Bottom-up Recognition and Parsing of the Human Body

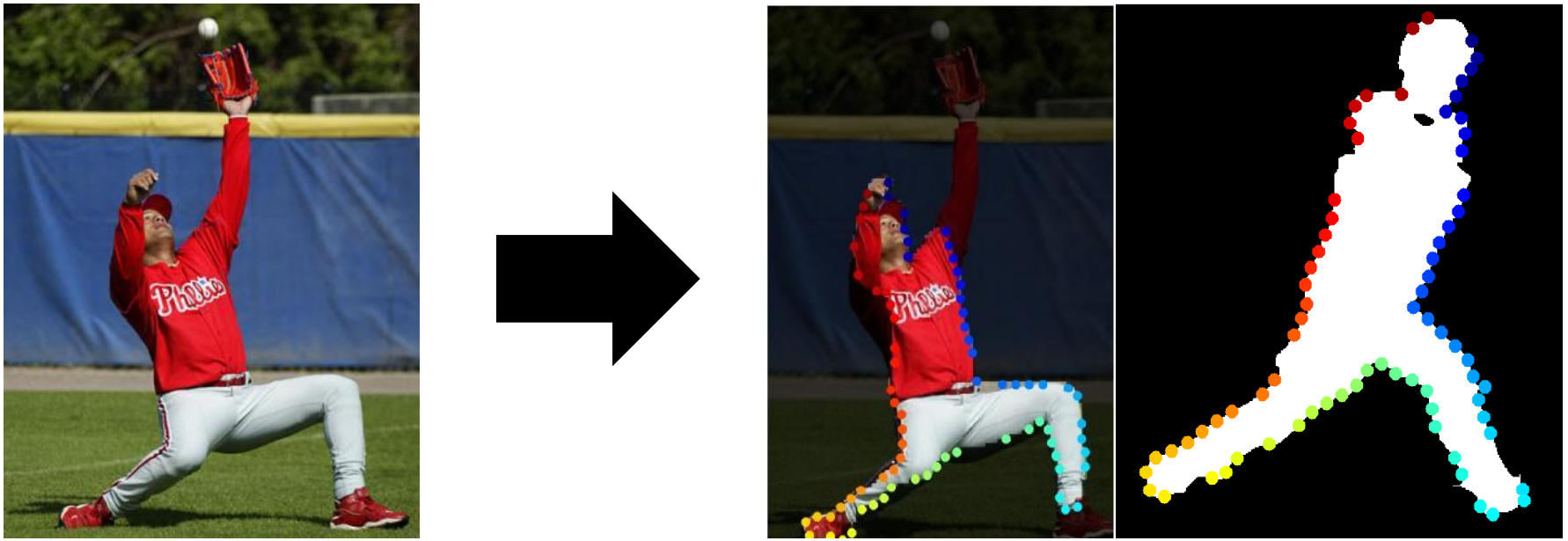
Praveen Srinivasan
Jianbo Shi

GRASP Laboratory,
University of Pennsylvania

GRASP

Penn
University of Pennsylvania

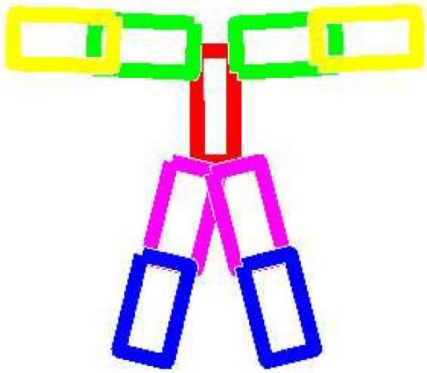
2-D Image: Segmentation, Pose



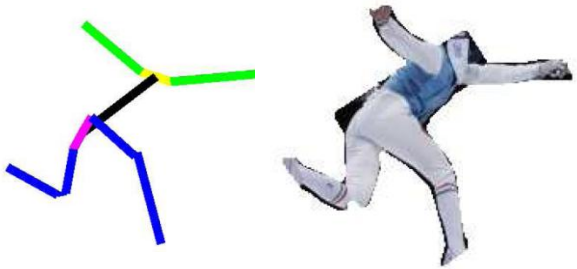
- **Challenge: Large variation in appearance and pose**

Previous work

- Top-down methods: Pictorial structures (Felzenszwalb 2005), (Ramanan 2006), non-parametric belief propagation (Isard 2003), (Sigal 2006)
- Top-down/bottom-up (Mori 2004), (Ren 2005), (Forsyth 1997), (Ioffe 1999)



From (Ramanan 2005)

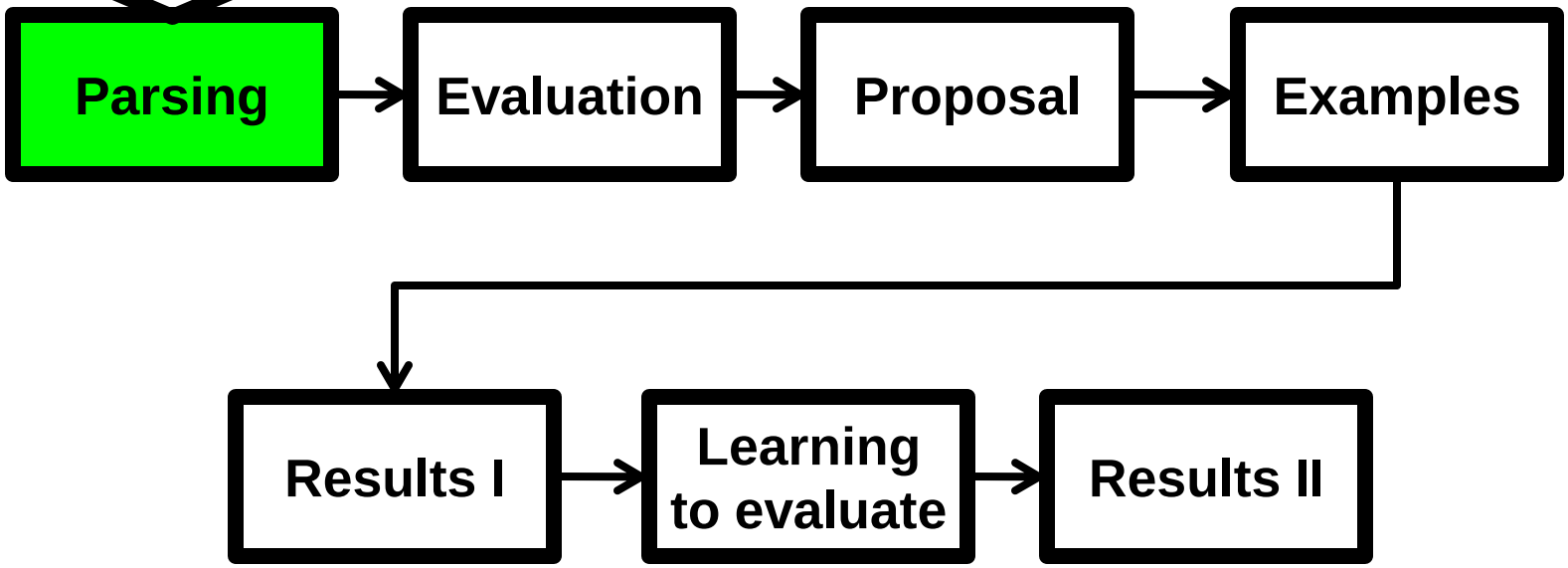


From (Ren 2005)

GRASP

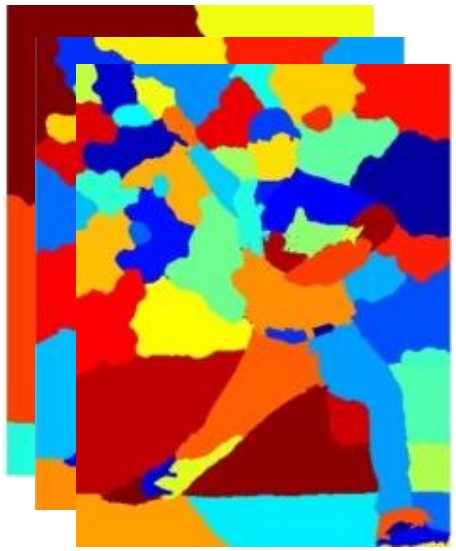
Penn
University of Pennsylvania

- Start with segments
- Segment groupings
- Proposal and evaluation
- Rules guide search



Parsing: We start with segments

- Multiple NCut segmentations provide initial shapes
- 5,10,…,60 segments
- How to group segments into a human figure?



