

# Computer Vision: Understanding, Interpreting and Learning from Visual Data



 Lawrence Livermore  
National Laboratory

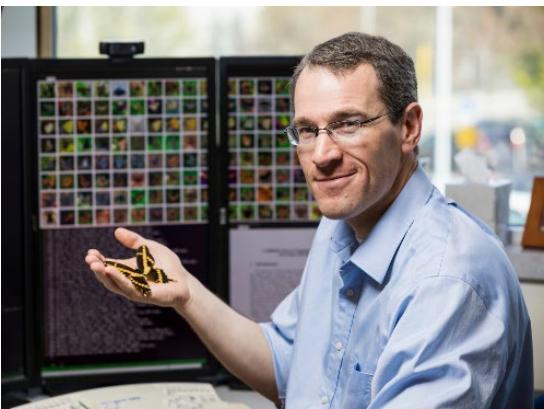
## DAY THREE

July 25, 2019

July 30, 2019

July 31, 2019

August 2, 2019



Ryan Farrell

**BYU**  
BRIGHAM YOUNG  
UNIVERSITY

# UPDATES

- Tomorrow – **Meet the Director** at 10AM
- Friday – **Our Final Class** at 10AM
- My email: [farrell@cs.byu.edu](mailto:farrell@cs.byu.edu)

# What is Fine-grained Recognition?

# What is Recognition??



DETECTION (WHERE)

# What is Recognition??



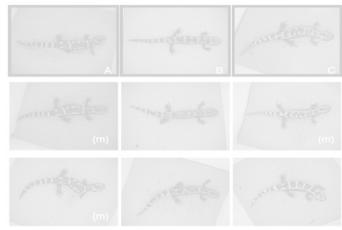
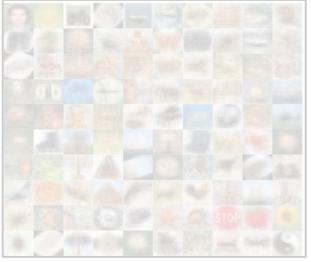
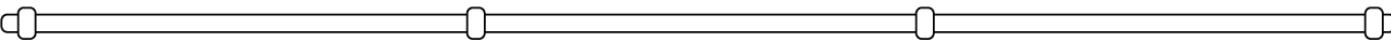
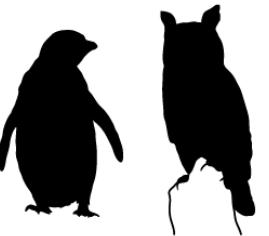
POSE ESTIMATION

# What is Recognition??



CLASSIFICATION (WHO/WHAT?)

# The Categorization Spectrum



Caltech/UCSD Birds 200 (2011)  
Welinder, Branson, Mita, Wah, Schroff, Belongie, Perona

# Fine-Grained Recognition



Photographs by Frode Jacobsen (CCUB NABirds)

# Fine-Grained Recognition

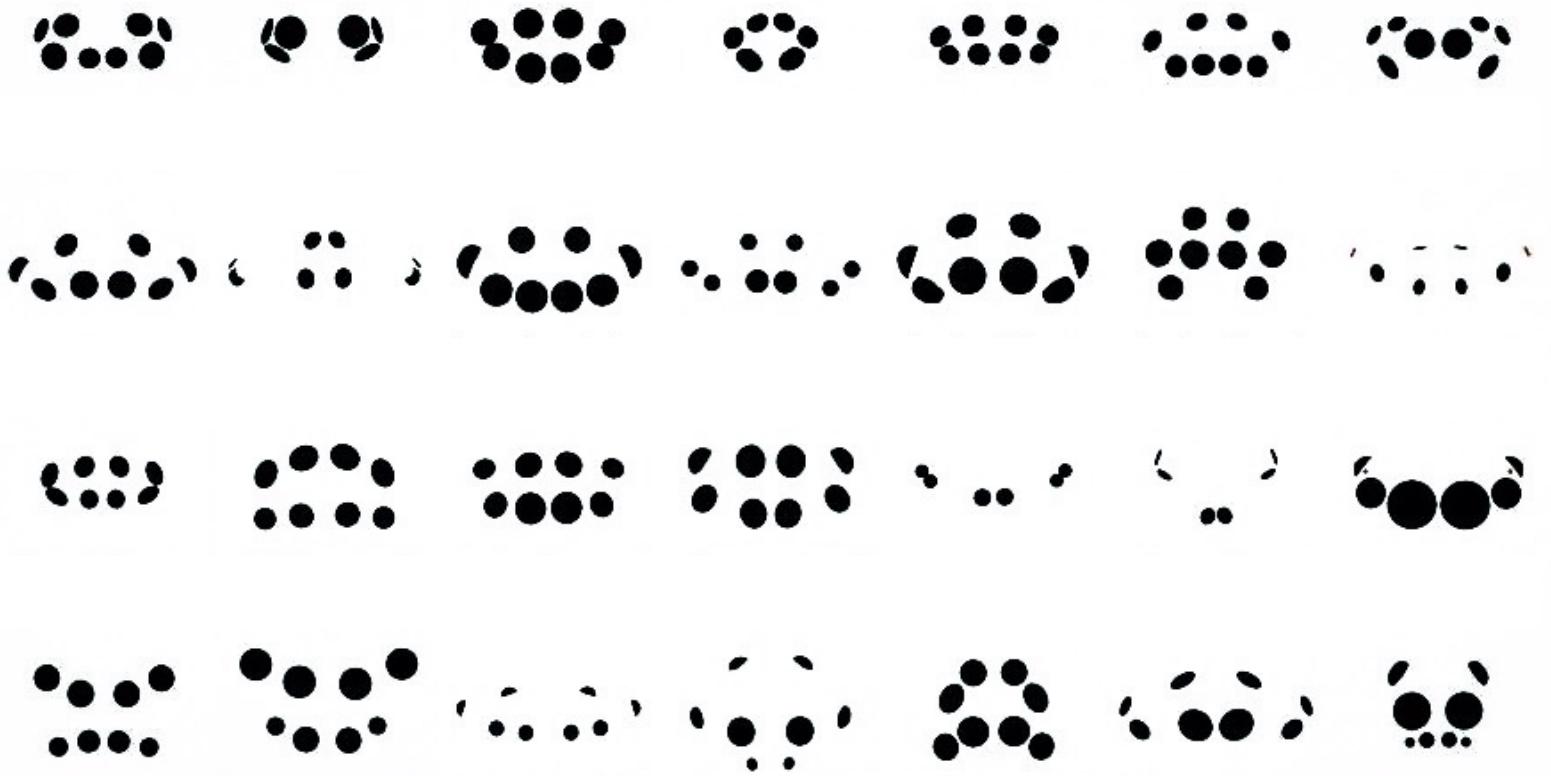


Photographs by Frode Jacobsen (CCUB NABirds)

# North American Warblers



# Nature's Domains



# Nature's Domains



*Eurycantha calcarata*



*Phyllium giganteum*



*Phyllium philippinicum*



*Phyllium westwoodii*



*Ramulus nematodes*



*Medauroides extradentata*



*Hermagoras hosei*



*Rhaphiderus spiniger*



*Phobaeticus serratipes*



*Diapherodes gigantea*



*Neohirasea maerens*



*Parapachymorpha zomproi*



*Paramenexenus laetus*



*Spinohirasea bengalensis*



*Heteropteryx dilatata*



*Haaniella dehaanii*



*Brasidas samarensis*



*Aretaon asperrimus*



*Trachyareaon carmeli*



*Mearnsiana bullosa*

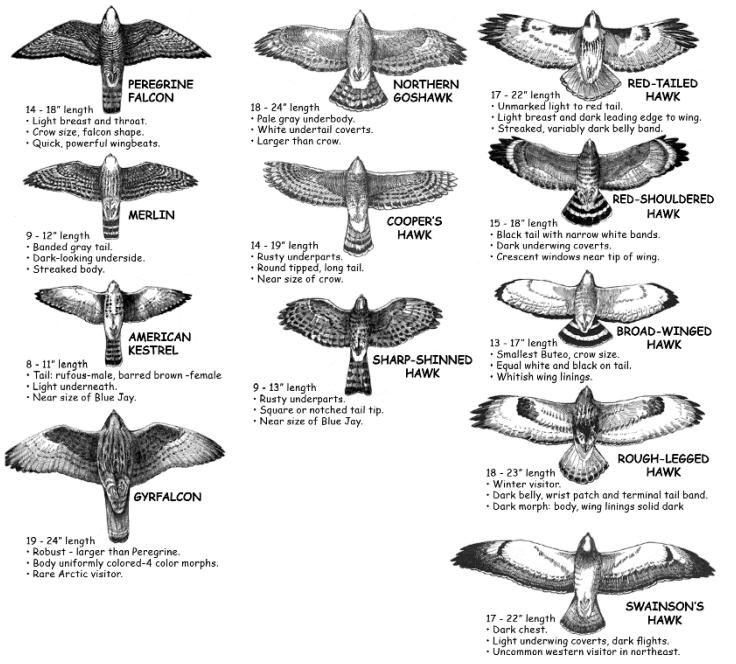


*Sungaya inexpectata*

# Nature's Domains



# Nature's Domains



**SIBLEY'S RAPTORS OF NORTH AMERICA**



# Manmade Domains

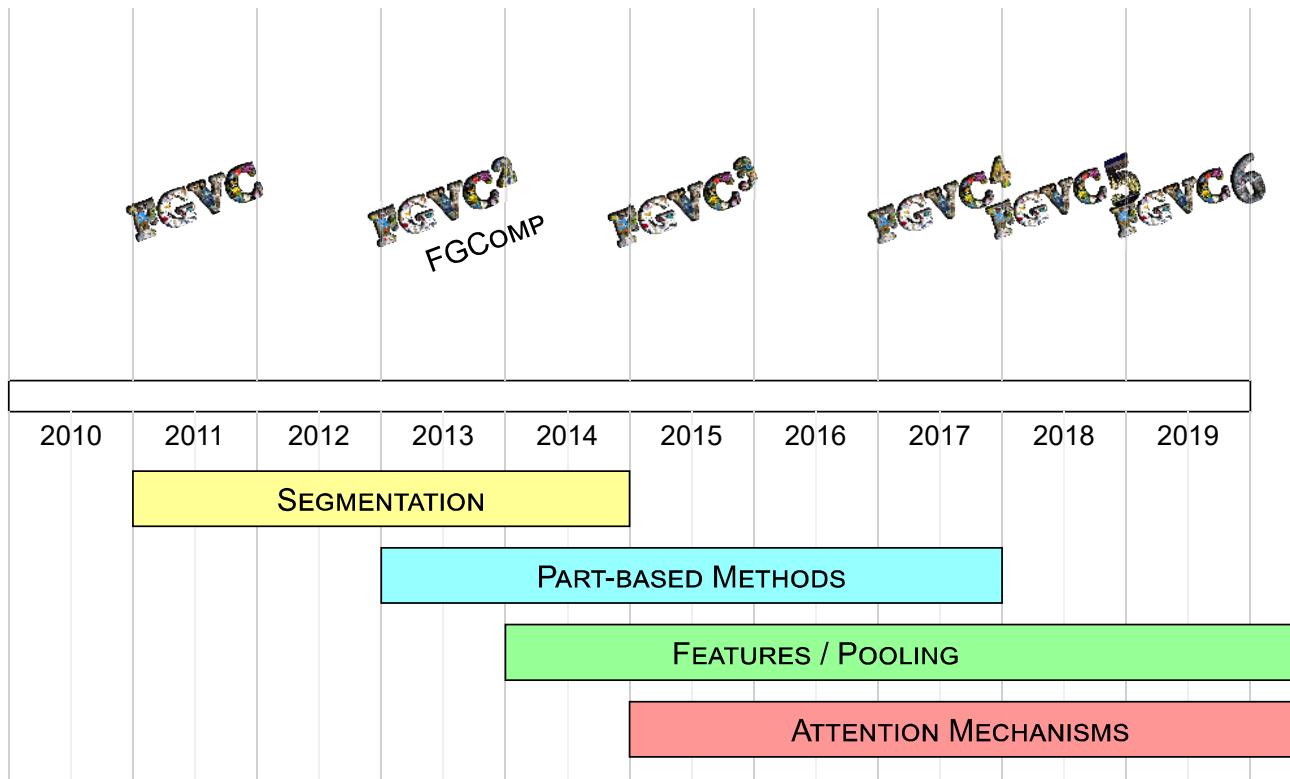
LACE UP	WHOLE CUT	PLAIN TOE	CAP TOE	WING TIP
THE OXFORD'S (AKA BALMORALS)				
THE DERBY'S				
SLIP ON	PENNY	BIT	TASSLE	KILTIE
THE LOAFER'S				
FORMAL	BLACK OXFORD (POLISHED CALFSKIN)	BLACK OXFORD (PATENT LEATHER)	OPERA PUMP (PATENT LEATHER)	RIBBON PUMP (PATENT LEATHER)
BLACK TIE				
BOOT	CHELSEA	CHUKKA	CAP TOE	WINGTIP
DRESS BOOTS				
STRAP	SINGLE	DOUBLE	TRIPLE	
MONK SHOES				
PERFORATION	QUARTER	SEMI	FULL	LONGWING
BROGUEING				



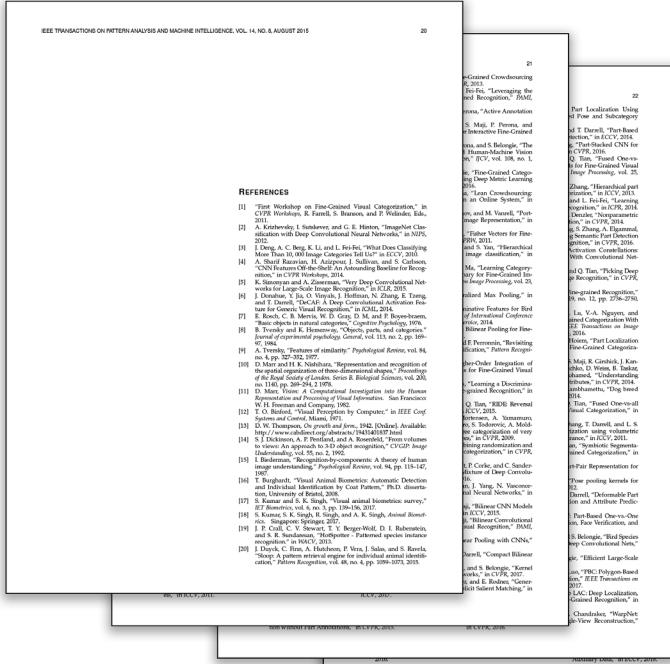
# Manmade Domains



# Timeline



# Many Papers/Approaches



Zhao, et al.  
IJASS 2017



Zheng, et al.  
IJASS 2018

# Recognition with Humans in the Loop

## Visual 20 Questions

(A) Easy for Humans



Chair? Airplane? ...

(B) Hard for Humans



Finch? Bunting? ...

