Tutorial outline

- Overview (this)
- Image representation (60 mins, 9:15 - 10:30)
  - motivation, local features, global features, break
- Learning (90 mins, 10:30 - 12:30)
  - discriminative models, tea-break, generative models, break
- Object detection and recognition (90 mins, 12:30 - 2:00)
  - Dalal & Triggs, lunch-break, PASCAL challenge, *poselets* and their applications, tea-break
- Cross-modal search (60 mins, 2:30 - 3:30)

lunch-break 60 mins, break 15 mins, tea-break 20-30 mins

ICVGIP 2012, IIT Bombay

The eight Indian Conference on Computer Vision, Graphics and Image Processing

Tutorial

Object detection and recognition

Subhransu Maji
Toyota Technological Institute at Chicago

Introduction

auto-focus based on face detection

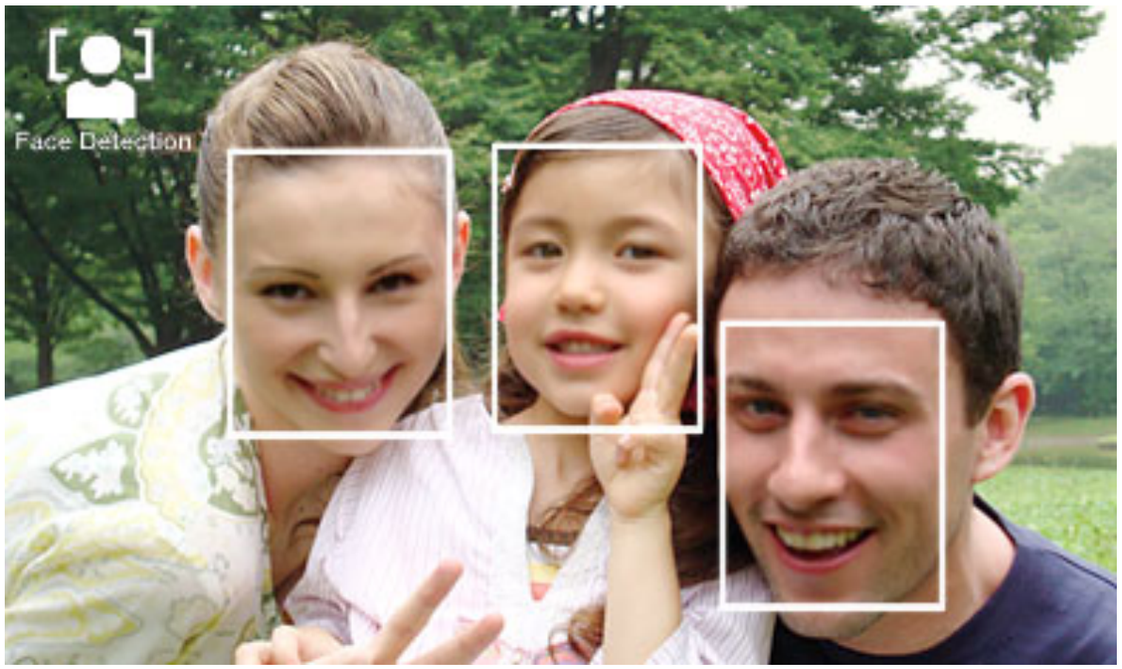


image credit : sony.co.in

pedestrian collision warning

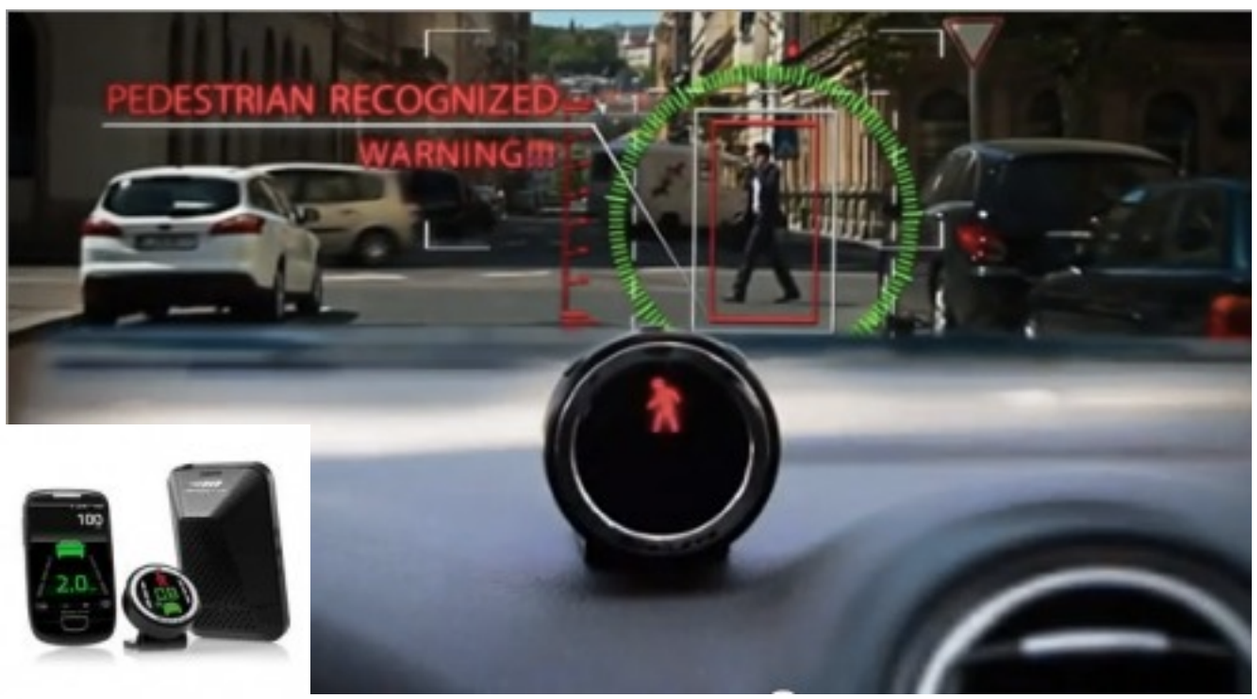


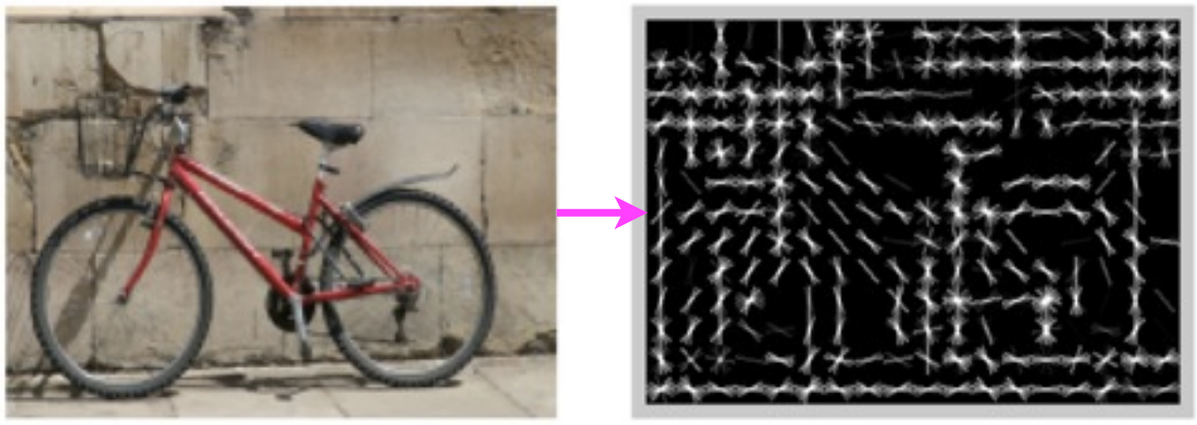
image credit : mobile eye

Outline

- Overview of Dalal and Triggs pedestrian detector
  - Histogram of Oriented Gradients (HOG) features
  - Training pipeline for detection
- PASCAL VOC challenge
- Overview of the Poselet-based detector
  - What is a poselet?
  - Training and selecting a library of poselets
  - Using poselets for detection and beyond

Histograms of Oriented Gradients (HOG)

- Introduce invariance
  - Bias / gain / nonlinear transformations
    - bias: gradients / gain: local normalization
    - nonlinearity: clamping magnitude, orientations
  - Small deformations
    - spatial subsampling
    - local "bag" models

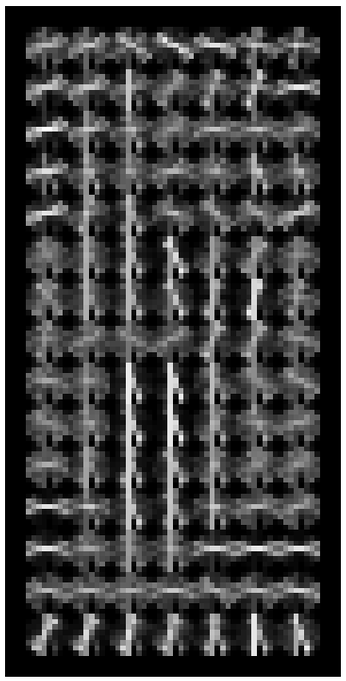


- References
  - "Histograms of oriented gradients for human detection." N. Dalal and B. Triggs, CVPR 2005.
  - "Finding people in images and videos." N. Dalal, Ph.D. Thesis, Institut National Polytechnique de Grenoble, 2006.

Classification training and testing

Pos = {…  …}

Cropped positive



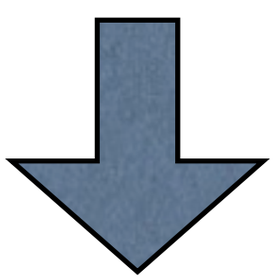
HOG

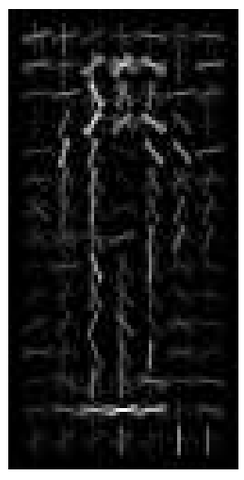
[Dalal06]

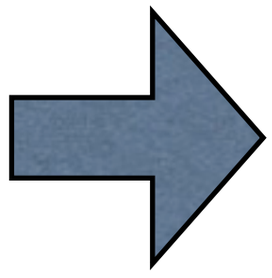
Classification training and testing

Pos = {…  …}

Neg = {… random background patches …}



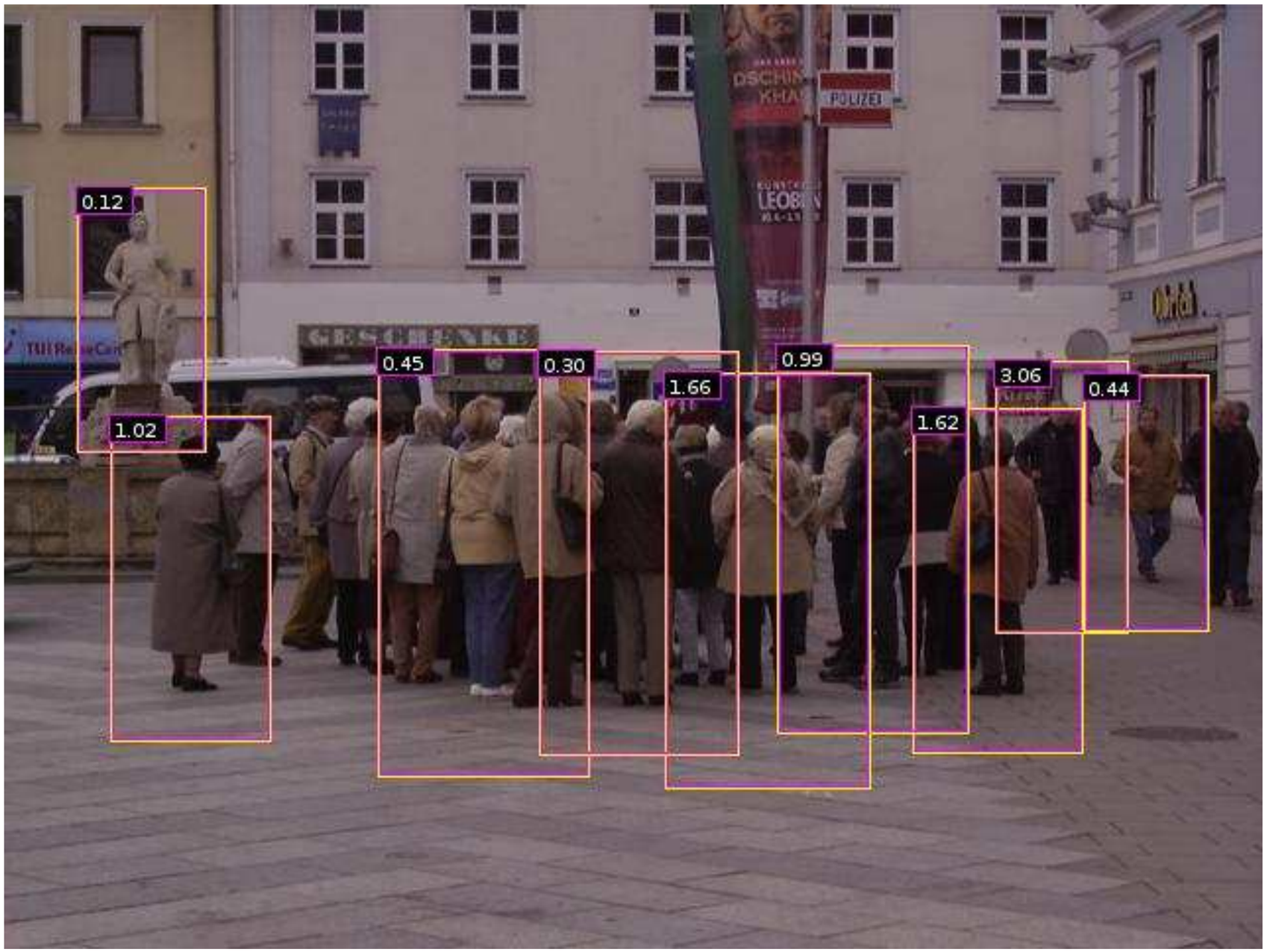




Test on cropped windows

SVM

Classification versus detection



[Dalal06]

Sliding window detection

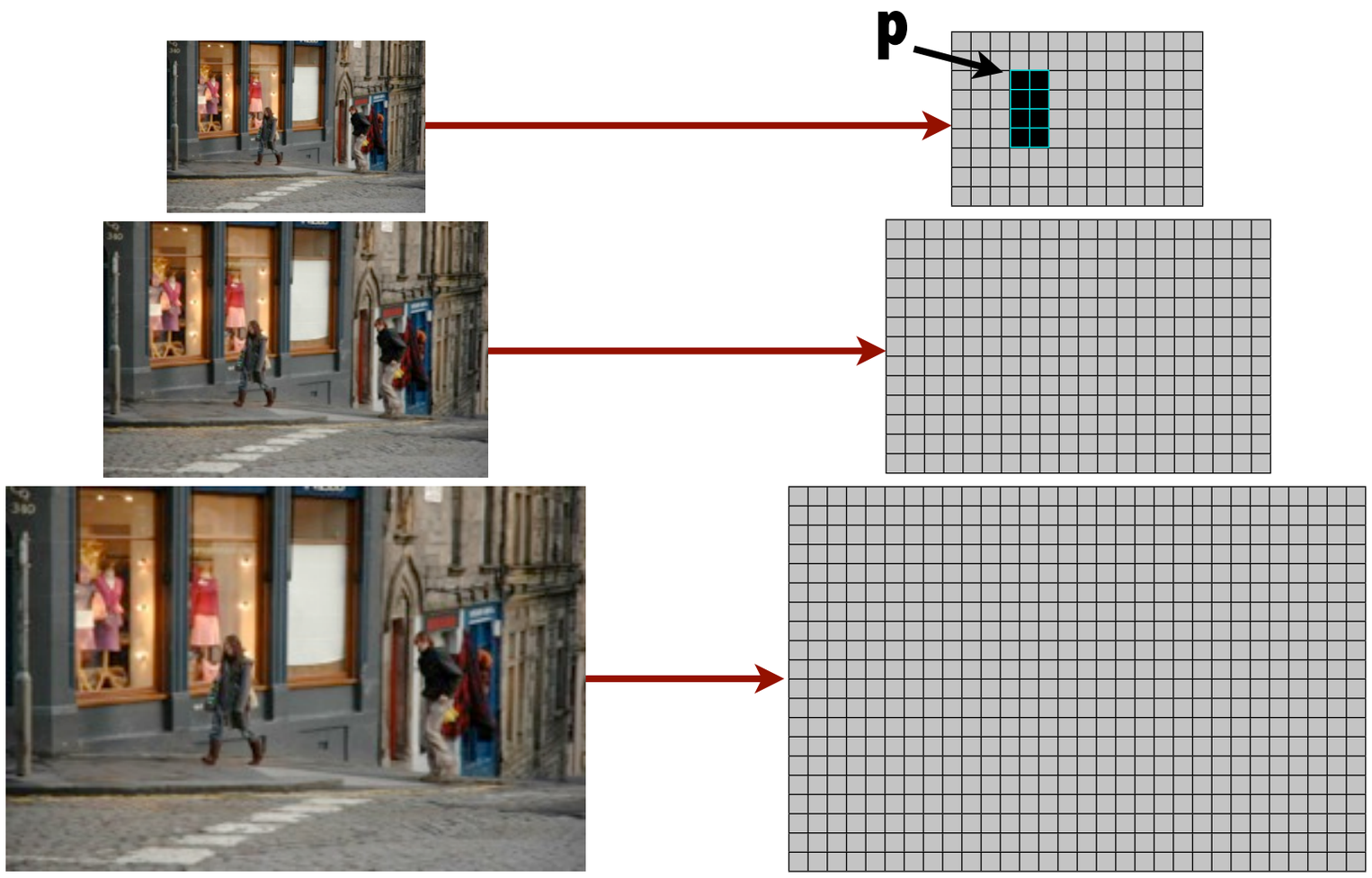
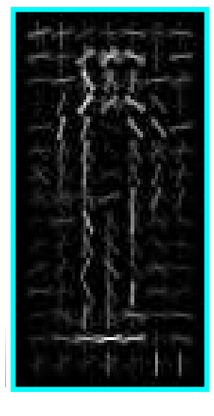


Image pyramid

HOG feature pyramid

$$score(I, p) = \mathbf{w} \cdot \phi(I, p)$$

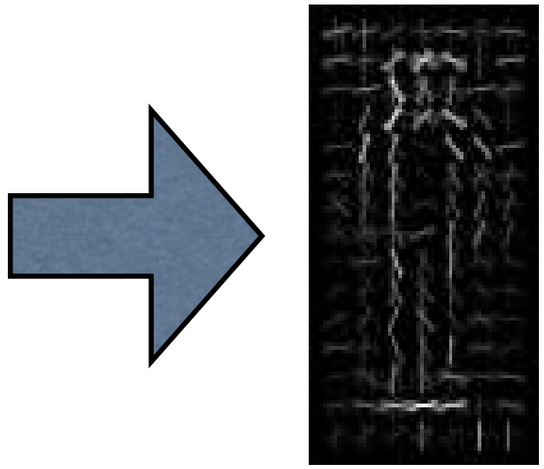


- Compute HOG of the whole image at multiple resolutions
- Score each subwindows of the feature pyramid

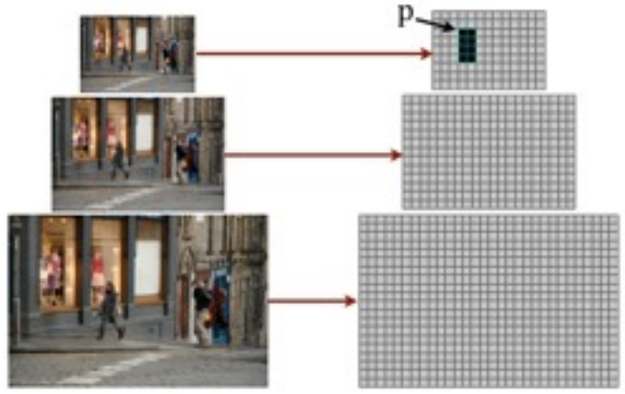
Mining hard negatives ("bootstrapping")

Pos = {…  …}

Neg_{rand} = {... random background patches ...}

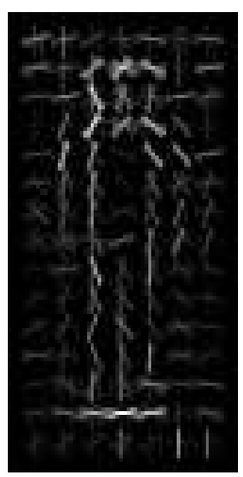


SVM



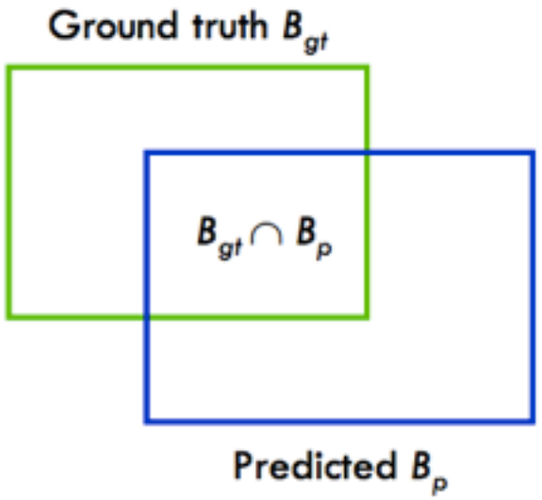
"Hard" negatives

+ Neg_{hard} = {... windows with score >= -1 ...}

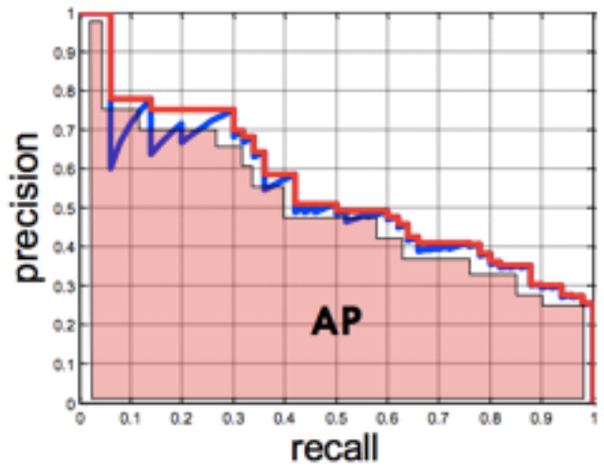


SVM

Detection evaluation

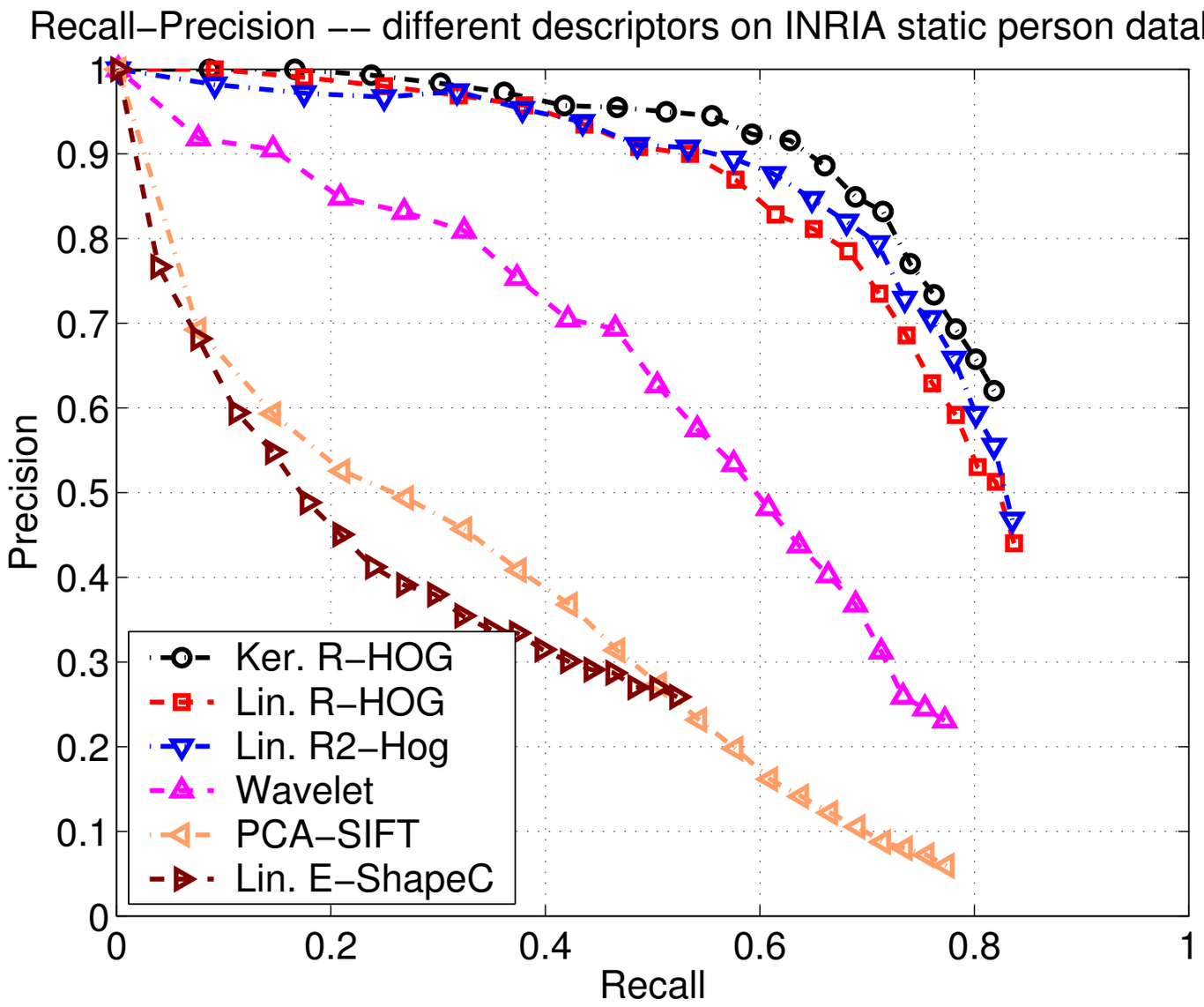


$$\text{overlap}(B_{\text{gt}}, B_{\text{p}}) = \frac{|B_{\text{gt}} \cap B_{\text{p}}|}{|B_{\text{gt}} \cup B_{\text{p}}|}$$



- Assign each prediction to
  - true positive (TP) or false positive (FP)
- $\text{Precision}_{@k}$ = $\#\text{TP}_{@k}$ / ($\#\text{TP}_{@k}$ + $\#\text{FP}_{@k}$)
- $\text{Recall}_{@k}$ = $\#\text{TP}_{@k}$ / #TotalPositives
- Average Precision (AP)

Dalal & Triggs detectors on INRIA

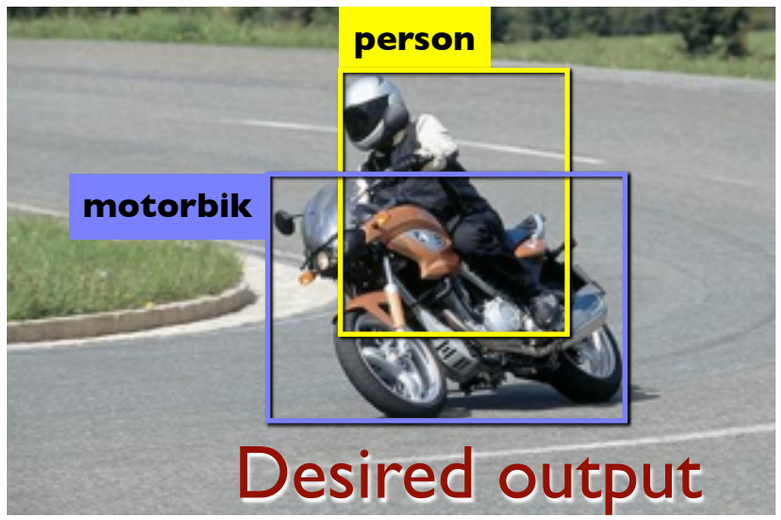


- AP = 0.75 with a linear SVM
- Very good, right?

PASCAL VOC Challenge

- Localize & name (*detect*) 20 basic-level object categories
  - Airplane, bicycle, bus, cat, car, dog, person, sheep, sofa, monitor, *etc*.


Input



Desired output

- Run from 2005 - 2012
- 11k training images with 500 to 8000 instances / category
- Substantially more challenging images
- Dalal & Triggs detector AP on ‘person’ category: 12%

