iNaturalist Competition 2018

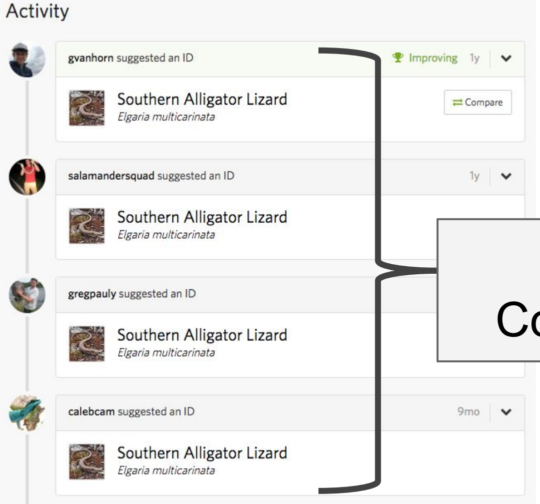
Grant Van Horn
Oisin Mac Aodha

iNat 2018 Dataset Overview

iNaturalist

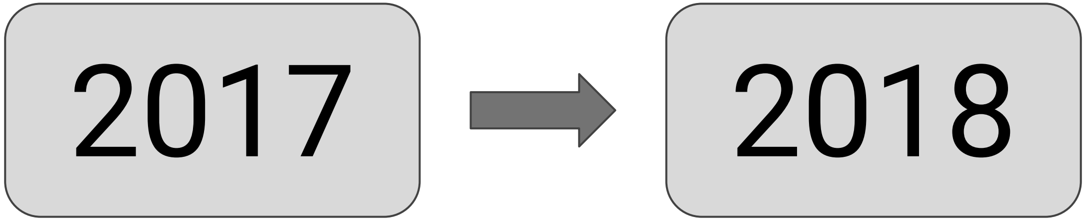
Southern Alligator Lizard

Research Grade



> ⅔ Consensus

Updates from 2017



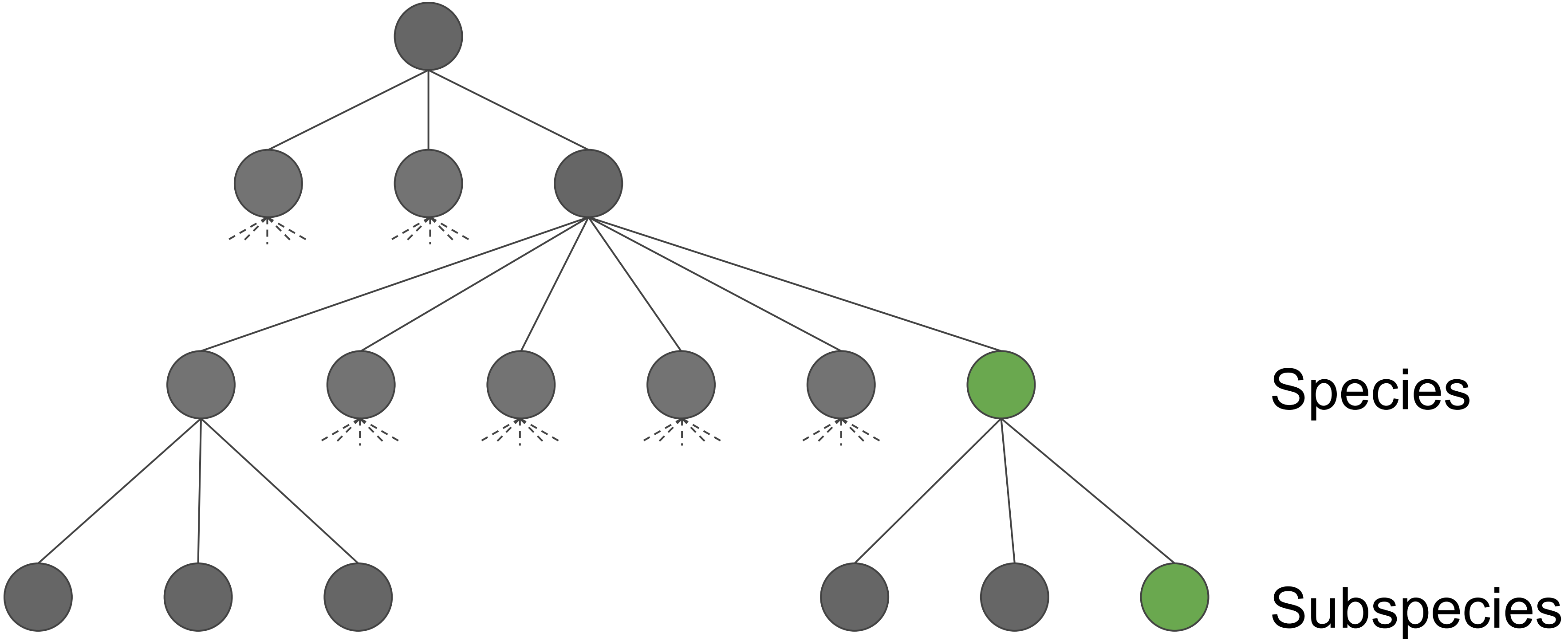
2017: 13 Super Categories

2018: 14 Super Categories

Star Slime (Nostoc commune)