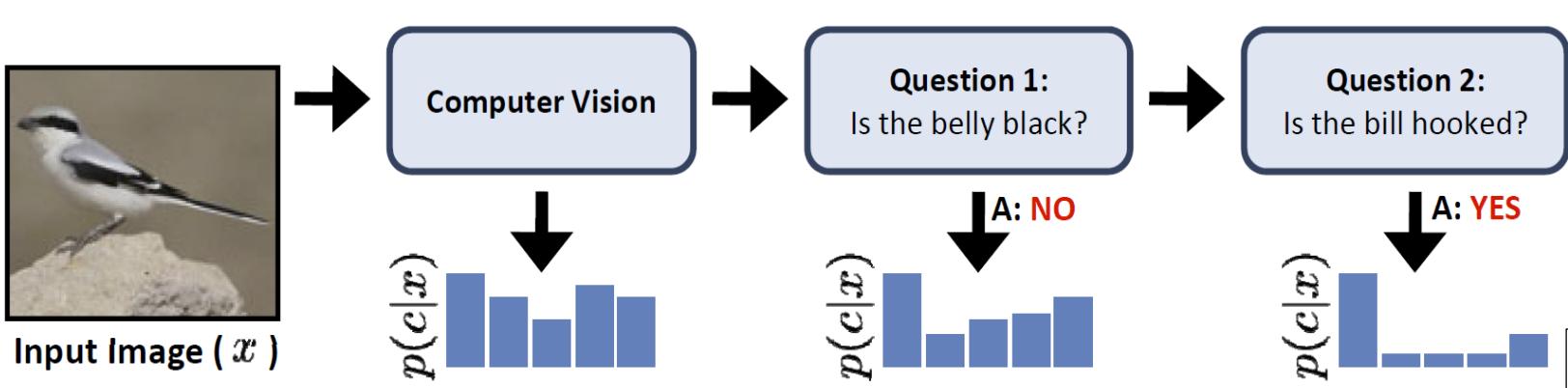
(C) Easy for Humans



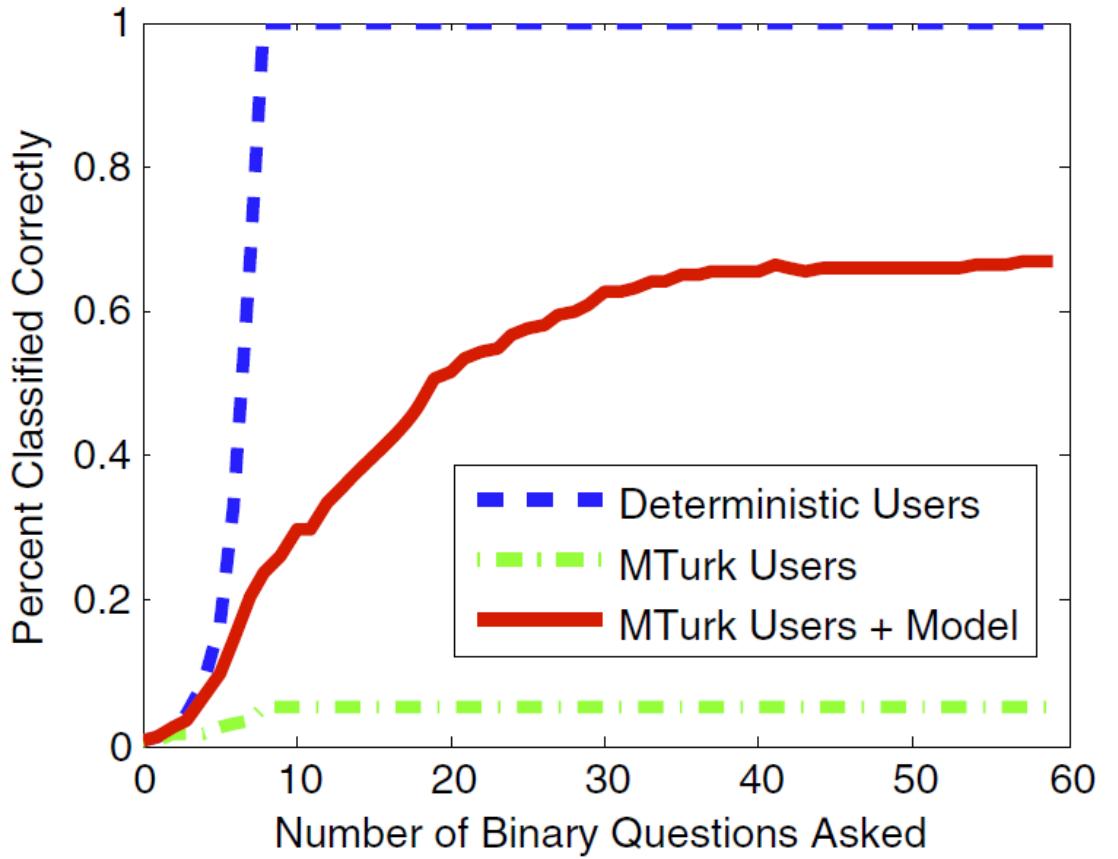
Yellow Belly? Blue Belly? ...

# Recognition with Humans in the Loop

## Visual 20 Questions



# Recognition with Humans in the Loop



Visual Recognition with Humans in the Loop

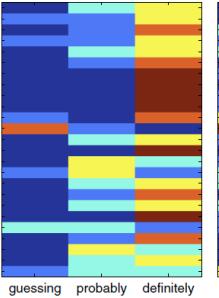
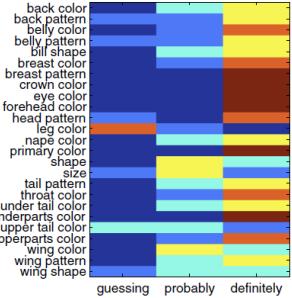
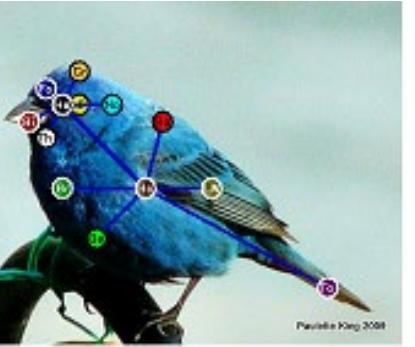
Steve Branson, Catherine Wah, Florian Schroff, Boris Babenko, Peter Welinder, Pietro Perona, Serge Belongie

# CUB-200 Dataset



Ivory Gull

Chuck-will's-widow

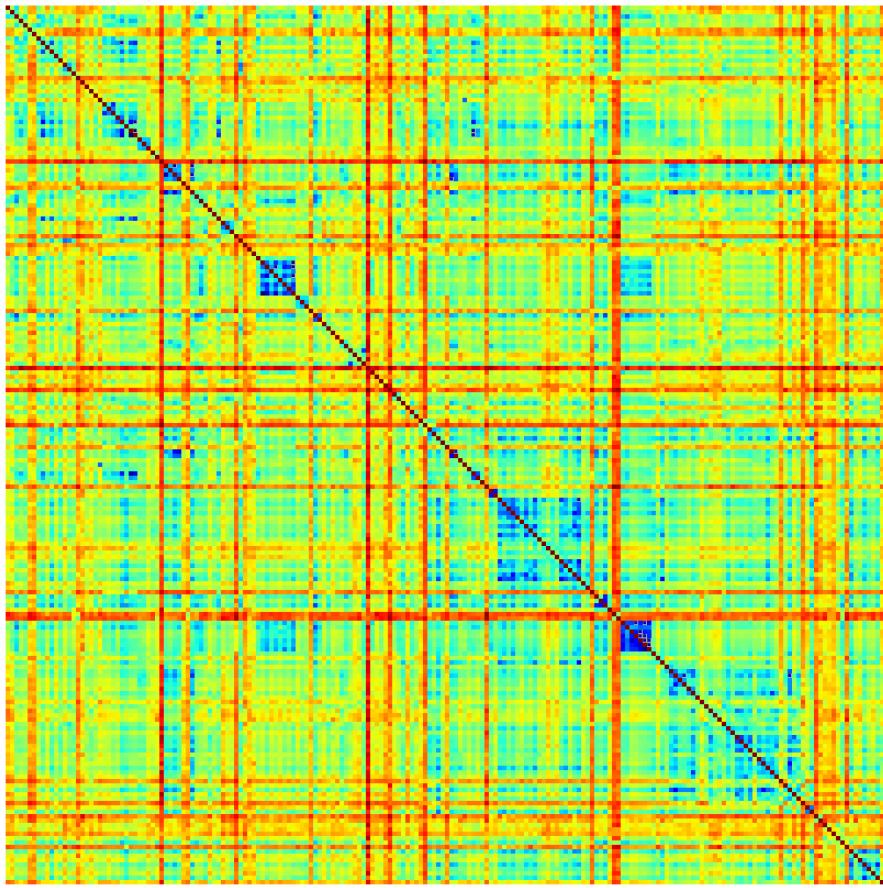


Visual Recognition with Humans in the Loop

Steve Branson, Catherine Wah, Florian Schroff, Boris Babenko, Peter Welinder, Pietro Perona, Serge Belongie

ECCV 2010

# CUB-200 Dataset



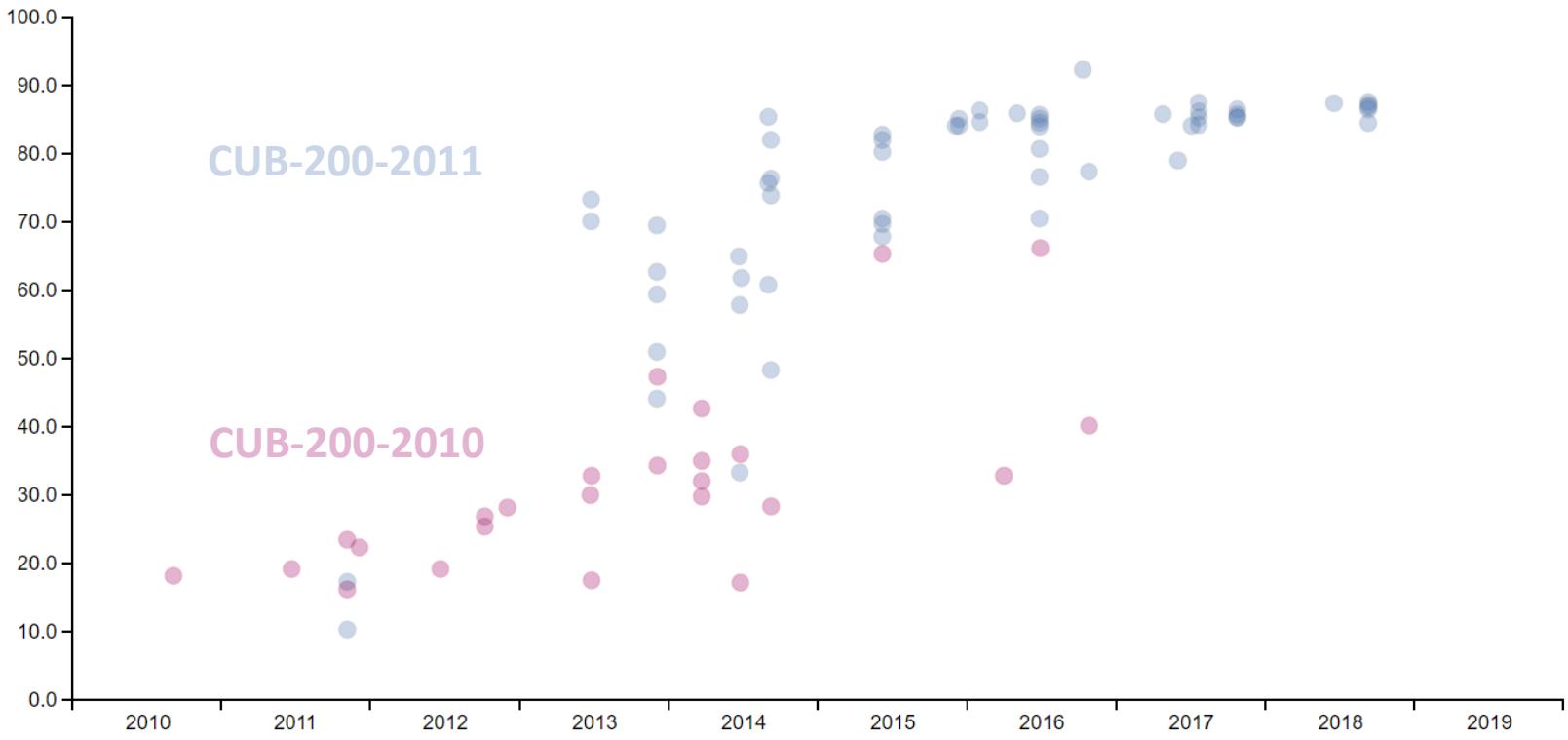
Visual Recognition with Humans in the Loop

Steve Branson, Catherine Wah, Florian Schroff, Boris Babenko, Peter Welinder, Pietro Perona, Serge Belongie



ECCV 2010

# CUB-200 Dataset

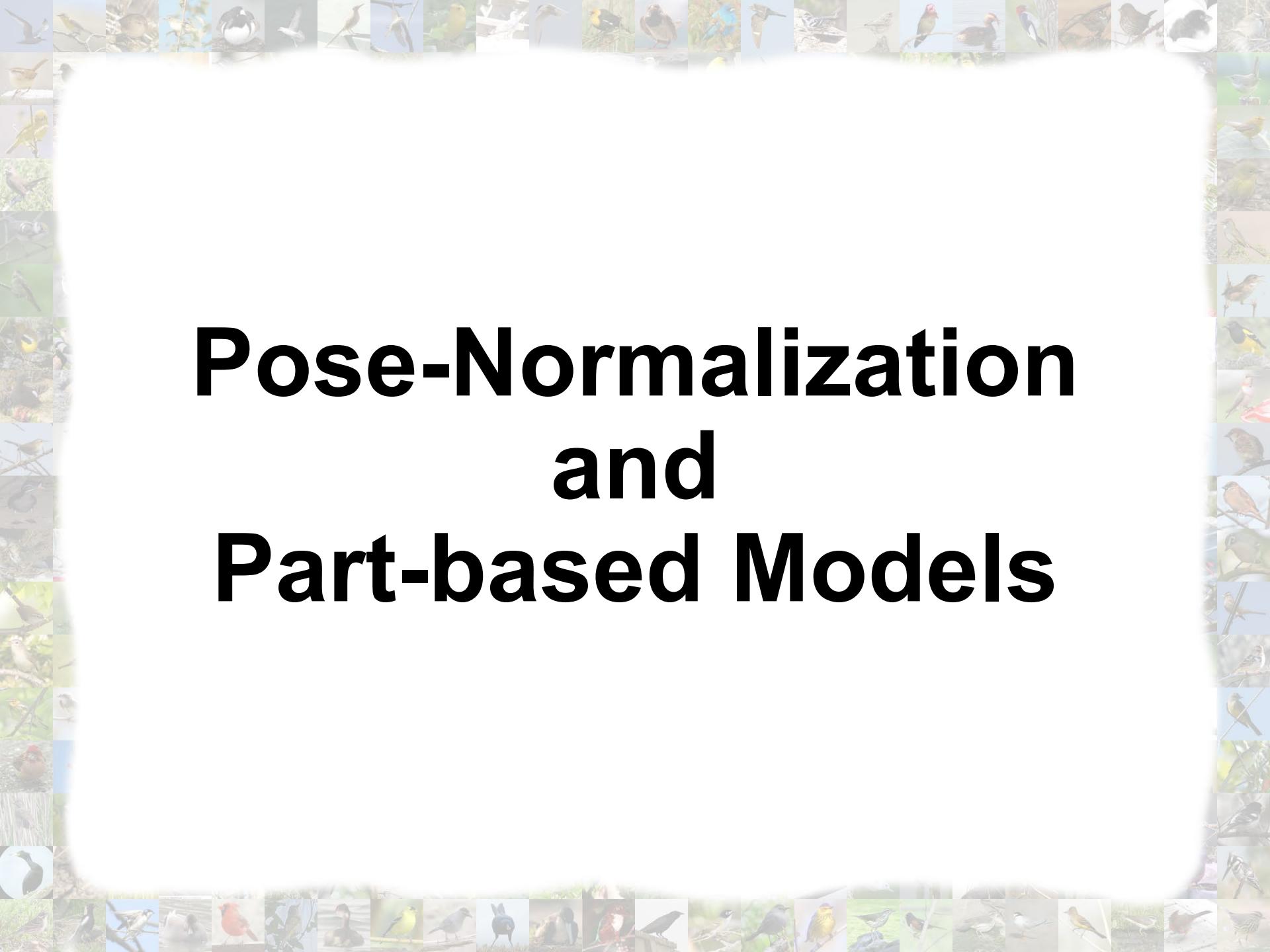


Visual Recognition with Humans in the Loop

Steve Branson, Catherine Wah, Florian Schroff, Boris Babenko, Peter Welinder, Pietro Perona, Serge Belongie



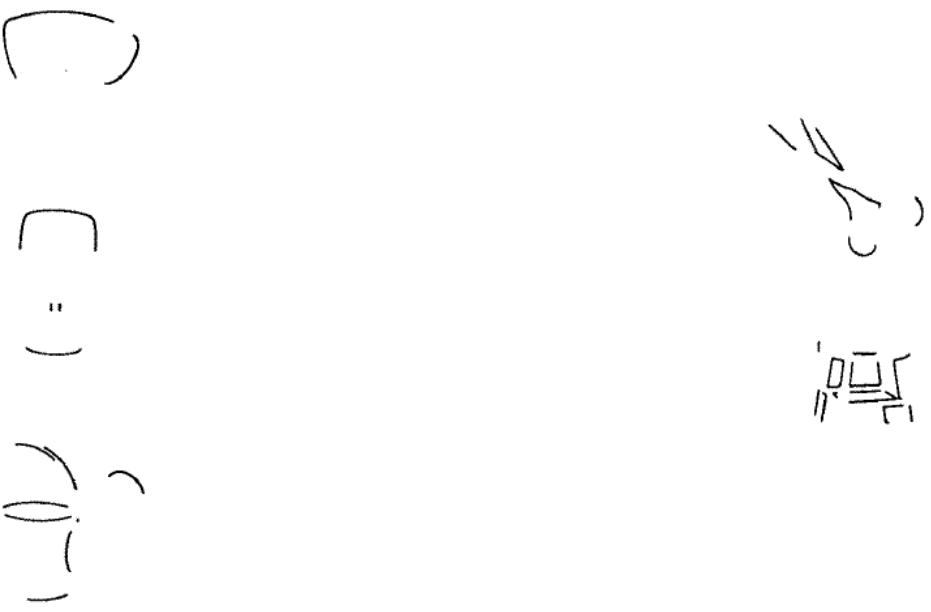
ECCV 2010



# Pose-Normalization and Part-based Models

# Recognition by Components

## Irving Biederman, 1987



Birdlets: Subordinate Categorization using Volumetric Primitives and Pose-normalized Appearance  
Ryan Farrell, Om Oza, Ning Zhang, Vlad I. Morariu, Trevor Darrell, Larry S. Davis



