Poselet Conditioned Pictorial Structures

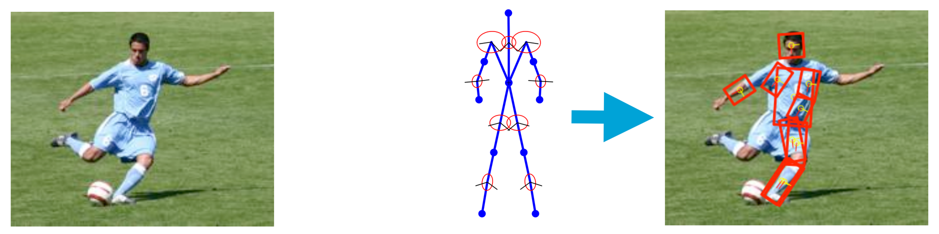
Leonid Pishchulin[1], Mykhaylo Andriluka[1], Peter Gehler[2] and Bernt Schiele[1]

[1]Max Planck Institute for Informatics, Saarbrücken, Germany

[2]Max Planck Institute for Intelligent Systems, Tübingen, Germany

State of the Art

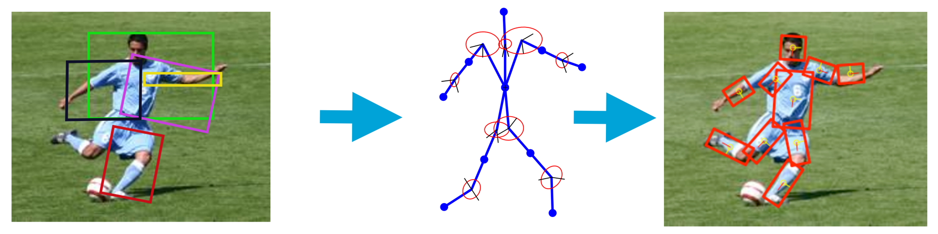
- Tree-structured pictorial structures models



– **generic** kinematic tree
– capture **adjacent** part dependencies **only**
+ exact and efficient

Contributions

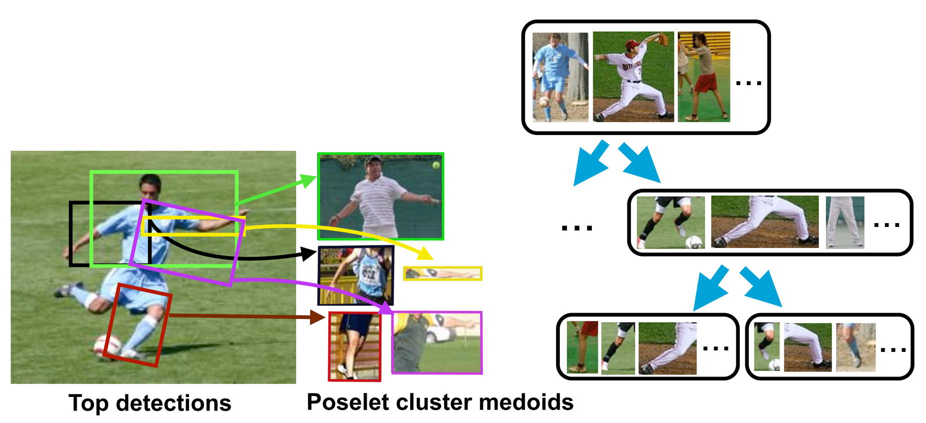
- Novel image conditioned pictorial structures model



+ **poselet conditioned** kinematic tree
+ **poselets** capture **non-adjacent** part dependencies
+ exact and efficient

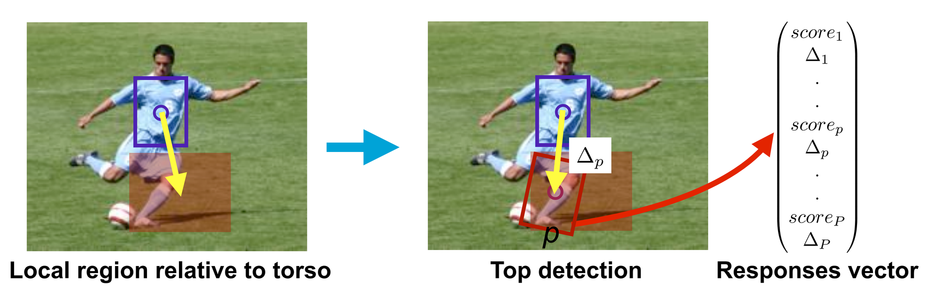
Poselets

- Detect joint part configurations [2]
- ⇒ **capture non-adjacent part dependencies**
- Trained for different levels of abstraction



