Part-based R-CNNs for Fine-grained Category Detection

Ning Zhang, Jeff Donahue, Ross Girshick, Trevor Darrell

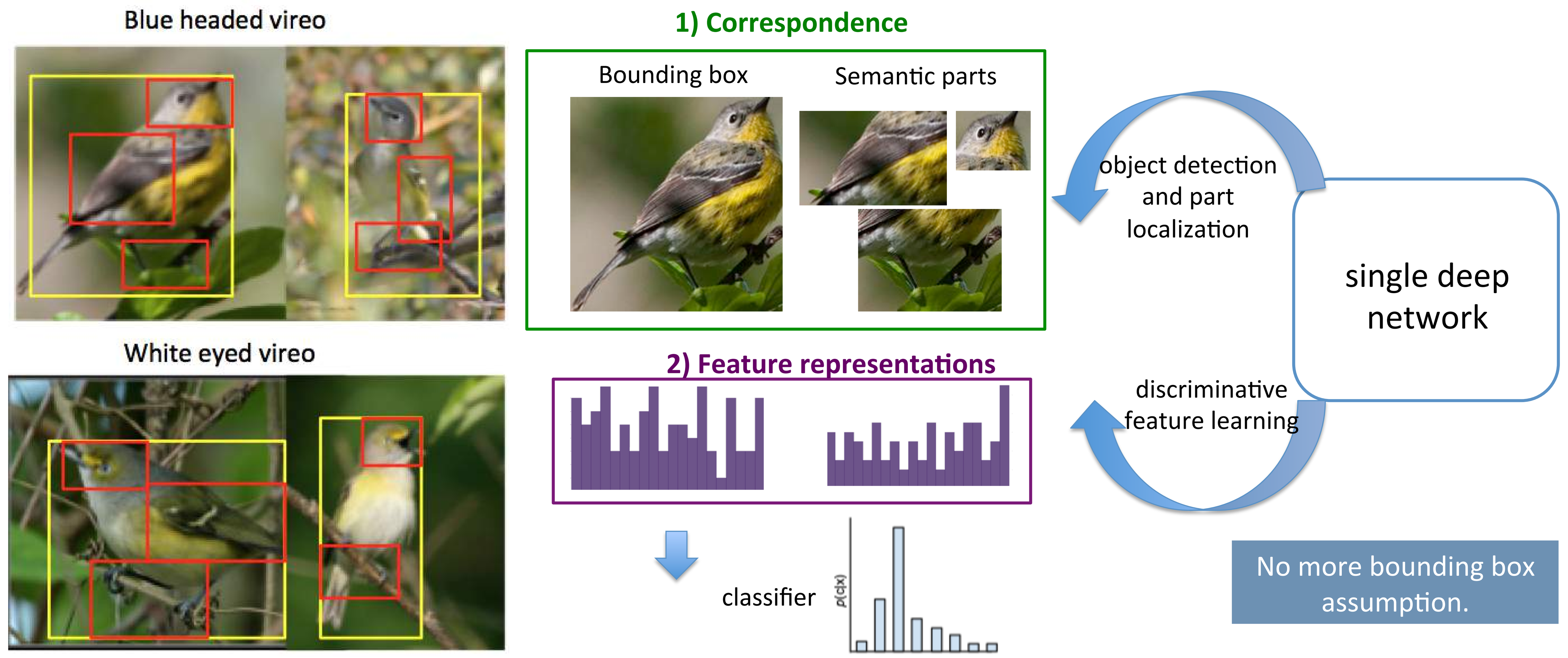
Fine-grained Categorization

- Challenging task due to pose variations and subtle differences assoicated with specific parts.