ICCV 2011

# Volumetric Representation



(Generalized Cylinders - Binford '71, Superquadrics - Pentland '86 , Geons - Biederman '87)

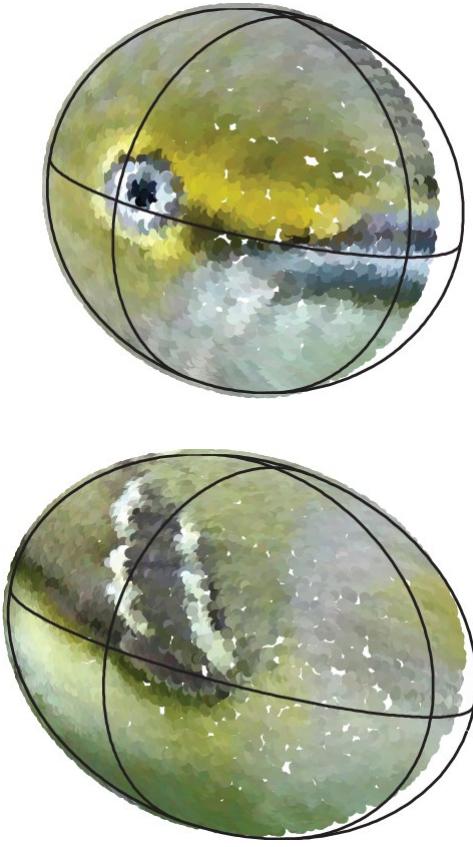
Birdlets: Subordinate Categorization using Volumetric Primitives and Pose-normalized Appearance

Ryan Farrell, Om Oza, Ning Zhang, Vlad I. Morariu, Trevor Darrell, Larry S. Davis



ICCV 2011

# Pose-Normalized Appearance



Birdlets: Subordinate Categorization using Volumetric Primitives and Pose-normalized Appearance

Ryan Farrell, Om Oza, Ning Zhang, Vlad I. Morariu, Trevor Darrell, Larry S. Davis



ICCV 2011

# Pose-specific Templates (Birdlets)



Inspired by Poselets – Bourdev and Malik, in *ICCV* 2009.

Birdlets: Subordinate Categorization using Volumetric Primitives and Pose-normalized Appearance

Ryan Farrell, Om Oza, Ning Zhang, Vlad I. Morariu, Trevor Darrell, Larry S. Davis



ICCV 2011

# Pose-specific Templates - Examples



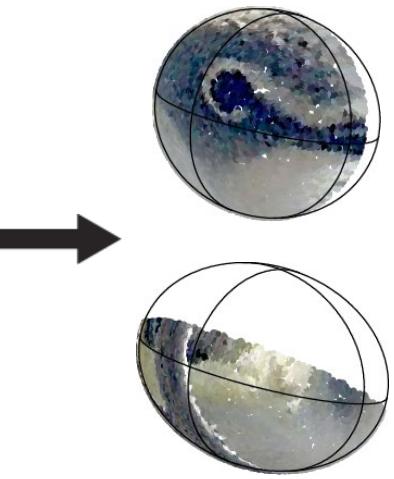
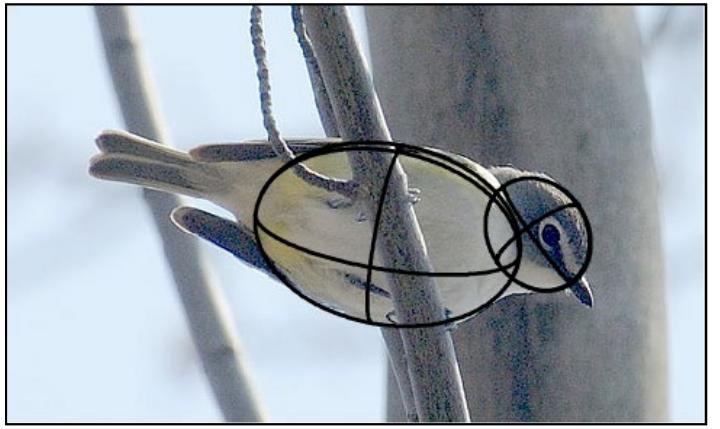
Birdlets: Subordinate Categorization using Volumetric Primitives and Pose-normalized Appearance

Ryan Farrell, Om Oza, Ning Zhang, Vlad I. Morariu, Trevor Darrell, Larry S. Davis

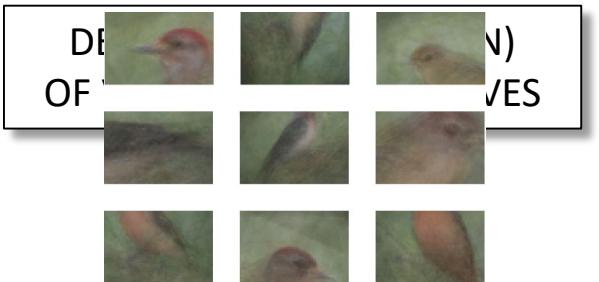
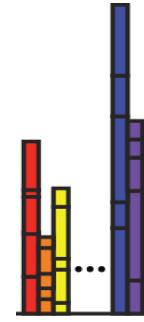


ICCV 2011

# Pose-normalized Representations



**“Blue-Headed  
Vireo”**



POSE-NORMALIZED  
APPEARANCE SPACE

CLASSIFICATION

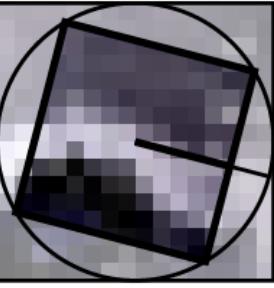
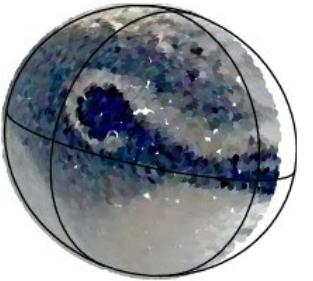
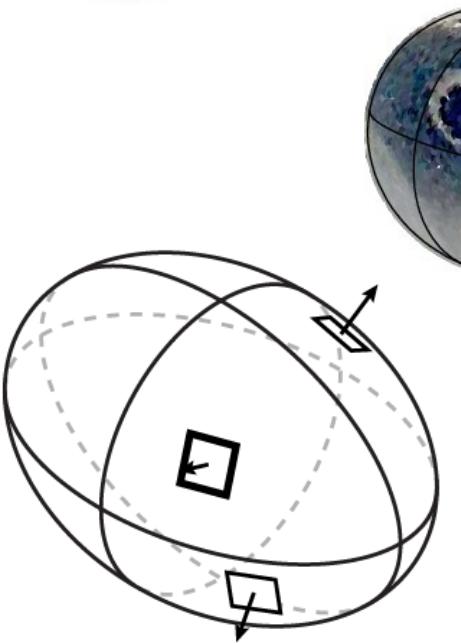
Birdlets: Subordinate Categorization using Volumetric Primitives and Pose-normalized Appearance

Ryan Farrell, Om Oza, Ning Zhang, Vlad I. Morariu, Trevor Darrell, Larry S. Davis

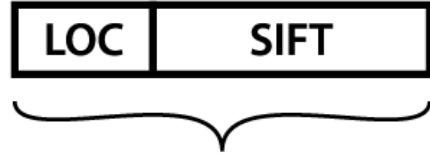


ICCV 2011

# Pose-Normalized Appearance Descriptors



Extracted Tangent Patch



Pose-normalized Appearance  
Descriptor (PNAD)

Birdlets: Subordinate Categorization using Volumetric Primitives and Pose-normalized Appearance

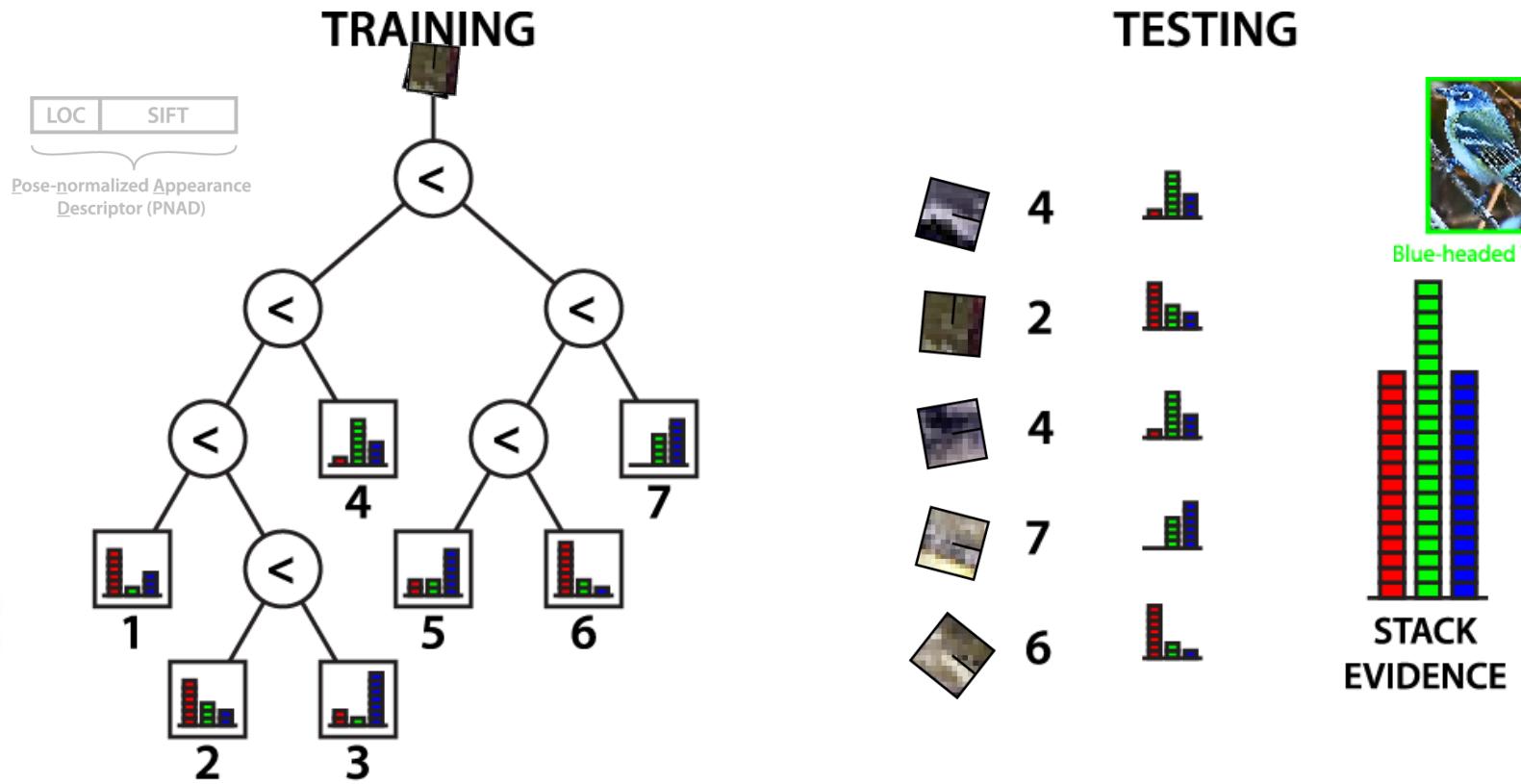
Ryan Farrell, Om Oza, Ning Zhang, Vlad I. Morariu, Trevor Darrell, Larry S. Davis



ICCV 2011

# Categorization

## Stacked-Evidence Random Forest



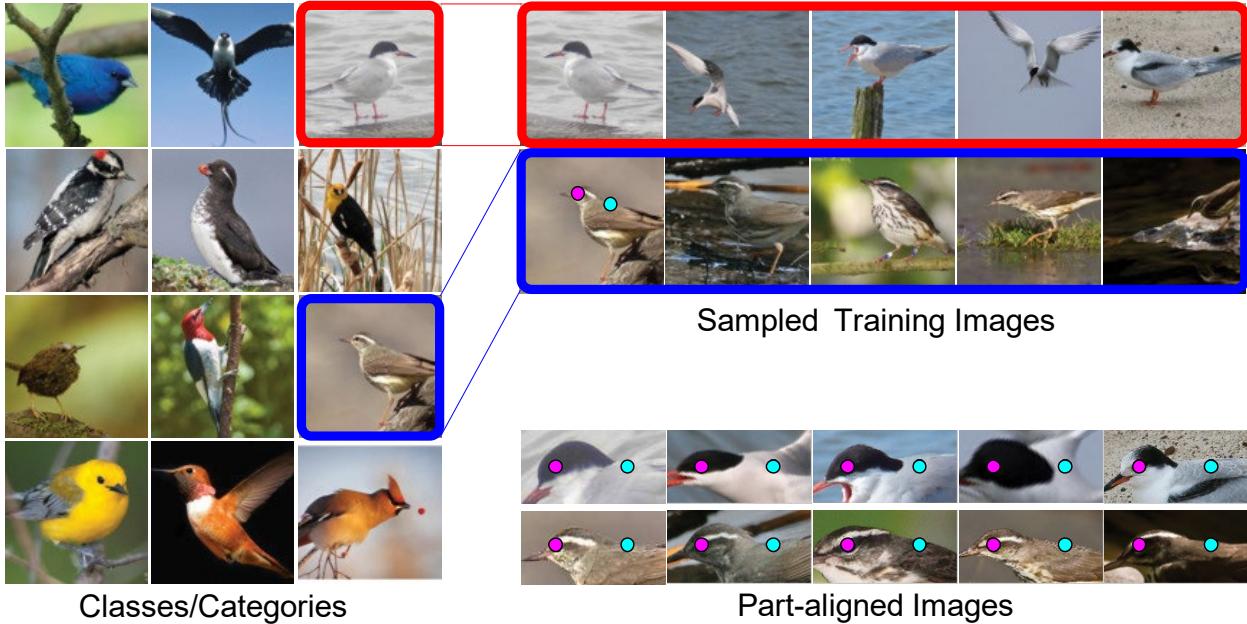
Birdlets: Subordinate Categorization using Volumetric Primitives and Pose-normalized Appearance  
Ryan Farrell, Om Oza, Ning Zhang, Vlad I. Morariu, Trevor Darrell, Larry S. Davis



ICCV 2011

# Pose-Estimation

# POOF



(e.g. with CUB-200, can build millions of POOFs, but thousands suffice)

$$\binom{200}{2} \text{ class pairs} \cdot (12 \cdot 11) \text{ part pairs} \cdot 2 \text{ base features} = 5,253,600 \text{ POOFs}$$

