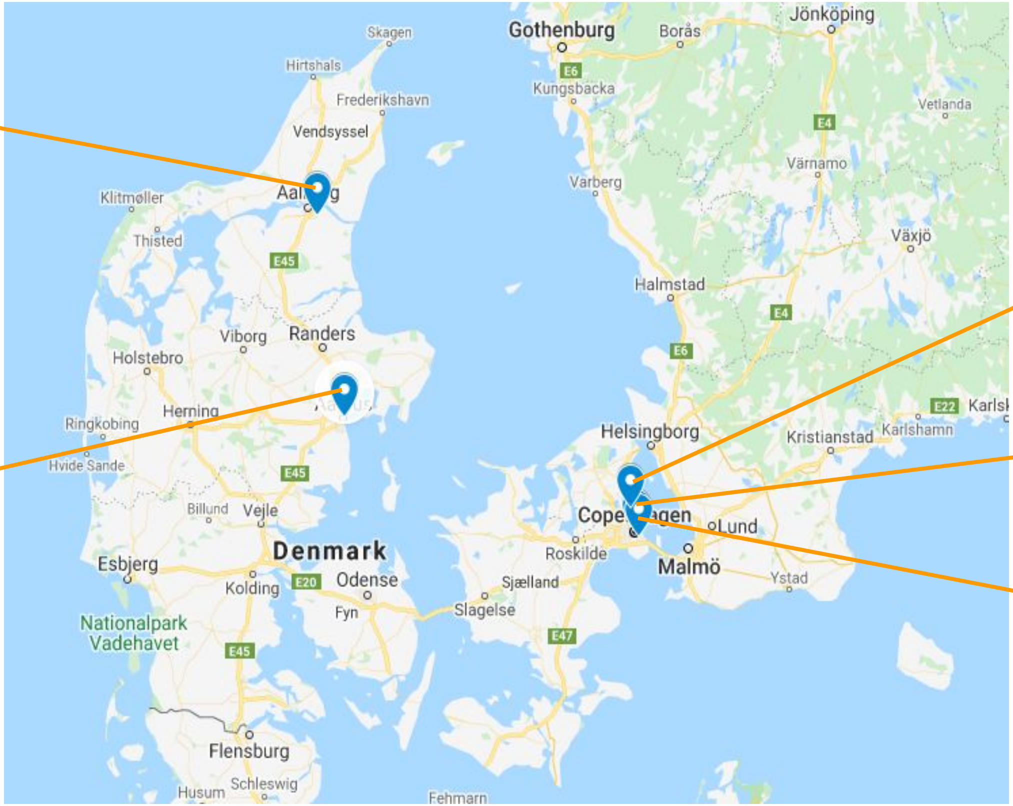
Challenges in Fine-Grained Visual Analysis

Serge Belongie

Cornell University

UNIVERSITY OF COPENHAGEN

Pioneer Centre for AI: DNRF Grant No. P1



ITU

CARLSBERG FOUNDATION

novo nordisk fonden

THE VELUX FOUNDATIONS

VILLUM FONDEN VELUX FONDEN

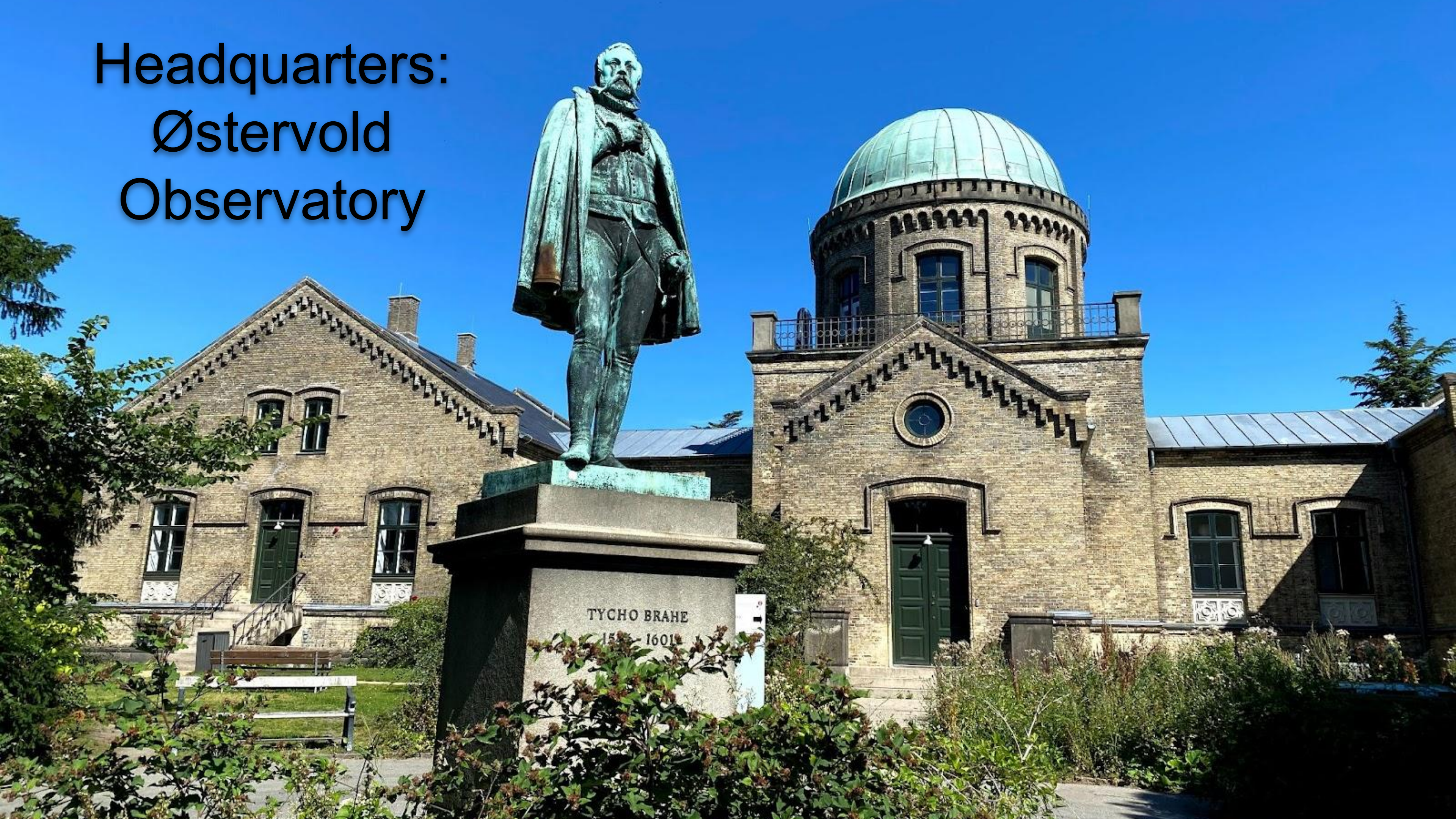
Danmarks Grundforskningsfond
Danish National Research Foundation

LUNDBECKFONDEN

Better lives through new knowledge

@AiCentreDK

Headquarters: Østervold Observatory



Collaboratory Themes & Co-Leads

Cx	Causality and Explainability	Jonas Peters, KU; Aasa Feragen, DTU, Ira Assent, AU
Xr	Extended Reality	Dan Witzner Hansen, ITU; Kasper Hornbæk, KU, Hans Gellersen, AU
Fg	Fine Grained Analysis	Mads Nielsen, KU; Thomas Moeslund, AAU
Lo	Learning Theory and Optimization	Ole Winther, DTU/KU; Christian Igel, KU
Sd	Signals and Decoding	Lars Kai Hansen, DTU; Zheng-Hua Tan, AAU
Sl	Speech and Language	Barbara Plank, ITU; Anders Søgaard, KU
Ng	Networks and Graphs	Sune Lehmann, DTU; David Dreyer Lassen, KU