ahuereca
ob: 12066494

2017: Sloppy Taxonomy Processing

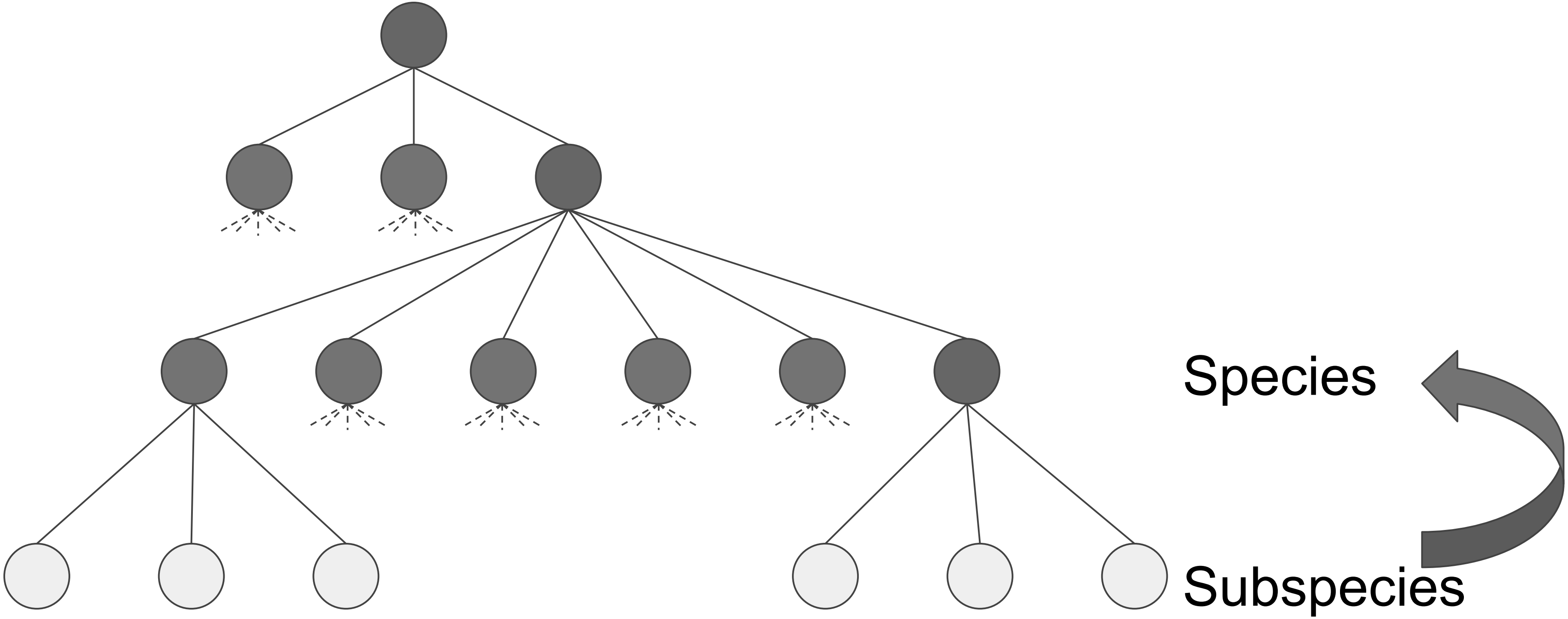


2017: Species & Subspecies

Callisaurus draconoides

Callisaurus draconoides rhodostictus

2018: Roll up to Species

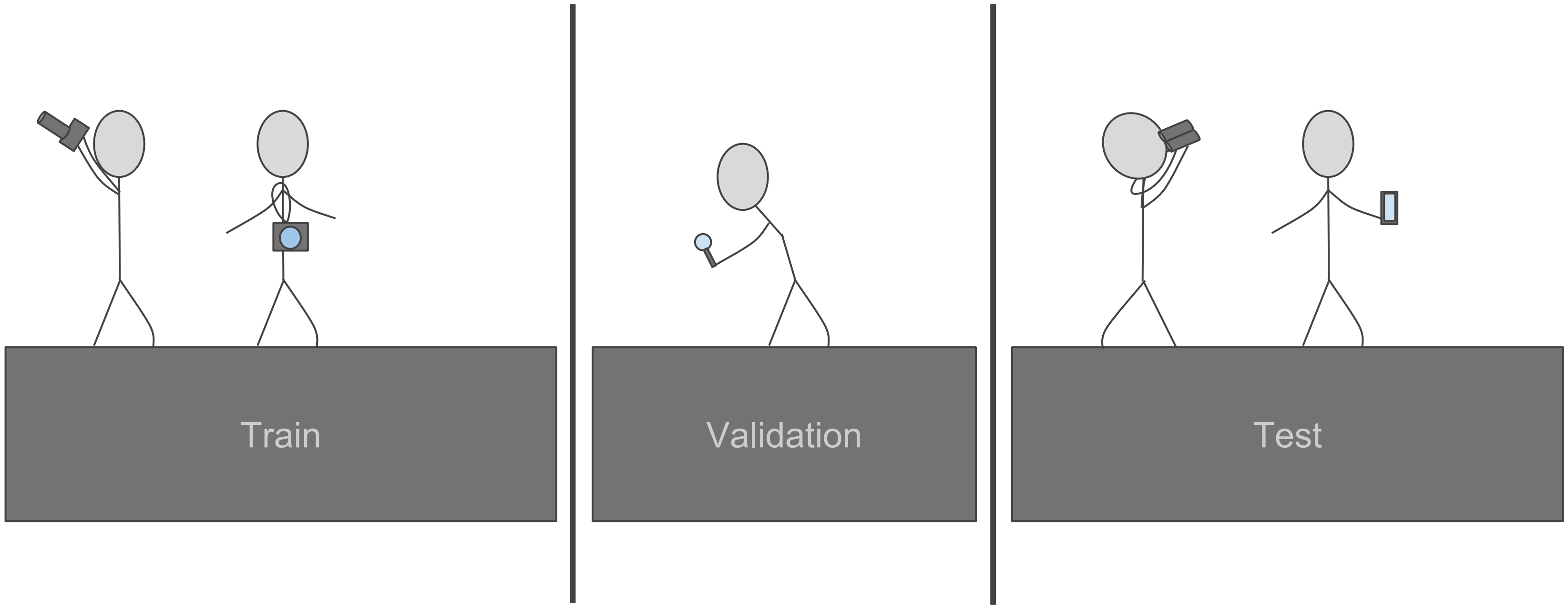


2018: Taxonomy Included

```
category{
  "id" : int,
  "name" : str,
  "supercategory" : str,
  "kingdom" : str,
  "phylum" : str,
  "class" : str,
  "order" : str,
  "family" : str,
  "genus" : str
}
```

