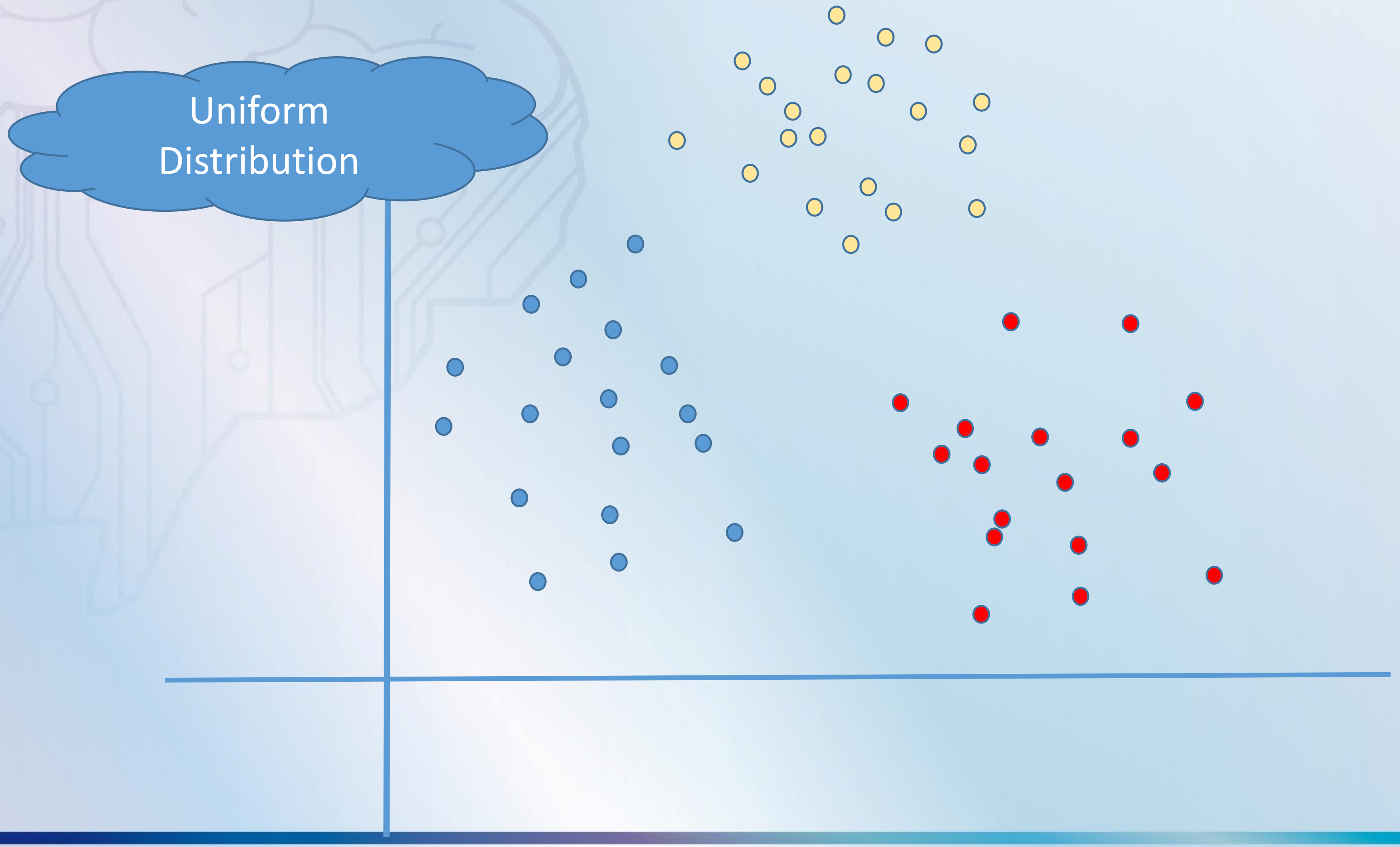
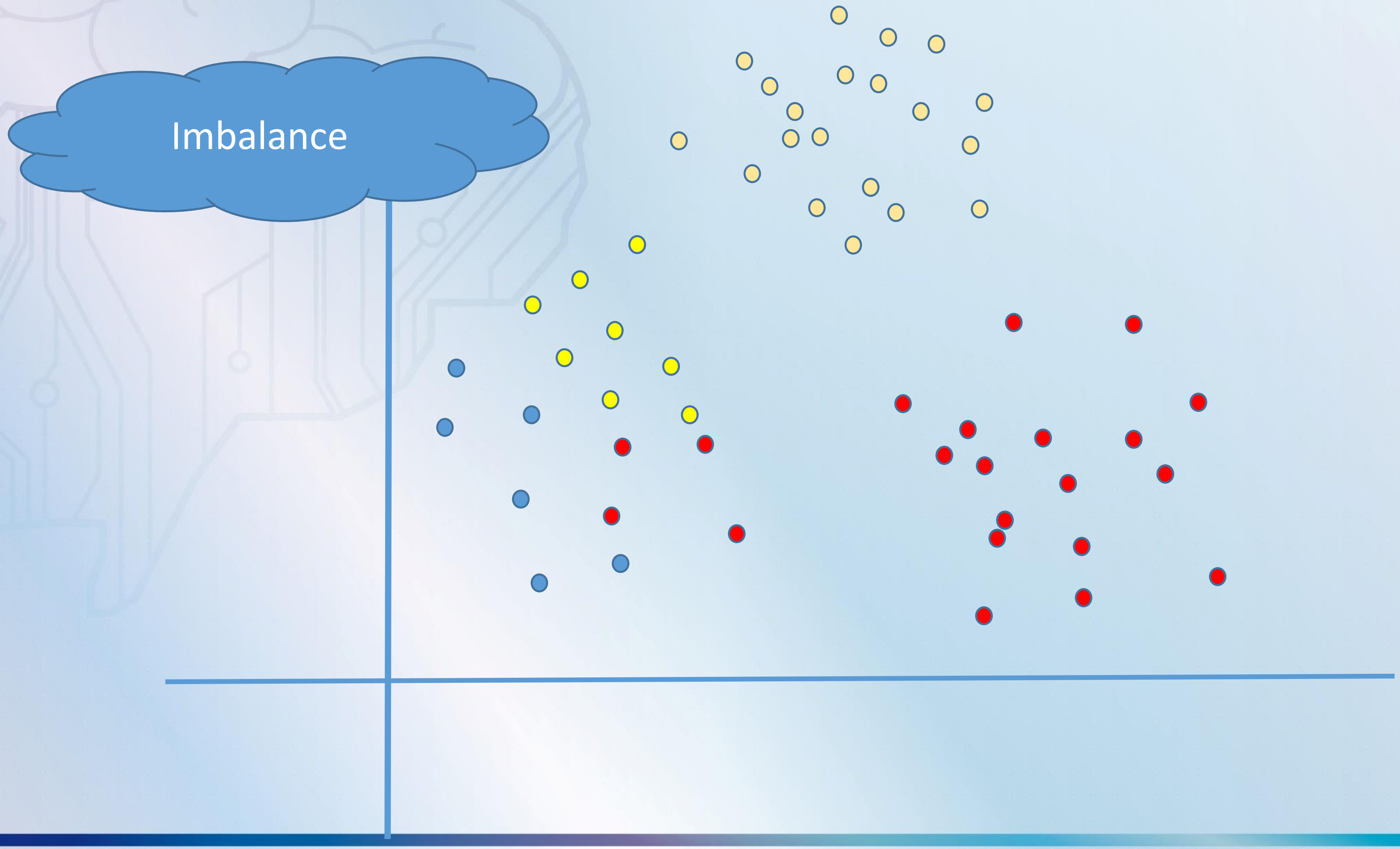
The iNaturalist Species Classification and Detection Dataset