Collaboratory Themes & Co-Leads

Cx	Interpretable AI, Patient Trajectories, Privacy, Fairness, Bias, Pandemic Prediction	Jonas Peters, KU; Aasa Feragen, DTU, Ira Assent, AU
Xr	AR/VR, Human-Centered Computing, Hand Tracking, Active Illumination, 3D Reconstruction, Simulation Environments, Synthetic Data, Accessibility	Dan Witzner Hansen, ITU; Kasper Hornbæk, KU, Hans Gellersen, AU
Fg	Species Identification, Medical Diagnosis, Anomaly Detection, Computational Pathology, Arts & Culture Informatics, Knowledge Bases	Mads Nielsen, KU; Thomas Moeslund, AAU
Lo	Algorithms & Architectures, Reinforcement Learning, Operations Research, Transportation Problems, Optimal Control	Ole Winther, DTU/KU; Christian Igel, KU
Sd	Telemedicine, Remote Sensing, Eye Tracking, Neuroscience, Brain Decoding, Environmental Monitoring, Biometrics, Egocentric Sensing, Consciousness	Lars Kai Hansen, DTU; Zheng-Hua Tan, AAU
Sl	Natural Language Processing, Speech Recognition, Misinformation Detection, Automated Translation, Predictive Models, Electronic Medical Records	Barbara Plank, ITU; Anders Søgaard, KU
Ng	Social Data Science, Federated Learning, Privacy-Preserving Contact Tracing, Mobility Analytics	Sune Lehmann, DTU; David Dreyer Lassen, KU

Collaboratories × Societal Impact Areas

	💊 Biotech, Life, and Health Sciences	🌥️ Climate and Conservation	🎓 Education and Capacity Building	♿ Equality and Inclusion	💹 Economic Growth and Entrepreneurship	🚒 Crisis Response	📑 Information Verification and Validation	🏢 Energy and Infrastructure	⚖️ Security, Ethics, and Justice	🤝 Public and Social Sector
💡 CX										
👓 XR										
🦜 FG										
🎯 LO										
⚡ SD										
💬 SL										
🔗 NG										

Collaboratories × Societal Impact Areas

	💊 Biotech, Life, and Health Sciences	🌤️ Climate and Conservation	🎓 Education and Capacity Building	♿ Equality and Inclusion	💹 Economic Growth and Entrepreneurship	C Res
💡 CX	✅					
👓 XR	✅					
🦜 FG						
🎯 LO						
⚡ SD	✅			✅	✅	
💬 SL						
🔗 NG						

Project: Democratization of EEG (Lars Kai)

Motivation: Neurotechnology can connect everyday behavior with brain dynamics and provide diagnostic support e.g. for epilepsy. WHO has identified a world wide epilepsy diagnosis gap.

Data: Wearable EEG, focus on low cost EEG data acquisition. EEG is entering the "ImageNet"-phase with marked increased access to data.

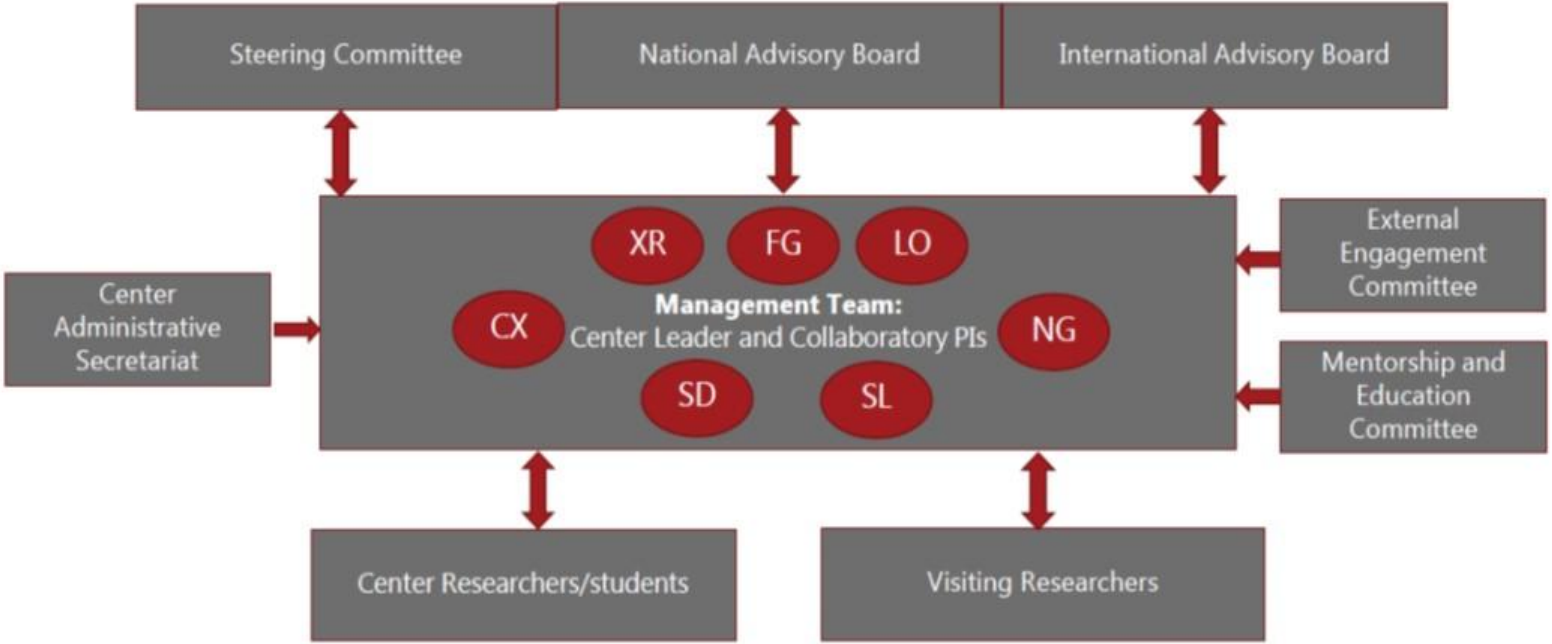
Challenges: Extreme signal-to-noise conditions. Real-time quality and control/interactivity.

Funding: EEG project eGAP funded by EU/Eurostars, BrainCapture, DTU. Funding history: NIH, Lundbeck, NNF, IFDK

SD Moonshot: Global access to neurotechnology.
Moonshot: Foundational EEG models with explainability

Collaborators in P1: Cogsys, Witzner, (Feragen)
Collaborators outside P1: Neurologists, Cognitive Scientists, Hearing Aid business sector and start-ups.