GRASP

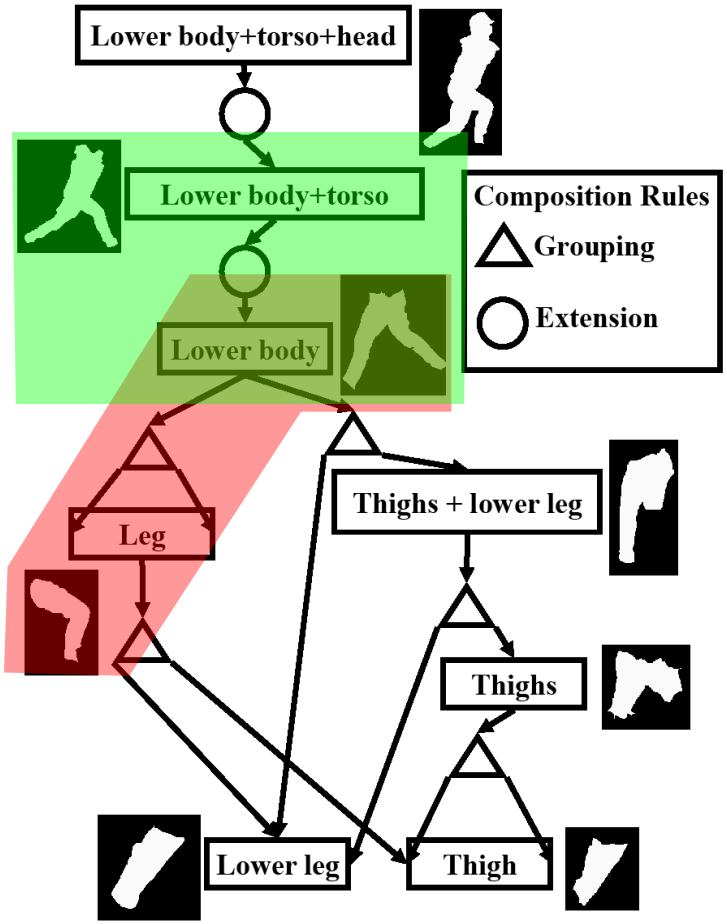
Penn
University of Pennsylvania

Parsing: Segment groupings

- Given body subparts, can group into whole body
- Need proposals for body subparts
- Recursive nature yields bottom-up formulation

Parsing: proposal and evaluation

- Parsing begins at leaves, continues upwards
- Parse rules create proposals for each part (**proposal**)
- Proposals scored according to shape, ranked/pruned (**evaluation**)



GRASP

Penn
University of Pennsylvania

Parsing: rules guide search

- {Lower leg, Thigh} → Leg
- {Thigh, Thigh} → Thighs
- {Thighs, Lower leg} → Thighs+Lower leg
- {Thighs+Lower leg, Lower leg} → Lower body
- {Leg, Leg} → Lower body
- {Lower body} → Lower body+torso
- {Lower body+torso} → Lower body+torso+head

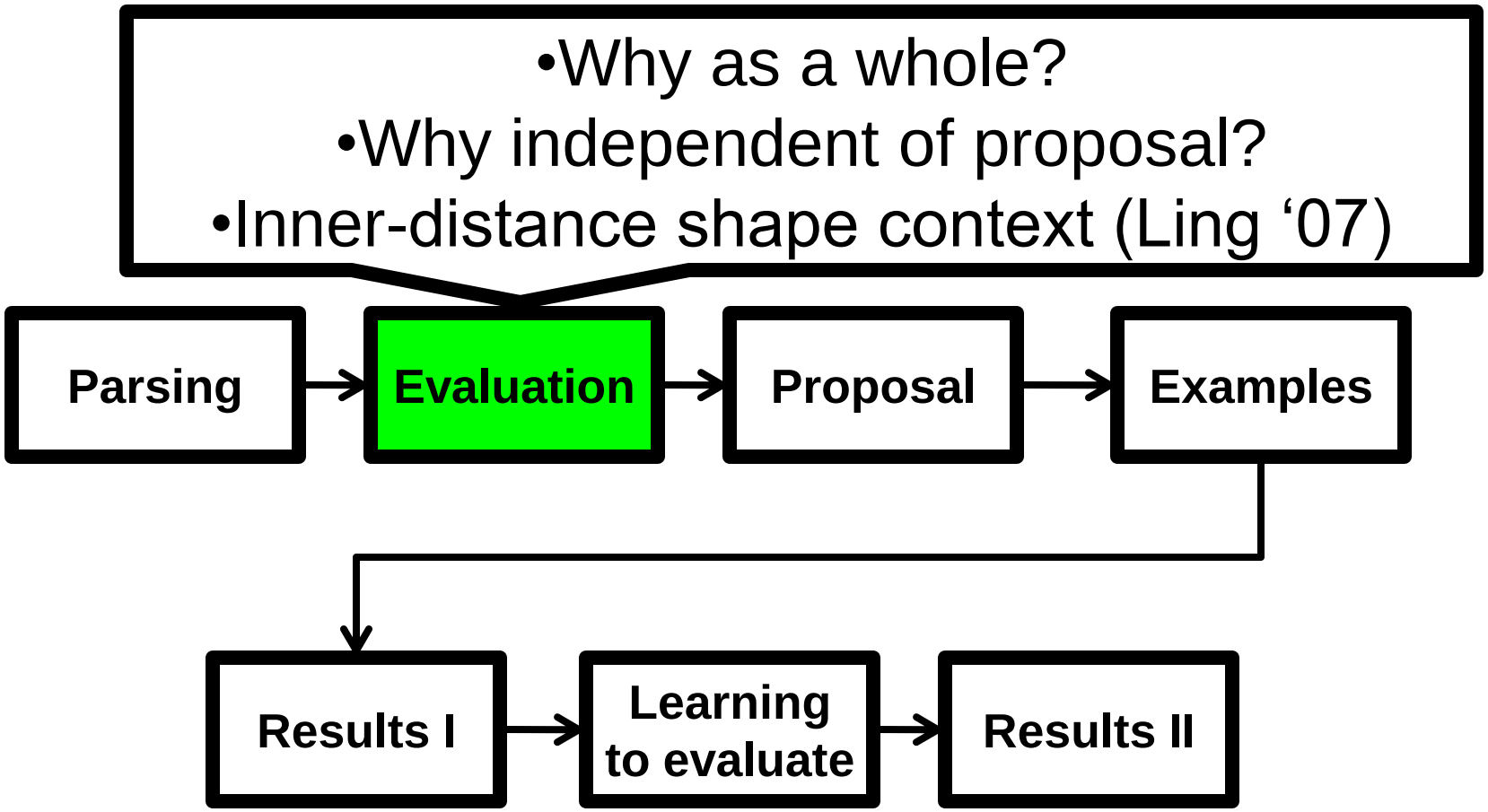
Figure 2. Our parse rules. We write them in reverse format to emphasize the bottom-up nature of the parsing.

GRASP

Penn
University of Pennsylvania

Proposal vs. Evaluation

- Proposals are evaluated as a whole by shape matching to exemplars
- Proposal and evaluation are **disjoint**
- Evaluation is as a **whole** and **independent of proposal**

