Human Pose Search using Deep Poselets

Nataraj Jammalamadaka [*]

Andrew Zisserman [§] C. V. Jawahar[*]

[*]CVIT,

IIIT Hyderabad, India

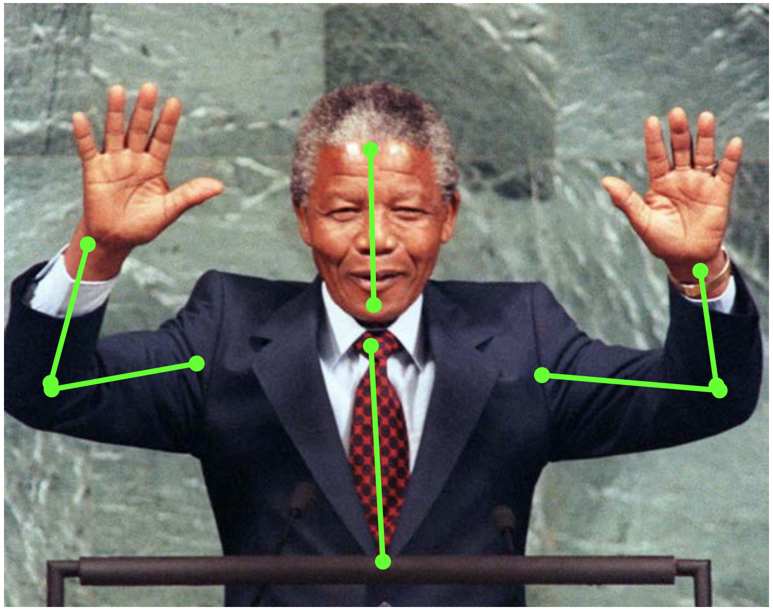
[§] Visual Geometry Group,

Department of Engineering,

University of Oxford

Human Pose: Gesture and action

Walking



Gesturing

Cover Drive

Human pose is a very important precursor to gesture and action

Pose Search: Motivation

Retrieve cover drive shots

Retrieve Bharatanatyam poses

Pose Search: System

Take a query

$$[x_1, \ldots, x_n]$$

Build a feature

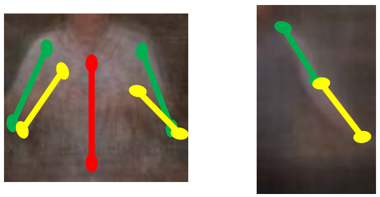
Search through video DB

Return the retrieved results

Overview

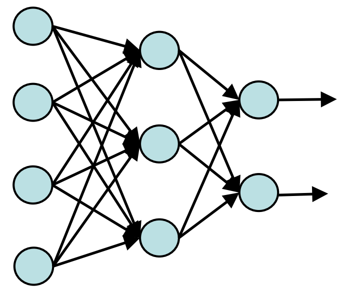
Deep Poselets

Poselet Discovery



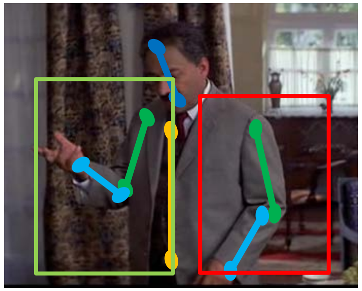
- Cluster pose space

Training



- Train poselets using convolutional neural networks

Detection

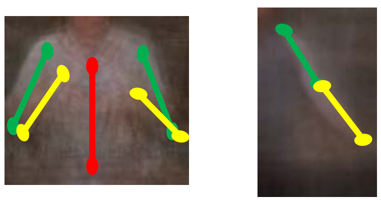


- Detect poselets

Overview

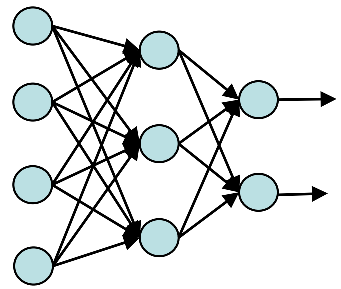
Deep Poselets

Poselet Discovery



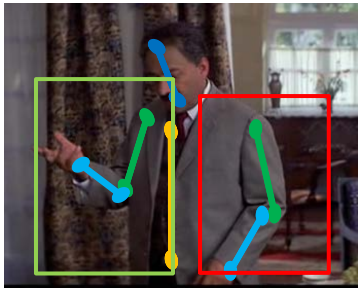
- Cluster pose space

Training



- Train poselets using convolutional neural networks

Detection



- Detect poselets

Pose retrieval

- Given a query image

- Build Bag of Deep poselets

- Return the retrieved results

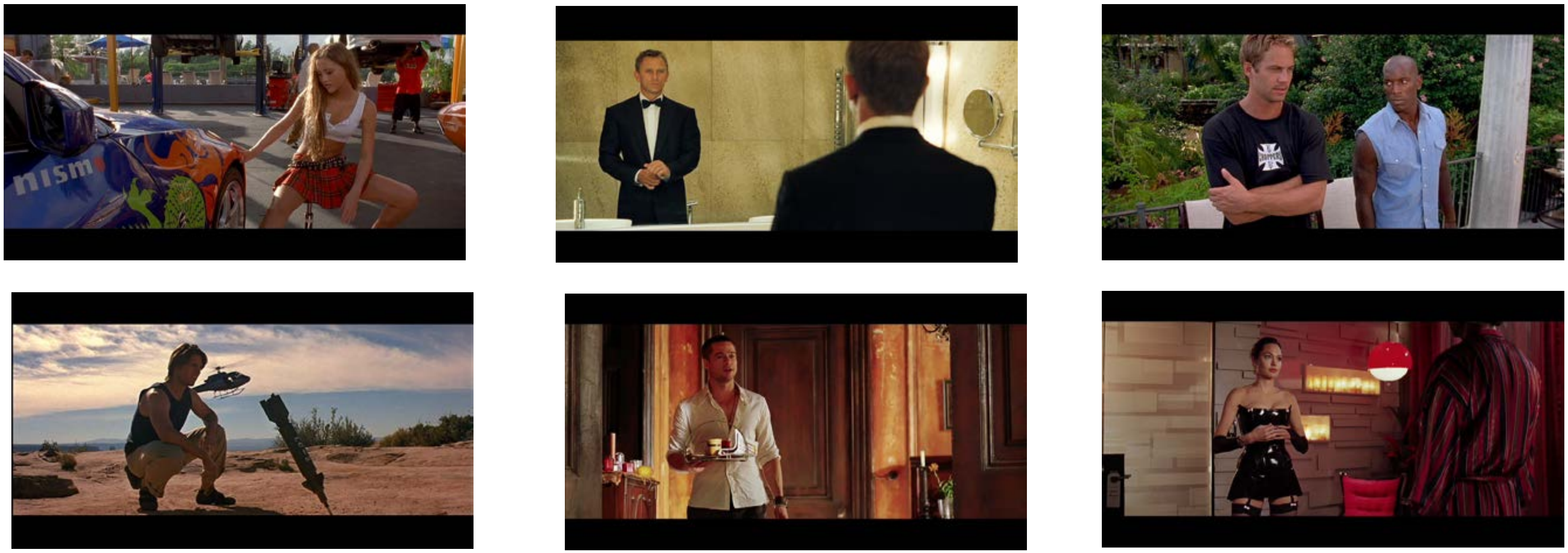
Datasets

Buffy Stickmen (Season 1, 5 episodes)

ETH Pascal dataset (Flickr Images)

H3D
(Flickr Images)

Datasets



FLIC dataset (30 Hollywood movies)

Movie dataset (Ours) (22 Hollywood movies)
No overlap with FLIC

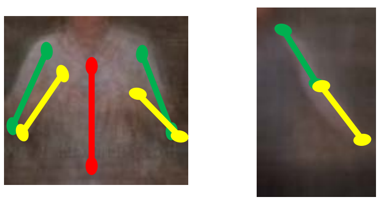
Datasets

Dataset	Train	Validation	Test	Total
H3D	238	0	0	238
ETHZ Pascal	0	0	548	548
Buffy	747	0	0	747
Buffy-2	396	0	0	396
Movie	1098	491	2172	3756
Flic	2724	2279	0	5003
Total stickmen annotations	5198	2764	2720	10682
+ Flipped version	10396	5528	5440	21364