P1: Organisational Structure



7 step plan - example for Visipedia

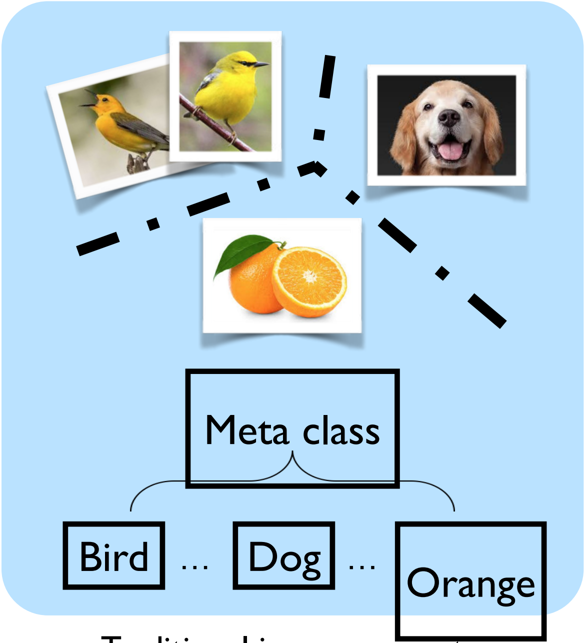
Phase	Activity
1. Inception	AI researchers observe that birding is a popular hobby around the globe, and birders pride themselves on being able to distinguish between bird species with very similar appearances. Social Science and Humanities colleagues who study public participation highlight the potential of motivated teams to take collective action …
2. Early Explorations	AI researchers build a scrappy dataset of labeled bird images from internet based resources and obtain baseline results with state-of-the-art Machine Learning techniques. It is clear that the problem is very difficult.
3. Painstorming	AI researchers travel to the Lab of Ornithology to learn about the community's needs. Birders don't need a machine to tell them the difference between a pigeon and a sparrow. They need the machine to tell them the difference between a blue grosbeak and an indigo bunting. If they help train the machine, they want the …
4. Deep Dive	AI researchers team with ornithologists to create large, world class dataset of labeled bird images, and invent new algorithms for discriminating among tightly related visual classes, thereby laying the foundations of a new subfield: Fine Grained Visual Categorization. Ornithologists release Merlin bird photo ID app for iPhone …
5. Branching Out	AI researchers and experts from domains including plant disease, entomology, nutrition science, and apparel design launch a new workshop featuring visual classification competitions on challenging datasets. AI researchers join with the California Academy of Sciences to add photo ID functionality to the iNaturalist …
6. Going Global	AI researchers visit the Global Biodiversity Information Facility (GBIF) to explore how to provide the tech stack behind the above apps to every area of biodiversity research in a socially responsible manner, with proper attribution and citation mechanisms. Together with Google's TensorFlow Hub team, they establish a new …
7. Moonshot	We aspire to create a system that can recognize every living organism on earth based on photos, sound, and video.

Outline

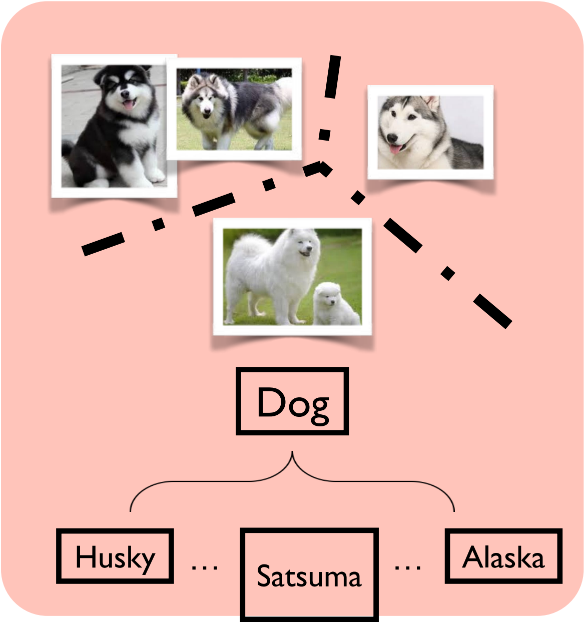
- Introducing granularity
- Subordinate categories
- Parts & Attributes
- Long-tailed distributions
- Popular datasets
- Beyond categorization
- Open problems

Introduction

Fine-grained image recognition *vs.* Generic image recognition



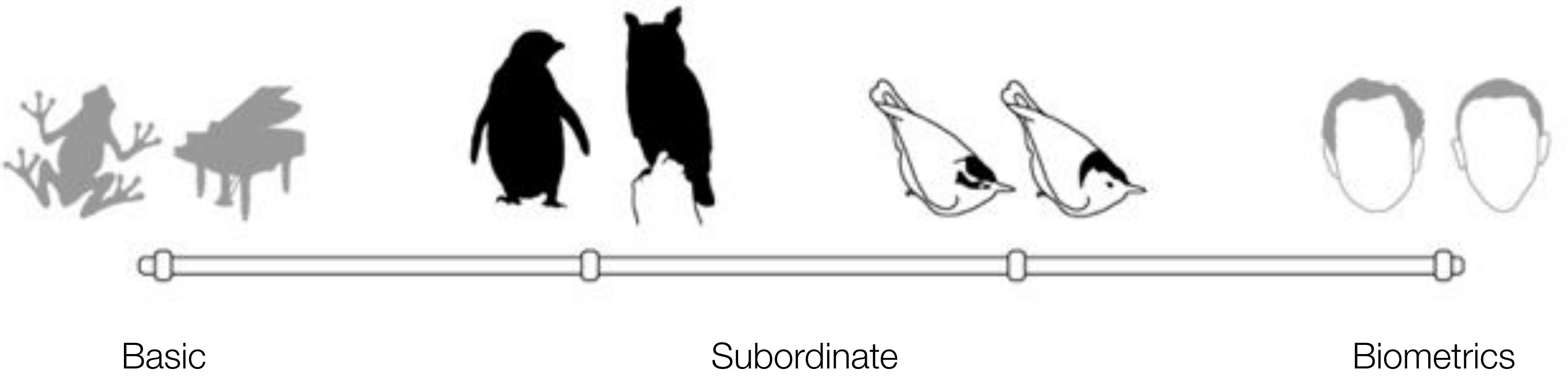
Traditional image recognition
(Coarse-grained)



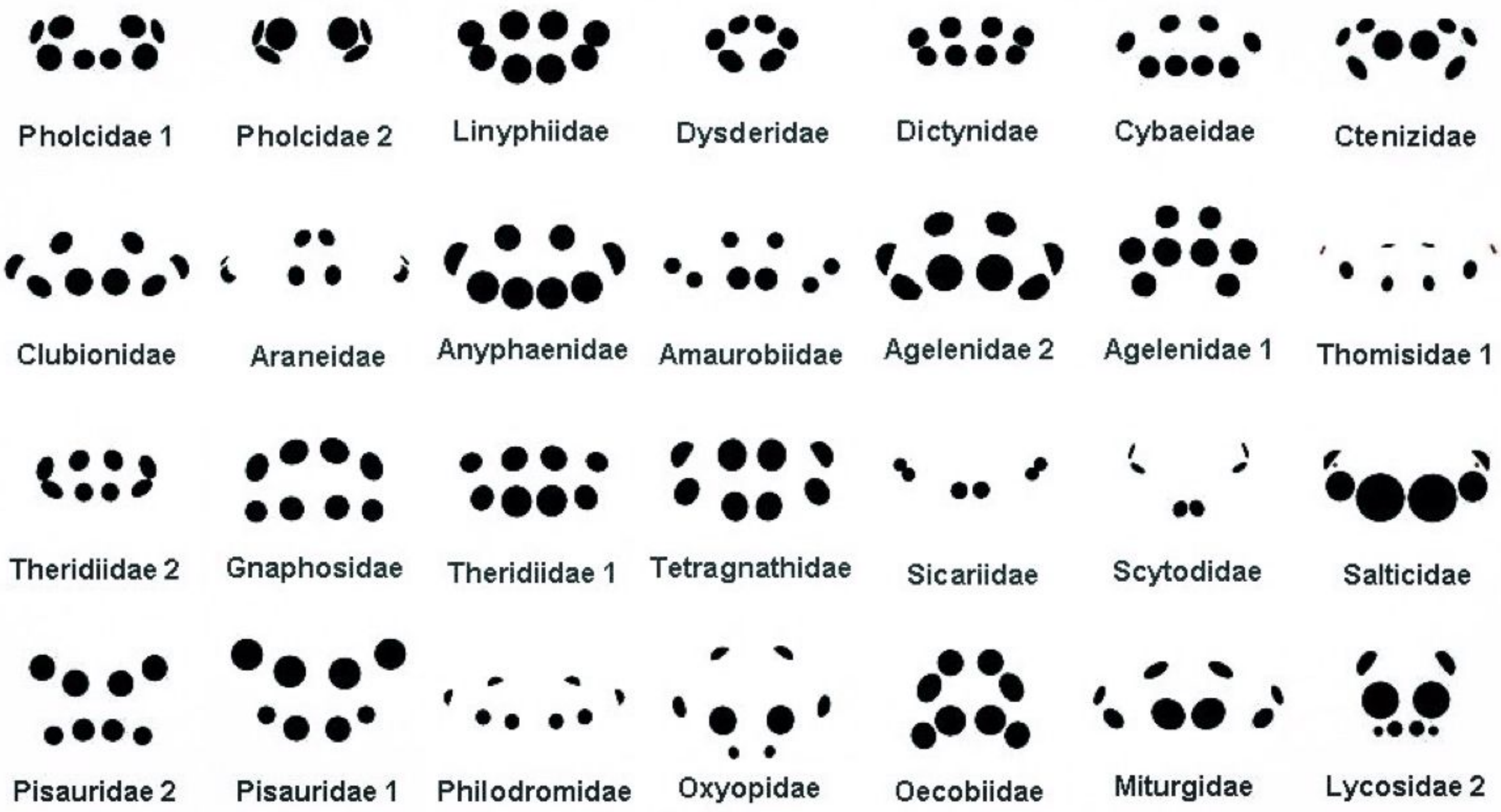
Fine-grained image recognition

[Xiu-Shen Wei]