Figure: Black footed Albatross

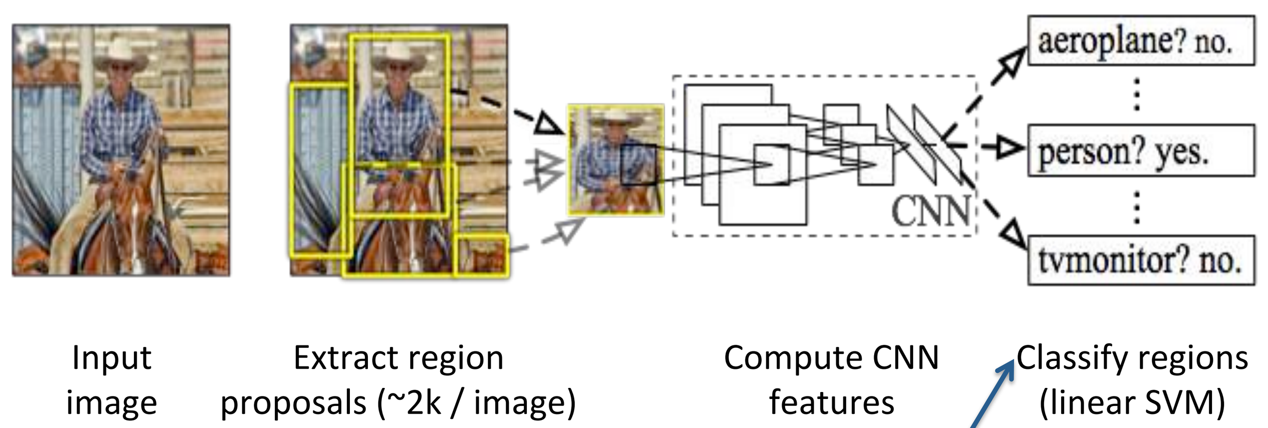
Figure: Laysan Albatross

Pose-normalized representation



Progress in deep learning

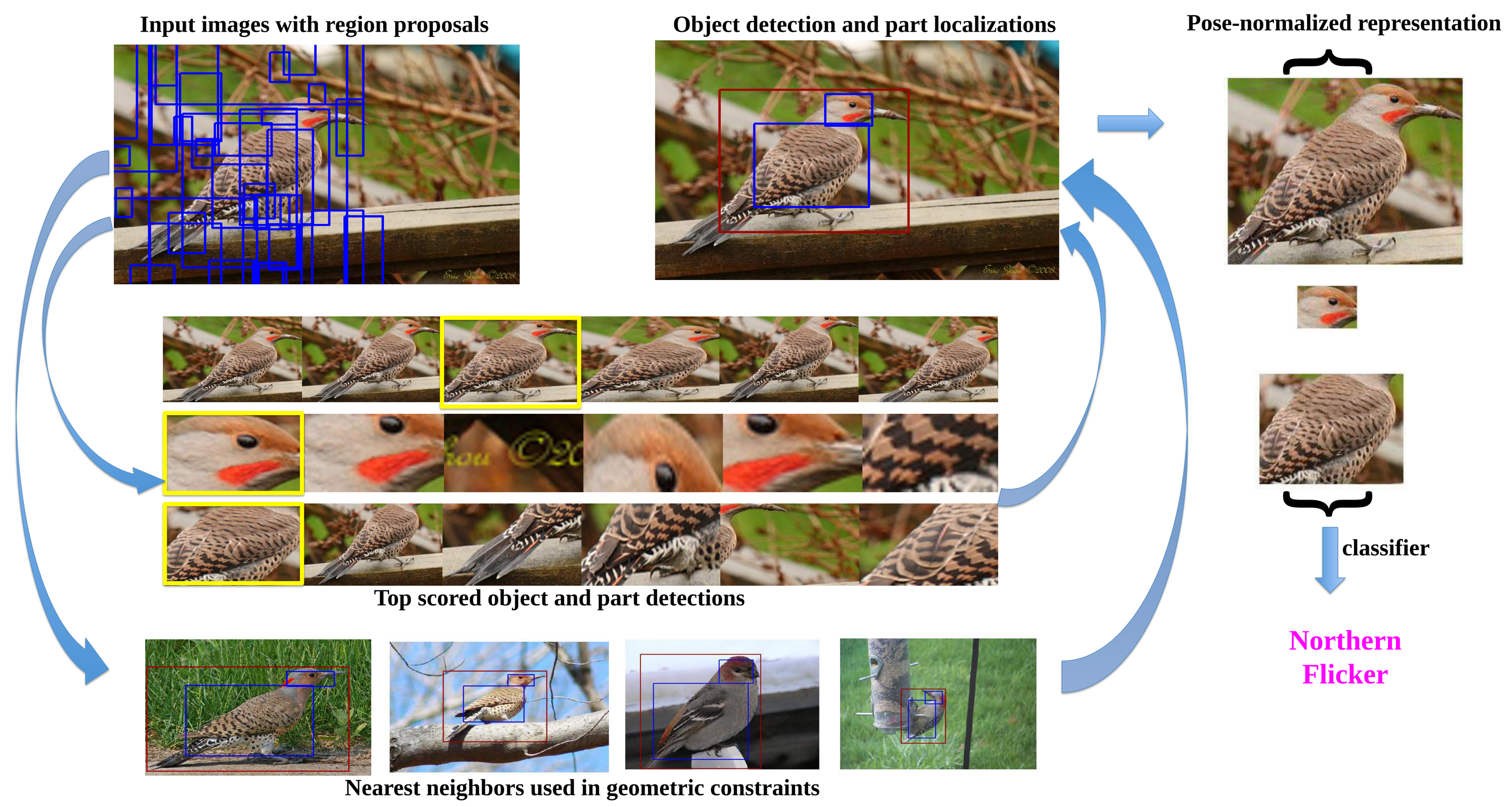
- Dramatic progress made in image classification and other computer vision applications
- R-CNN is the state of the art in object detection
- Can we simultaneously detect objects and find part correspondece?