Overview

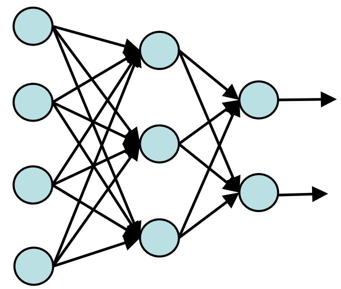
Deep Poselets

Poselet Discovery



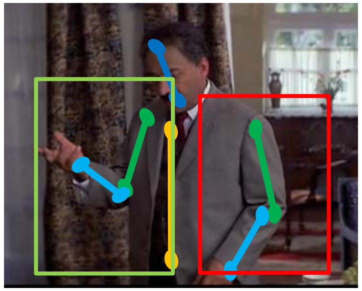
- Cluster pose space

Training



- Train poselets using convolutional neural networks

Detection



- Detect poselets

Pose retrieval

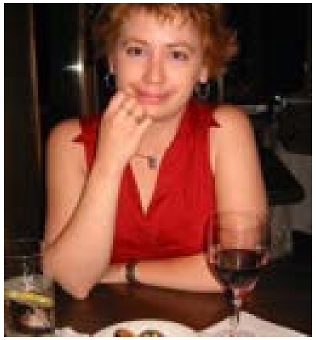
- Given a query image

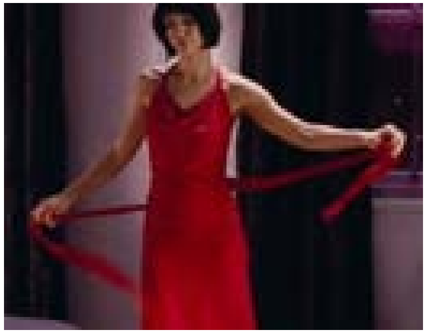
- Build Bag of Deep poselets

- Return the retrieved results

Poselets

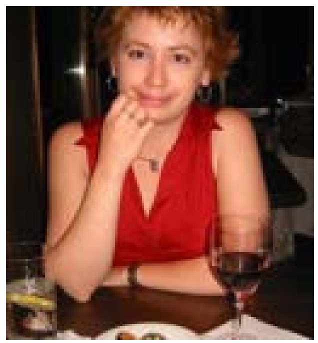
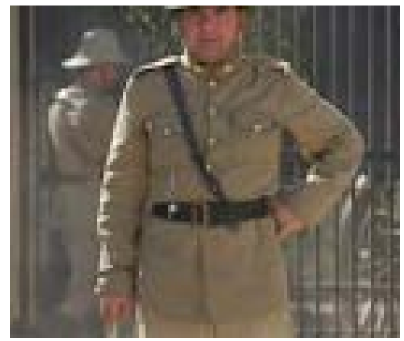
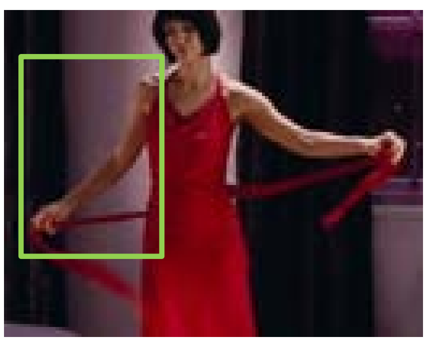
Poselets model body parts in a particular spatial configuration.

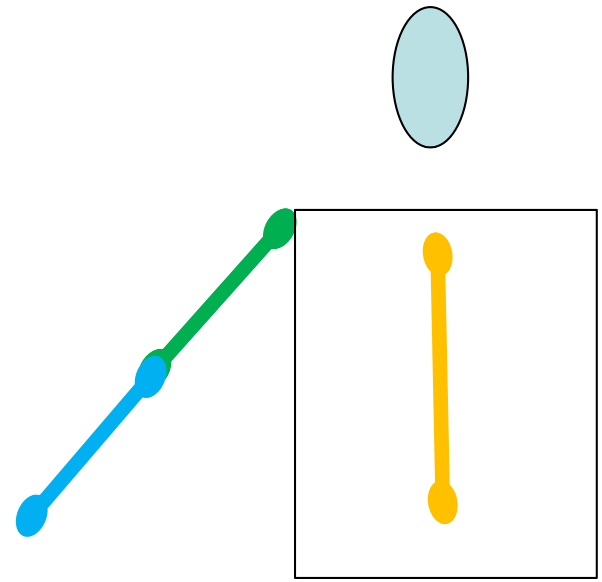





Poselets

Poselets model body parts in a particular spatial configuration.

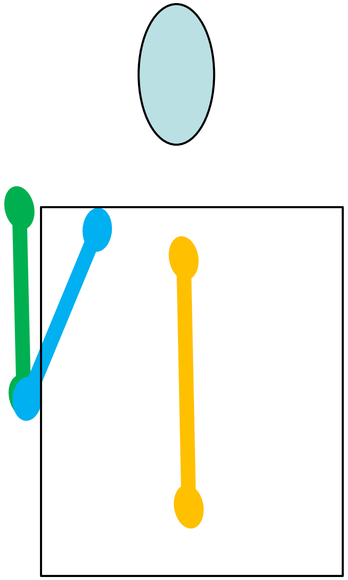





Poselet 1

Poselets

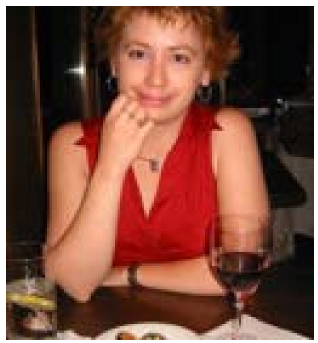
Poselets model body parts in a particular spatial configuration.

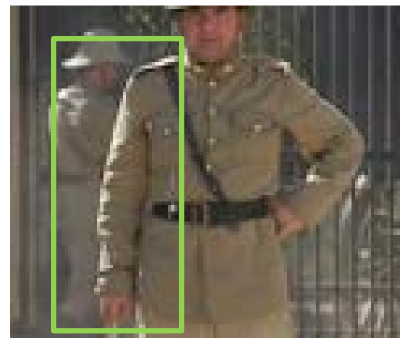


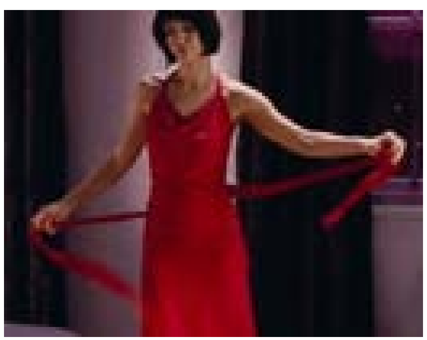
Poselet 2

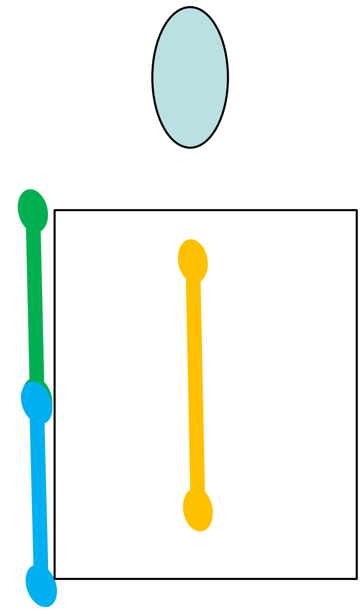
Poselets

Poselets model body parts in a particular spatial configuration.