2017: Train, Validation, and Test Splits

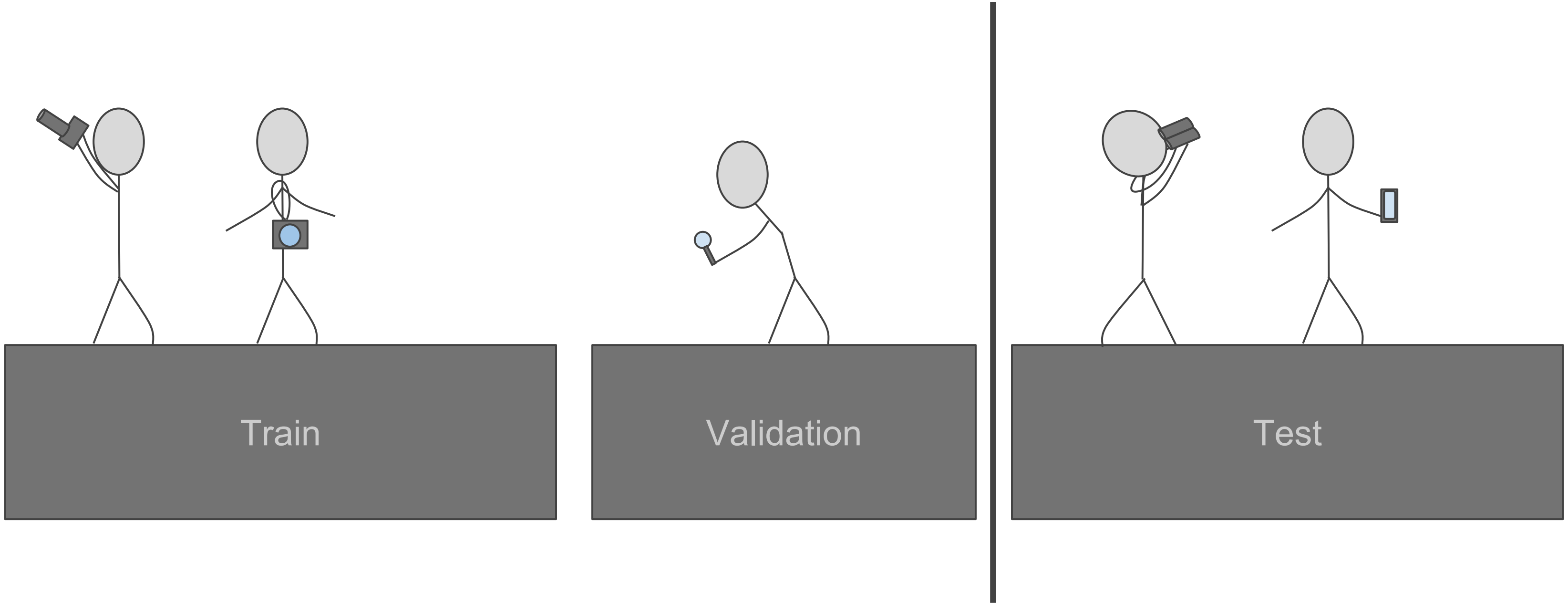
Unique Photographers in Each Split



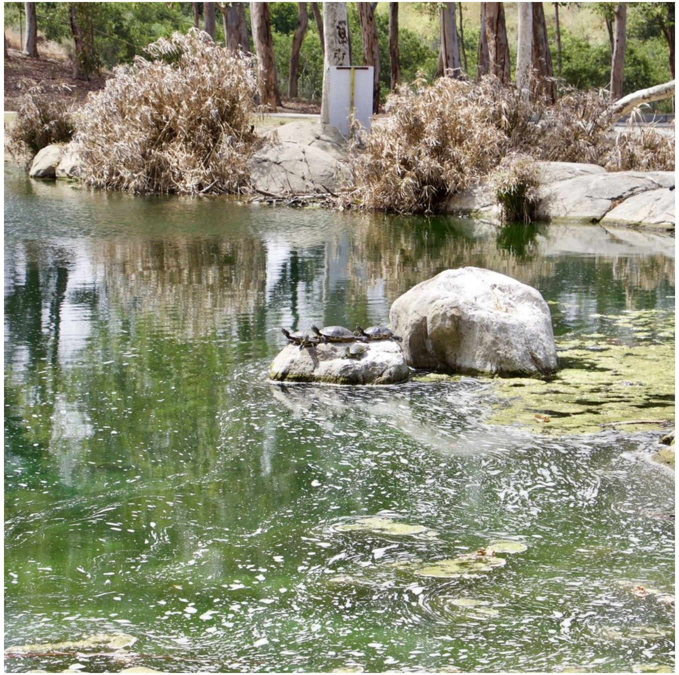
Isolate Photographers to a Single Split

2018: Train, Validation, and Test Splits

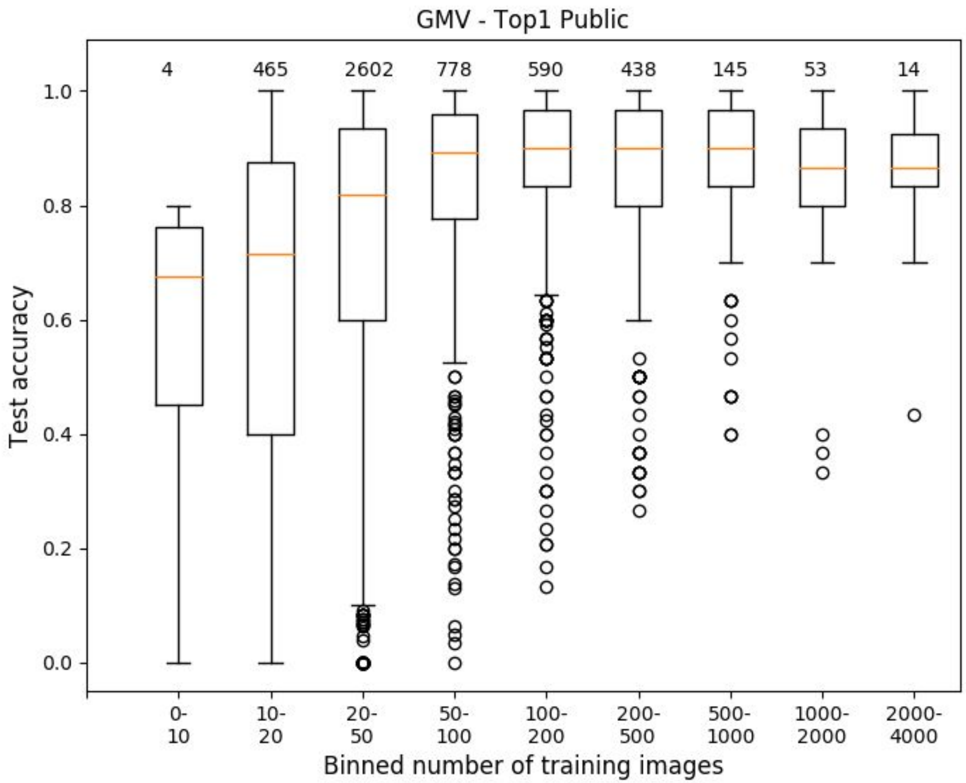
Unique Photographers in the Test Split



2018: 1 Photo per Observation



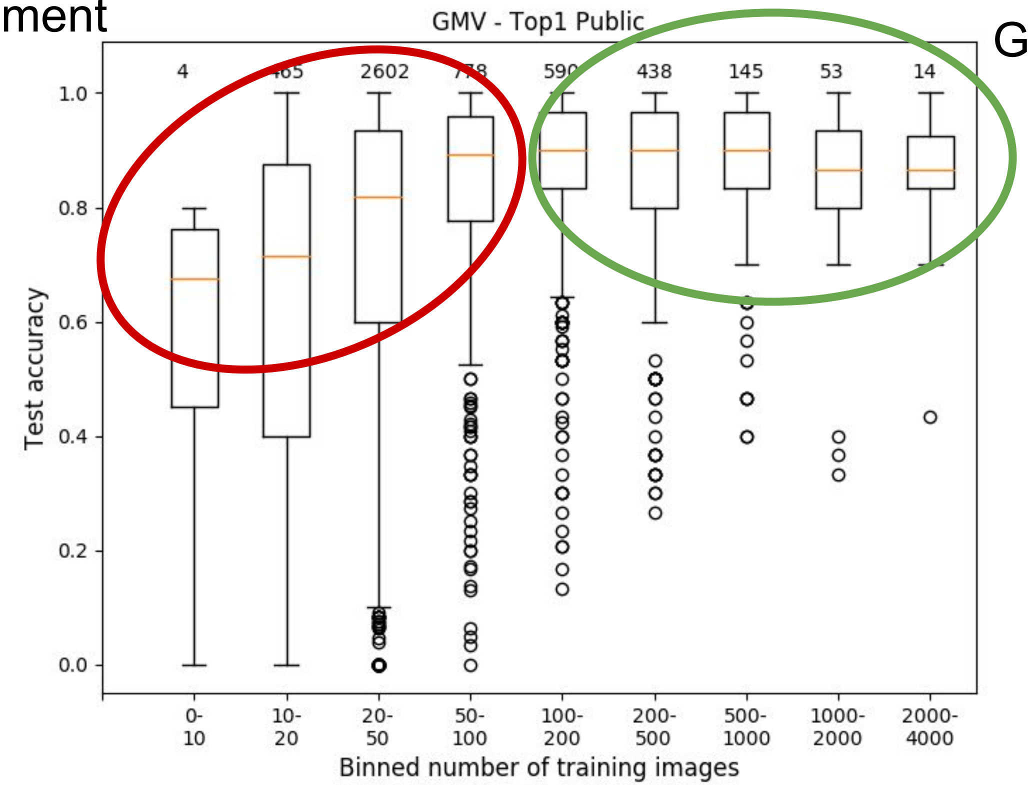
2017: Top 1 Accuracy of the Winning Team



2017: Top 1 Accuracy of the Winning Team

Needs Some Improvement

Looking Pretty Good



2018: Training Image Bounds

2017

3000+ Images

8 Images Min

2018

1000 Images Max

2 Images Min

2018: Validation Set

Overview

- Validation set split: 90% validation → val; 10% validation → minival.
- Inception models trained on train + val (666k images), evaluated on minival (9.6k images).
- Using higher resolution image improves the performance.
- To deal with label imbalance, fine-tune on val further with small learning rate after training.
  - Training learns good feature, fine-tuning on validation gives the network information on the label distribution.

Slide from Yin Cui

2018: Validation Set

Overview

- Validation set split: 90% validation → val; 10% validation → minival.
- Inception models trained on train + val (666k images), evaluated on minival (9.6k images).
- Using hig
- To deal … all learning rate after tra
    - Trai … nformation on the label

Cap the validation set at 3 images per species.

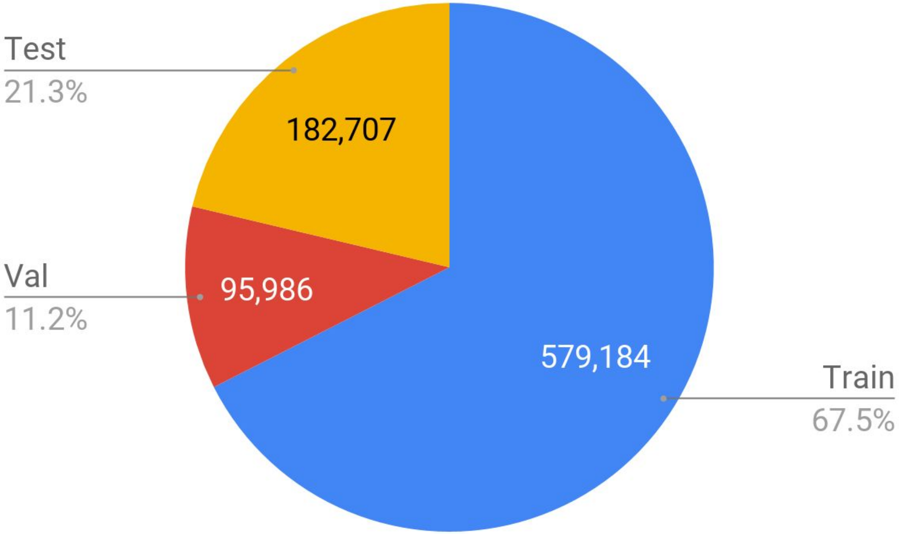
Slide from Yin Cui

Dataset Stats

2017

Number of Classes: 5,089

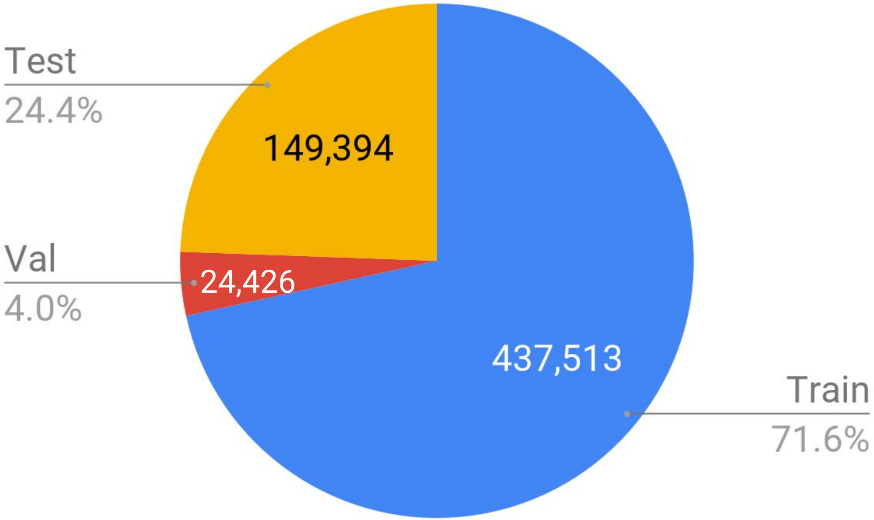
Number of Images: 857,877



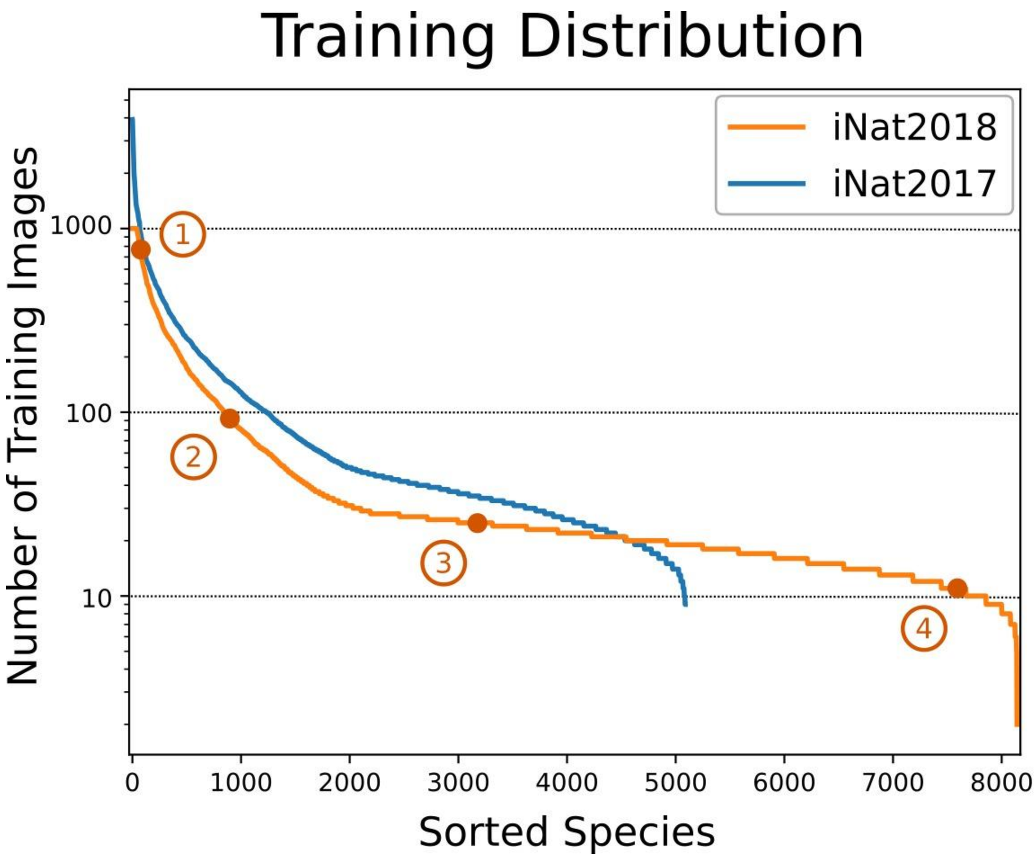
2018

Number of Classes: 8,142

Number of Images: 611,333



Training Image Distribution



Dataset Hosting

Big thanks to Yang Song!

