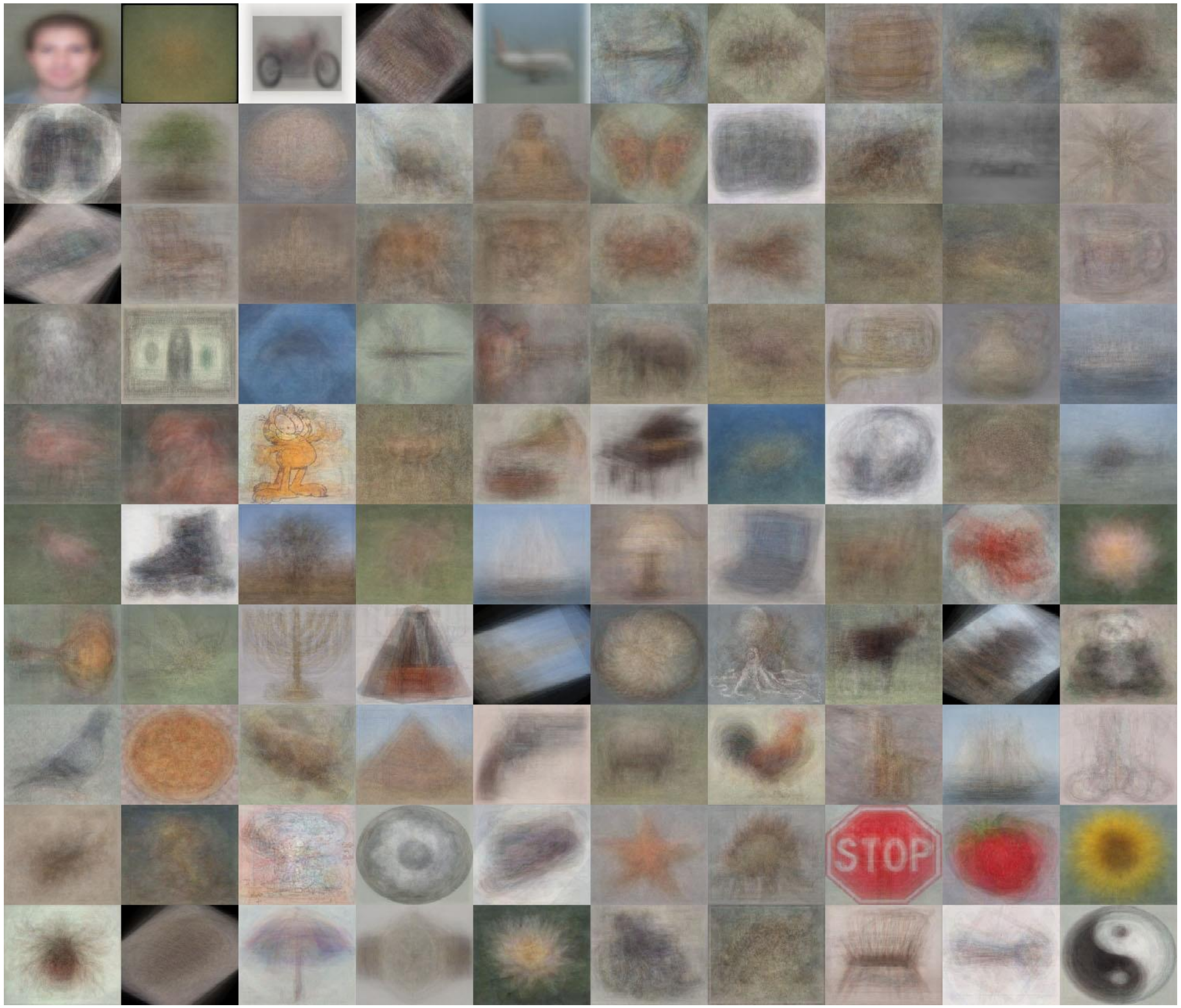
Image



Template Visualization

Images from Felzenszwalb

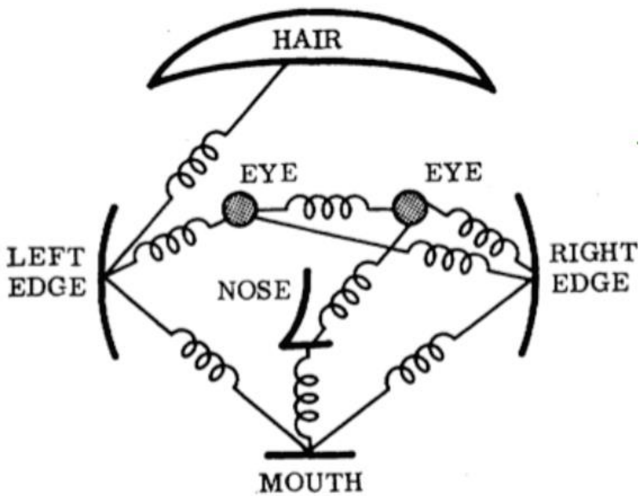
When do statistical templates make sense?

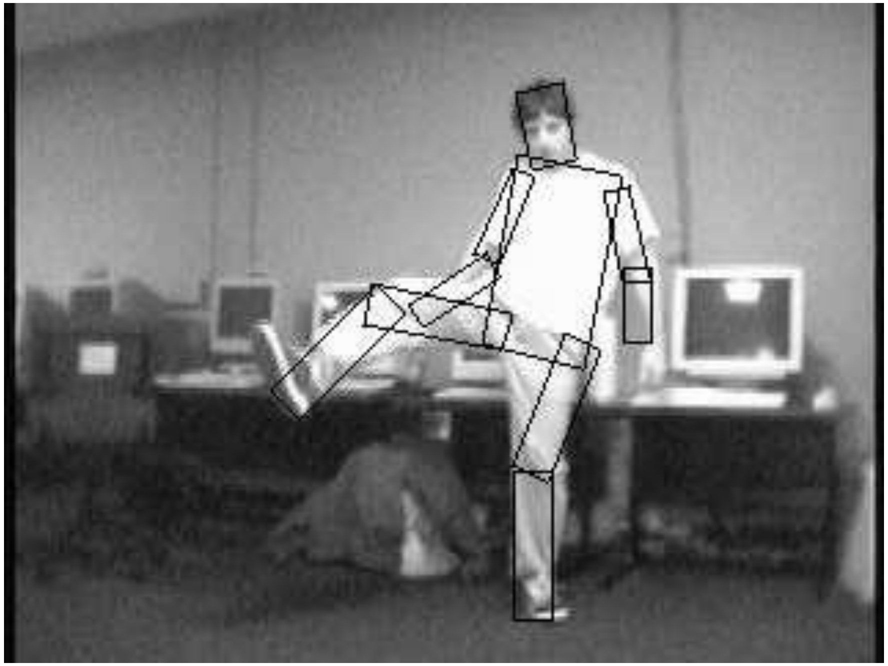


Caltech 101 Average Object Images

Object models: this class

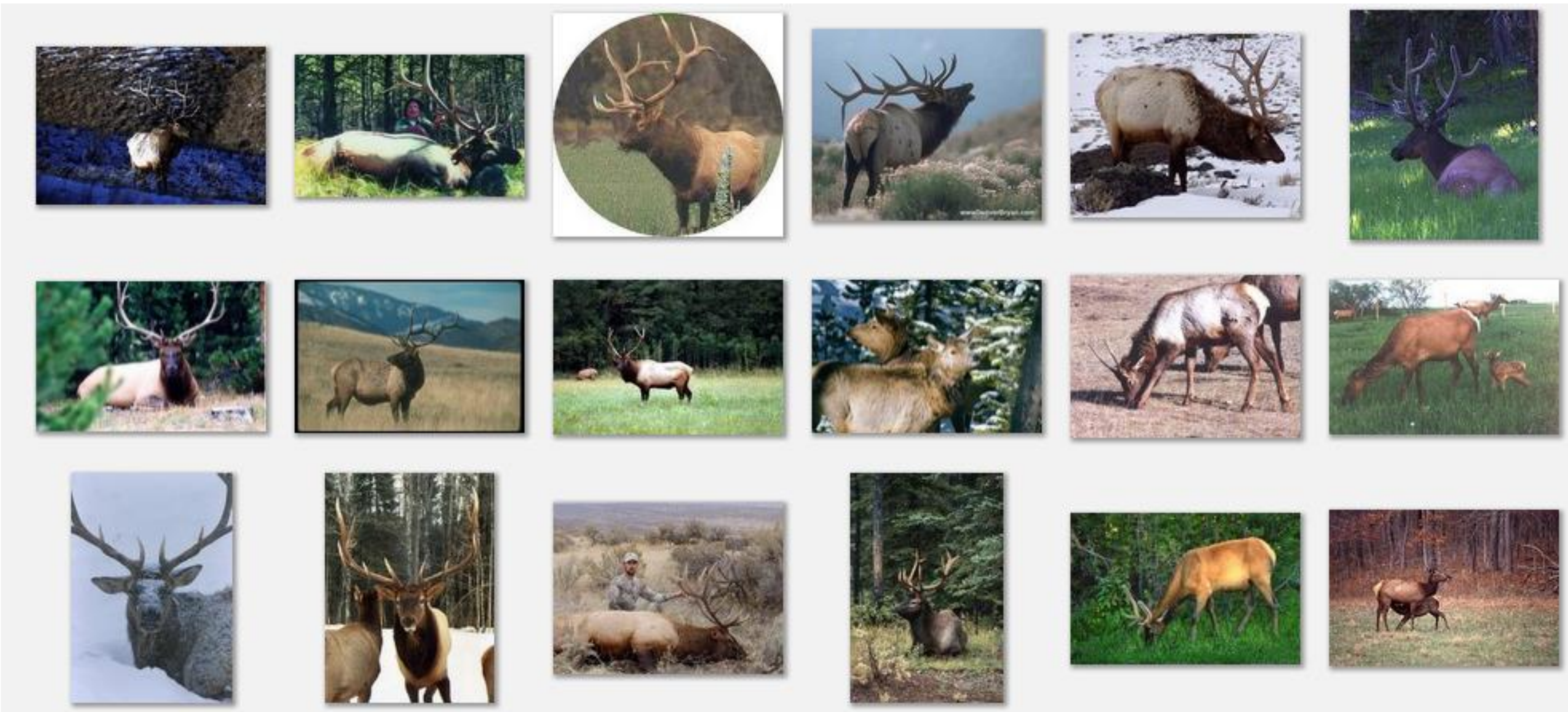
- Articulated parts model
  - Object is configuration of parts
  - Each part is detectable