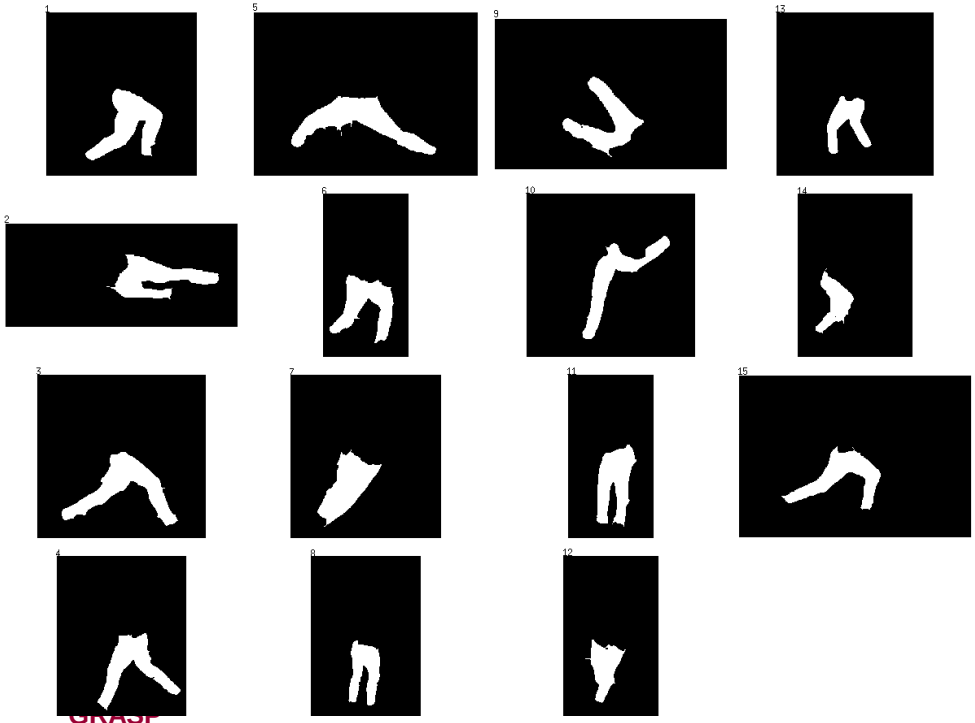


GRASP

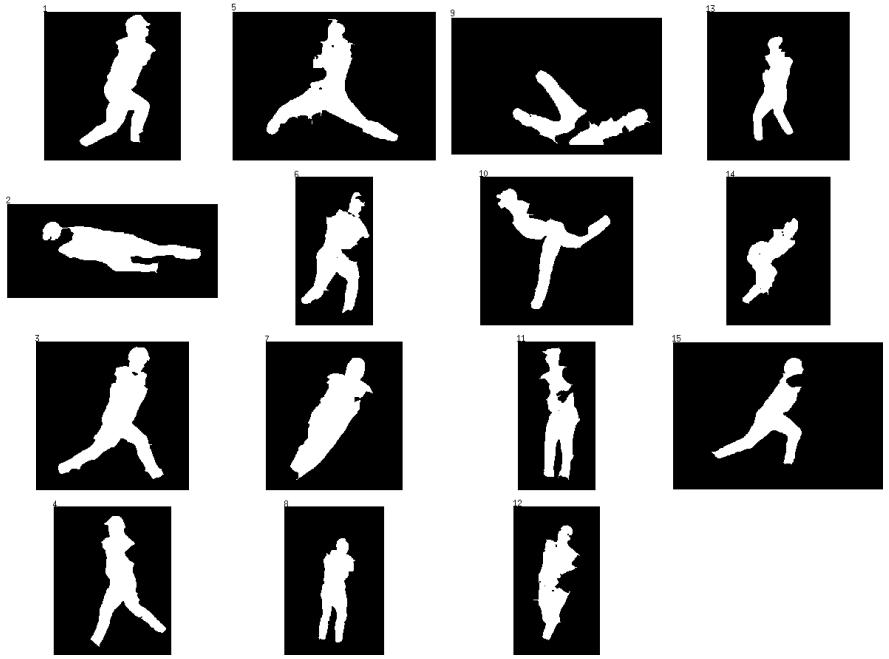
Penn
University of Pennsylvania

Evaluation: why as a whole?
Shape is **locally ambiguous, globally distinctive**

Lower body exemplars



Lower body+torso+head exemplars



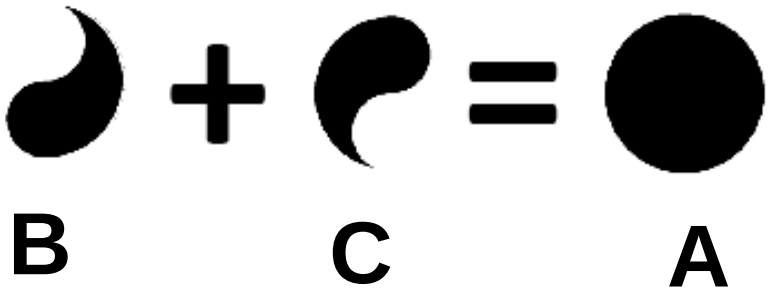
GRASP

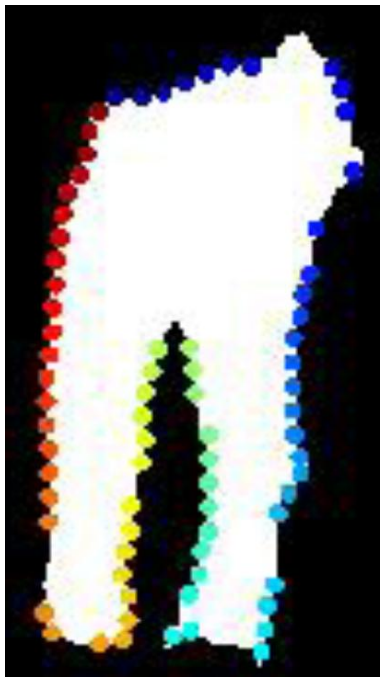
Penn
University of Pennsylvania

Evaluation: why independent of proposal?

Whole ≠ sum of its parts

- Shapes B, C alone do not appear disk-like
- Viewed together, disk perception is clear
- **Shape needs to be evaluated in large context**
- **Want evaluation of A to depend only on A**



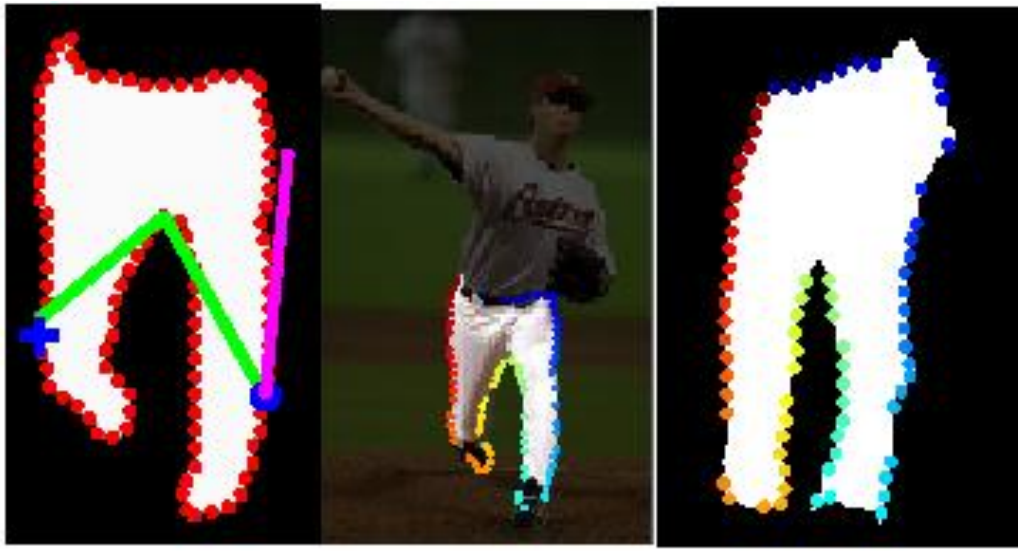


GRASP

Penn
University of Pennsylvania

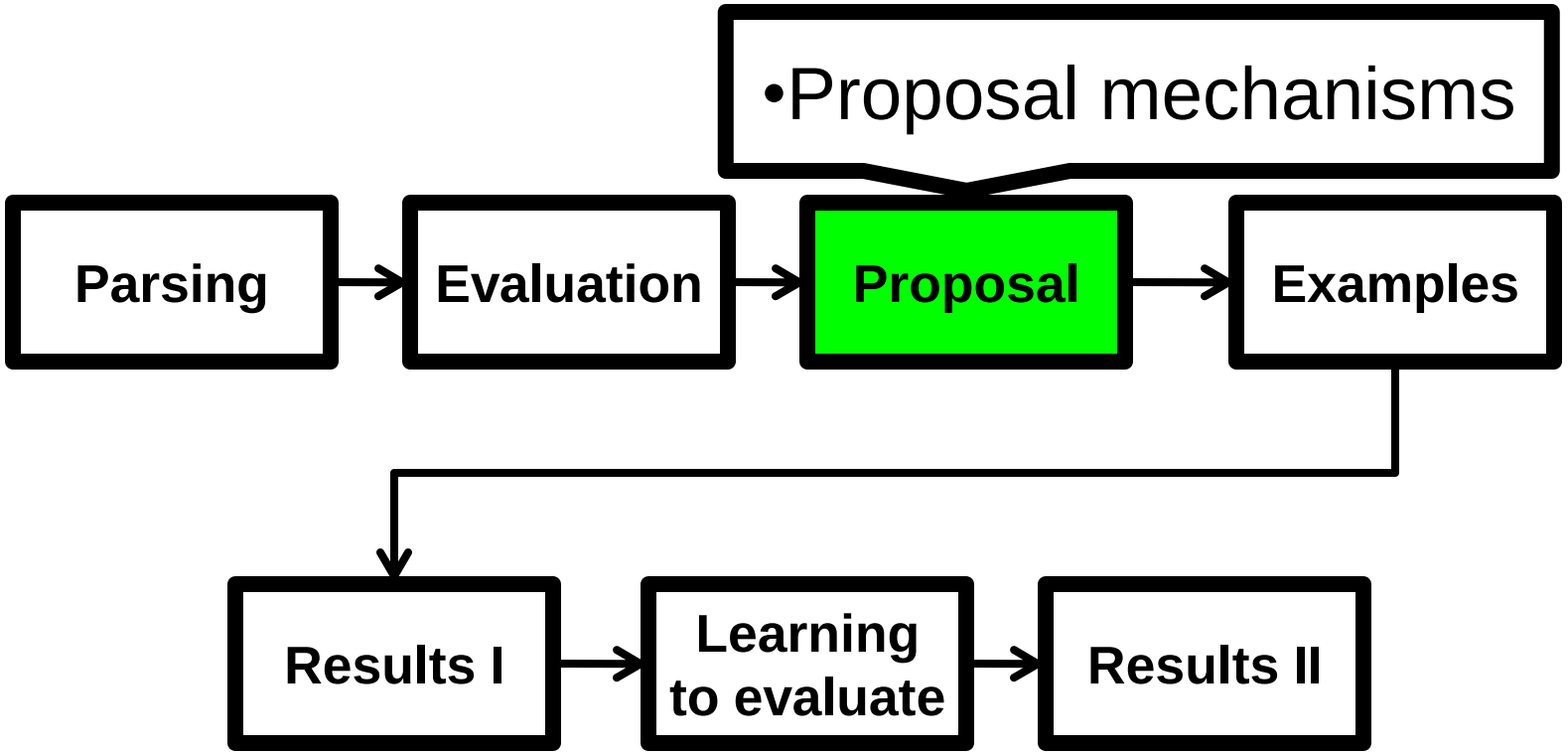
Evaluation: Inner-distance shape context (Ling 2007)