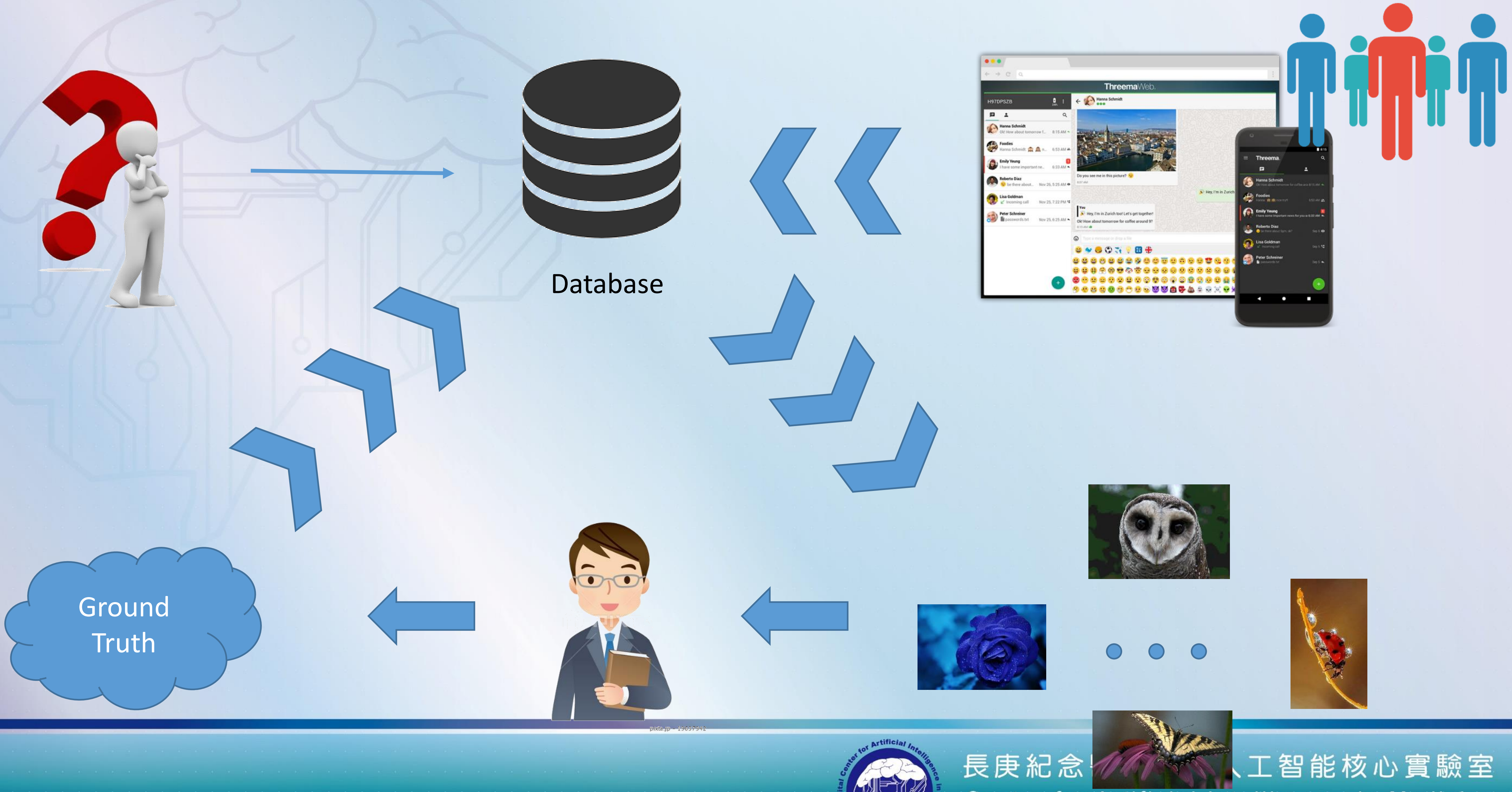
Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun Alex Shepard, Hartwig Adam, Pietro Perona[1] Serge Belongie