Images from Felzenszwalb

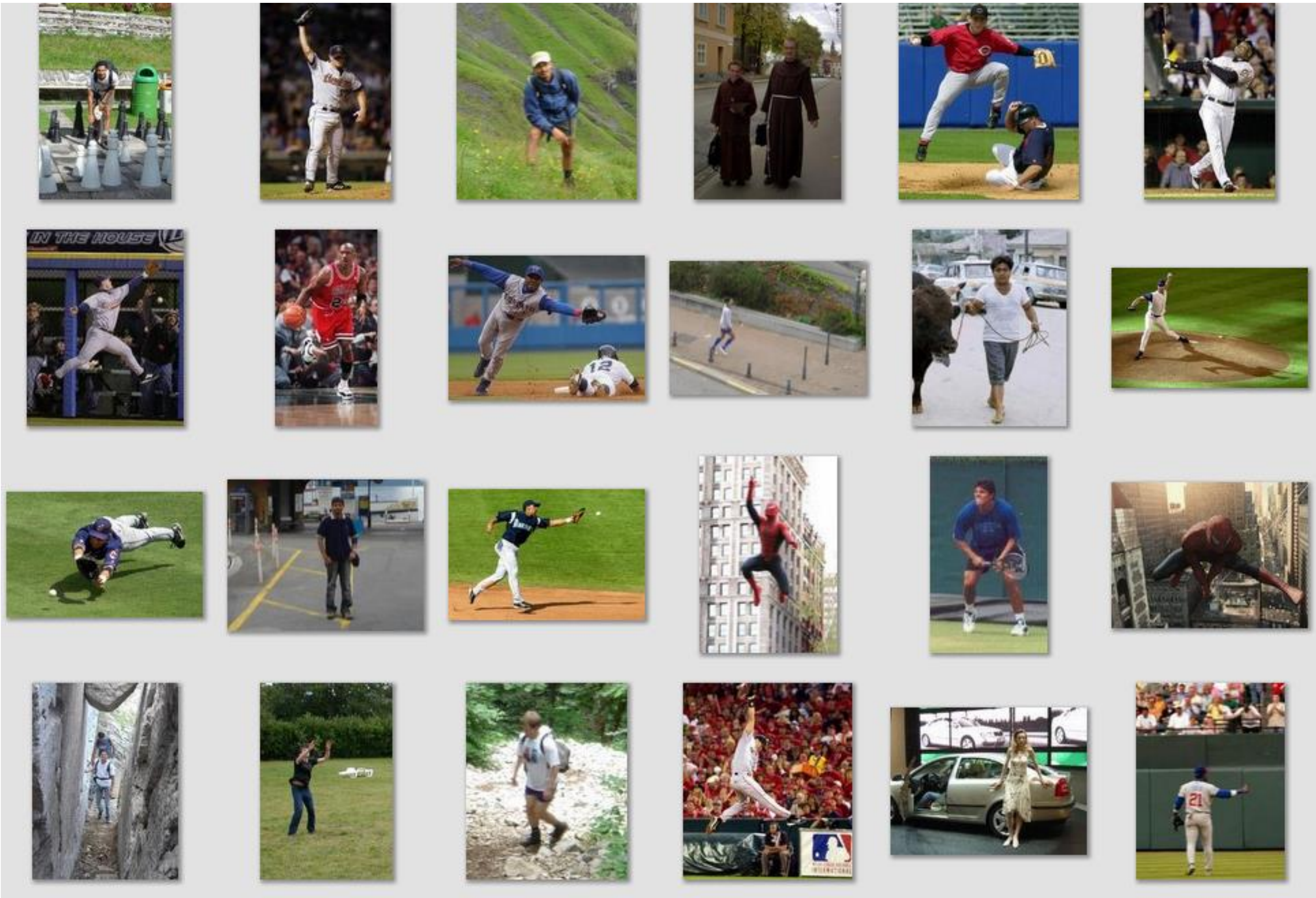
Deformable objects



Images from Caltech-256

Slide Credit: Duan Tran

Deformable objects



Images from D. Ramanan's dataset

Slide Credit: Duan Tran

Compositional objects







Parts-based Models

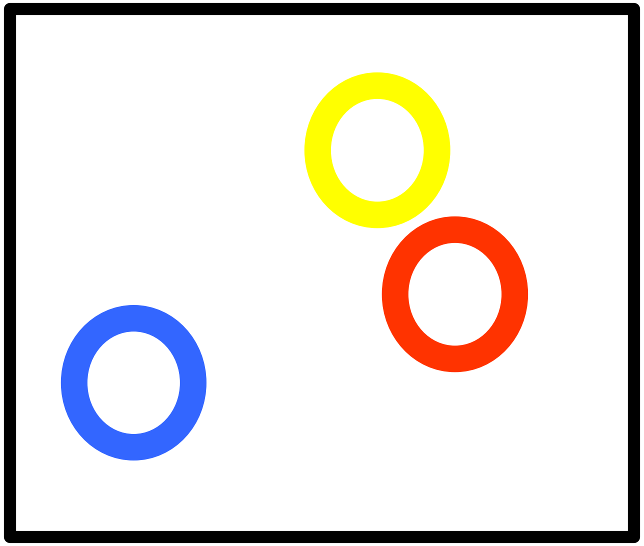
Define object by collection of parts modeled by

1. Appearance
2. Spatial configuration

Slide credit: Rob Fergus

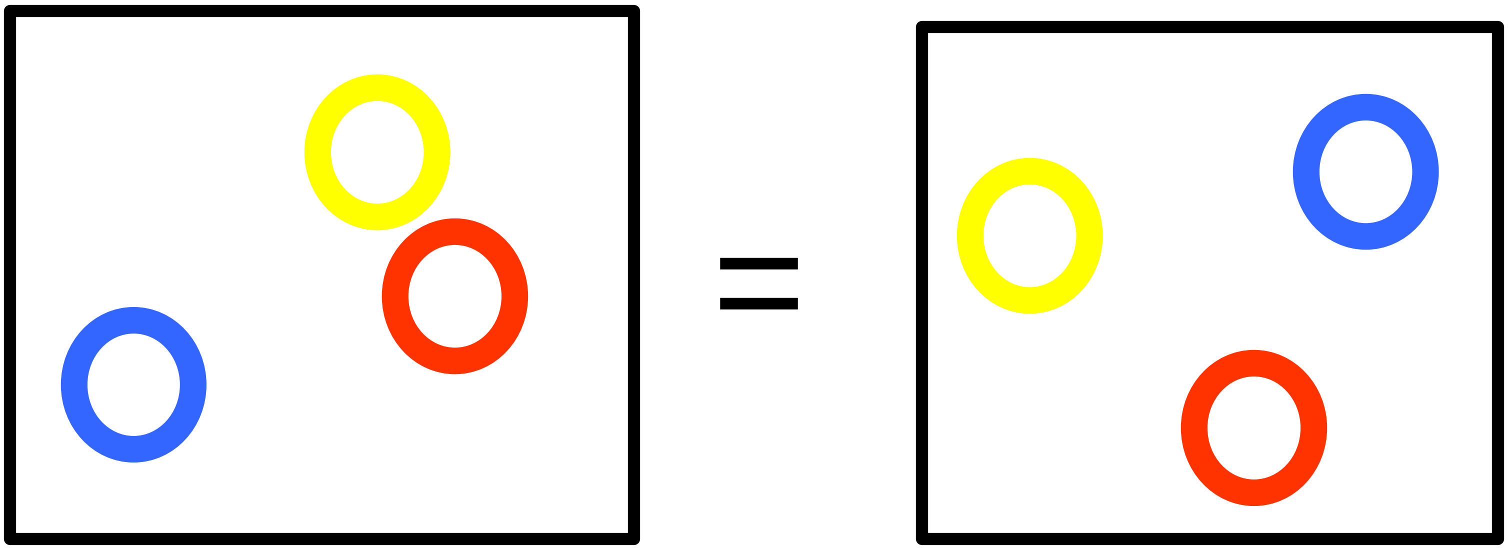
How to model spatial relations?

- One extreme: fixed template



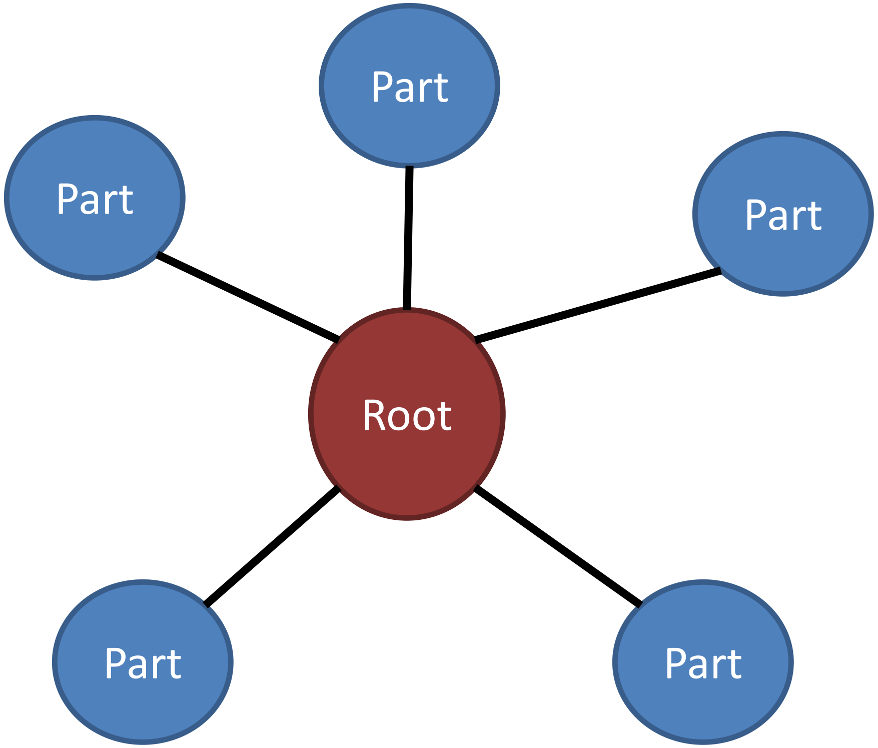
How to model spatial relations?