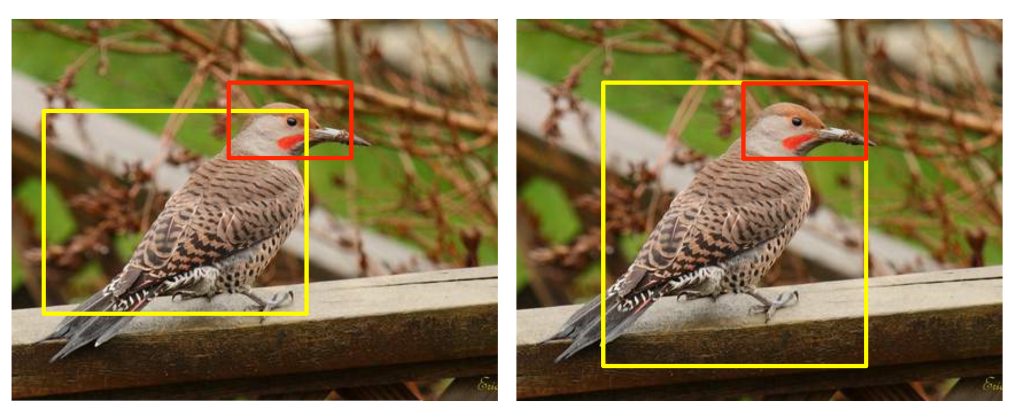
Try R-CNN https://github.com/rbgirshick/rcnn