長庚紀念醫院醫療人工智能核心實驗室

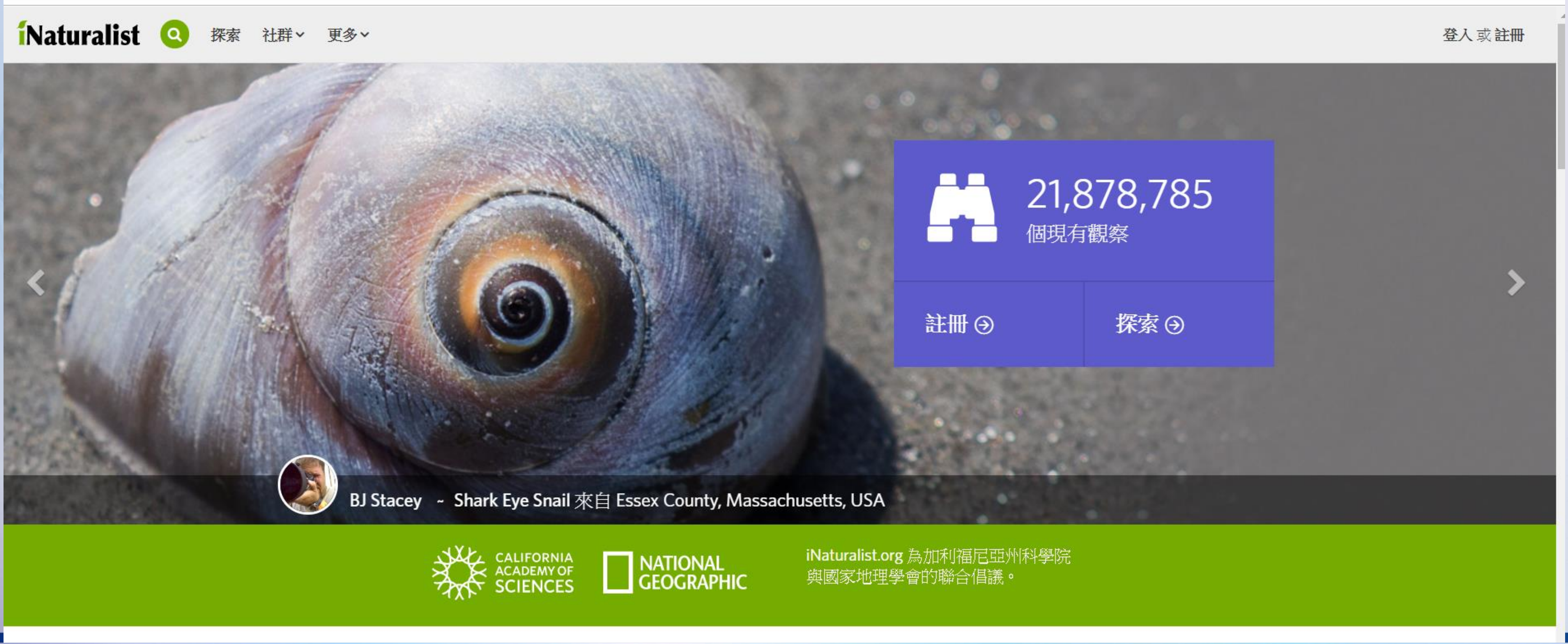
Center for Artificial Intelligence in Medicine
Chang Gung Memorial Hospital