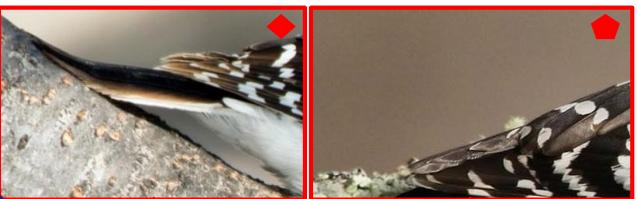


# Motivation

Caspian Tern



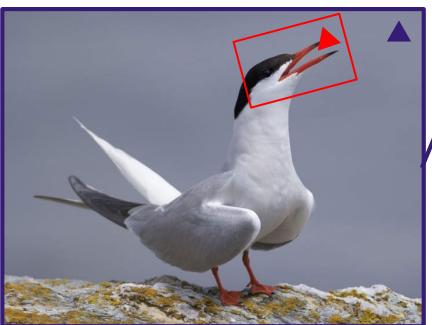
Pose Aligned Regions



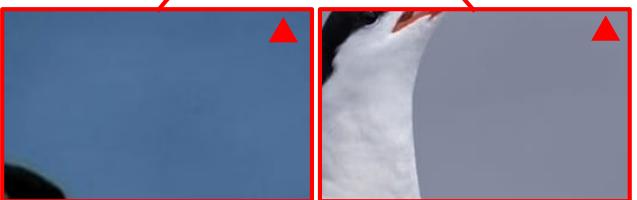
Hairy Woodpecker



Caspian Tern



Feature Embedding Space



Downy Woodpecker

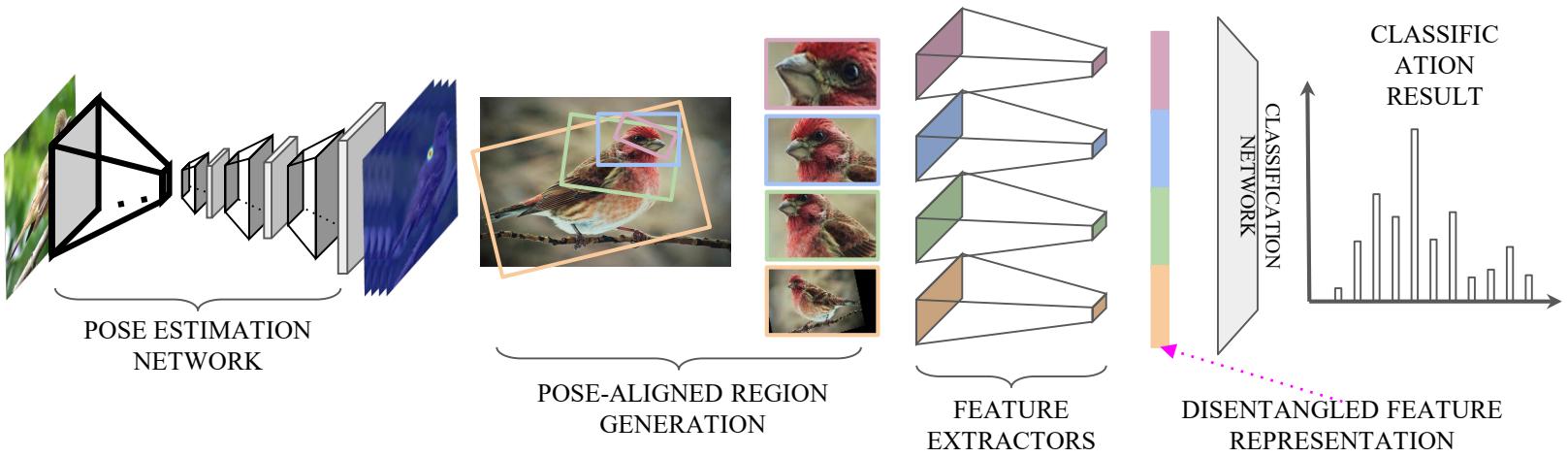


Aligned to the Object, not to the Image: A Unified Pose-aligned Representation for Fine-grained Recognition

Pei Guo, Ryan Farrell

WACV 2019

# Pipeline



# Pose Aligned Patches



Beak | Left-eye

Breast | Belly

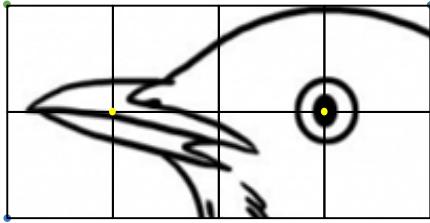
Left-leg | Tail



Beak | Left-eye

Breast | Belly

Left-leg | Tail



$\mathbf{p}_i, \mathbf{p}_j$

•  $(\mathbf{p}_i - h\hat{\mathbf{r}}_{ij}) + h\hat{\mathbf{t}}_{ij}$

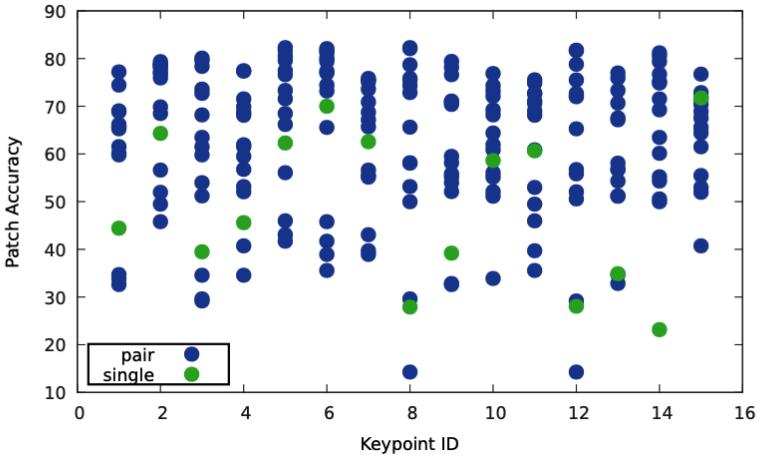
•  $(\mathbf{p}_j + h\hat{\mathbf{r}}_{ij}) - h\hat{\mathbf{t}}_{ij}$

•  $(\mathbf{p}_i - h\hat{\mathbf{r}}_{ij}) - h\hat{\mathbf{t}}_{ij}$

# Keypoint Detection Results

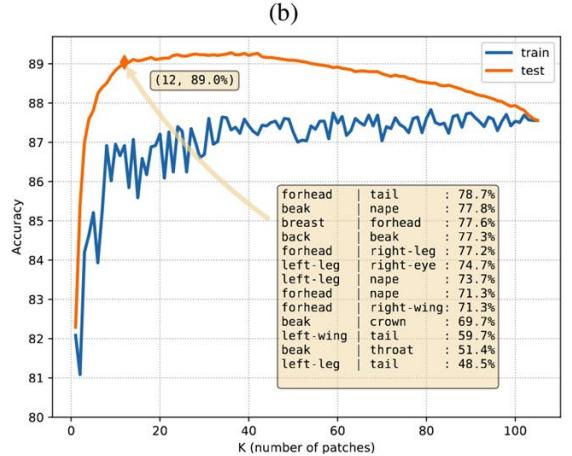
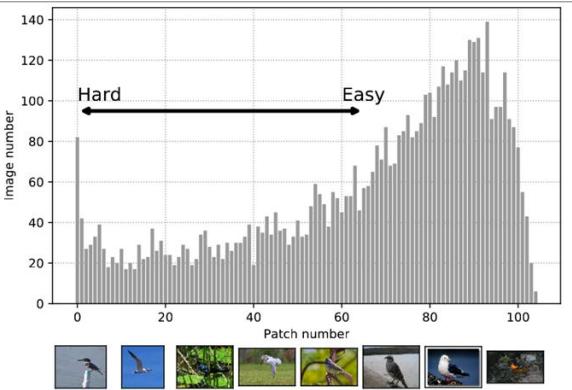
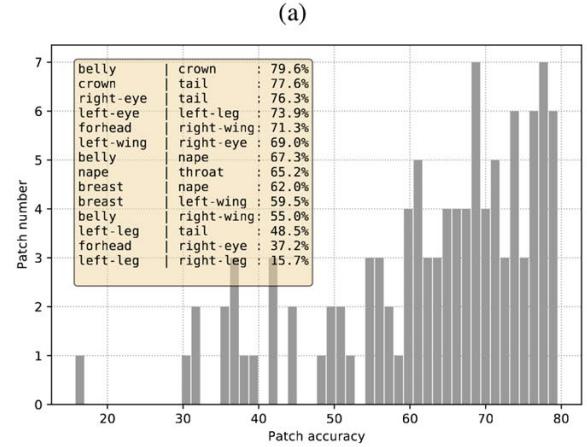
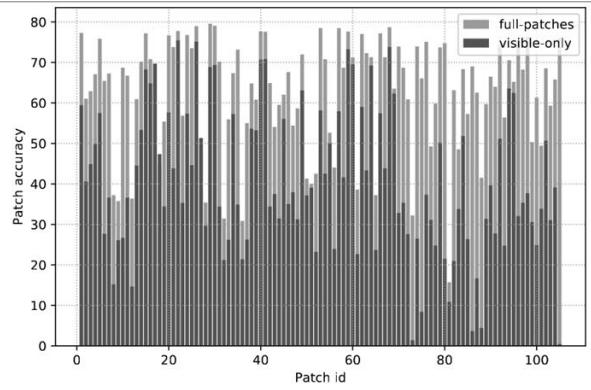


# Pose Aligned Patches



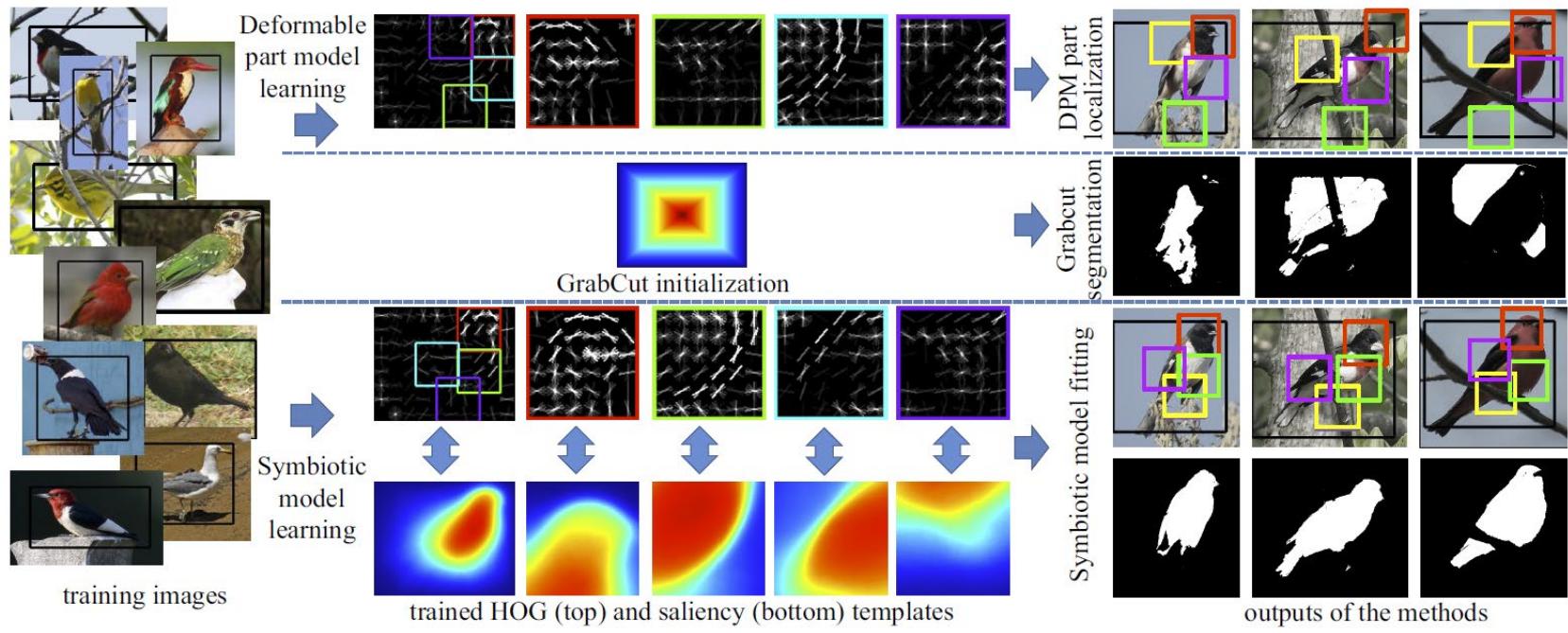
Approach	Annotations	Accuracy
Huang <i>et al.</i> [21]	GT+BB+KP	76.2
Zhang <i>et al.</i> [52]	GT + BB	76.4
Krause <i>et al.</i> [26]	GT+BB	82.8
Jaderberg <i>et al.</i> [23]	GT	84.1
Shu <i>et al.</i> [24]	GT	84.2
Zhang <i>et al.</i> [56]	GT	84.5
Xu <i>et al.</i> [49]	GT+BB+KP+WEB	84.6
Lin <i>et al.</i> [32]	GT+BB	85.1
Cui <i>et al.</i> [9]	GT	86.2
Lam <i>et al.</i> [29]	GT+KP	87.5
PAIRS Only	GT + KP	88.7
PAIRS+Single	GT + KP	<b>89.2</b>

# Pose Aligned Patches

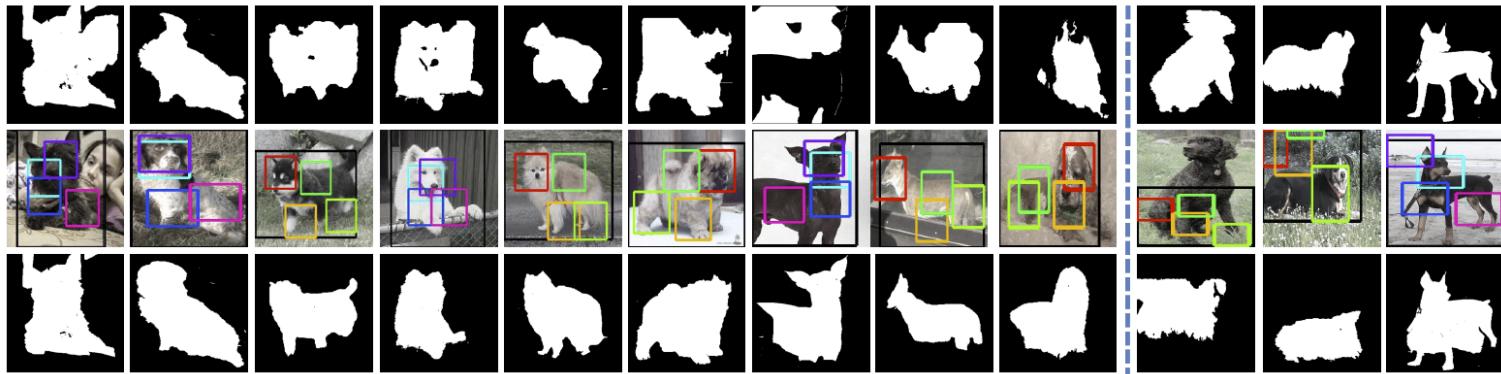
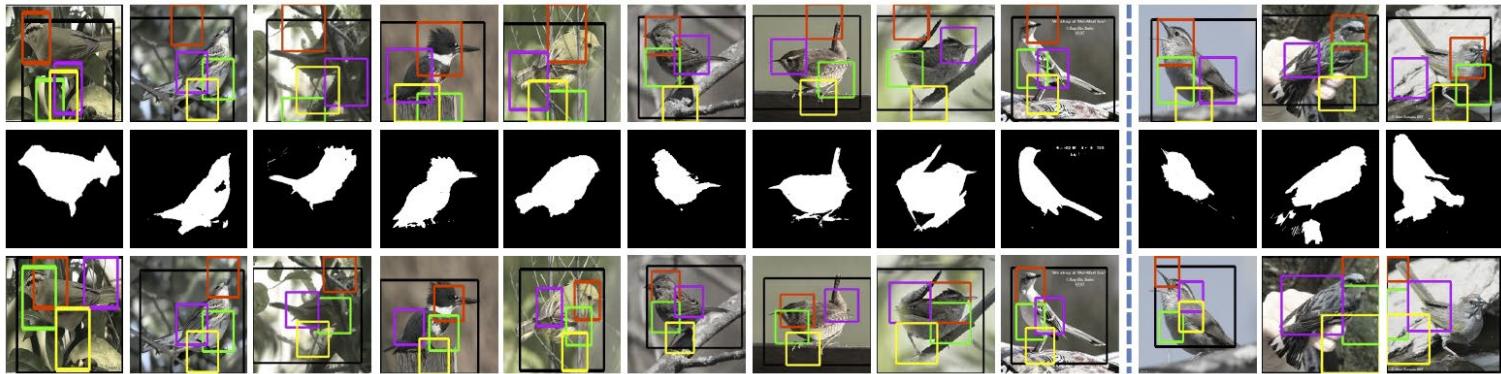


# Segmentation

# Segmentation and Part Localization

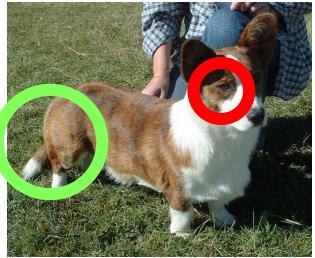


# Segmentation and Part Localization

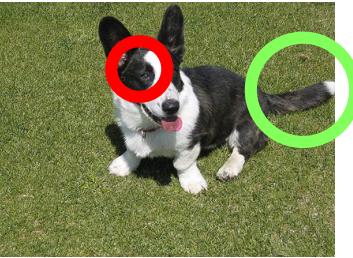


# Human Perception for FGVC

# Why is Fine-Grained Difficult?

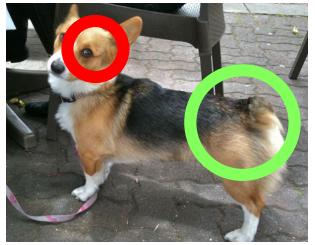


Cardigan Welsh Corgi

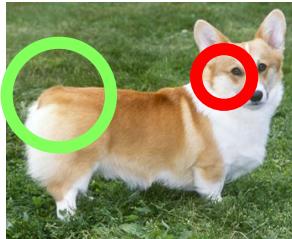


...

Learning



Pembroke Welsh Corgi



...

How to help computers  
select features?