Try CAFFE http://caffe.berkeleyvision.org

Overview of our approach

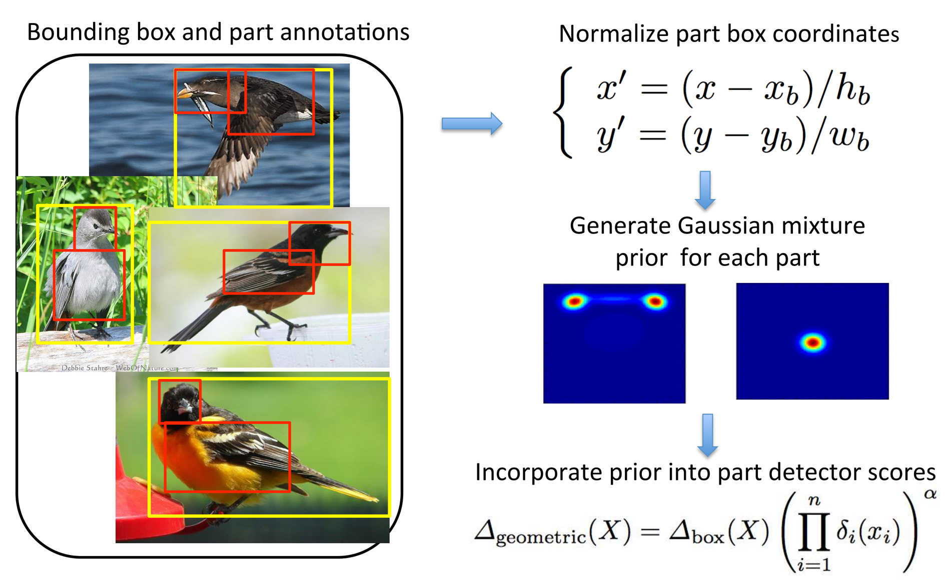


Box Constraint

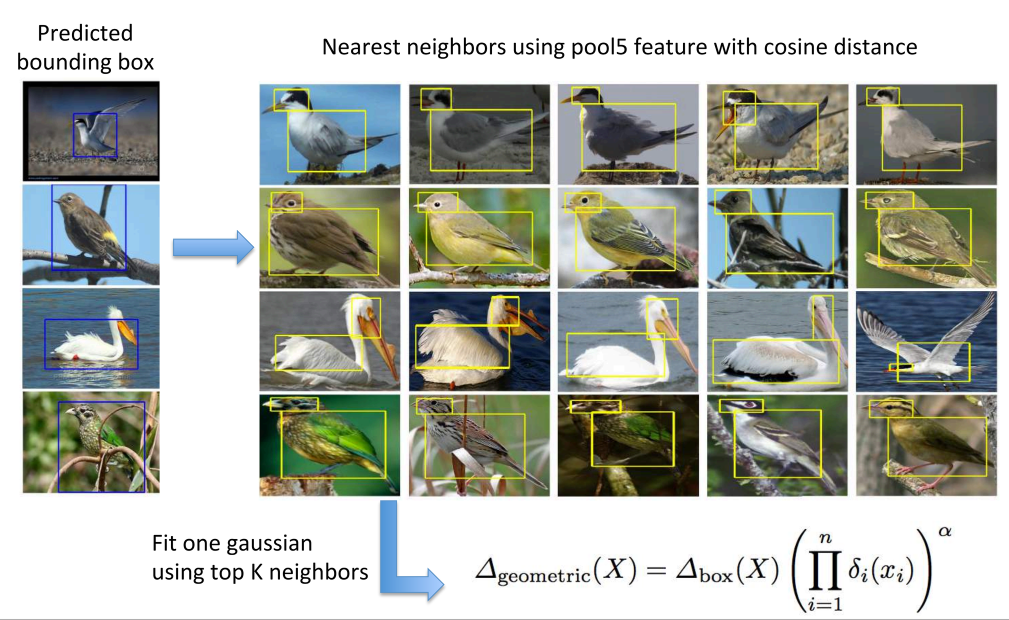


$$\Delta_{\text{box}}(X) = \prod_{i=1}^{n} c_{x_0}(x_i) \qquad c_x(y) = \begin{cases} 1 \text{ if region } y \text{ falls outside region } x \\ 0 \text{ otherwise} \end{cases}$$

