Part-Pair Representation for Part Localization

Jiongxin Liu, Yinxiao Li, and Peter N. Belhumeur {liujx09, yli, belhumeur}@cs.columbia.edu Columbia University

Contribution

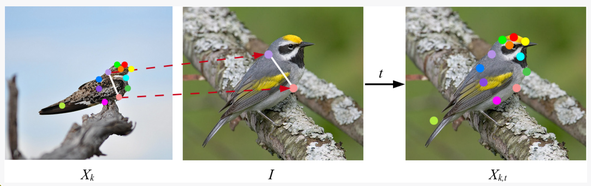
We propose a novel part-pair representation to localize the parts of deformable objects.

Our work focuses on two aspects:
1. Rich appearance models
2. Flexible and reliable spatial models

Our method achieves good performance on two challenging datasets: CUB-200-2011 [1] and Leeds Sports Poses [2].

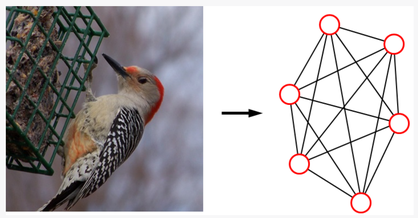
Background [3]

Exemplar-based method: local part detectors model the appearance of an object, and exemplars encode the geometry of the object. Consensus of highest scoring exemplars predicts the part locations.



Part-Pair Representation

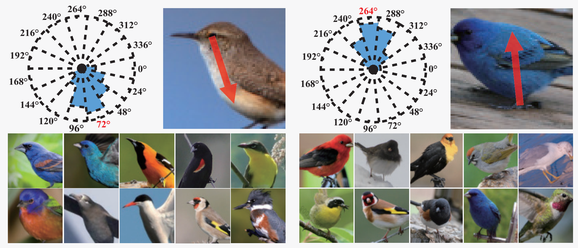
An object is represented as a complete set of part pairs.



Properties:
1. Redundancy in the appearance models (the appearance of each part pair is modeled)
2. Each pair carries the orientation and scale information
3. We can customize the graph structure for different localization tasks

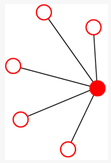
Pair Detector

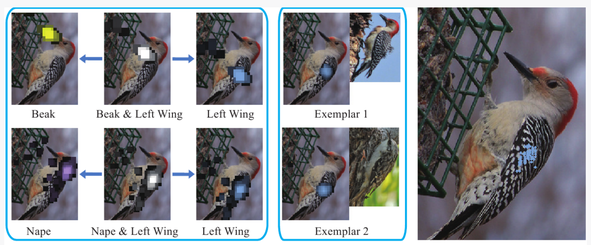
We build mixtures of pair detectors. Each detector targets a specific orientation of the part pair. Each pair activation casts a vote for its two corresponding parts.



Super Part Detector

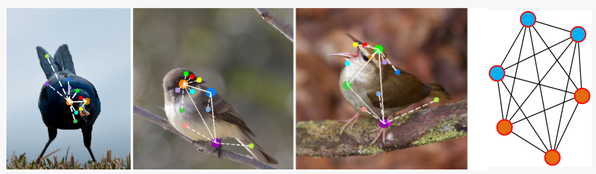
- The goal of super part detector is to localize an individual part accurately
- A **star graph** is suitable for such task
- A super part detector is conditioned on an exemplar



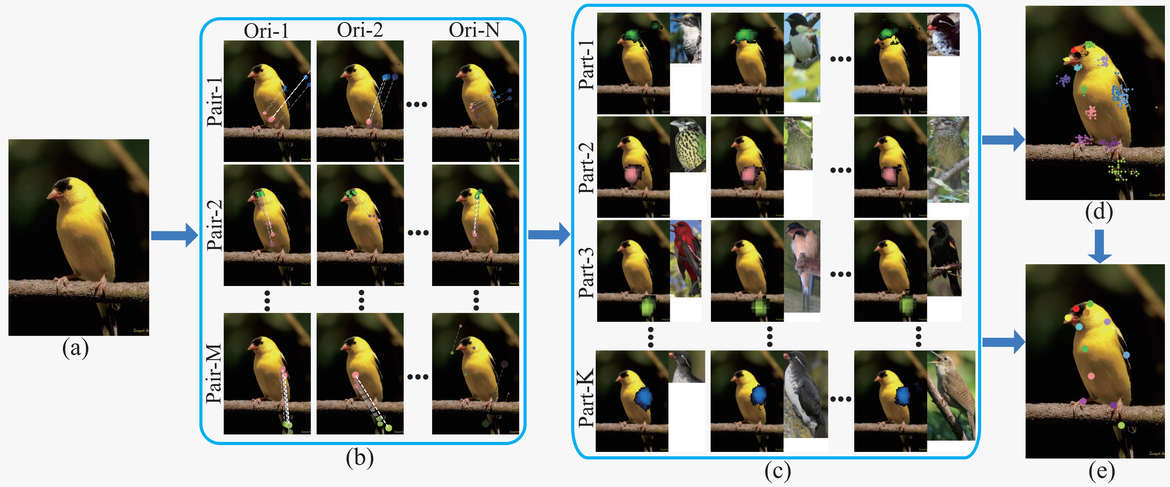


Predicting the Part Configuration

We can do this in two ways:
1. Consensus of exemplars using pair detection (Ours-r)
- Place exemplars at the candidate part locations
- Evaluate the exemplars by averaging the scores of all its pairs (**complete graph**)
- Make the consensus
2. Flexible composition of part hypotheses (Ours-f)
- Generate groups of compatible part hypotheses & exemplars
- Construct the graph structure for each group
- Pool the scores from the constituting pairs
- Evaluate the score vector using a regression model
- The best group is used to predict the configuration
- Graph examples:



Pipeline



(a)Testing image (b) Pair detection (c) Part detection (d) Part hypotheses (e) Output

Bird Part Localization

Back Beak Belly Breast Crown Forehead Left Eye Left Leg Left Wing Nape Right Eye Right Leg Right Wing Tail Throat

PCP	Ba	Bk	Be	Cr	Le	Ta	All
CoE [3]	62.1	49.0	69.0	72.9	40.7	40.2	59.7
Ours-r	59.7	59.0	69.5	**77.1**	39.9	34.7	63.1
Ours-f	**64.5**	**61.2**	**71.7**	76.8	**45.0**	**46.2**	**66.7**

Performance of three methods on localizing the parts jointly

Human Pose Estimation

PCP	To	Ul	Ll	Ua	Fa	He	All
CoE [3]	83.4	69.0	61.7	47.5	28.1	79.3	57.5
Ours-r	84.2	69.3	61.5	48.7	28.5	79.9	58.0
Ours-f	**87.6**	**76.4**	**69.7**	**55.4**	**37.6**	**82.0**	**64.8**

Acknowledgements

This work was supported by NSF award 1116631, ONR award N00014-08-1-0638, and Gordon and Betty Moore Foundation grant 2987.

References

[1] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. *Computation & Neural Systems Technical Report*, CNS-TR-2011-001, 2011

[2] S. Johnson and M. Everingham. Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation. In *BMVC '10*

[3] J. Liu and P. N. Belhumeur. Bird Part Localization Using Exemplar-Based Models with Enforced Pose and Subcategory Consistency. In *ICCV '13*