The Categorization Spectrum



[R. Farrell]


Pholcidae 1
Pholcidae 2
Linyphiidae
Dysderidae
Dictynidae
Cybaeidae
Ctenizidae
Clubionidae
Araneidae
Anyphaenidae
Amaurobiidae
Agelenidae 2
Agelenidae 1
Thomisidae 1
Theridiidae 2
Gnaphosidae
Theridiidae 1
Tetragnathidae
Sicariidae
Scytodidae
Salticidae
Pisauridae 2
Pisauridae 1
Philodromidae
Oxyopidae
Oecobiidae
Miturgidae
Lycosidae 2

[R. Farrell]


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
(PATIENT LEATHER)
OPERA PUMP
(PATIENT LEATHER)
RIBBON PUMP
(PATIENT LEATHER)
BLACK TIE
BOOT
CHELSEA
CHUKKA
CAP TOE
WINGTIP
DRESS BOOTS
STRAP
SINGLE
DOUBLE
TRIPLE
MONK SHOES
PERFORATION
QUARTER
SEMI
FULL
LONGWING
BROGUEING


Pigalle
Lady Peep
Simple Pump
Decolette 554
Bianca
Bana
Batignolles
Daffodile
Fifi
Corneille
Highness
Ron Ron
Piou Piou
Very Prive
Lady Lynch

[R. Farrell]



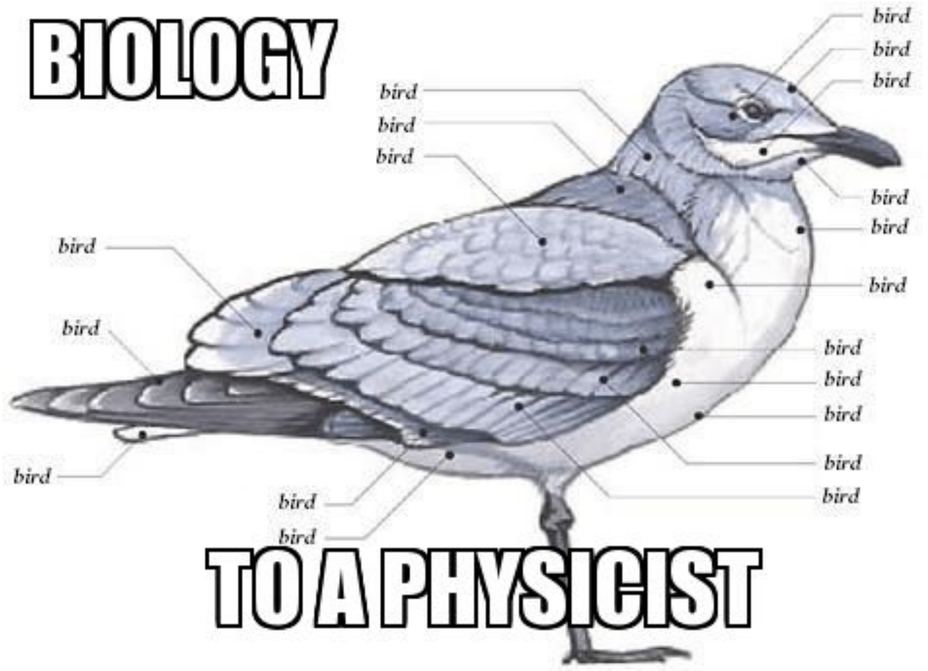




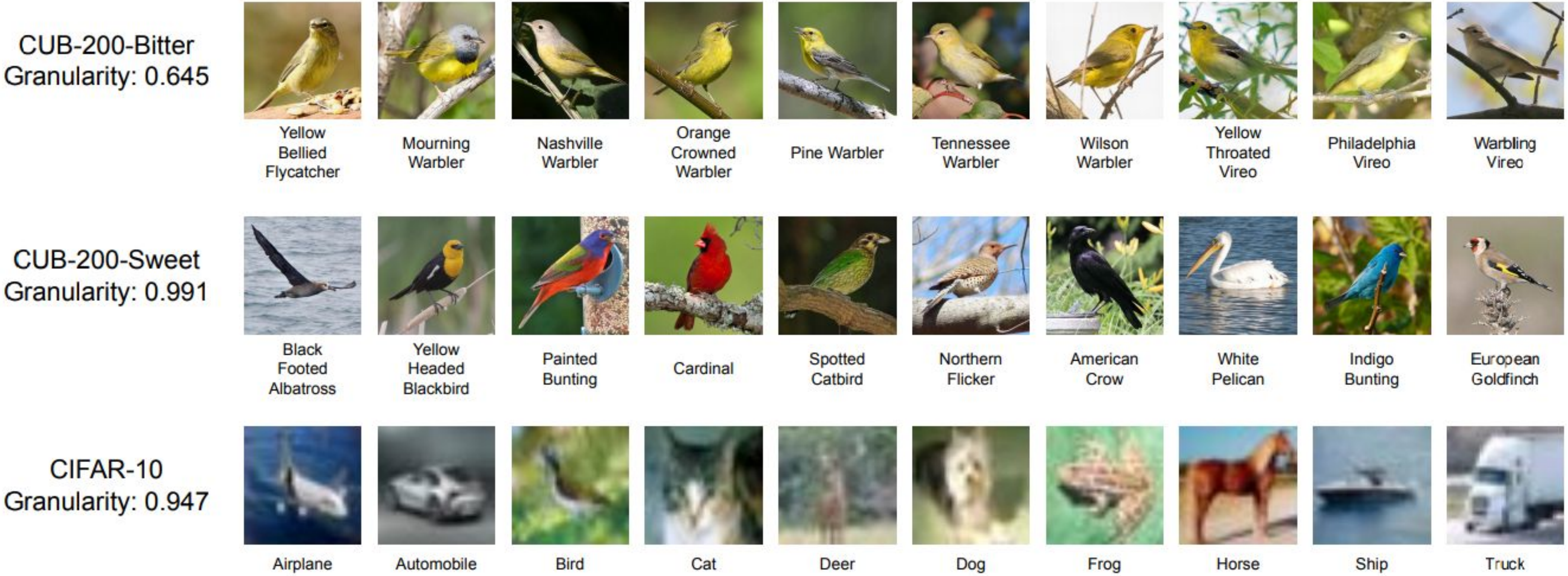



Granularity: human vs. machine perspective

- Dataset granularity depends on:
  - the ground truth labeling
  - the distance function
- Important to consider role of human expertise
- Some datasets are “fine grained in name only”
- Machine perspective: embedding vectors in high-dim. space