- IDSC used for shape comparison
- Invariant to scale, rotation
- Robust to articulation
- Used to evaluate proposals against exemplar shapes



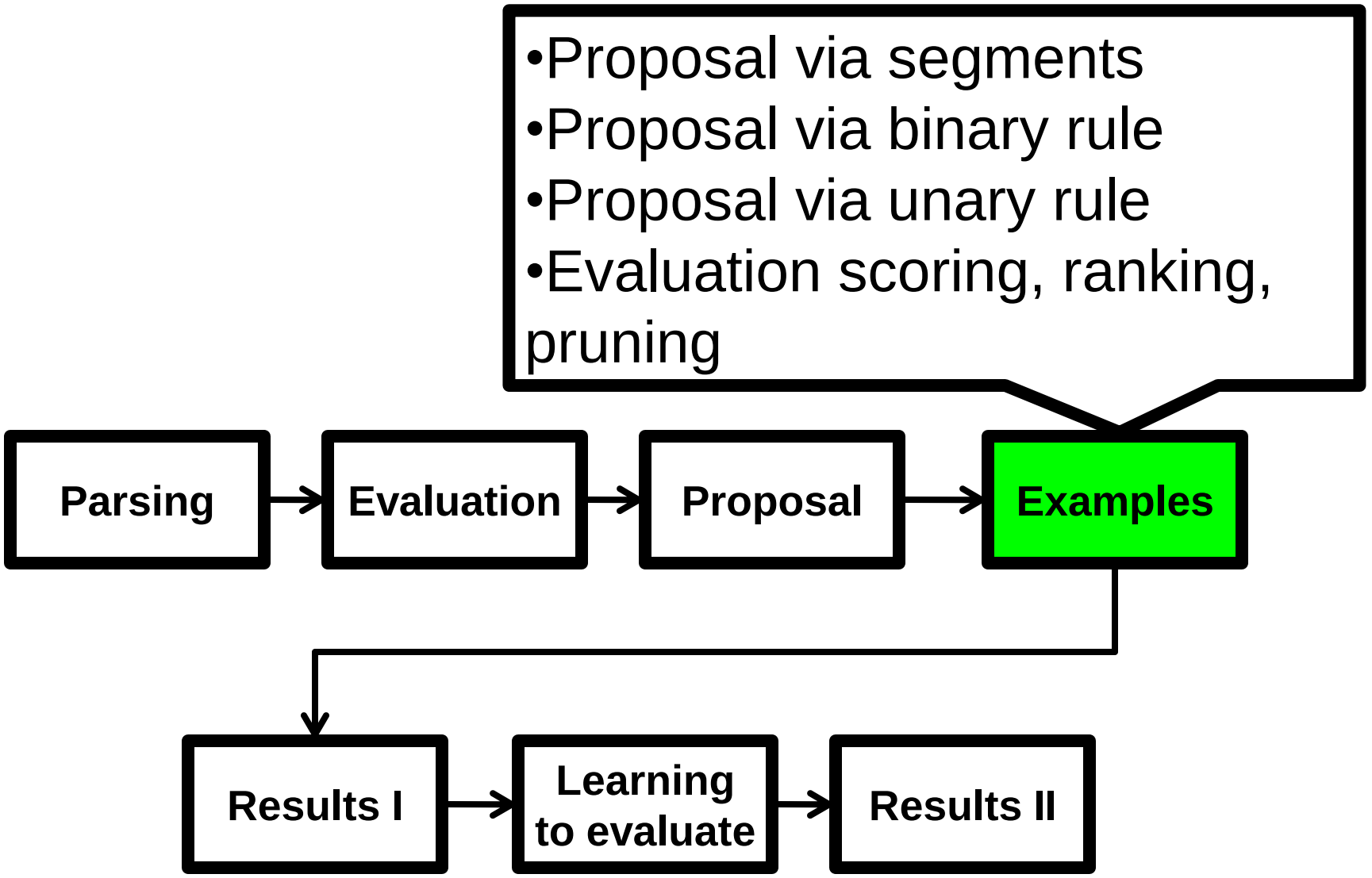
GRASP

Penn
University of Pennsylvania


•Proposal mechanisms
Parsing
Evaluation
Proposal
Examples
Results I
Learning to evaluate
Results II

Proposal: mechanisms

- Proposals from individual segments (applies to all parts)
- Proposals from a pair of subparts (B,C -> A, e.g. Leg,Leg->Lower body)
- Proposals from a single subpart (B->A, e.g. Lower body -> Lower body+torso)



GRASP

Penn
University of Pennsylvania

Proposal: segments

Multiple segmentations

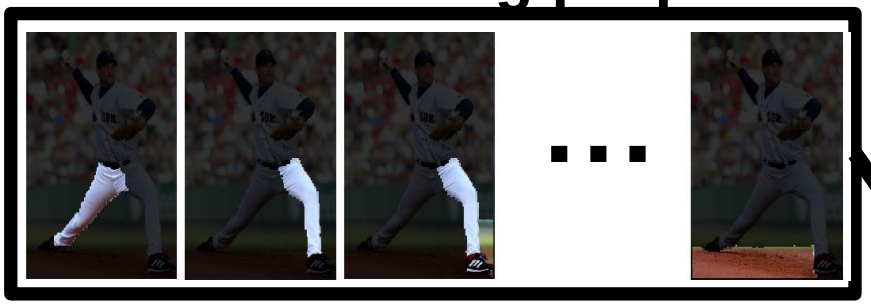
Proposals from segments

GRASP

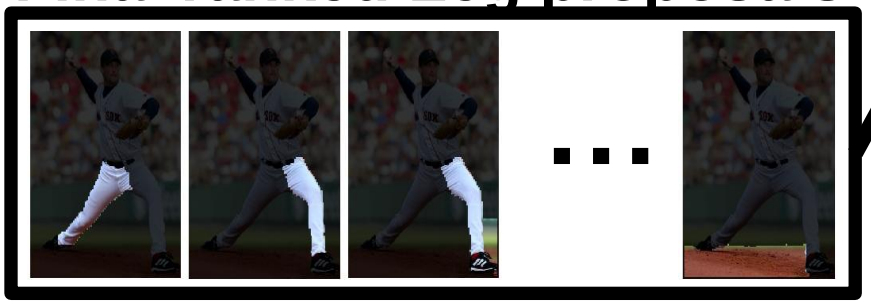
Penn
University of Pennsylvania

Proposal: binary rule (Leg,Leg->Lower body)

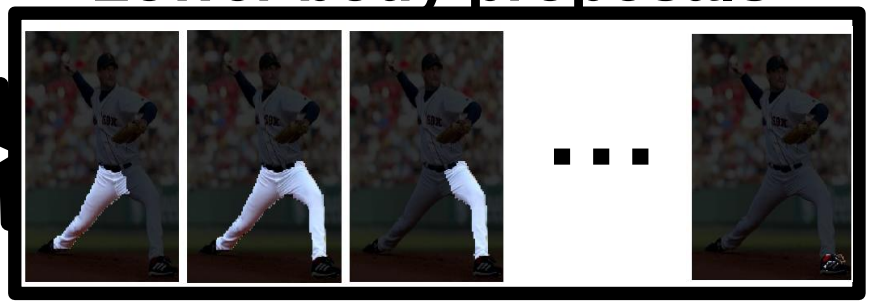
Final ranked Leg proposals



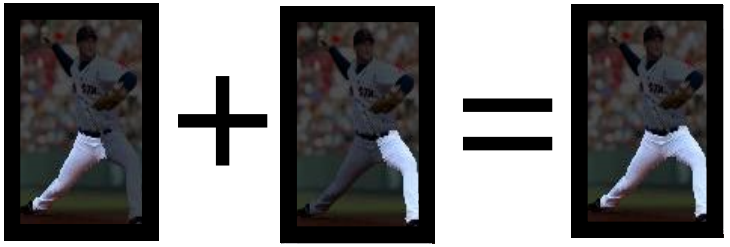
Final ranked Leg proposals



Lower body proposals



Grouping example

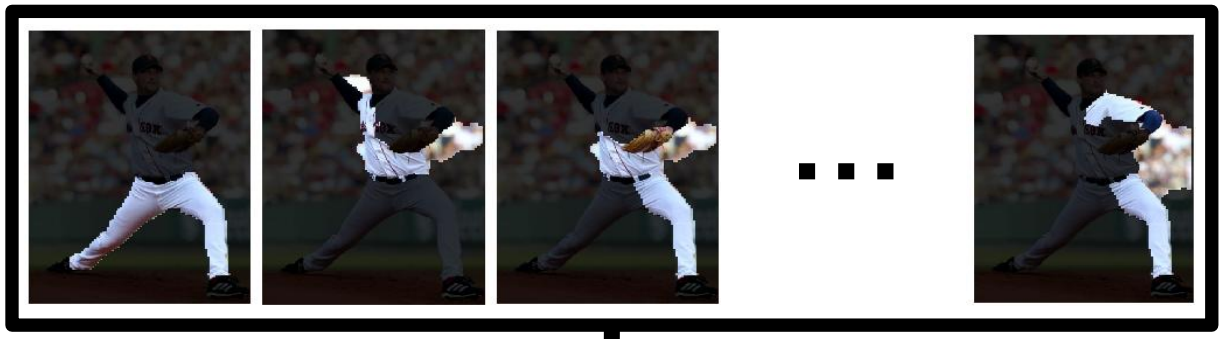


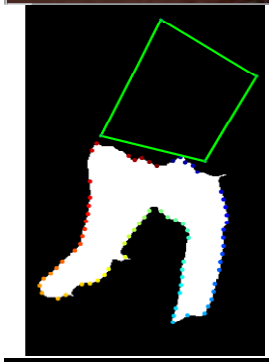
GRASP

Penn
University of Pennsylvania

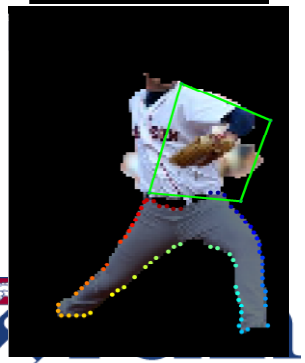
Proposal: unary rule (Lower body-> Lower body+torso