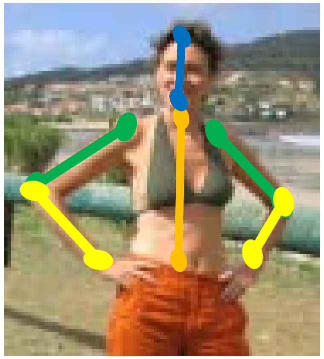






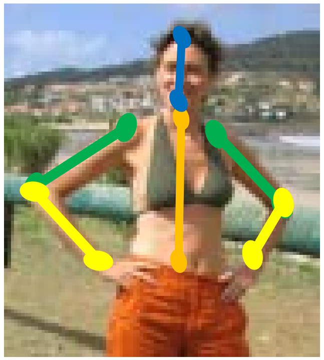
Poselet 3

Poselets: Discovery

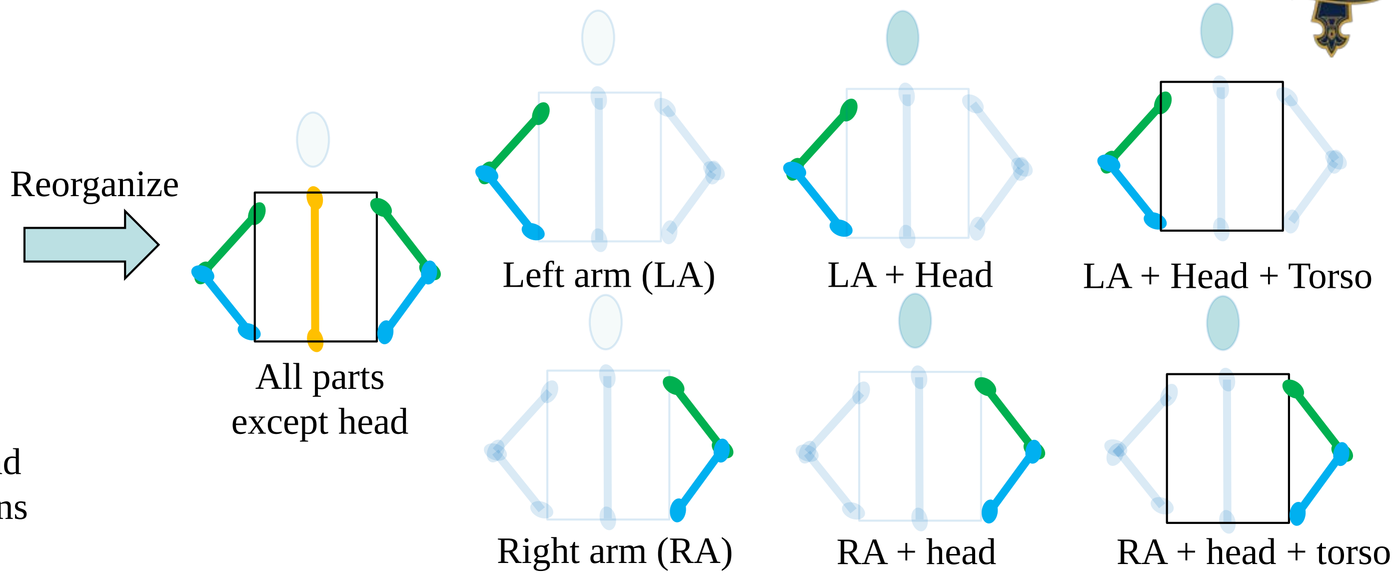


Training data with ground truth stickmen annotations

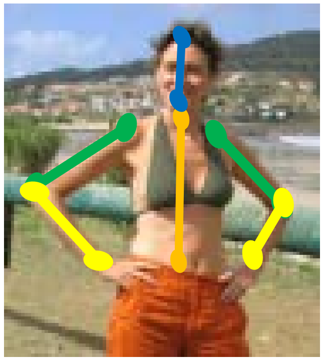
Poselets: Discovery



Training data with ground truth stickmen annotations

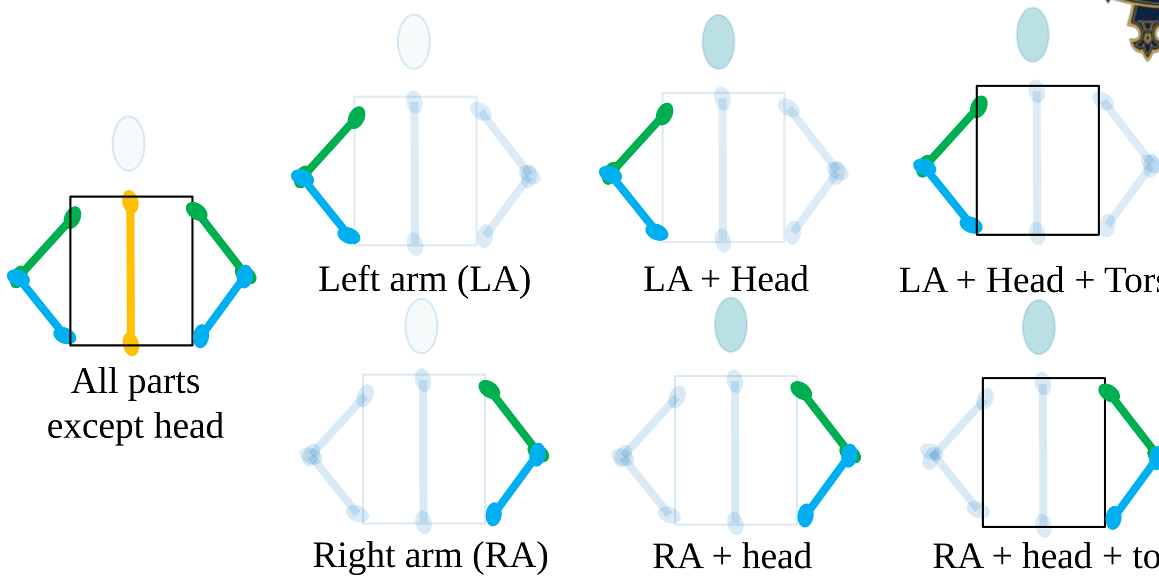


Poselets: Discovery



Training data with ground truth stickmen annotations

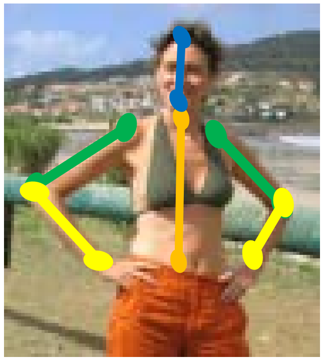
Reorganize



For each set, get pose descriptors

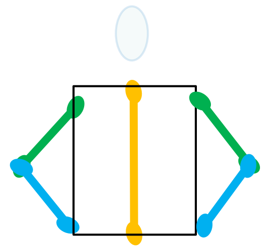
- For each body part, note the angle
- Cluster on the angles

Poselets: Discovery

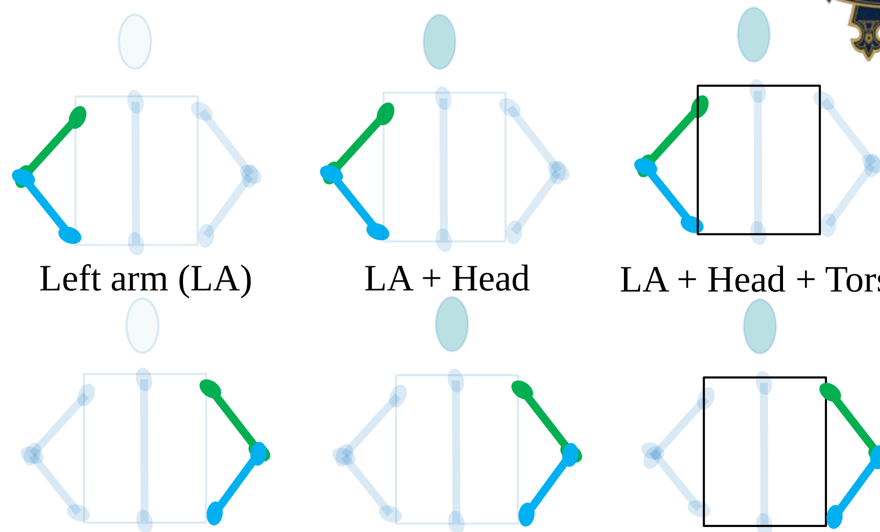


Training data with ground truth stickmen annotations

Reorganize



All parts except head

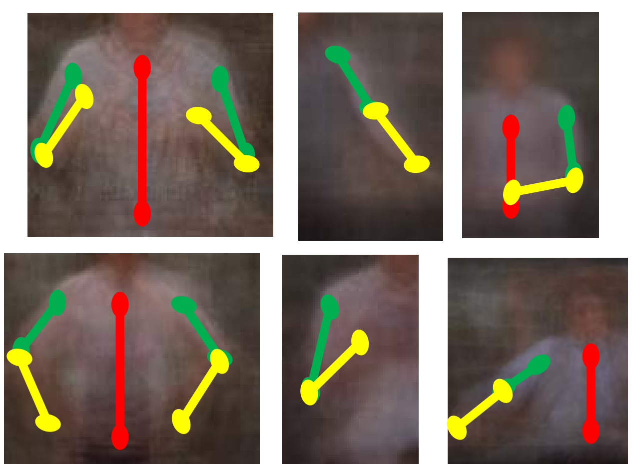


For each set, get pose descriptors

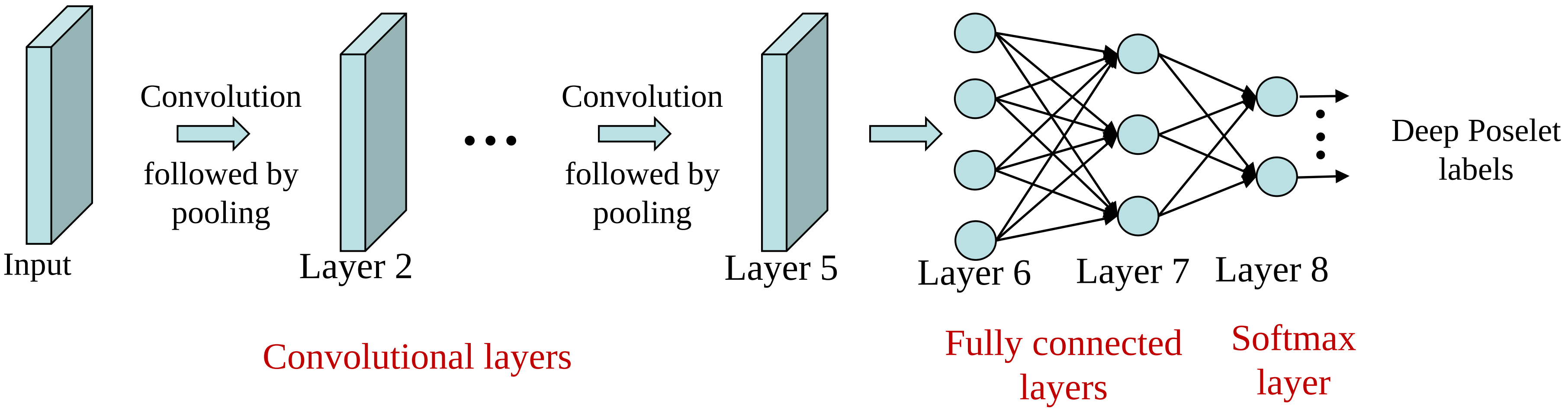
- For each body part, note the angle
- Cluster on the angles