- Another extreme: bag of words



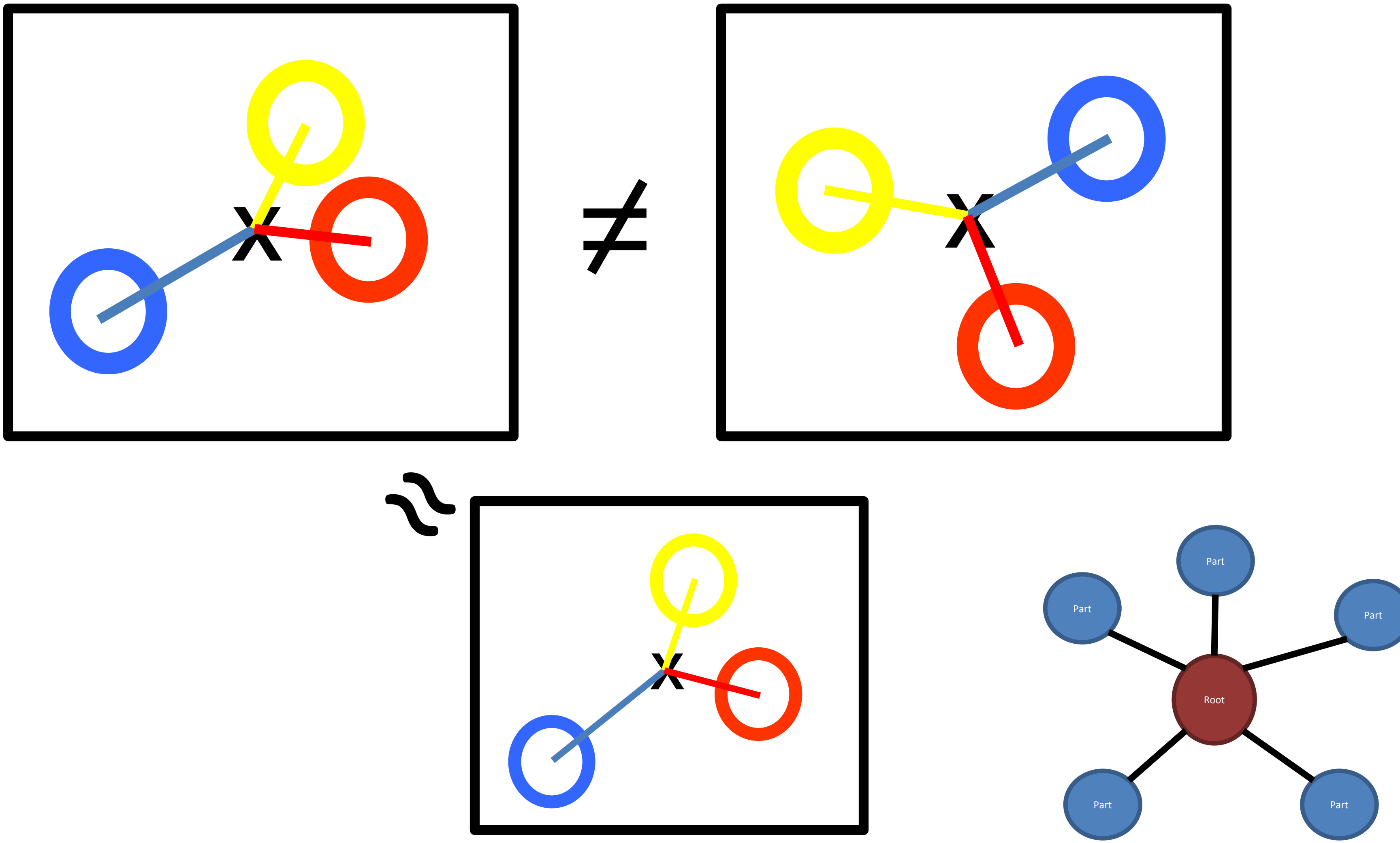
How to model spatial relations?

- Star-shaped model



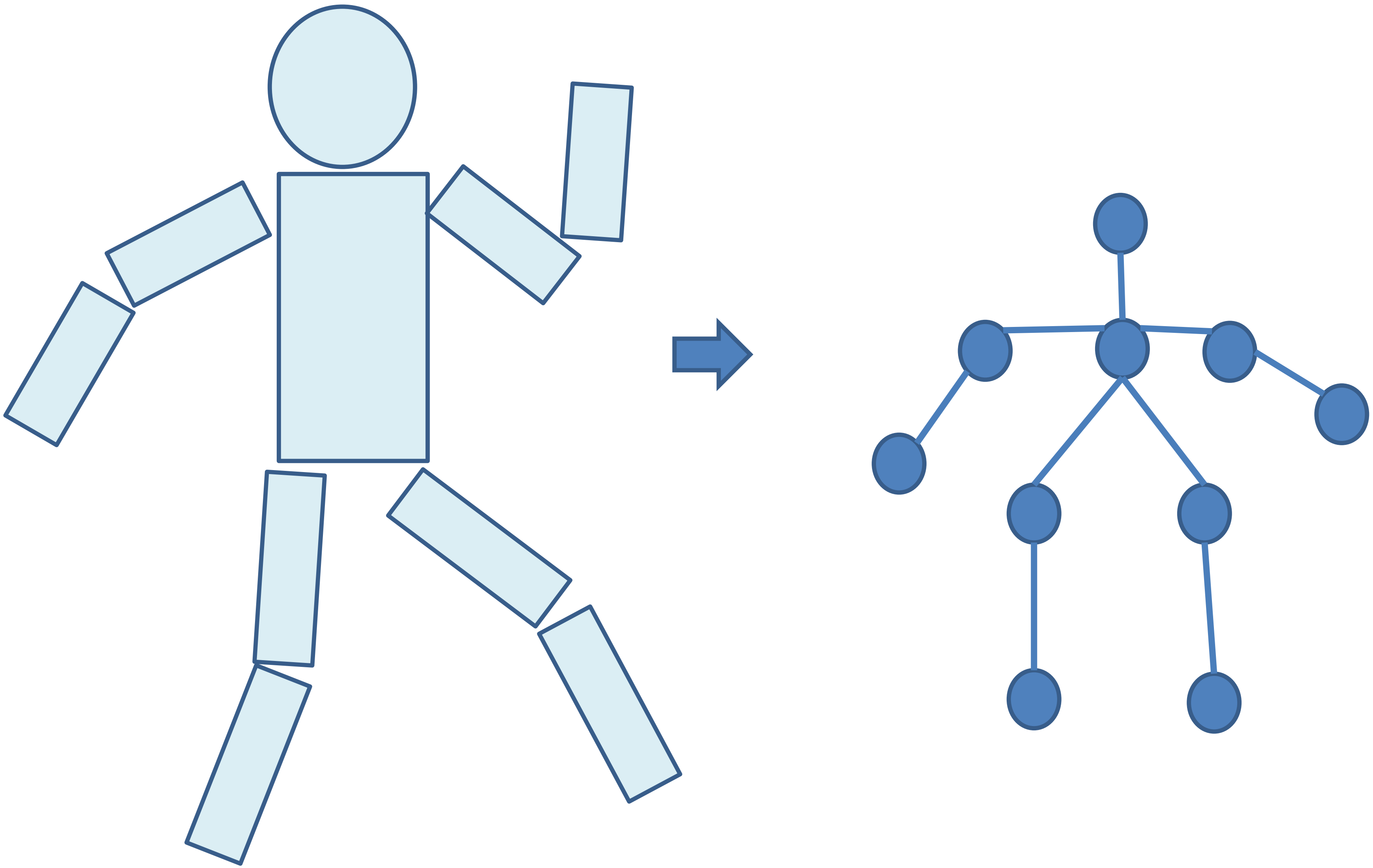
How to model spatial relations?

- Star-shaped model



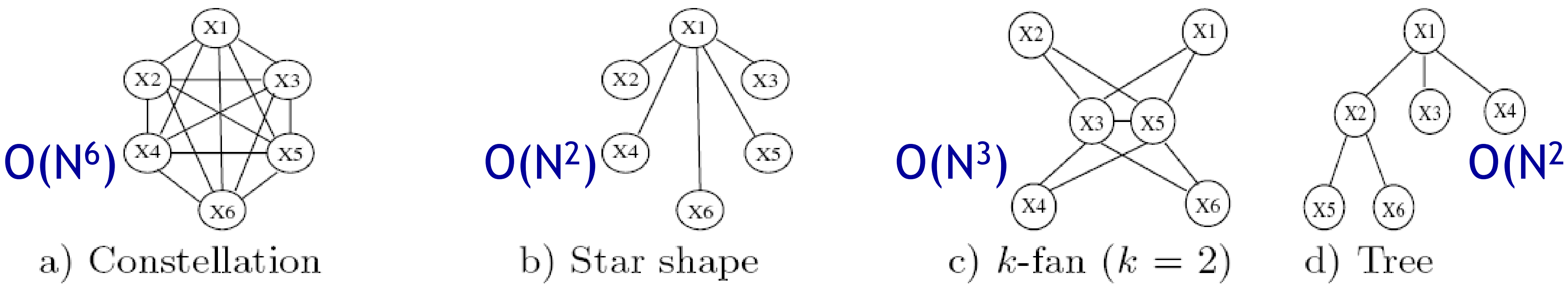
How to model spatial relations?

- Tree-shaped model



How to model spatial relations?

- Many others...