Final ranked Lower body proposals





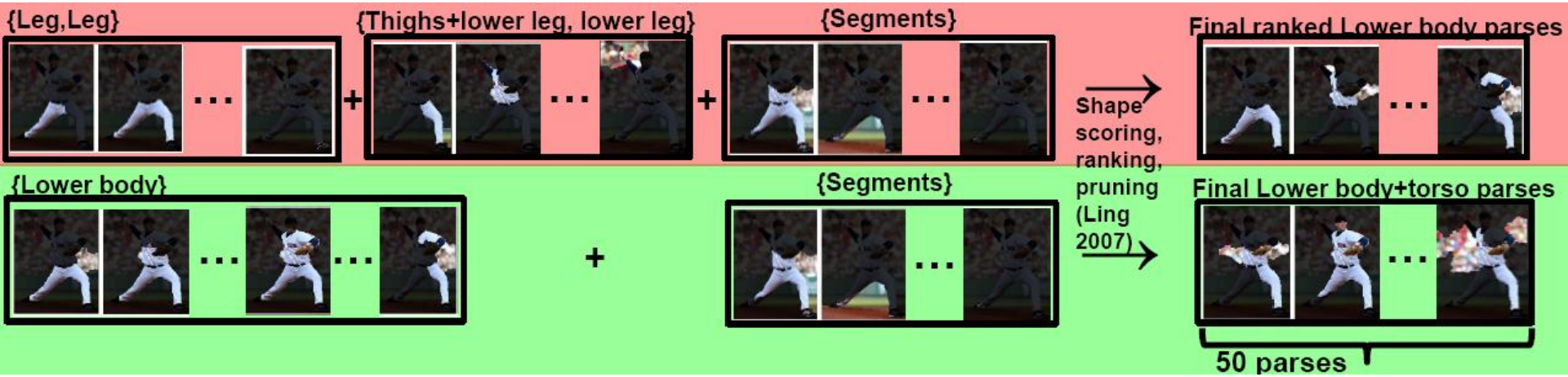
Lower body+ torso proposals



GRASP

University of Pennsylvania

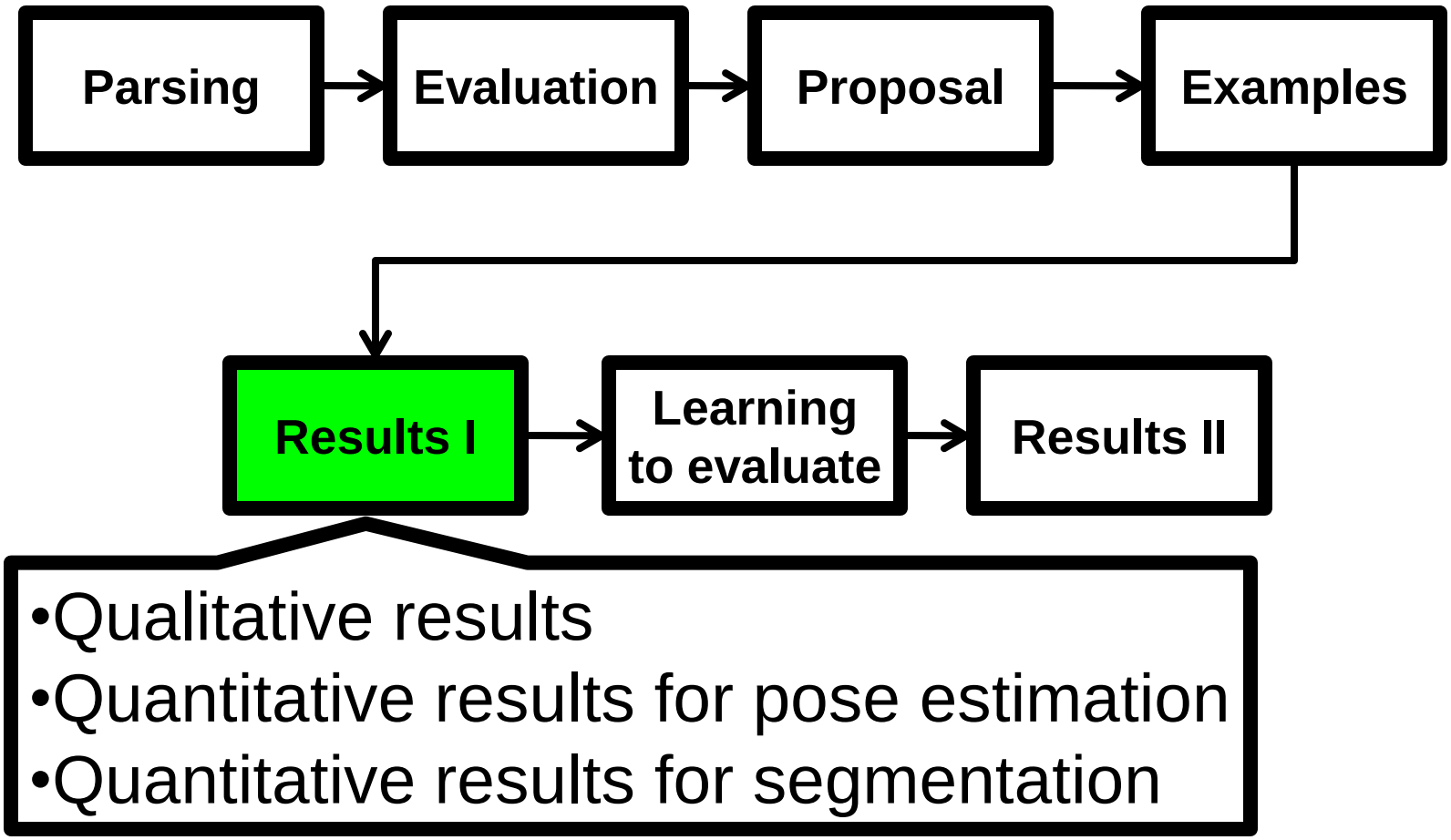
Evaluation: scoring, ranking, pruning



- After all proposals generated for a part, they are shape scored, ranked, pruned to a constant # of proposals

GRASP

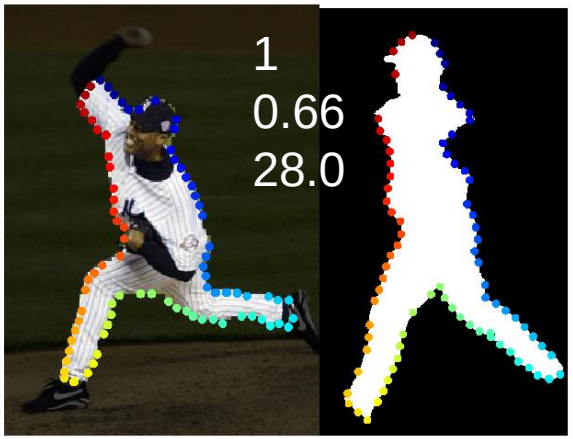
Penn
University of Pennsylvania

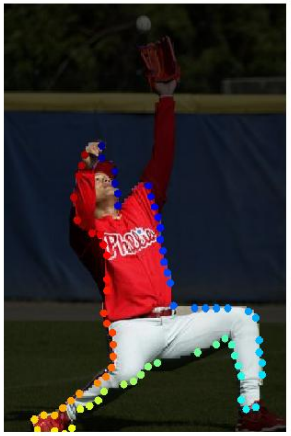

Parsing
Evaluation
Proposal
Examples
Results I
Learning to evaluate
Results II
•Qualitative results
•Quantitative results for pose estimation
•Quantitative results for segmentation