Uniform
Distribution


Imbalance


Database
Ground Truth

https://www.inaturalist.org/



它是如何運作的

記錄您的觀察

與關注的博物學友分享

討論您所發現的

觸手可及的自然

保持追蹤

在雲端上記錄您與各種生物的相遇，以及維護生物清單。

創建有用的資料

協助科學家與資源管理者了解生物在何時及何處出現。

群眾外包鑑定

聯絡可以鑑定您所觀察到生物的專家。

成為一位公眾科學家

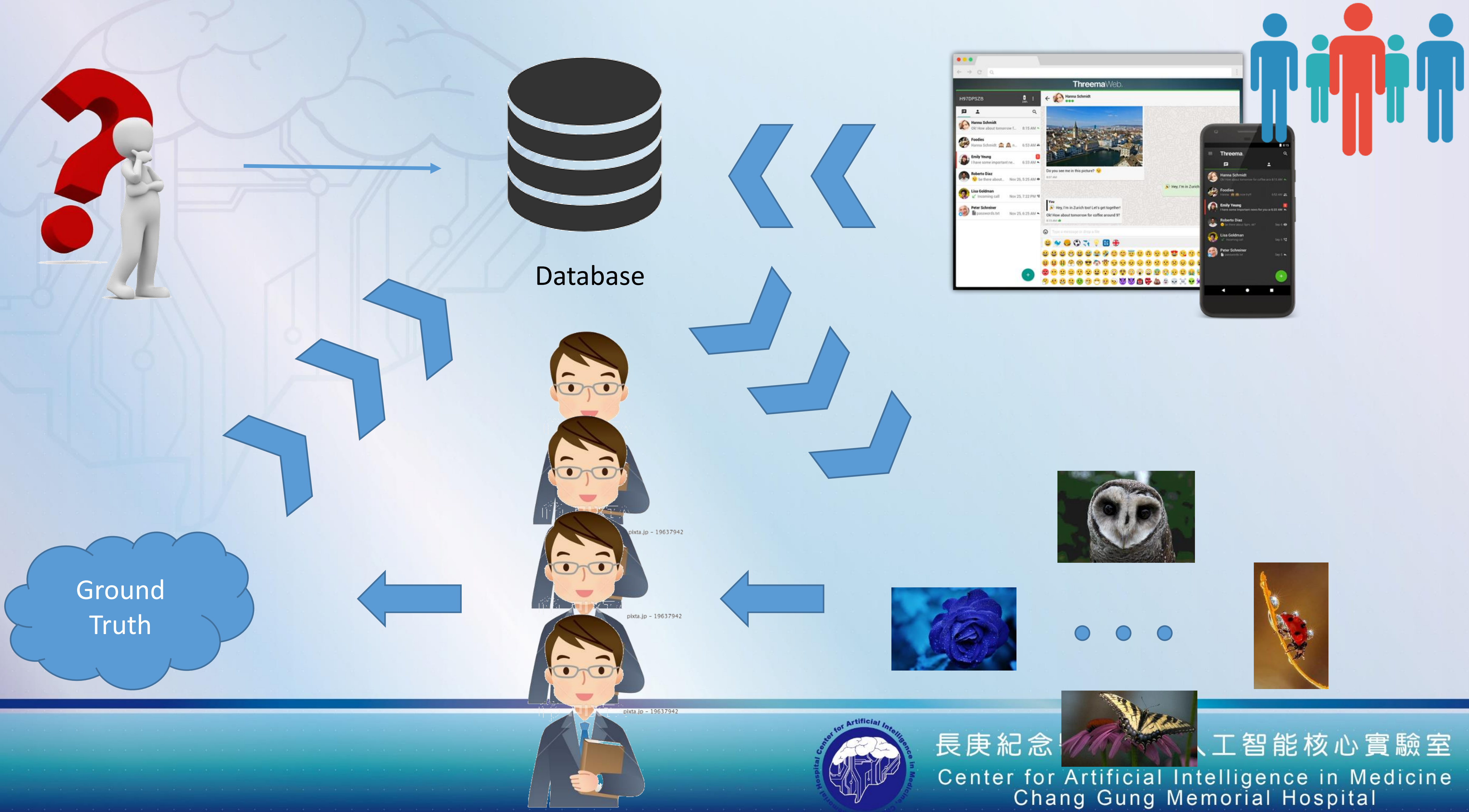
尋找一個有著您感興趣任務的專案，或是開始您自己擁有的專案。

了解關於自然

透過與其他博物學友的交談，以及協助他人來建立您的知識。

舉行生態速查（Bioblitz）

舉行一場能讓人們盡情來找尋物種的活動。


Database
Ground Truth

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*.

Dataset Name	# Train	# Classes	Imbalance
Flowers 102 [27]	1,020	102	1.00
Aircraft [24]	3,334	100	1.03
Oxford Pets [29]	3,680	37	1.08
DogSnap [23]	4,776	133	2.85
CUB 200-2011 [42]	5,994	200	1.03
Stanford Cars [19]	8,144	196	2.83
Stanford Dogs [16]	12,000	120	1.00
Urban Trees [43]	14,572	18	7.51
NABirds [40]	23,929	555	15.00
LeafSnap* [20]	30,866	185	8.00
CompCars* [48]	136,727	1,716	10.15
VegFru* [10]	160,731	292	8.00
Census Cars [7]	512,765	2,675	10.00
ILSVRC2012 [32]	1,281,167	1,000	1.78
iNat2017	579,184	5,089	435.44

Maximum / Minimum

4% error

Table 1. Summary of popular general and fine-grained computer vision classification datasets. 'Imbalance' represents the number of images in the largest class divided by the number of images in the smallest. While susceptible to outliers, it gives an indication of the imbalance found in many common datasets. *Total number of train, validation, and test images.

	Super-Class	Class	Train	Val	BBoxes
	Plantae	2,101	158,407	38,206	-
	Insecta	1,021	100,479	18,076	125,679
	Aves	964	214,295	21,226	311,669
	Reptilia	289	35,201	5,680	42,351
	Mammalia	186	29,333	3,490	35,222
	Fungi	121	5,826	1,780	-
	Amphibia	115	15,318	2,385	18,281
	Mollusca	93	7,536	1,841	10,821
	Animalia	77	5,228	1,362	8,536
	Arachnida	56	4,873	1,086	5,826
	Actinopterygii	53	1,982	637	3,382
	Chromista	9	398	144	-
	Protozoa	4	308	73	-
	Total	5,089	579,184	95,986	561,767

Table 2. Number of images, classes, and bounding boxes in iNat2017 broken down by super-class. 'Animalia' is a catch-all category that contains species that do not fit in the other super-classes. Bounding boxes were collected for nine of the super-classes. In addition, the public and private test sets contain 90,427 and 92,280 images, respectively.