PASCAL examples

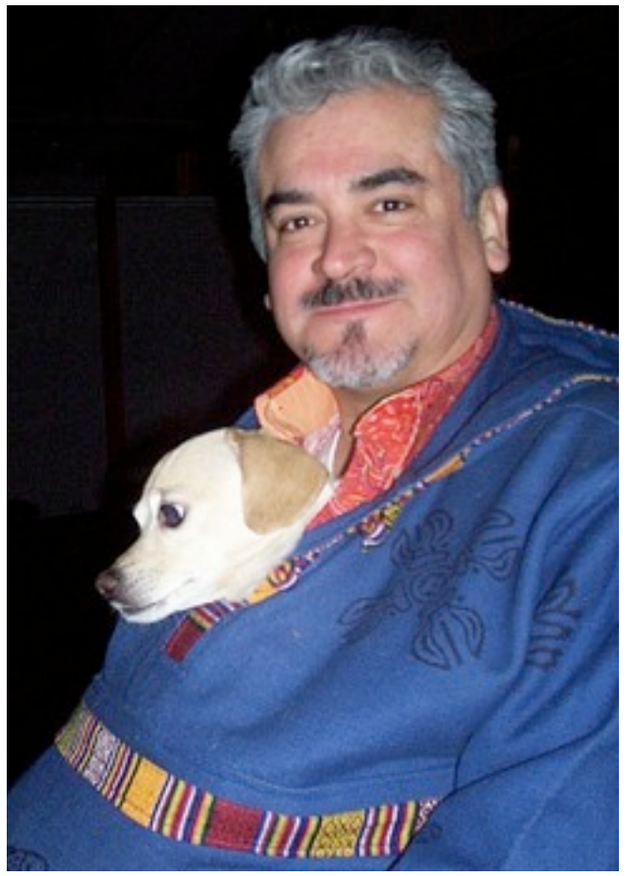


Image credits: PASCAL VOC

PASCAL examples

Image credits: PASCAL VOC

PASCAL examples

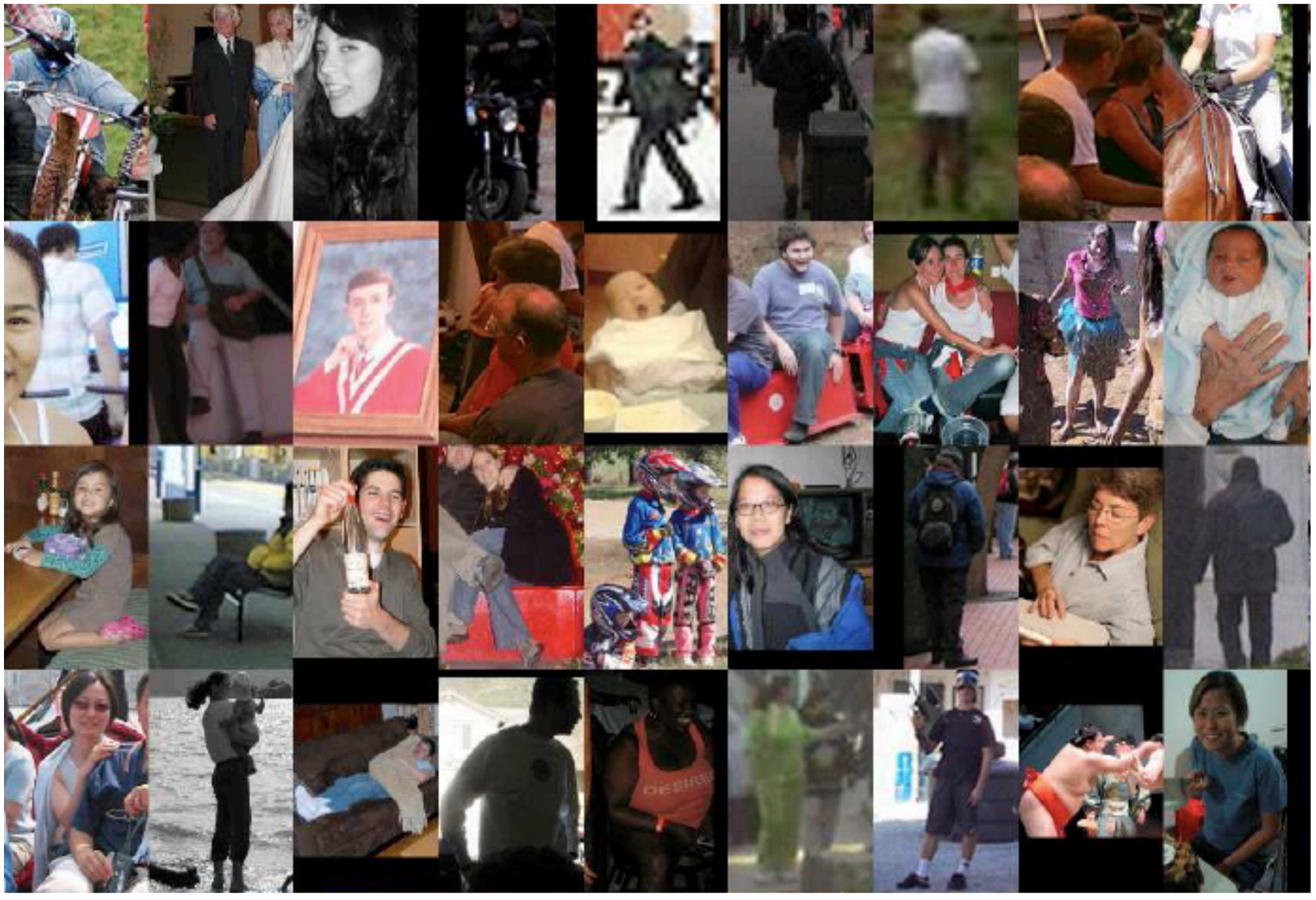
Image credits: PASCAL VOC

PASCAL examples

Image credits: PASCAL VOC

Diversity in visual apperance

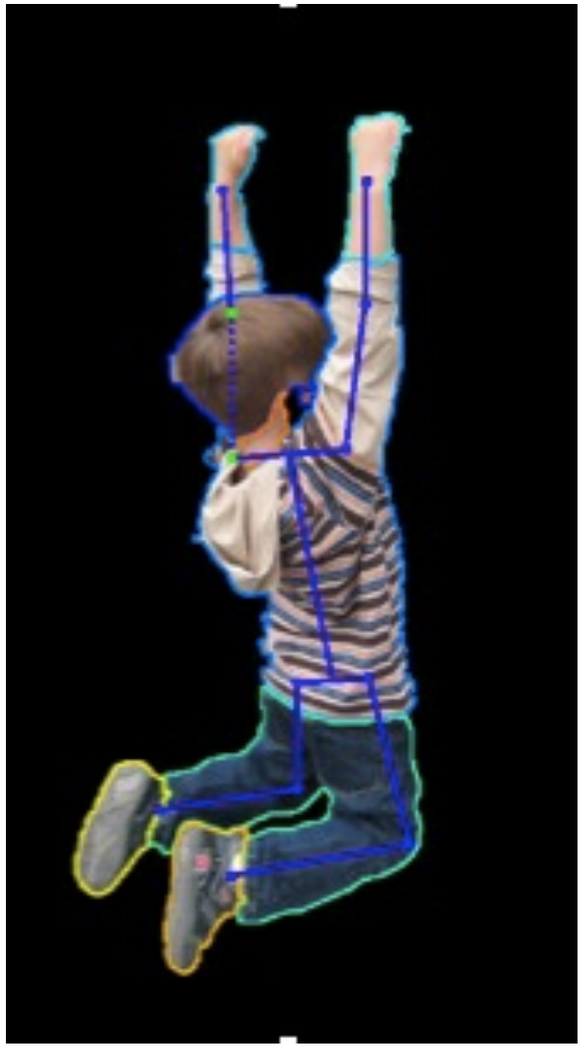
examples from the person category



many sub-categories: pose, viewpoint, occlusion

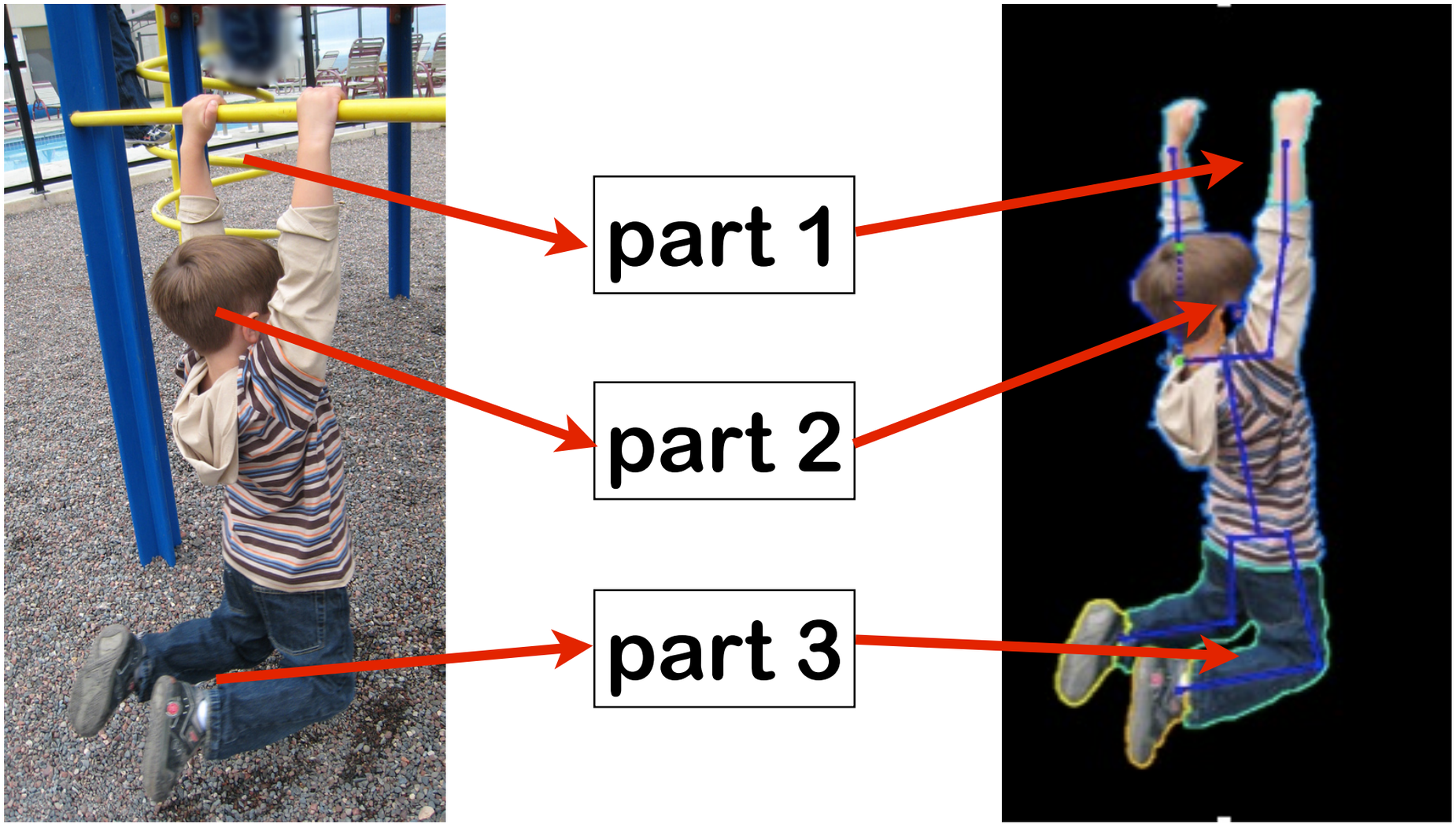
a single template is not enough (D&T AP=12%)

Beyond detection...

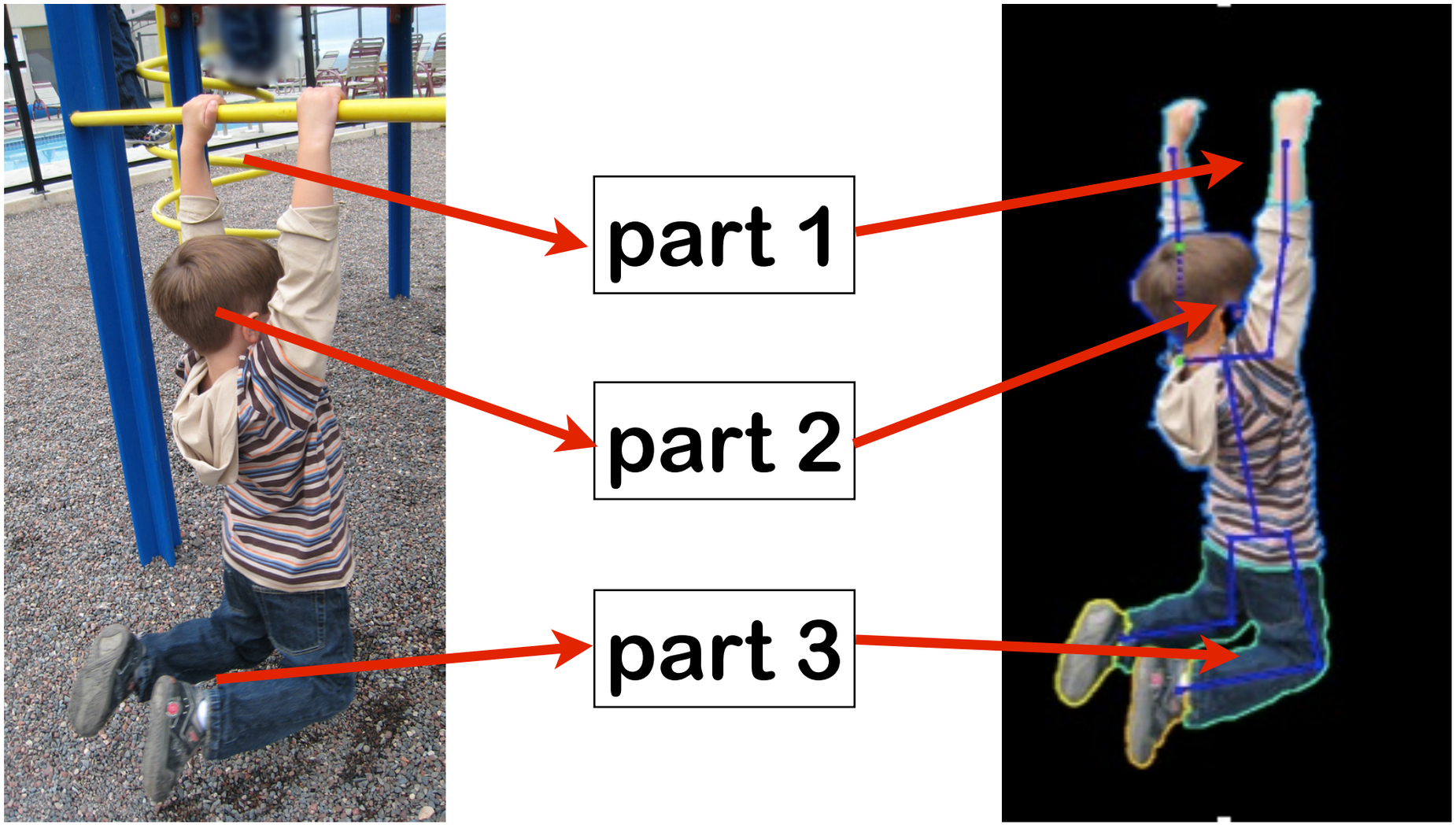


estimate pose, segmentation, gender, clothing, age, action, hair-style, etc.

Solution: part-based detectors

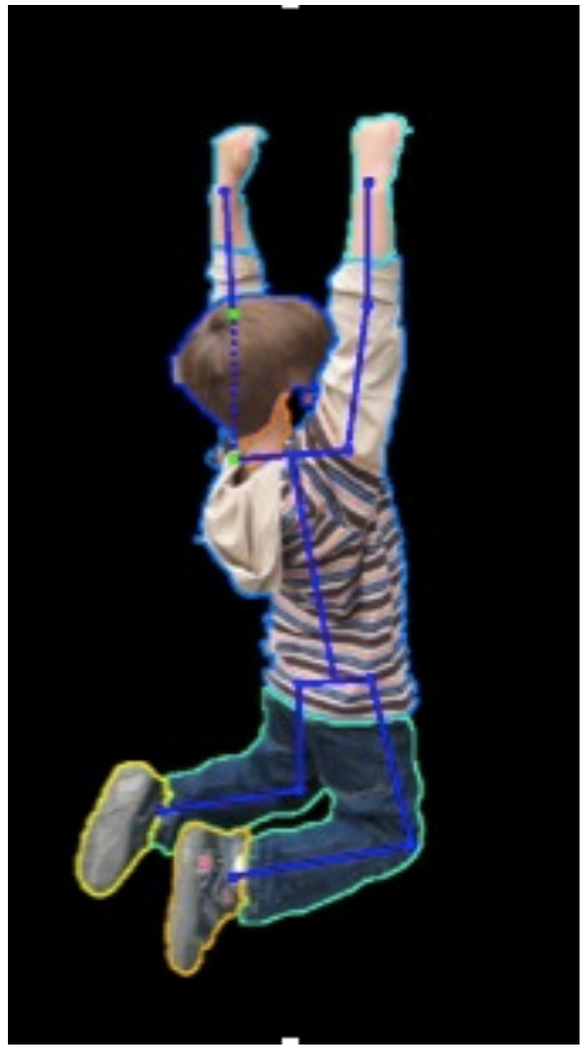


Solution: part-based detectors



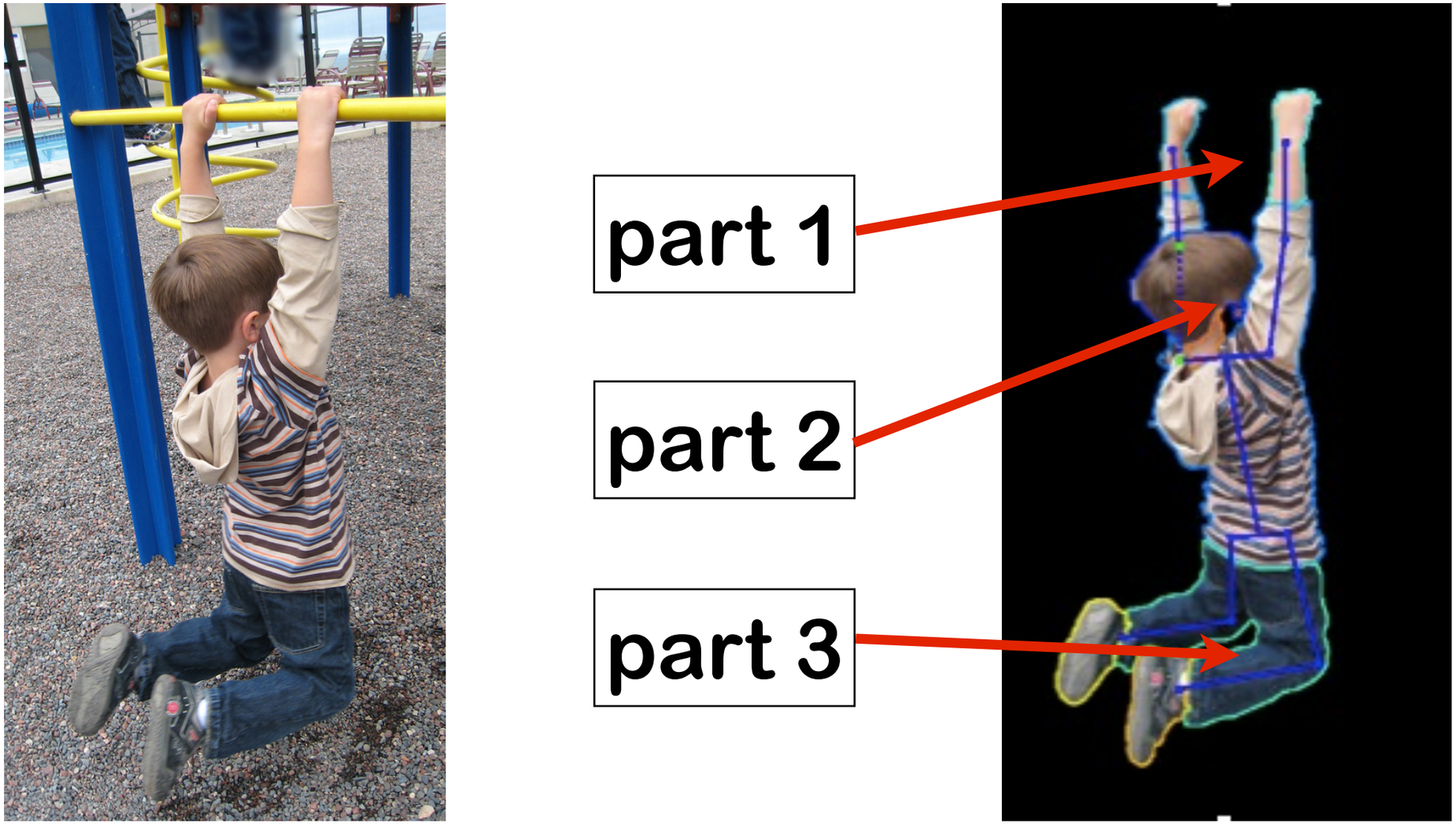
But how should we select good parts?

Properties of good parts



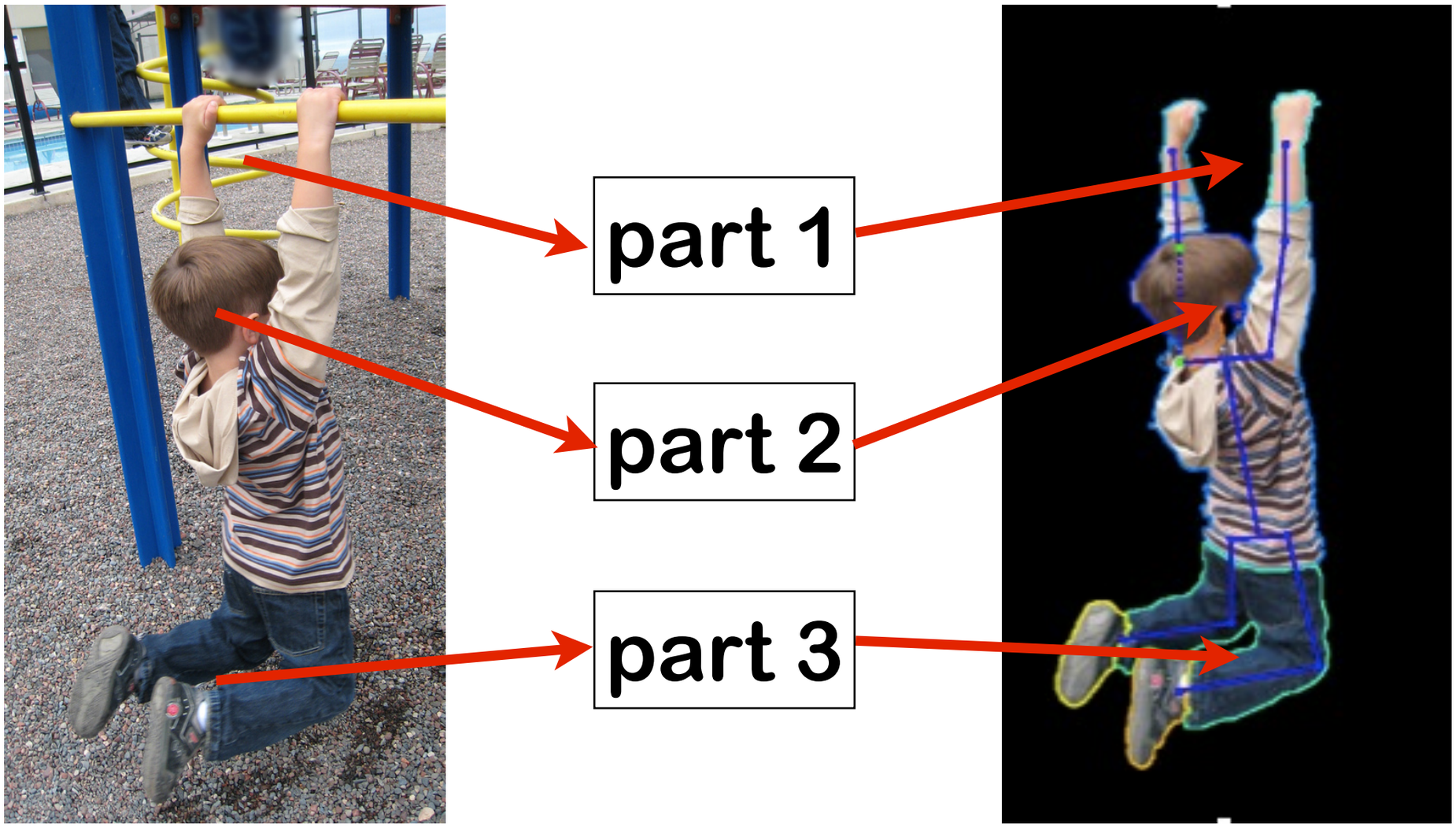
It should be easy to detect the part from the image
i.e., want discriminative parts such as frontal faces

Properties of good parts



It should be easy to predict the pose given the part
i.e., want parts tightly clustered in pose space

Properties of good parts



want parts that are (1) visually discriminative and are (2) semantically meaningful

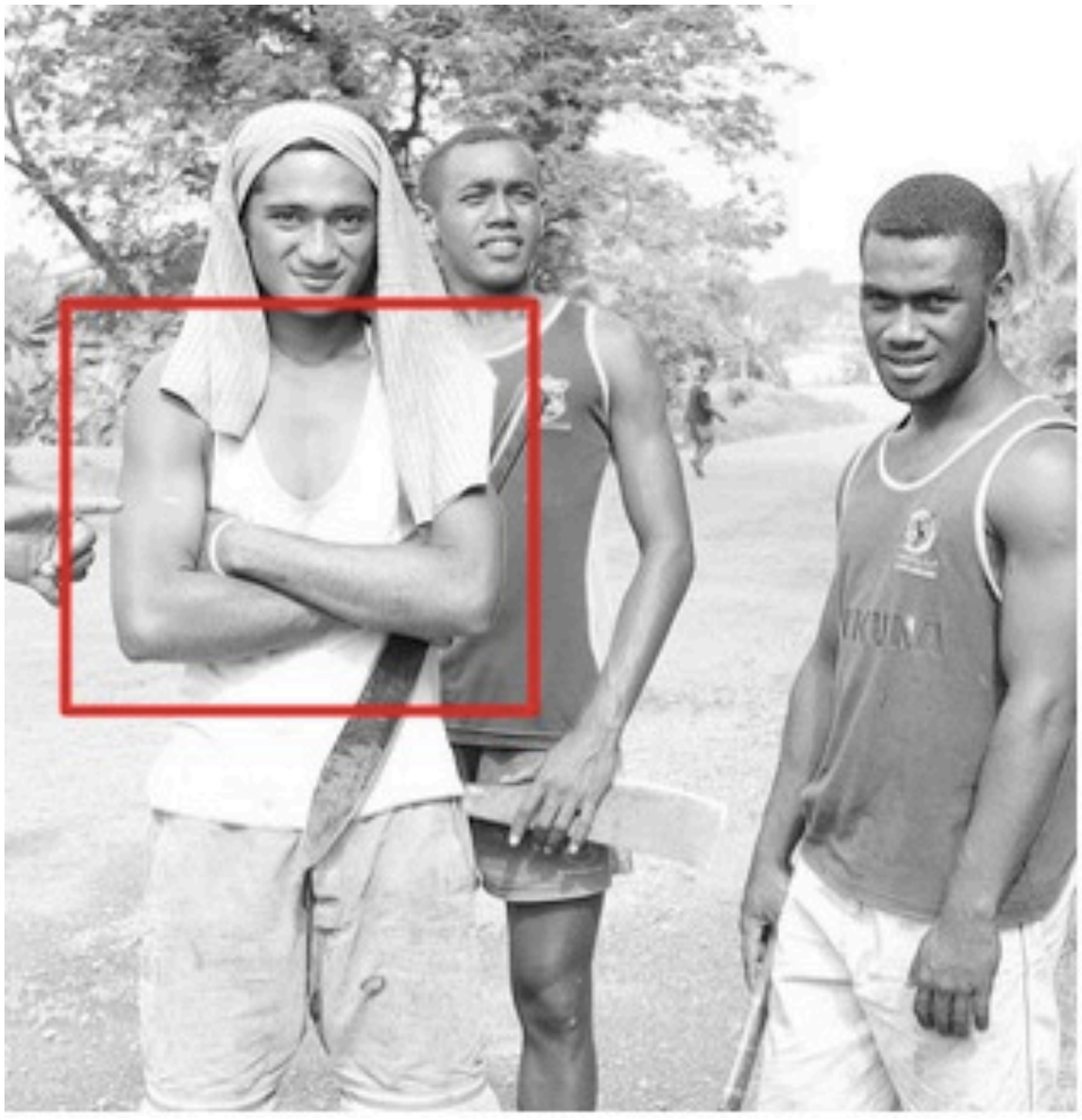
Examples of good parts

parts are often far *visually*, but they are close *semantically*

We call such parts *poselets*

Bourdev & Malik 09
Bourdev et al. 10

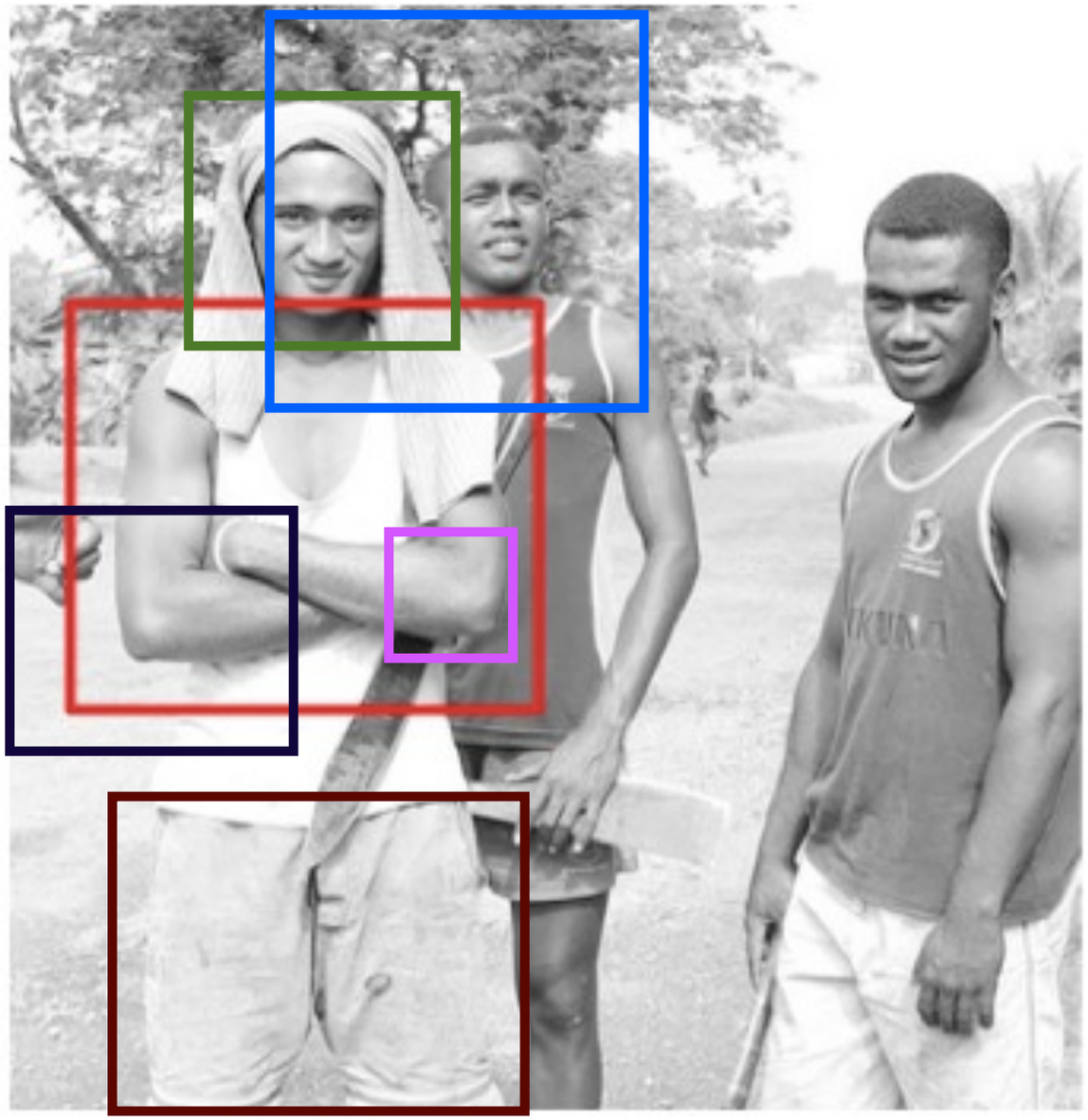
Key problem : How to find *semantically* similar patches?



Given a part of the human pose

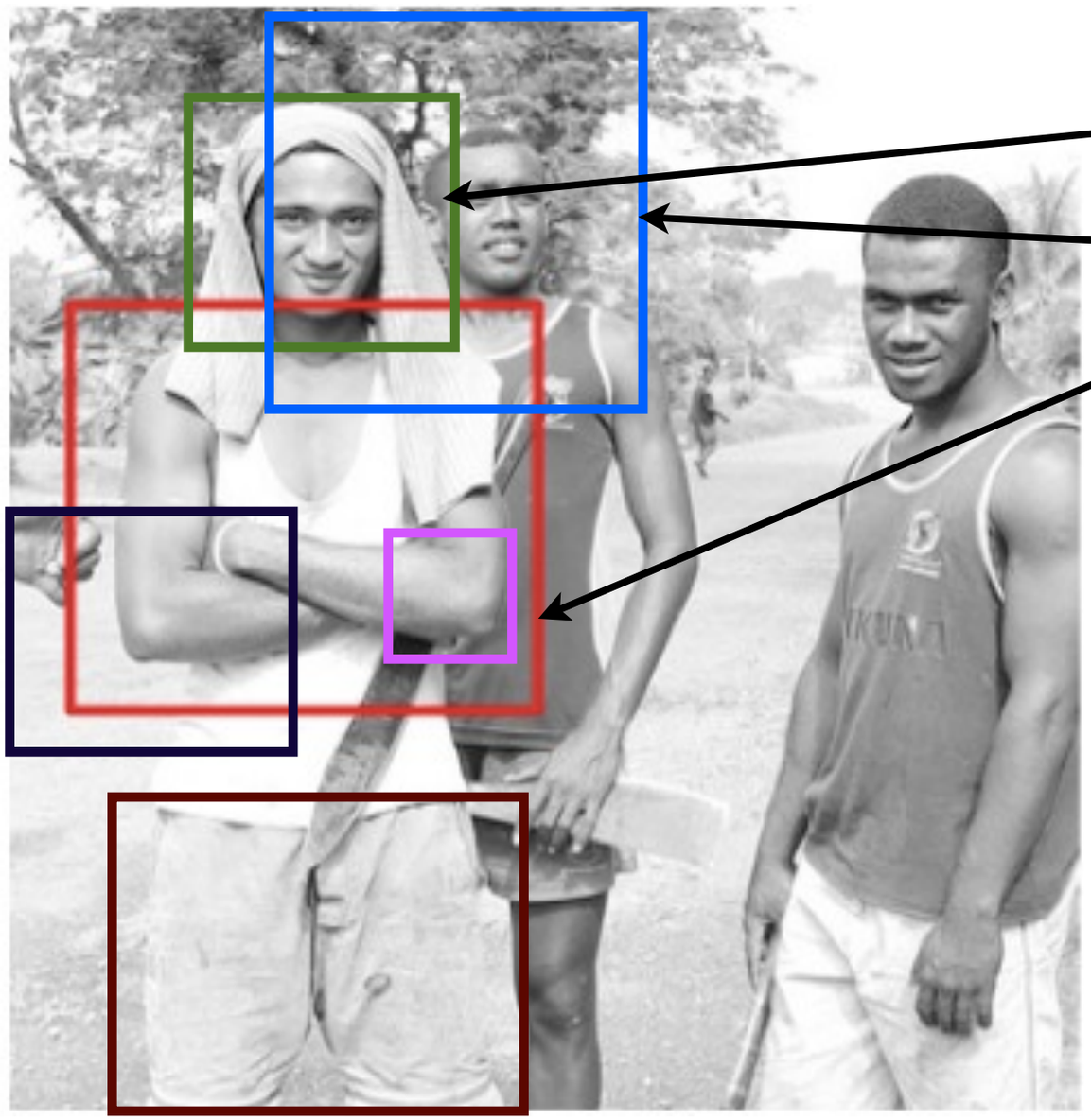
How do we find a similar pose configuration in another image?

Key Problem : what poses to consider?



Combinatorial number of poses are possible

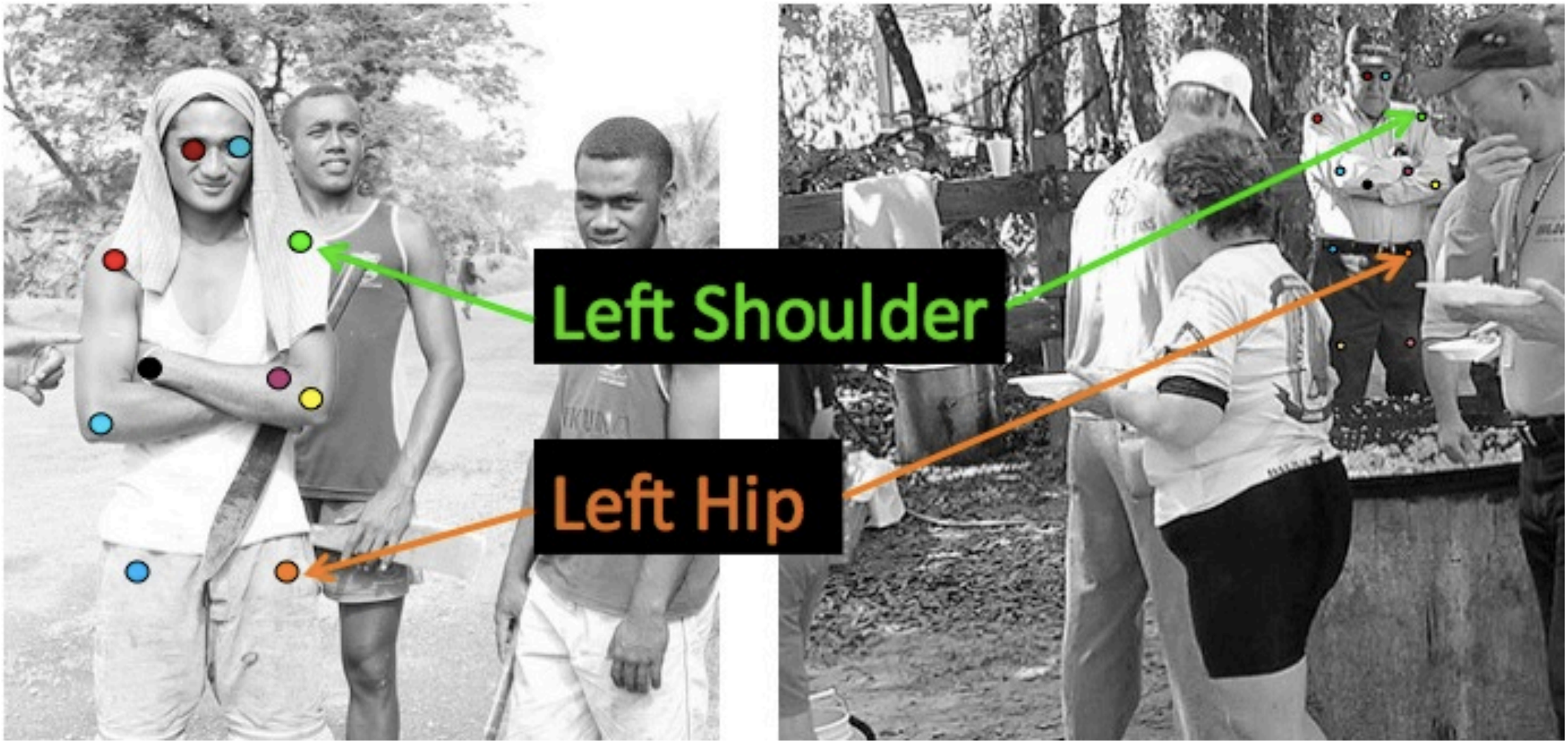
Key Problem : what poses to consider?