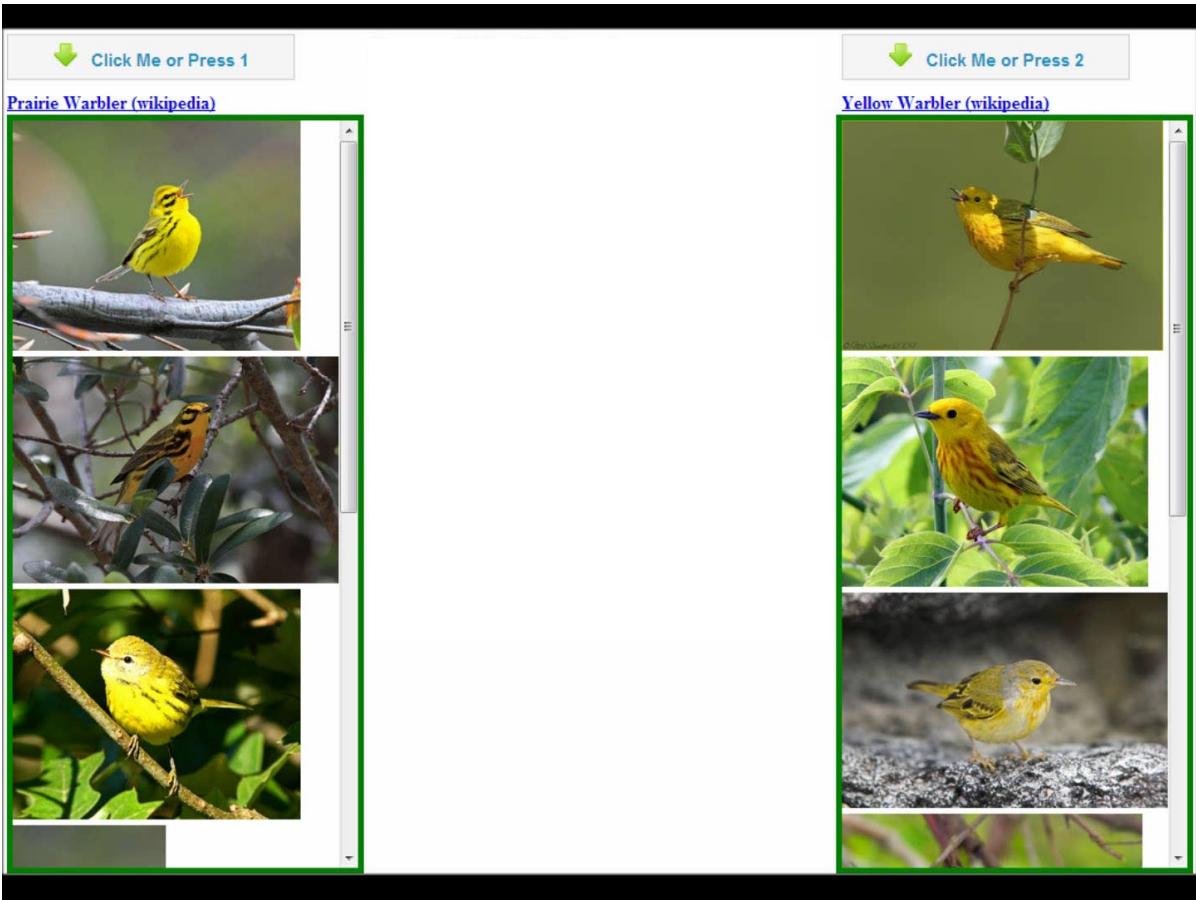
Leveraging the Wisdom of the Crowd for Fine-Grained Recognition

Jia Deng, Jonathan Krause, Michael Stark, Li Fei-Fei



(CVPR 2013) PAMI 2015

# Bubbles Game



Leveraging the Wisdom of the Crowd for Fine-Grained Recognition

Jia Deng, Jonathan Krause, Michael Stark, Li Fei-Fei

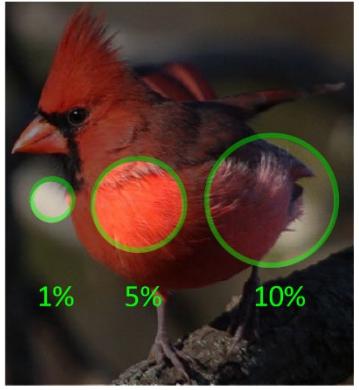
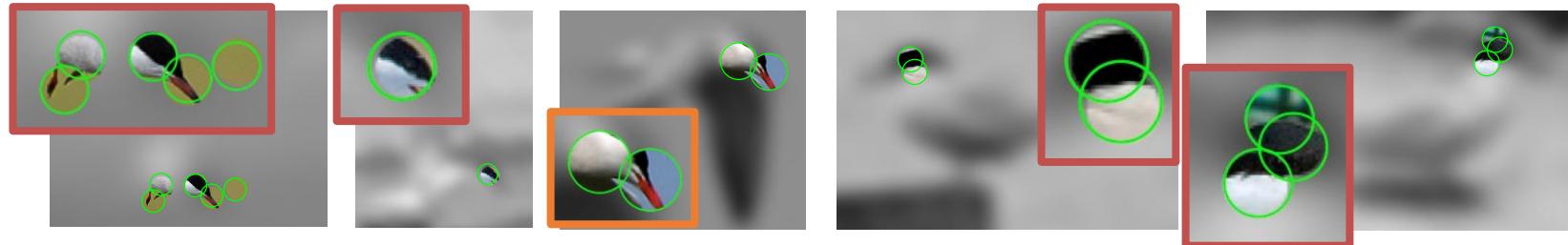


(CVPR 2013) PAMI 2015

# Crowd Picked Bubbles (AMT)

800 top confusing class pairs (via cross-validation)

90K games on Amazon Mechanical Turk



>90% of successful games use <10% of the bounding box

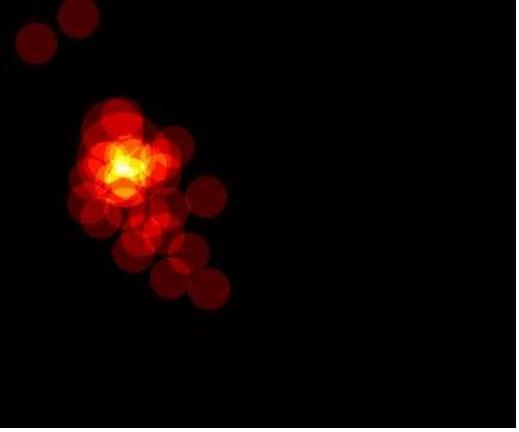
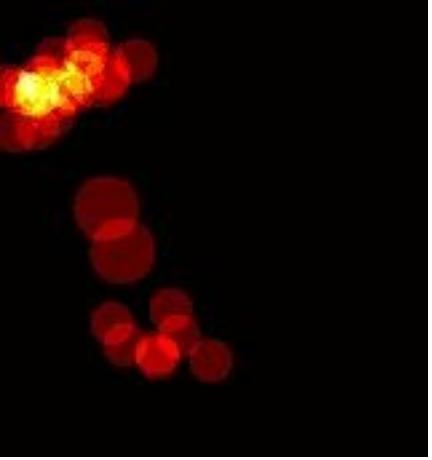


Leveraging the Wisdom of the Crowd for Fine-Grained Recognition

Jia Deng, Jonathan Krause, Michael Stark, Li Fei-Fei

(CVPR 2013) PAMI 2015

# Bubble Heatmaps



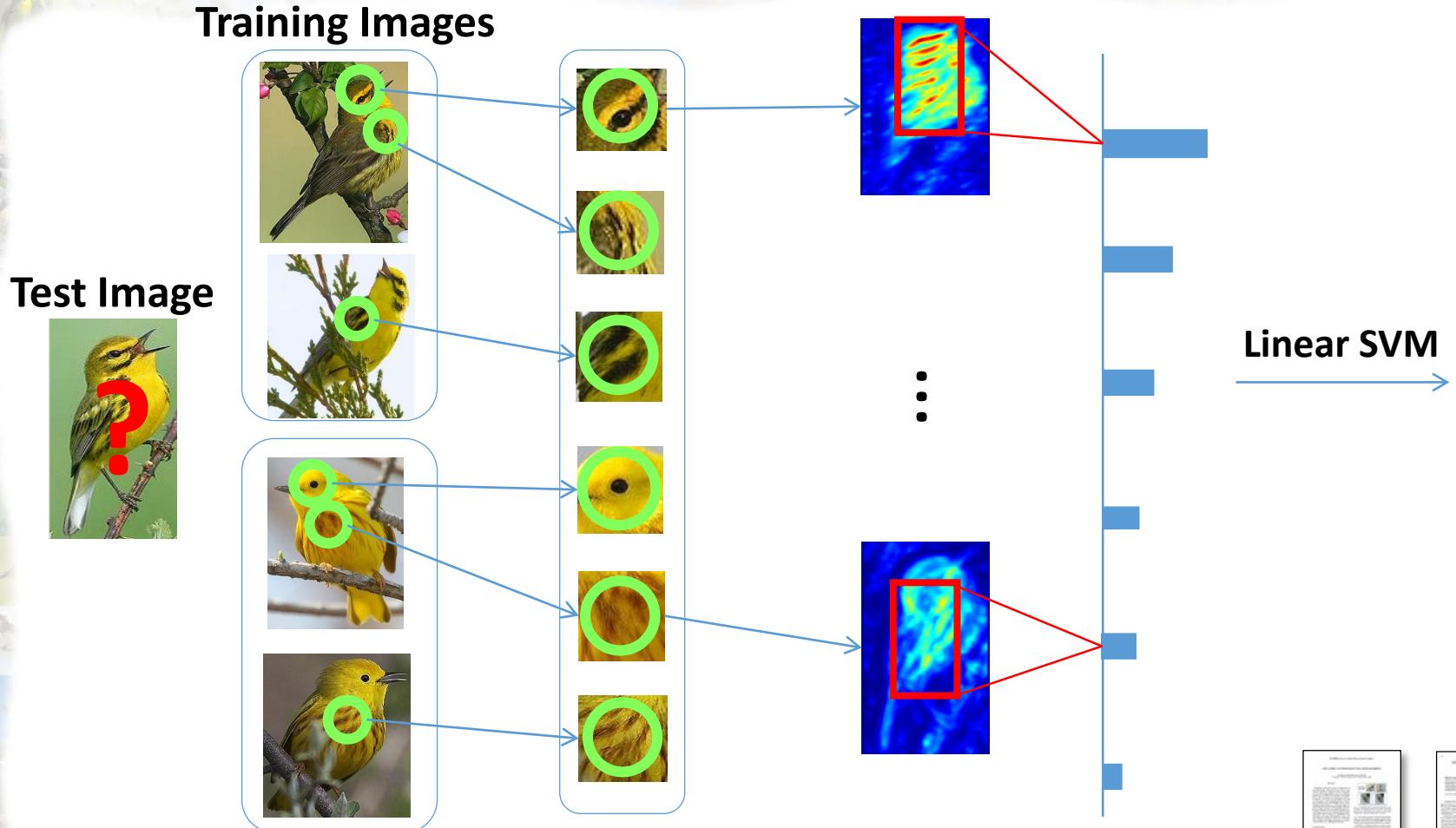
Leveraging the Wisdom of the Crowd for Fine-Grained Recognition

Jia Deng, Jonathan Krause, Michael Stark, Li Fei-Fei



(CVPR 2013) PAMI 2015

# The BubbleBank Representation



Leveraging the Wisdom of the Crowd for Fine-Grained Recognition

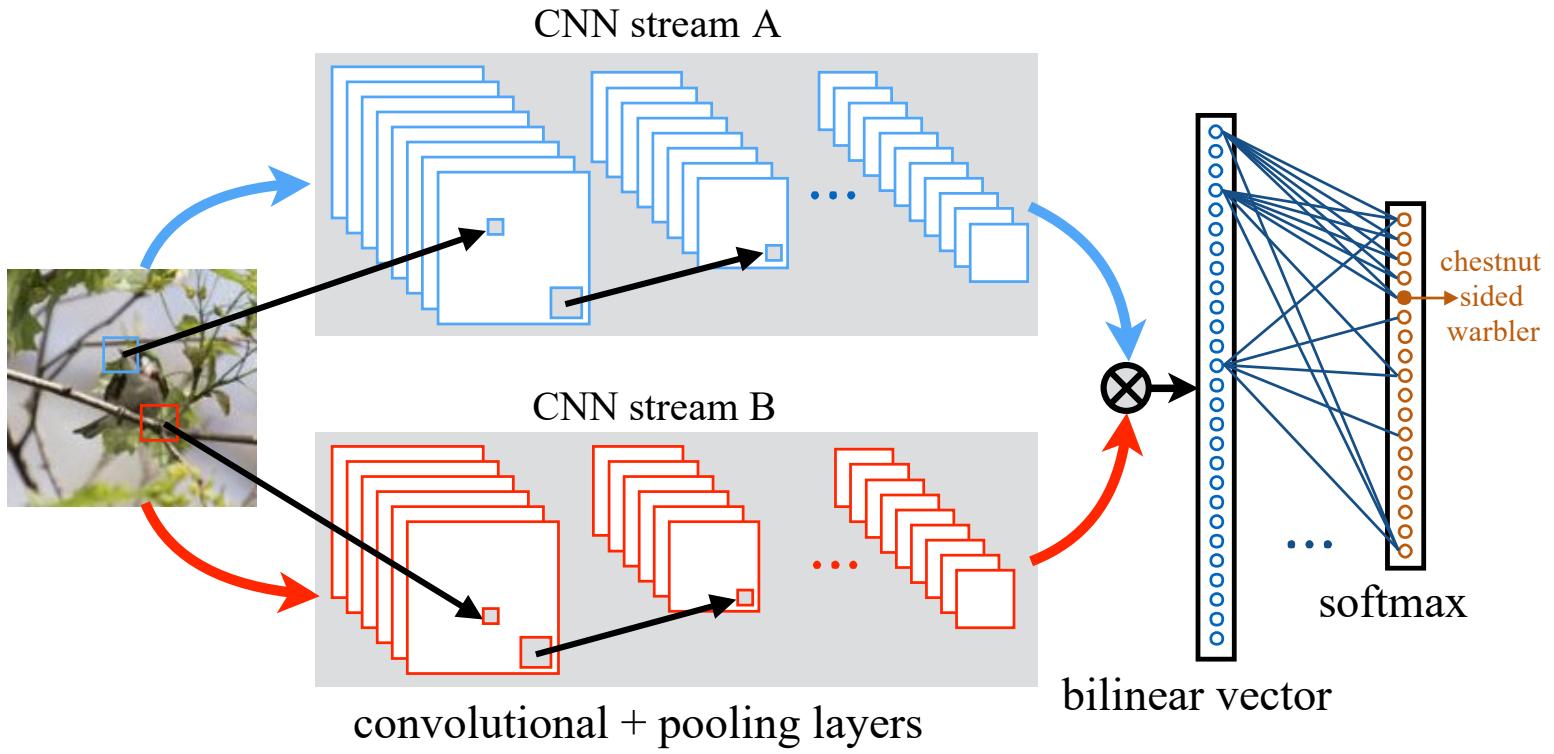
Jia Deng, Jonathan Krause, Michael Stark, Li Fei-Fei



(CVPR 2013) PAMI 2015

# Pooling

# Bilinear Pooling



Bilinear CNN Models for Fine-grained Visual Recognition

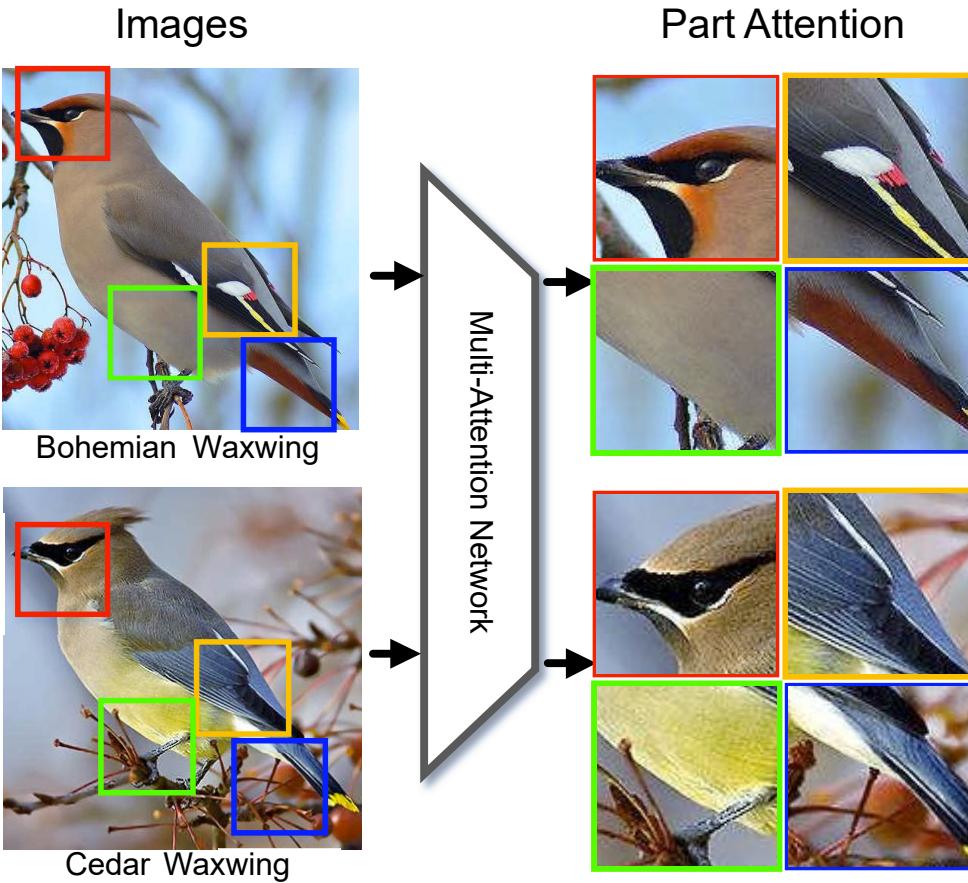
Tsung-Yu Lin, Aruni RoyChowdhury and Subhransu Maji