K-Means Clustering

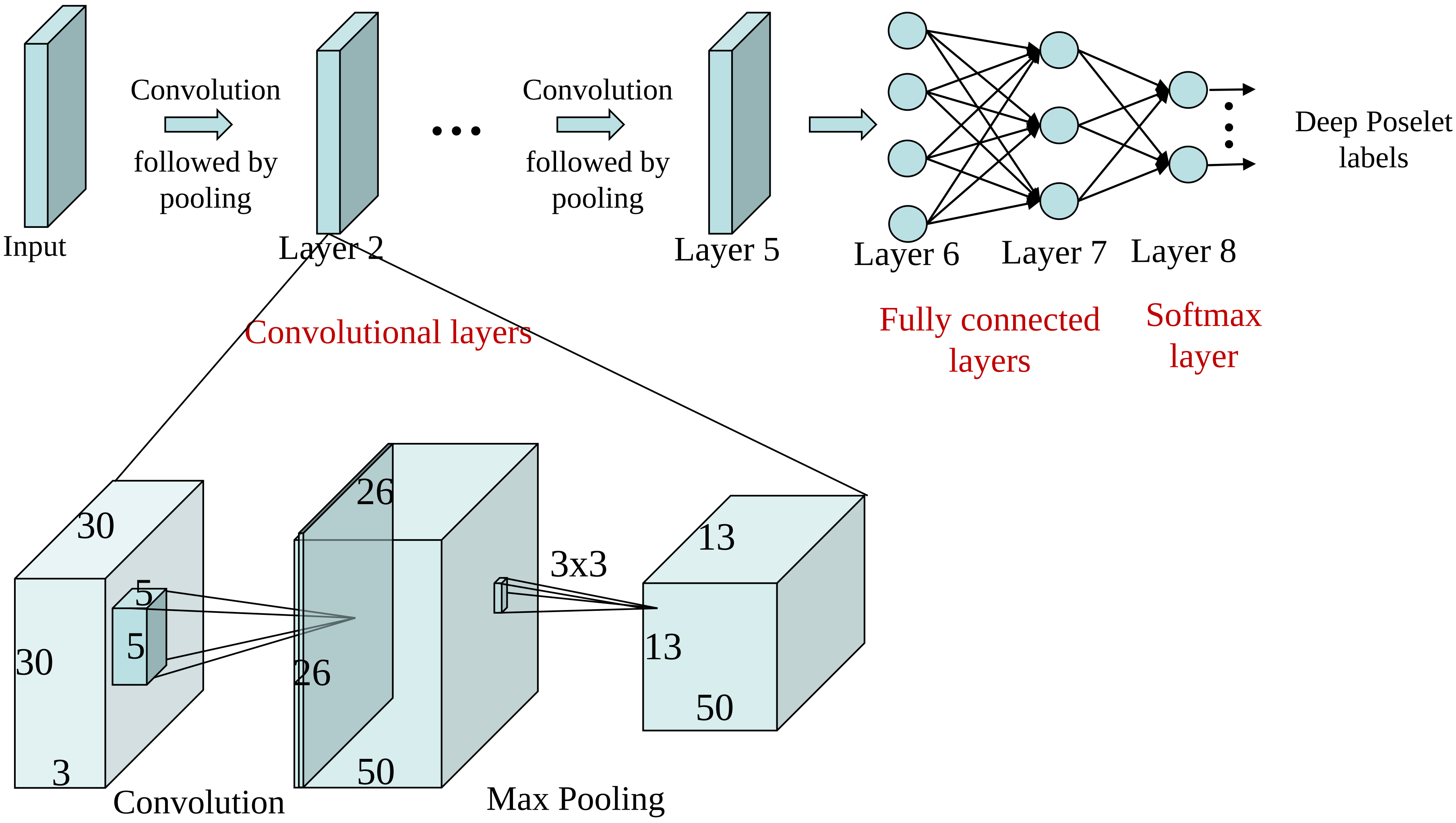
Poselet Average Images



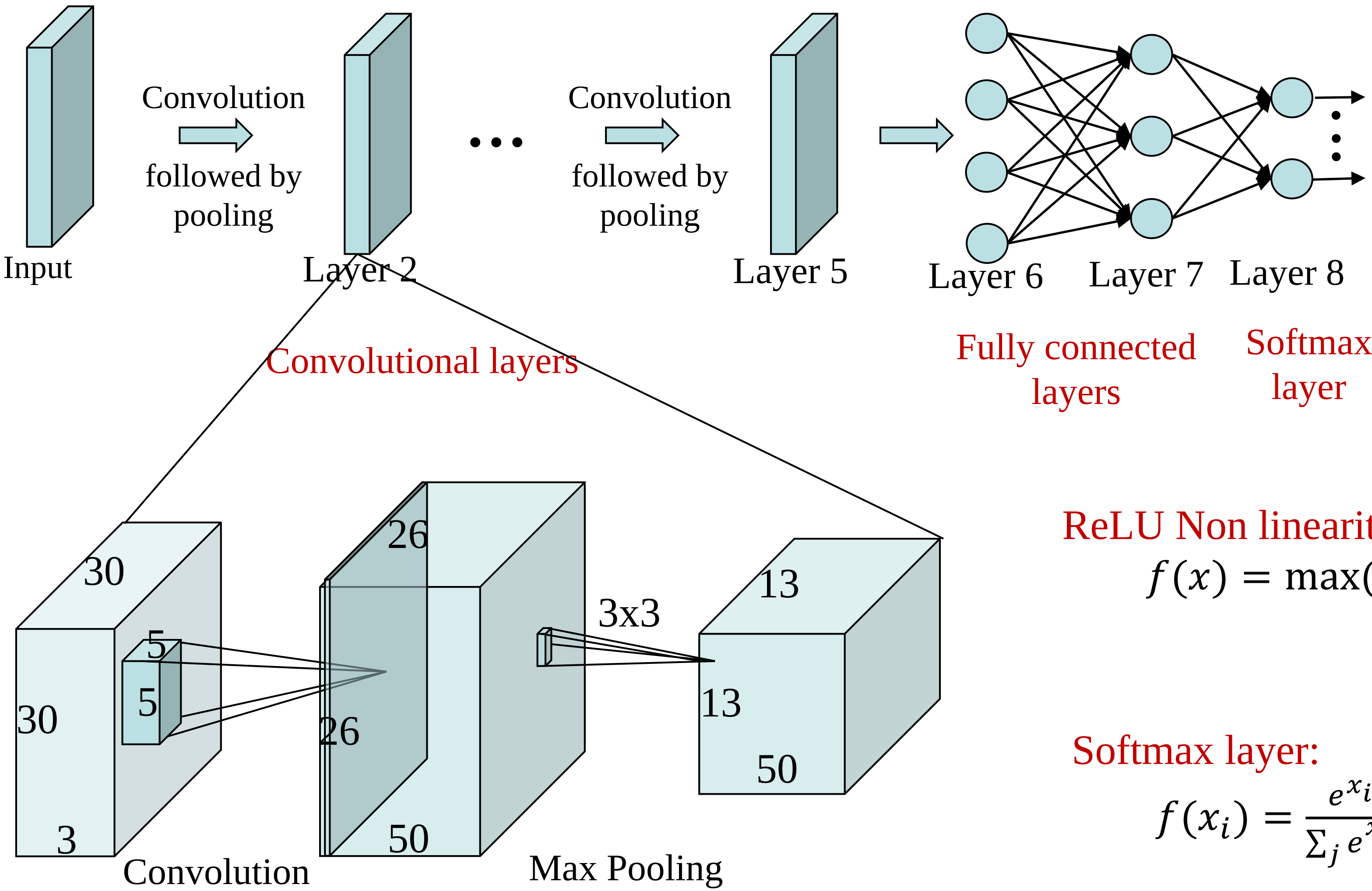
Deep Poselets: CNNs



Deep Poselets: CNNs



Deep Poselets: CNNs



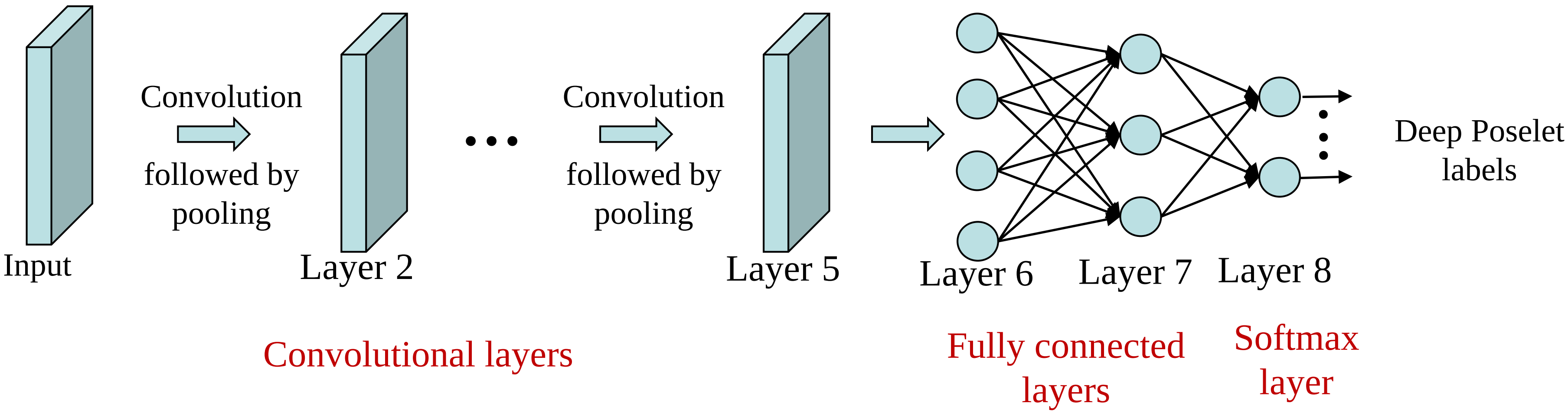
ReLU Non linearity:

$$f(x) = \max(0, x)$$

Softmax layer:

$$f(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

Deep Poselets: Training



Input image: x Model parameters: w Ground truth: g Output: $y = f(x, w)$

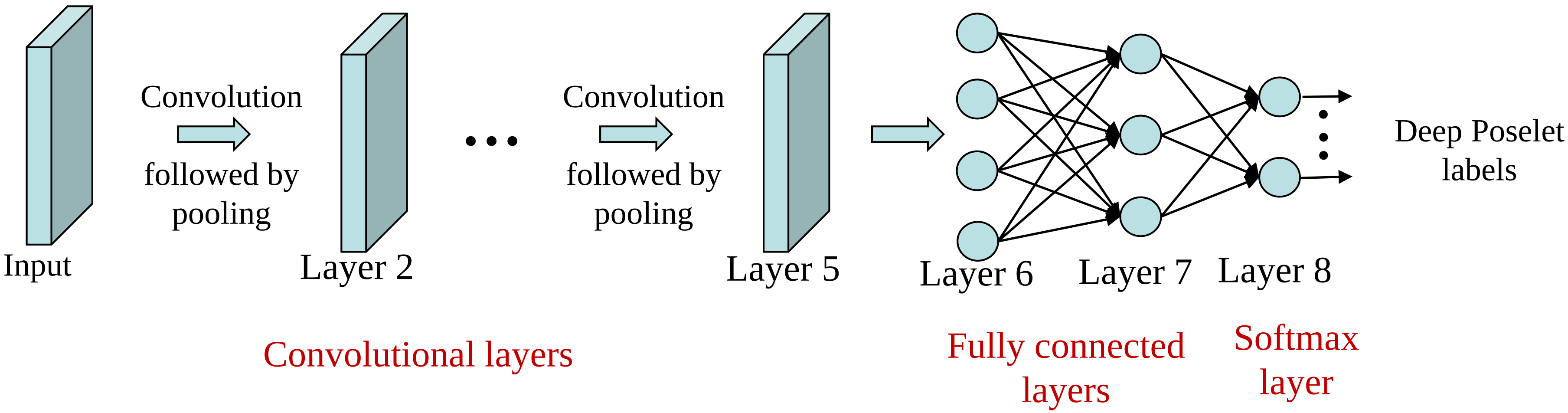
Loss function: $L = \sum_j g_j \log(y_j)$

Training: Stochastic Gradient Descent

$$w = w - \frac{\eta \partial L}{\partial w}$$

Architecture from Krizhevsky et al., NIPS 2012

Deep Poselets: Fine tuning



Challenge:

-- Network has 40 million parameters.
-- Required training data ~1-2 million.
-- Available training data ~50K.

Solution:

-- Train the network on a task with enough data present.
-- Fine-tune the network to the current task.

Fine tuning procedure:

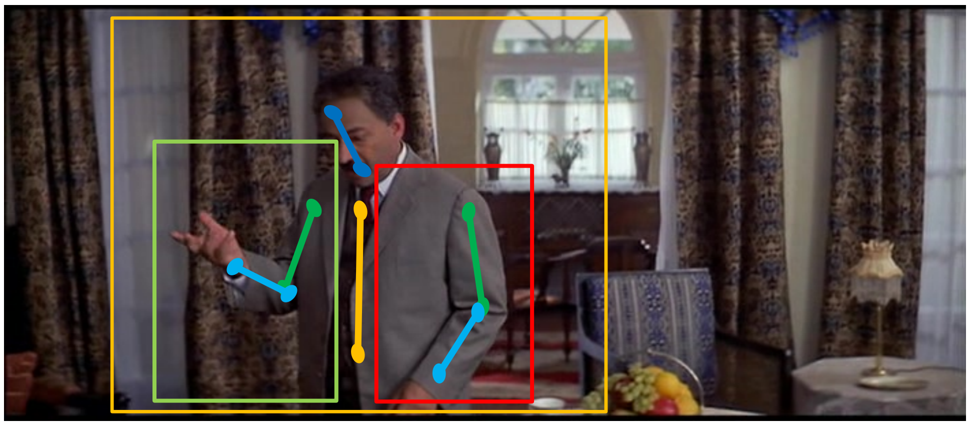
-- Train image classification task using imagenet data of size 1.2 million.

-- Replace the softmax layer with random initialization.

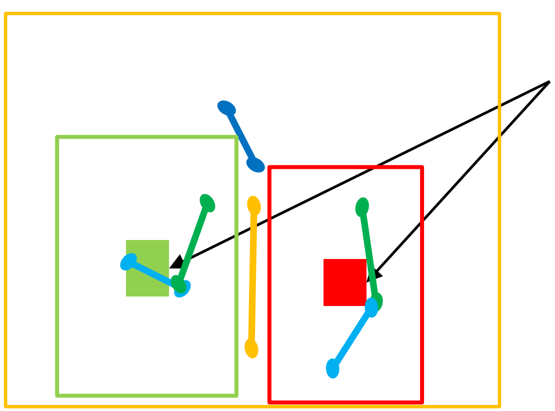
-- Run the gradient descent.

Deep Poselets: Detection

Given a test image, run all the deep poselets.

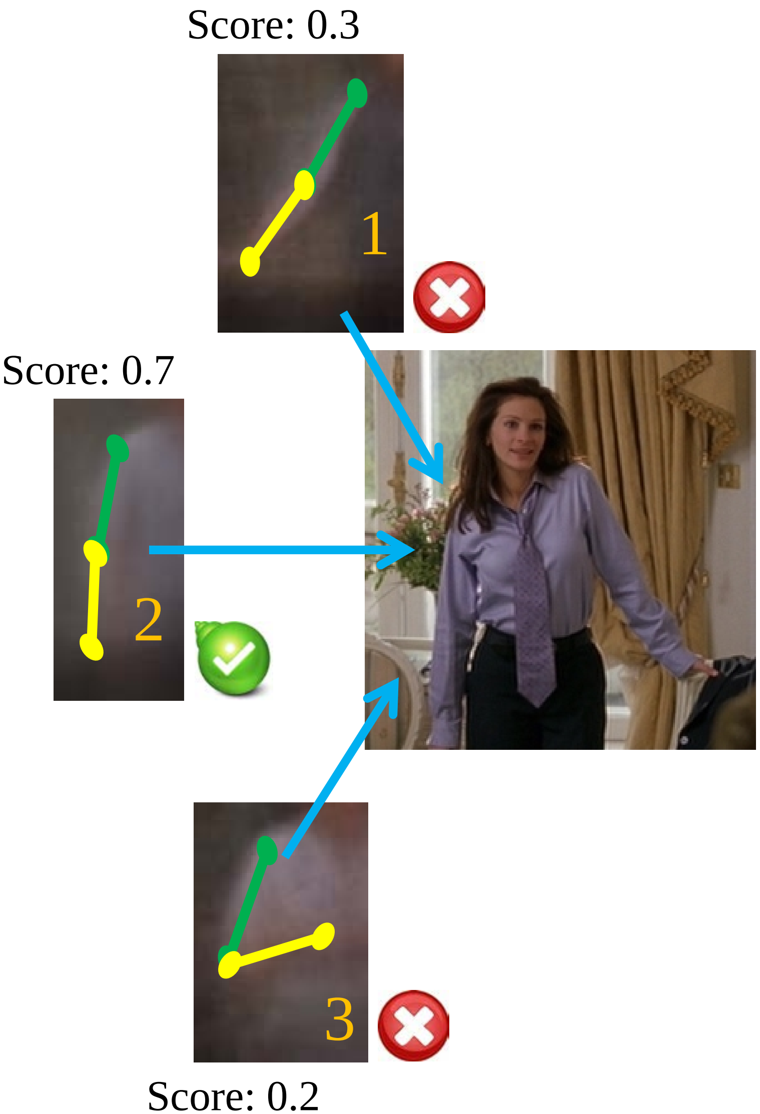


- Each poselet occurs in a localized regions within a upper body detection.
- Run the classifiers on the “Expected center points of poselets”.
- This improves both the speed and accuracy.



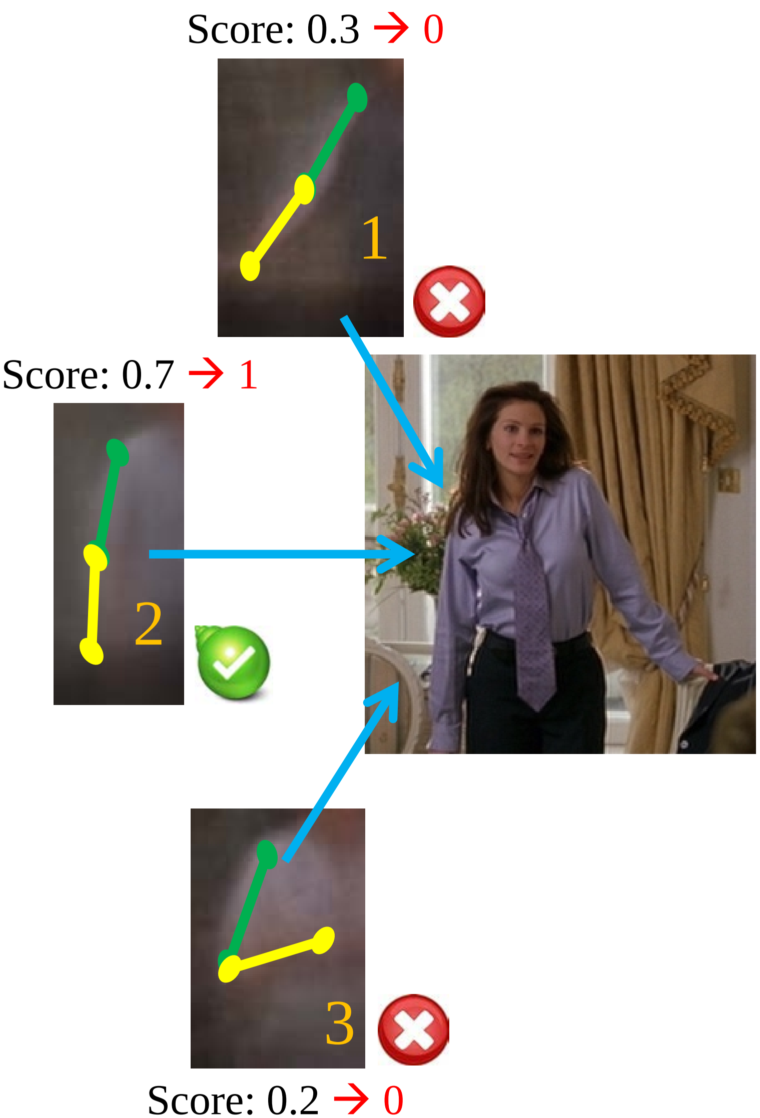
Expected center points of poselets.