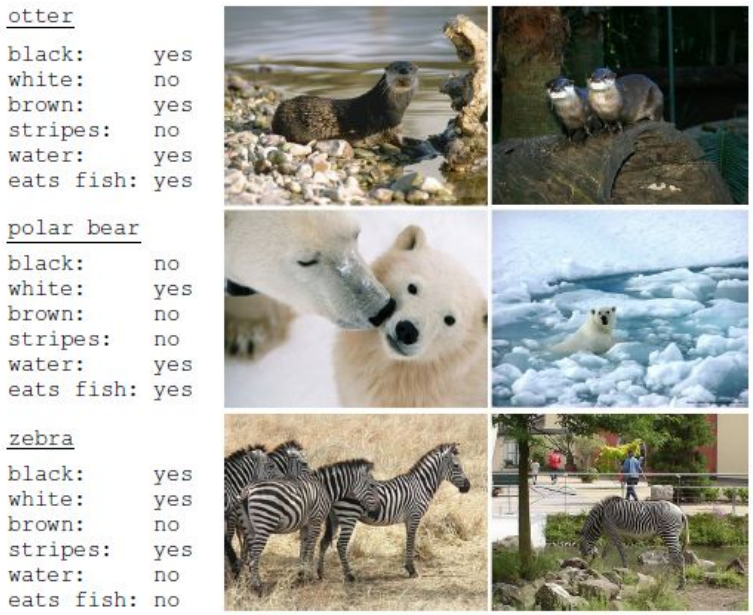
Quantifying Granularity



[Cui et al, arXiv 2019 https://arxiv.org/abs/1912.10154]

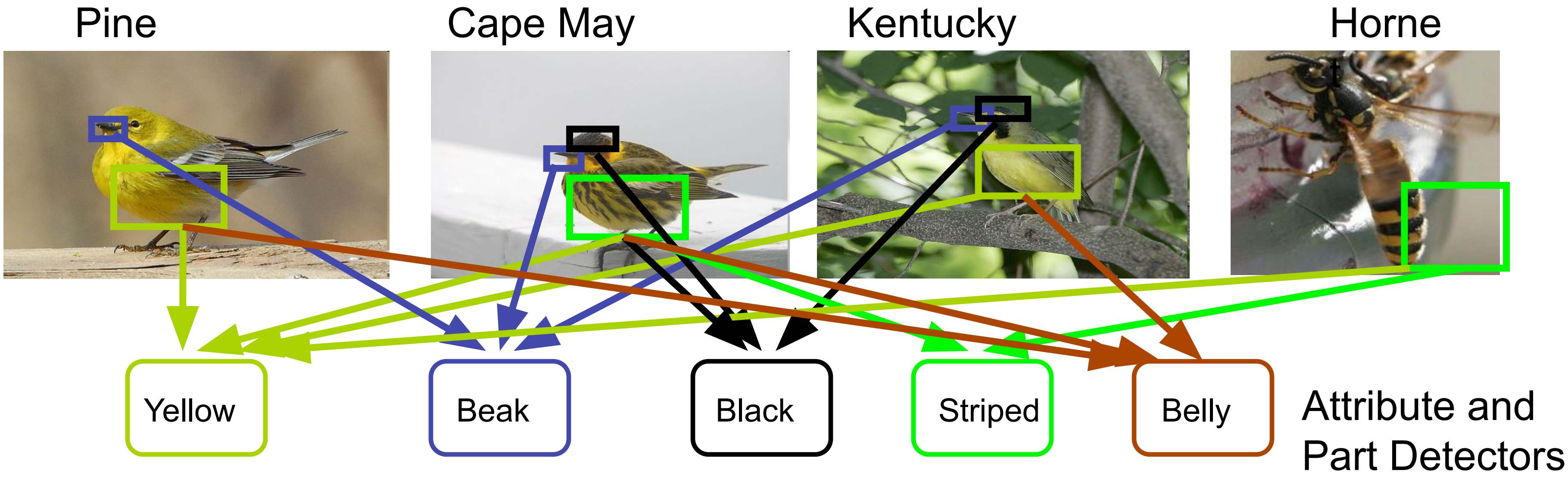
Attribute-Based Classification

- Train classifiers on attributes instead of objects
- Attributes are shared by different object classes
- Attributes provide the ingredients necessary to recognize each object class