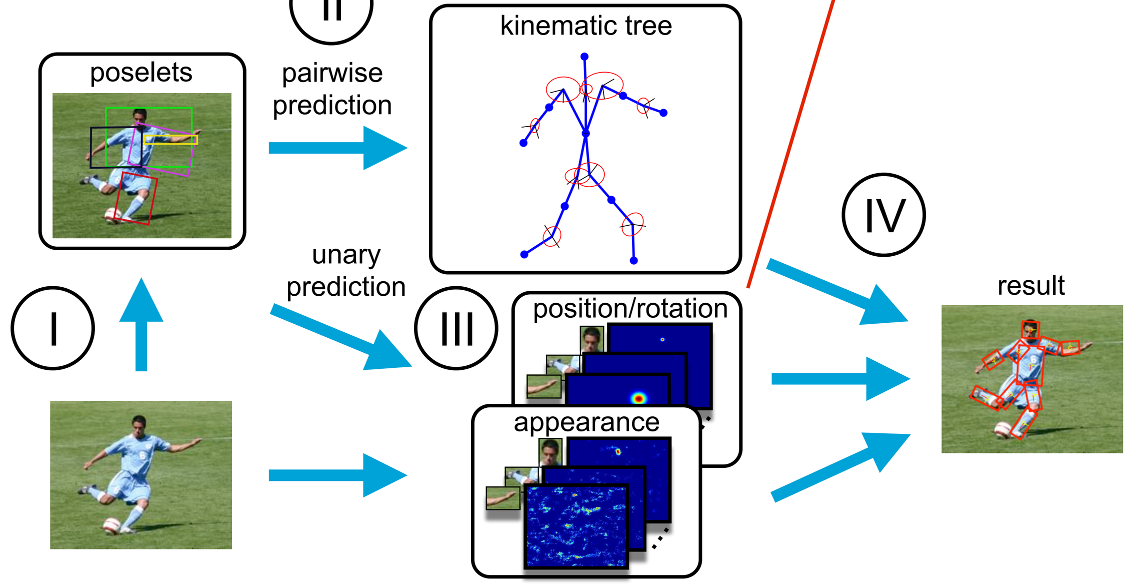
Poselets responses vector as **mid-level representation**:

- detect torso using strong detector [7]
- poselet offset w.r.t torso defines center of pooling region
- top response and offset contribute to vector



Poselet Conditioned Pictorial Structures

$$E(L;D) = \sum_{m=1}^{M} E^{u}(l_m;D) + \sum_{n\sim m} E^{p}(l_n,l_m;D) + w_p \sum_{m=1}^{M} E^{u,poselet}(l_m;D)$$

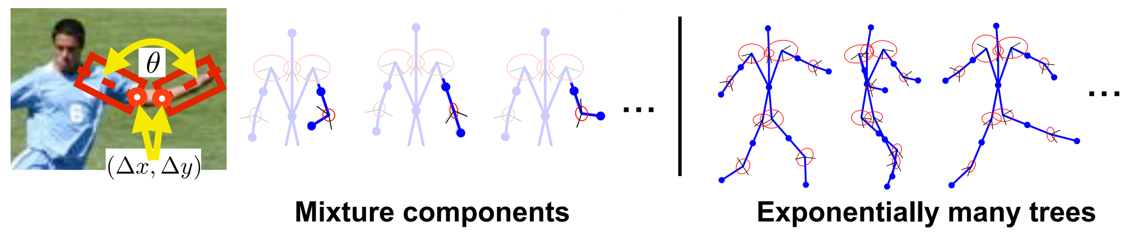


I. Mid-level representation based on poselets

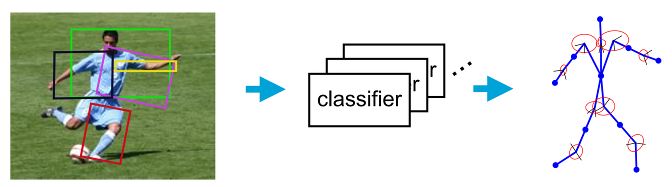
- compute poselets responses vector

II. Predicting pairwise parameters

- pairwise: relative offset $(\Delta x, \Delta y)$ and rotation θ
- learn mixtures *per pairwise* from clustering θ
- ⇒ allows to model exponentially many trees