Can annotate a few parts such as faces, but we may miss many discriminative parts

Combinatorial number of poses are possible

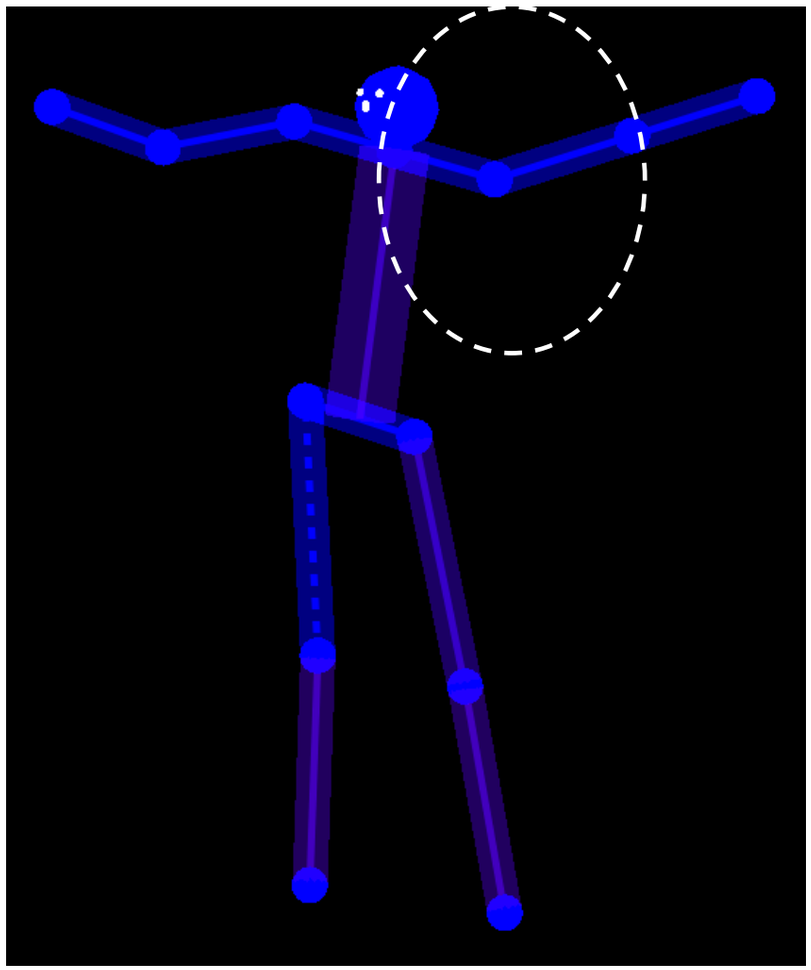
How to find *semantically* similar patches at training time



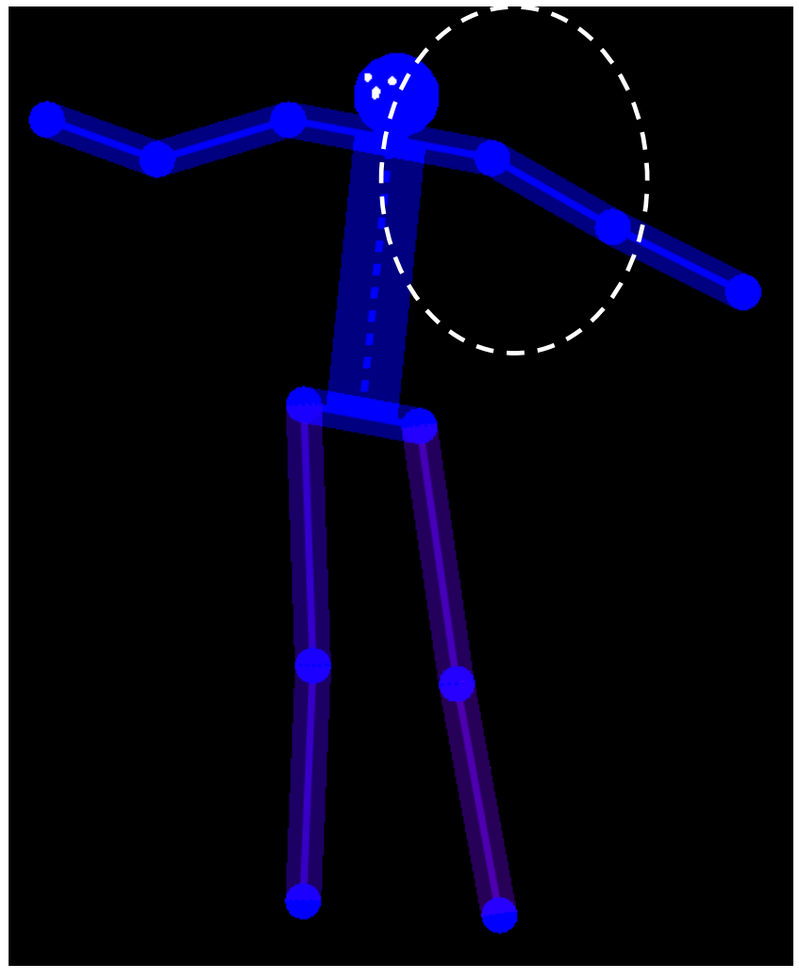
We annotated the locations of various joints such as eyes, nose, shoulders and limbs for each *training* instances

Distance in configuration space

Procrustes analysis



s



t

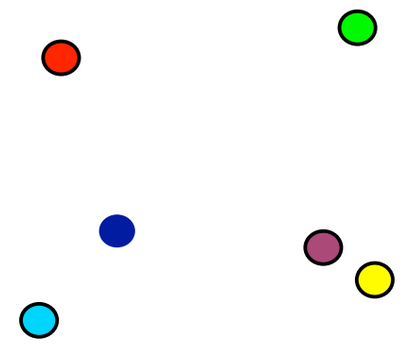
sum of the squared errors after transformation

$$d_s(r) = \sum_i ||\mathbf{x}_s(i) - T\mathbf{x}_r(i)||_2^2$$

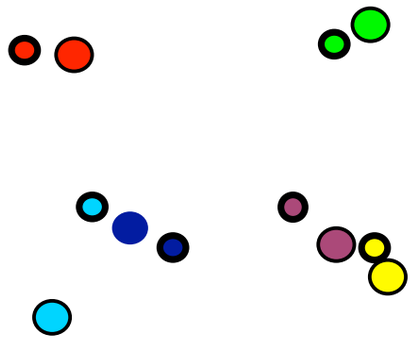
Question : how to find the optimal transformation?

How to find *semantically* similar patches at training time

How to find *semantically* similar patches at training time

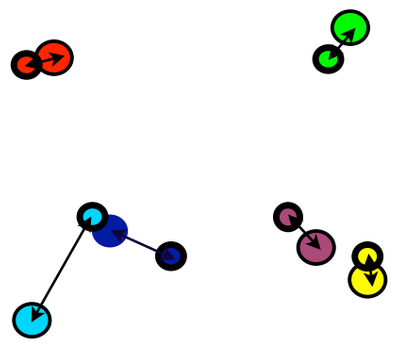


How to find *semantically* similar patches at training time



How to find *semantically* similar patches at training time

How to find *semantically* similar patches at training time



residual error

Training poselet classifiers

residual error

- Given a source patch

Training poselet classifiers

residual error

- Given a source patch
- Find the closest patch in every other instance

Training poselet classifiers

residual error

0.15

- Given a source patch
- Find the closest patch in every other instance

Training poselet classifiers

residual error

0.15

0.20

- Given a source patch
- Find the closest patch in every other instance

Training poselet classifiers

residual error

0.15

0.20

0.15

- Given a source patch
- Find the closest patch in every other instance

Training poselet classifiers

residual error

0.15

0.20

0.15

0.85

- Given a source patch
- Find the closest patch in every other instance

Training poselet classifiers

residual error

0.15

0.20

0.15

0.85

0.35

- Given a source patch
- Find the closest patch in every other instance

Training poselet classifiers

residual error

0.15

0.20

0.15

0.85

0.35

0.10

- Given a source patch
- Find the closest patch in every other instance

Training poselet classifiers


residual error


0.10


0.15


0.15


0.20


0.35


0.85

- Given a source patch
- Find the closest patch in every other instance
- Sort them by residual error

Training poselet classifiers

residual error

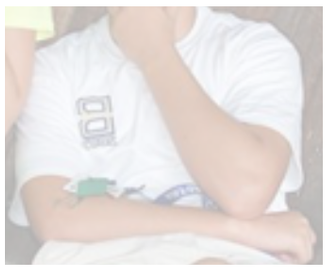
0.10

0.15

0.15

0.20

0.35



0.85

- Given a source patch
- Find the closest patch in every other instance
- Sort them by residual error
- Threshold the list

Training poselet classifiers

Pos ={…  …}

- Given a source patch
- Find the closest patch in every other instance
- Sort them by residual error
- Threshold the list
- Use these patches to train a standard “Dalal & Triggs” detector, i.e. HOG + linear SVMs with data mining

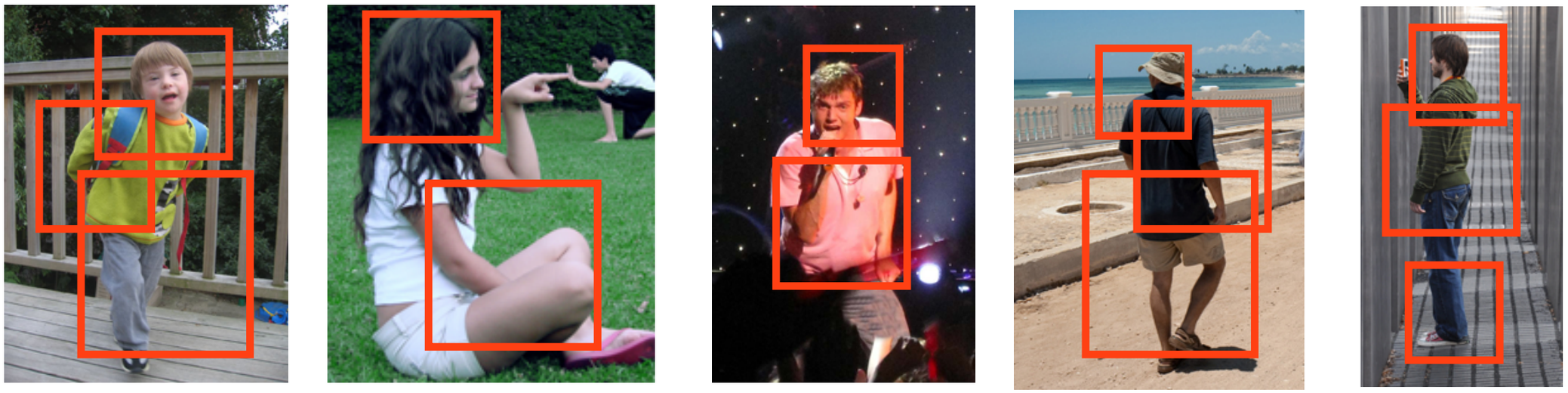
Which poselets should we train?

- *Machine learning solution* : train a large number of possible poselets and select a subset based on the task
- Generate thousands of *random windows*, generate poselet candidates using the earlier method and train detectors using HOG + linear SVMs (Dalal & Triggs)

Bourdev & Malik 09
Bourdev et al. 10

Which poselets should we train?

- *Machine learning solution* : train a large number of possible poselets and select a subset based on the task
- Generate thousands of *random windows*, generate poselet candidates using the earlier method and train detectors using HOG + linear SVMs (Dalal & Triggs)



Bourdev & Malik 09
Bourdev et al. 10

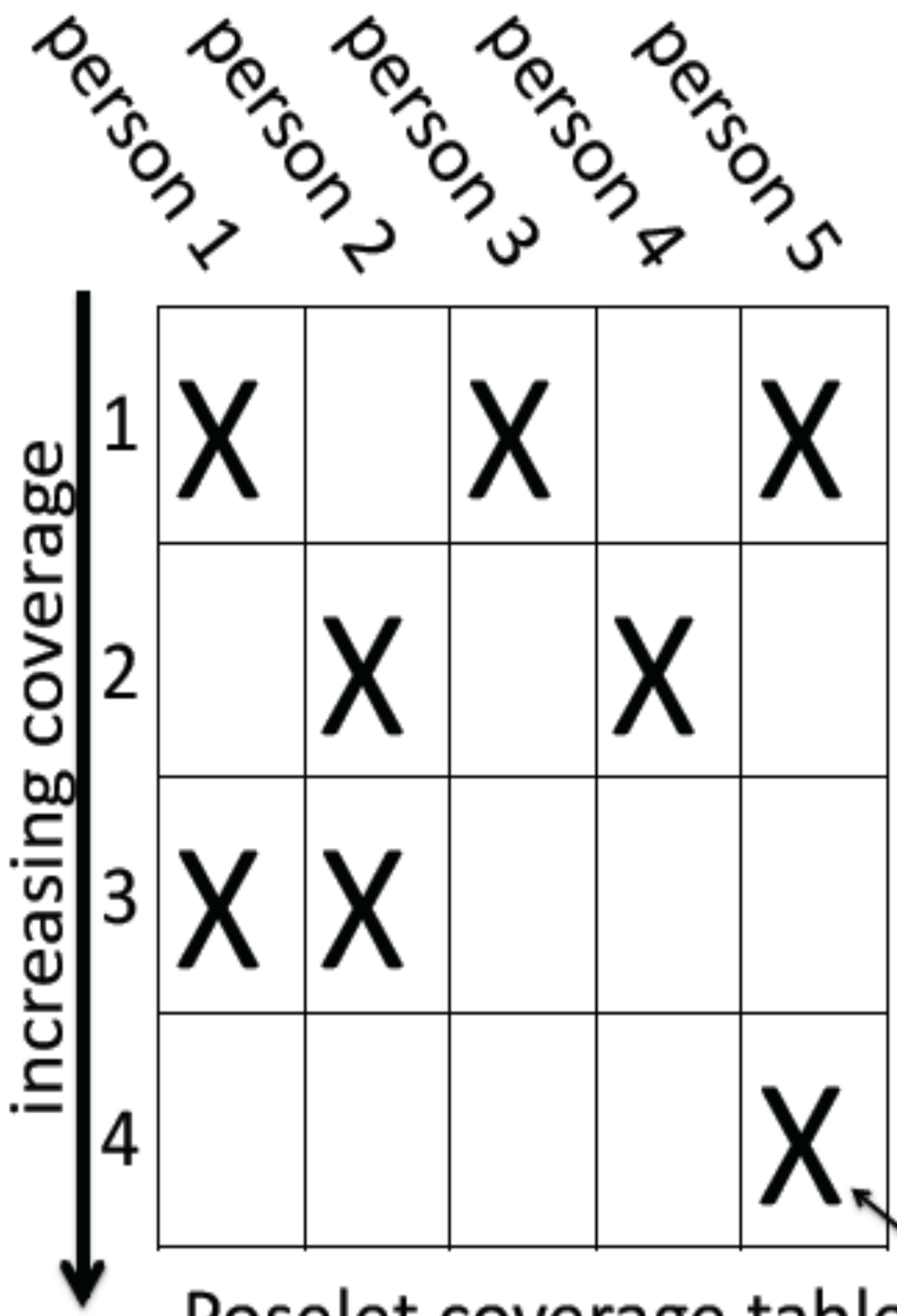
Which poselets should we train?

- *Machine learning solution* : train a large number of possible poselets and select a subset based on the task
- Generate thousands of *random windows*, generate poselet candidates using the earlier method and train detectors using HOG + linear SVMs (Dalal & Triggs)

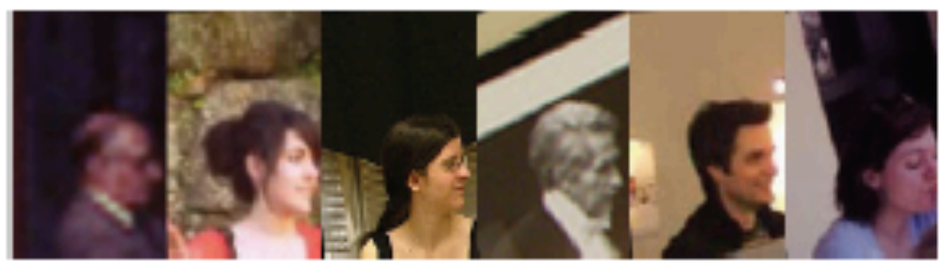
- Select a set of poselets that are
  - individually effective
  - complimentary

Bourdev & Malik 09
Bourdev et al. 10

Selecting poselets for detection



Selected poselets



Poselet 4 activates on person 5

Poselets selected for PASCAL VOC person detection

Bourdev et al. 10

Poselets selected for PASCAL VOC person detection

Bourdev et al. 10

Poselets selected for PASCAL VOC person detection

Bourdev et al. 10

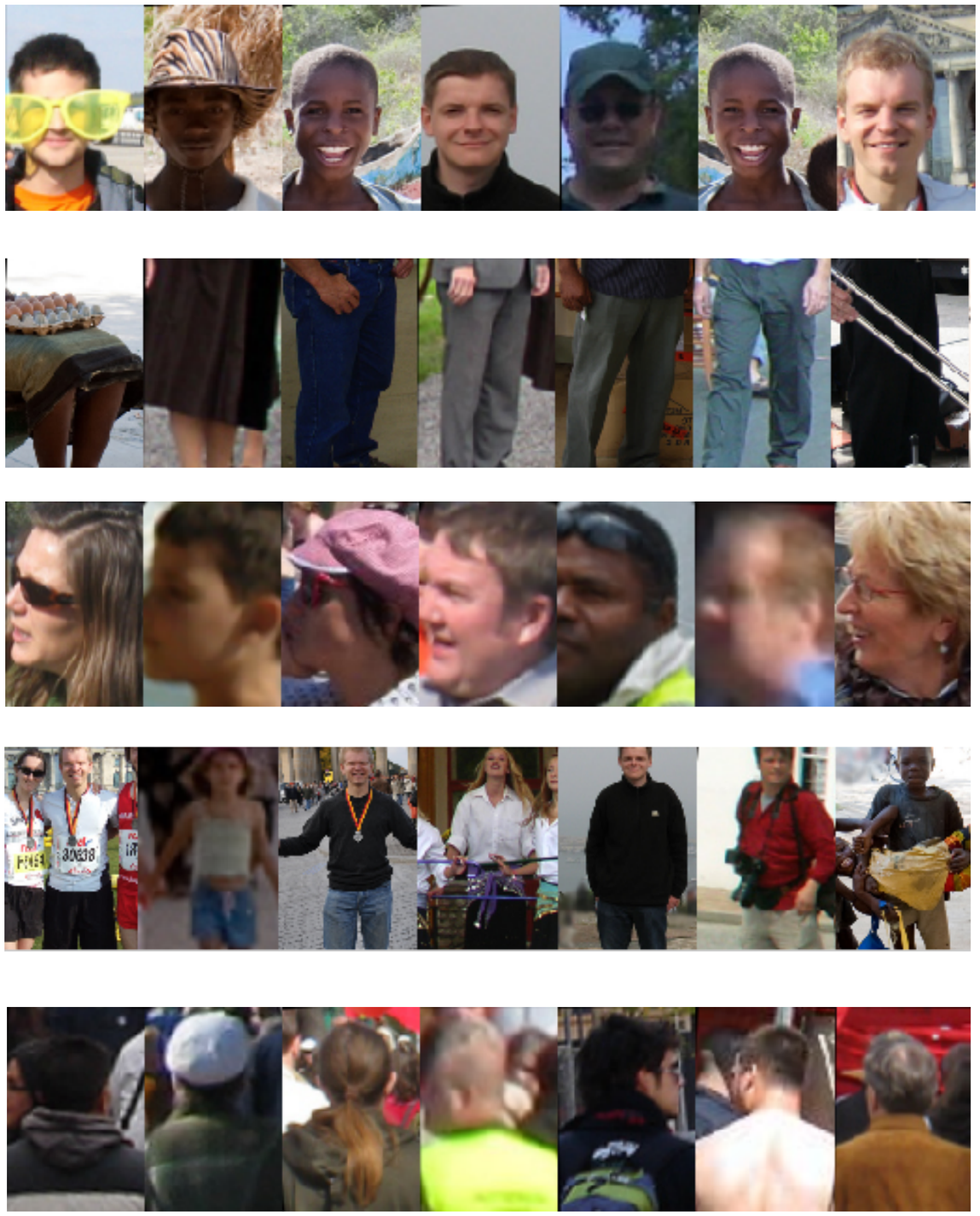
Poselets selected for PASCAL VOC person detection

Bourdev et al. 10

Poselets selected for PASCAL VOC person detection

Bourdev et al. 10

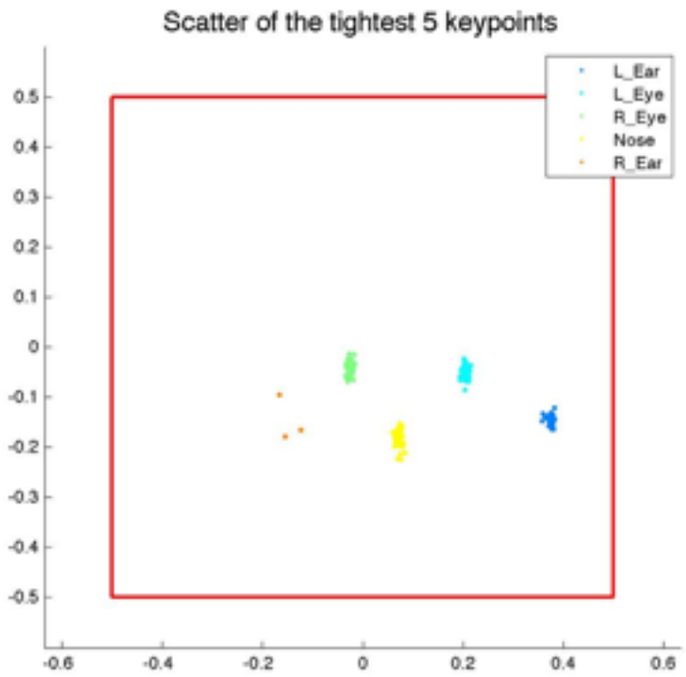
Poselets selected for PASCAL VOC person detection

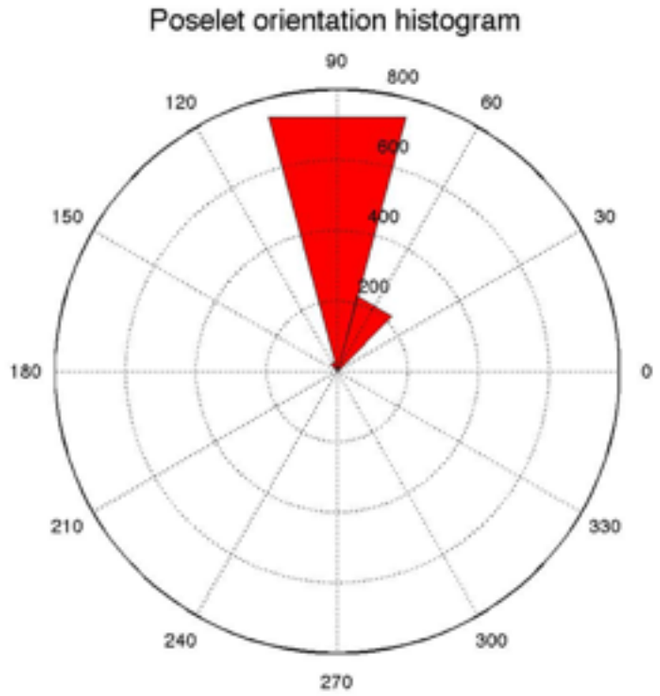


Bourdev et al. 10

What does a poselet tell us?

frontal face





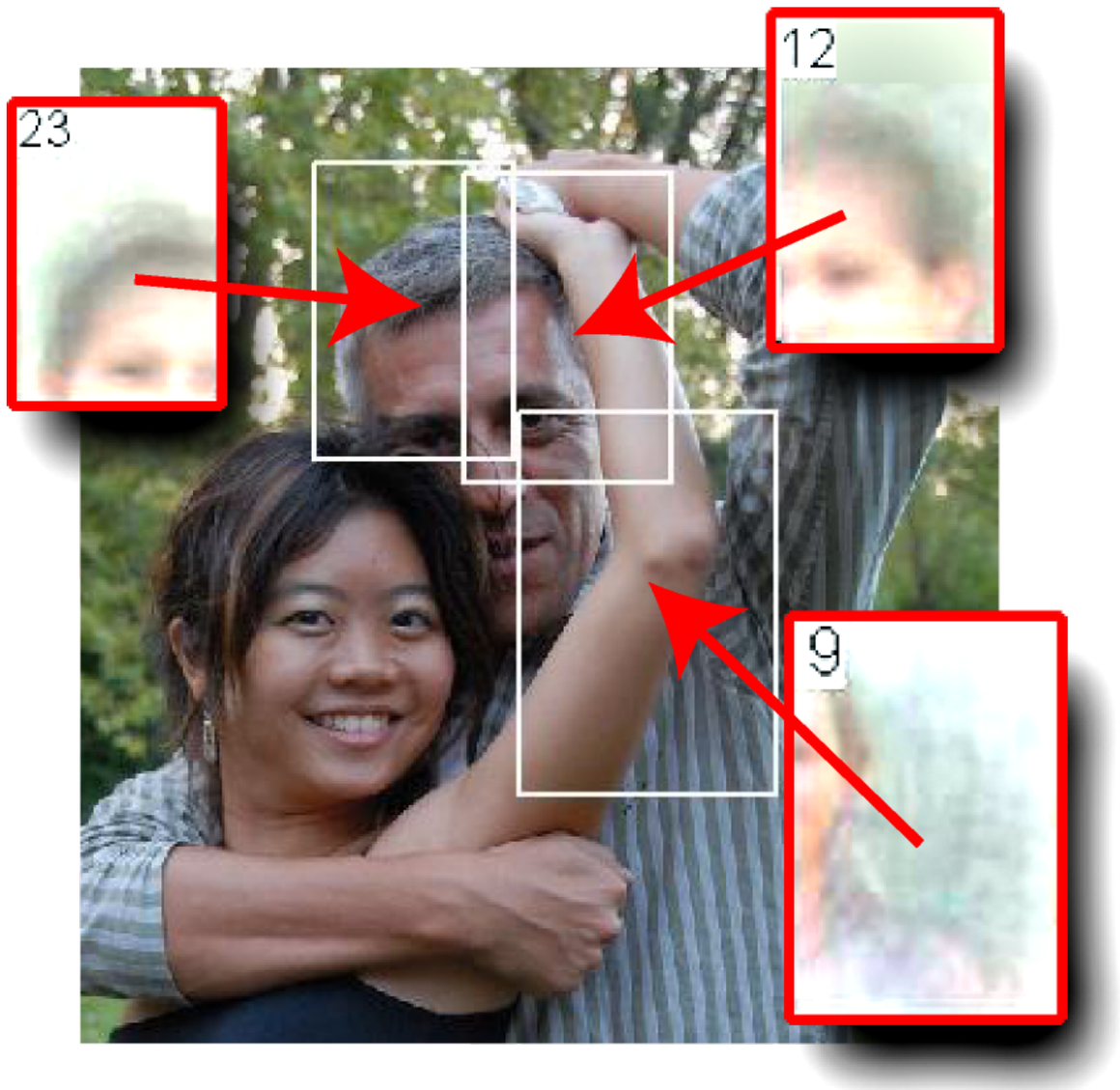
top detections on the training set

Person detection using poselets

- Example of a poselet

- Estimate relative bounding box on the training set

Person detection using poselets



- Detect each poselet in an image
- Vote for the person bounding box (Hough transform)
- Find non-overlapping clusters
- Score each cluster using a weighted combination of poselet detection scores