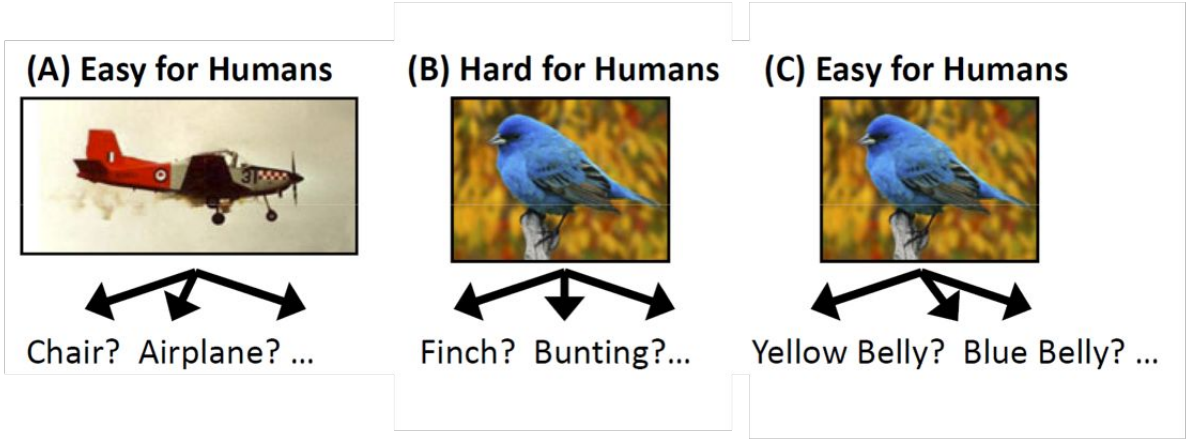
Lampert et al. 2009
Farhadi et al. 2009

Shared Parts and Attributes



Recognition with Humans in the Loop

Visual 20 Questions

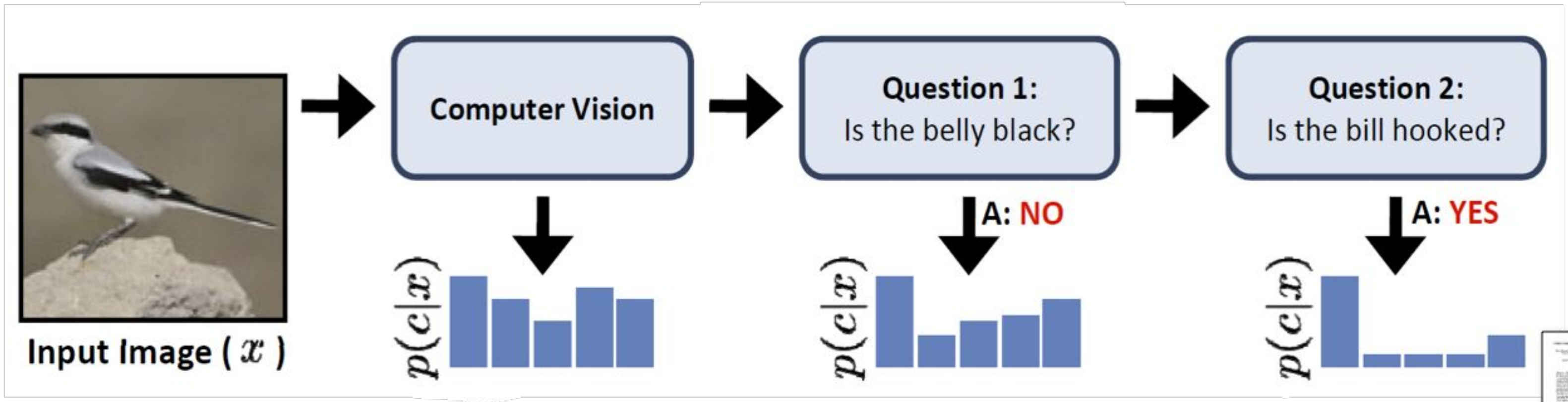


Visual Recognition with Humans in the Loop

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

ECCV 2010

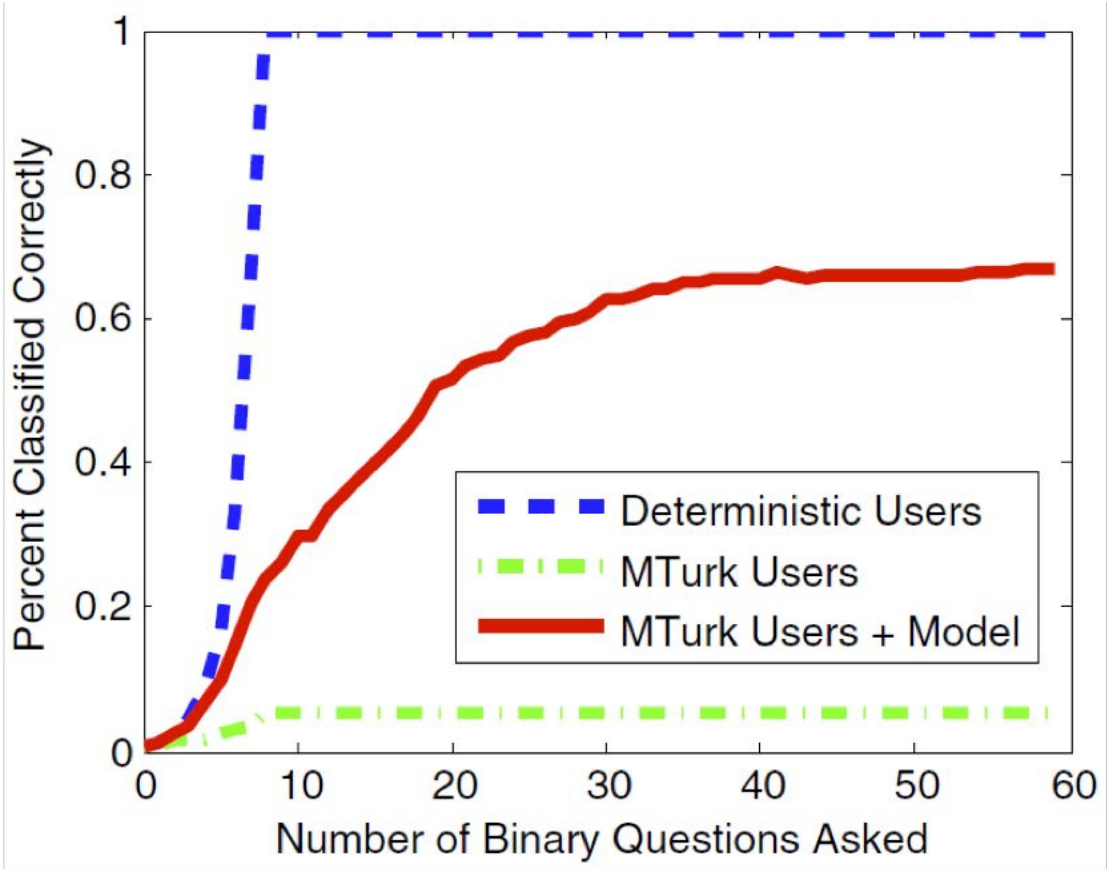
Visual 20 Questions



Visual Recognition with Humans in the Loop

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

ECCV 2010

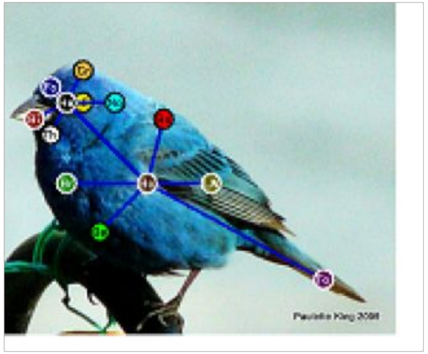


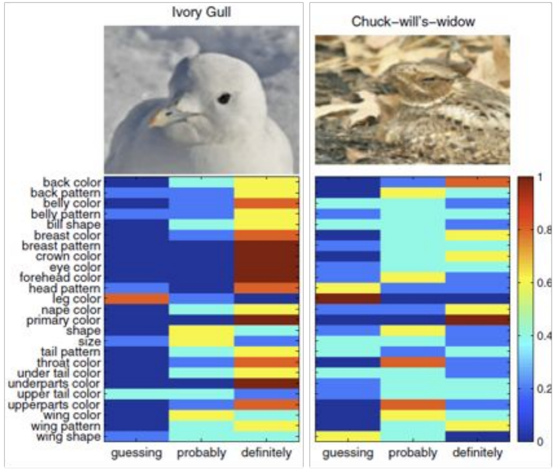
Visual Recognition with Humans in the Loop

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

antedeepluvian

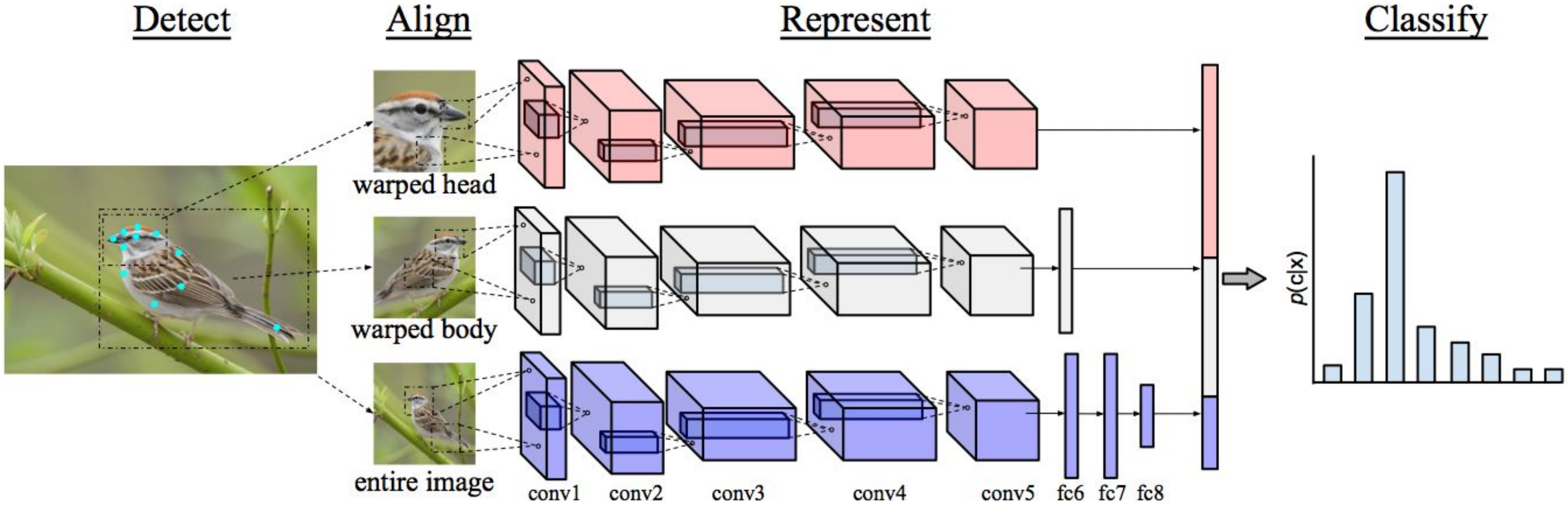
an·te·deep·lu·vi·an

ˌan(t)ēdēpˈlo͞ovēən/

adjective

1. before the flood of deep learning papers
2. “Histograms of vector quantized filter responses are *antedeepluvian* features.”