(ICCV 2015) PAMI 2017

# Attention Mechanisms

# Multi-Attention CNN

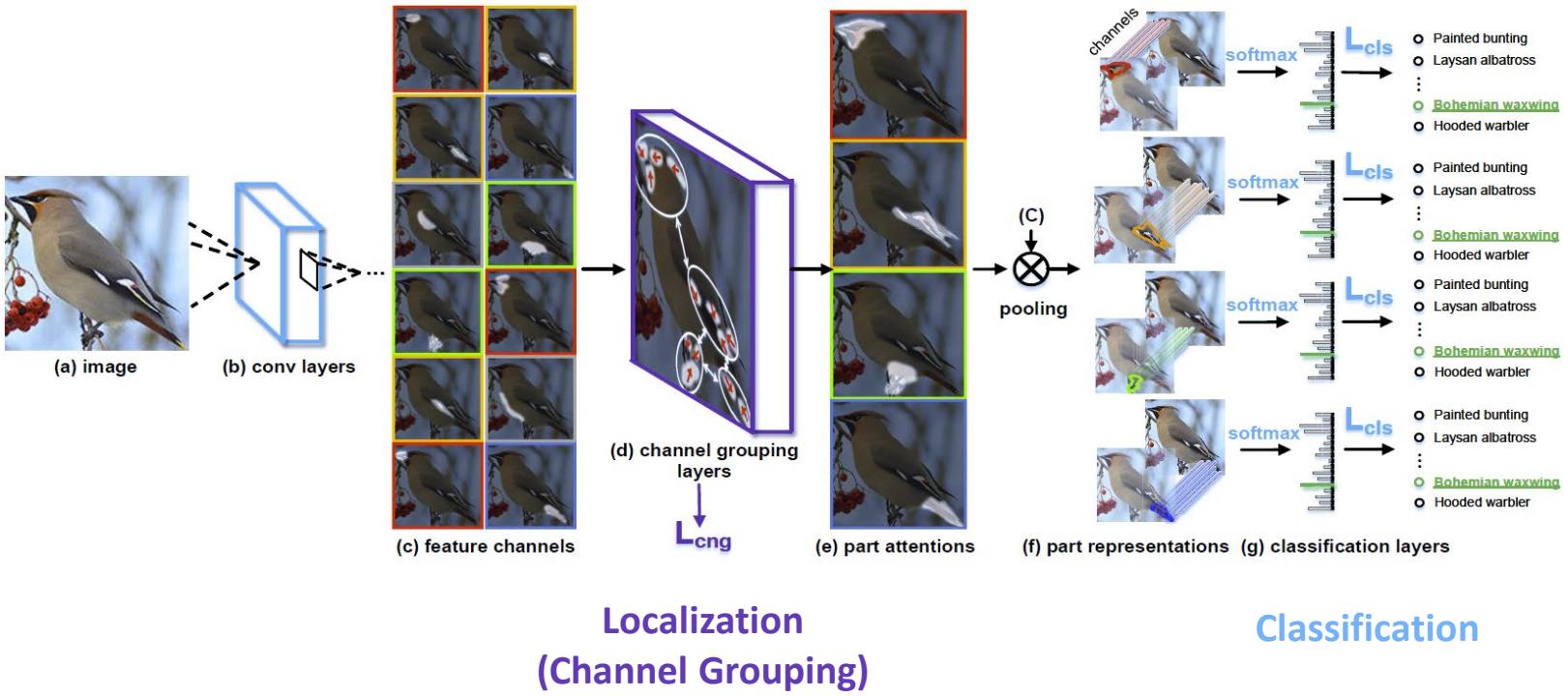


Learning Multi-Attention Convolutional Neural Network for Fine-Grained Image Recognition  
Heliang Zheng, Jianlong Fu, Tao Mei, Jiebo Luo



ICCV 2017

# Multi-Attention CNN



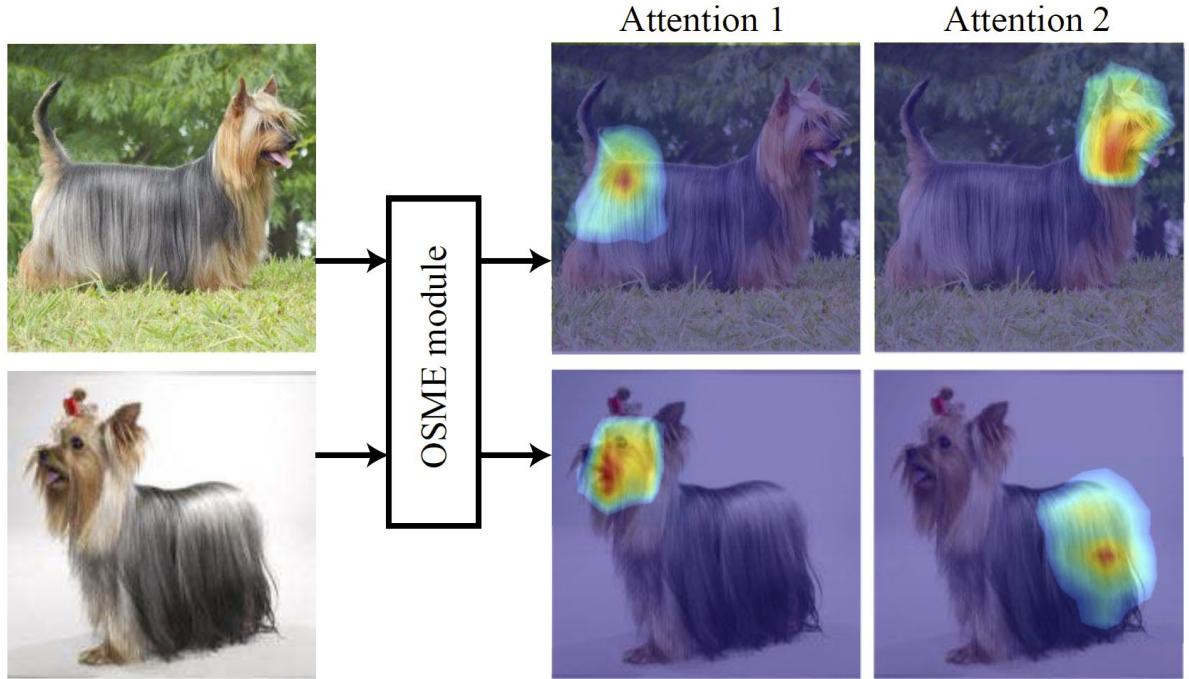
Learning Multi-Attention Convolutional Neural Network for Fine-Grained Image Recognition

Heliang Zheng, Jianlong Fu, Tao Mei, Jiebo Luo



ICCV 2017

# Multi-Attention Multi-Class Constraint



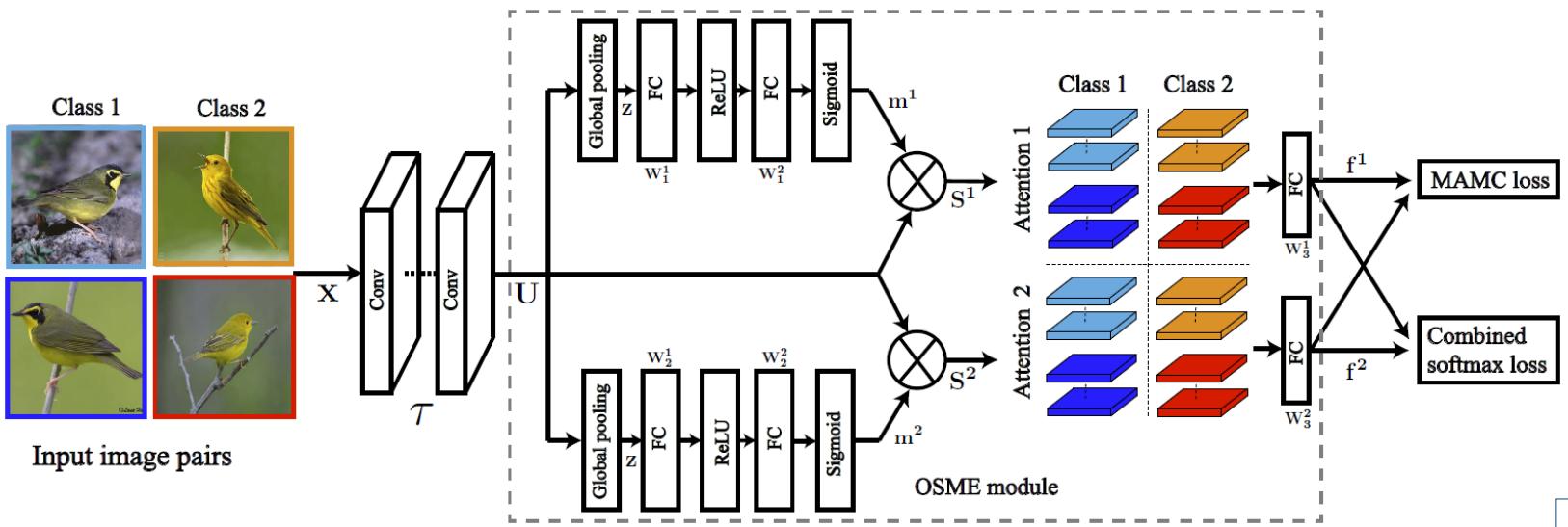
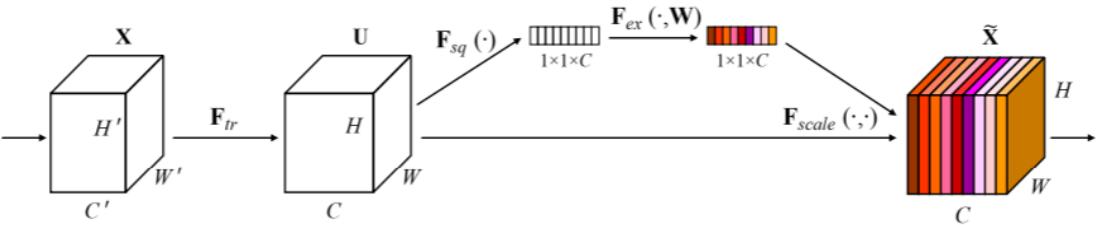
Multi-Attention Multi-Class Constraint for Fine-grained Image Recognition

Ming Sun, Yuchen Yuan, Feng Zhou, Errui Ding



ECCV 2018

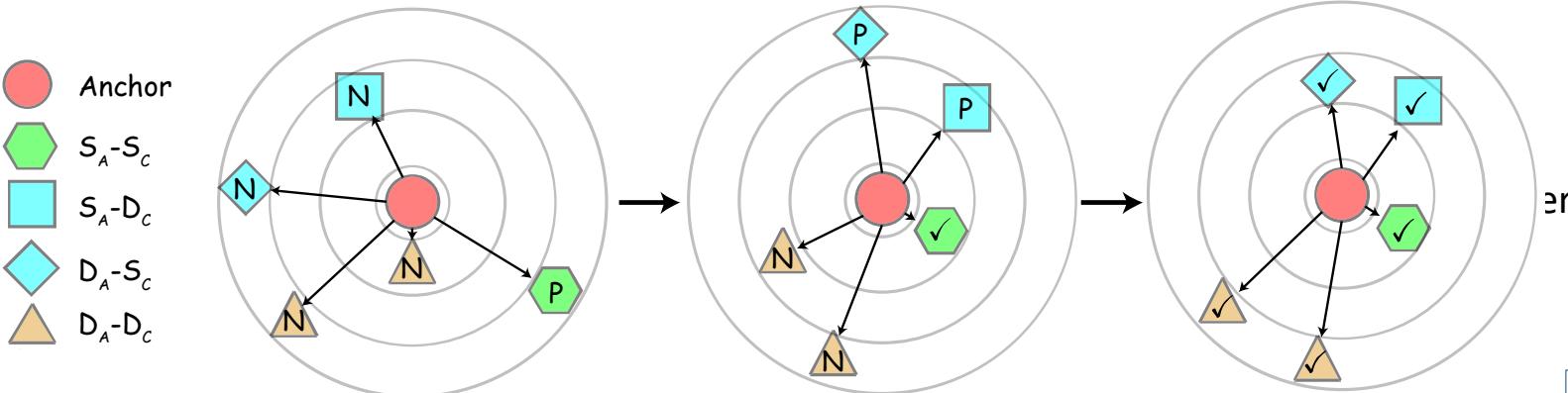
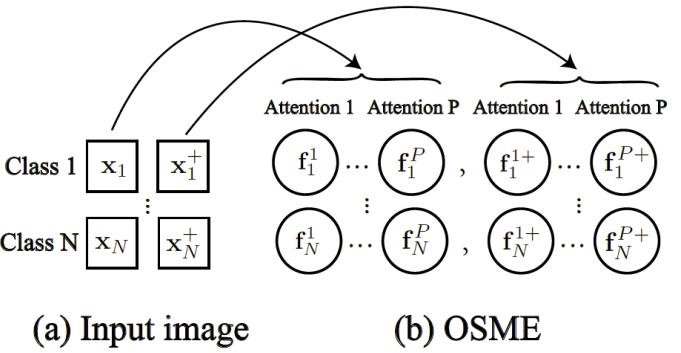
# Multi-Attention Multi-Class Constraint



Multi-Attention Multi-Class Constraint for Fine-grained Image Recognition

Ming Sun, Yuchen Yuan, Feng Zhou, Errui Ding

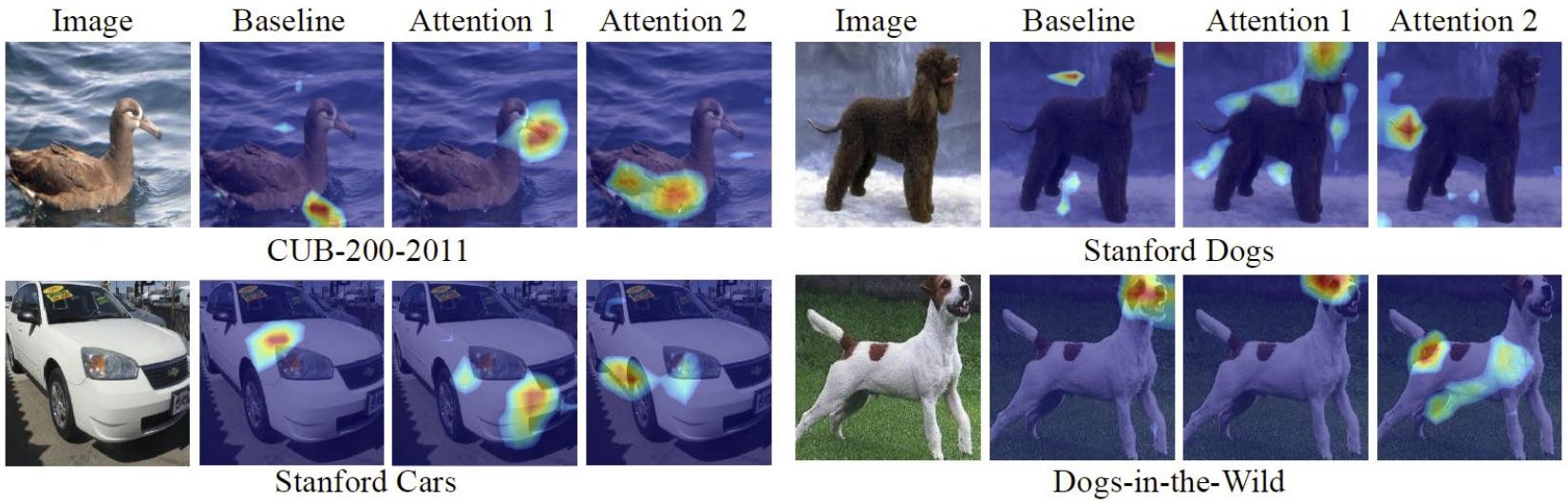
# Multi-Attention Multi-Class Constraint