$$s_i = \sum_{p \in C_i} w_p a_p$$

person detection score

weight of each poselet

poselet detection score

Bourdev & Malik 09, Bourdev et al.10, Maji & Malik 10

Performance on PASCAL VOC detection challenge

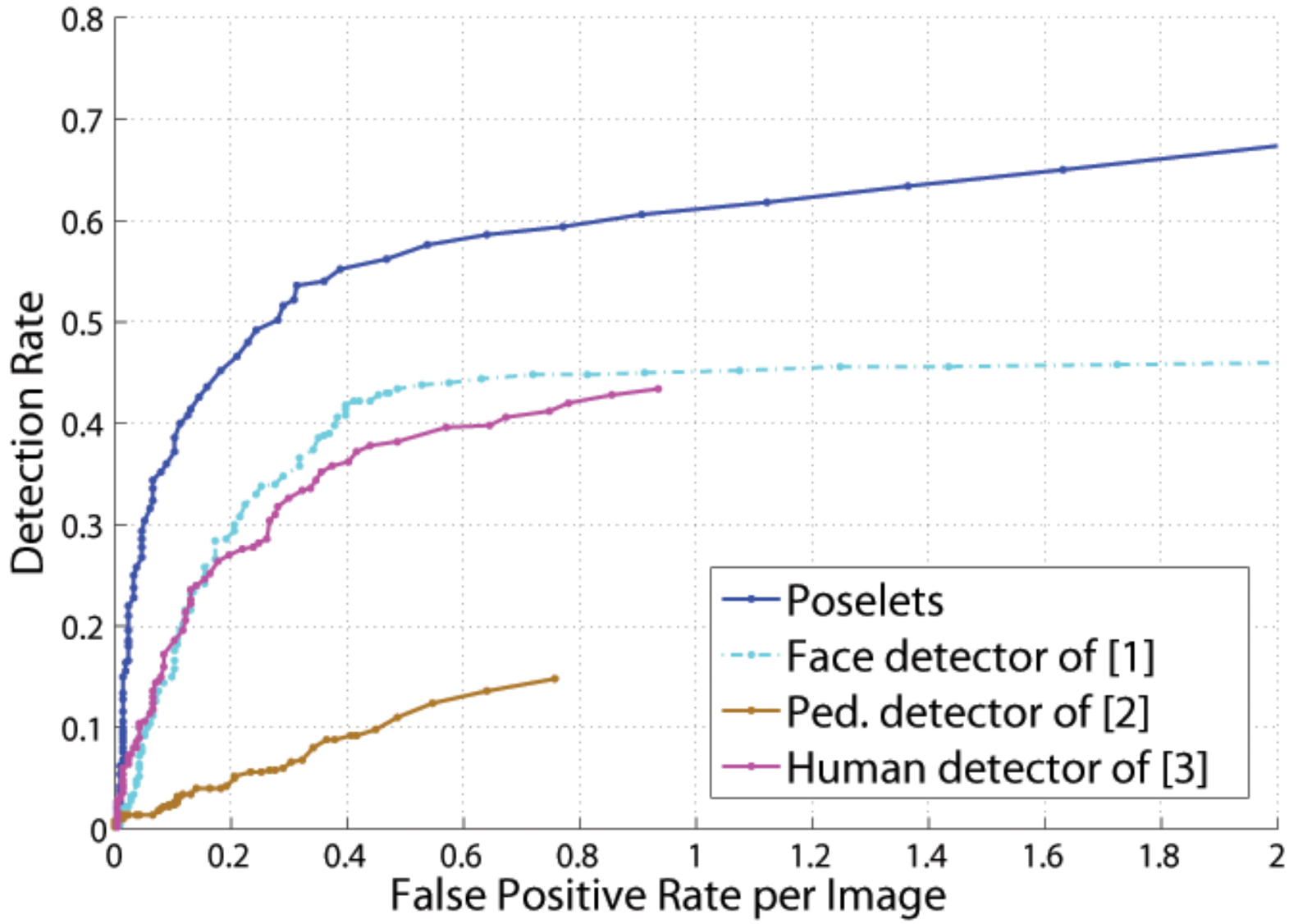
Person category VOC 2010 *test set*

Method	**Detection AP**
Poselets	48.5%
Dalal & Triggs	~12%

L. Bourdev, S. Maji, T. Brox, J. Malik
Detecting people using mutually consistent poselet activations, ECCV 2010

http://www.cs.berkeley.edu/~lbourdev/poselets/

Performance on the H3D dataset

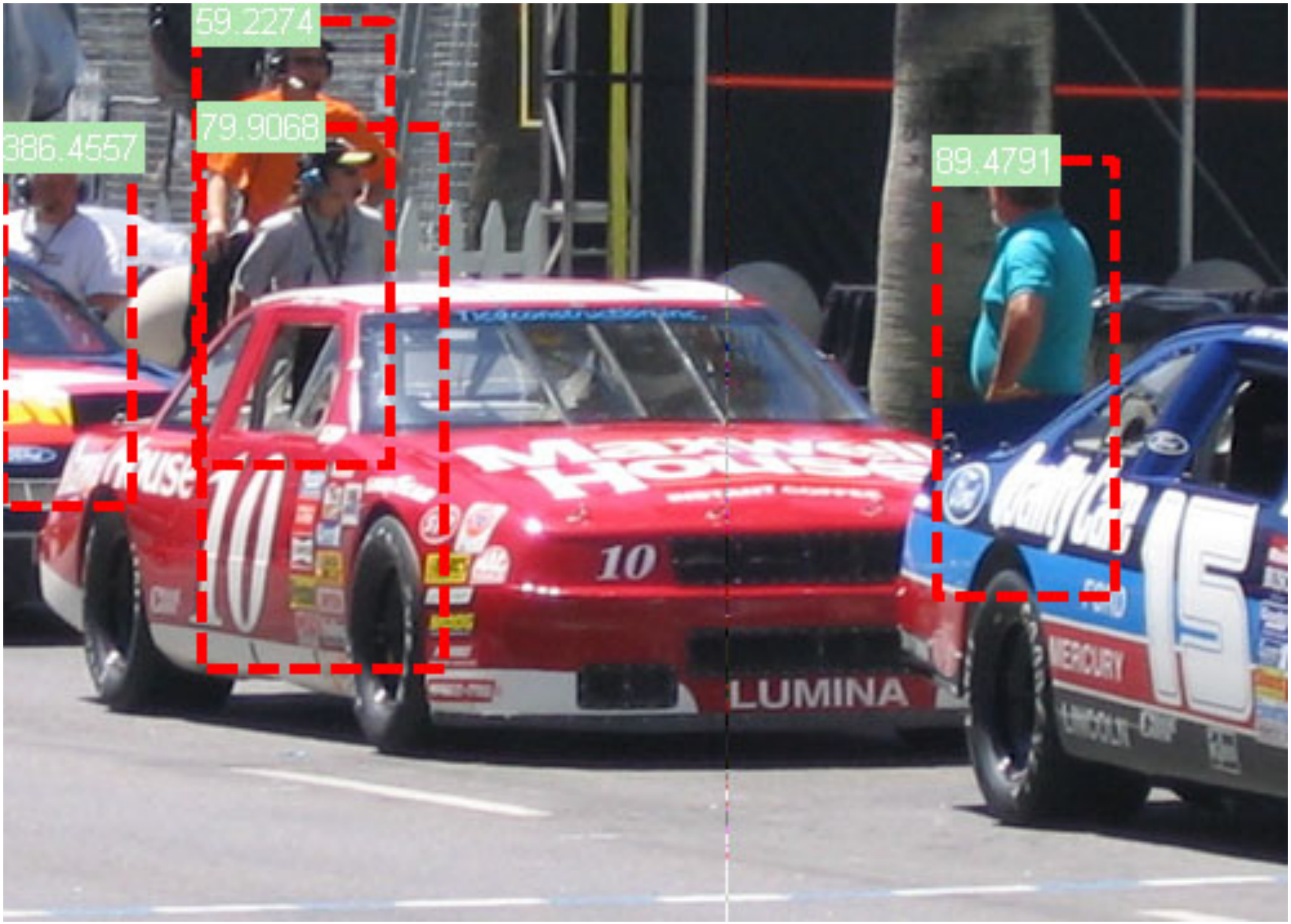


[1] Bourdev and Brandt, CVPR 2005

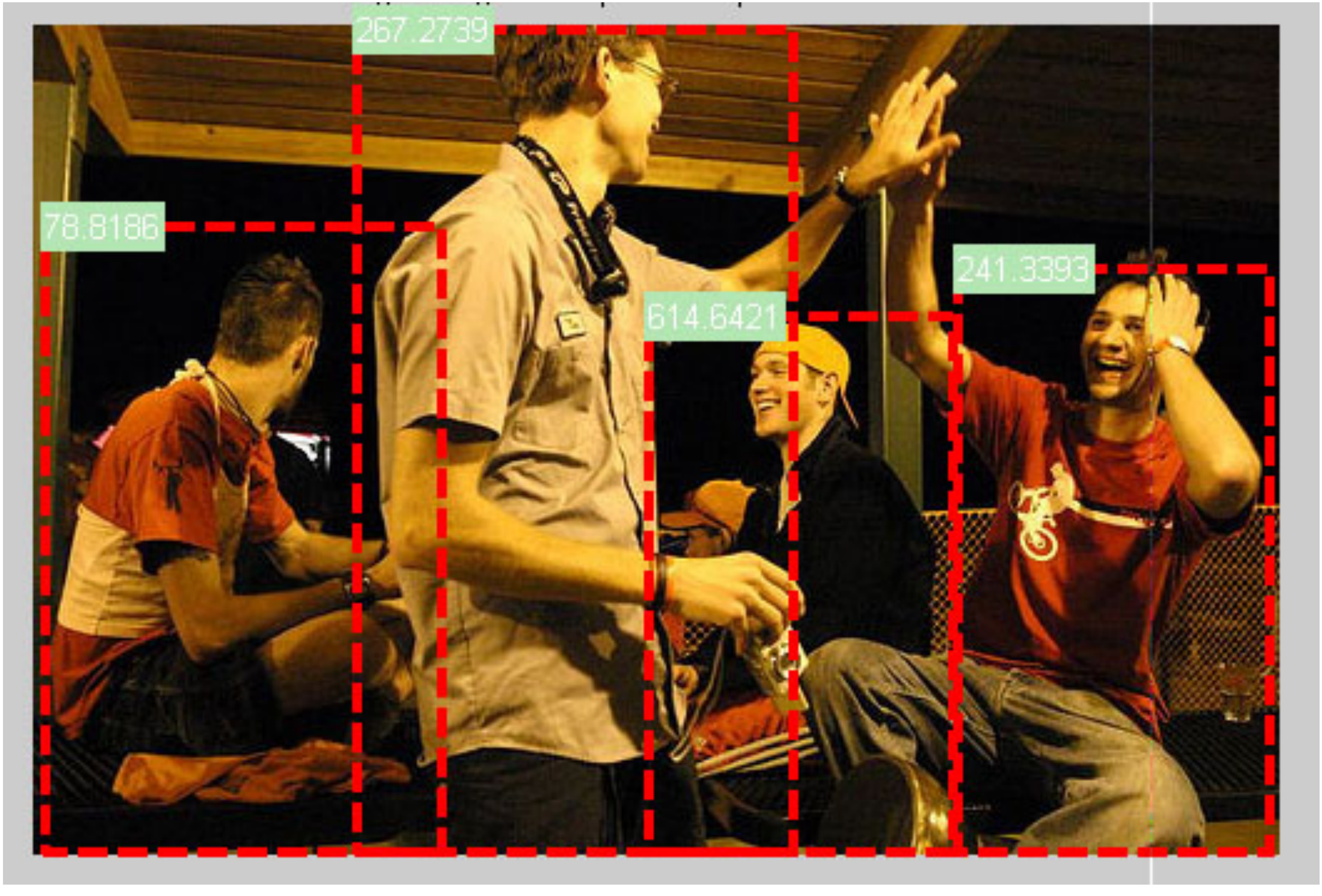
[2] Dalal & Triggs, CVPR 2005

[3] Felzenszwalb, McAllester & Ramanan, CVPR 2008

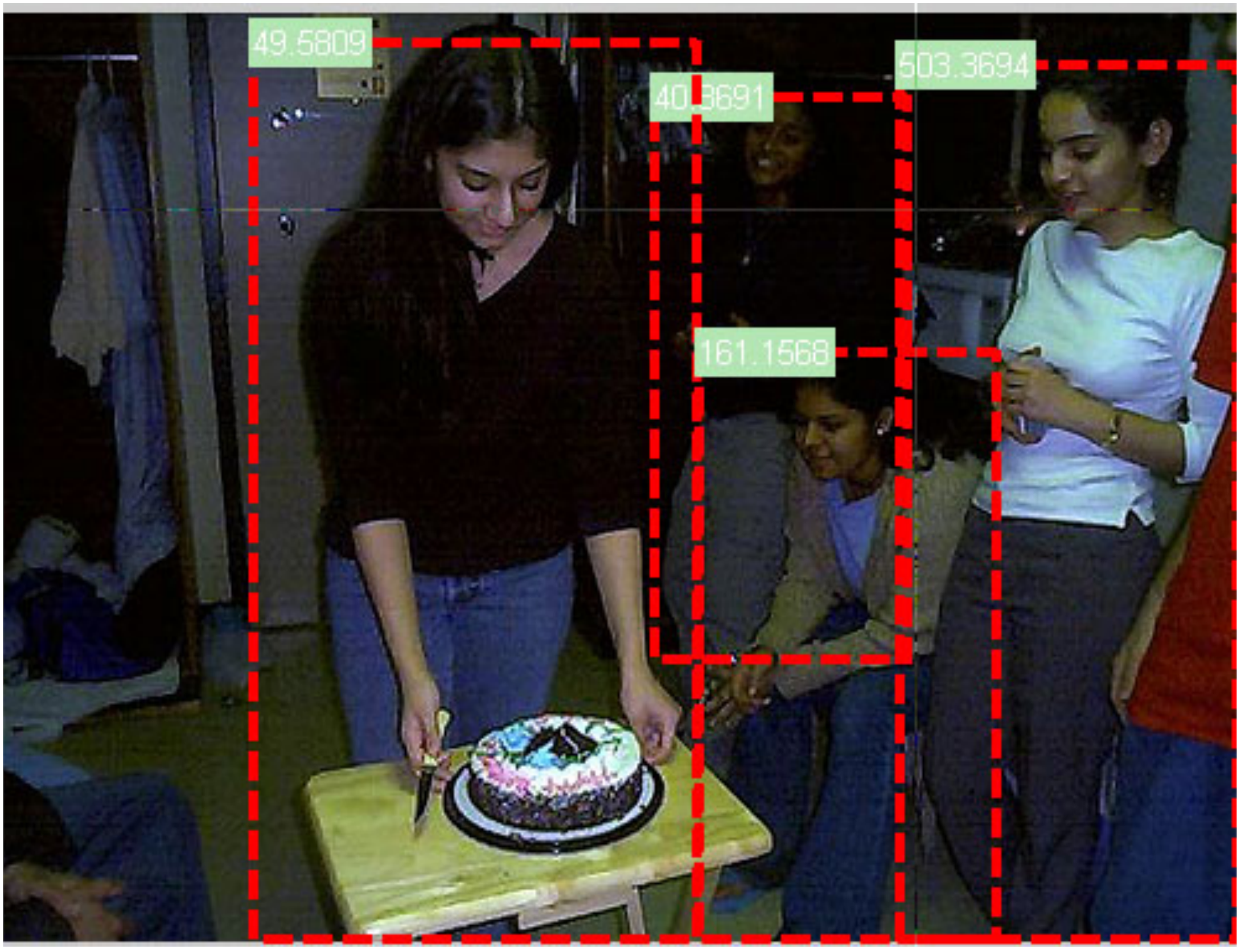
Example detections



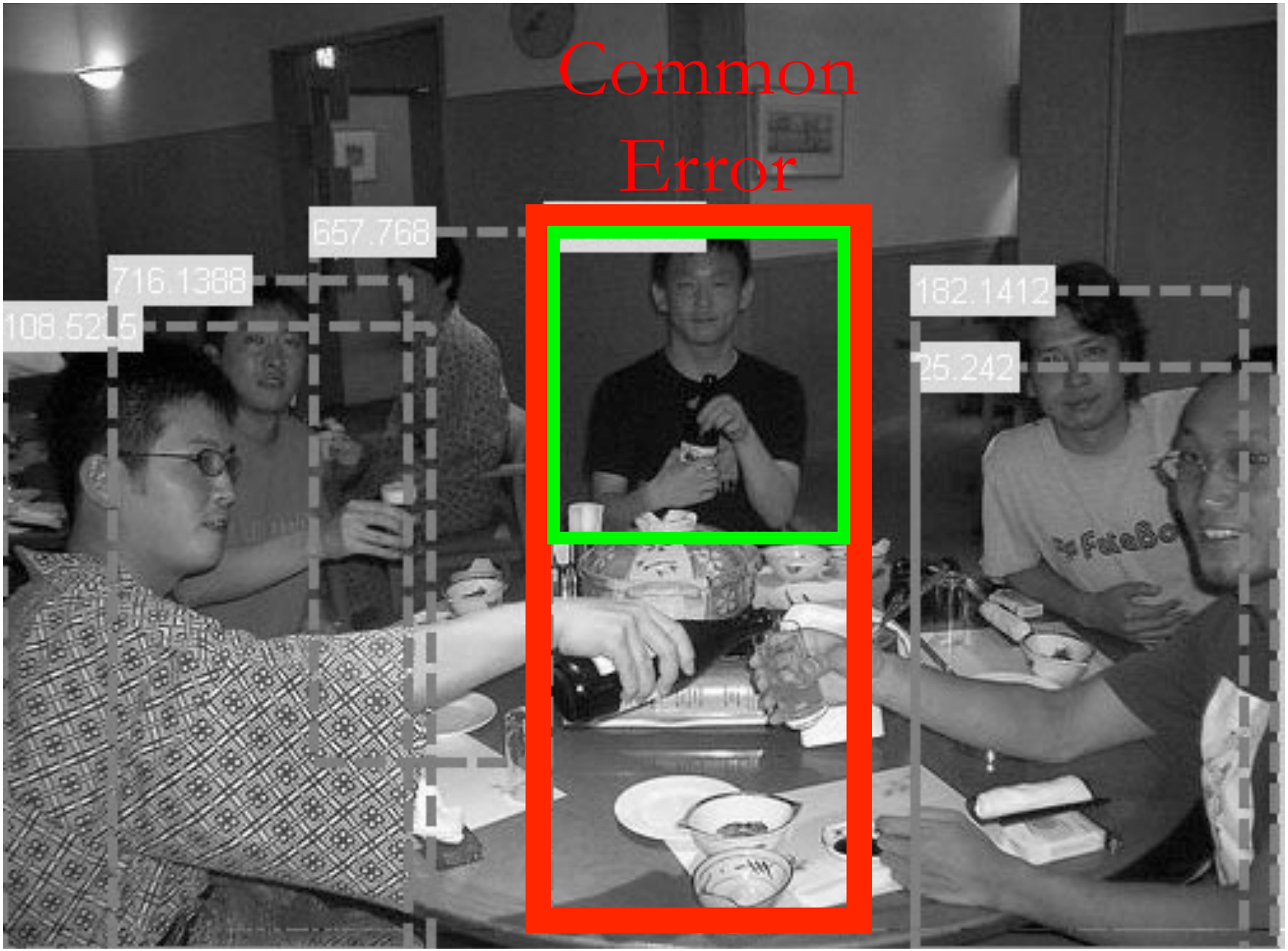
Example detections



Example detections



Example detections

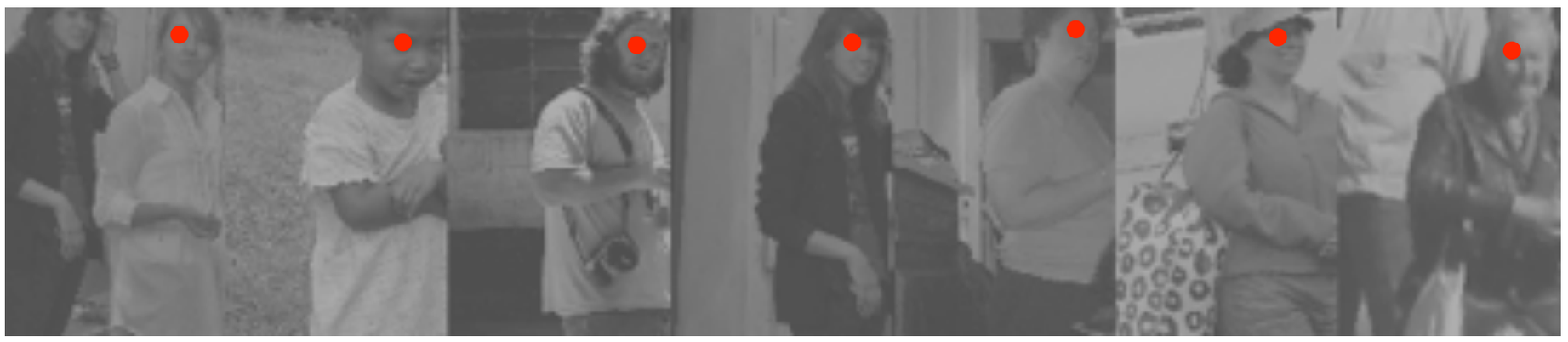


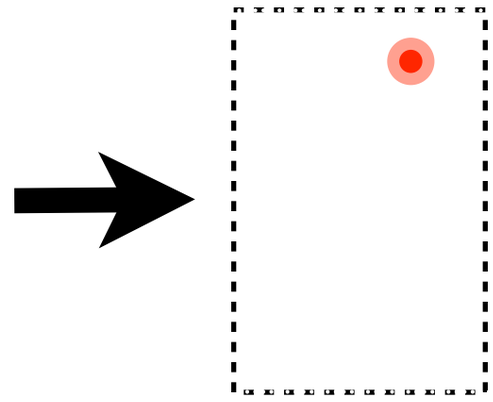
keypoint prediction using poselets

- Example of a poselet

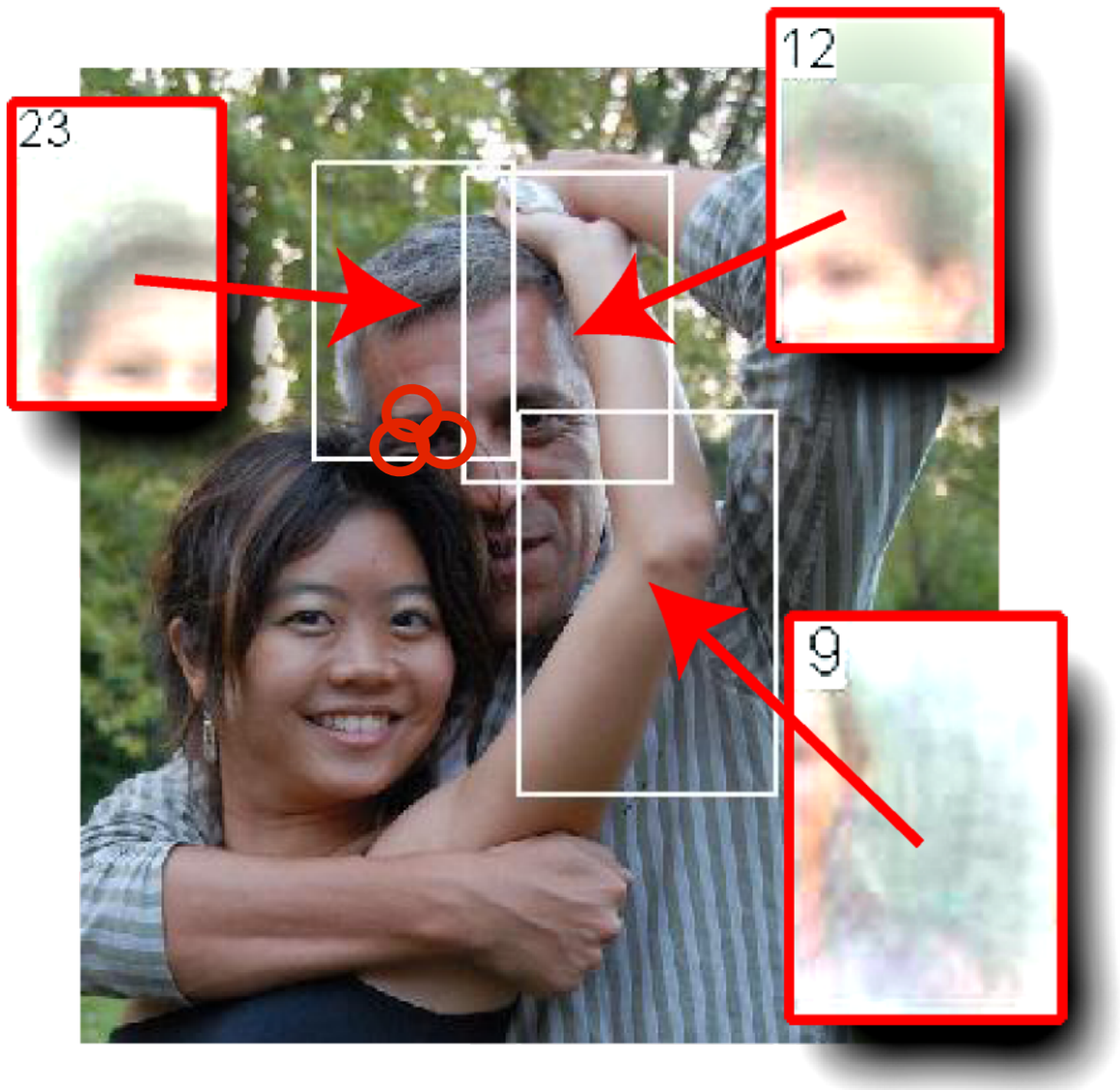
- Estimate average right-eye location

expected location





right-eye detection using poselets



- Detect each poselet in an image
- Vote for the ~~bounding box~~ eye location
- Find non-overlapping clusters
- Score each cluster using a weighted combination of poselet detection scores

$$s_i = \sum_{p \in C_i} w_p a_p$$

keypoint detection score

weight of each poselet

poselet detection score

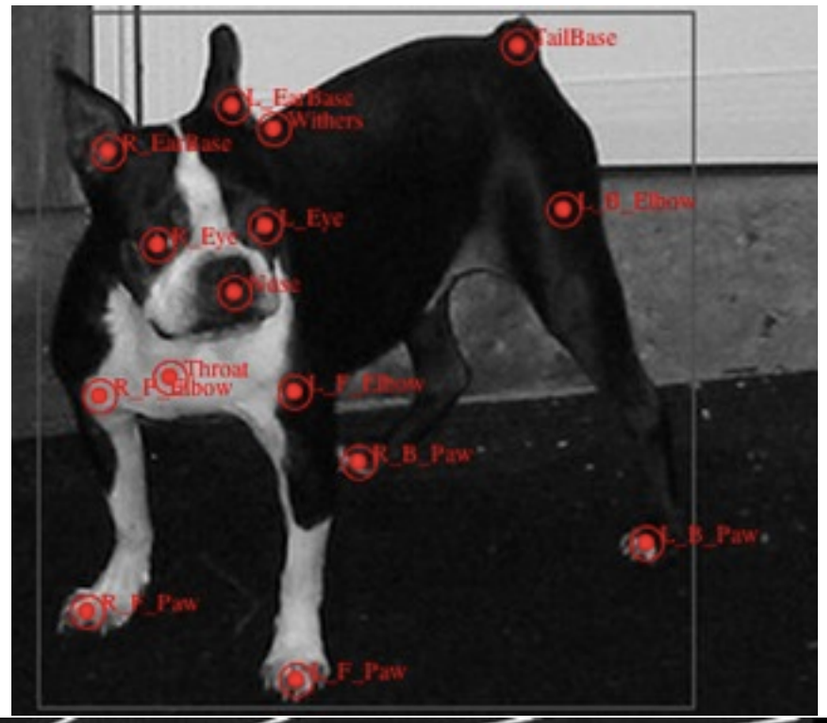
Bourdev & Malik 09, Bourdev et al.10, Maji & Malik 10

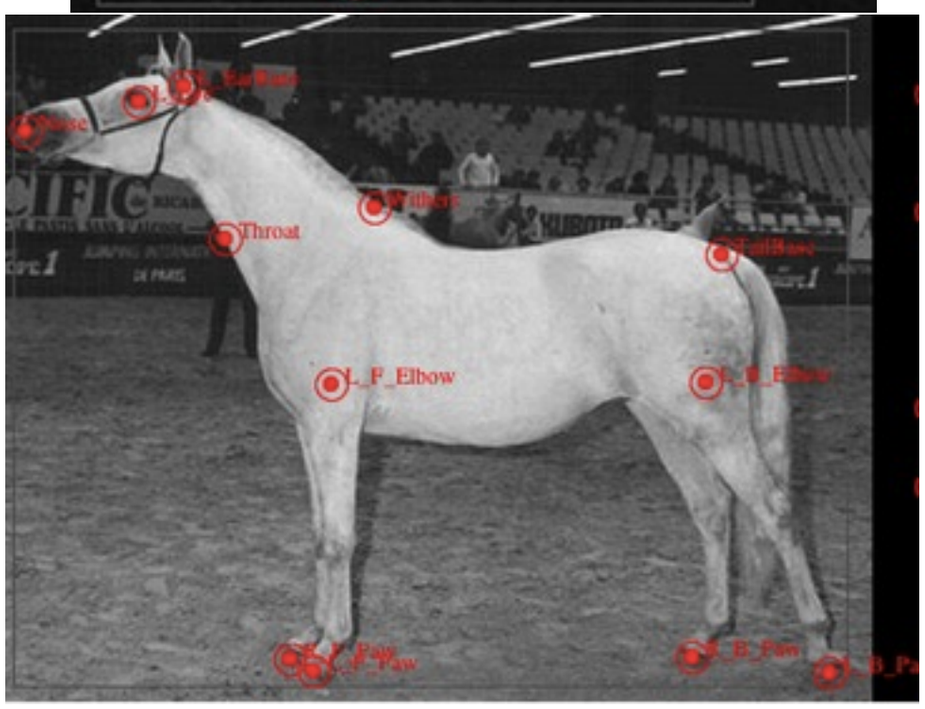
Any questions so far?

Poselets for other categories

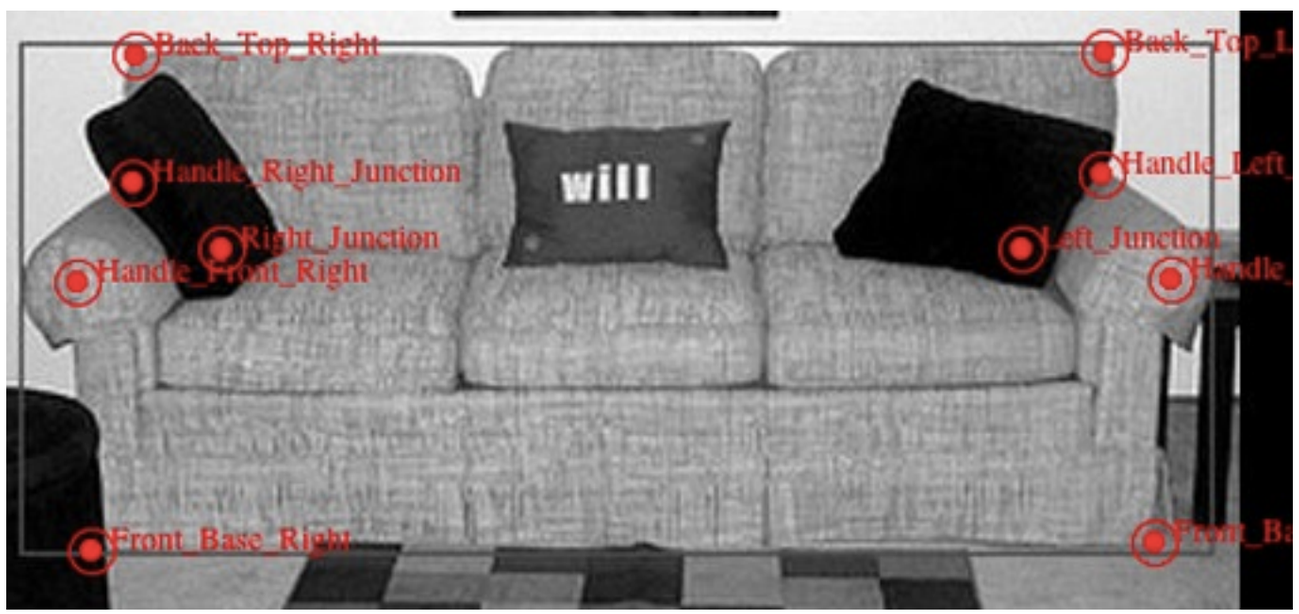
Identify a set of *keypoints*

animals





man-made objects

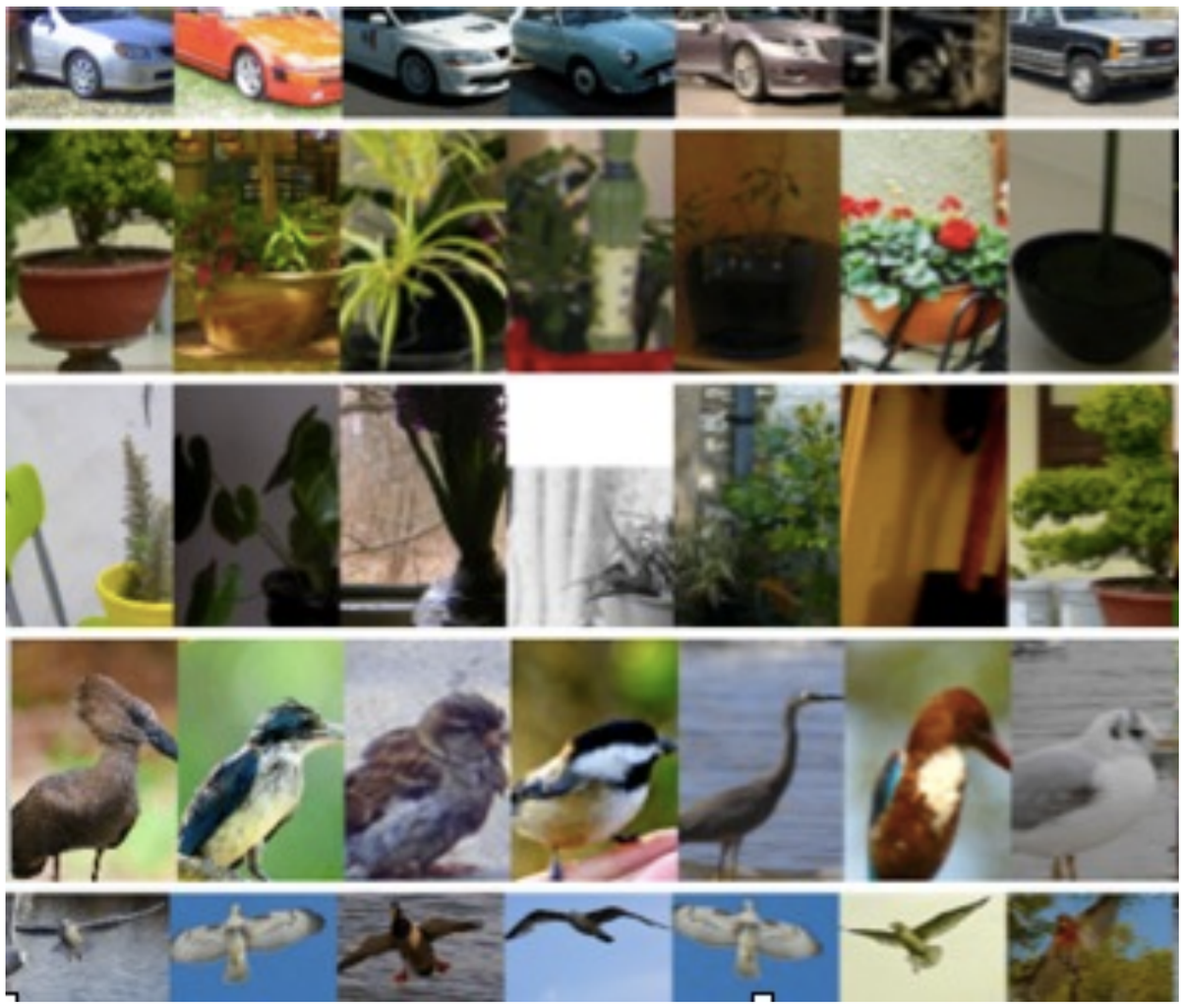


not easy sometimes

Poselets for *horses*

each *poselet* (row) captures the appearance of the object at a fixed viewpoint and pose

Poselets for other categories



each *poselet* (row) captures the appearance of the object at a fixed viewpoint and pose

How much do we gain over a *single template*?

PASCAL VOC 2007 *test set*

Category	Dalal&Triggs	Poselets
aeroplane	12.7	24.4
bicycle	25.3	57.9
bird	0.5	15.8
boat	1.5	14.8
bottle	10.7	41.7
bus	20.5	40.6
car	23.0	61.4
cat	0.5	22.5
chair	2.1	18.9
cow	12.8	32.3

Category	Dalal&Triggs	Poselets
diningtable	1.4	21.4
dog	0.4	17.8
horse	12.2	60.1
motorbike	10.3	37.8
person	10.1	46.9
pottedplant	2.2	14.2
sheep	5.6	29.6
sofa	5.0	26.8
train	12.0	22.7
tvmonitor	24.8	41.3

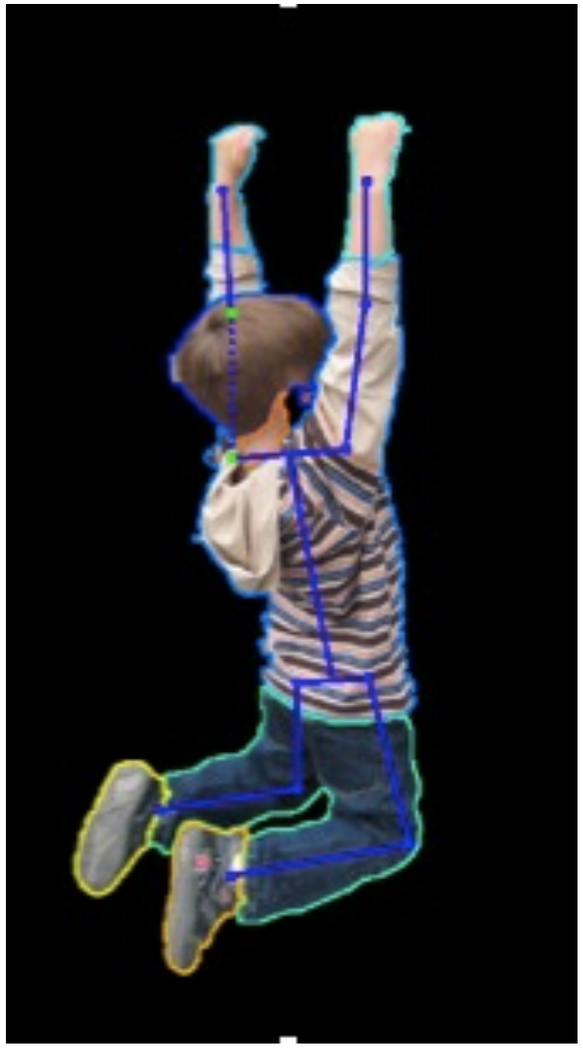
Poselets (Mean AP=32.2%)

Dalal & Triggs (Mean AP=9.7%)

Break

- Any questions?

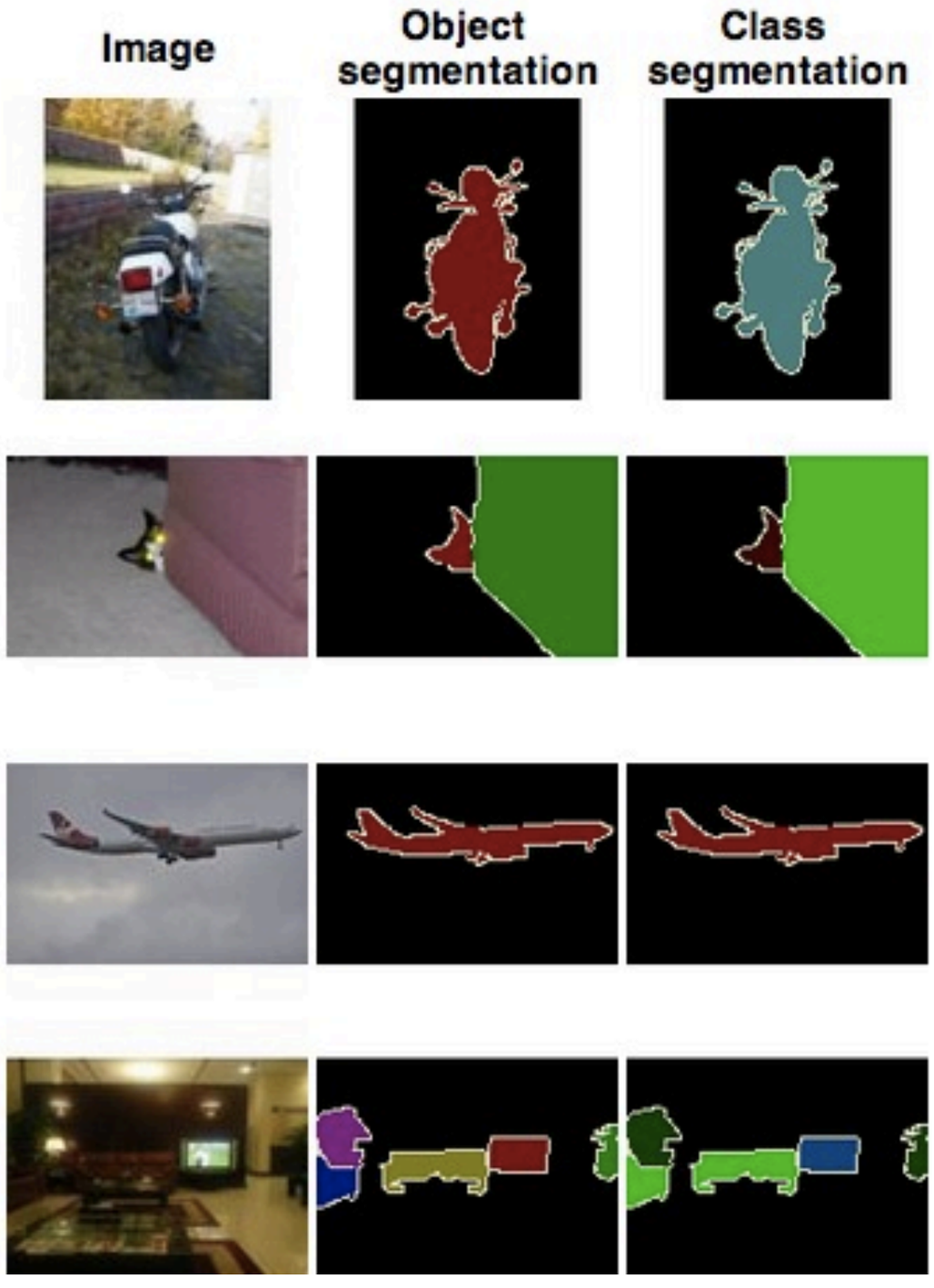
Beyond detection...



estimate pose, segmentation, gender, clothing, age, action, hair-style, etc.

