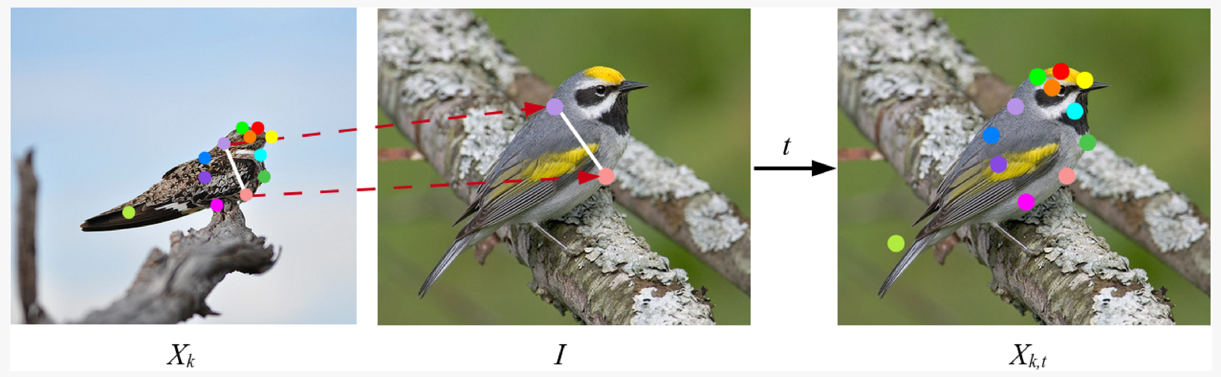
Bird Part Localization Using Exemplar-Based Models with Enforced Pose and Subcategory Consistency

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

Problem

The goal of our work is to localize the parts automatically and accurately for fine-grained categories. We evaluate our method on bird images in the CUB-200-2011 [1] dataset.

Approach



Does $X_{k,t}$ match the image I? $\Longleftrightarrow P(X_{k,t}|I) = ?$

$$P(X_{k,t}|I) = P(X_{k,t}|D_p)^{\alpha} P(X_{k,t}|D_s)^{1-\alpha} \quad (1)$$

$$P(X_{k,t}|D_p) = \mathrm{G}_{\text{avg}}\{P(x^i_{k,t}|d^i_p[c^i_k, s^i_{k,t}])\} \quad (2)$$

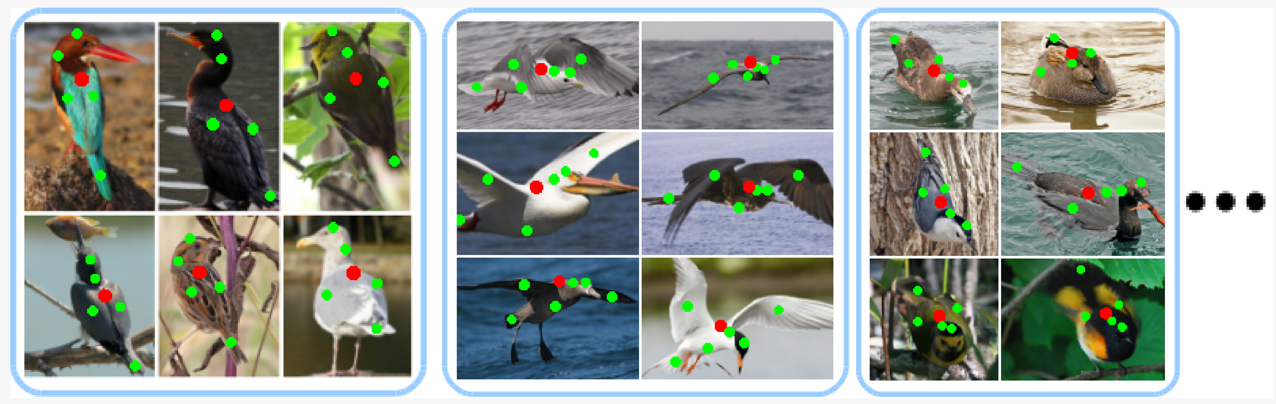
$$P(X_{k,t}|D_s) = \max_l P(X_{k,t}|l, D_s) \quad (3)$$

$$P(X_{k,t}|l, D_s) = \mathrm{G}_{\text{avg}}\{P(x^i_{k,t}|d^i_s[l, s^i_{k,t}, \theta^i_{k,t}])\} \quad (4)$$