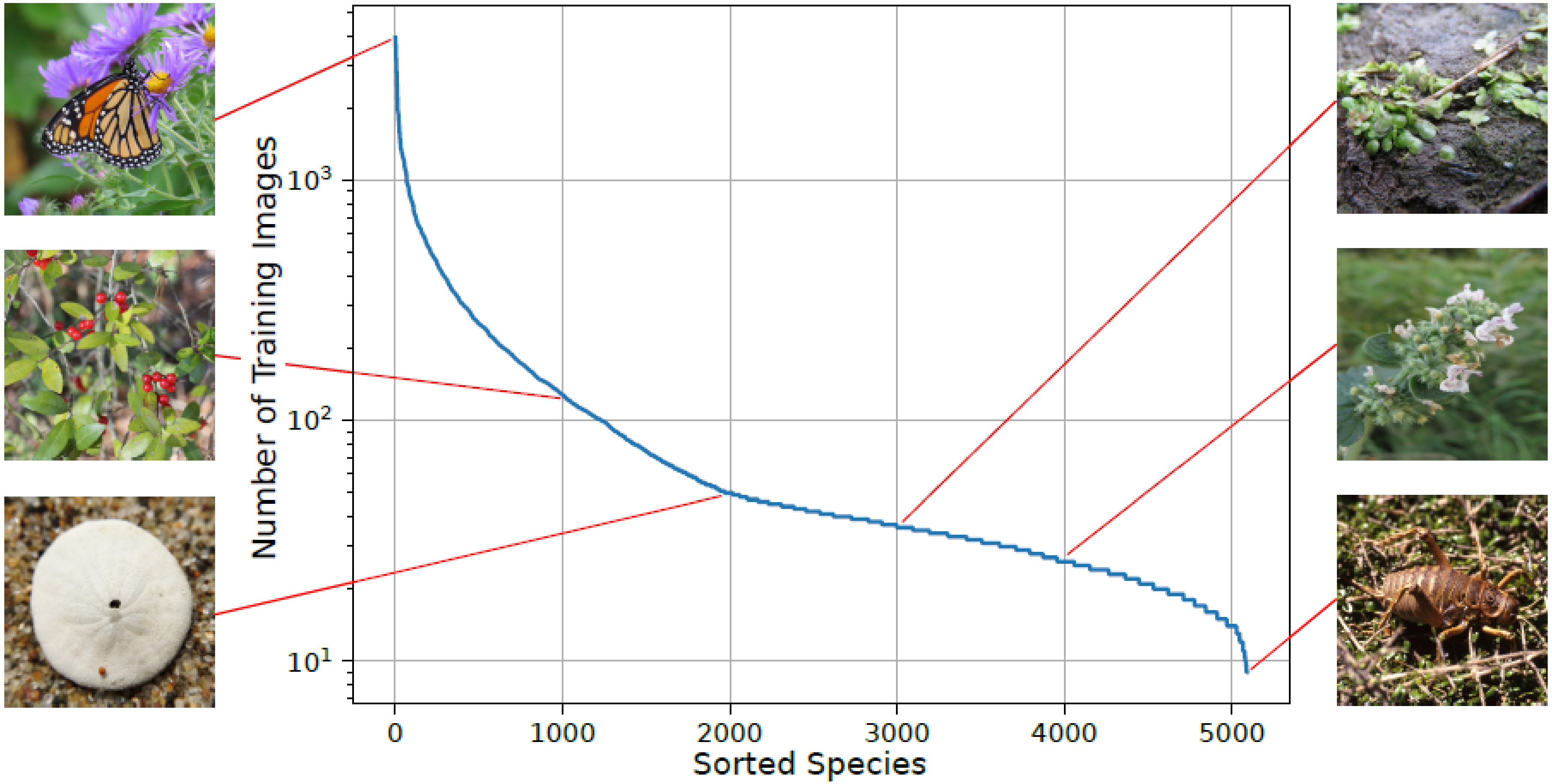


Figure 2. Distribution of training images per species. iNat2017 contains a large imbalance between classes, where the top 1% most populated classes contain over 16% of training images.

	Validation		Public Test		Private Test	
	Top1	Top5	Top1	Top5	Top1	Top5
IncResNetV2 SE	**67.3**	**87.5**	**68.5**	**88.2**	67.7	**87.9**
IncResNetV2	67.1	**87.5**	68.3	88.0	**67.8**	87.8
IncV3 SE	65.0	85.9	66.3	86.7	65.2	86.3
IncV3	64.2	85.2	65.5	86.1	64.8	85.7
ResNet152 drp	62.6	84.5	64.2	85.5	63.1	85.1
ResNet101 drp	60.9	83.1	62.4	84.1	61.4	83.6
ResNet152	59.0	80.5	60.6	81.7	59.7	81.3
ResNet101	58.4	80.0	59.9	81.2	59.1	80.9
MobileNet V1	52.9	75.4	54.4	76.8	53.7	76.3

Table 3. Classification results for various CNNs trained on only the training set, using a single center crop at test time. Unlike some current datasets where performance is near saturation, iNat2017 still poses a challenge for state-of-the-art classifiers.

Super-Class	Avg Train	Public Test	
		Top1	Top5
Plantae	75.4	69.5	87.1
Insecta	98.4	77.1	93.4
Aves	222.3	67.3	88.0
Reptilia	121.8	45.9	80.9
Mammalia	157.7	61.4	85.1
Fungi	48.1	74.0	92.3
Amphibia	67.9	51.2	81.0
Mollusca	81.0	72.4	90.9
Animalia	67.9	73.8	91.1
Arachnida	87.0	71.5	88.8
Actinopterygii	37.4	70.8	86.3
Chromista	44.2	73.8	92.4
Protozoa	77.0	89.2	96.0

Table 4. Super-class level accuracy (computed by averaging across all species within each super-class) for the best performing model Inception ResNetV2 SE [12]. "Avg Train" indicates the average number of training images per class for each super-class. We observe a large difference in performance across the different super-classes.

療人工智能核心實驗室
al Intelligence in Medicine
Chang Gung Memorial Hospital

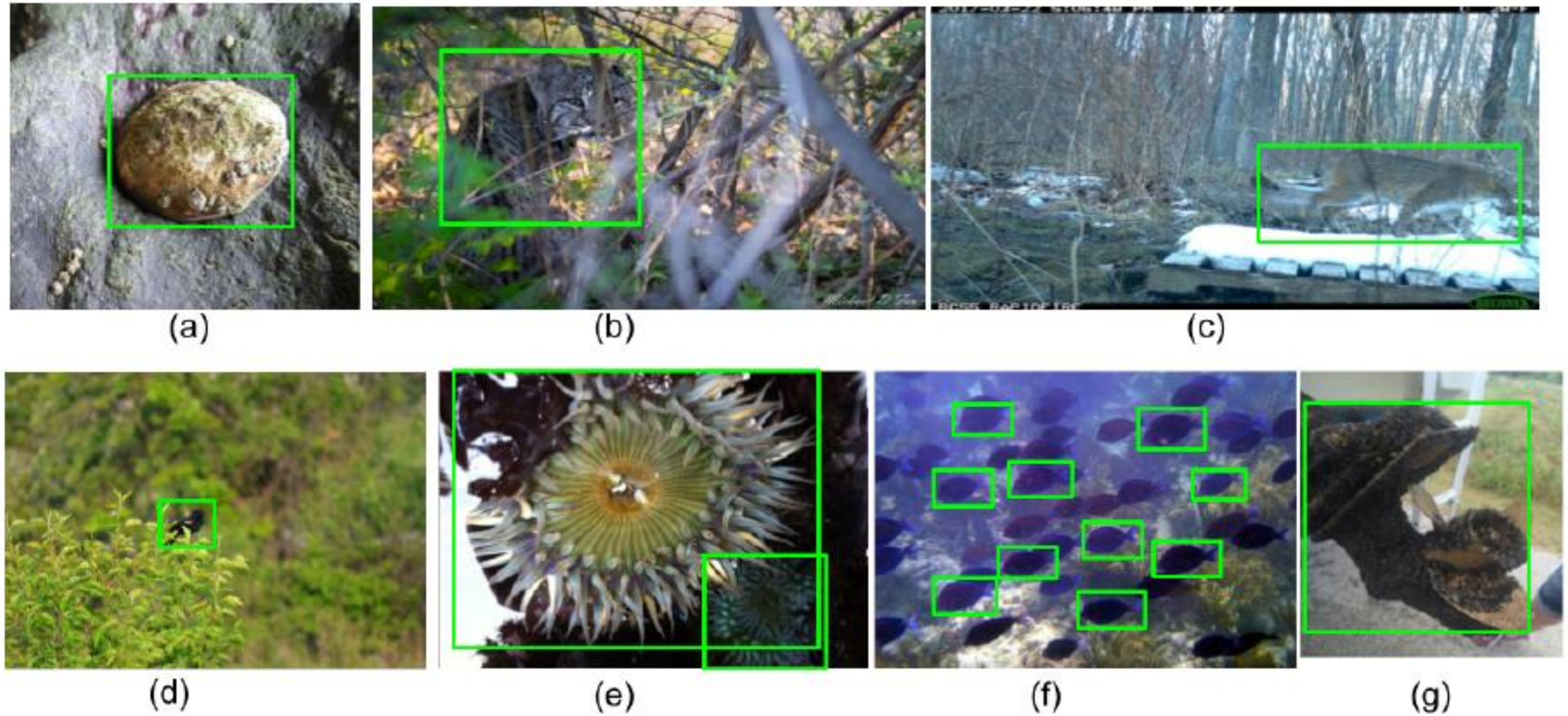


Figure 3. Sample bounding box annotations. Annotators were asked to annotate up to 10 instances of a super-class, as opposed to the fine-grained class, in each image.