Semantic segmentation

PASCAL VOC : 20 classes + "background"

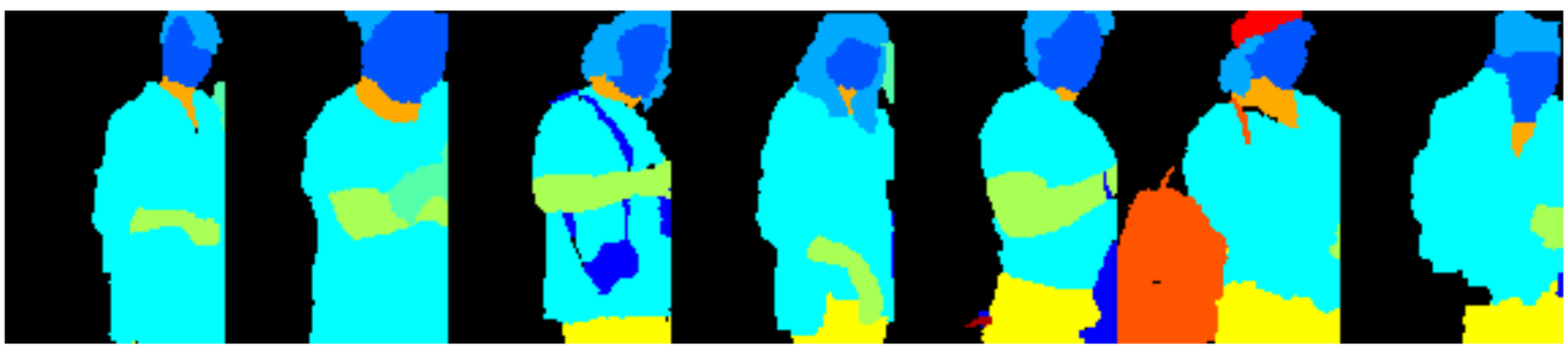


Person *segmentation* using poselets

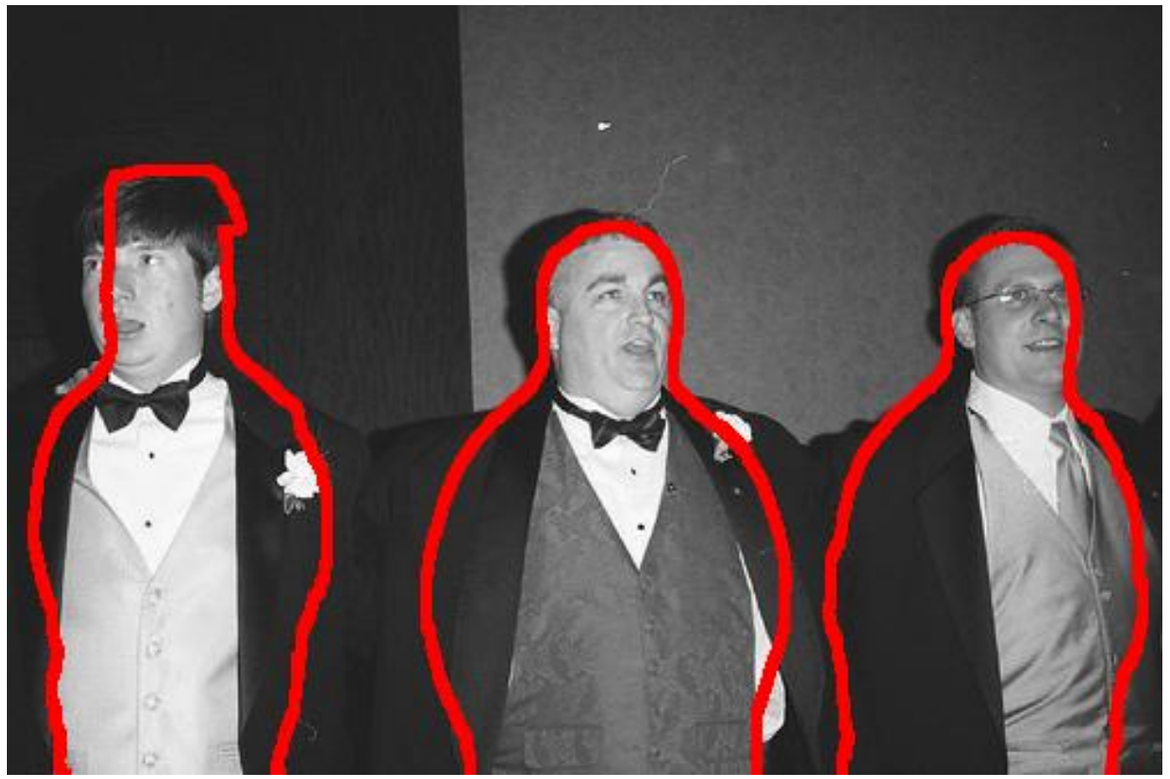
- Example of a poselet

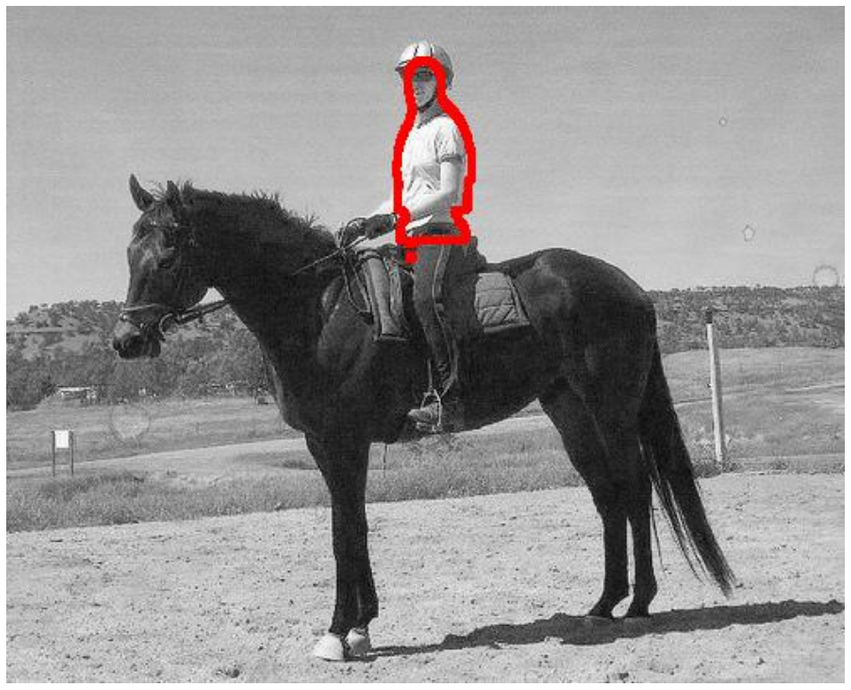
- Estimate average segmentation mask

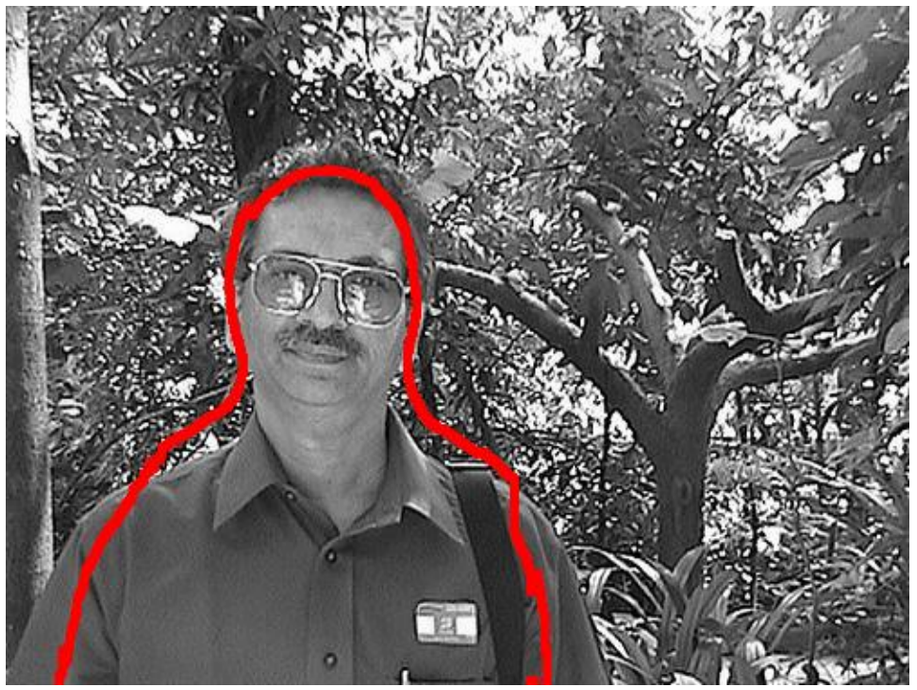
average ‘person’ mask



Example person segmentations



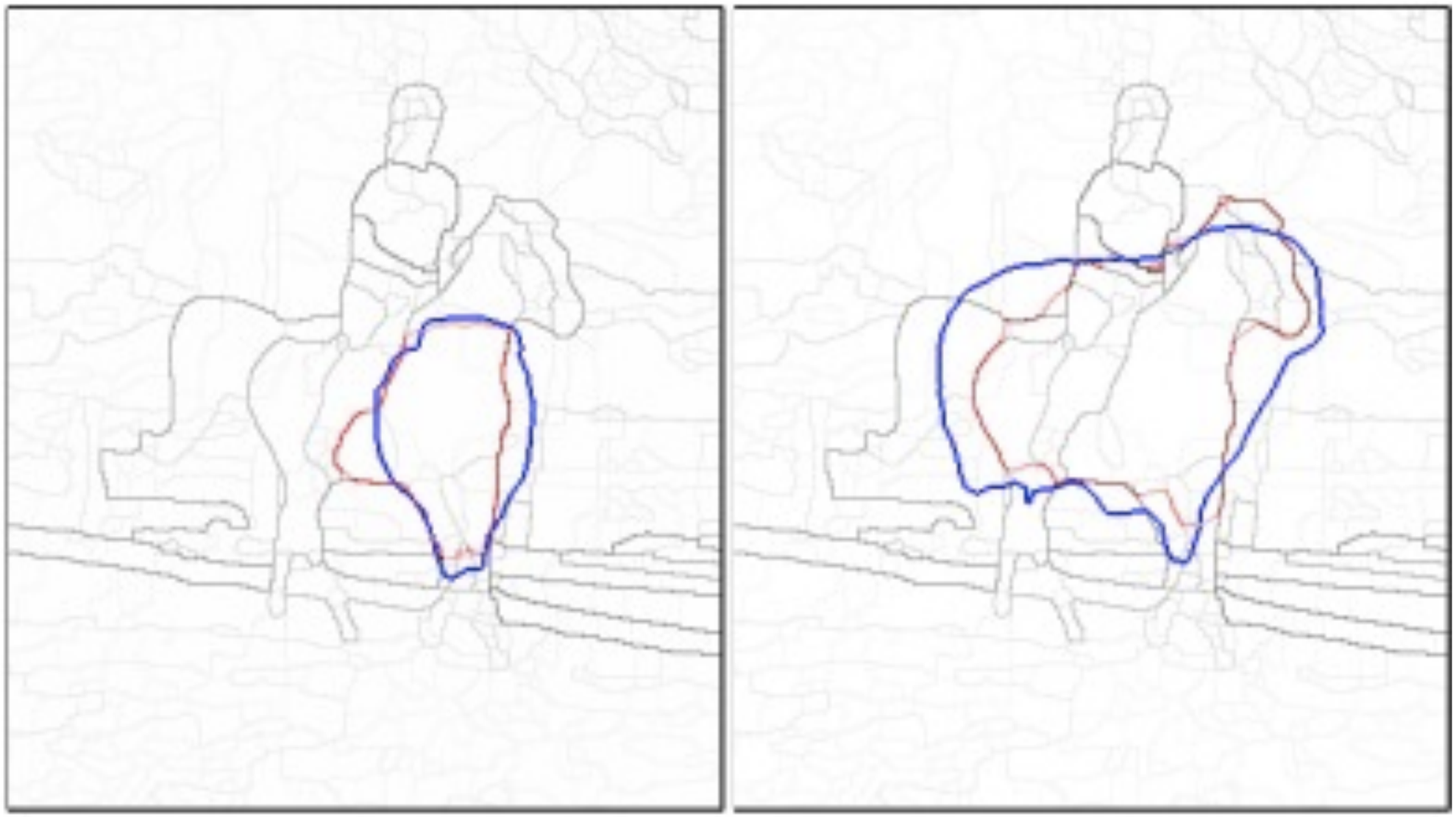




Alignment to image contours

Averaging causes blurring
Solution: align to image boundaries

Alignment to image contours



poselet contours before and after alignment

Multi-class segmentation

Combine predictions from *poselets* from 20 classes

Images from the PASCAL VOC *segmentation challenge*

Brox et al., CVPR 11

PASCAL VOC 2010 *segmentation* challenge

Method	**Accuracy**
Our [1]	34.9%
Oxford Brookes [2]	30.3%
Bonn [3]	39.7%
Barcelona [4]	40.1%

Rank #3 in the challenge
Best performance on 4/20 categories

[1] Brox, Bourdev, Maji and Malik, CVPR 11
[2] Associative Hierarchical CRFS, Ladicky, Russell, Kohli and Torr, ICCV 09
[3] Contrained CPMC, Carreira and Sminchisescu, CVPR 10
[4] Harmony potentials, Gonfaus et al., CVPR 10

note : current state of the art is around 45%

Any questions on semantic segmentation?

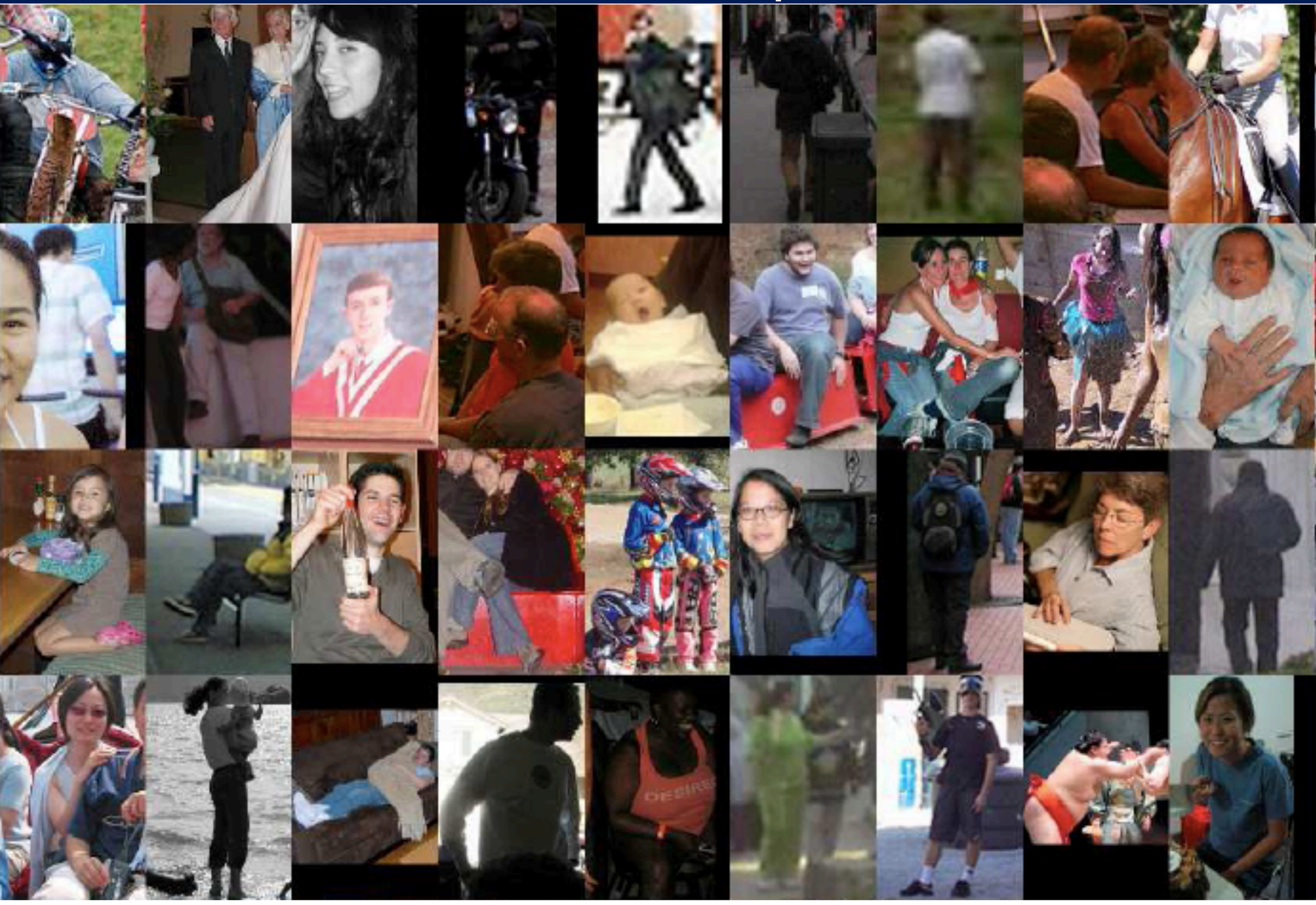
- There are many methods that work well. Most of them combine some bottom-up region signal with top-down classifications.

Recognizing attributes

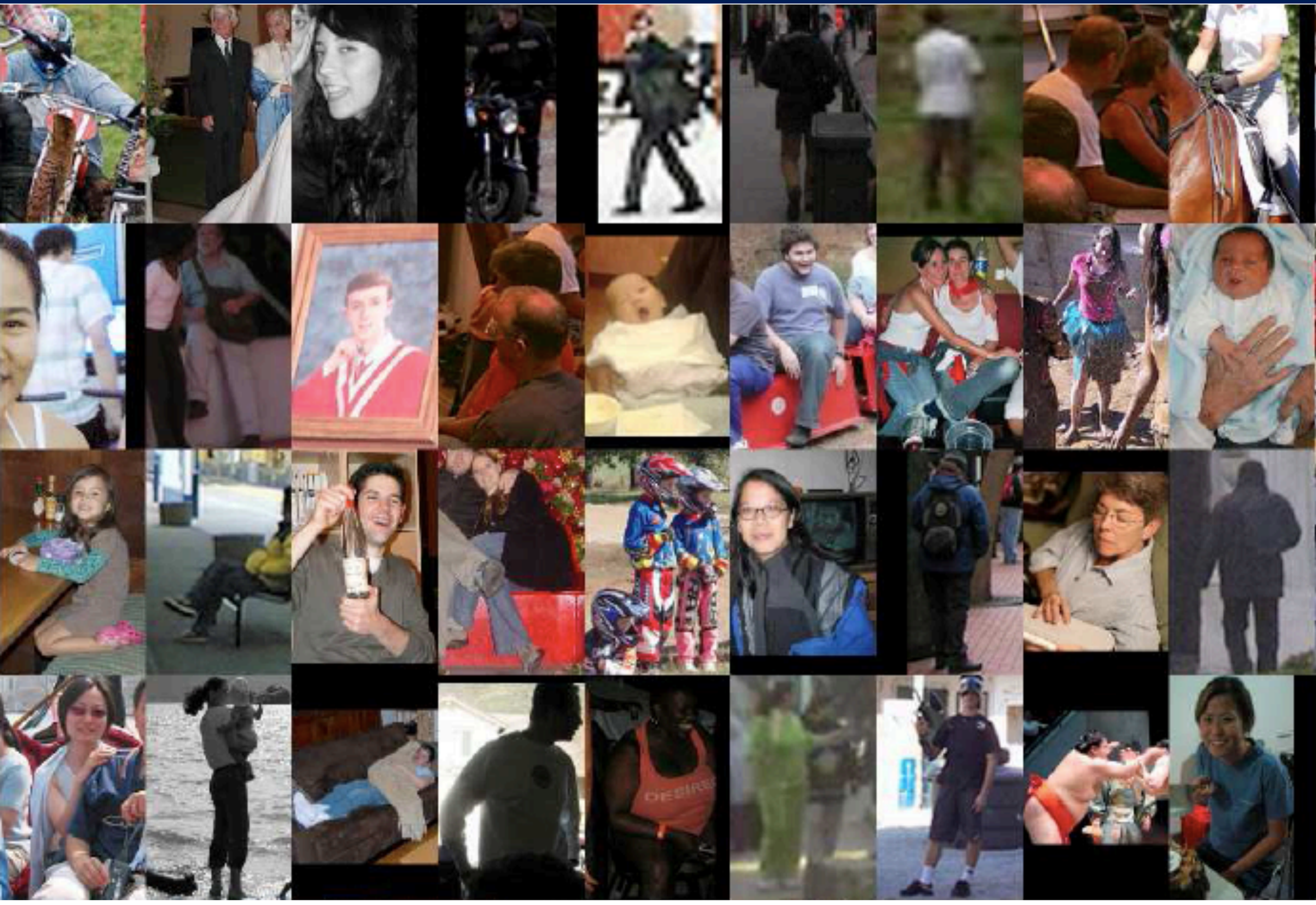
(Ready for a quiz?)

Who has long hair?

Who has short pants?



male or female?



male or female?

Classification is easier if we factor out the pose

male or female?

Classification is easier if we factor out the pose

male or female?

this is exactly what *poselets* allow us to do

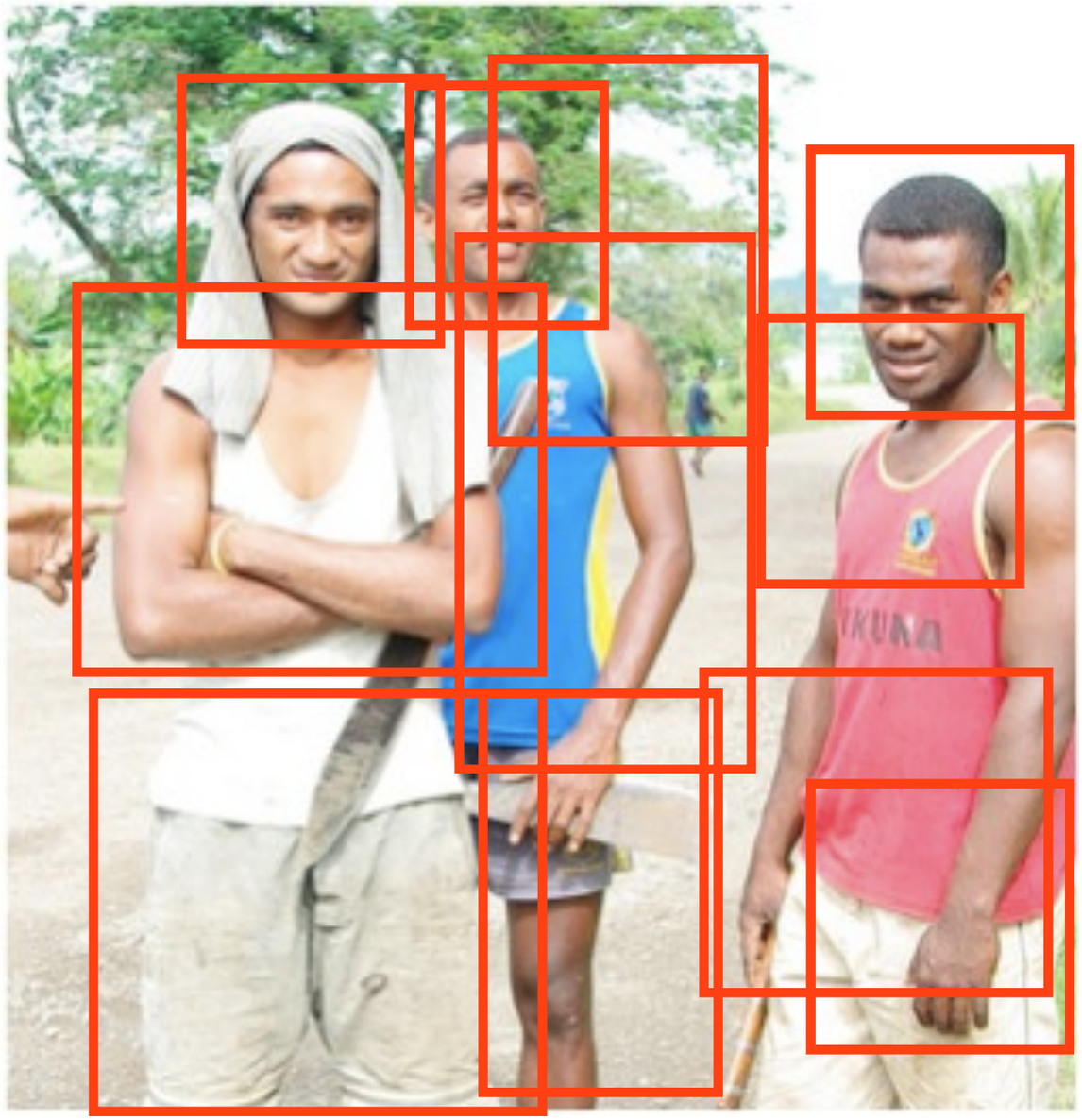
Goal : extract attributes of this person

Goal : extract attributes of this person

Given: *image* and *bounds* of the desired and other humans in the image

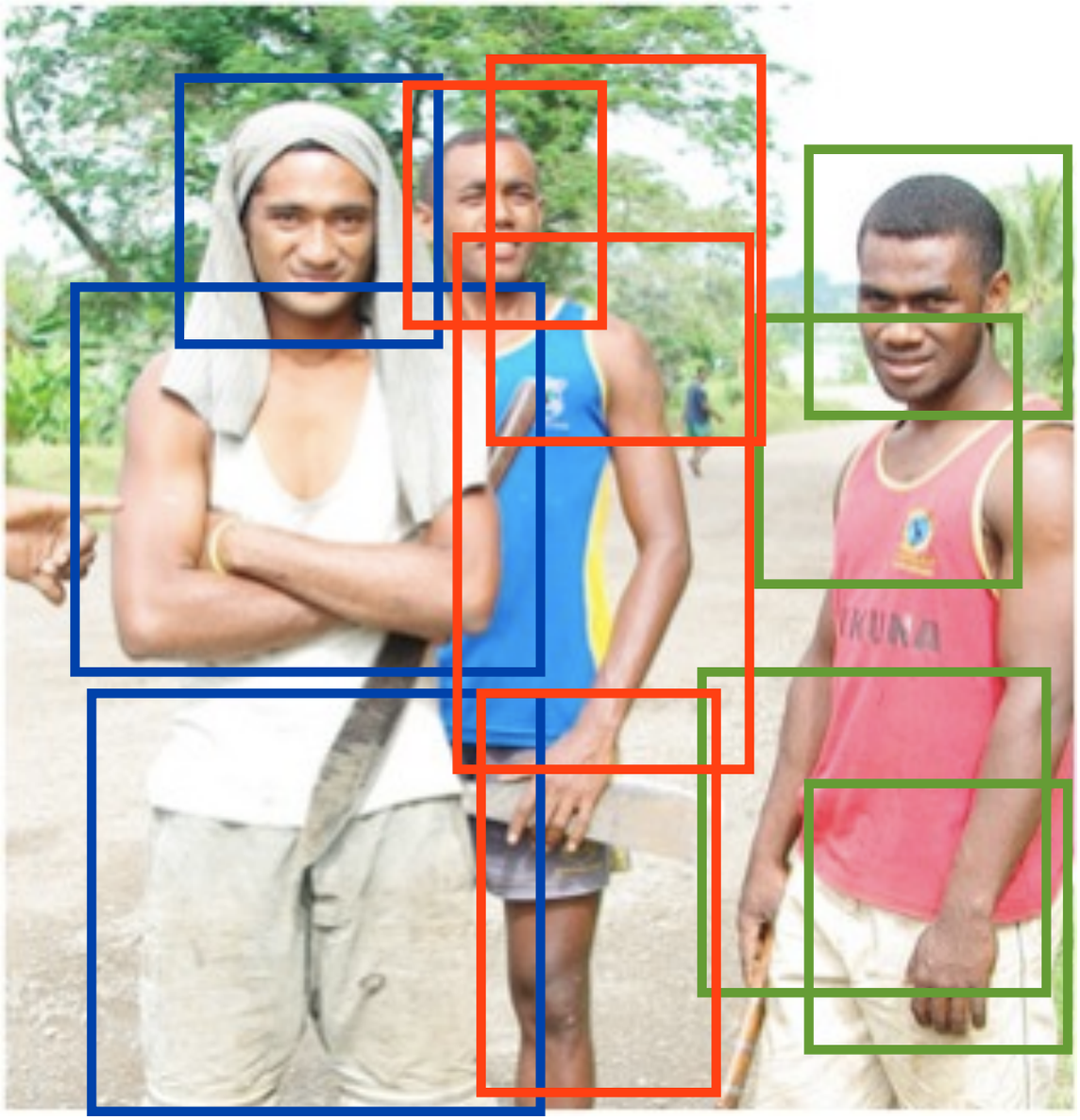
A *poselet*-based approach for attribute recognition