Pose Normalized Deep ConvNets

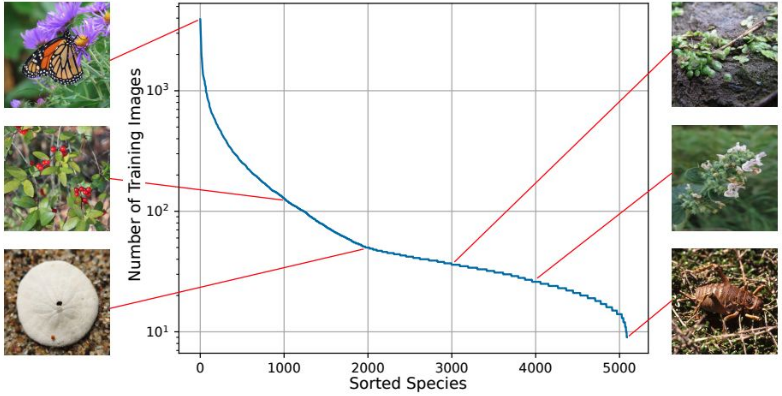


[Van Horn, Branson, Perona, Belongie BMVC 2014]

Categorization vs. Retrieval

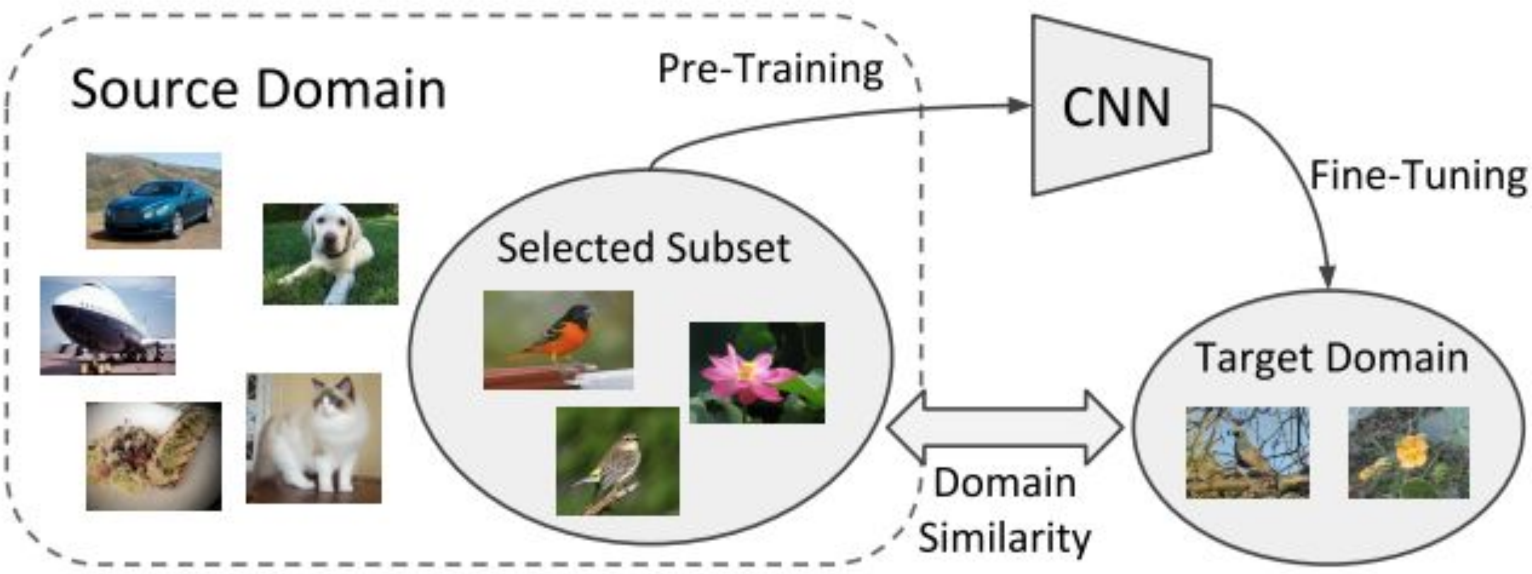
- Retrieval metrics, top k, psychometric factors
- Recognition via retrieval, and vice versa

Long-tailed fine-grained datasets



https://www.kaggle.com/c/inaturalist-challenge-at-fgvc-2017

Scaling to large numbers of domains

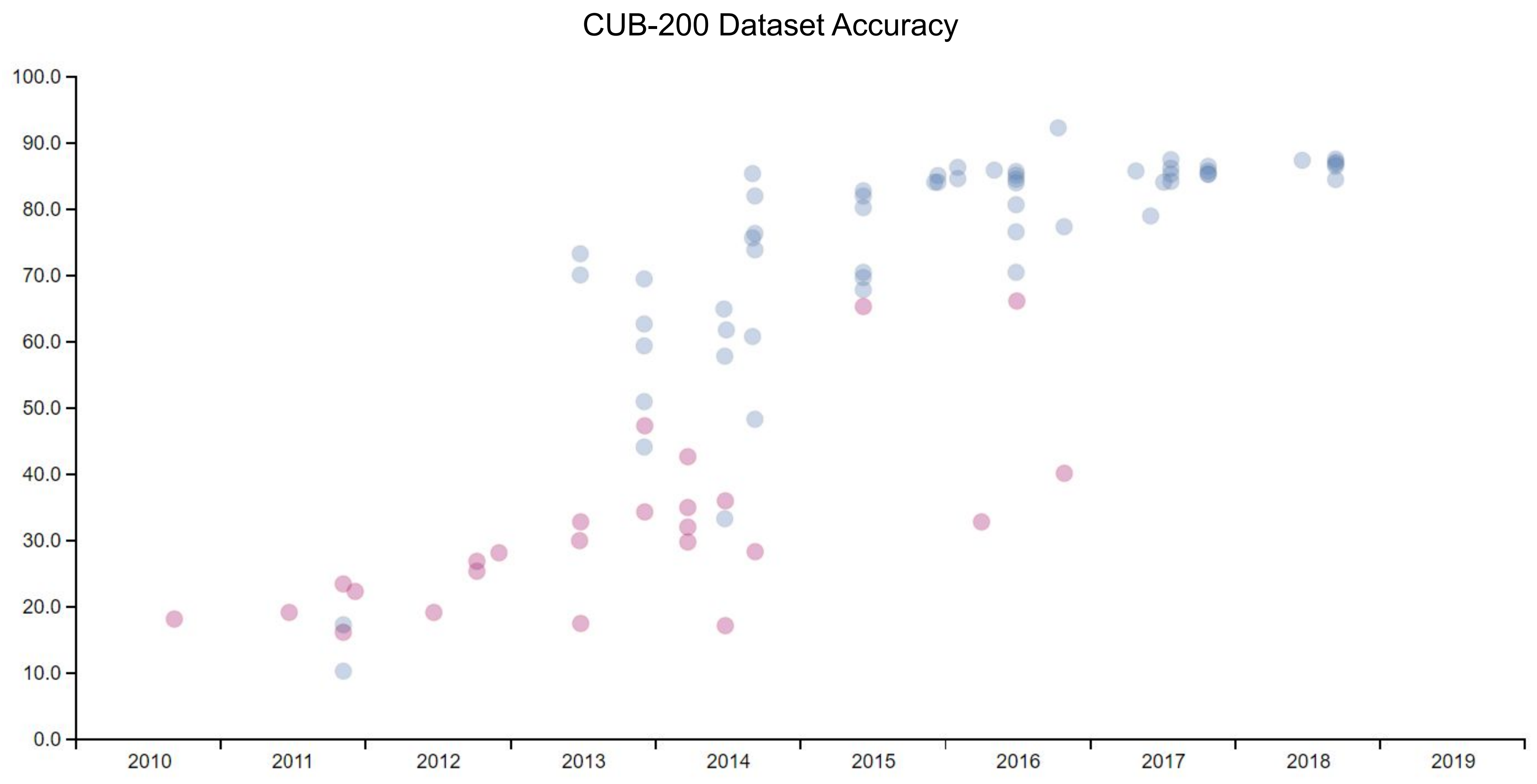


Fine-grained benchmark datasets

CUB200-2011

- 11,788 images, 200 fine-grained classes

[Catherine Wah et al., CNS-TR-2011-001, 2011]


CUB-200 Dataset Accuracy
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

Various real-world applications

Identify plant species from herbarium specimens.