We use the most likely models $\mathcal{M}$ to predict the part locations of the testing sample:

$$\hat{x}^i = \arg\max_{x^i} \sum_{k,t \in \mathcal{M}} P(\triangle x^i_{k,t}) P(x^i|d^i_p[c^i_k, s^i_{k,t}]) \quad (5)$$

Pose Detectors

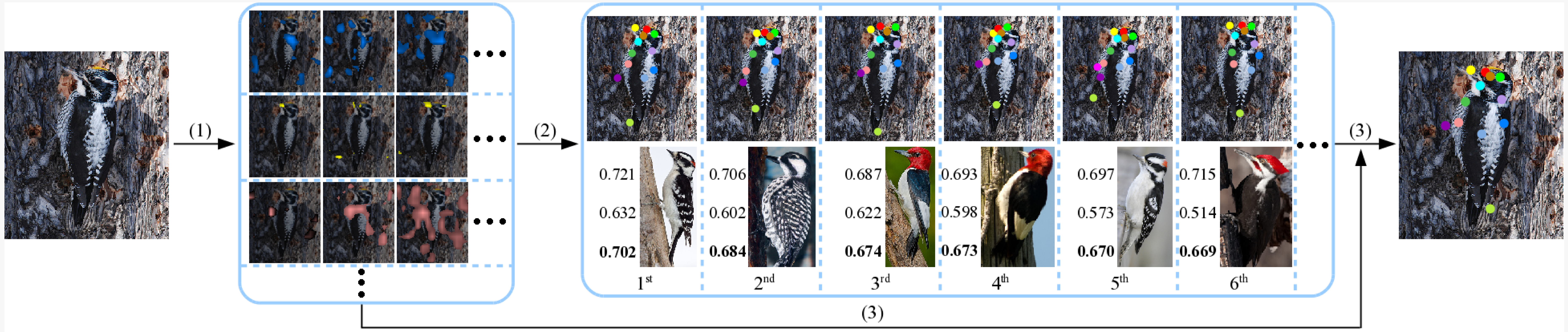


Pose 1 Pose 2 Pose 3

Poses clusters of Back

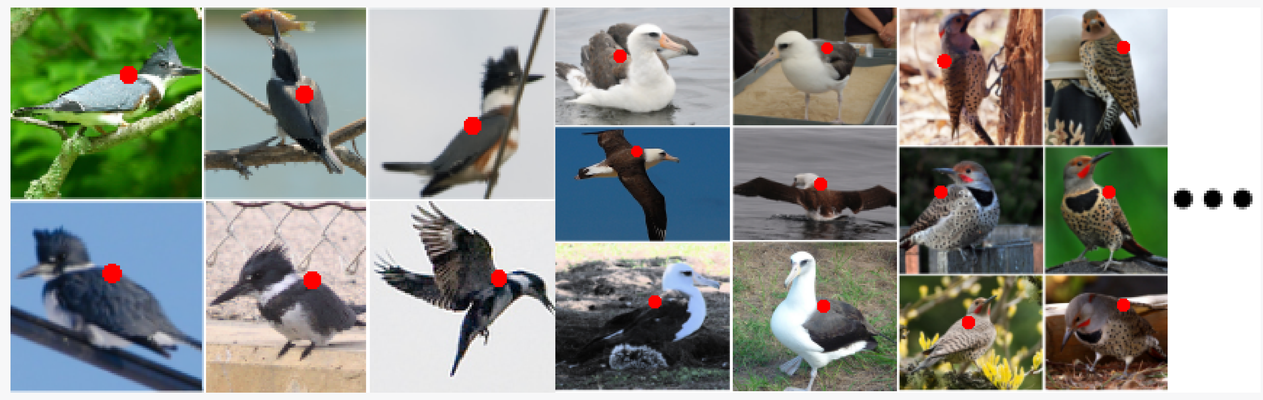
For each pose cluster c^i of part i, we build a detector. The detector scans the image over scales, and the response map of this detector at a particular scale s^i is denoted as $d^i_p[c^i, s^i]$.