North America

Europe

Asia

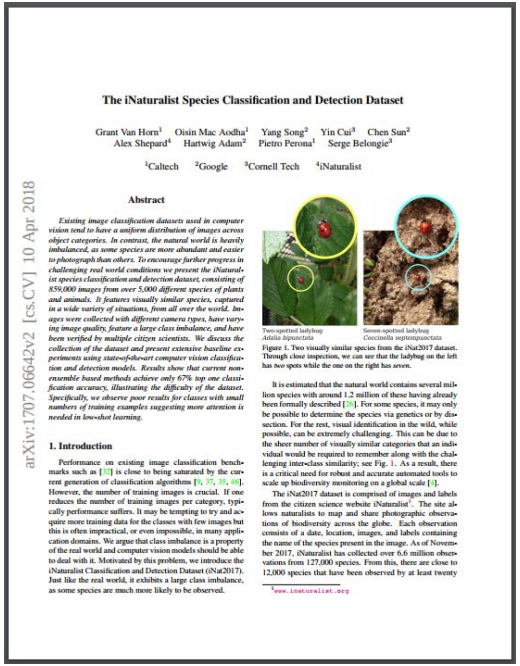
Google

More Details

CVPR 2018 Paper

https://arxiv.org/abs/1707.06642

- Detailed overview of 2017 dataset
- Additional experiments on classification and with bounding boxes



The iNaturalist Species Classification and Detection Dataset

Grant Van Horn[1] Oisin Mac Aodha[1] Yang Song[2] Yin Cui[3] Chen Sun[2]
Alex Shepard[4] Hartwig Adam[2] Pietro Perona[1] Serge Belongie[3]

[1]Caltech [2]Google [3]Cornell Tech [4]iNaturalist

Abstract

Existing image classification datasets used in computer vision tend to have a uniform distribution of images across object categories. In contrast, the natural world is heavily imbalanced, as some species are more abundant and easier to photograph than others. To encourage further progress in challenging real world conditions we present the iNaturalist species classification and detection dataset, consisting of 859,000 images from over 5,000 different species of plants and animals. It features visually similar species, captured in a wide variety of situations, from all over the world. Images were collected with different camera types, have varying image quality, feature a large class imbalance, and have been verified by multiple citizen scientists. We discuss the collection of the dataset and present extensive baseline experiments using state-of-the-art computer vision classification and detection models. Results show that current non-ensemble based methods achieve only 67% top one classification accuracy, illustrating the difficulty of the dataset. Specifically, we observe poor results for classes with small numbers of training examples suggesting more attention is needed in low-shot learning.

1. Introduction

Performance on existing image classification benchmarks such as [32] is close to being saturated by the current generation of classification algorithms [9, 37, 35, 46]. However, the number of training images is crucial. If one reduces the number of training images per category, typically performance suffers. It may be tempting to try and acquire more training data for the classes with few images but this is often impractical, or even impossible, in many application domains. We argue that class imbalance is a property of the real world and computer vision models should be able to deal with it. Motivated by this problem, we introduce the iNaturalist Classification and Detection Dataset (iNat2017). Just like the real world, it exhibits a large class imbalance, as some species are much more likely to be observed.

Two-spotted ladybug *Adalia bipunctata* — Seven-spotted ladybug *Coccinella septempunctata*

Figure 1. Two visually similar species from the iNat2017 dataset. Through close inspection, we can see that the ladybug on the left has *two* spots while the one on the right has *seven*.

It is estimated that the natural world contains several million species with around 1.2 million of these having already been formally described [26]. For some species, it may only be possible to determine the species via genetics or by dissection. For the rest, visual identification in the wild, while possible, can be extremely challenging. This can be due to the sheer number of visually similar categories that an individual would be required to remember along with the challenging inter-class similarity; see Fig. 1. As a result, there is a critical need for robust and accurate automated tools to scale up biodiversity monitoring on a global scale [4].

The iNat2017 dataset is comprised of images and labels from the citizen science website iNaturalist[1]. The site allows naturalists to map and share photographic observations of biodiversity across the globe. Each observation consists of a date, location, images, and labels containing the name of the species present in the image. As of November 2017, iNaturalist has collected over 6.6 million observations from 127,000 species. From this, there are close to 12,000 species that have been observed by at least twenty

[1] www.inaturalist.org

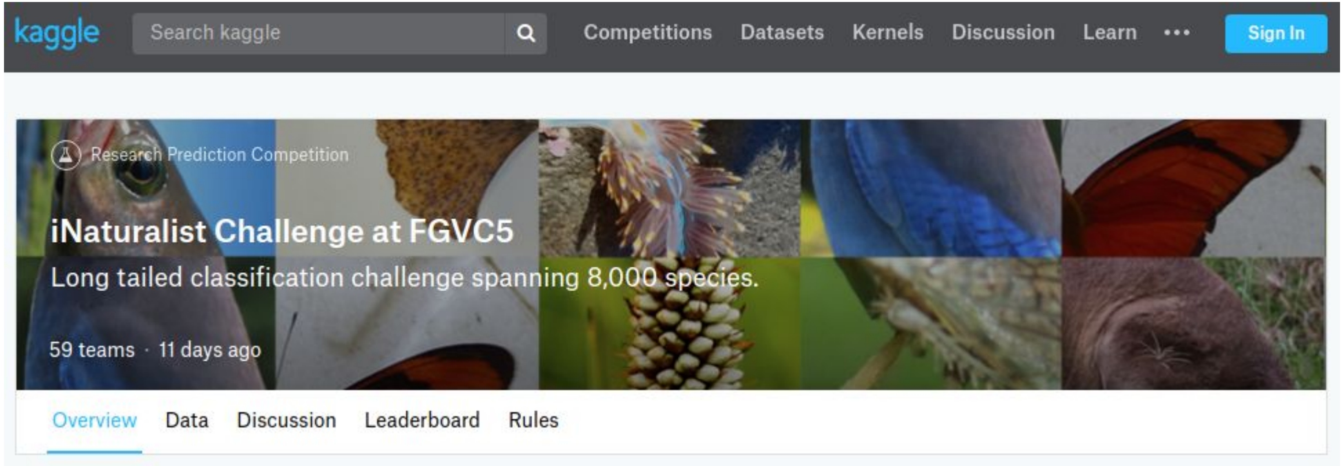
Competition Results

Competition

Hosted on Kaggle

Submission server open from Feb to June 4th

40 Teams with full submissions



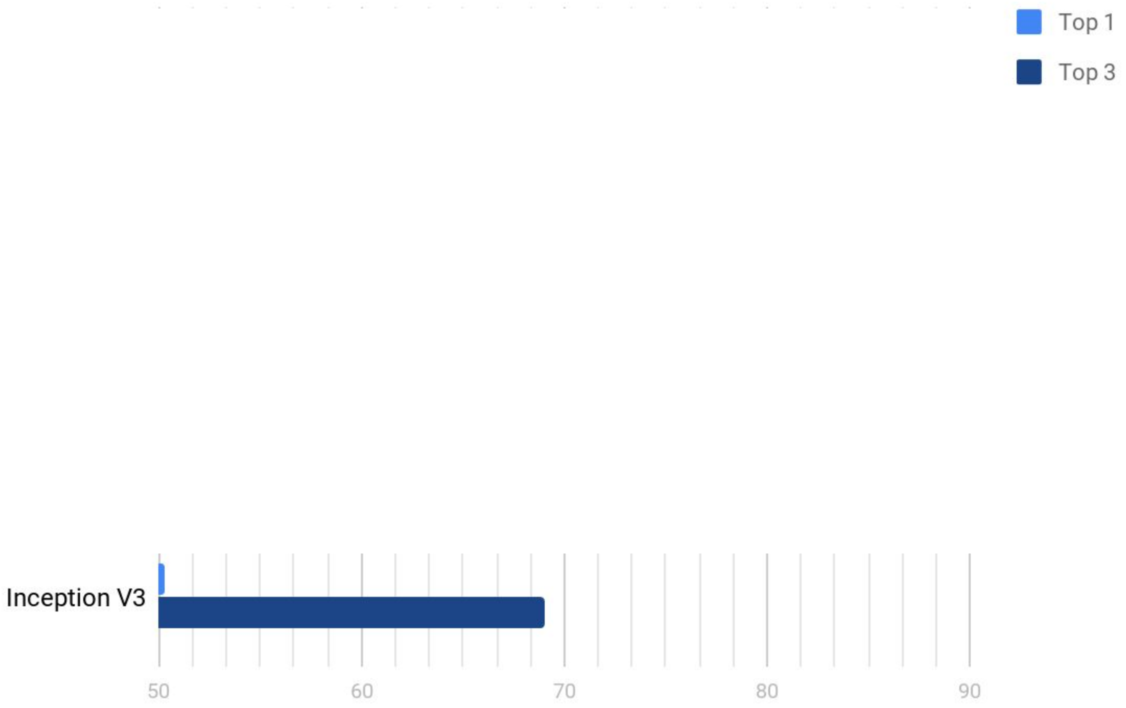
Final Leaderboard - Top 3 Error Private Split

#	△pub	Team Name	Kernel	Team Members	Score	Entries	Last
1	—	DLUT_VLG (Dalian University ...			0.12858	133	11d
2	—	Deep Learning Analytics			0.13981	93	11d
3	—	fadivugibs			0.14618	79	11d
4	—	CMP			0.16076	14	11d
5	▲1	fISHpAM			0.16892	3	23d
6	▼1	traveler			0.16988	30	11d
7	—	yen			0.17201	20	11d
8	—	Shuang			0.18357	15	11d
9	—	Mr.M			0.20092	7	13d
10	—	Dequan Wang			0.20814	10	2mo
11	—	Du Ang			0.21469	4	15d
12	—	pingjing			0.21764	8	18d
13	—	FiberHome AI R&D Center			0.22311	26	1mo
14	—	1111			0.23253	37	2mo
15	—	w______g			0.23417	24	3mo
16	—	Val An			0.23899	8	15d
17	—	HongX			0.24833	74	13d
18	—	Ziyi Lin			0.25301	6	2mo
19	—	杨威			0.26707	6	3mo
20	—	Igor Krashenyi			0.26877	13	3mo