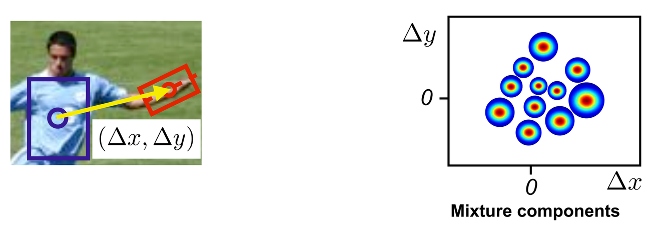
- **Prediction:** multi-class classifier on poselets responses



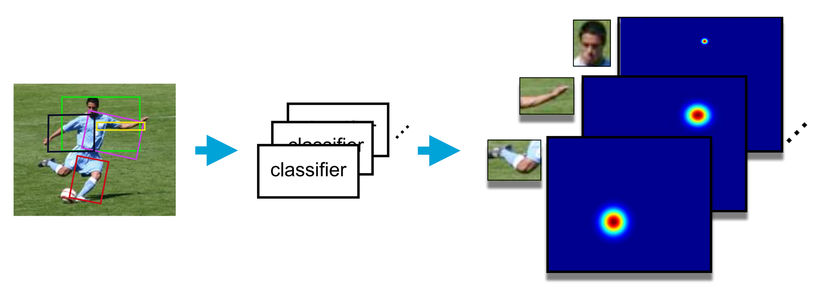
⇒ prediction **before** pose inference: exact and efficient inference

III. Predicting part position and rotation

- Part position relative to torso
  - learning: cluster offsets into mixture components