Pipeline



(1) Sliding-window detection. (2) Matching and ranking exemplars. (3) Predicting the final part configuration.

Subcategory Detectors

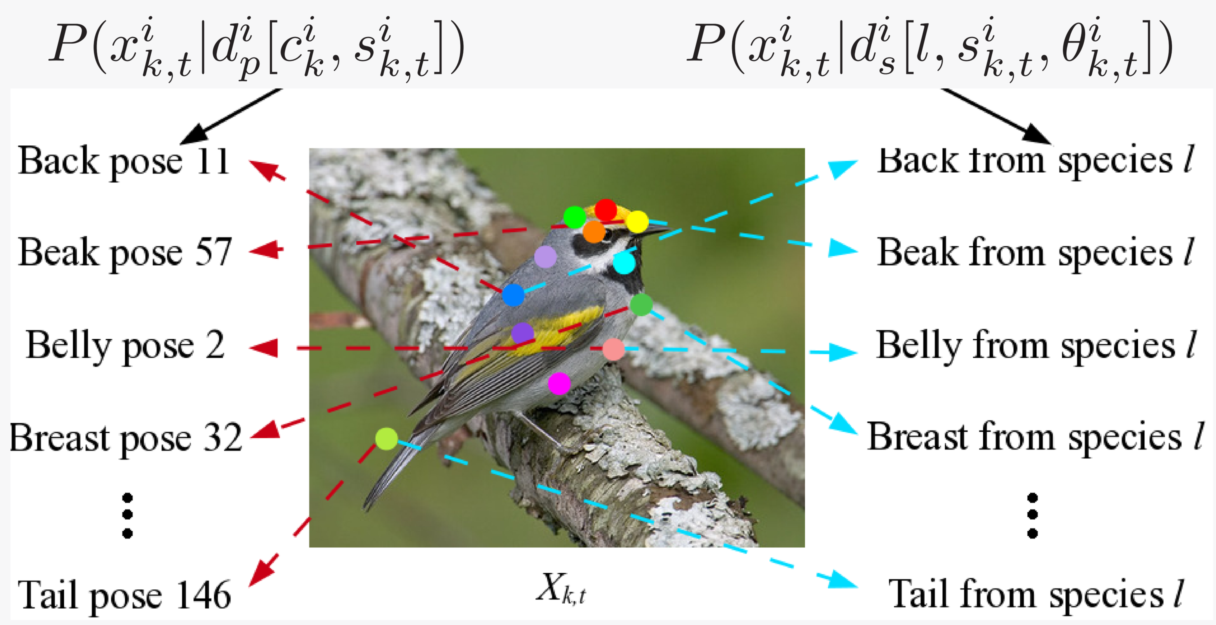


Species 1 Species 2 Species 3

Subcategory clusters of Back

For each species l of part i, we build a detector after aligning the samples. Assuming the detector scans the image over scales and orientations, then the response map of this detector at a particular scale s^i and orientation θ^i is denoted as $d^i_s[l, s^i, \theta^i]$.