Baselines (single crop)

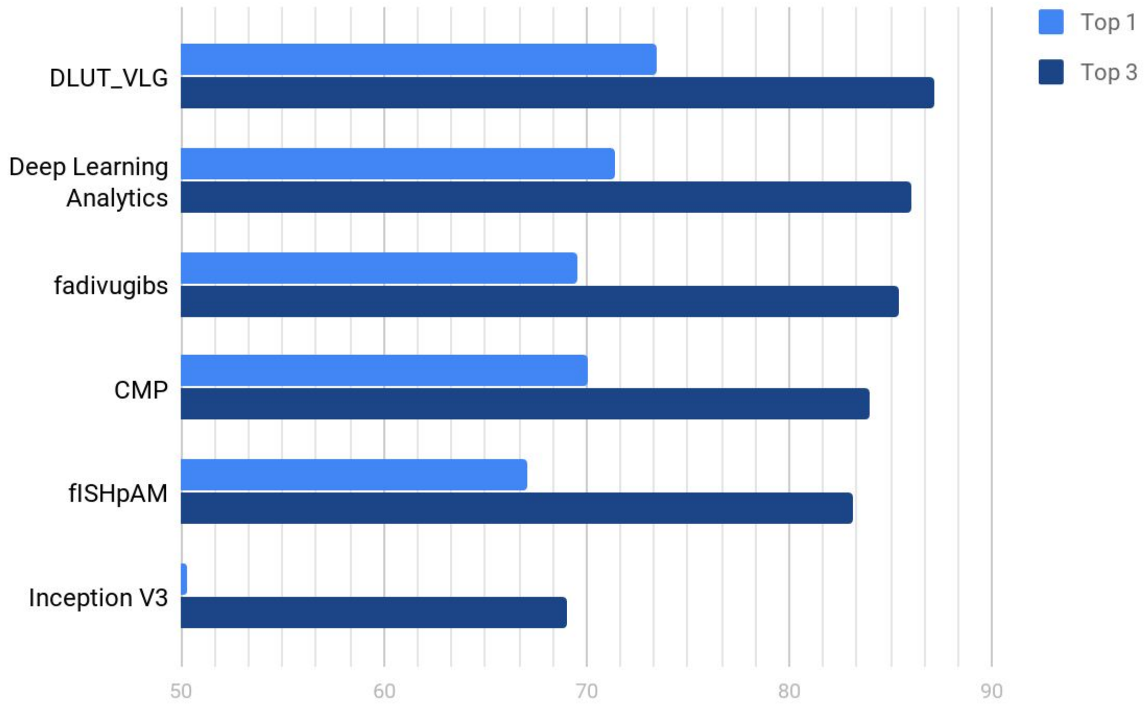
InceptionV3 0.3096

Accuracy



Results reported on private split

Accuracy



Results reported on private split

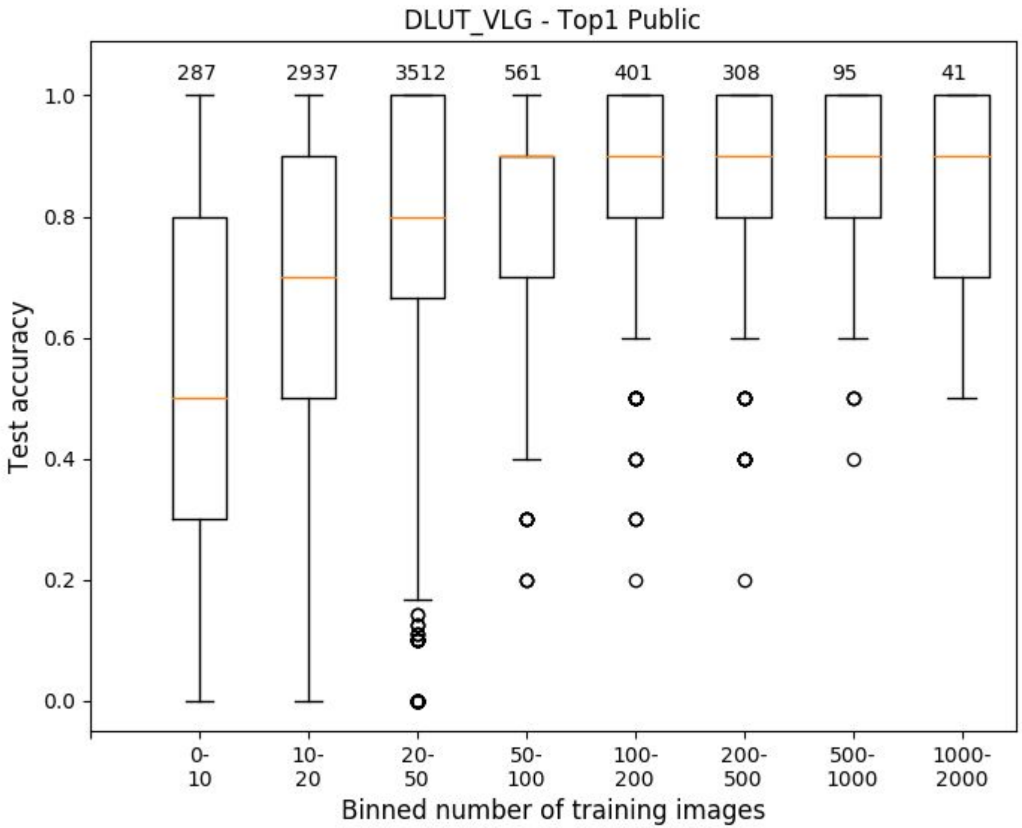
Per Super Category Averaged Top1 Accuracy - DLUT_VLG

Average accuracy for each class in each super-class

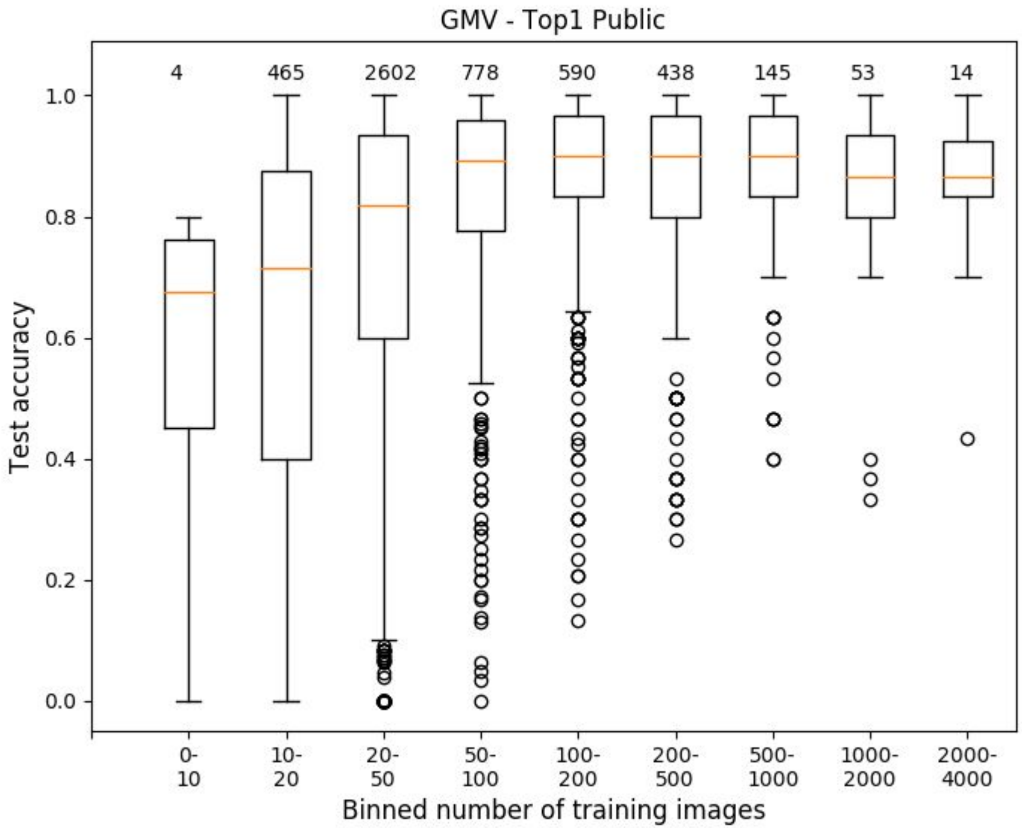
Super Cat	Num Classes	Accuracy
Bacteria	1	100.0
Protozoa	4	85.29
Insecta	2,031	79.88
Aves	1,258	76.07
Plantae	2,917	73.53
Chromista	25	69.08
Animalia	178	67.95
Arachnida	114	65.67
Fungi	321	64.35
Actinopterygii	369	64.12
Mammalia	234	62.82
Reptilia	284	62.03
Mollusca	262	61.87
Amphibia	144	54.97

Results reported on public split

Accuracy by Number Training Examples - DLUT_VLG



Accuracy by Number Training Examples - 2017 Best Team



Commonly Confused Classes

Aphonopelma hentzi

Aphonopelma chalcodes

Commonly Confused Classes

Acronicta lobeliae

Acronicta hasta

Commonly confused classes Top1 Public for DLUT_VLG

Commonly Confused Classes

Papilio thoas

Papilio rumiko

Commonly confused classes Top1 Public for DLUT_VLG

Commonly Confused Classes

Sphyraena jello

Sphyraena barracuda

Commonly confused classes Top1 Public for DLUT_VLG

Commonly Confused Classes

Lepus alleni

Lepus californicus

Commonly confused classes Top1 Public for DLUT_VLG

Winning Teams

DLUT_VLG

Dalian University of Technology

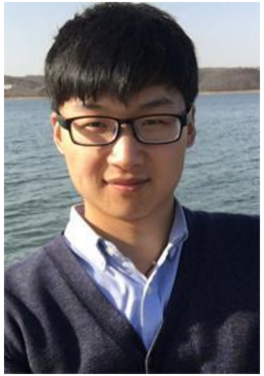
iNaturalist Challenge 2018 **1st Place**