A *poselet*-based approach for attribute recognition



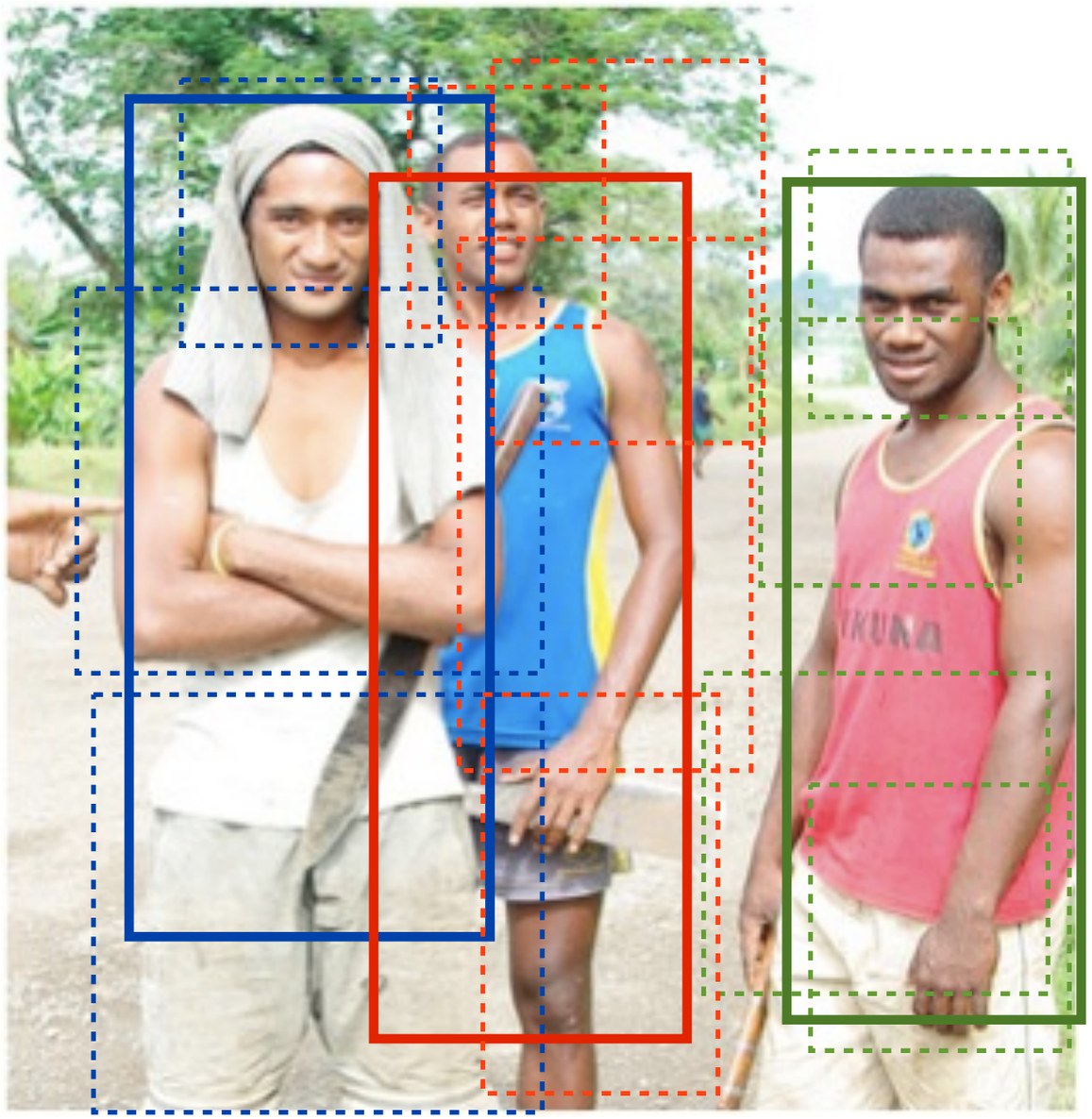
Find *poselet* activations

A *poselet*-based approach for attribute recognition



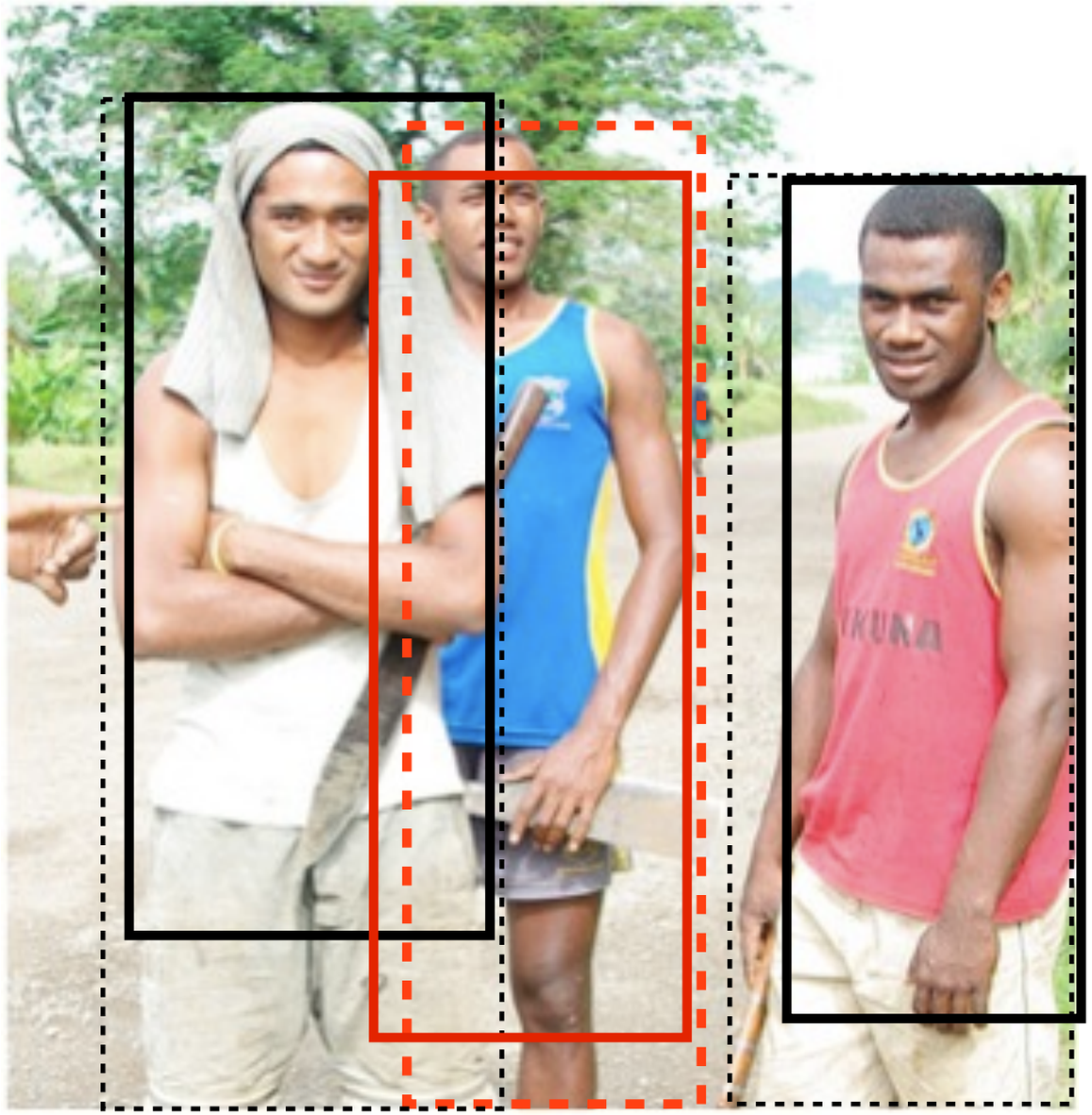
Cluster *poselet* activations

A *poselet*-based approach for attribute recognition



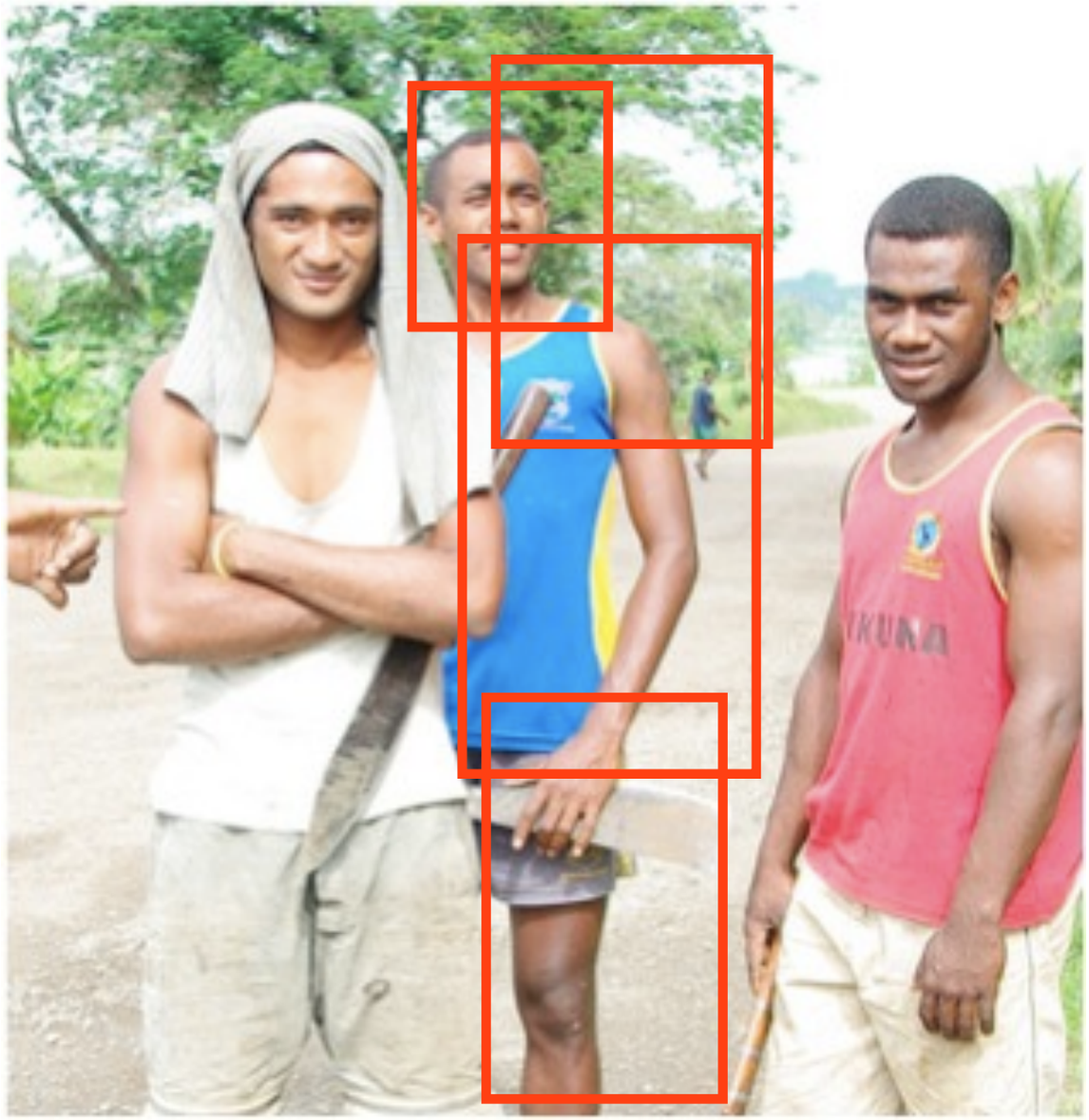
Predict *bounds* from *activations*

A *poselet*-based approach for attribute recognition



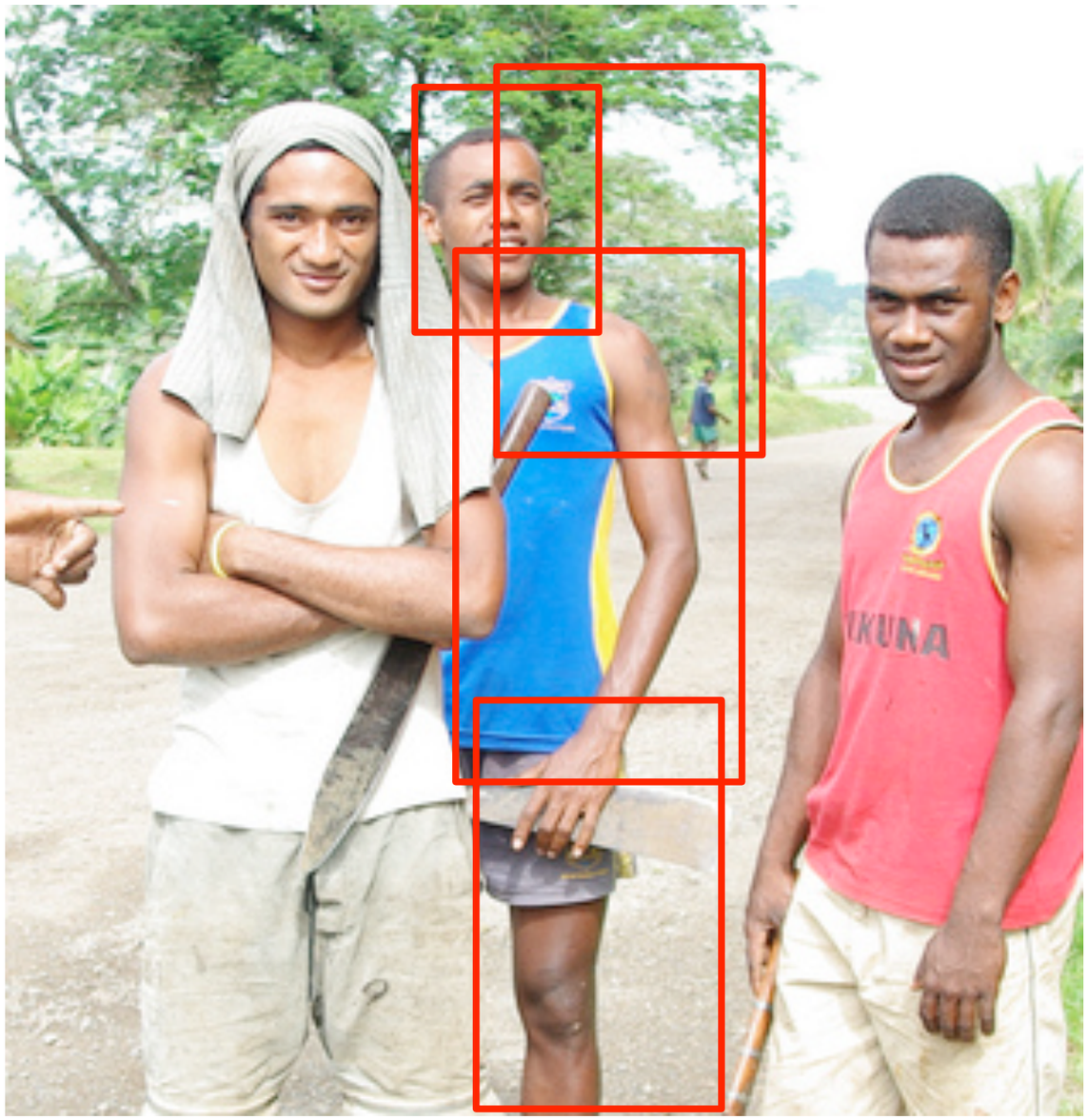
match *predicted* bounds to *ground-truth* bounds
max-flow in a bipartite graph

A *poselet*-based approach for attribute recognition

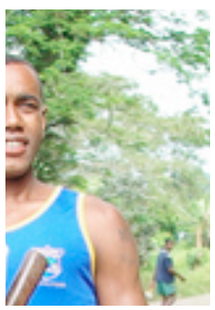


obtain *poselet* activations corresponding to the desired person

Start with *poselet* activations

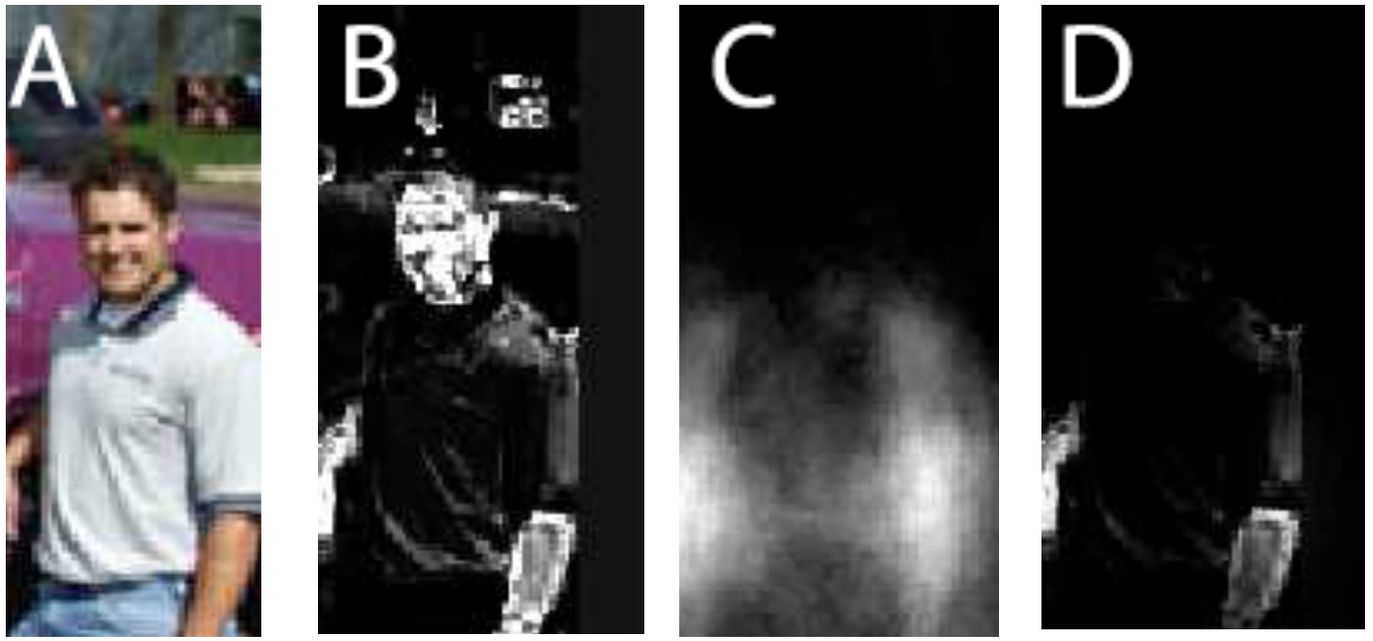


Poselet Activations



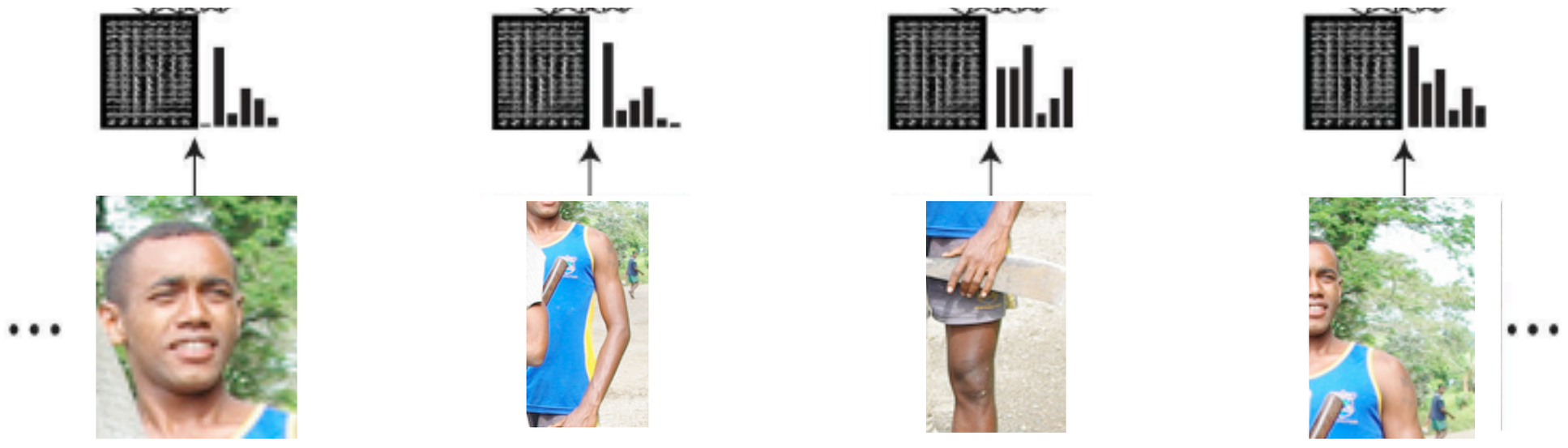
Features

- pyramid HOG
- LAB histogram
- skin color features



Features

Poselet Activations



Attribute classification overview

Poselet-level classifier

Features

Poselet Activations

Estimate attribute from each poselet

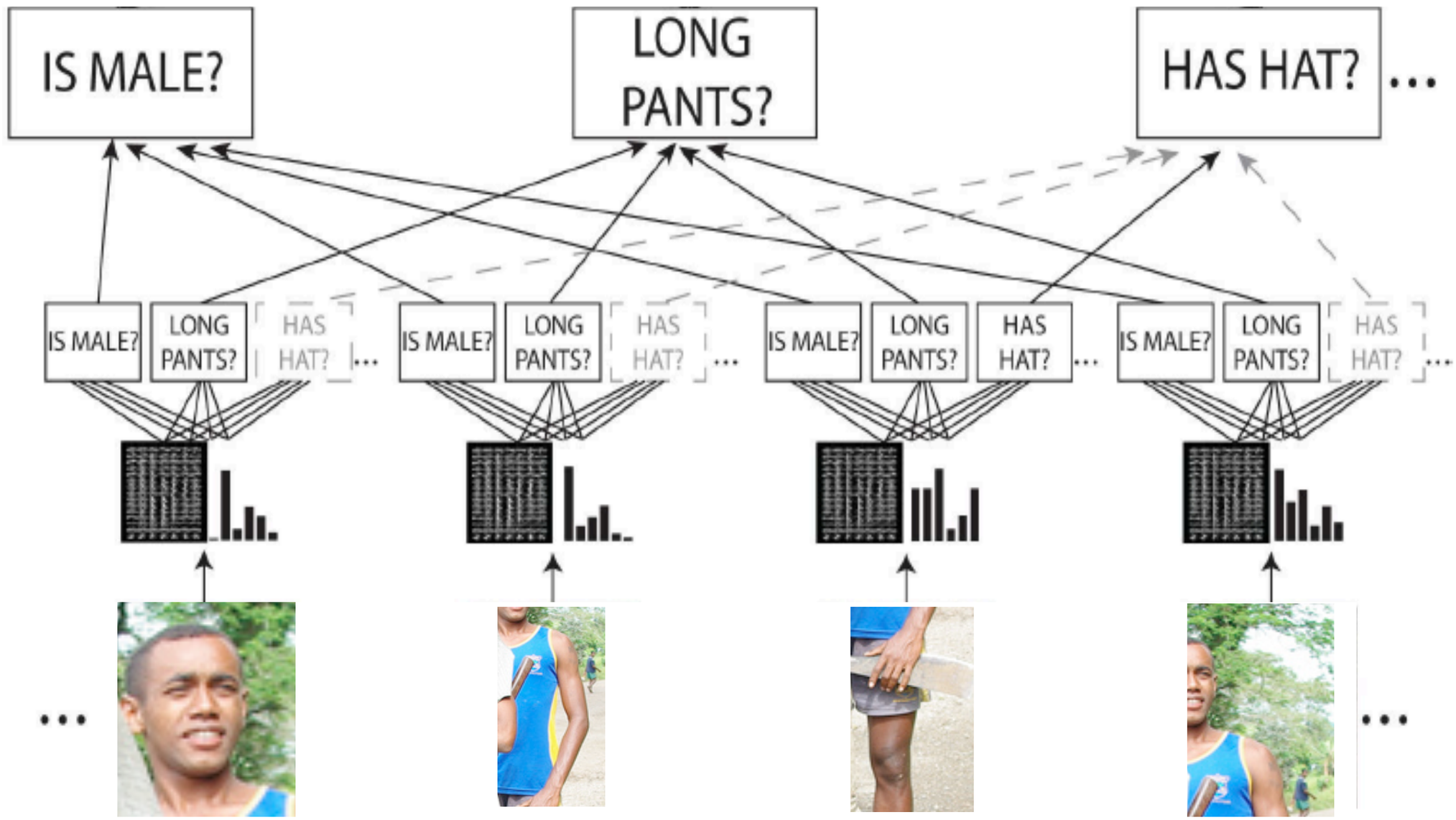
Attribute classification overview

Person-level classifier

Poselet-level classifier

Features

Poselet Activations



Combine evidence from all poselets

Attribute classification overview

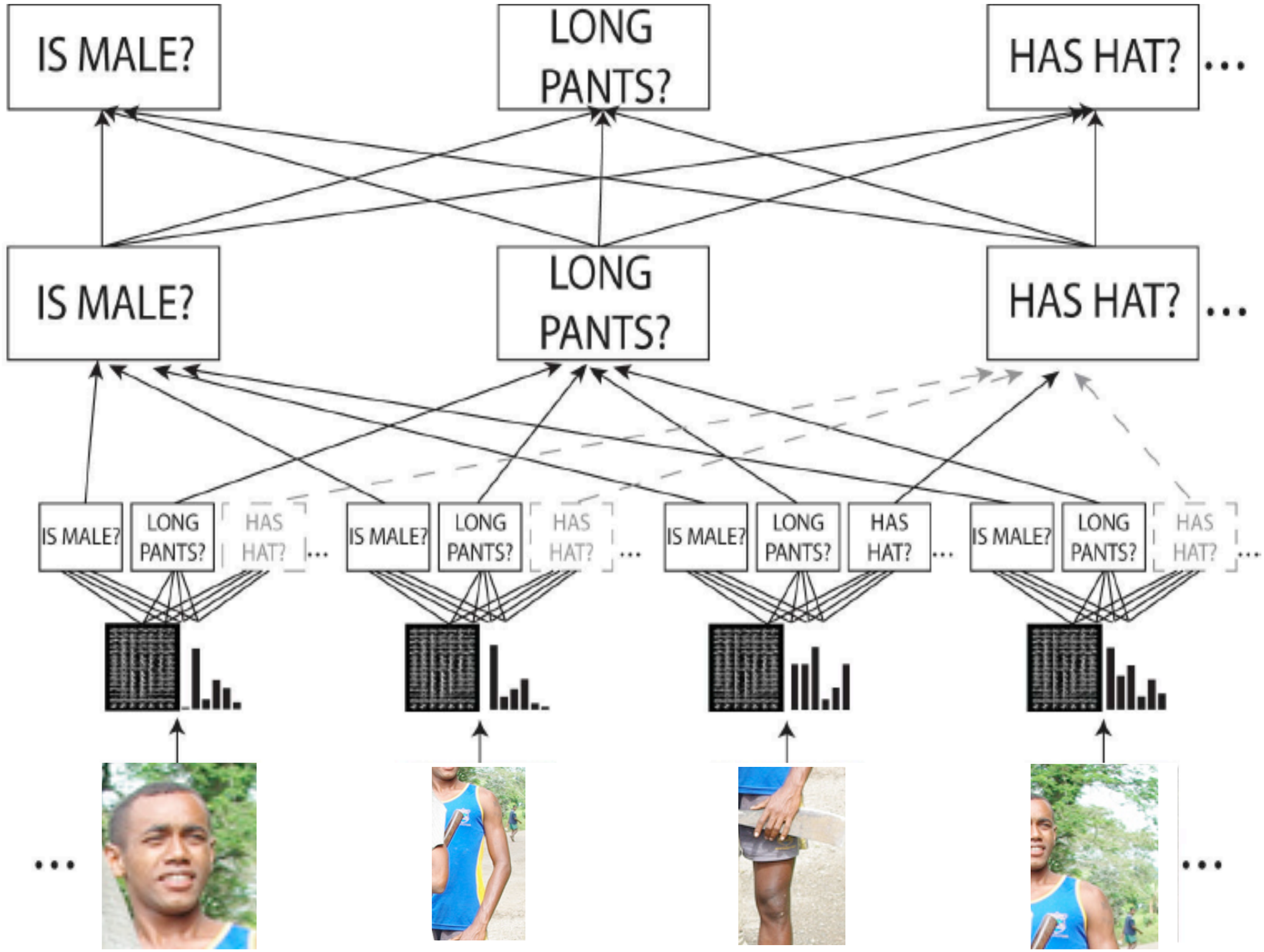
Context re-scoring

Person-level classifier

Poselet-level classifier

Features

Poselet Activations



Re-score attributes based on other attribute predictions

Any questions?

- ... before we proceed to experimental evaluation.

Our dataset

- Image source : PASCAL VOC 2010 *trainval* images for the person category (high-resolution equivalents) + H3D dataset
- Annotations collected on Amazon Mechanical Turk
- Dataset details:
  - ~8000 person instances (4000 *train*, 4000 *test*)
  - 9 binary attributes: *is-male*, *has-long-hair*, *has-glasses*, *has-hat*, *has-long-sleeves*, *has-t-shirt*, *has-long-pants*, *has-jeans*, *has-shorts*
- Dataset is publicly available at :
  - *http://www.cs.berkeley.edu/~lbourdev/poselets*

Visual search on the *test* set

Visual search on the *test* set

wears hat

Visual search on the *test* set

wears hat

female

Visual search on the *test* set

Visual search on the *test* set

has long hair

Visual search on the *test* set

has long hair

wears glasses

Visual search on the *test* set

Visual search on the *test* set

wears shorts

Visual search on the *test* set

wears shorts

has long sleeves

Visual search on the *test* set

wears shorts

has long sleeves

Visual search on the *test* set

Visual search on the *test* set

doesn't have long sleeves

Visual search on the *test* set

doesn't have long sleeves

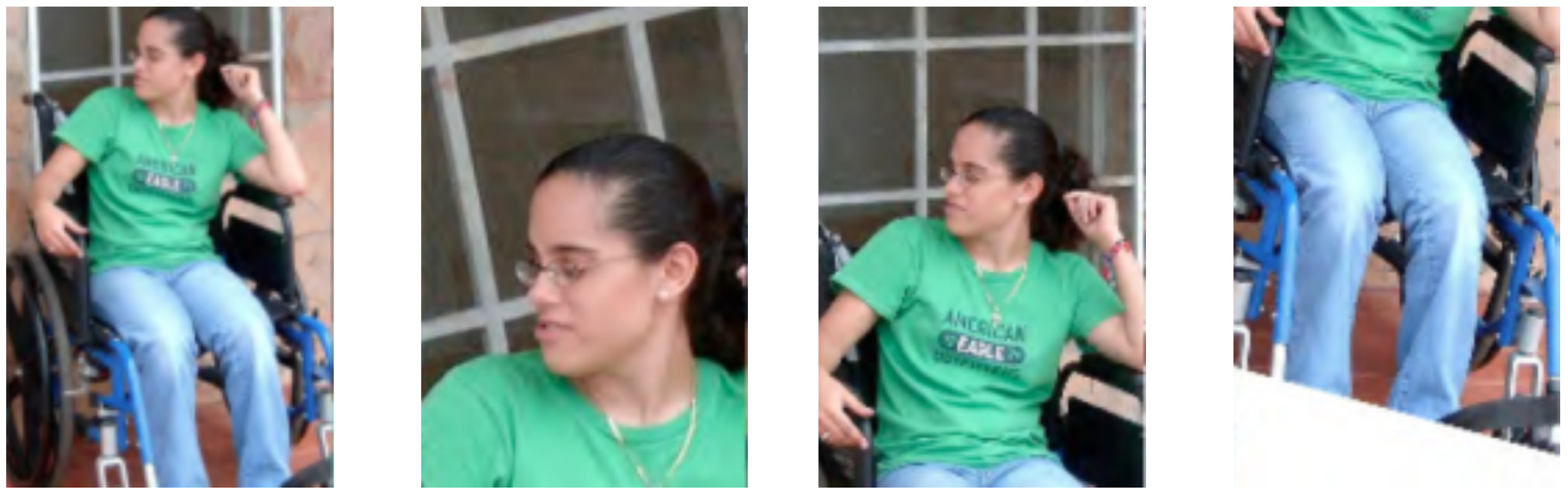
Baseline algorithms

- How important is the decomposition using pose?
  - i.e., how well can we do using a classifier trained on the entire bounding box
- How important is the information from multiple parts?
  - suppose we were given a perfect face detector, can we do well on the task?

Baseline algorithms

- How important is the decomposition using pose?
  - i.e., how well can we do using a classifier trained on the entire bounding box
- How important is the information from multiple parts?
  - suppose we were given a perfect face detector, can we do well on the task?

train on ground truth



Full view Head zoom Upper body Legs

HOG + Spatial pyramid matching (Lazebnik et al., CVPR 06)

Most informative *region* for attribute recognition

Attribute	**Prior**	**head**	**lower**	**upper**	**bbox**
is male	59.3	**74.9**	63.9	71.3	68.1
has long hair	30.0	**60.1**	34.0	45.2	40.0
has glasses	22.0	**33.4**	22.6	25.5	25.9
has hat	16.6	**53.0**	24.3	32.3	35.3
has t-shirt	23.5	**32.2**	25.4	30.0	30.6
has long sleeves	49.0	53.4	52.1	56.6	**58.0**
has shorts	17.9	22.9	24.8	22.9	**31.4**
has jeans	33.8	38.5	38.5	34.6	**39.5**
has pants	74.7	79.9	80.4	76.9	**84.3**
Mean AP	*36.3*	*49.81*	*40.66*	*43.94*	*45.91*

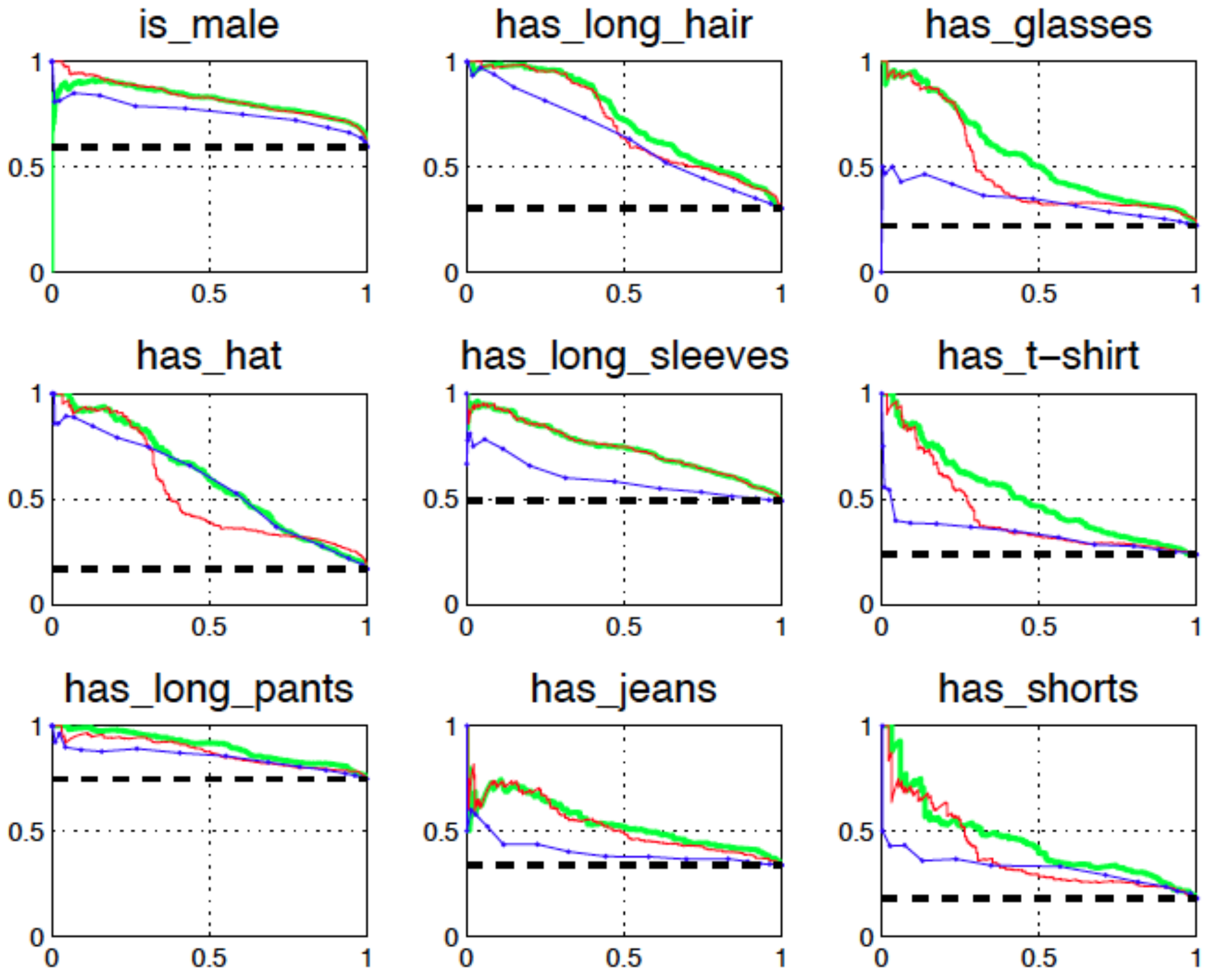
Results (precision vs. recall curves)

Label frequency

SPM

No context

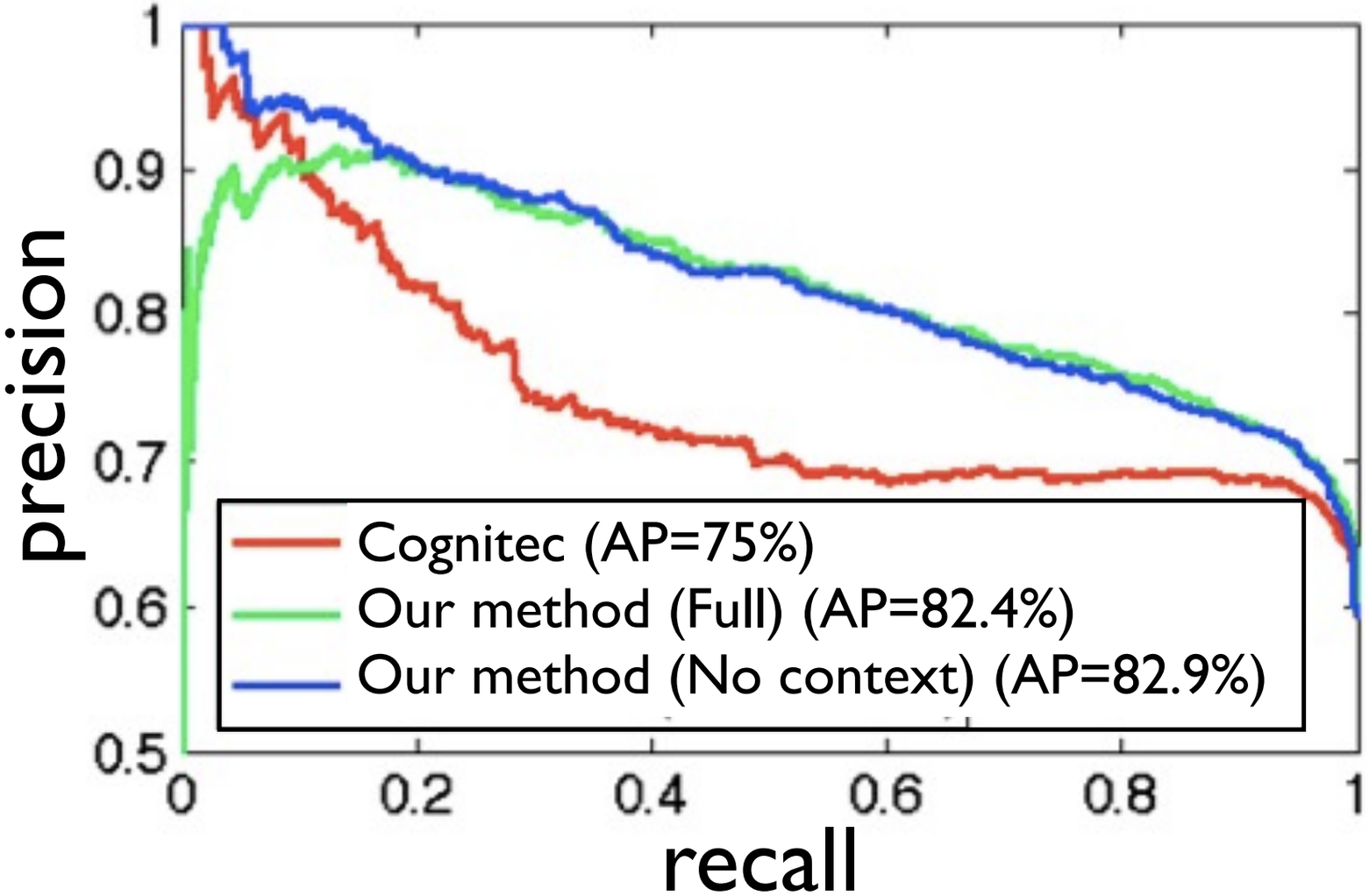
Full Model



we pick the best baseline algorithm

State-of-the-art on *gender recognition*

- We achieve AP=**82.4%,** out-perform Cognitec's gender recognizer (a commercial face recognition system)
- We out-perform any frontal face-based gender recognizer :
  - **61%** of our data has frontal faces. A perfect result on frontal faces *only*, achieves AP=**80.5%**