Enforcing Consistency

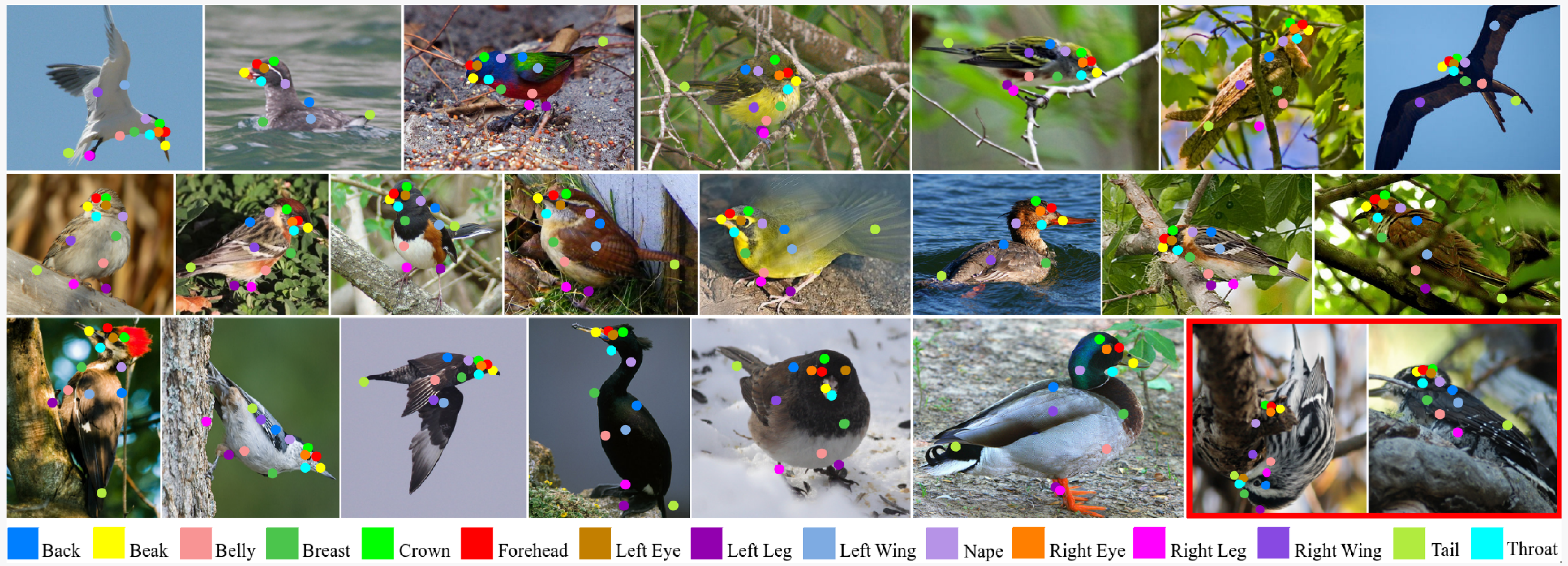


References

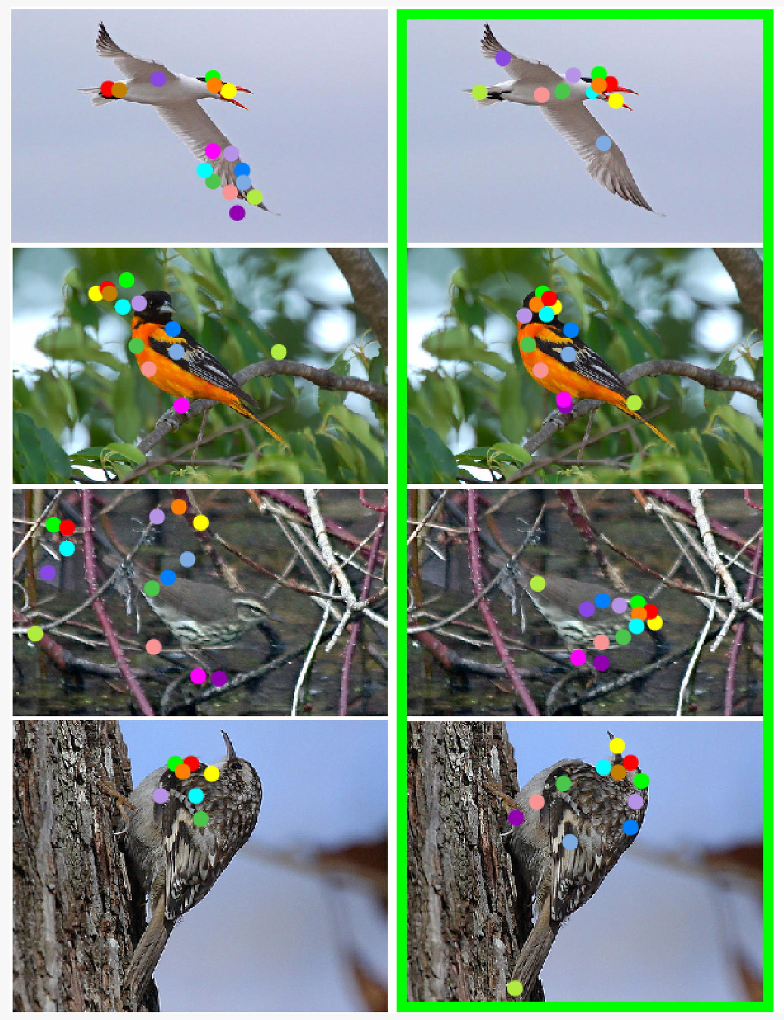
[1] C. Wah, S. Branson, P. Welinder, P. Perona, S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. *Computation & Neural Systems Technical Report*, CNS-TR-2011-001, 2011