Deep Poselets: Spatial reasoning



Problem: The three detections fired in the same area.

Deep Poselets: Spatial reasoning



Problem: The three detections fired in the same area.

Objective: Rescore detection 2 to 1 and the detections 1,3 to 0.

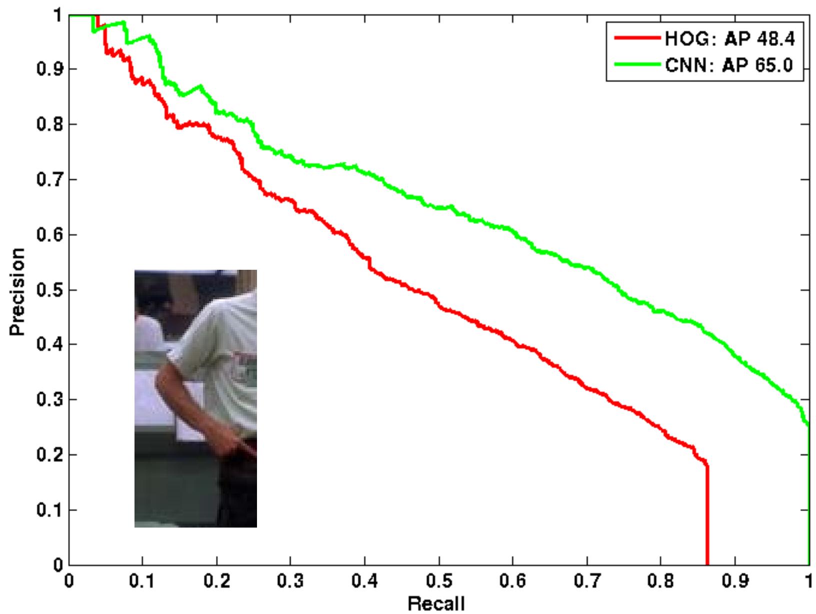
Solution:

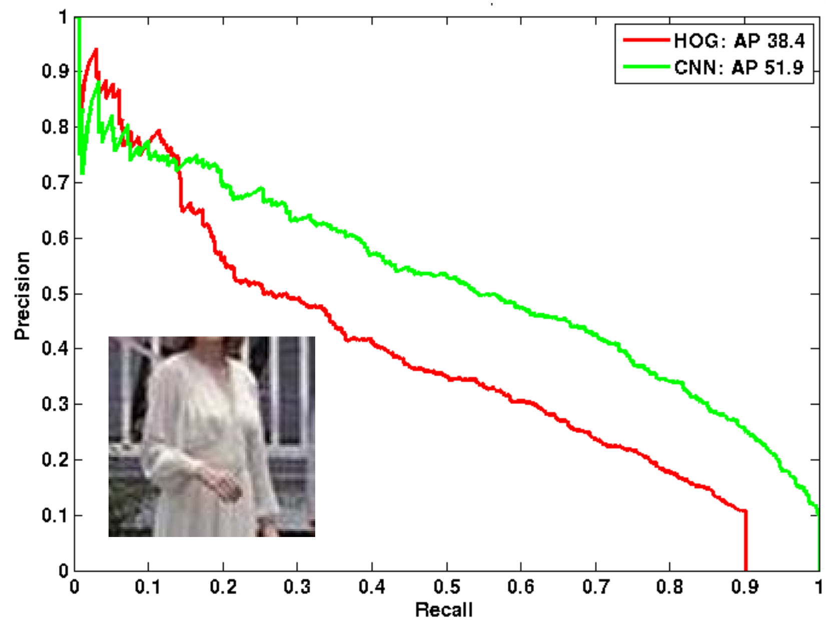
For each poselet, learn regression function whose

-- Input: Scores of other poselet detections

-- Output: New score

Deep Poselets: Results

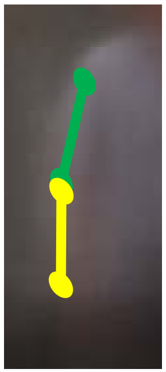




Method	MAP-test
HOG	32.6
CNN before fine-tuning	48.6
CNN after fine-tuning	56.0

- *Evaluation measure:* Mean average precision.
- *Comparison:* Poselets are trained using HOG feature.

Deep Poselets: Results



AP 78.1

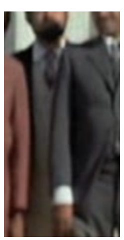
#positives in train set 1863

Rank 1

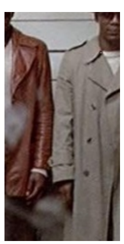
Rank 6

Rank 11

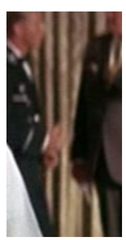
Rank 16



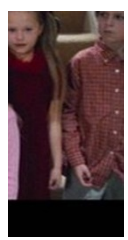
Rank 21



Rank 26



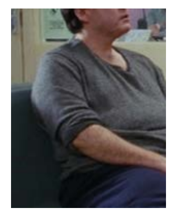
Rank 31



Rank 36

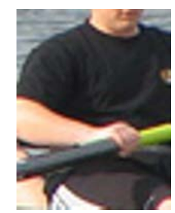
AP 40.4

#positives in train set 698



Rank 1

Rank 6



Rank 11

Rank 16

Rank 21

Rank 26

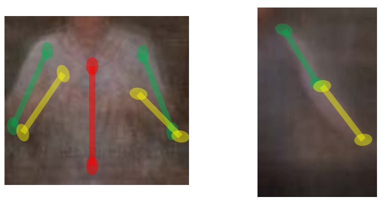
Rank 31

Rank 36

Overview

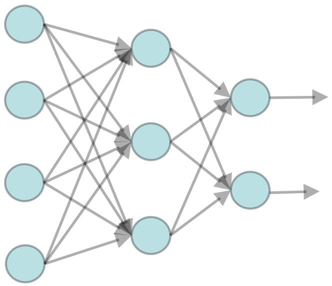
Deep Poselets

Poselet Discovery



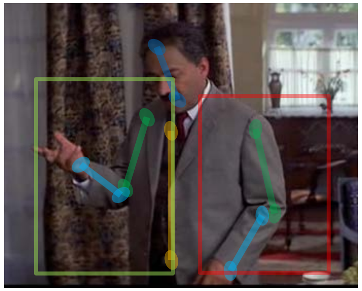
- Cluster pose space

Training



- Train poselets using convolutional neural networks

Detection



- Detect poselets

Pose retrieval

- Given a query image

- Build Bag of Deep poselets

- Return the retrieved results

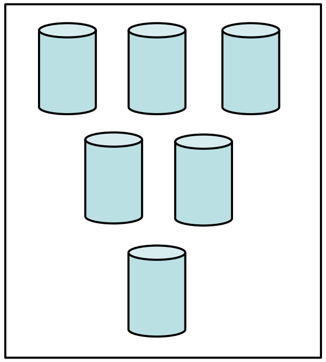
Pose Search: Indexing

For each frame in the video DB collection

- Detect the upper body.
- Run all the poselets.
- Perform spatial reasoning.

Descriptor: Max pool the Deep Poselet detections

122D vector

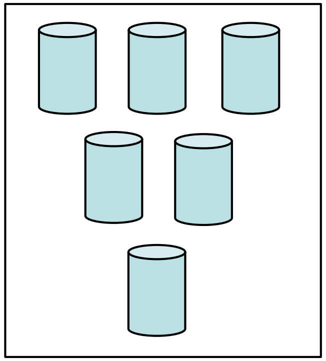


Index in a database

Pose Search: Retrieval

Given a query image

Build Bag of
Deep poselets



Using *cosine distance*, search through the database

Return the retrieved results

Pose Search: Results

Experimental setup

- Database: Test data of size 5440 is used as the database.
- Queries: All the samples in the test data are used as query.
- Evaluation metric: Mean average precision (MAP).