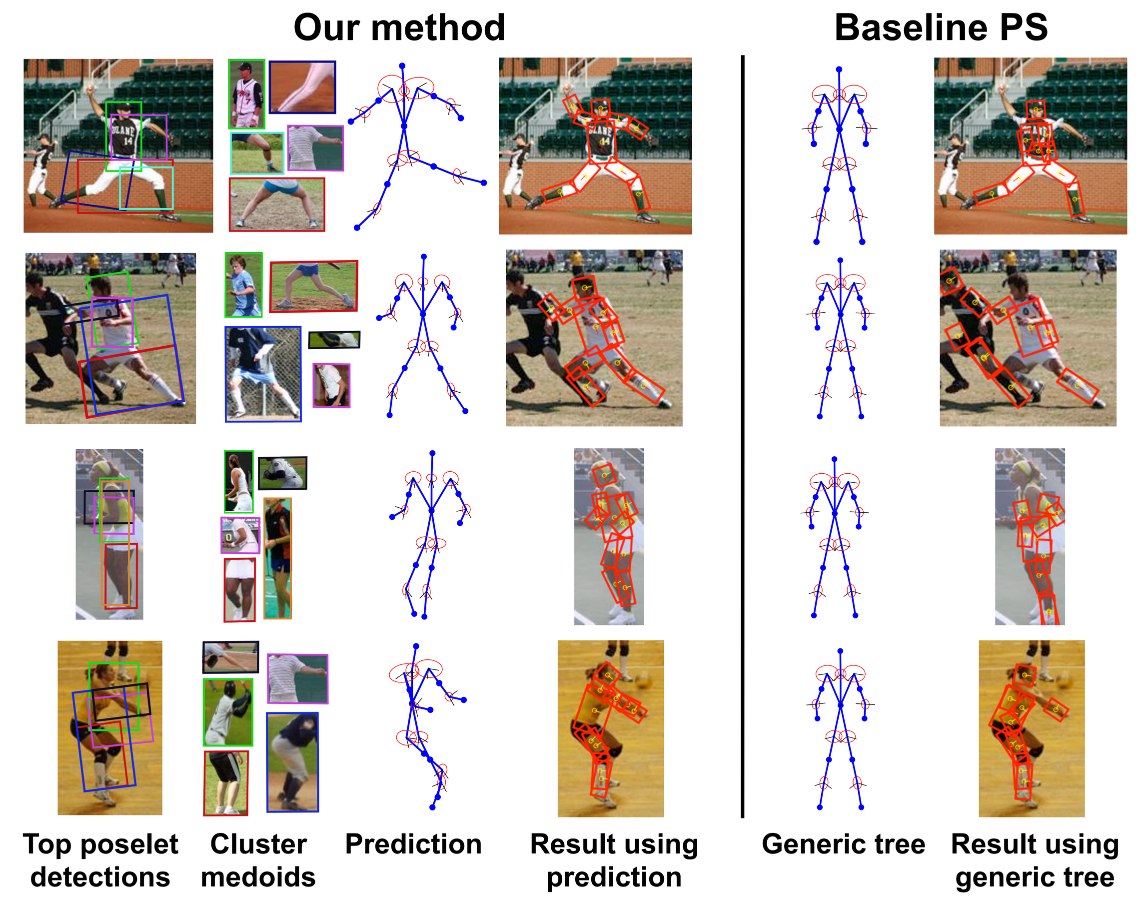


  - prediction: multi-class classifier



- Absolute part rotation:
  - learning: bin rotation to get mixture components
  - prediction: similar to predicting part position

Qualitative Results



Quantitative Results

Leeds Sports Poses (LSP) [5]

- 1,000 train, 1,000 test images
- set parameters using validation set
- observer-centric annotations for testing [4]
- *Percentage Correct Parts (PCP)* criterion

Method	Torso	Upper leg	Lower leg	Upper arm	Fore arm	Head	Total
Andriluka et al., [1]	80.9	67.1	60.7	46.5	26.4	74.9	55.7
predict pairwise	85.8	74.0	66.1	51.7	30.9	78.0	60.9
predict unary	86.1	73.3	65.8	52.8	31.0	76.0	60.8
predict pairwise+unary	**87.5**	**75.7**	68.0	54.2	33.9	78.1	62.9
Yang&Ramanan [9]	84.1	69.5	65.6	52.5	35.9	77.1	60.8
Eichner&Ferrari [4]	84.9	73.1	**68.3**	**55.8**	**38.6**	**80.1**	**63.7**

Image Parse (IP) [8]

- 100 train, 205 test images

Method	Torso	Upper leg	Lower leg	Upper arm	Fore arm	Head	Total
ours	**92.2**	74.6	63.7	54.9	39.8	70.7	62.9
ours + [7]	90.7	**80.0**	**70.0**	59.3	37.1	77.6	66.1
Andriluka et al. [1]	86.3	66.3	60.0	54.6	35.6	72.7	59.2
Yang&Ramanan, [9]	82.9	69.0	63.9	55.1	35.4	77.6	60.7
Duan et al., [3]	85.6	71.7	65.6	57.1	36.6	**80.4**	62.8
Pishchulin et al., [7]	88.8	77.3	67.1	53.7	36.1	73.7	63.1
Johnson&Everingham, [6]	87.6	74.7	67.1	**67.3**	**45.8**	76.8	**67.4**

Limitations

- prediction
  - prediction: 62.9% PCP; oracle: 88.1% PCP (on LSP)
- typical failure cases

Self-occlusion   Foreshortening   Rare poses

References

[1] M. Andriluka, S. Roth, and B. Schiele. Pictorial structures revisited: People detection and articulated pose estimation. In *CVPR*, 2009.
[2] L. Bourdev and J. Malik. Poselets: Body part detectors trained using 3D human pose annotations. In *ICCV'09*.
[3] K. Duan, D. Batra, and D. Crandall. A multi-layer composite model for human pose estimation. In *In BMVC'12*.
[4] M. Eichner and V. Ferrari. Appearance sharing for collective human pose estimation. In *In ACCV'12*.
[5] S. Johnson and M. Everingham. Clustered pose and nonlinear appearance models for human pose estimation. In *BMVC'10*.
[6] S. Johnson and M. Everingham. Learning Effective Human Pose Estimation from Inaccurate Annotation. In *CVPR'11*.
[7] L. Pishchulin, A. Jain, M. Andriluka, T. Thormaehlen, and B. Schiele. Articulated people detection and pose estimation: Reshaping the future. In *CVPR*, 2012.
[8] D. Ramanan. Learning to parse images of articulated objects. In *NIPS'06*.
[9] Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In *CVPR'11*.