Multi-Attention Multi-Class Constraint for Fine-grained Image Recognition

Ming Sun, Yuchen Yuan, Feng Zhou, Errui Ding

# Multi-Attention Multi-Class Constraint



Multi-Attention Multi-Class Constraint for Fine-grained Image Recognition

Ming Sun, Yuchen Yuan, Feng Zhou, Errui Ding



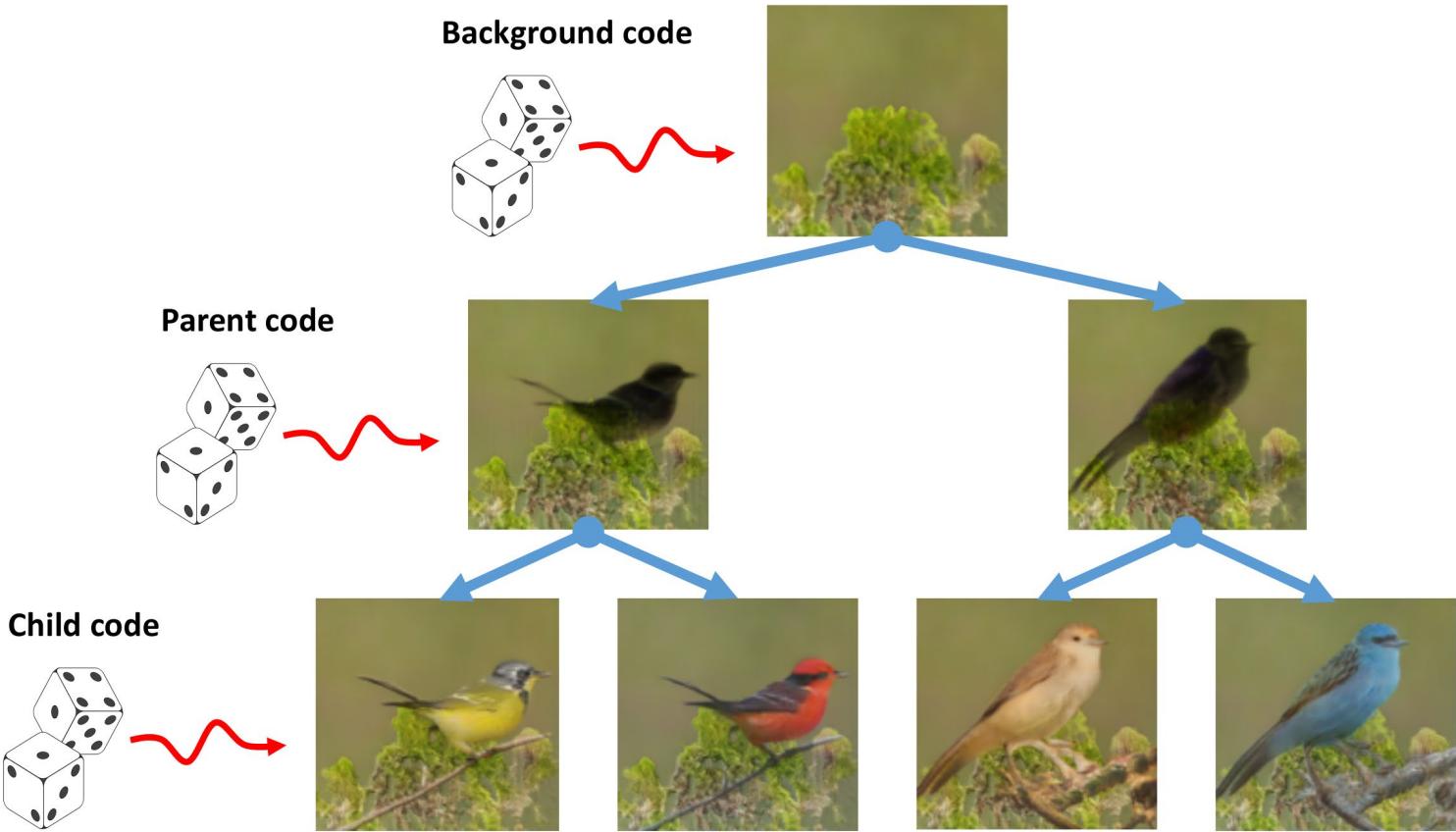
ECCV 2018

# Other Methods

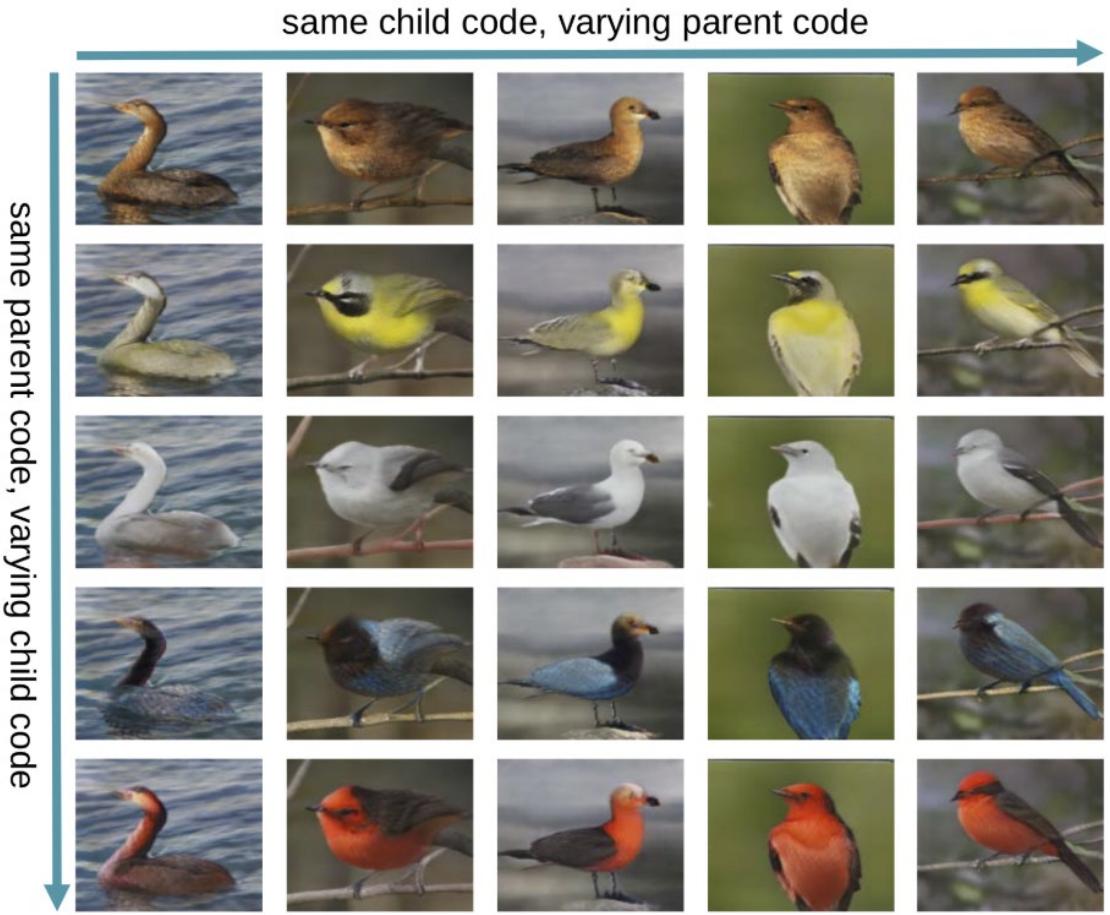
- Taxonomy
- Language
- Few-shot / Zero-shot
- Regularization
- Attributes