Team: DLUT_VLG

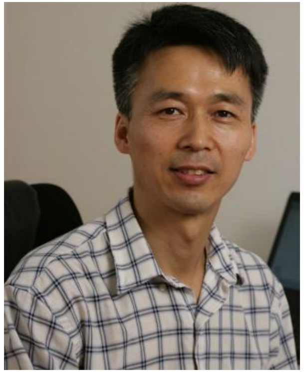
Dalian University of Technology, China

Shuyu Ge
MS Student

Qiule Sun
MS Student



Jiangtao Xie
BS Student



Prof. Peihua Li
Team Leader

iNat2018 Challenge Result

DLUT_VLG performs relatively 8% in top-3 error better than 2nd place

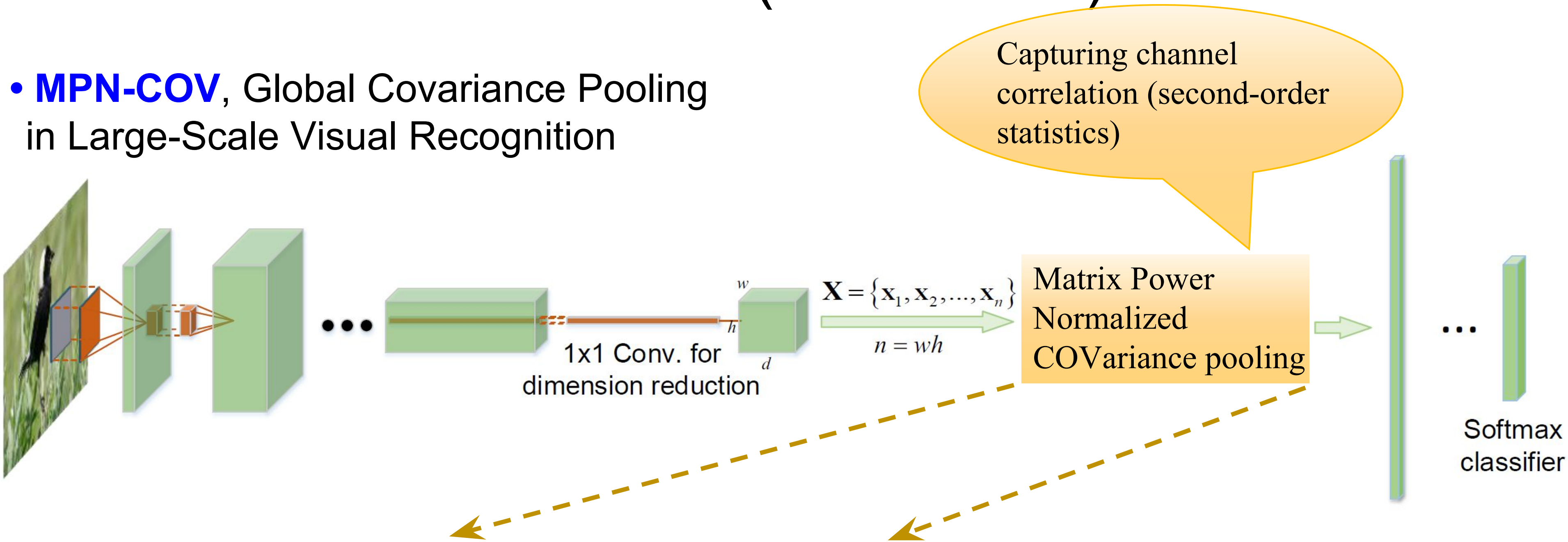
Public Leaderboard   Private Leaderboard

The private leaderboard is calculated with approximately 50% of the test data.   Refresh

#	△pub	Team Name	Kernel	Team Members	Score	Entries	Last
1	—	DLUT_VLG (Dalian University ...			0.12858	133	11d
2	—	Deep Learning Analytics			0.13981	93	11d
3	—	fadivugibs			0.14618	79	11d
4	—	CMP			0.16076	14	11d
5	▲1	fISHpAM			0.16892	3	23d
6	▼1	traveler			0.16988	30	11d
7	—	yen			0.17201	20	11d

Backbone of Our Method (MPN-COV)

- **MPN-COV**, Global Covariance Pooling in Large-Scale Visual Recognition



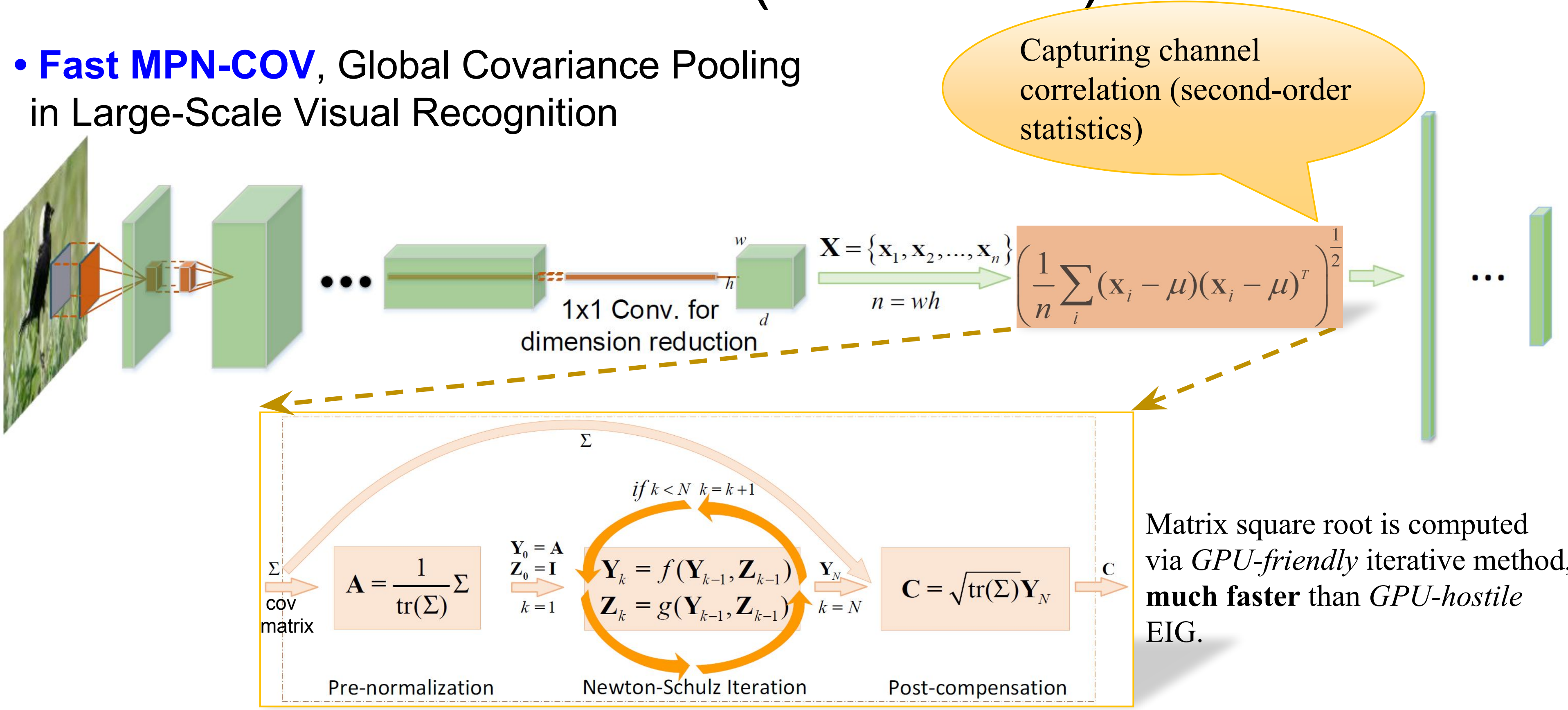
$$\left(\frac{1}{n}\sum_i (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T\right)^{\alpha}$$

where $\alpha = 1/2$ performs best, μ is mean vector of $\mathbf{x}_i$; matrix power is computed via EIGen-decomposition

[MPN-COV] Peihua Li, Jiangtao Xie, Qilong Wang and Wangmeng Zuo. Is Second-order Information Helpful for Large-scale Visual Recognition? In *ICCV*, 2017.

Backbone of Our Method (MPN-COV)

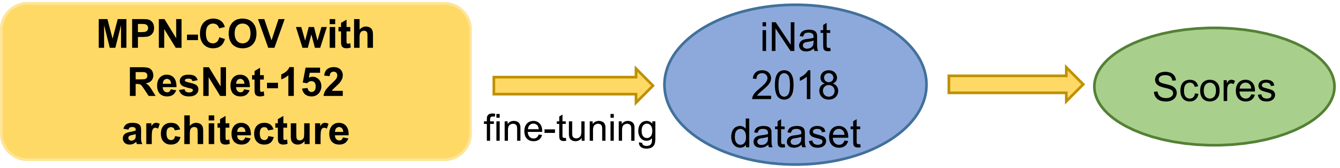
- **Fast MPN-COV**, Global Covariance Pooling in Large-Scale Visual Recognition



Matrix square root is computed via *GPU-friendly* iterative method, **much faster** than *GPU-hostile* EIG.

[Fast MPN-COV] Peihua Li, Jiangtao Xie, Qilong Wang and Zilin Gao. Towards Faster Training of Global Covariance Pooling Networks by Iterative Matrix Square Root Normalization. In *CVPR*, 2018.