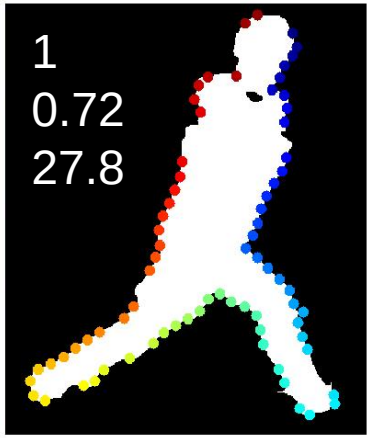
Results

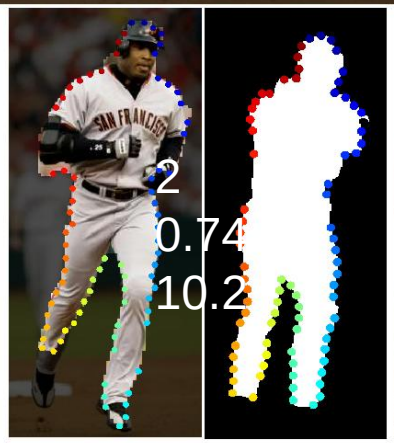
- End result of parsing: 50 ranked proposals for each part
- Shape only cue; typically find good proposal in top 10
- Exclude arms – could add separate layers for arms (self-occlusion)

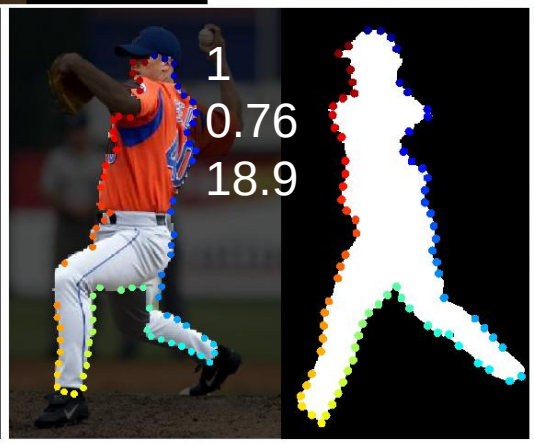

Rank: 1
Overlap: 0.83
Jt. err: 12.3

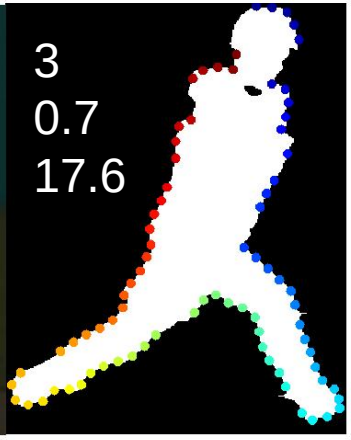

1
0.66
28.0

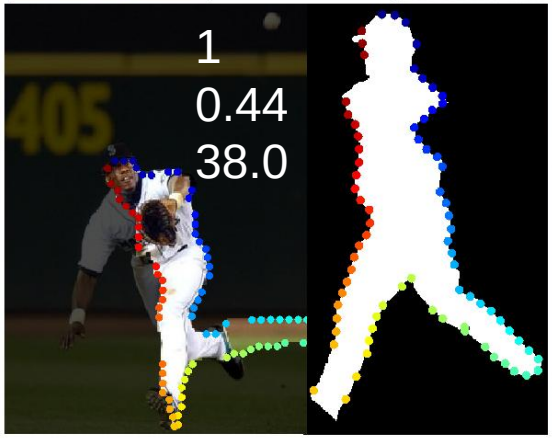



1
0.72
27.8

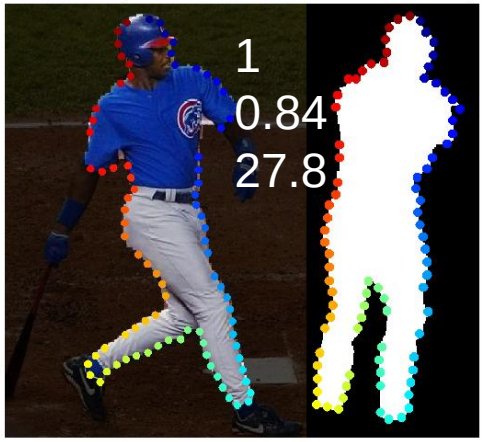

2
0.74
10.2


1
0.76
18.9


3
0.7
17.6


1
0.44
38.0

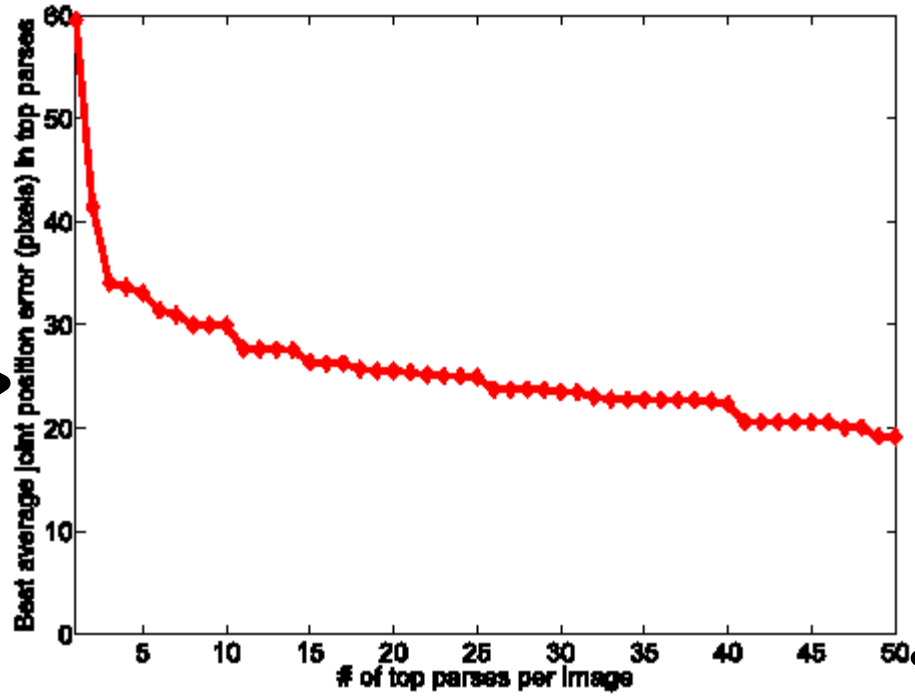

3
0.57
18.2


1
0.84
27.8

Quantitative results – pose estimation

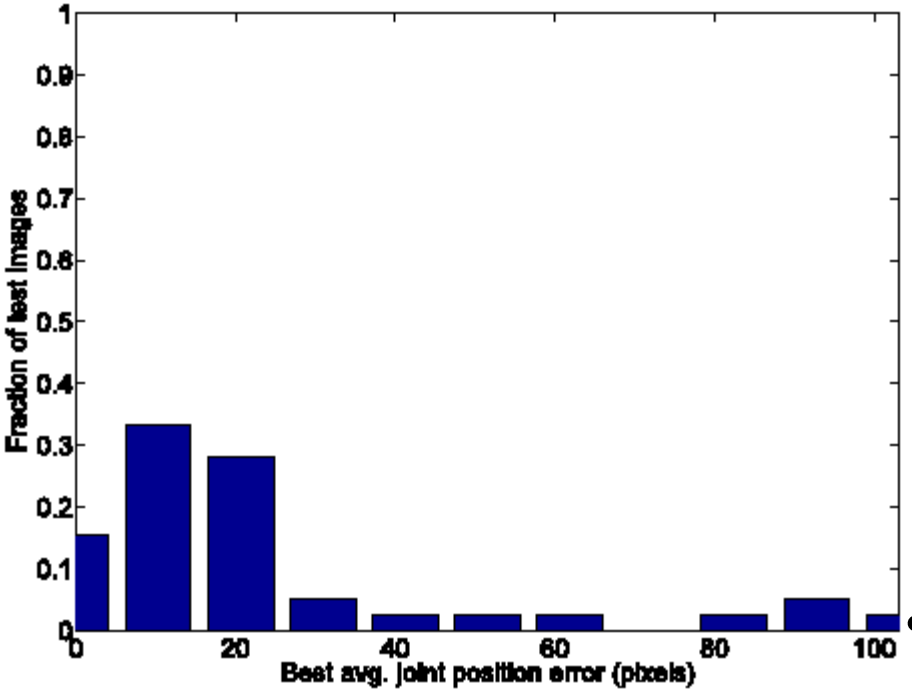
- For full body proposals, projected positions of 5 joints (knees, hips, neck)
- Computed average error in pixels
- For top k proposals in an image, took minimum error
- Plot average of minimum error as k varied
- Histogram of minimum errors for k = 10