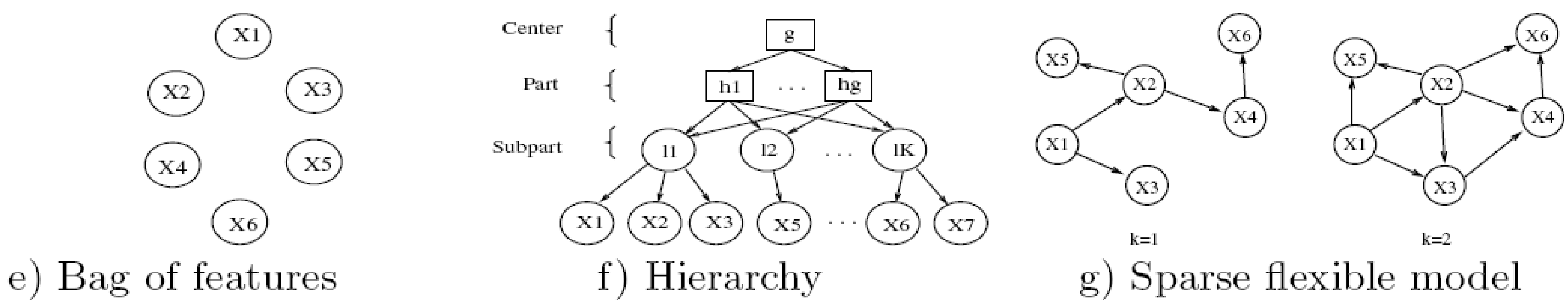


Fergus et al. '03
Fei-Fei et al. '03

Leibe et al. '04, '08
Crandall et al. '05
Fergus et al. '05

Crandall et al. '05

Felzenszwalb & Huttenlocher '05



Csurka '04
Vasconcelos '00

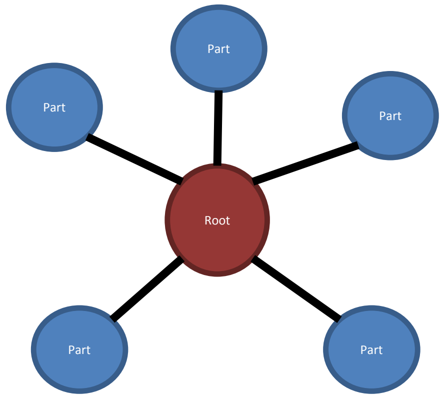
Bouchard & Triggs '05

Carneiro & Lowe '06

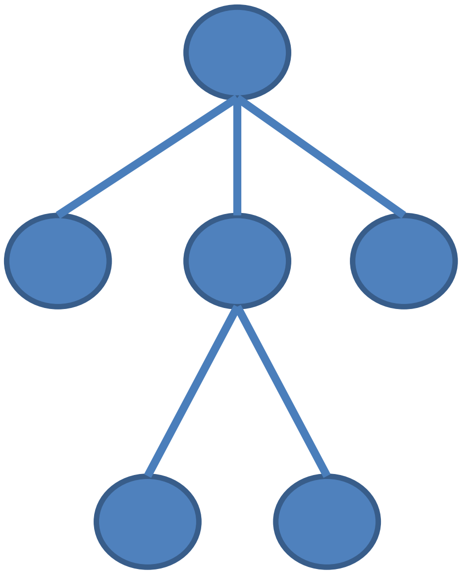
from [Carneiro & Lowe, ECCV'06]

Today's class

1. Star-shaped model
   - Example: Deformable Parts Model
     - [Felzenswalb et al. 2010]



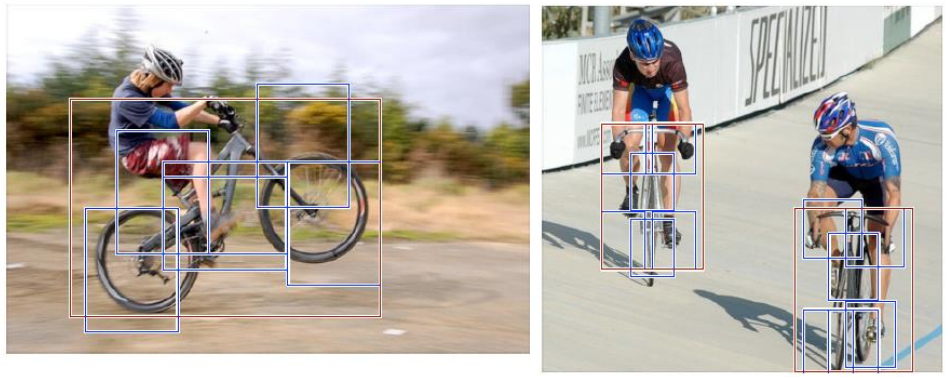
2. Tree-shaped model
   - Example: Pictorial structures
     - [Felzenszwalb Huttenlocher 2005]



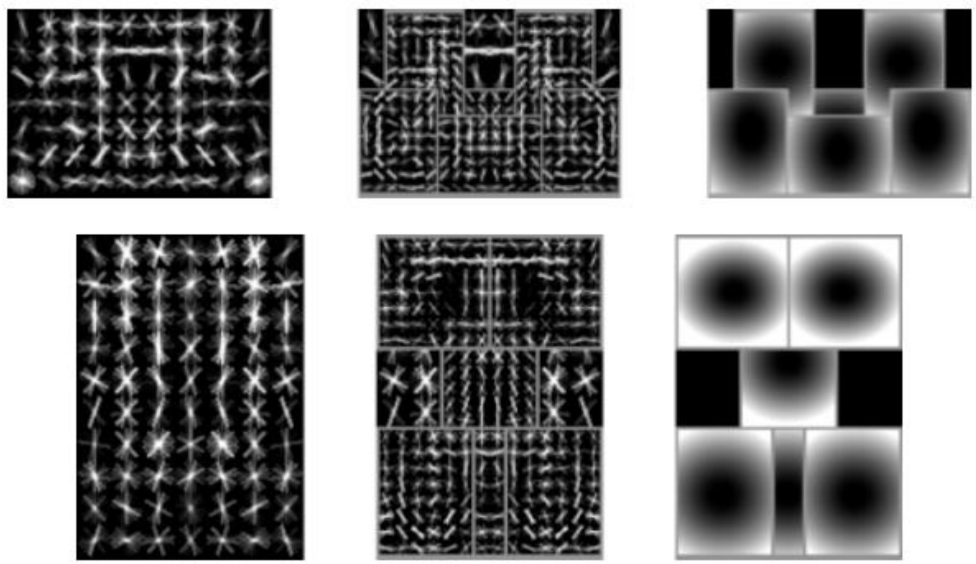
3. Sequential prediction models

Deformable Latent Parts Model (DPM)

Detections



Template Visualization



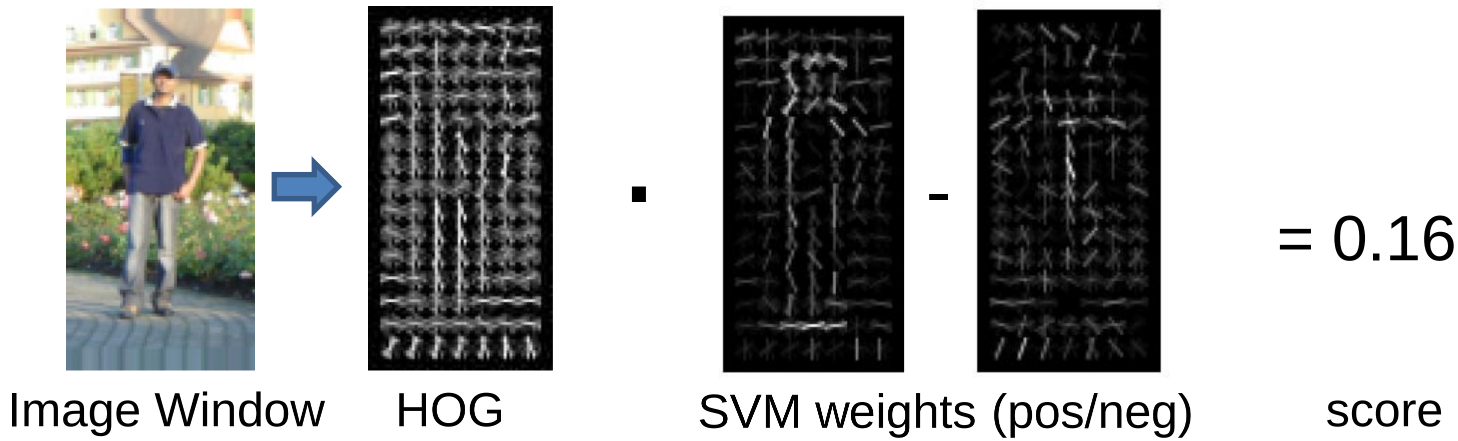
root filters
coarse resolution

part filters
finer resolution

deformation
models

Felzenszwalb et al. 2008, 2010

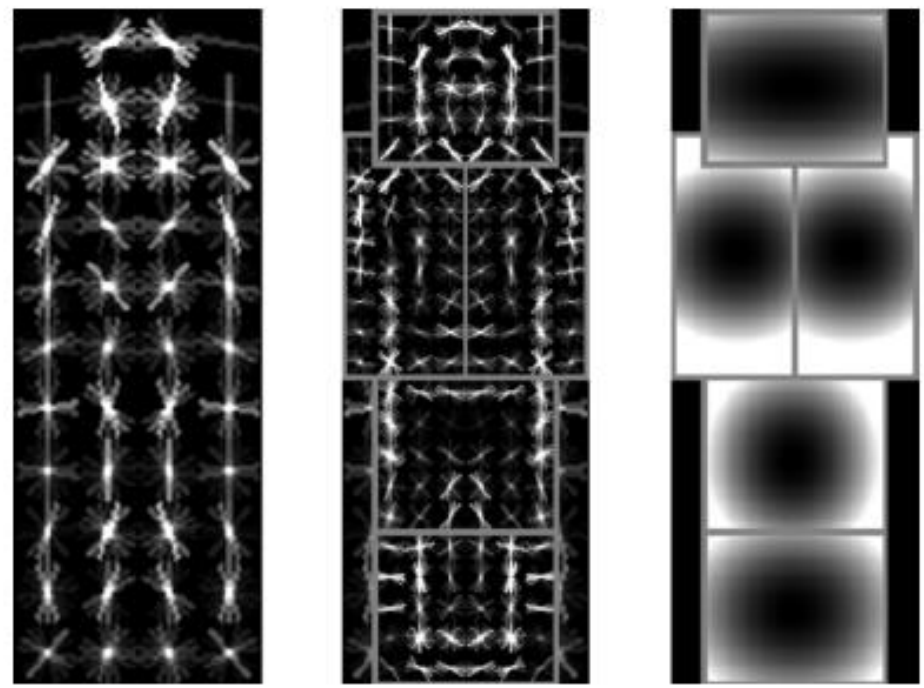
Review: Dalal-Triggs detector



1. Extract fixed-sized (64x128 pixel) window at each position and scale
2. Compute HOG (histogram of gradient) features within each window
3. Score the window with a linear SVM classifier
4. Perform non-maxima suppression to remove overlapping detections with lower scores