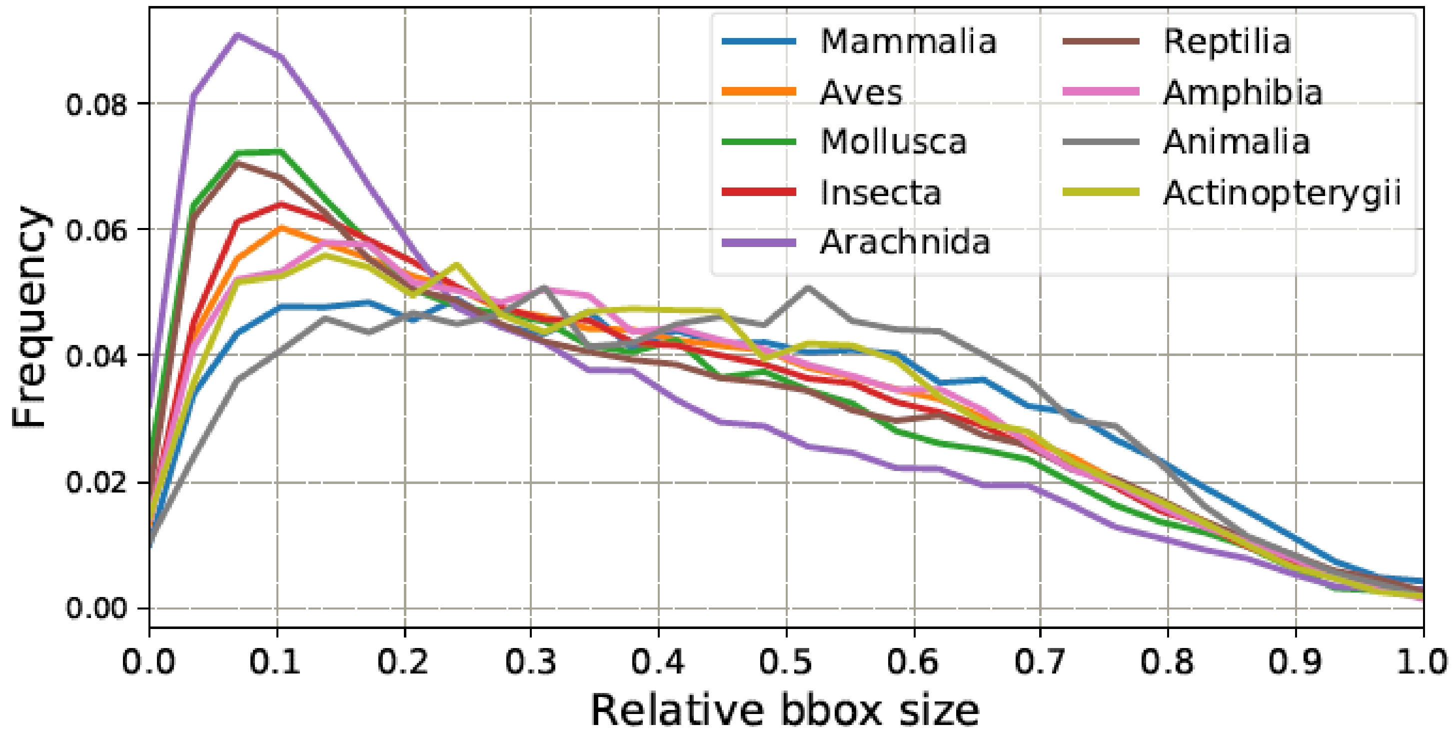


Figure 4. The distribution of relative bounding box sizes (calculated by $\sqrt{w_{bbox} \times h_{bbox}} / \sqrt{w_{img} \times h_{img}}$) in the training set, per super-class. Most objects are relatively small or medium sized.