Methods compared against

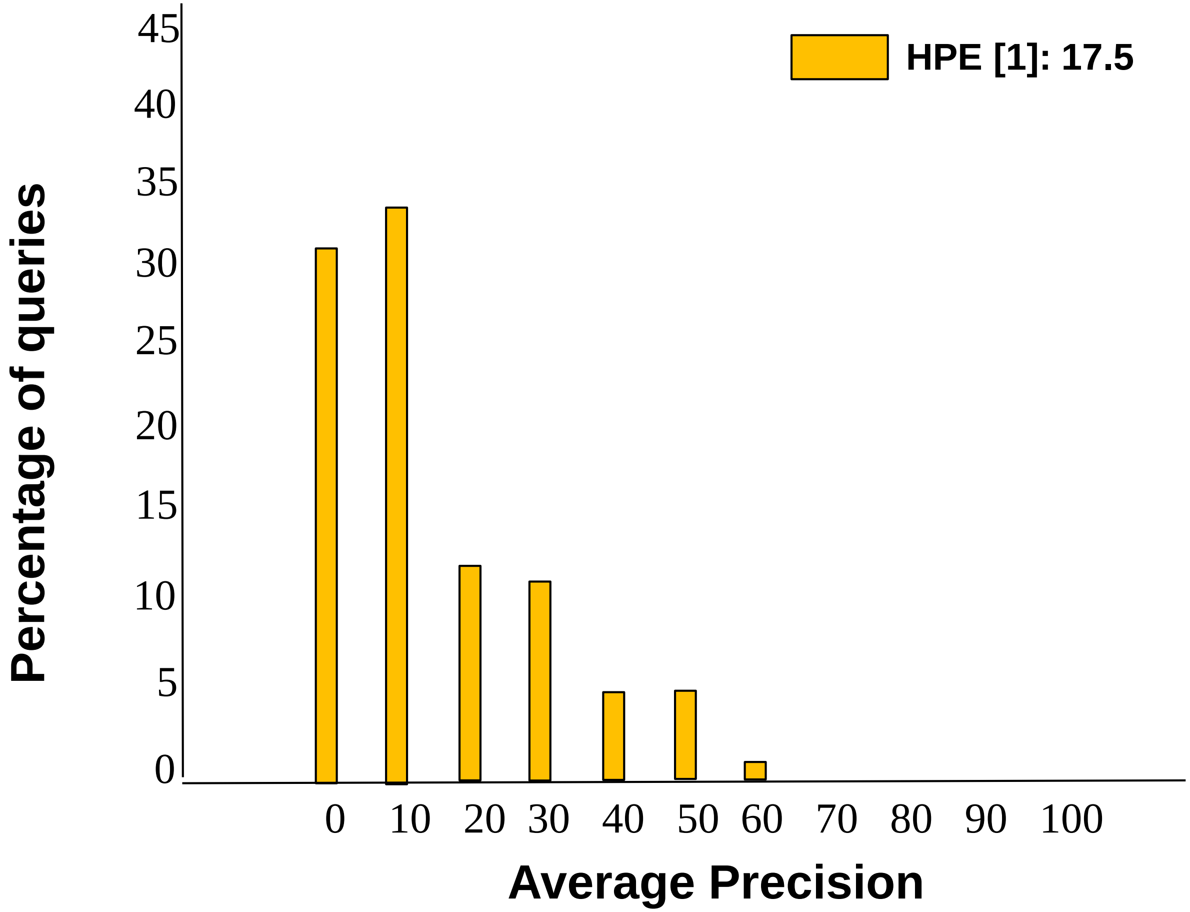
- *Bag of visual words (BOVW)*
  - *Detect sift → K means (K = 1000) → VQ.*
- *Berkeley Poselets (BPL)*
  - *Run poselets → Bag of parts.*
- *Human pose estimation* [1] *(HPE)*
  - *Run human pose estimation algorithms*
  - *Concatenate (sin(x),cos(x)) of all the body part angles.*

Results

Method	MAP
BOVW	14.2
BPL	15.3
HPE [1]	17.5
Ours	34.6

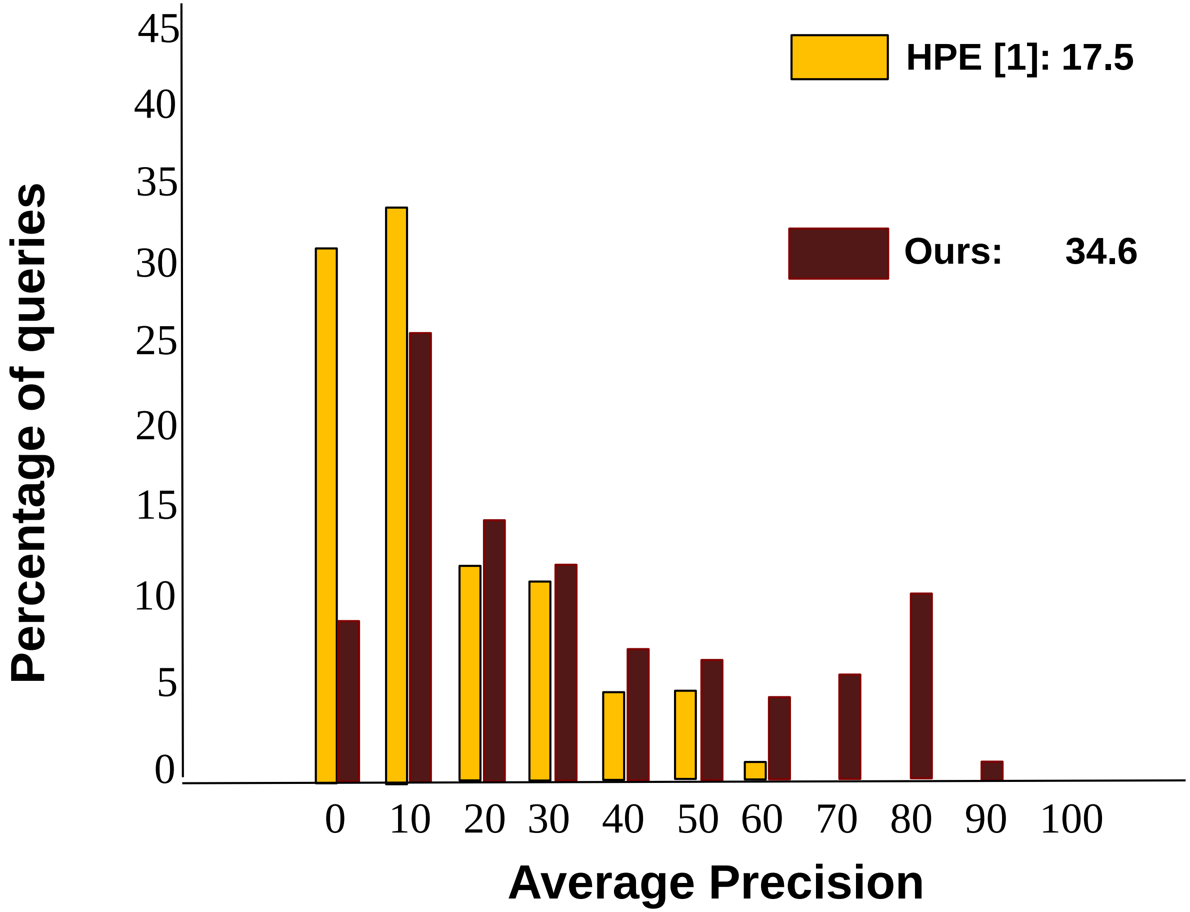
[1] Y. Yang and D. Ramanan. "Articulated pose estimation with flexible mixtures-of-parts." In CVPR, 2011.