Deformable parts model

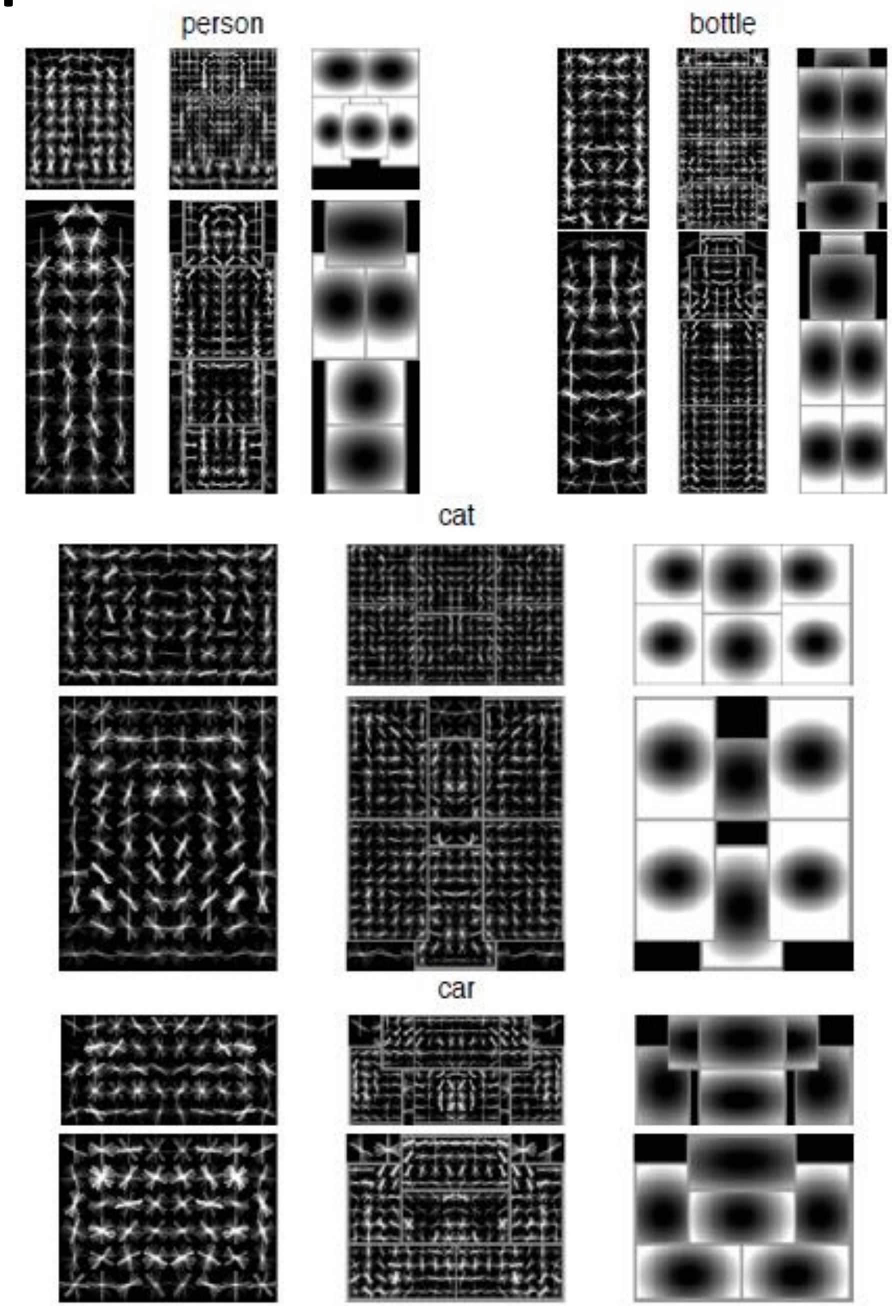
- Root filter models coarse whole-object appearance
- Part filters model finer-scale appearance of smaller patches
- For each root window, part positions that maximize appearance score minus spatial cost are found
- Total score is sum of scores of each filter and spatial costs



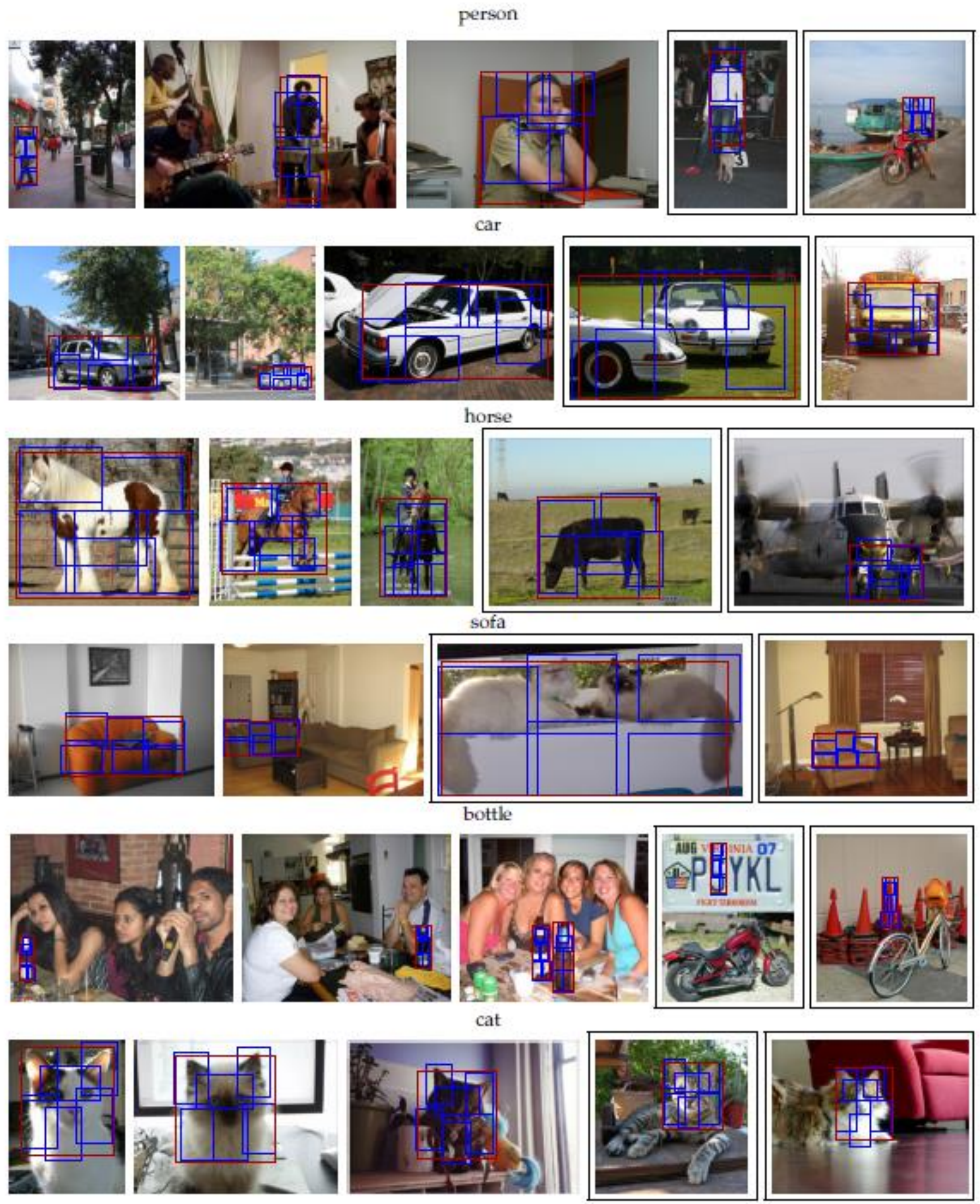
Root filter　　Part filters　　Spatial costs

DPM: mixture model

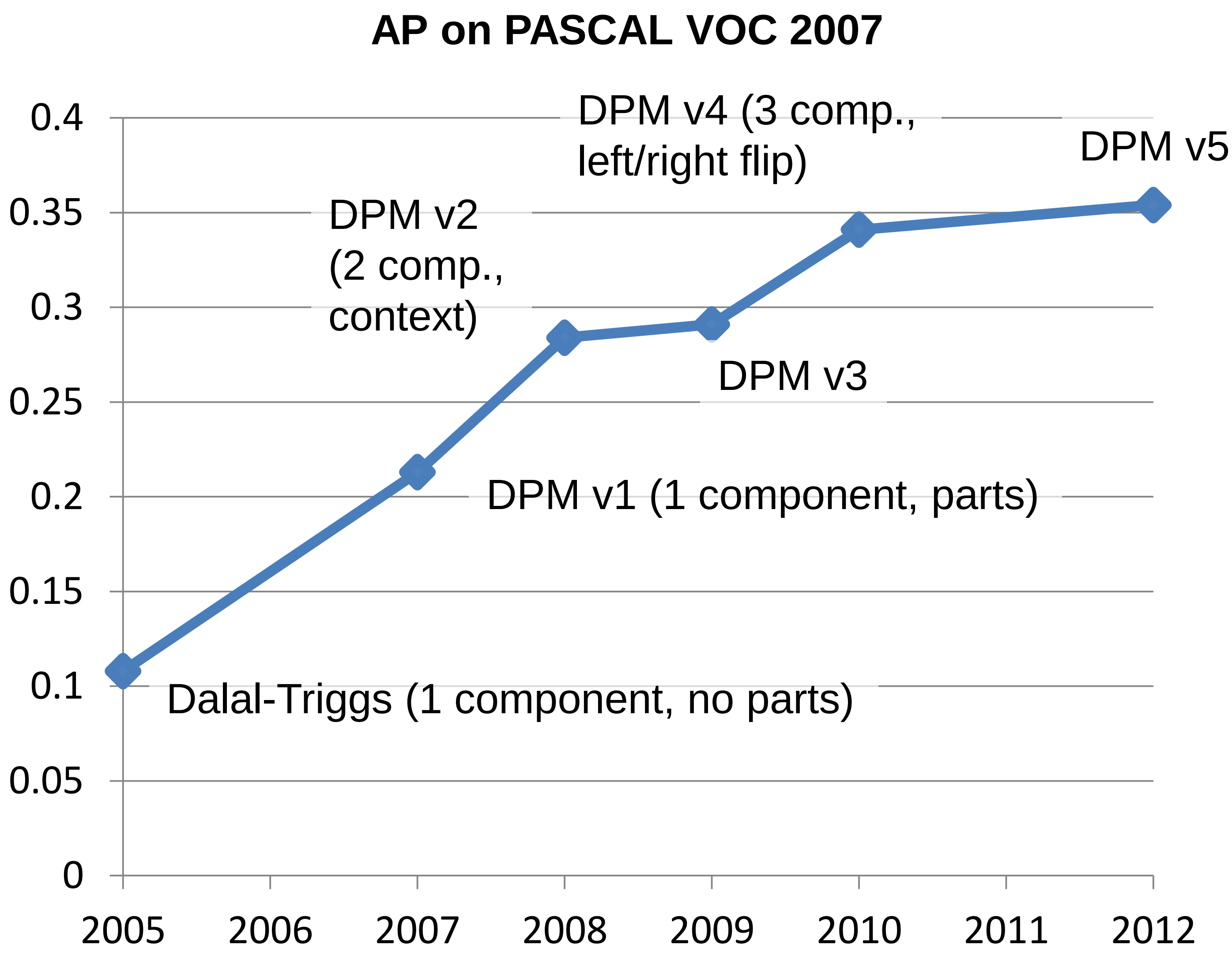
- Each positive example is modeled by one of M detectors
- In testing, all detectors are applied with non-max suppression