# Very Recent Work (CVPR – June 2019)

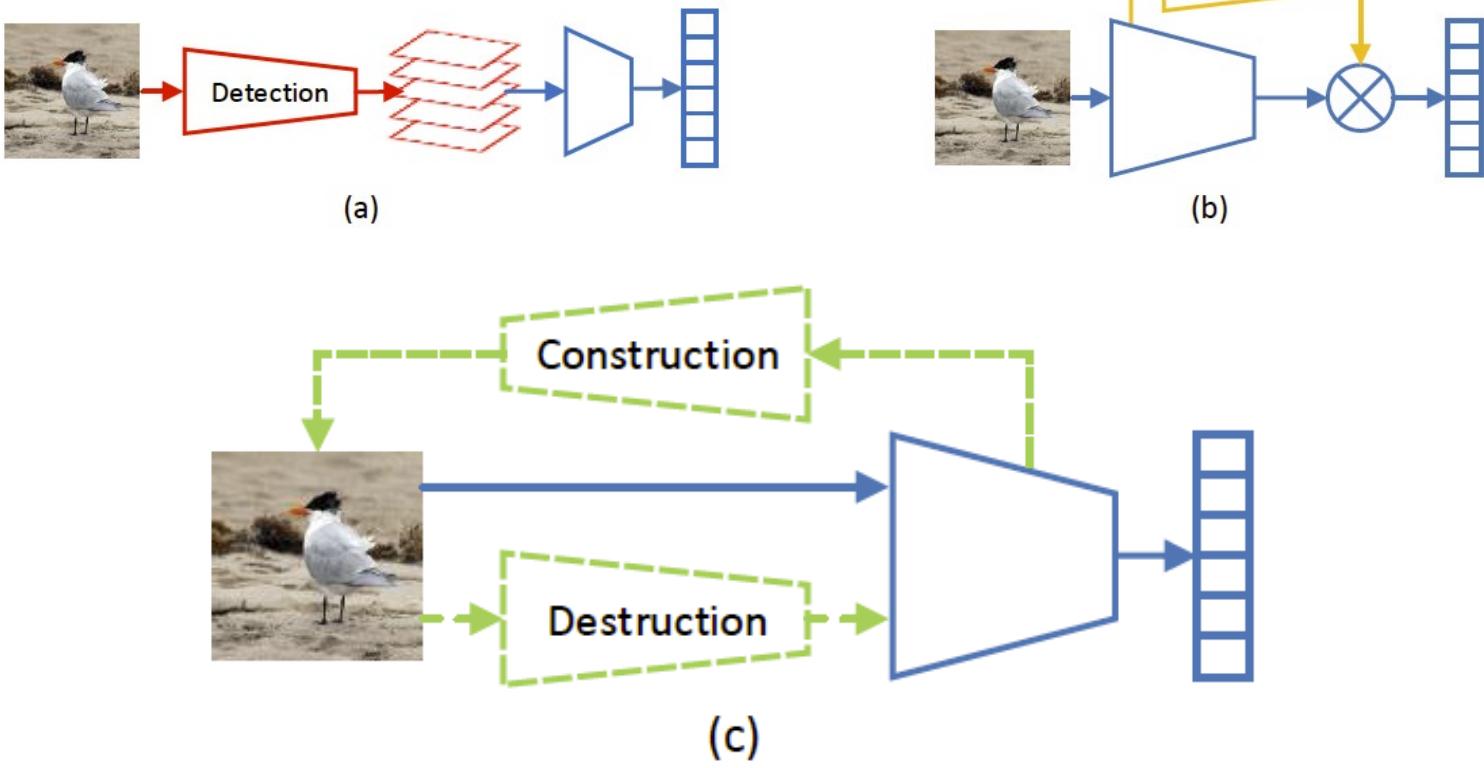
# Disentangling Pose and Appearance



# Disentangling Pose and Appearance



# Destruction/Construction Learning

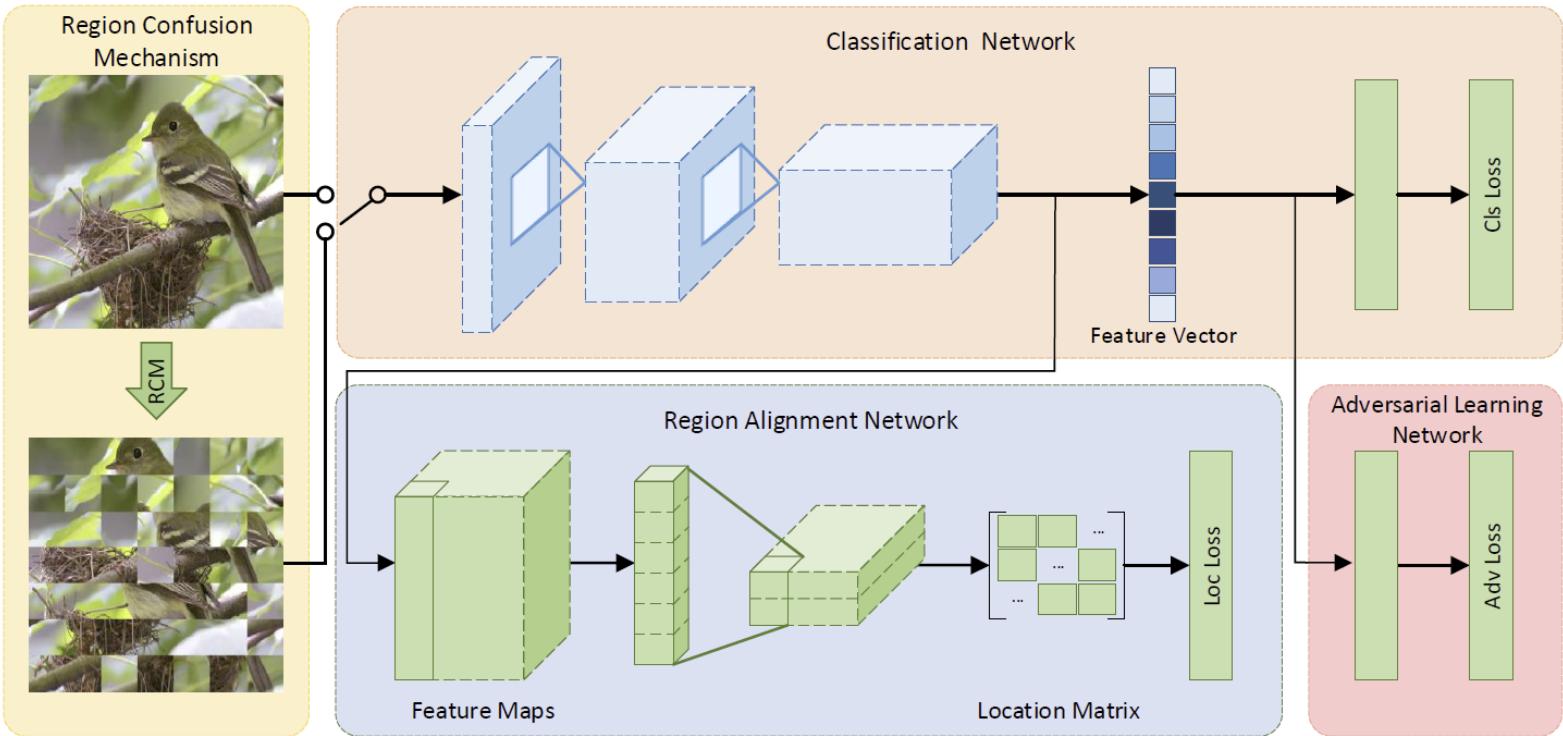


Destruction and Construction Learning for Fine-grained Image Recognition

Yue Chen, Yalong Bai, Wei Zhang, Tao Mei

CVPR 2019

# Destruction/Construction Learning

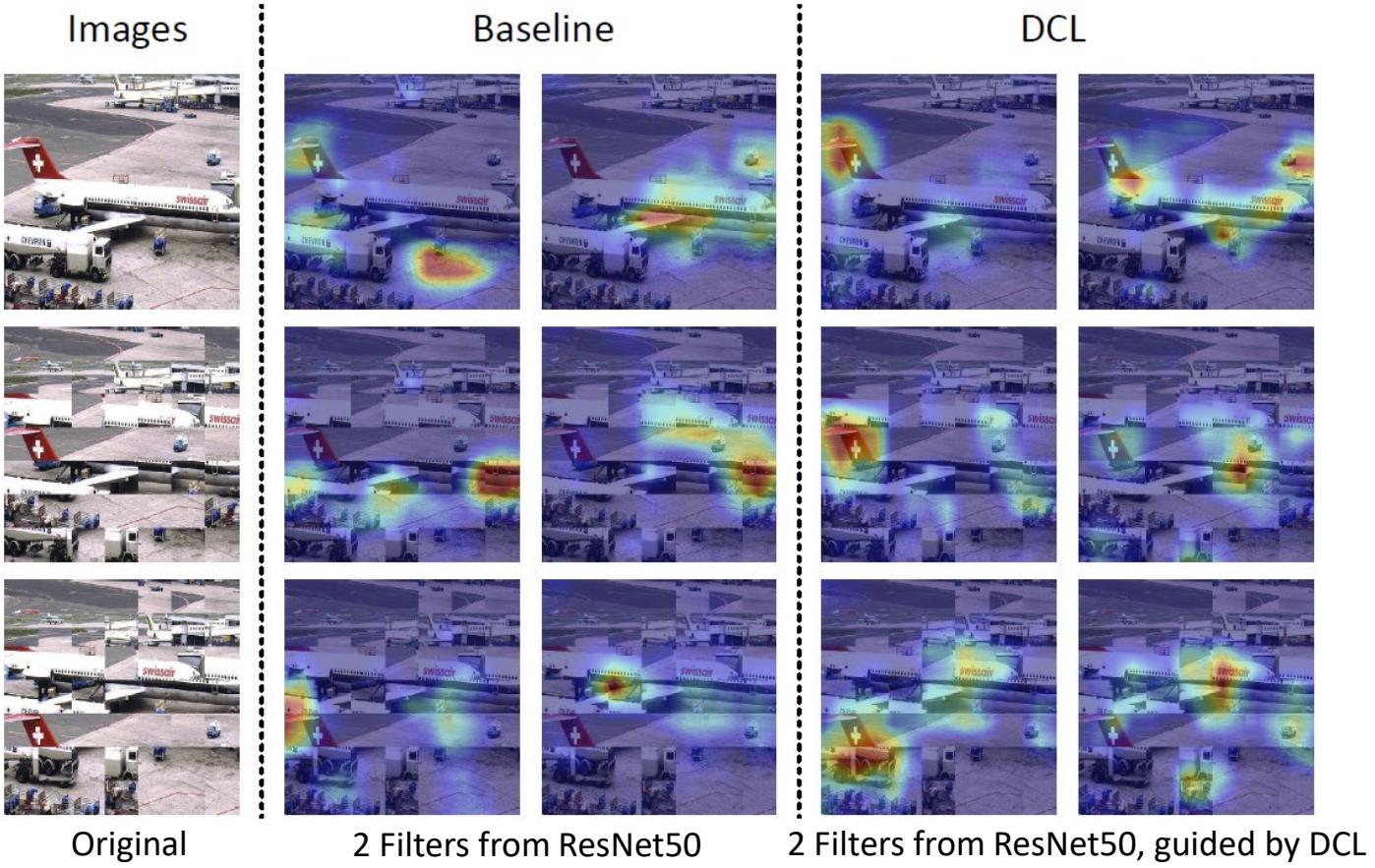


Destruction and Construction Learning for Fine-grained Image Recognition

Yue Chen, Yalong Bai, Wei Zhang, Tao Mei

CVPR 2019

# Destruction/Construction Learning



Destruction and Construction Learning for Fine-grained Image Recognition

Yue Chen, Yalong Bai, Wei Zhang, Tao Mei

CVPR 2019

# Destruction/Construction Learning

Method	Base Model	Accuracy (%)		
		CUB-200-2011	Stanford Cars	FGVC-Aircraft
CoSeq(+BBox) [13]	VGG-19	82.8	92.8	-
FCAN(+BBox) [17]	ResNet-50	84.7	93.1	-
B-CNN [16]	VGGnet	84.1	91.3	84.1
HIHCA [2]	VGG-16	85.3	91.7	88.3
RA-CNN [7]	VGG-19	85.3	92.5	88.2
OPAM [21]	VGG-16	85.8	92.2	-
Kernel-Pooling [5]	VGG-16	86.2	92.4	86.9
Kernel-Pooling [5]	ResNet-50	84.7	91.1	85.7
MA-CNN [42]	VGG-19	86.5	92.8	89.9
DFL-CNN [33]	ResNet-50	87.4	93.1	91.7
DCL	VGG-16	86.9	94.1	91.2
DCL	ResNet-50	<b>87.8</b>	<b>94.5</b>	<b>93.0</b>

Table 1. Comparison results on three different standard datasets. Base Model means the backbone network used in the method.

# Datasets

# Natural World

- Birds
  - CUB-200 (2010/2011)
  - Birdsnap
  - NABirds
- Dogs
  - Stanford
  - Columbia
  - Oxford-IIIT PETS
- Insects
  - Butterflies
- Flowers
  - Oxford Flowers
  - NEC
  - Xingse
- Fungi
  - Competition
- Plants/Trees
  - Caltech
- General
  - iNaturalist 2017/18/19

# Man-made Domains

- Cars
  - Stanford Cars
  - Comp-Cars
- Aircraft
  - FGVC Aircraft
- Clothing/Apparel
  - iMaterialist Fashion 2017/18/19
- Furniture
  - iMaterialist Furniture 2018/19
- Food
  - Food-101

# FGVC Resources



# FGVC Workshops

FGVC

@ CVPR 2011, Colorado Springs, CO

FGVc<sup>2</sup>

@ CVPR 2013, Portland, OR

FGVc<sup>3</sup>

@ CVPR 2015, Boston, MA

FGVc<sup>4</sup>

@ CVPR 2017, Honolulu, HI

FGVc<sup>5</sup>

@ CVPR 2018, Salt Lake City, UT



# FGVC Workshops

# FGVC6

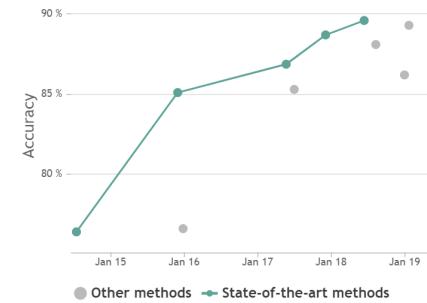
- Was held June 17<sup>th</sup> at CVPR 2019
- Workshop Website (<http://fgvc.org/FGVC6/>)
- 10 Challenges! (<https://www.kaggle.com/FGVC6/>)

# fgvc.org/ and Survey Article



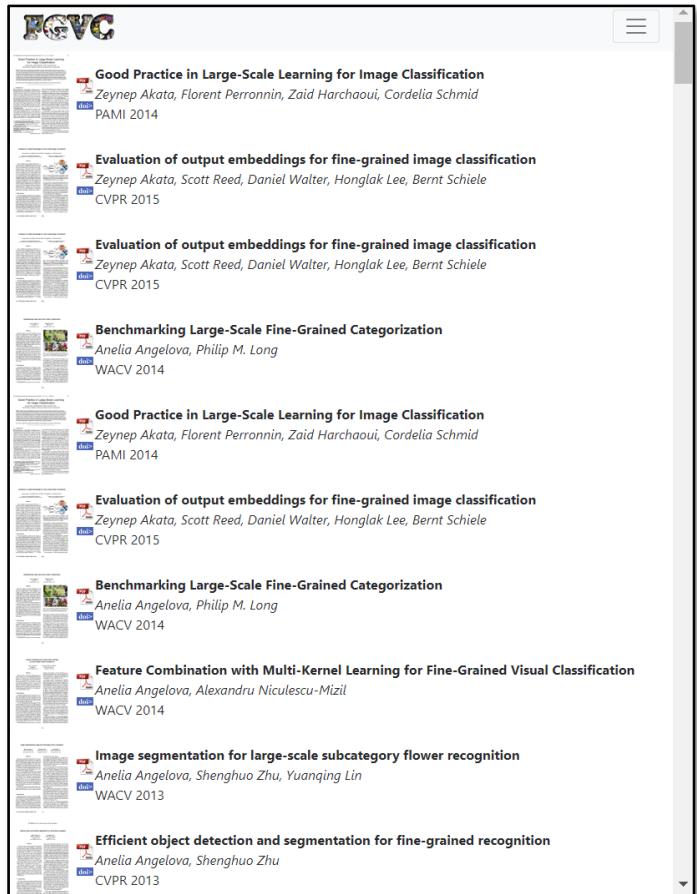
[paperswithcode.com](https://paperswithcode.com)

Fine-Grained Image Classification on  
CUB-200-2011



Rank	Method	Accuracy	Paper Title	Year	Paper	Code
1	Inception-V3	89.6	Large Scale Fine-Grained Categorization and Domain-Specific Transfer Learning	2018	<a href="#">Paper</a>	<a href="#">Code</a>
2	WS-DAN	89.3%	See Better Before Looking Closer: Weakly Supervised Data Augmentation Network for Fine-Grained Visual Classification	2019	<a href="#">Paper</a>	<a href="#">Code</a>
3	MPN-COV	88.7%	Towards Faster Training of Global Covariance Pooling Networks by Iterative Matrix Square Root Normalization	2017	<a href="#">Paper</a>	<a href="#">Code</a>
4	Hierarchical Semantic Embedding	88.1%	Fine-Grained Representation Learning and Recognition by Exploiting Hierarchical Semantic Embedding	2018	<a href="#">Paper</a>	<a href="#">Code</a>
5	PC-DenseNet-161	86.87%	Pairwise Confusion for Fine-Grained Visual Classification	2017	<a href="#">Paper</a>	<a href="#">Code</a>
6	A3M	86.2%	Attribute-Aware Attention Model for Fine-grained Representation Learning	2019	<a href="#">Paper</a>	<a href="#">Code</a>
7	RA-CNN	85.3%	Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-Grained Image Recognition	2017	<a href="#">Paper</a>	<a href="#">Code</a>
8	Bilinear-CNN	85.1%	Bilinear CNN Models for Fine-Grained Visual Recognition	2015	<a href="#">Paper</a>	<a href="#">Code</a>
9	PS-CNN	76.6%	Part-Stacked CNN for Fine-Grained Visual Categorization	2015	<a href="#">Paper</a>	<a href="#">Code</a>
10	Part RCNN	76.4%	Part-based R-CNNs for Fine-grained Category Detection	2014	<a href="#">Paper</a>	<a href="#">Code</a>

# fgvc.org/ and Survey Article



Zhao, et al.  
IJAC 2017



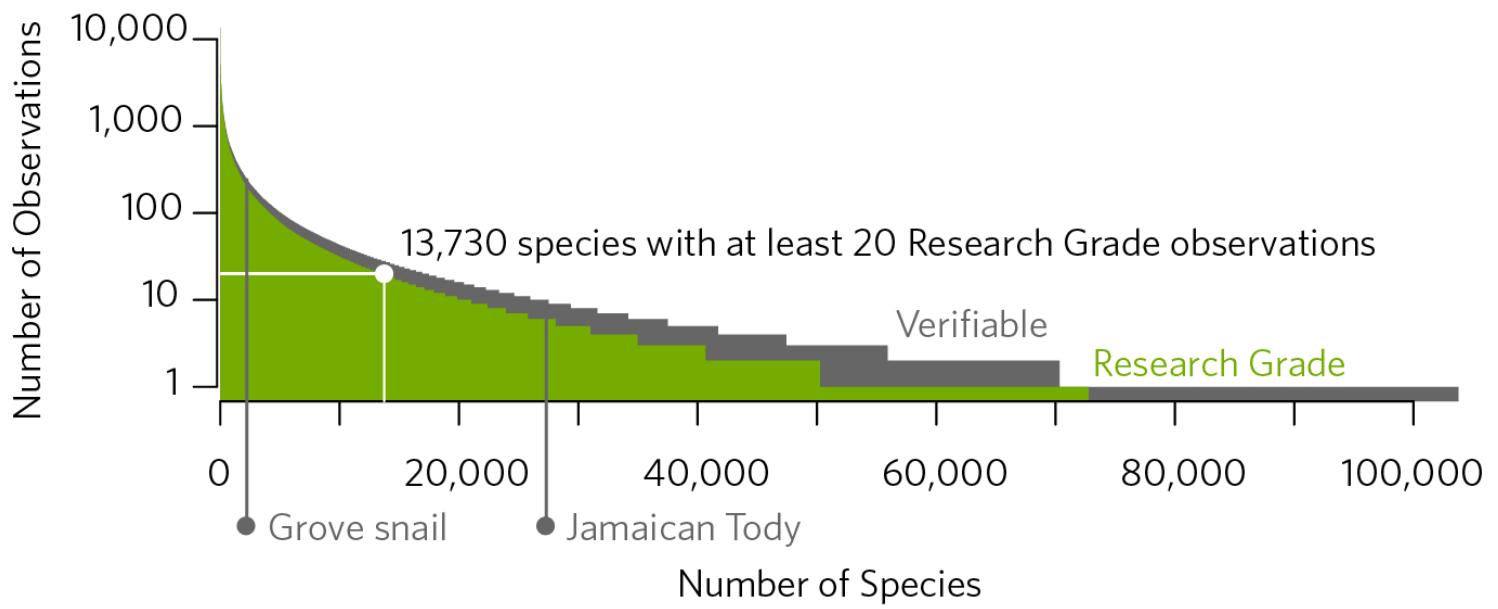
Zheng, et al.  
ICSP 2018



# **What's Missing, Future Directions**

# Future Directions

# Long Tail

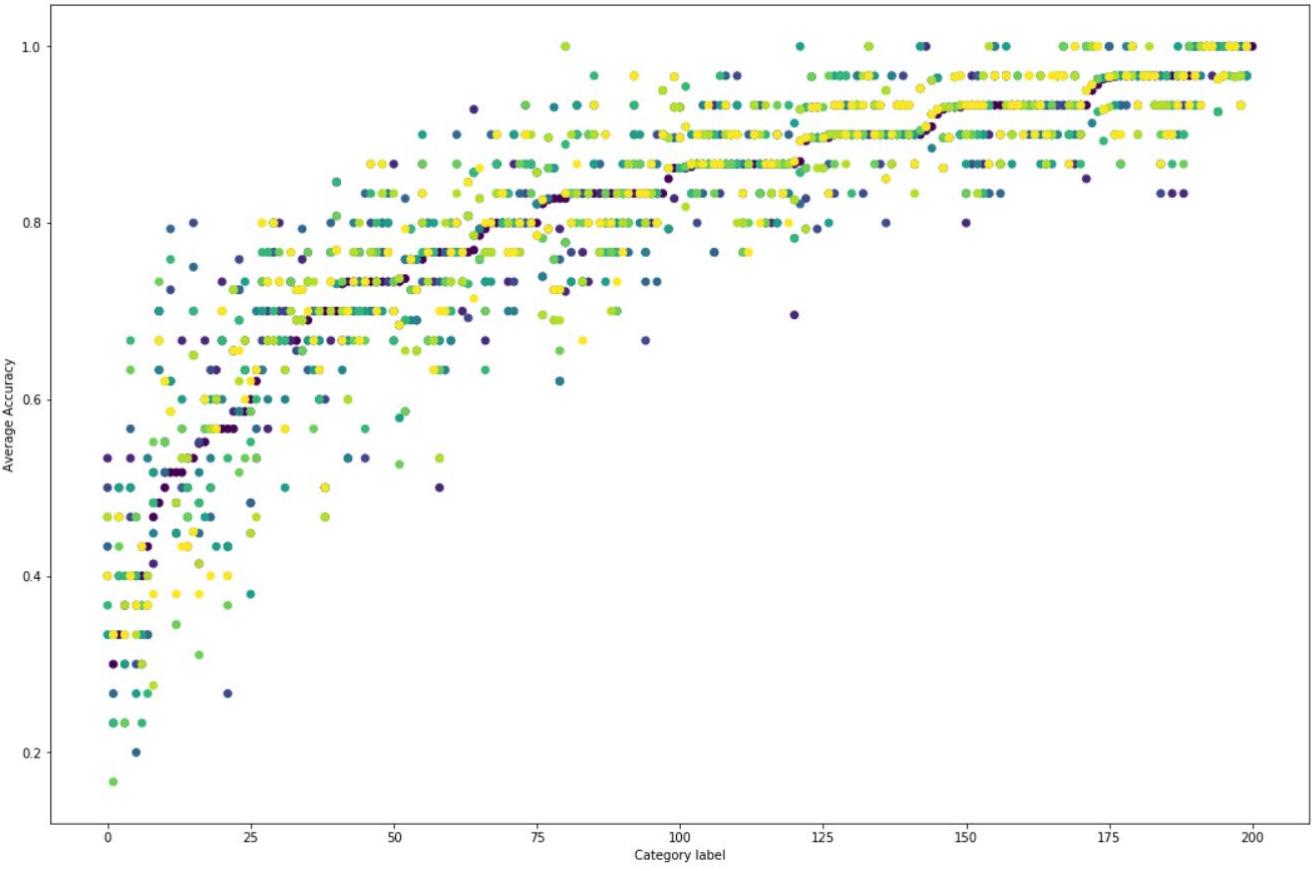


iNaturalist Dataset

# Future Directions

- Long Tail
- Dataset Availability
- 1-Shot/Few-shot

# Variability of Deep Learning



# Future Directions

- Long Tail
- Dataset Availability
- 1-Shot/Few-shot
- Variability of CNN
- Human Perceptual Expertise
- Qualitative Identification