Faster-RCNN

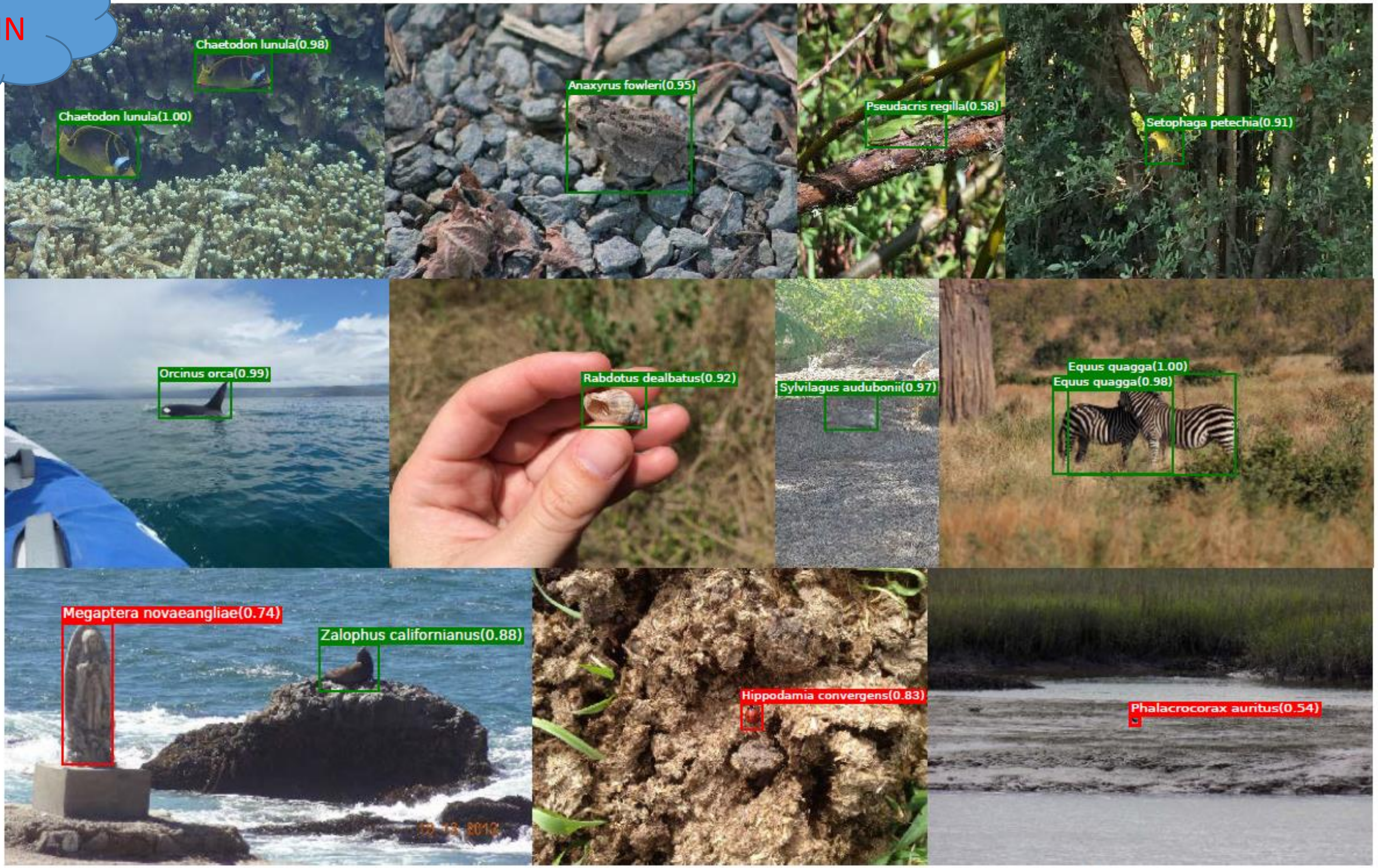


Figure 7. Sample detection results for the 2,854-class model that was evaluated across all validation images. Green boxes represent correct species level detections, while reds are mistakes. The bottom row depicts some failure cases. We see that small objects pose a challenge for classification, even when localized well.

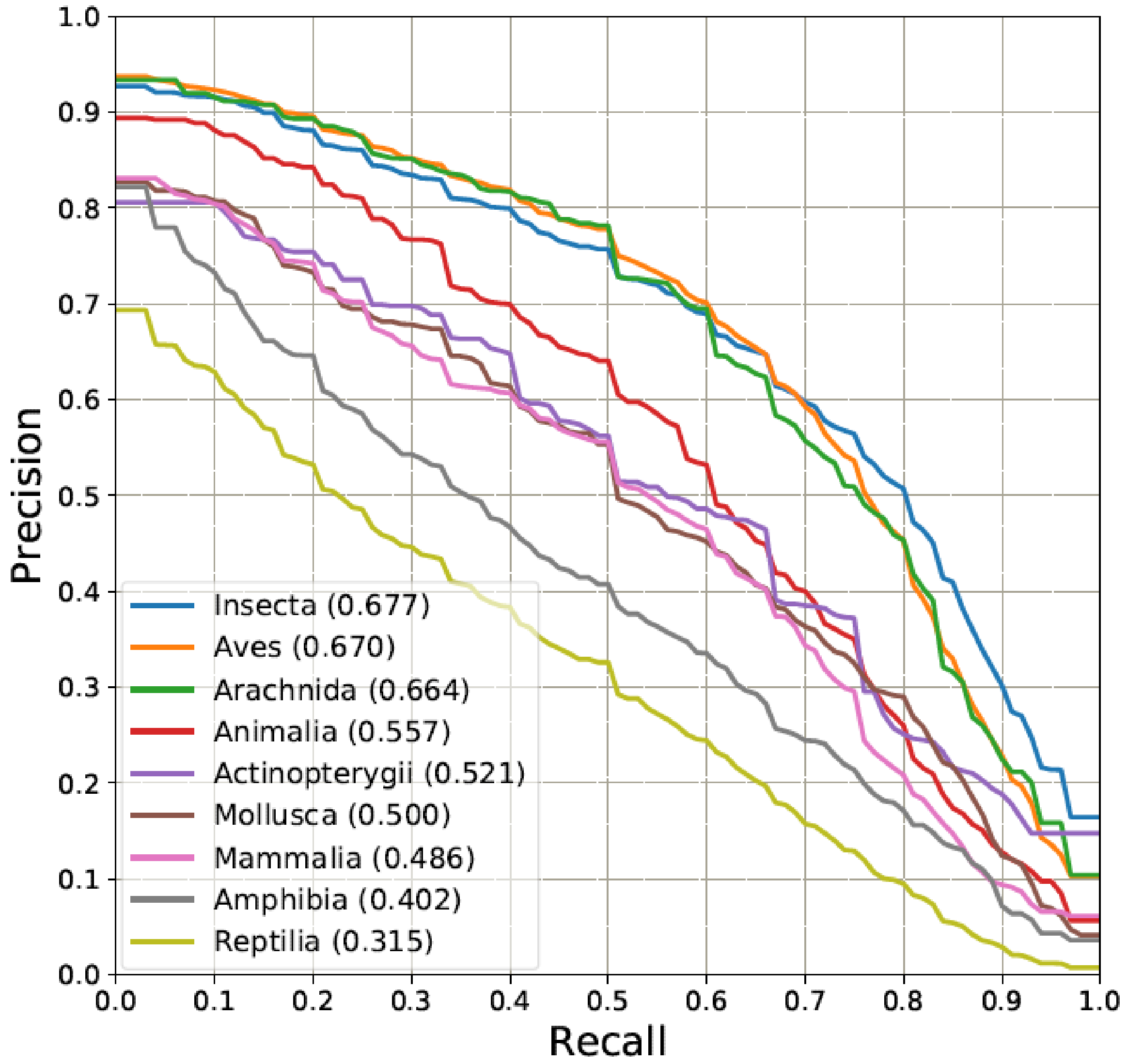


Figure 6. Precision-Recall curve with 0.5 IoU for each super-class, where the Area-Under-Curve (AUC) corresponds to AP^{50} in Table 5. Super-class performance is calculated by averaging across all fine-grained classes. We can see that building a detector that works well for all super-classes in iNat2017 will be a challenge.

