

## Fine-grained Categorization

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

Figure: Black footed Albatross



Figure: Laysan Albatross



## **Pose-normalized representation**

### Blue headed vireo



1) Correspondence Bounding box Semantic parts object detection White eyed vireo 2) Feature representations 



## **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?



![](_page_0_Picture_17.jpeg)

Input image

Extract region proposals (~2k / image)

![](_page_0_Figure_20.jpeg)

Classify regions (linear SVM)

Jse part annotations as regions

Try R-CNN <a href="https://github.com/rbgirshick/rcnn">https://github.com/rbgirshick/rcnn</a> Try CAFFE <a href="http://caffe.berkeleyvision.org">http://caffe.berkeleyvision.org</a>

## **Box Constraint**

![](_page_0_Picture_25.jpeg)

![](_page_0_Picture_26.jpeg)

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

EECS, University of California, Berkeley

## Geometric constraint: Gaussian mixture

![](_page_0_Picture_30.jpeg)

**Overview of our approach** 

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

![](_page_0_Picture_36.jpeg)

classifier

Normalize part box coordinates  $x' = (x - x_b)/h_b$  $y' = (y - y_b)/w_b$ Generate Gaussian mixture prior for each part Incorporate prior into part detector scores  $\Delta_{\text{geometric}}(X) = \Delta_{\text{box}}(X) \left(\prod_{i=1}^{n} \delta_{i}(x_{i})\right)$ 

## Geometric constraint: Non-para

![](_page_0_Picture_39.jpeg)

![](_page_0_Picture_40.jpeg)

![](_page_0_Picture_41.jpeg)

Fit one gaussian using top K neighbors  $\Delta_{\text{geometric}}(X) = \Delta_{\text{box}}(X) \left(\prod_{i=1} \delta_i(x_i)\right)$ 

![](_page_0_Figure_43.jpeg)

Ning Zhang, Jeff Donahue, Ross Girshick, Trevor Darrell

	Experimental Results	
	Fine-grained categorization results on CUB200-2011 bird dataset	
	Bounding Box Given	
object detection	DPD [1] DPD+DeCAF feature [2]	50.98% 64.96%
localization	POOF [3] Symbiotic Segmentation [4]	56.78% 59.40%
single deep	Alignment [5]	62.70%
network	Oracle Oracle-ft	72.83%
discriminative	Ours $(\Delta_{\text{box}})$	67.55%
feature learning	Ours ( $\Delta_{\text{geometric}}$ with $\delta^{MG}$ )	67.98%
	Ours ( $\Delta_{\text{geometric}}$ With $\delta'''$ ) Ours-ft ( $\Lambda_1$ )	68.07% 75.34%
	Ours-ft ( $\Delta_{\text{geometric}}$ with $\delta^{MG}$ )	76.37%
No more bounding box	Ours-ft ( $\Delta_{\text{geometric}}$ with $\delta^{NP}$ )	76.34%
assumption	Bounding Box Unknown	
	DPD+DeCAF [2] with no bounding bo	)X 44.94%
	Ours $(\Delta_{null})$ Ours $(\Lambda_{hor})$	65 22%
	Ours ( $\Delta_{\text{geometric}}$ with $\delta^{MG}$ )	65.98%
	Ours ( $\Delta_{\text{geometric}}$ with $\delta^{NP}$ )	65.96%
ocalizations Pose-normalized representation	Ours-ft ( $\Delta_{box}$ )	72.73%
	Ours-ft ( $\Delta_{\text{geometric}}$ with $\delta^{NO}$ )	72.95%
	► Fine-grained categorization results of CUB200-2011 bird dataset with no p	on Sarts
	Oracle (ground truth bounding box)	57.94%
	Oracle-II Strong DPM [6]	68.29% 38.02%
	Ours $(\Delta_{hov})$	50.02 %
	Ours ( $\Delta_{\text{geometric}}$ with $\delta^{MG}$ )	51.83%
classifier	Ours ( $\Delta_{\text{geometric}}$ with $\delta^{NP}$ )	52.38%
	Ours-ft ( $\Delta_{box}$ )	62.13%
Northern	Ours-ft ( $\Delta_{\text{geometric}}$ with $\delta^{MC}$ )	62.06%
	<ul> <li>Part localization accuracy in terms c (Percentage of Correctly Localized F</li> </ul>	of PCP Parts)
straint: Non-parametric prior	Bounding Box Given Head	Body
Nearest neighbors using pool5 feature with cosine distance	Strong DPM [6] 43.49%	75.15%
	Ours $(\Delta_{box})$ 61.40%	65.42%
	Ours $(\Delta_{\text{geometric}} \text{ With } \delta^{MO})$ 66.03% Ours $(\Lambda  \text{with } \delta^{NP})$ 69.03%	/ b.b2% 79 82%
	Bounding Box Unknown	
	Head	Body
	Strong DPM [6] 37.44%	47.08%
	Ours $(\Delta_{\text{null}})$ 60.50%	64.43%
	Ours $(\Delta_{box})$ (0.56%) Ours $(\Lambda \times with SMG)$ <b>61 0.7%</b>	00.31% 70 16%
	Ours ( $\Delta_{\text{geometric}}$ with $\delta^{NP}$ ) 61.42%	70.68%
Theorem $A = (X) - A = (X) \left( \prod_{i=1}^{n} \delta_i(x_i) \right)^{\alpha}$		

![](_page_0_Picture_47.jpeg)

## Part Localization Samples

Strong DPM

Ours  $(\Delta_{box})$ 

Ours ( $\delta^{NP}$ )

![](_page_0_Picture_52.jpeg)

## 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.