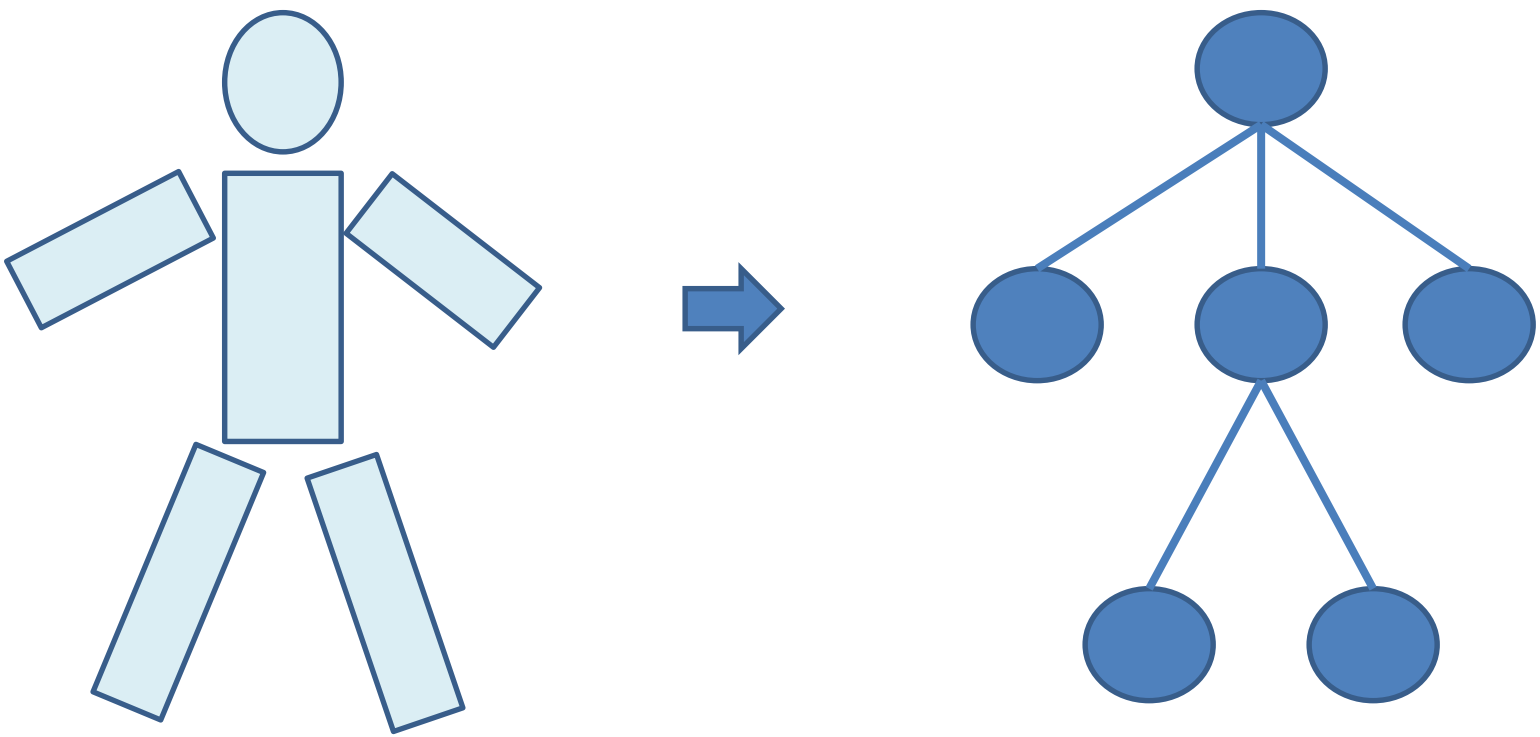
Results



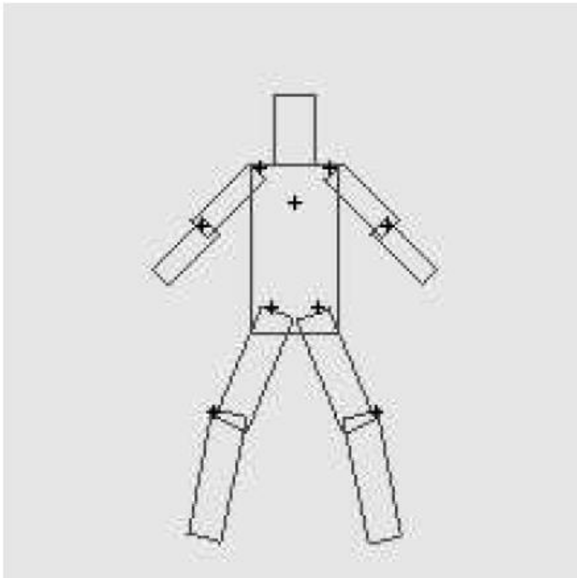
Improvement over time for HOG-based detectors

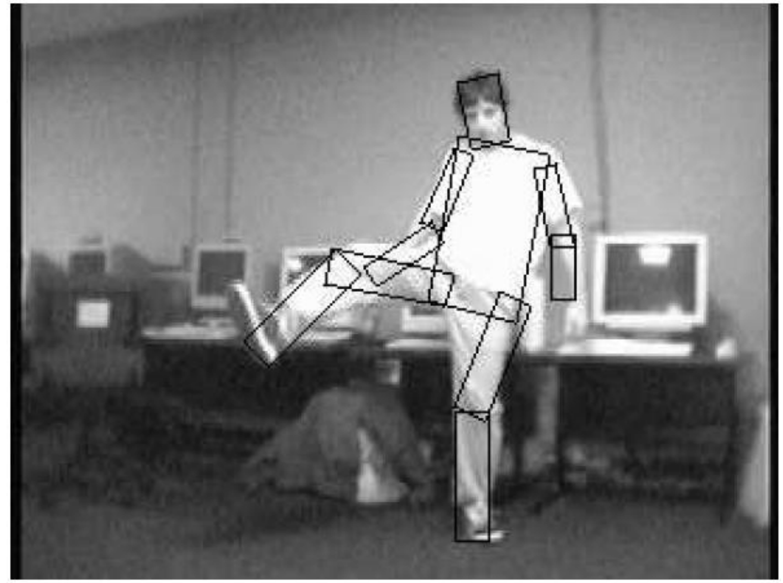


Tree-shaped model



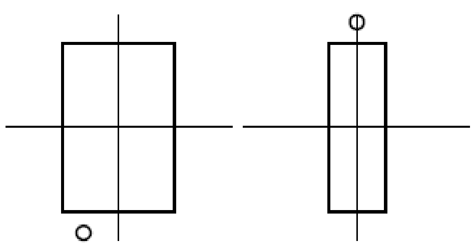
Pictorial Structures Model



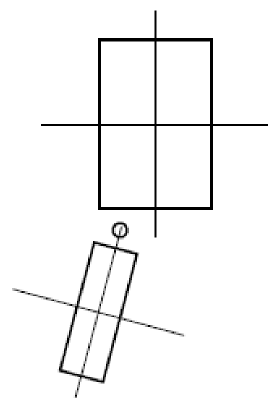


Part = oriented rectangle

Spatial model = relative size/orientation



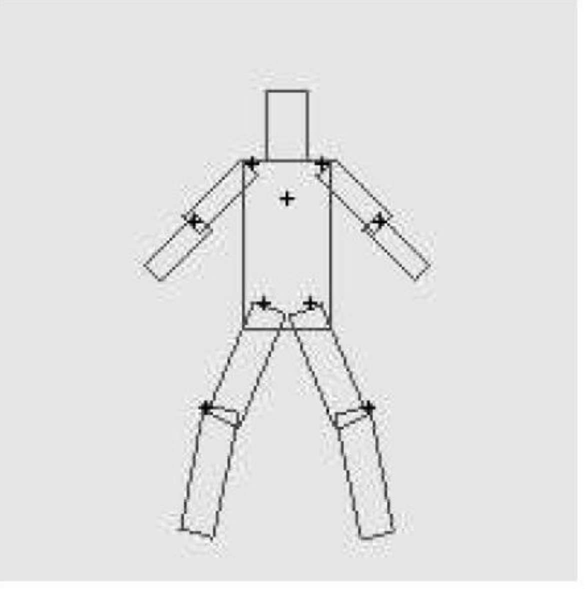
a



b

Felzenszwalb and Huttenlocher 2005

Pictorial Structures Model



$$P(L|I,\theta) \propto \left(\prod_{i=1}^{n} p(I|l_i,u_i) \prod_{(v_i,v_j)\in E} p(l_i,l_j|c_{ij}) \right)$$

Appearance likelihood

Geometry likelihood

Modeling the Appearance

- Any appearance model could be used
  - HOG Templates, etc.
  - Here: rectangles fit to background subtracted binary map

- Can train appearance models independently (easy, not as good) or jointly (more complicated but better)

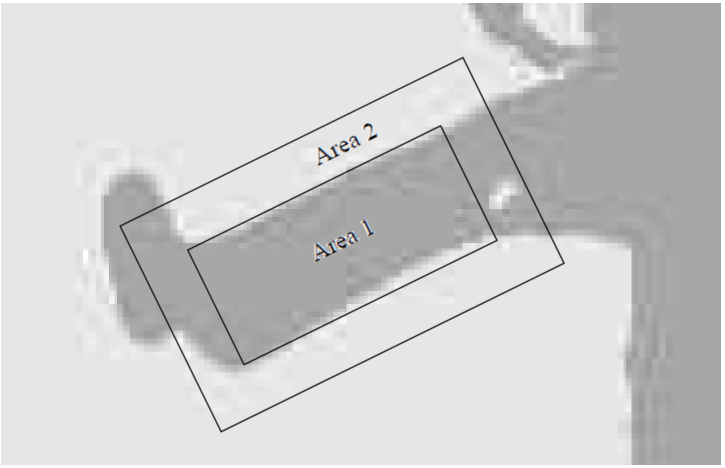
$$P(L|I,\theta) \propto \left(\prod_{i=1}^{n} p(I|l_i,u_i) \prod_{(v_i,v_j)\in E} p(l_i,l_j|c_{ij}) \right)$$