	AP	AP^{50}	AP^{75}	AR^{1}	AR^{10}
Insecta	49.4	**67.7**	**59.3**	**64.5**	**64.9**
Aves	**49.5**	67.0	59.1	63.3	63.6
Reptilia	21.3	31.5	24.9	44.0	44.8
Mammalia	33.3	48.6	39.1	49.8	50.6
Amphibia	28.7	40.2	35.0	52.0	52.3
Mollusca	34.8	50.0	41.6	52.0	53.0
Animalia	35.6	55.7	40.8	48.3	50.5
Arachnida	43.9	66.4	49.6	57.3	58.6
Actinopterygii	35.0	52.1	41.6	49.1	49.6
Overall	43.5	60.2	51.8	59.3	59.8

Table 5. Super-class-level Average Precision (AP) and Average Recall (AR) for object detection, where AP, AP^{50} and AP^{75} denotes AP@[IoU=.50:.05:.95], AP@[IoU=.50] and AP@[IoU=.75] respectively; AR^{1} and AR^{10} denotes AR given 1 detection and 10 detections per image.

https://paperswithcode.com/paper/the-inaturalist-species-classification-and

Code

✎ Edit

tensorflow/models ★ 54,245 TensorFlow

Tasks

✎ Edit

IMAGE CLASSIFICATION

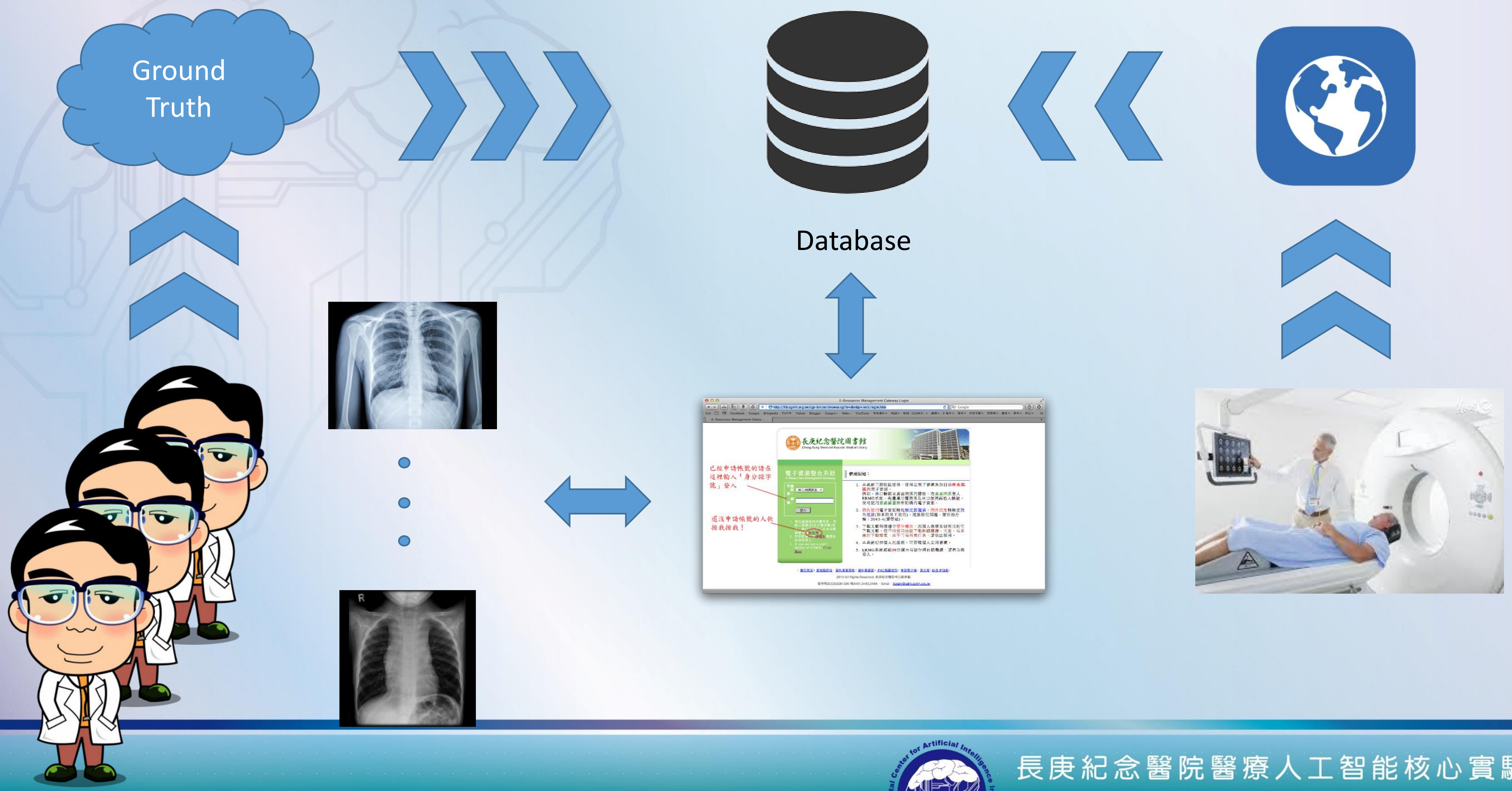
Evaluation results from the paper

✎ Edit

SOTA for Image Classification on iNaturalist

→| Get a GitHub badge

Task	Dataset	Model	Metric name	Metric value	Global rank	Compare
Image Classification	iNaturalist	IncResNetV2 SE	Top 1 Accuracy	67.3%	# 1	See all
Image Classification	iNaturalist	IncResNetV2 SE	Top 5 Accuracy	87.5%	# 1	See all


Ground Truth
Database

Thanks for listening