Pose Search: Results



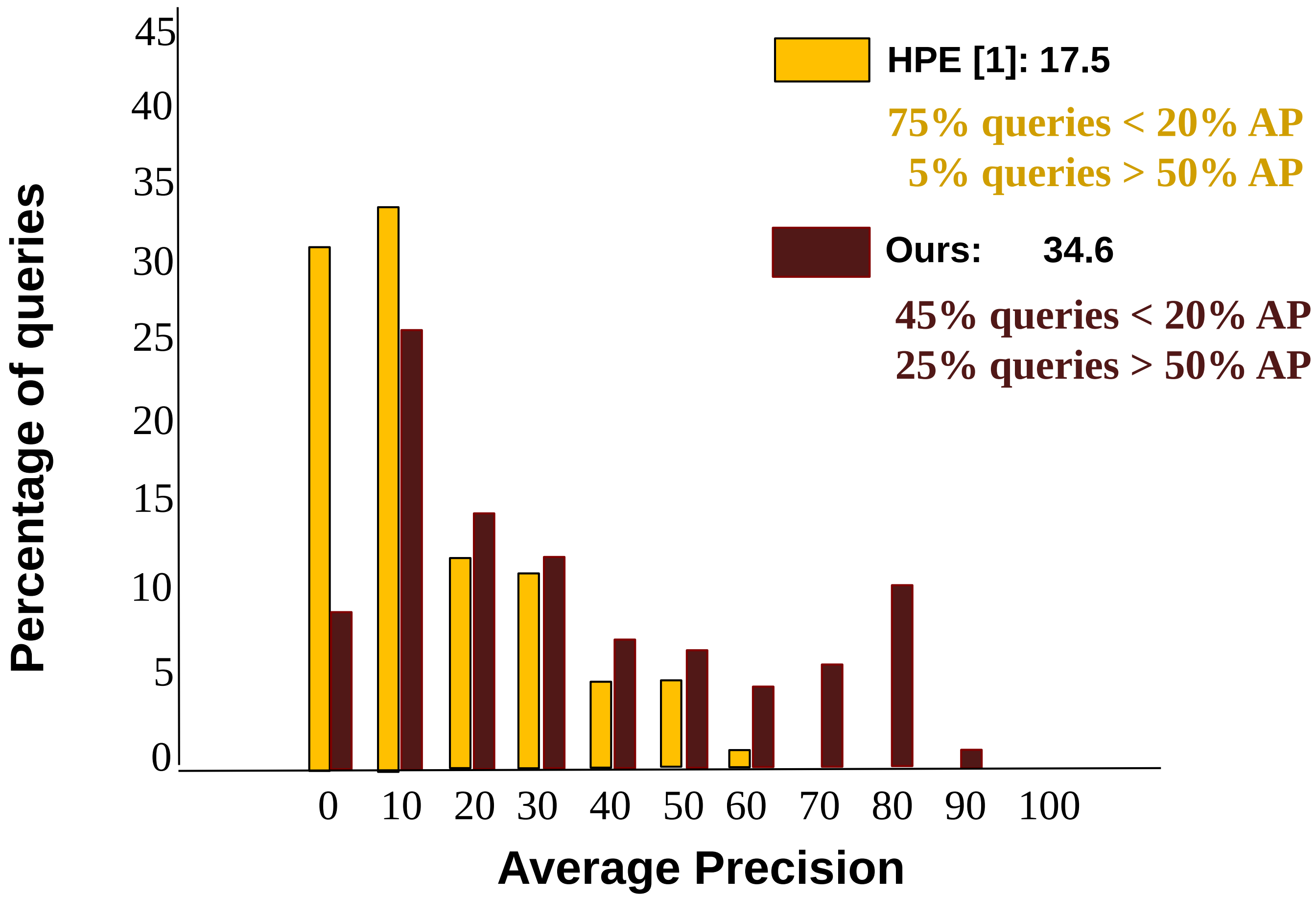
Comparison with the state-of-the-art

Pose Search: Results



Comparison with the state-of-the-art

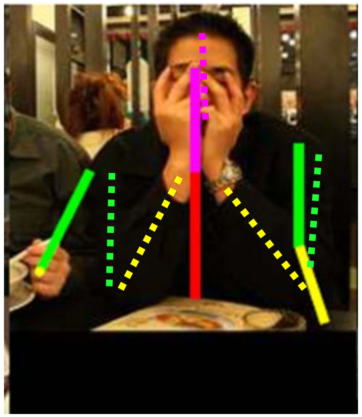
Pose Search: Results

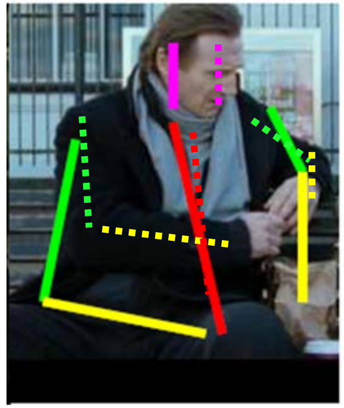


Comparison with the state-of-the-art

Pose Search: Analysis

HPE





Ground truth   Detection

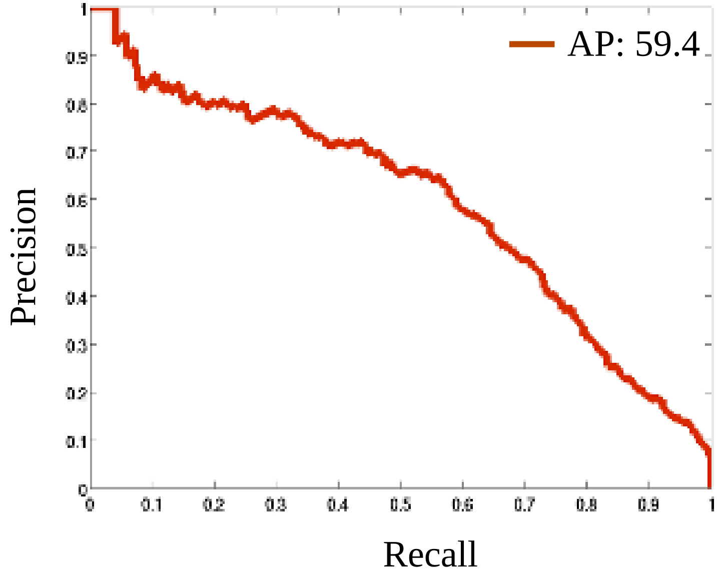
- Pose detection algorithms often commit to wrong pose.
- Pose search systems based on them perform poorly.

OURS

- Bag of poselets descriptor encodes multiple proposals weighted by their likelihood
- Hence it can recover when some of the detections are wrong.

Pose Search: Results

Query

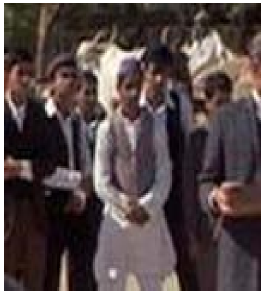


Rank 1

Rank 5

Rank 10

Rank 15

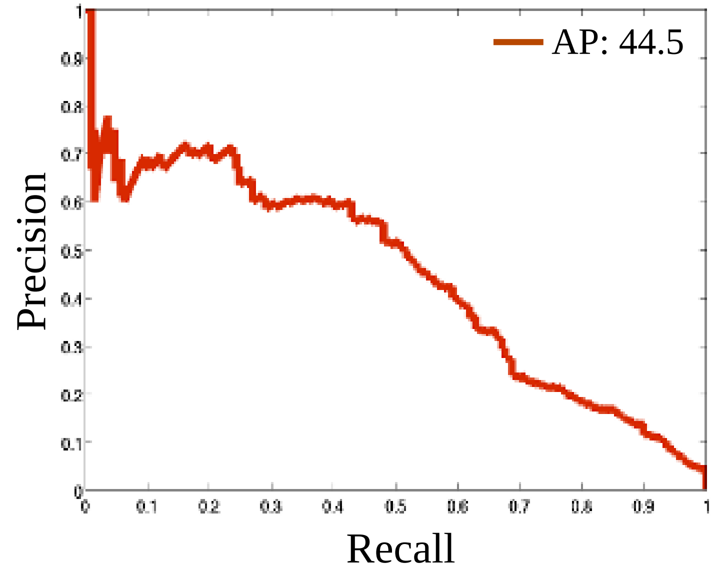


Rank 20

Rank 25

Pose Search: Results

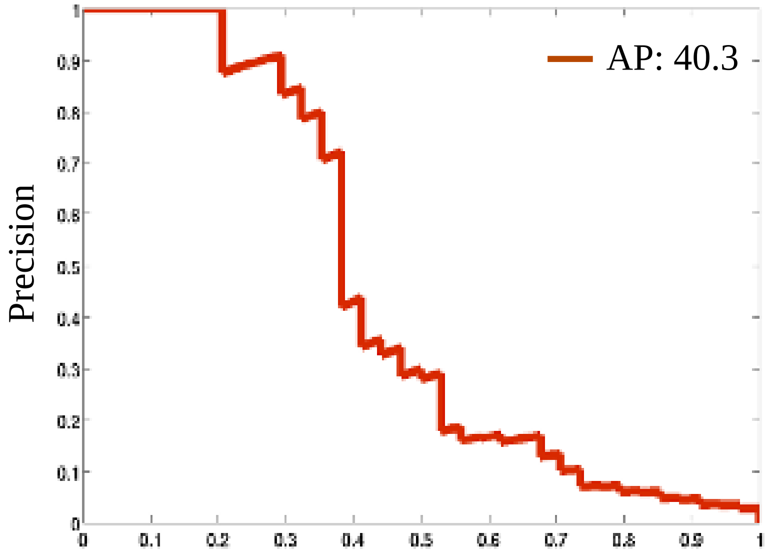
Query

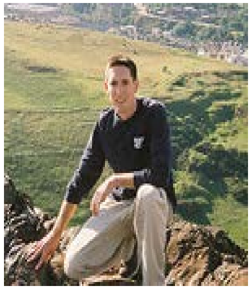


					
Rank 1	Rank 5	Rank 10	Rank 15	Rank 20	Rank 25

Pose Search: Results

Query





Rank 1

Rank 5

Rank 10

Rank 15

Rank 20

Rank 25

Summary

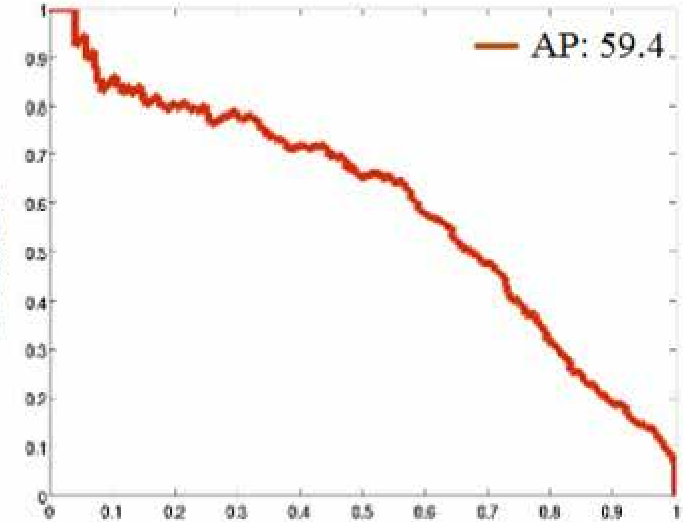
- We propose a novel *Deep Poselets* based method for human pose search system.
- Our *Deep Poselet* method outperforms HOG based poselets by 25% MAP.
- *Our pose retrieval method* improves the performance of the current state-of-art system by 17% MAP.

Thank you.

Questions?

Pose Search: Results

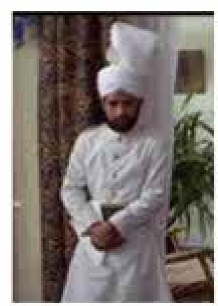
Query



Rank 1

Rank 5

Rank 10



Rank 15

Rank 20

Rank 25