Appearance likelihood

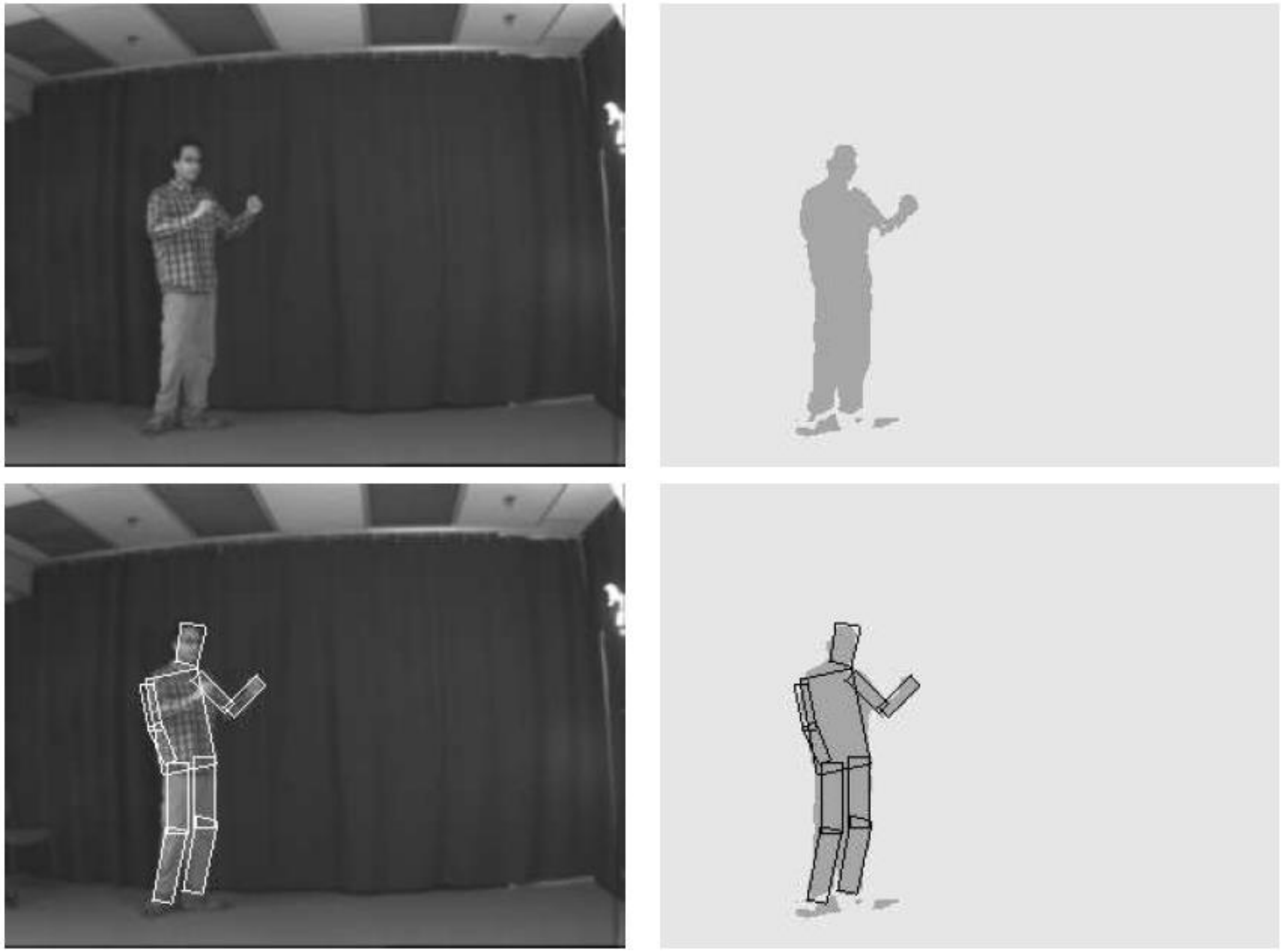
Geometry likelihood

Part representation

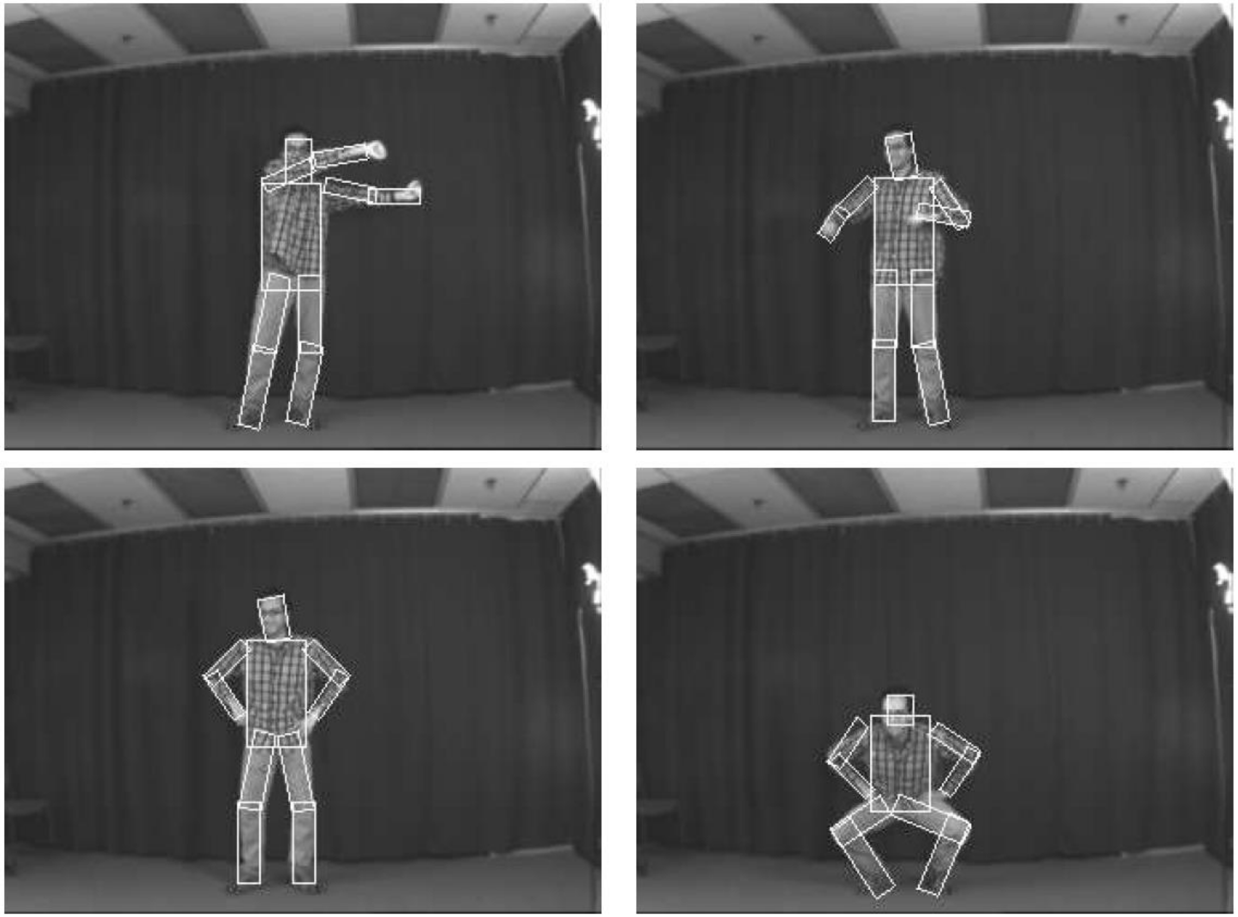
- Background subtraction



Results for person matching



Results for person matching



Enhanced pictorial structures

- Learn spatial prior
- Color models from soft segmentation (initialized by location priors of each part)

BMVC 2009

2 minute break

Which patch corresponds to a body part?

Example from Ramakrishna

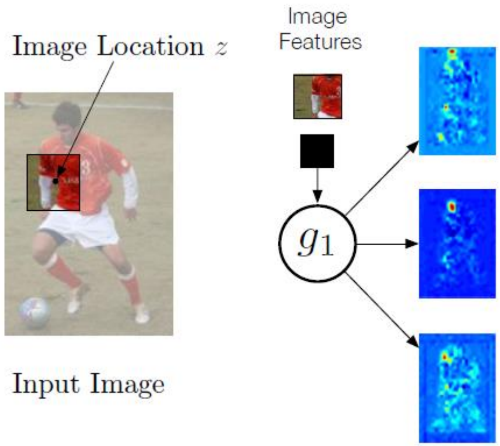
Sequential structured prediction

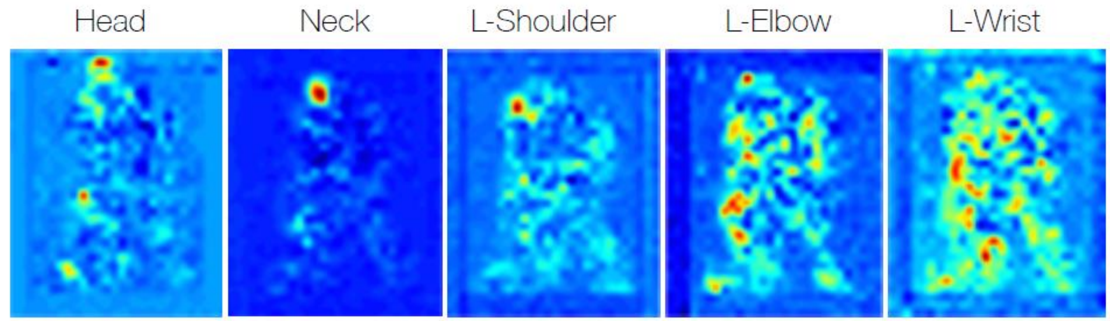
- Can consider pose estimation as predicting a set of related variables (called structured prediction)
  - Some parts easy to find (head), some are hard (wrists)

- One solution: jointly solve for most likely variables (DPM, pictorial structures)

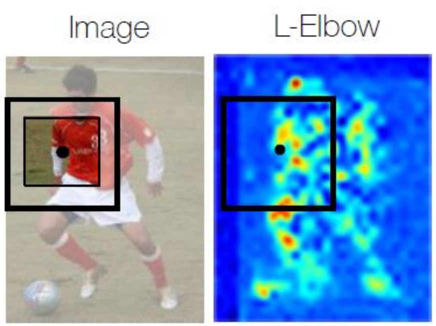
- Another solution: iteratively predict each variable based in part on previous predictions