Best average jt. error (pixels) in top k proposals



# of top proposals per image (k)

Fraction of test images



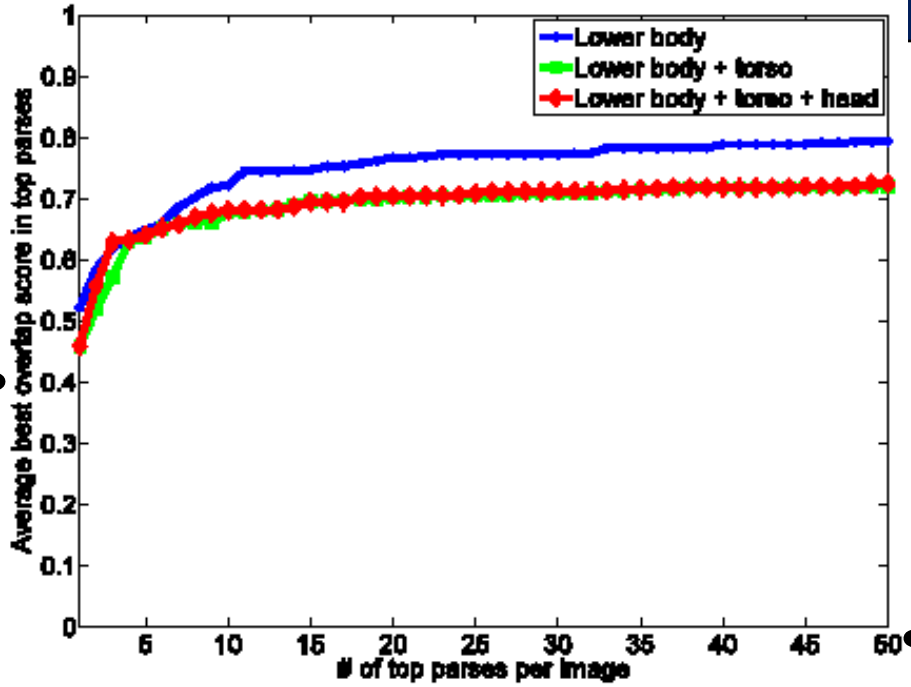
Best avg. joint position error (pixels) in top 10 proposals

GRASP

Quantitative results - segmentation

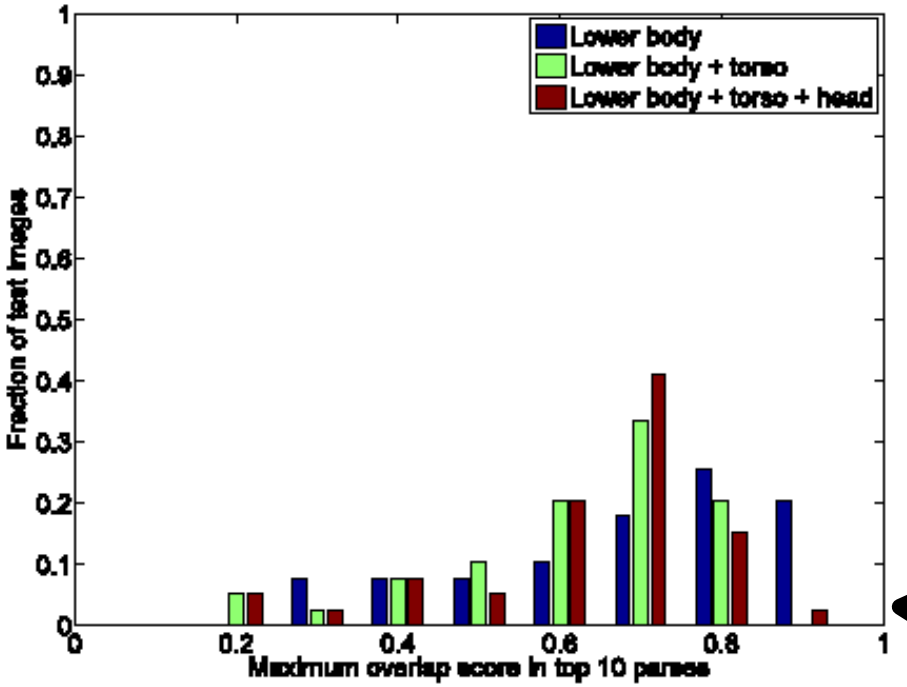
- For top k proposals, computed overlap score with ground truth mask
- Took maximum over top k proposals in each image
- Plotted average of max values across all images
- Histogram of max values for k = 10

Best average overlap score in top k proposals



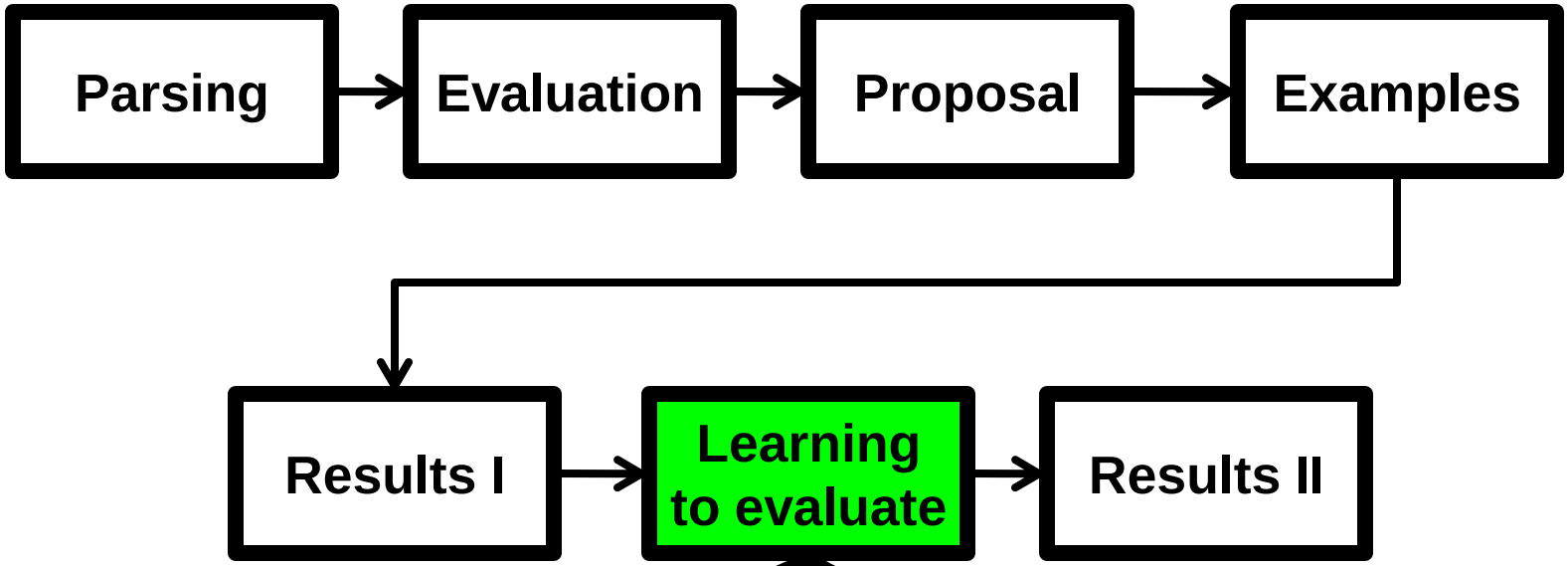
# of top proposals per image (k)

Fraction of test images



Best overlap score in top 10 proposals

GRASP



- Evaluation with additional features
- Learning a ranking function
- Energy function
- Learning methodology
- Learning results

GRASP

Penn
University of Pennsylvania

Evaluation with additional features

- Shape is not the only meaningful feature
- Boundary information
- Internal texture
- Can rank proposals by a variety of different features

Learning a ranking function for each part

- Given images $j = 1, ...m$, feature vector f_i^j and score $s_i^j \geq 0$ for each proposal i
- Learn ranking function $F(f_i^j) = w^\mathsf{T} f_i^j$
- Want $s_a^j > s_b^j \Rightarrow F(f_a^j) > F(f_b^i)$
- Ranking function trained using proposals generated by previously learned ranking functions