Experiment



MPN-COV with ResNet-152 architecture, fine-tuned on iNat 2018

- Implementation
- Three useful tricks for performance boost
    - — Exploit higher resolution images
    - — Deal with long tailed distribution
    - — Usage of iNat2017 dataset
- Results

[MPN-COV] Peihua Li, Jiangtao Xie, Qilong Wang and Wangmeng Zuo. Is Second-order Information Helpful for Large-scale Visual Recognition? In *ICCV*, 2017.

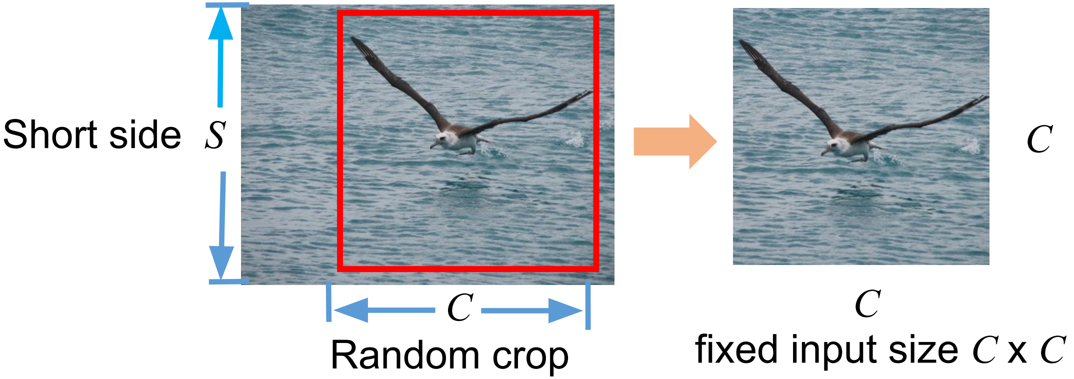
[Fast MPN-COV] Peihua Li, Jiangtao Xie, Qilong Wang and Zilin Gao. Towards Faster Training of Global Covariance Pooling Networks by Iterative Matrix Square Root Normalization. In *CVPR*, 2018.

Implementation with MatConvNet

- Pre-trained ResNet-152 on **ImageNet-11k** and then finetuned on iNaturalist 2017 dataset
- Two stage MPN-COV training on iNaturalist 2018 dataset
  - Firstly, fine-tune MPN-COV module—1x1 Conv. layer for dimension reduction, and 8142-way FC connecting normlized COV to output.
  - Next, fine-tune final 9 residual blocks and the subsequent MPN-COV module.
- Fast MPN-COV module
  - Iteration number: 3
  - Dimension of input: 160 (after a 1x1x2048x160 Conv. layer)
  - Dimension of image representation: 12,880
- Dense crop + multiple scales on test images for inference
  - Analogous to the method in Simonyan & Zisserman (ICLR 2015).
- Data augmentation with MatConvNet default jittering (color, aspect ratio, location, scale etc.)

Useful tricks — Exploit higher resolution images

- Randomly crop a *C*x*C* image from resized image with shorter side *S*, *where C=224,320,392 and S=256, 360, 448, respectively*

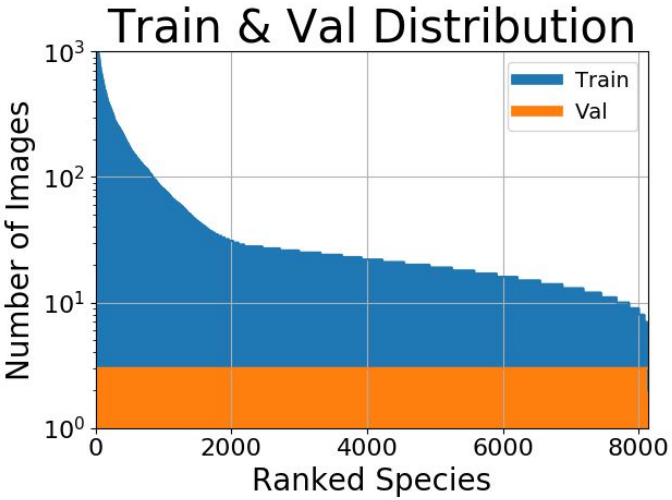


Top-3 errors (%) with single model on test set using varying input size

Input size *C*	MPN-COV with ResNet-152	Vanilla ResNet-152
320x320	**15.038**	16.623
392x392	**14.704**	16.024

Useful tricks — Deal with Long tailed distribution

Following Cui et al., we use validation set with balanced distribution to fine-tune with smaller learning rate (2e-4)



Top-3 errors (%) on test set (center 320x320 crop, *S*=360)

Finetuning on val. set	MPN-COV with ResNet-152	Vanilla ResNet-152
NO	**17.875**	18.770
YES	**15.038 (2.837↑)**	16.623 (2.147↑)

Yin Cui, Yang Song, Chen Sun, Andrew Howard, Serge Belongie. Large Scale Fine-Grained Categorization and Domain-Specific Transfer Learning. In *CVPR*, 2018

Useful tricks — Usage of iNat2017

We fine-tune on iNat Challenge 2017 dataset before training on iNat Challenge 2018 dataset

Top-3 errors (%) on test set (center 224x224 crop, *S*=256)

Finetuning on iNat2017	MPN-COV with ResNet-152	Vanilla ResNet-152
NO	N/A	25.451
YES	N/A	24.660 (0.7911↑)

Results

Evaluation on test set with *single model* using varying image resolution

Method	Input size *C*	Fusion scales *S*	Top-3 error (%)
MPN-COV with ResNet-152	320x320	360,480,512	15.038
	392x392	448,544,608	**14.704**
Vanilla ResNet-152	320x320	380,480	16.623
	392x392	480,576	16.024

Ensemble of three models with input size 392x392

Method	Top-3 error (%)
MPN-COV with ResNet-152	**13.499**
Vanilla ResNet-152	14.625
MPN-COV+Vanilla ResNet-152	**13.103**

Note: Fusion of 6 MPN-COV models (3 w/ *C*=392, 3 w/ *C*=320) and 6 vanilla ResNet-152 models (3 w/ *C*=392, 3 w/ 320), the error decreases to 13.068%.

Summary

- ➢ Matrix Power Normalized COVariance pooling (MPN-COV) networks are compelling for large-scale classification
- ➢ **Potentially**
  - — Pretrained MPN-COV on *ImageNet* will generalize better
  - — Ensemble of MPN-COV with varying architectures (e.g. ResNet and Inception) will further improve
  - — Fast MPN-COV with more iterations (>3) will benefit
  - — Higher resultion images (>392) will be more helpful

[MPN-COV] Peihua Li, Jiangtao Xie, Qilong Wang and Wangmeng Zuo. Is Second-order Information Helpful for Large-scale Visual Recognition? In *ICCV*, 2017.

[Fast MPN-COV] Peihua Li, Jiangtao Xie, Qilong Wang and Zilin Gao. Towards Faster Training of Global Covariance Pooling Networks by Iterative Matrix Square Root Normalization. In *CVPR*, 2018.

Deep Learning Analytics

Achieving <14% Test Error on the iNaturalist 2018 Competition

M.J. Trammell[1], P. Oberoi[1], J. Kaufhold[1]

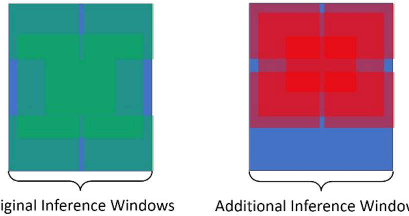
[1] DEEP LEARNING ANALYTICS

CVPR 2018 SALT LAKE CITY • JUNE 18-22

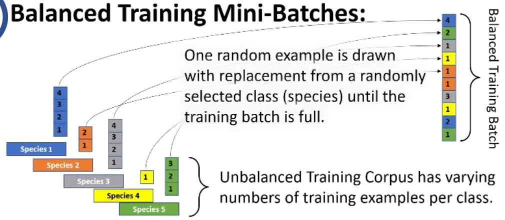
FGVC5

1 Denser viewing during inference:

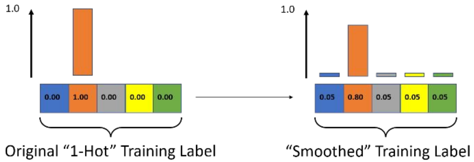
Focusing inference on the upper portions of the image reduced error.



2 Balanced Training Mini-Batches:

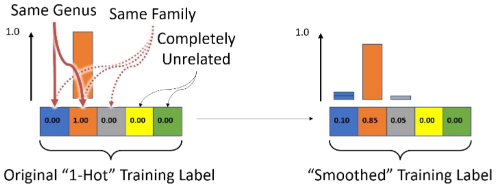


3 Generic Label Smoothing:



Regularizing Neural Networks by Penalizing Confident Output Distributions (https://arxiv.org/abs/1701.06548)

4 Taxon-Aware Label Smoothing:



Do Convolutional Neural Networks Learn Class Hierarchy? (https://arxiv.org/abs/1710.06501)

iNaturalist 2018 Competition Solution

Baseline: Our baseline ensemble consisted of a pair of networks trained on the iNaturalist training corpus and then fine-tuned on the iNaturalist validation corpus. Both models were initialized with parameters learned from the ILSVRC-2012-CLS corpus.