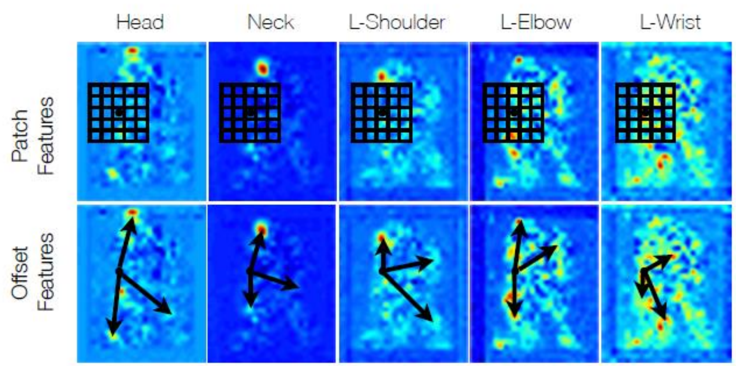
Pose machines



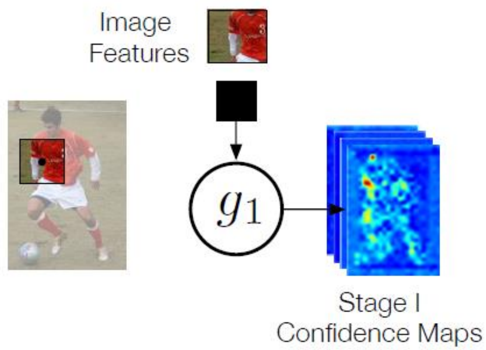


Local image evidence is weak
Certain parts are easier to detect than others

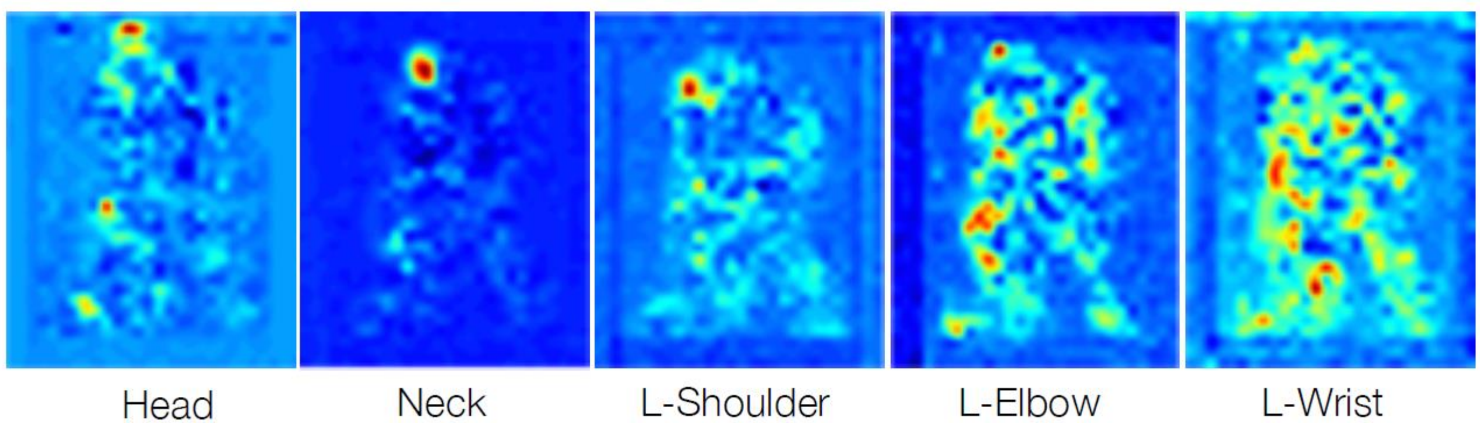


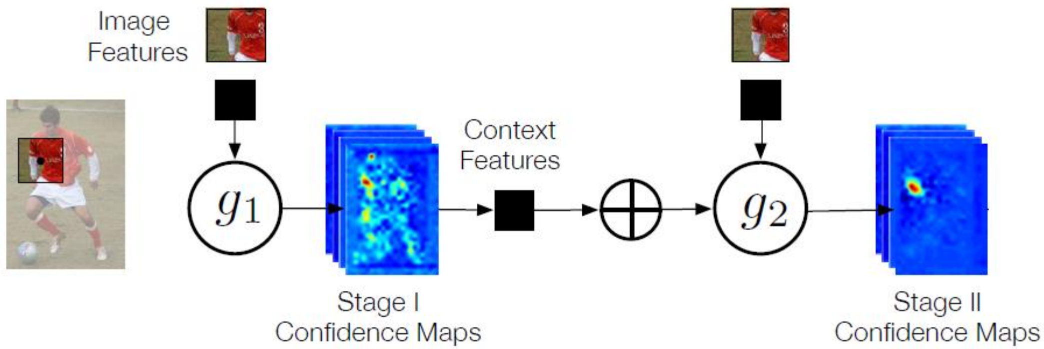


Ramakrishna et al. ECCV 2014

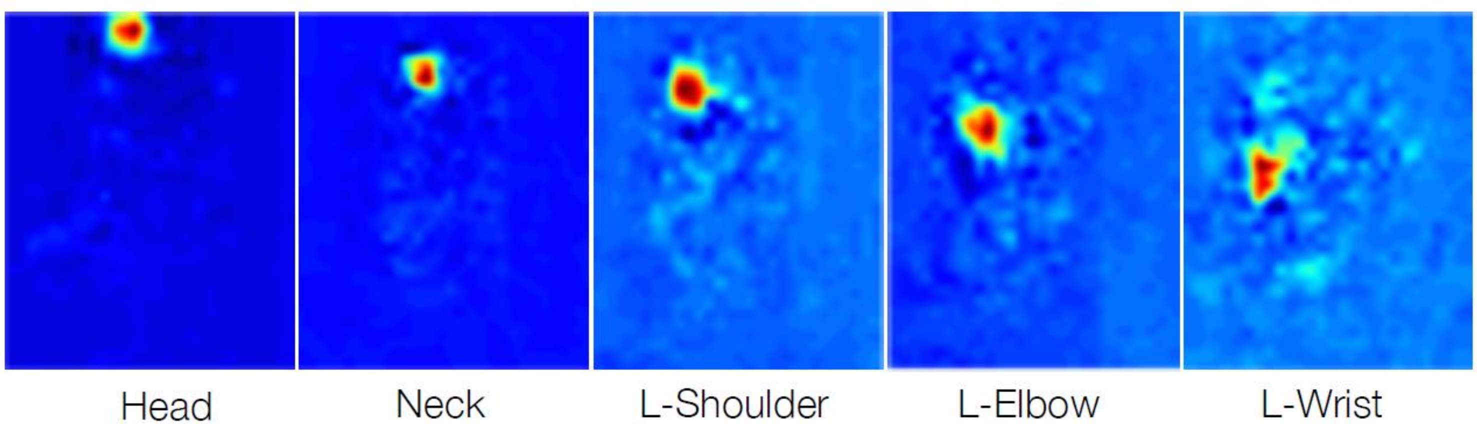


Stage I Confidence





Stage II Confidence



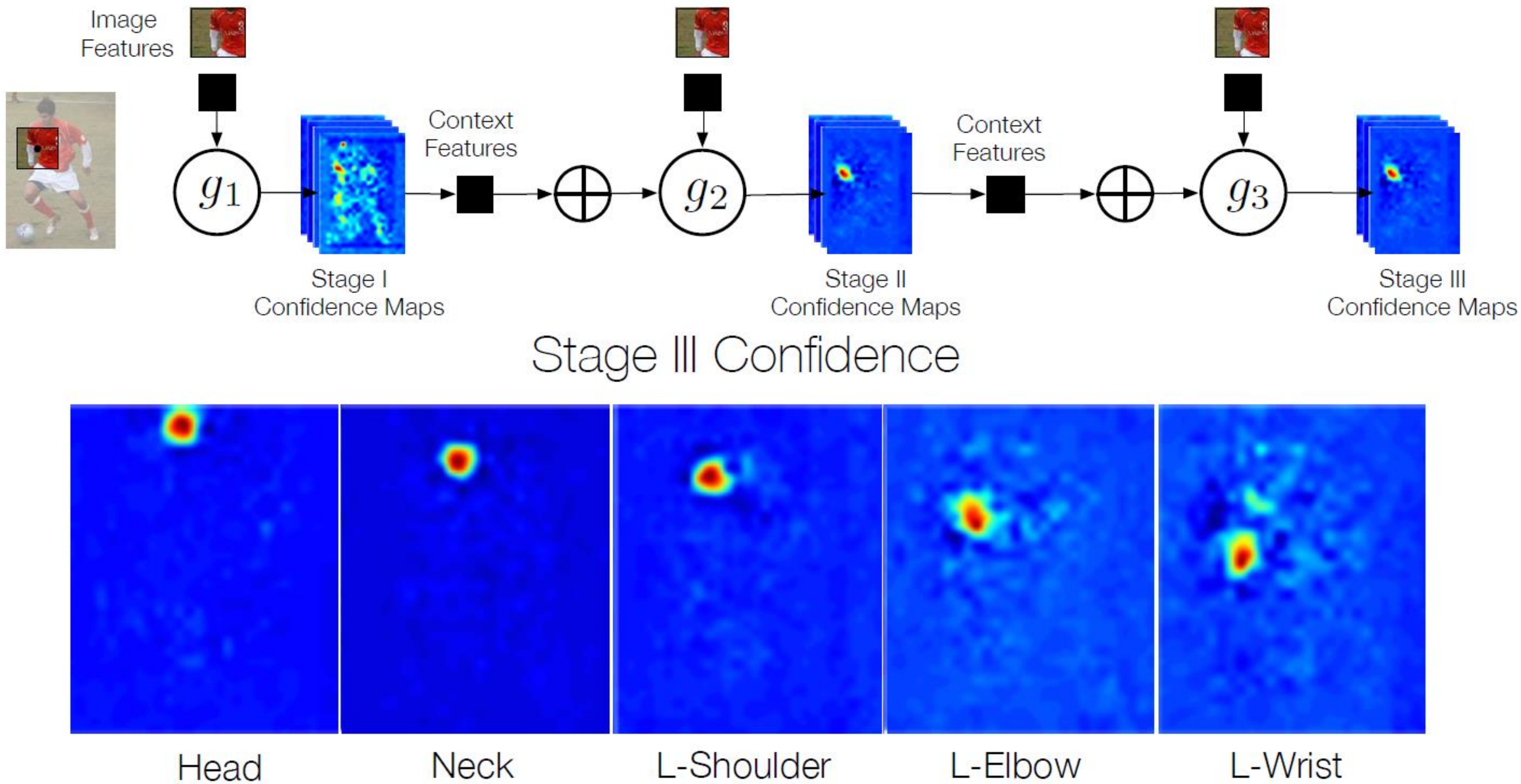


Image
Features
Context
Features
Context
Features
g_1
g_2
g_3
Stage I
Confidence Maps
Stage II
Confidence Maps
Stage III
Confidence Maps
Stage III Confidence
Head
Neck
L-Shoulder
L-Elbow
L-Wrist

Example results

Graphical models vs. structured prediction

- Advantages of sequential prediction
  - Simple procedures for training and inference
  - Learns how much to rely on each prediction
  - Can model very complex relations

- Advantages of BP/graphcut/etc
  - Elegant
  - Relations are explicitly modeled
  - Exact inference in some cases