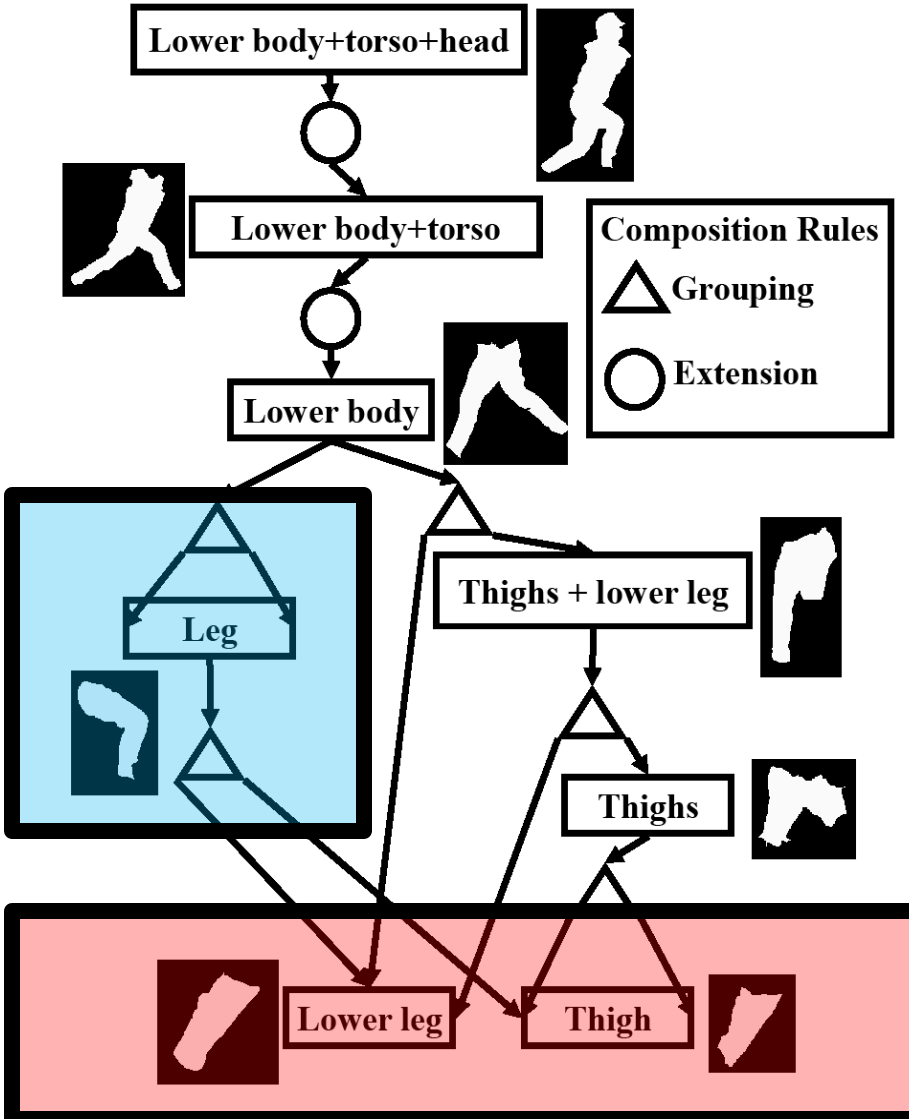
w

GRASP

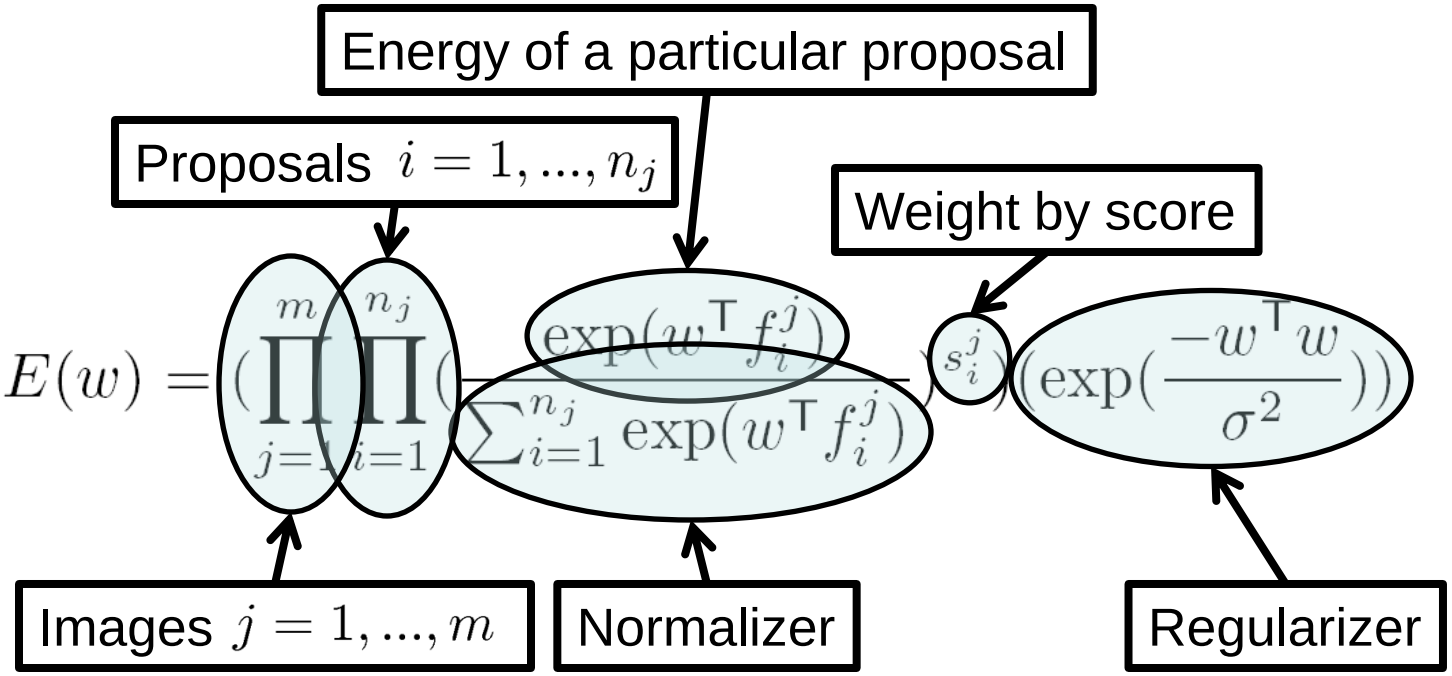
Penn
University of Pennsylvania

Learning a ranking function

- w for Leg trained using proposals for lower leg, thigh ranked using respective w



Energy function (weighted multinomial logistic regression)



GRASP

Penn
University of Pennsylvania

Learning methodology

- Texture:
  - Built a 200 entry codebook of SIFT features from ~400 additional baseball images via K-Means
  - For each proposal, computed histogram of SIFT codebook occurrences inside
- Boundary: Average PB along boundary
- Shape: IDSC distance to best matching exemplar

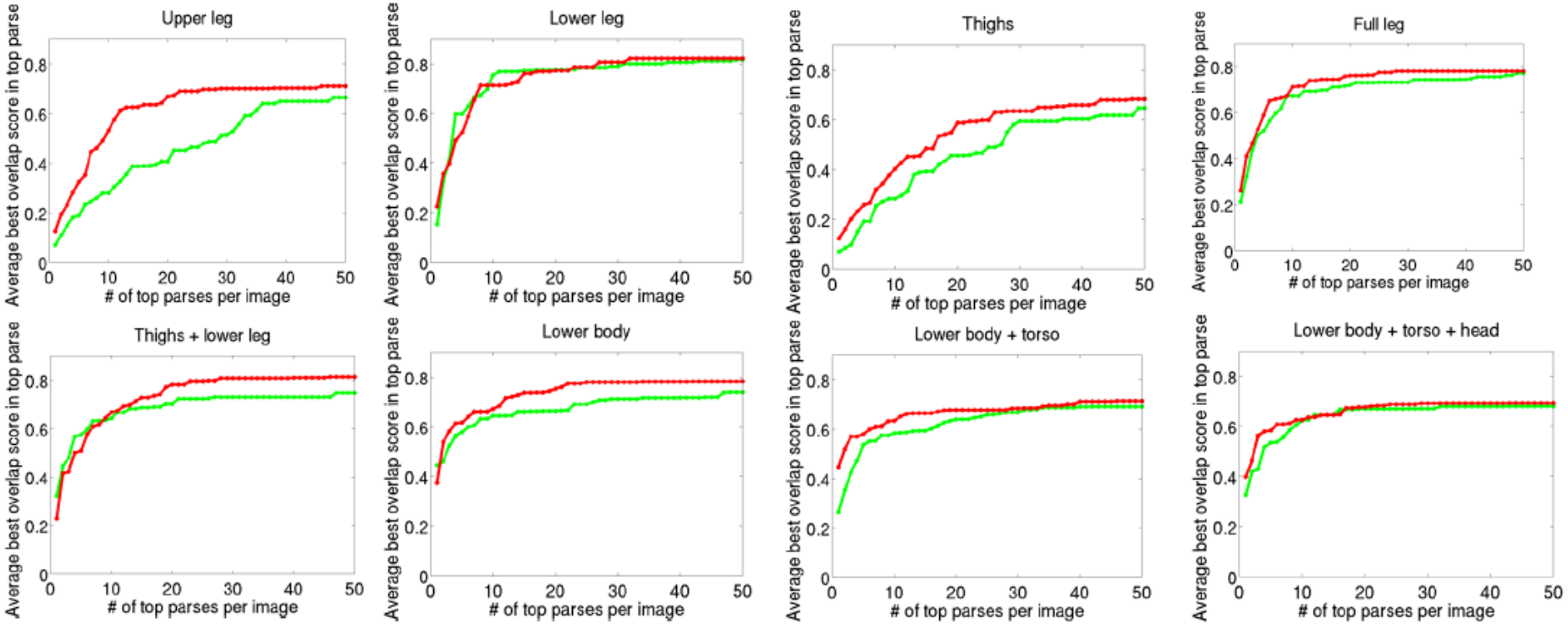
GRASP

Penn
University of Pennsylvania

Learning methodology, cont.

- In addition to previous 15 training images, used an additional 16 images for learning w for each part
- 10-fold cross validation was performed to select sigma (in regularizer) for each part
- Tested on 26 more images

Learning results



With learning

Shape only

GRASP

Penn
University of Pennsylvania

Future work

- Phrase search in A* framework
- Incorporate contour cues into grouping, extension and scoring
- Explore other shape scoring cues
- Incorporate other features

GRASP

Penn
University of Pennsylvania

Summary and conclusion

- Shape needs to be evaluated in larger context
- Proposal and evaluation can be separated to improve parsing