Geometric constraint: Gaussian mixture



Geometric constraint: Non-parametric prior



Experimental Results

- Fine-grained categorization results on CUB200-2011 bird dataset

Bounding Box Given	
DPD [1]	50.98%
DPD+DeCAF feature [2]	64.96%
POOF [3]	56.78%
Symbiotic Segmentation [4]	59.40%
Alignment [5]	62.70%
Oracle	72.83%
Oracle-ft	82.02%
Ours (Δ_{box})	67.55%
Ours ($\Delta_{\text{geometric}}$ with δ^{MG})	67.98%
Ours ($\Delta_{\text{geometric}}$ with δ^{NP})	68.07%
Ours-ft (Δ_{box})	75.34%
Ours-ft ($\Delta_{\text{geometric}}$ with δ^{MG})	**76.37%**
Ours-ft ($\Delta_{\text{geometric}}$ with δ^{NP})	76.34%
Bounding Box Unknown	
DPD+DeCAF [2] with no bounding box	44.94%
Ours (Δ_{null})	64.57%
Ours (Δ_{box})	65.22%
Ours ($\Delta_{\text{geometric}}$ with δ^{MG})	65.98%
Ours ($\Delta_{\text{geometric}}$ with δ^{NP})	65.96%
Ours-ft (Δ_{box})	72.73%
Ours-ft ($\Delta_{\text{geometric}}$ with δ^{MG})	72.95%
Ours-ft ($\Delta_{\text{geometric}}$ with δ^{NP})	**73.89%**

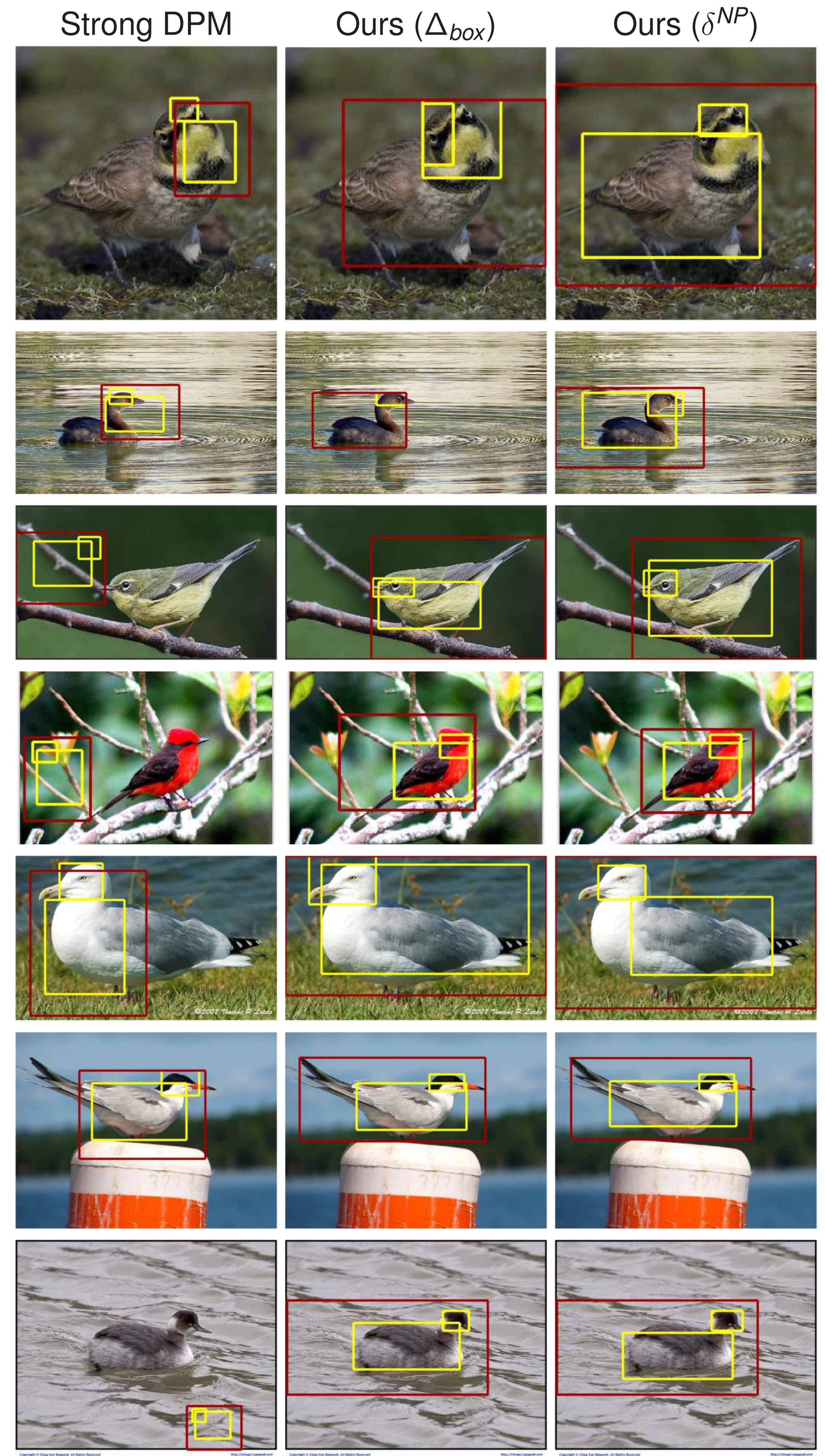
- Fine-grained categorization results on CUB200-2011 bird dataset with *no parts*