ENSEMBLE SCORE
public / private

0.16983
0.16982

Several new techniques were developed to improve accuracy on this dataset compared to our baseline:

1. Denser viewing during inference
2. Balanced training batches
3. Generic label smoothing
4. Taxonomy-aware label smoothing

DEEP LEARNING ANALYTICS

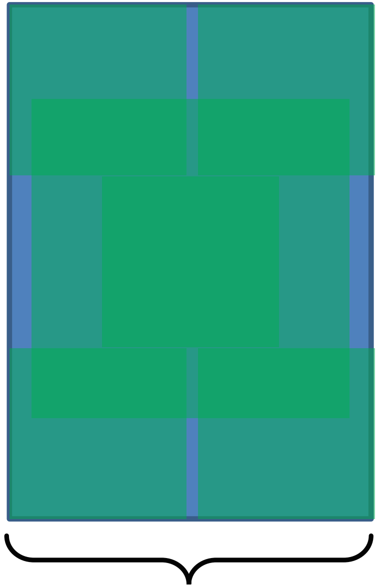
iNaturalist 2018 Competition Solution

Denser Viewing: We found that by focusing the inference stage on the upper portions of the image we could reduce our error rate.

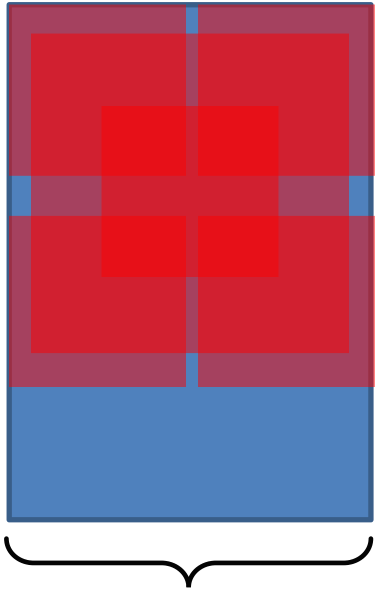
ENSEMBLE SCORE
public / private

0.16983
0.16982

0.16211
0.15991



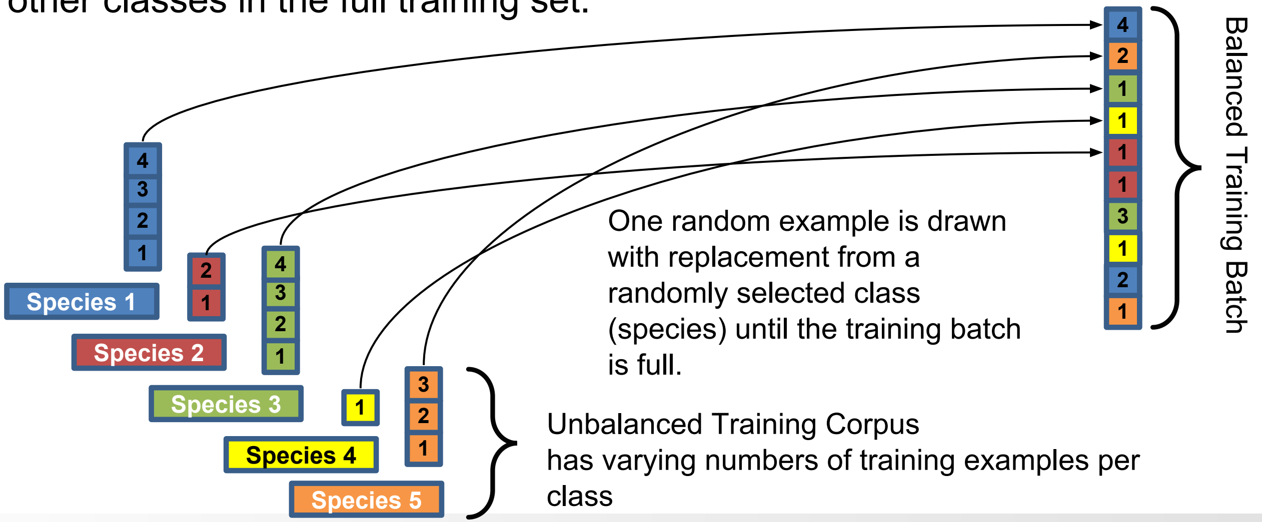
Original Inference Windows



Additional Inference Windows

iNaturalist 2018 Competition Solution

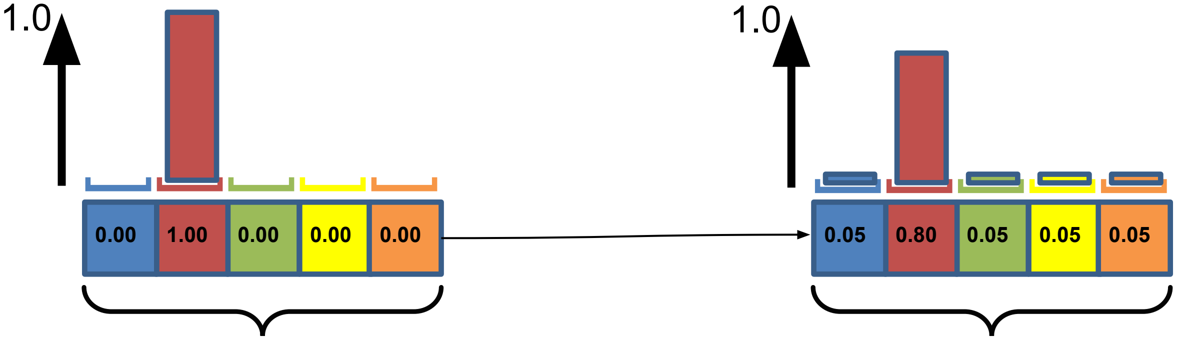
Balanced training batches: Both the 2017 and 2018 training sets have highly unbalanced class distributions. Classifier training is often done with batches of examples randomly constructed before the training process begins. By dynamically constructing these training batches during training, we can ensure that each class is evenly seen during training, even if it has fewer examples than other classes in the full training set.



ENSEMBLE SCORE public / private
0.16983
0.16982
0.16211
0.15991
0.14703
0.14536

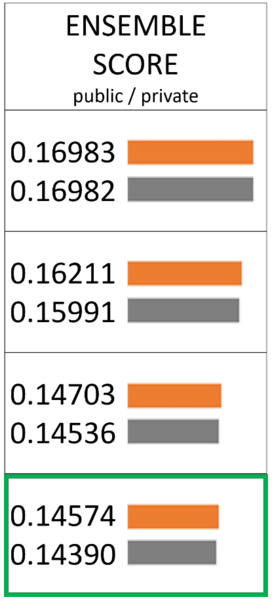
iNaturalist 2018 Competition Solution

Label Smoothing: One symptom of networks which are overfitting to their training data is unreasonably high output confidence. One way to reduce this form of overfitting is to “smooth” the labels during training so that instead of one “correct” answer and multiple “incorrect” answers the network is provided with one “mostly correct” label and multiple “mostly incorrect” labels.

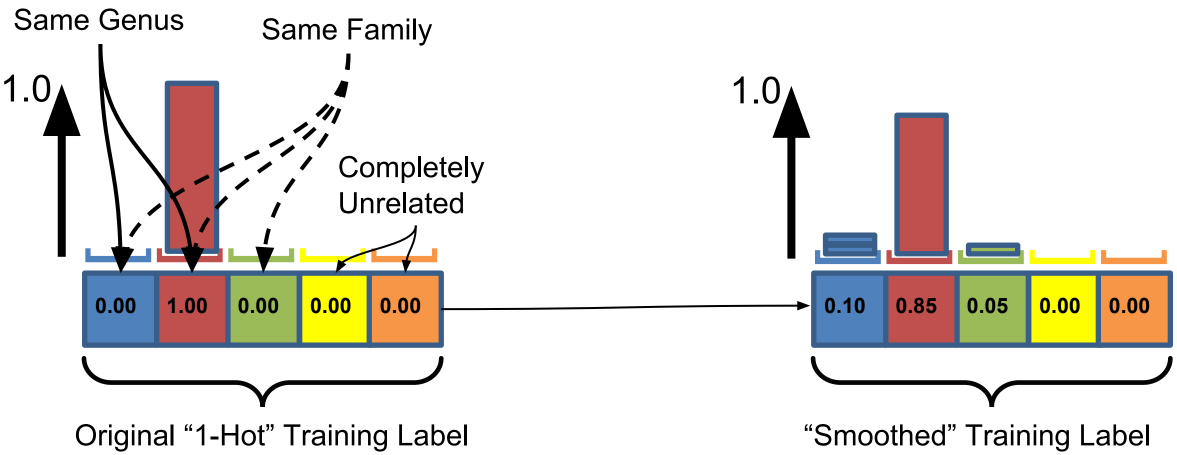


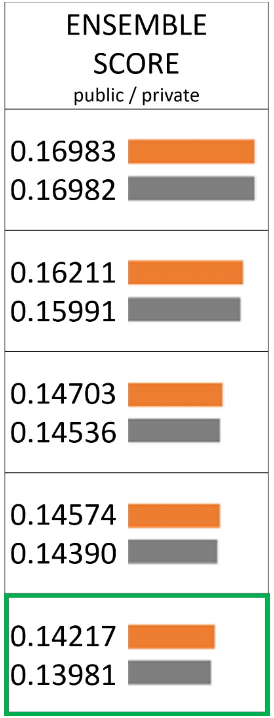
Original “1-Hot” Training Label

“Smoothed” Training Label



Taxon-aware Label Smoothing: Because living organisms exist within a hierarchical taxonomy, we hypothesized that a more intelligent form of label smoothing could be employed. Rather than uniformly smoothing all labels, we developed a system to only smooth labels for species within the same family and genus.





iNaturalist 2018 Competition Solution

Future Exploration: Ensembling is a major component of any kaggle competition solution. Because of our specific research interests, we did not spend much time developing a maximally effective ensembling technique for this competition. It is our belief that there is substantial room for improvement in this direction.

Our study of ROIs was also very limited. While we found that adding more inference windows to certain fixed regions of each image was beneficial, a more nuanced approach to generating per image ROIs is likely to show an improvement in performance.

Thanks!