Confusions

Confusions

men most confused to be *women*

Confusions

men most confused to be *women*

long hair

Confusions

long hair

men most confused to be *women*

women most confused to be *men*

Confusions

men most confused to be *women*

long hair

baseball hat

women most confused to be *men*

Confusions

men most confused to be *women*

women most confused to be *men*

Confusions

Confusions

non t-shirt most confused to be *t-shirt*

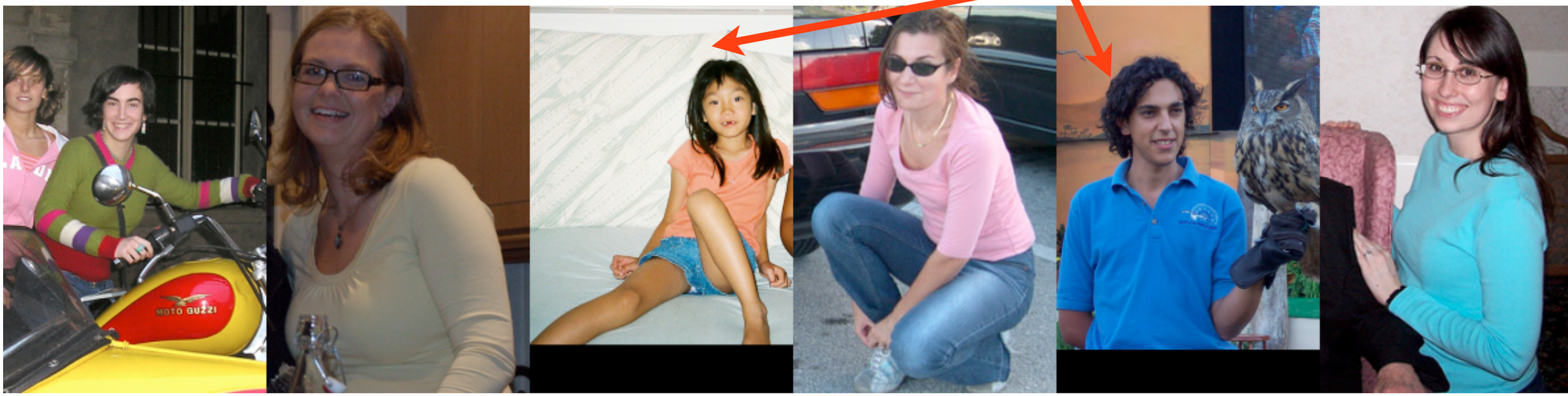




Confusions

non t-shirt most confused to be *t-shirt*

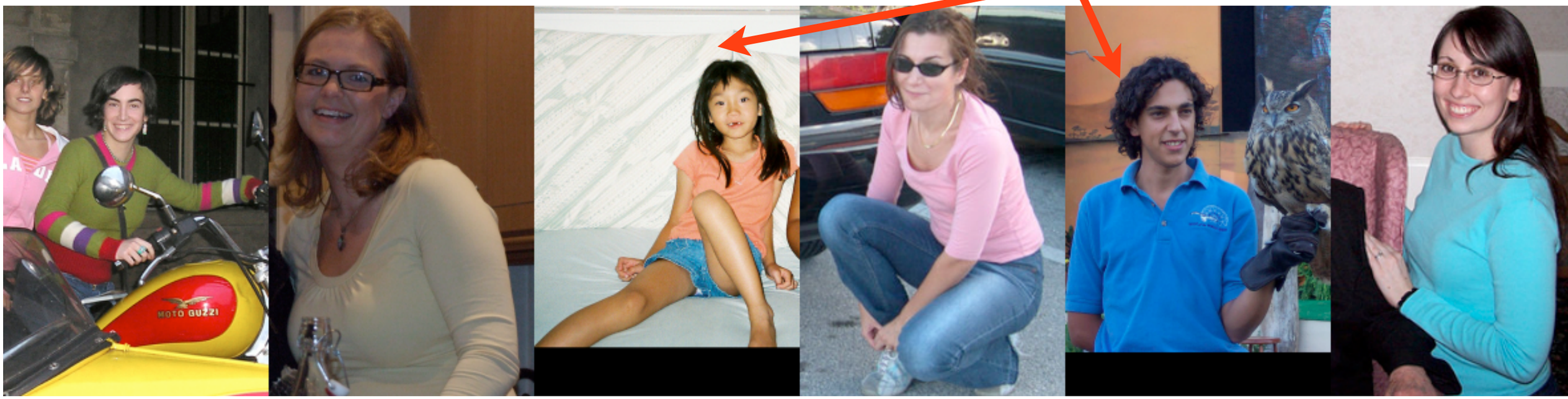
annotation errors



Confusions

annotation errors

non t-shirt most confused to be *t-shirt*

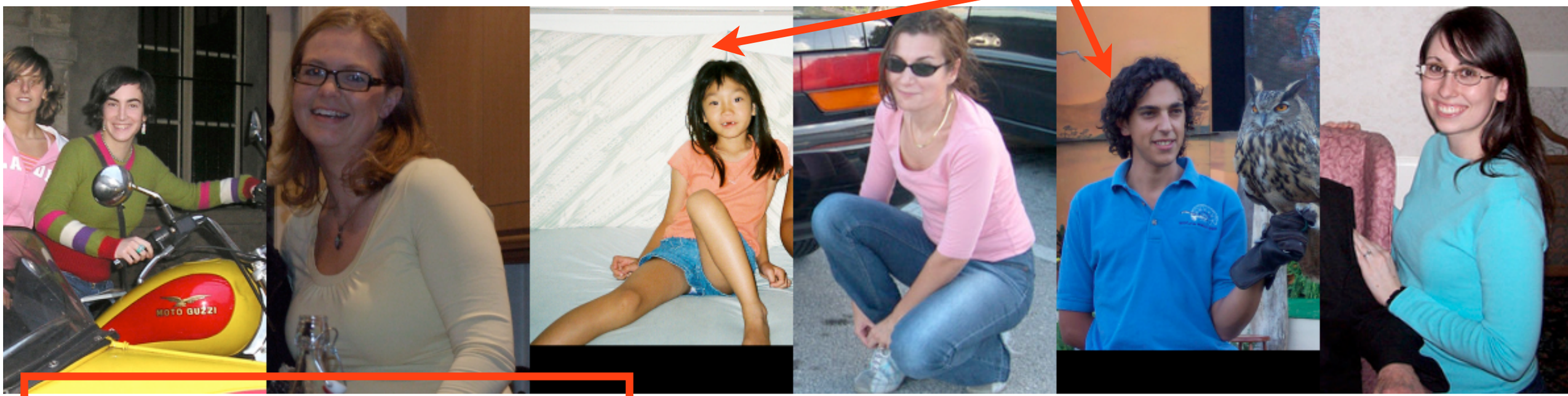


short pants most confused to be *long pants*

Confusions

annotation errors

non t-shirt most confused to be *t-shirt*



are these pants short?

short pants most confused to be *long pants*

Confusions

annotation errors

non t-shirt most confused to be *t-shirt*






are these pants short?

wrong person

short pants most confused to be *long pants*






Confusions

non t-shirt most confused to be *t-shirt*

annotation errors

are these pants short?

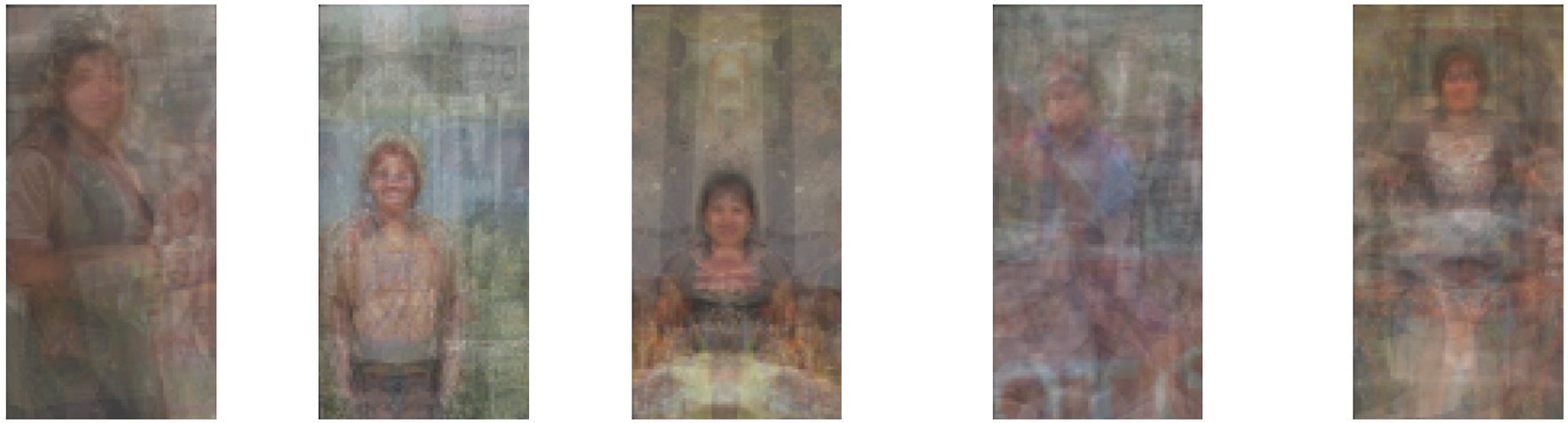
wrong person

occlusion

short pants most confused to be *long pants*

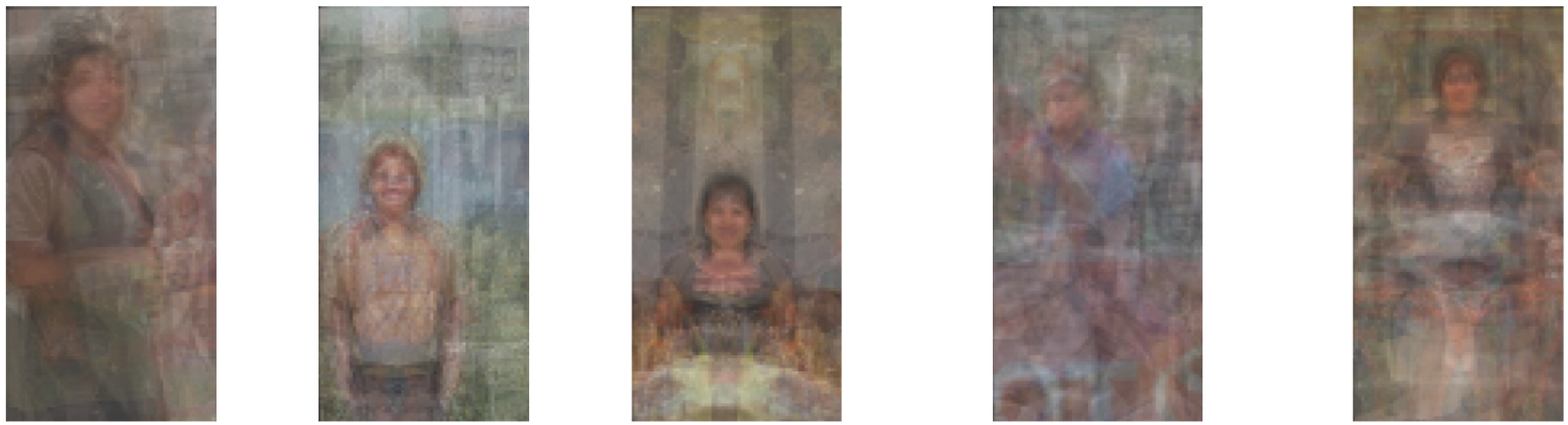
Most informative *poselets* for attribute prediction

Gender

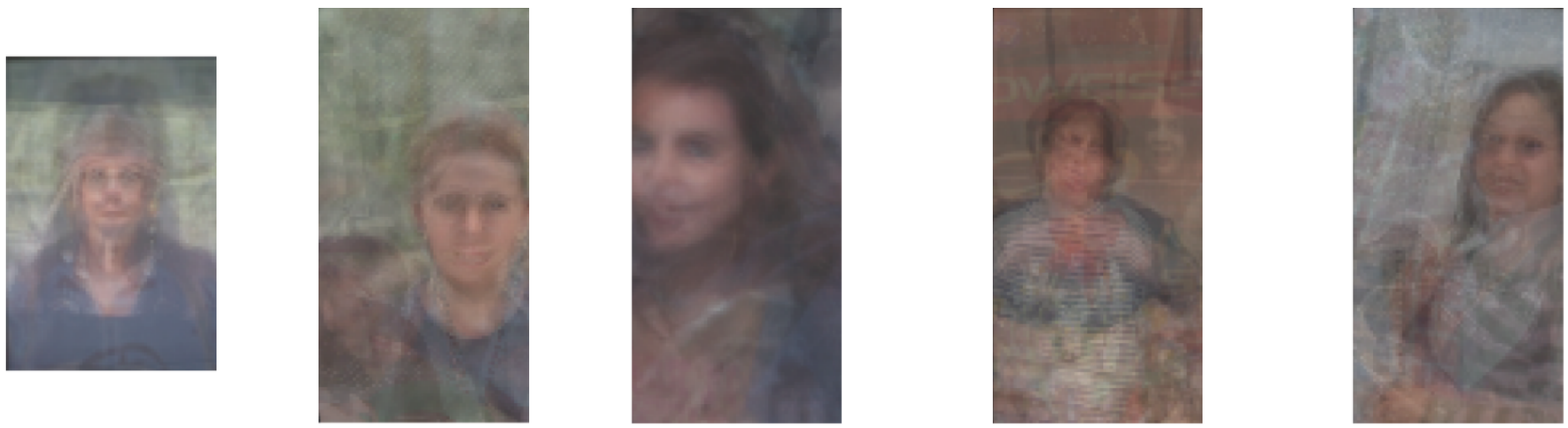


Most informative *poselets* for attribute prediction

Gender



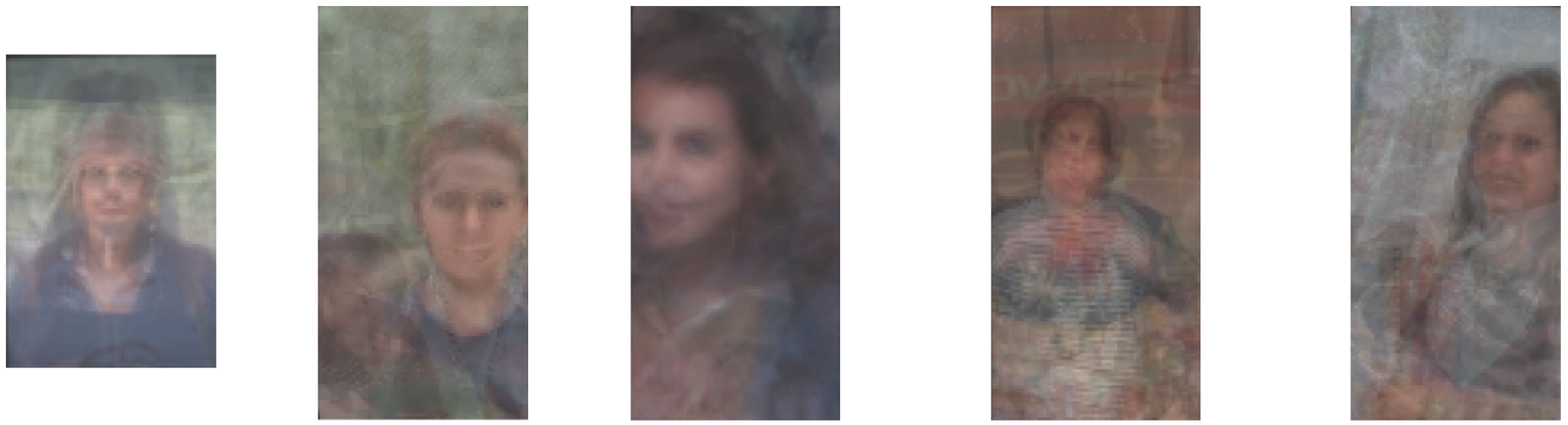
Long hair



Most informative *poselets* for attribute prediction

Gender

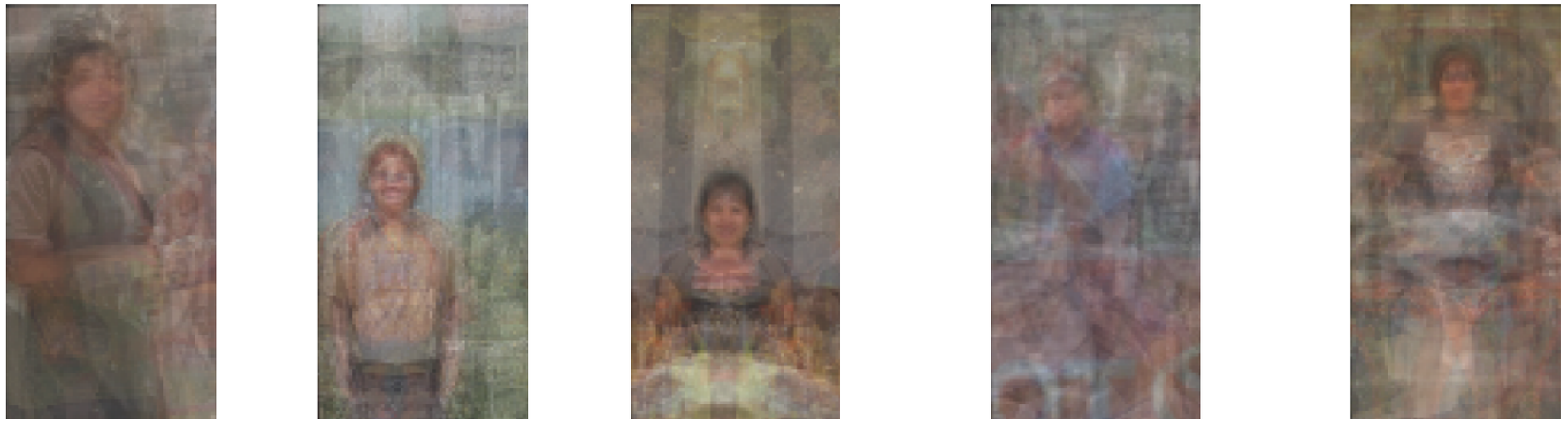
Long hair



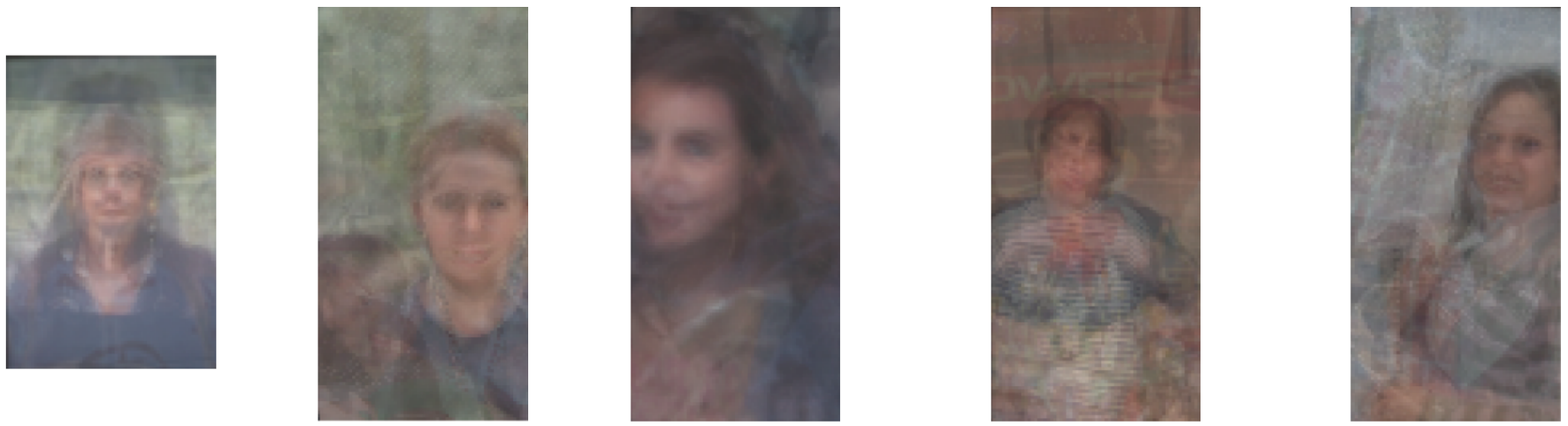
Glasses

Most informative *poselets* for attribute prediction

Gender



Long hair



Glasses

increasing zoom

Describing people

"A man with short hair, glasses, short sleeves and shorts"

Describing people

"A person with long pants"

Describing people

"A man with short hair and long sleeves"

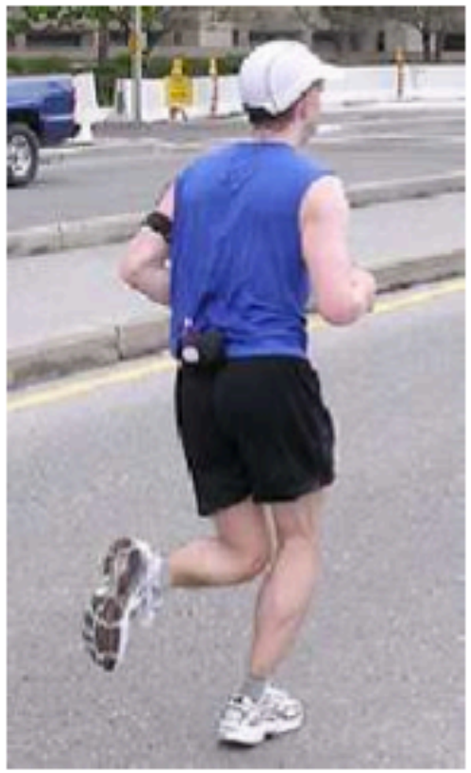
Any questions on attribute recognition?

Action recognition

(or recognizing unusual poses)

Action recognition

what are the people doing?




unusual and *characteristic poses*; which means face and pedestrian detectors may not work well

actions = discriminative pose + appearance

pose

actions = discriminative pose + appearance

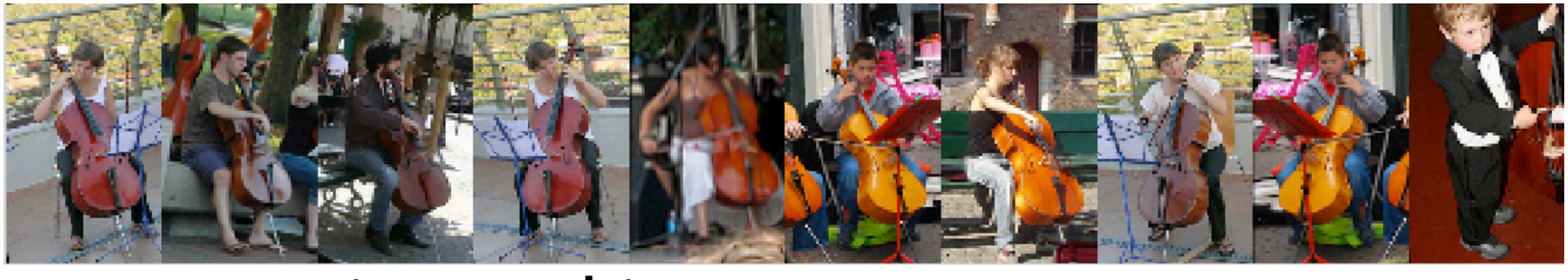
pose

pose + action

actions = discriminative pose + appearance

pose

pose + action



pose + action + object

PASCAL VOC action classification challenge

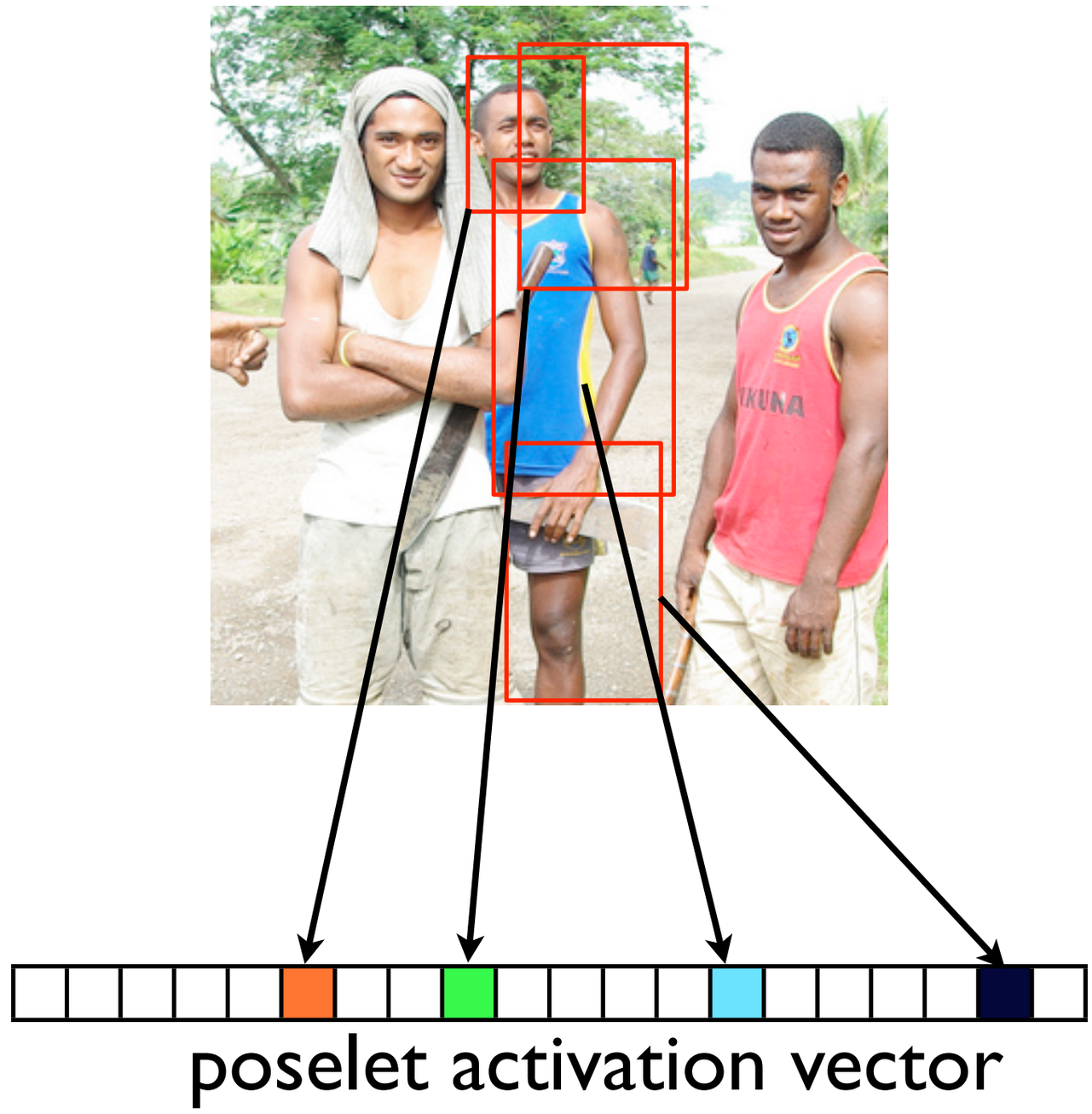
- Input : *input image* and *bounding box* of all persons in an image
- Output : predict which of the 9 actions are being performed

Poselets for action classification

- Train a large number of *poselets* using information from:
  - pose only
  - pose + action
  - pose + action + object

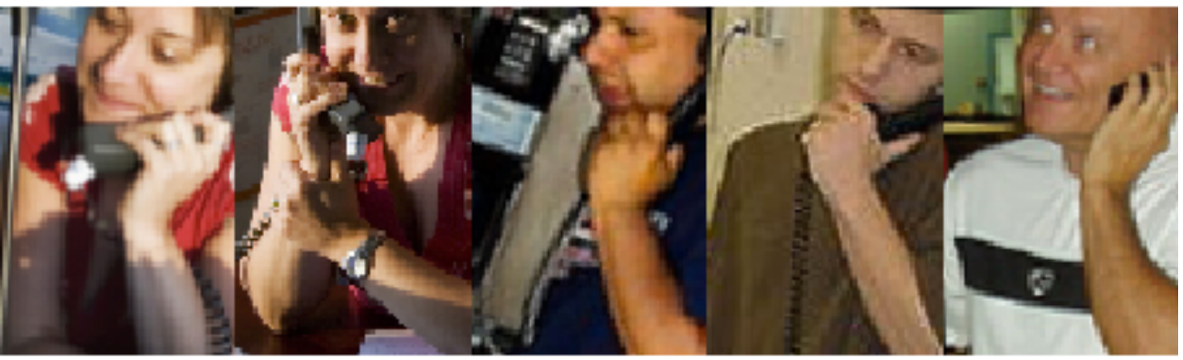
Poselets for action classification

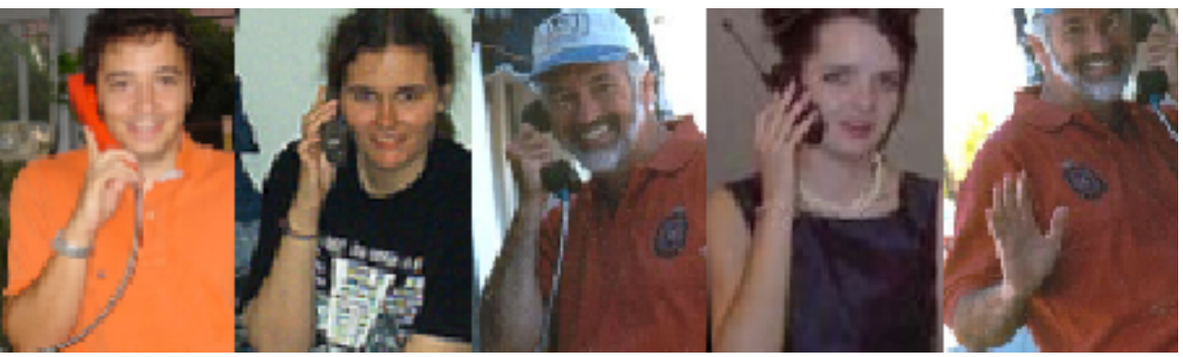
- Train a large number of *poselets* using information from:
  - pose only
  - pose + action
  - pose + action + object
- Action classification
  - *poselet activation vector* - score of all the poselet activations that *belong* to the person
  - linear SVMs trained in 1-vs-all manner