https://www.kaggle.com/c/herbarium-2020-fgvc7/

Fine-grained benchmark datasets

Stanford Dogs

- 20,580 images
- 120 fine-grained classes

[Aditya Khosla et al., CVPR Workshop 2011]

Fine-grained benchmark datasets

Oxford Flowers

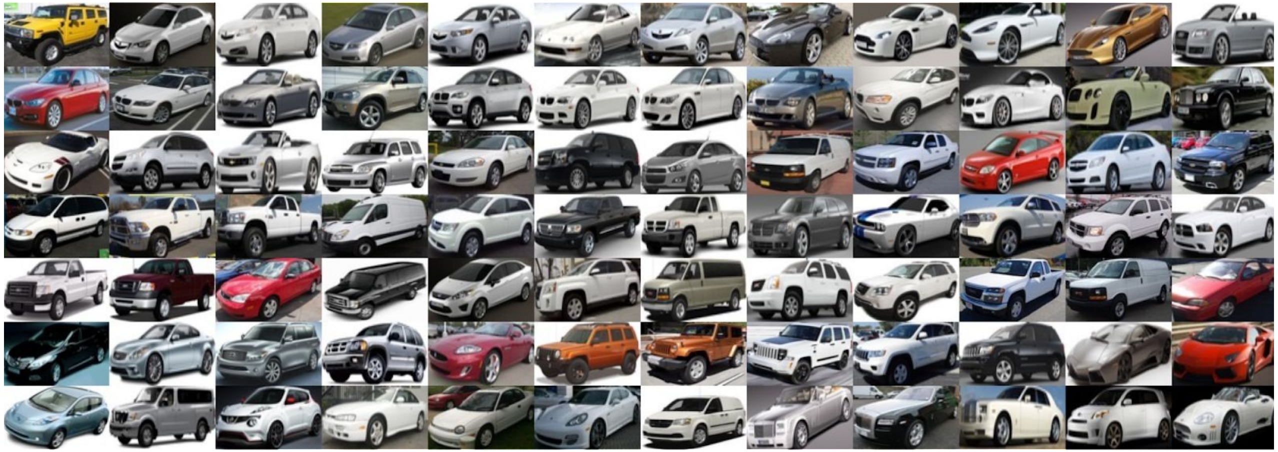
- 8,189 images, 102 fine-grained classes

[Maria-Elena Nilsback and Andrew Zisserman, CVGIP 2008]

Fine-grained benchmark datasets

Stanford Cars

- 16,185 images, 196 fine-grained classes

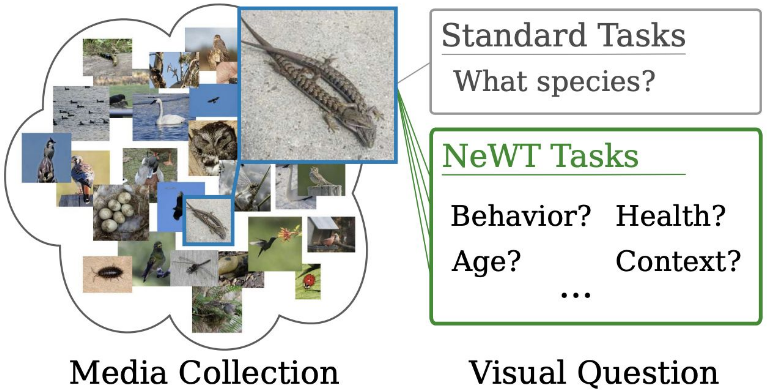


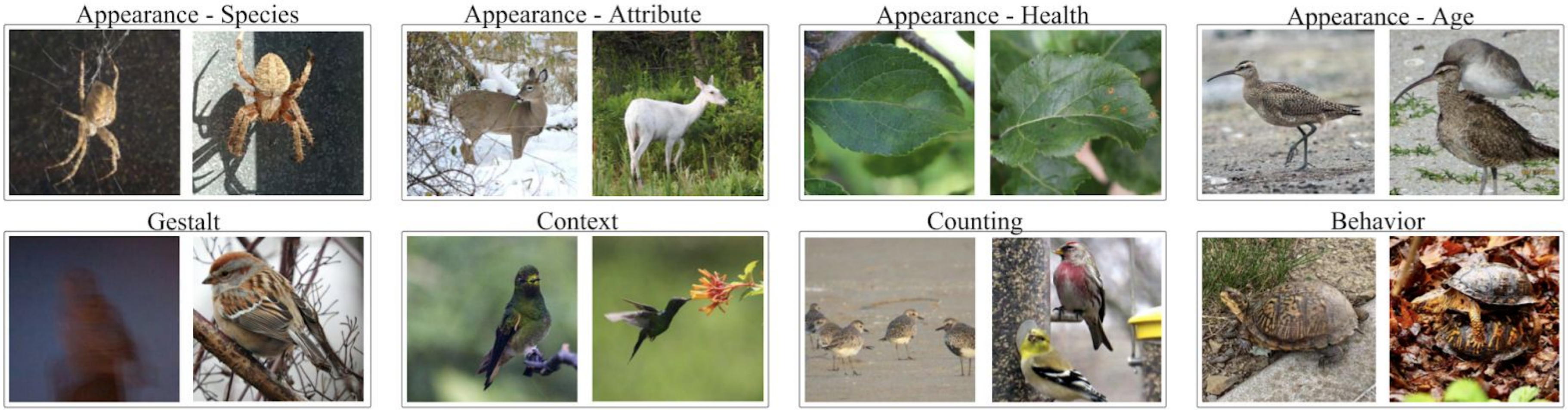
[Jonathan Krause et al., ICCV Workshop 2013]


675K IMAGES
iNaturalist Google
Large Scale Species Classification Competition
Results presented at CVPR 2017 Hawaii at 4th Workshop on Fine-Grained Visual Categorization
5000 SPECIES
https://github.com/visipedia/inat_comp
FGVC4
iNaturalist.org
Google
iNaturalist Competition 2018
8,000 specie
Long Tail Dis
FGVC5
iWildCam 2020
iFood 2019
iWildCam Competition 2018
FGVC5
Fungi Classification Challenge 2018
Flower Clas
Challenge 2018
1,000 species
FGVC5
PictureThis

Beyond fine grained image ID

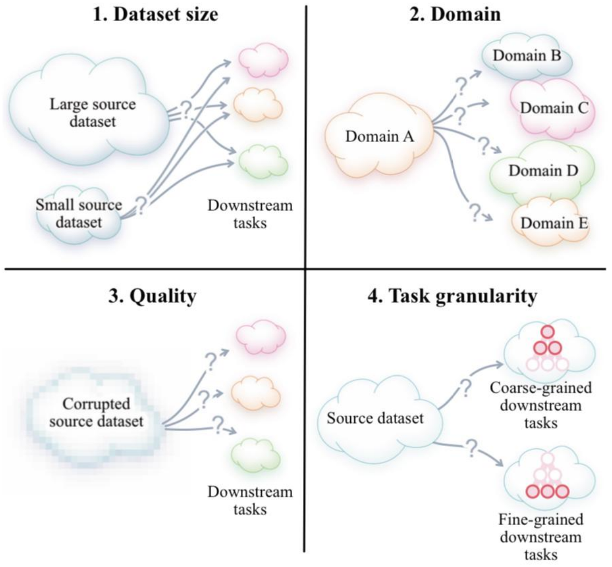
- Natural World Tasks (NeWT)
  Van Horn et al. CVPR 2021





Open Problems in Fine Grained Image Analysis

- Formal characterization of the problem
  - What, exactly, does “fine grained” mean?
- Data/label-efficient approaches
  - Targeted engagement with human expertise
- Self-supervision in the fine grained setting
  - Dataset augmentation for contrastive learning
- Beyond static images
  - Multimodal/video+audio
- Synthetic and augmented data
  - Devil in the details



[E. Cole et al. CVPR 2022]