[2] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, N. Kumar. Localizing Parts of Faces Using a Consensus of Exemplars. In *CVPR '11*

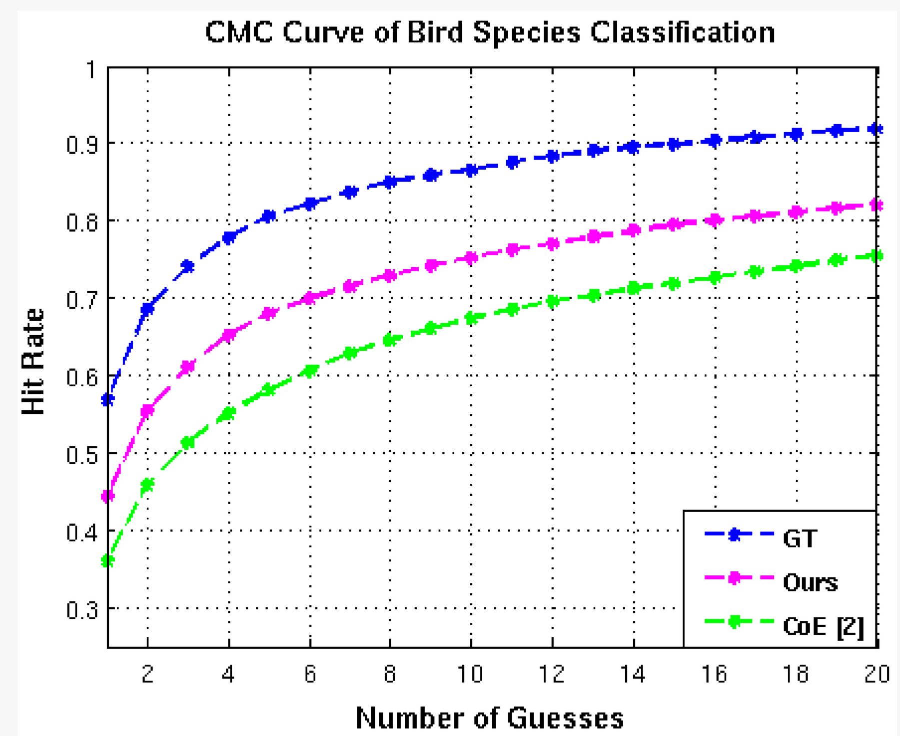
Localization Examples



Comparisons



PCP	CoE [2]	**Ours**
Back	46.29	**62.08**
Beak	43.08	**49.02**
Belly	54.44	**69.02**
Breast	54.19	**66.98**
Crown	64.69	**72.85**
Forehead	51.48	**58.46**
Left Eye	47.53	**55.78**
Left Leg	29.67	**40.94**
Left Wing	59.58	**71.57**
Nape	58.91	**70.78**
Right Eye	46.50	**55.51**
Right Leg	29.03	**40.52**
Right Wing	58.47	**71.56**
Tail	27.77	**40.16**
Throat	58.89	**70.83**
Average	48.70	**59.74**



mAP	200 species	14 species
Birdlets	-	40.25
Template bagging	-	44.73
Pose pooling	28.18	57.44
Ours	**44.13**	**62.42**