Examples of learned *poselets*

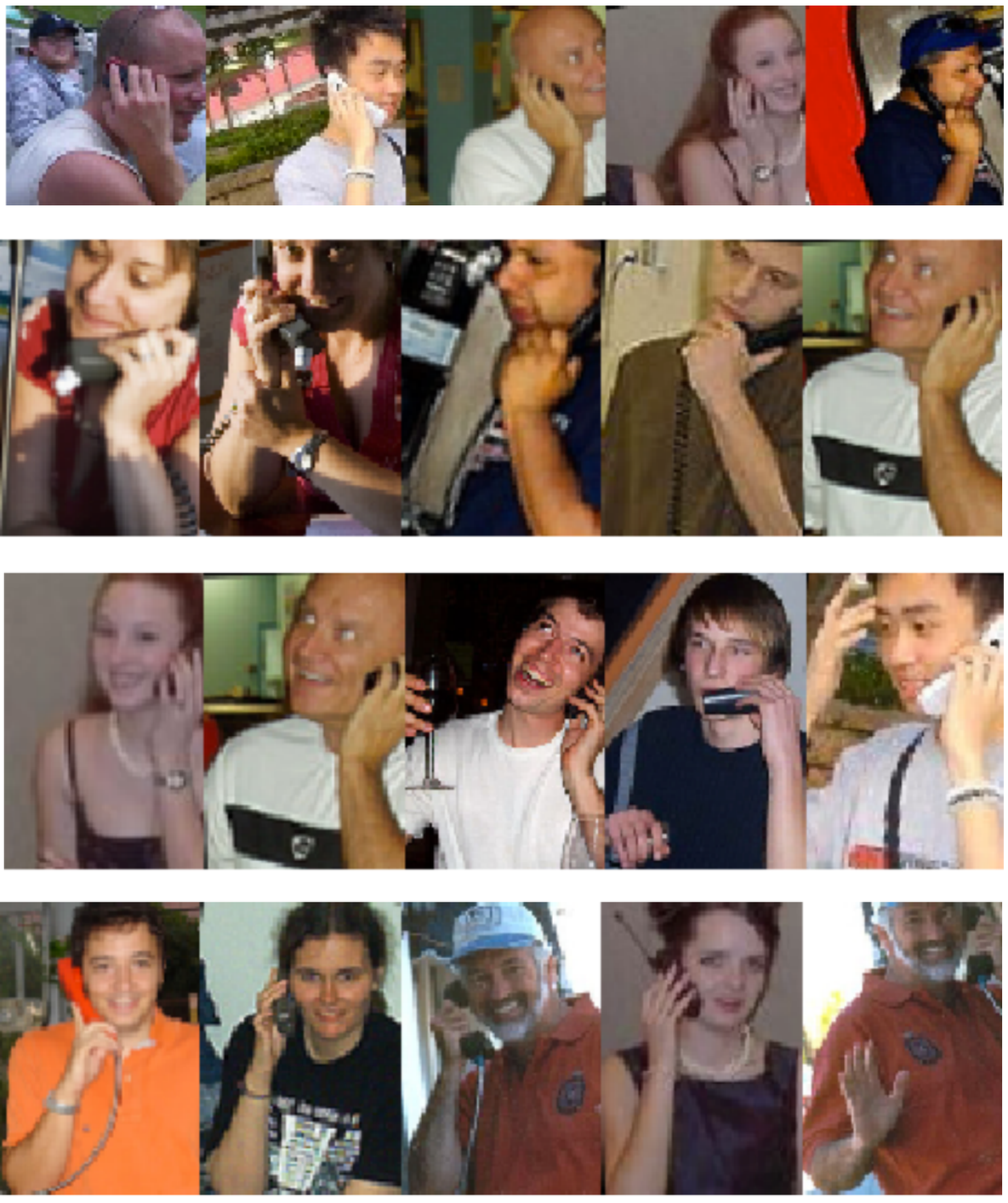
phoning





Examples of learned *poselets*

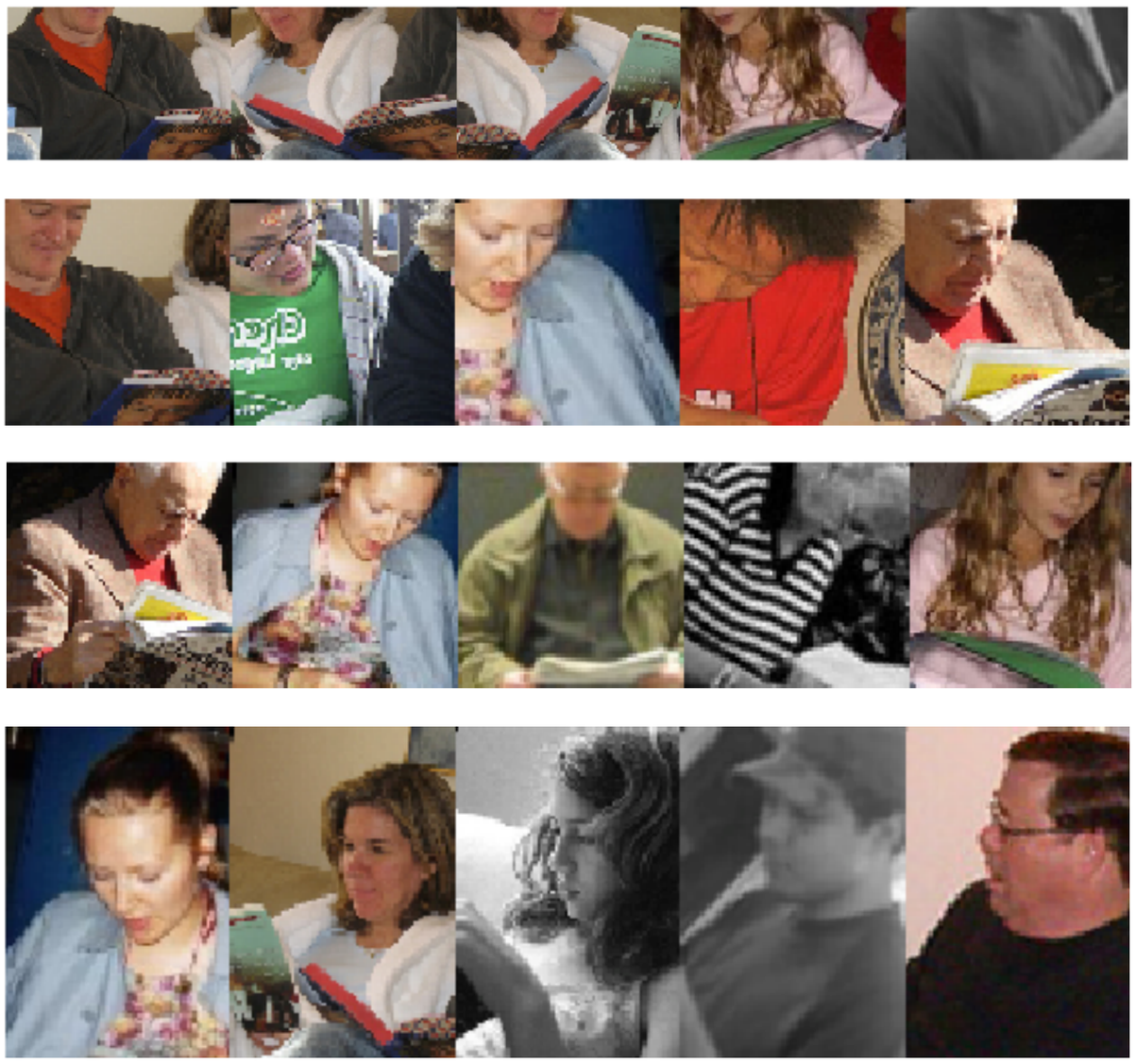
phoning



playing instrument

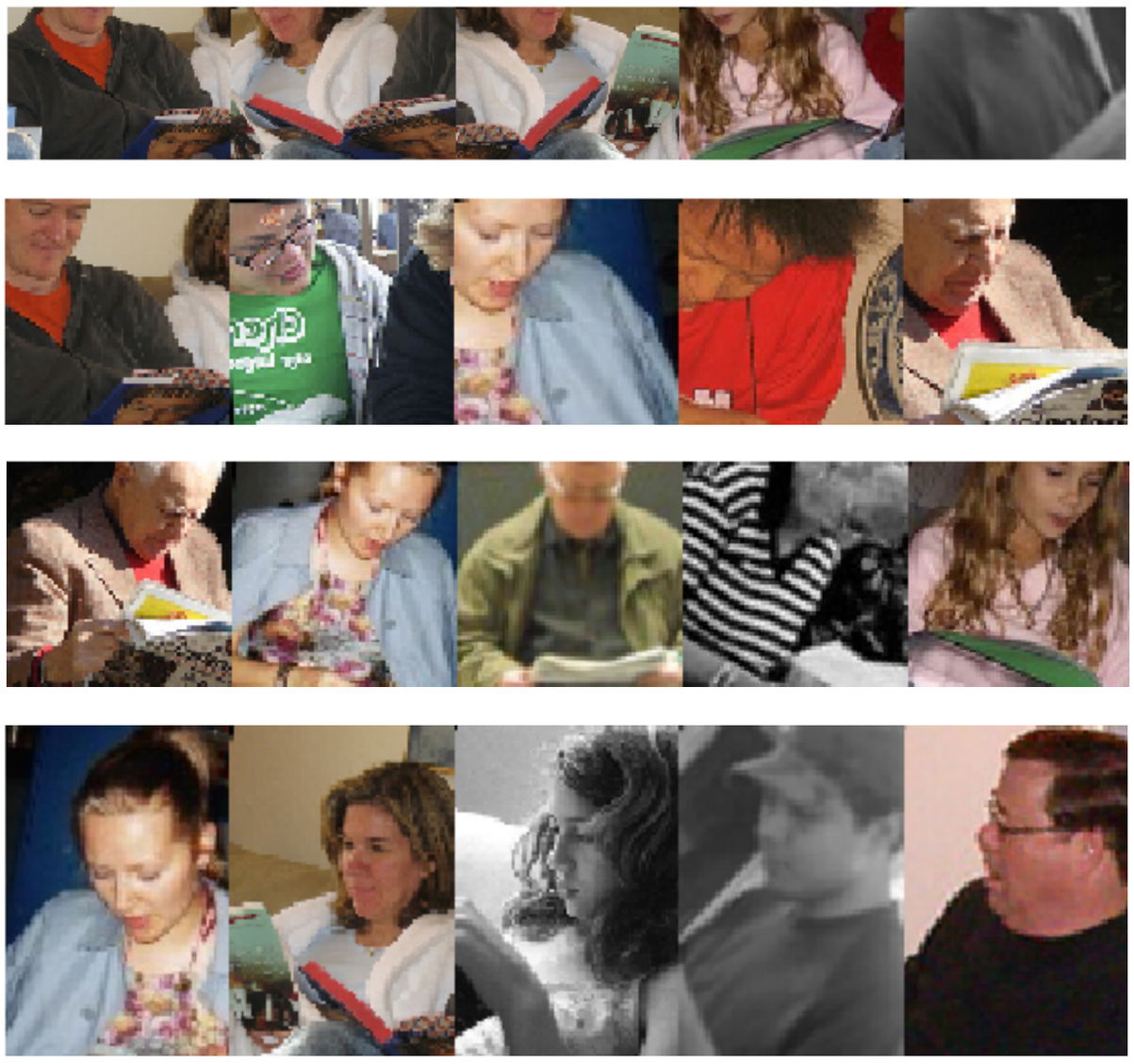
Examples of learned *poselets*

reading

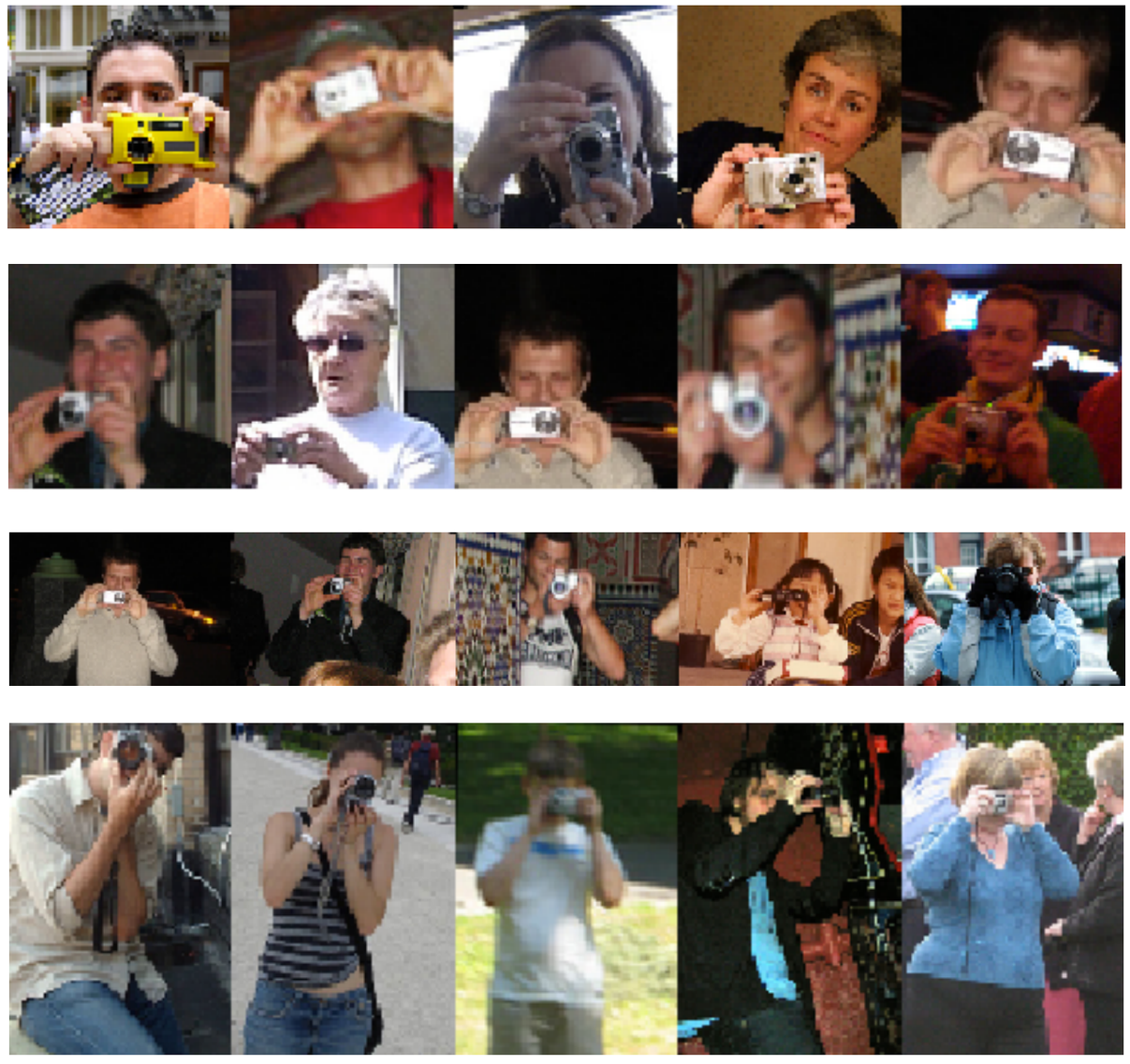


Examples of learned *poselets*

reading

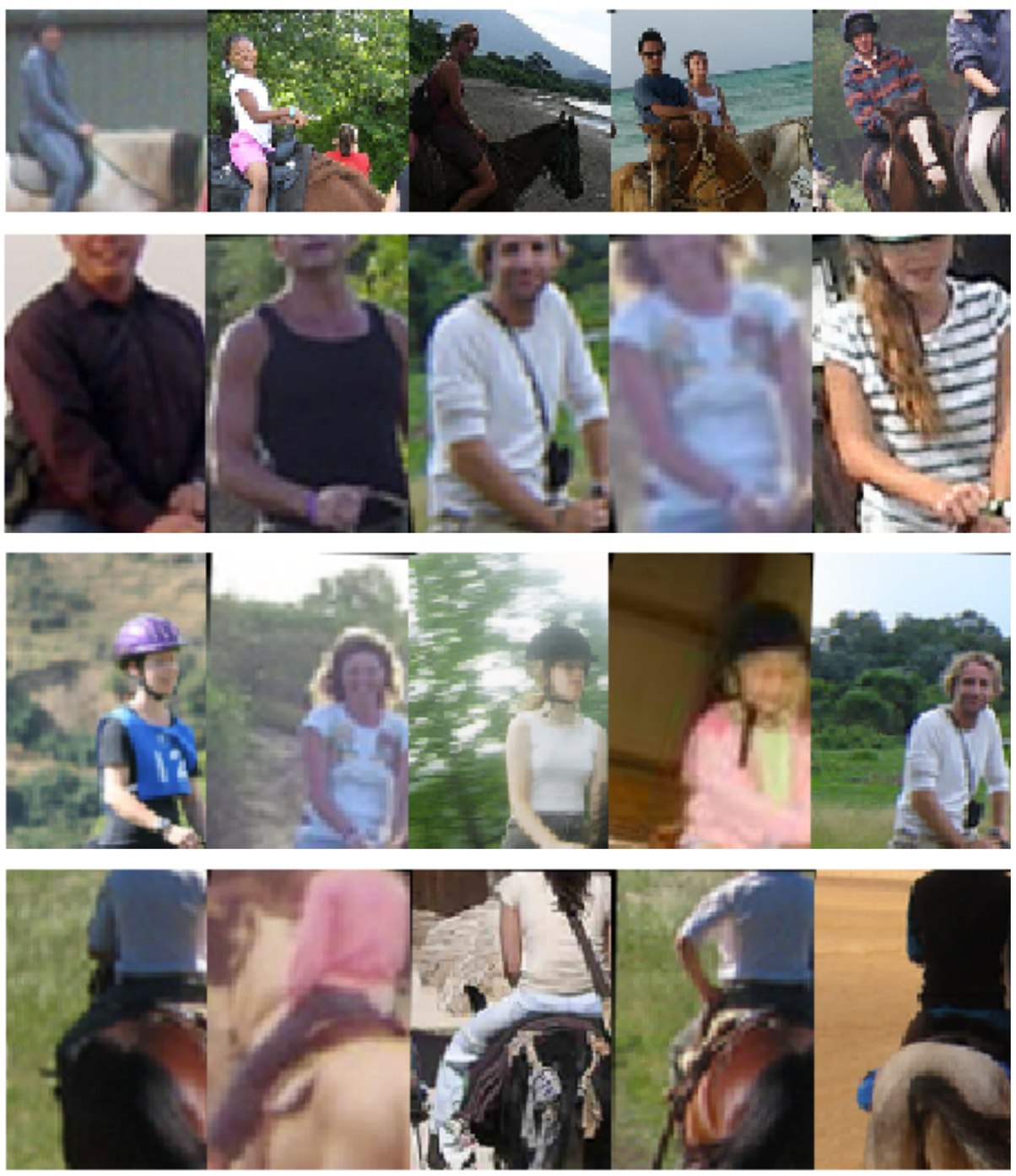


taking photo



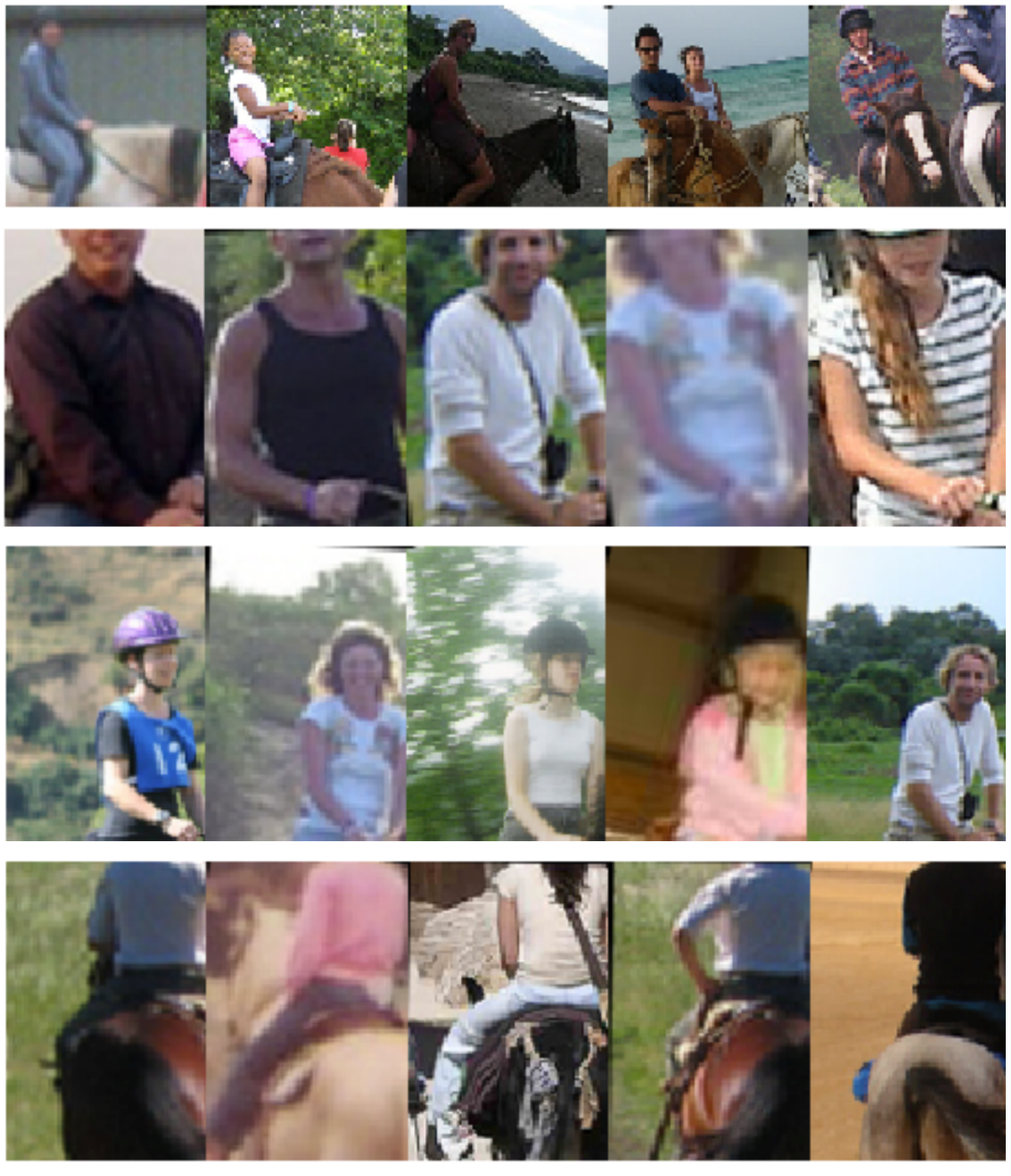
Examples of learned *poselets*

riding horse



Examples of learned *poselets*

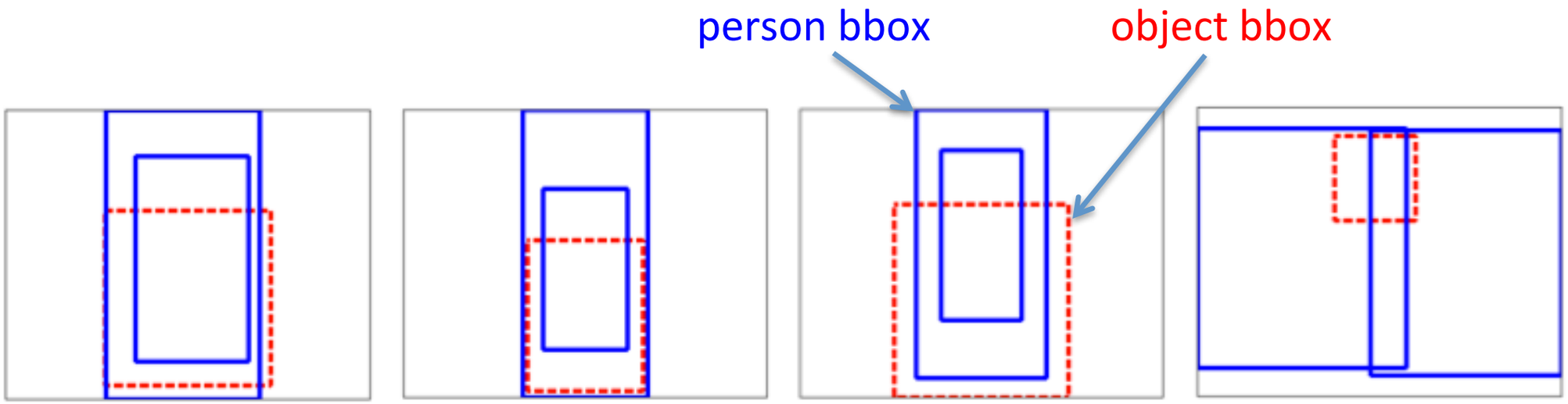
riding horse



running

Object context

spatial model of person-object interaction



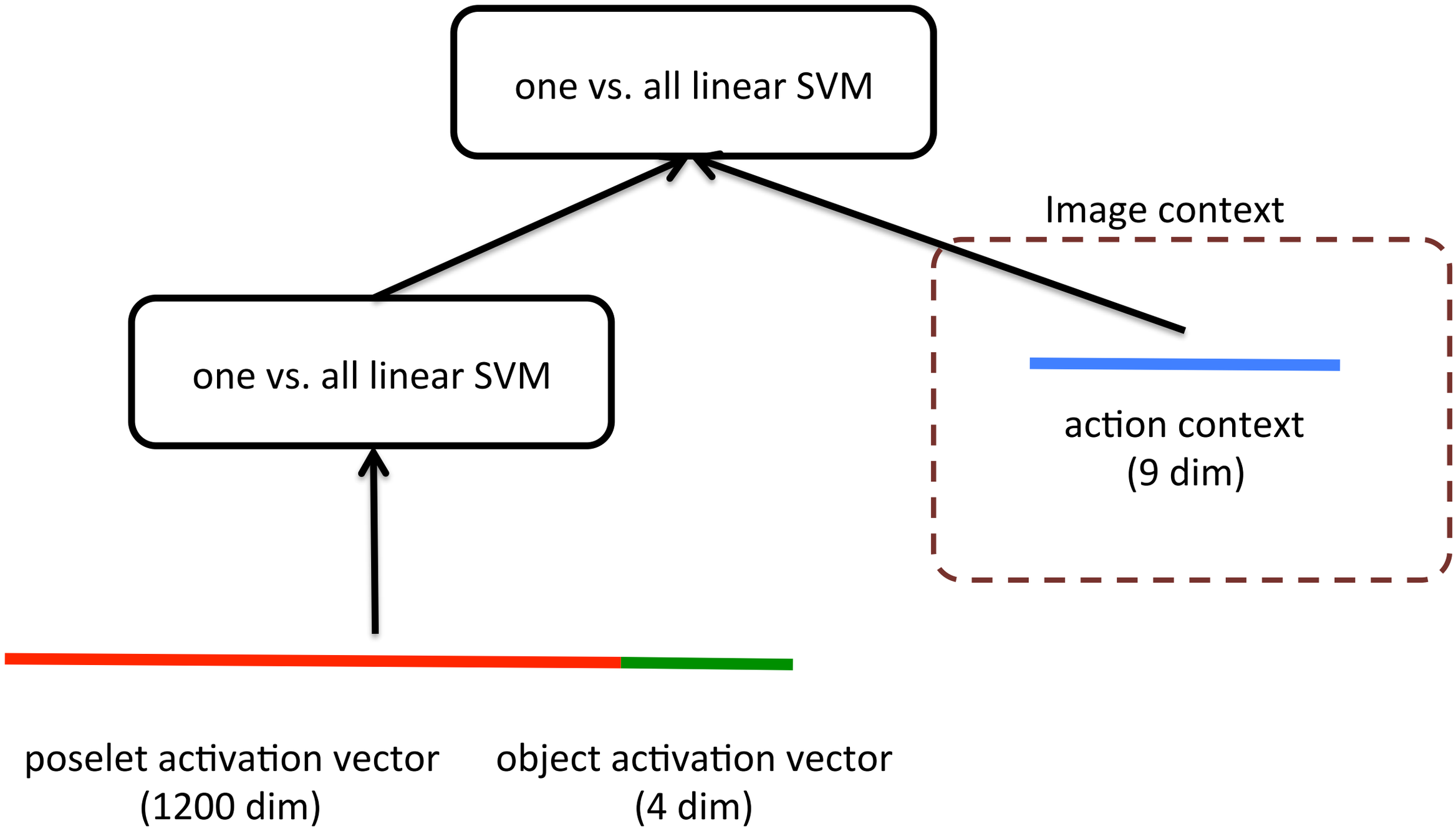
motorbike *bicycle* *horse* *tvmonitor*

Action context

what are other people doing in the scene?

group activities

Overall action classification



Action classification results

Object Context

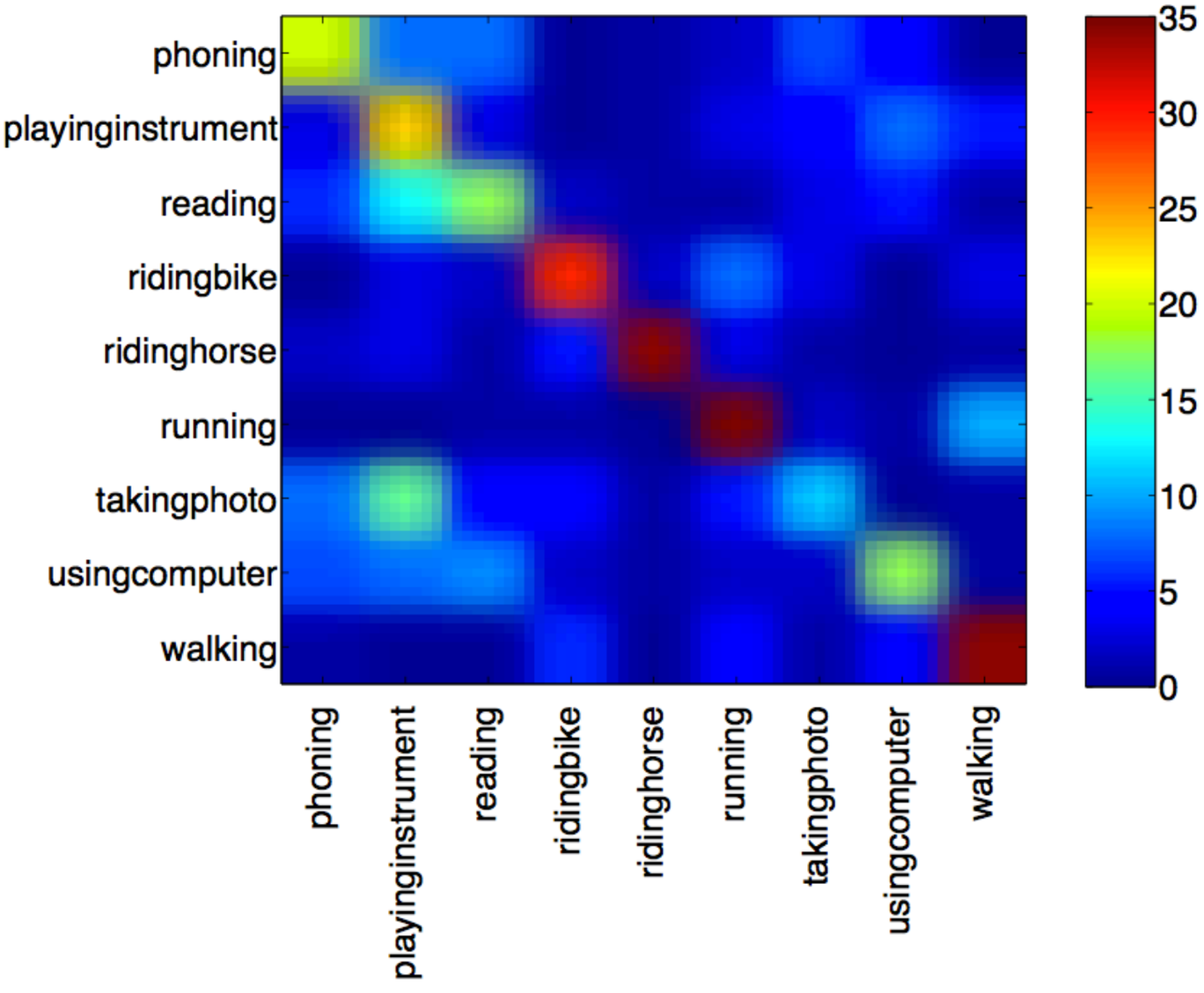
Image Context

PASCAL VOC 2010 challenge

	Validation			Test
category	PAV	w/ OAV	w/ C	w/ C
phoning	63.3	62.0	62.0	49.6
playinginstrument	44.2	44.4	45.6	43.2
reading	37.4	44.4	44.3	27.7
ridingbike	62.0	84.7	85.5	83.7
ridinghorse	91.1	97.7	97.5	89.4
running	82.4	84.1	86.0	85.6
takingphoto	21.1	22.9	24.6	31.0
usingcomputer	54.2	64.9	64.3	59.1
walking	82.0	83.6	80.8	67.9
average	59.8	65.3	65.6	59.7

Confusion matrix

class confusion matrix



Poselets: 59.7
INRIA_SPM_HT : 60.1
CVC_BASE : 60.3

"**CVC_BASE** : …. Standard BoW model over multiple features including PHOG, grayscale SIFT and (various) color SIFT descriptors. Foreground/background modeled separately, spatial pyramid over several features for foreground representation…."

"**INRIA_SPM_HT** : ..Spatial Pyramids on the bounding box, on the image and a hough transform for taking into account the object-person interactions for bicycle, horse and tvmonitor…."

poselet activation vector is a compact representation pose and appearance

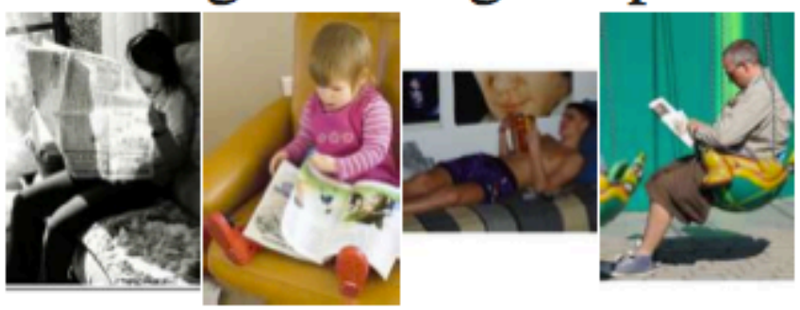
Maji, Bourdev & Malik, CVPR'11

Some confusions

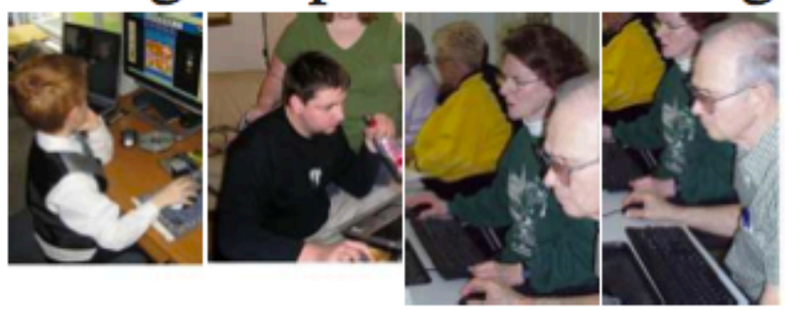
phoning → takingphoto

takingphoto → phoning

reading → usingcomputer



usingcomputer → reading

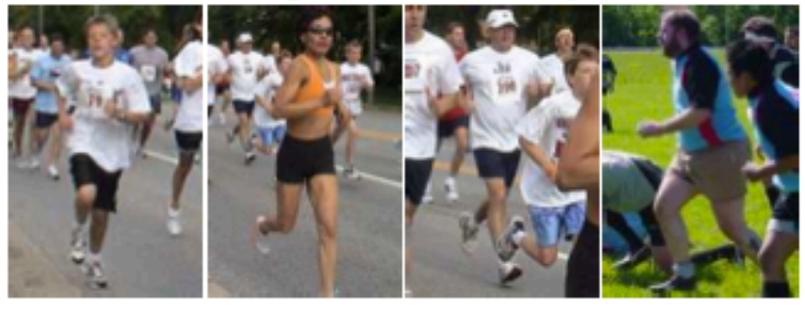


walking → running

running → walking

ridingbike → running

running → ridingbike



Papers Discussed

- N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, CVPR 2005
- L. Bourdev and J. Malik, *Poselets: Body part detectors trained using 3D human pose annotations*, ICCV 2009
- L. Bourdev, S. Maji, T. Brox and J. Malik, *Detecting people using mutually consistent poselet activations*, ECCV 2010
- S. Maji, L. Bourdev and J. Malik, *Action recognition using a distributed representation of pose and appearance*, CVPR 2010
- T. Brox, L. Bourdev, S. Maji and J. Malik, *Object segmentation by alignment of poselet activations to image contours*, CVPR 2011
- L. Bourdev, S. Maji and J. Malik, *Describing people: A poselet-based approach to attribute classification*, ICCV 2011