Oracle (ground truth bounding box)	57.94%
Oracle-ft	68.29%
Strong DPM [6]	38.02%
Ours (Δ_{box})	50.17%
Ours ($\Delta_{\text{geometric}}$ with δ^{MG})	51.83%
Ours ($\Delta_{\text{geometric}}$ with δ^{NP})	52.38%
Ours-ft (Δ_{box})	62.13%
Ours-ft ($\Delta_{\text{geometric}}$ with δ^{MG})	62.06%
Ours-ft ($\Delta_{\text{geometric}}$ with δ^{NP})	**62.75%**

- Part localization accuracy in terms of PCP (Percentage of Correctly Localized Parts)

Bounding Box Given	Head	Body
Strong DPM [6]	43.49%	75.15%
Ours (Δ_{box})	61.40%	65.42%
Ours ($\Delta_{\text{geometric}}$ with δ^{MG})	66.03%	76.62%
Ours ($\Delta_{\text{geometric}}$ with δ^{NP})	**68.19%**	**79.82%**
Bounding Box Unknown	**Head**	**Body**
Strong DPM [6]	37.44%	47.08%
Ours (Δ_{null})	60.50%	64.43%
Ours (Δ_{box})	60.56%	65.31%
Ours ($\Delta_{\text{geometric}}$ with δ^{MG})	**61.94%**	70.16%
Ours ($\Delta_{\text{geometric}}$ with δ^{NP})	61.42%	**70.68%**

Part Localization Samples



References

[1] Zhang et.al. Deformable Part Descriptors for Fine-grained Recognition and Attribute Modeling. In ICCV 2013.
[2] Donahue et.al. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. In ICML 2014.
[3] Berg et.al. Part-based one-vs-one features for fine-grained categorization, face verification, and attribute estimation. In CVPR 2013.
[4] Chai et.al. Symbiotic segmentation and part localization for fine-grained categorization. In ICCV 2013.
[5] Gavves et.al. Fine-grained categorization by alignments. In ICCV 2013.
[6] Azizpour et.al. Object detection using strongly supervised deformable part models. In ECCV 2012.

EECS, University of California, Berkeley

Mail: {nzhang, jdonahue, rbg, trevor}@eecs.berkeley.edu