Data Imbalance and Fine-Grained Classification

6.S954 Computer Vision and Planetary Health

Justin Kay 02/20/25

PASCAL

ILSVRC

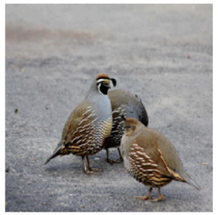
birds

bird

flamingo

cock

ruffed grouse



quail

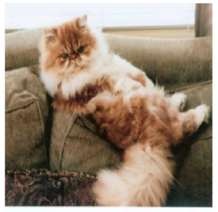
partridge

...

cats

cat

Egyptian cat



Persian cat

Siamese cat

tabby

lynx

...

dogs

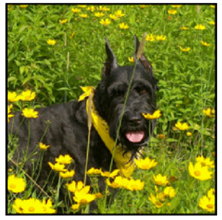
dog

dalmatian

keeshond

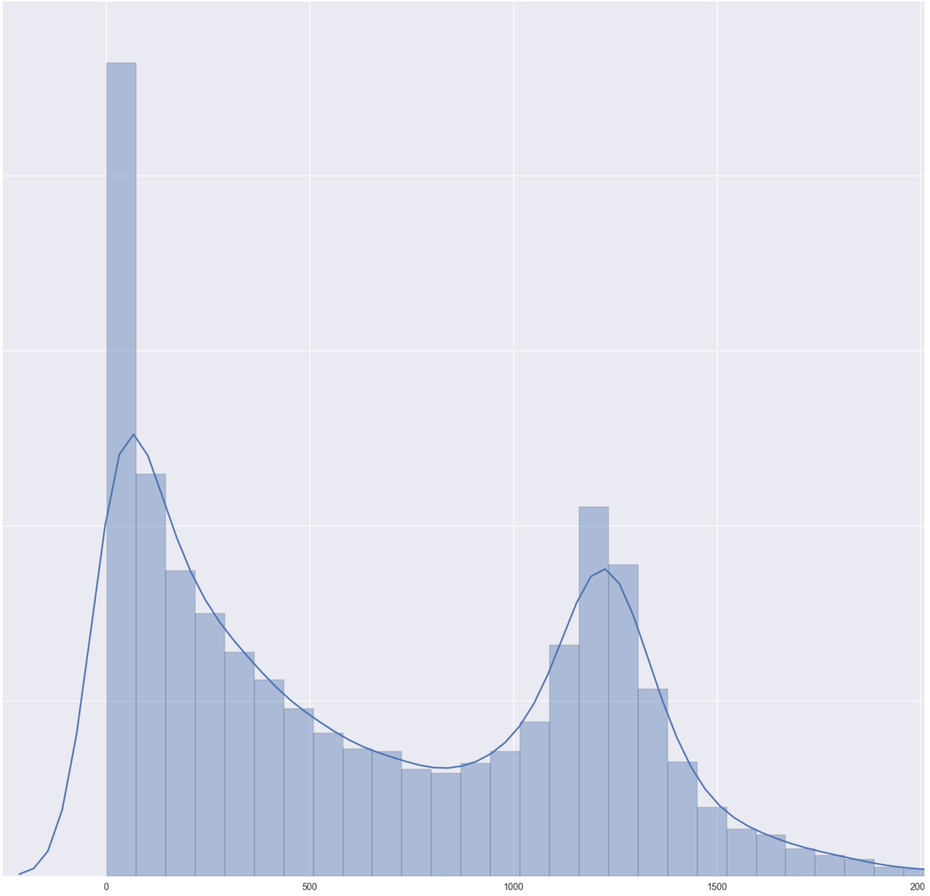
miniature schnauzer

standard schnauzer

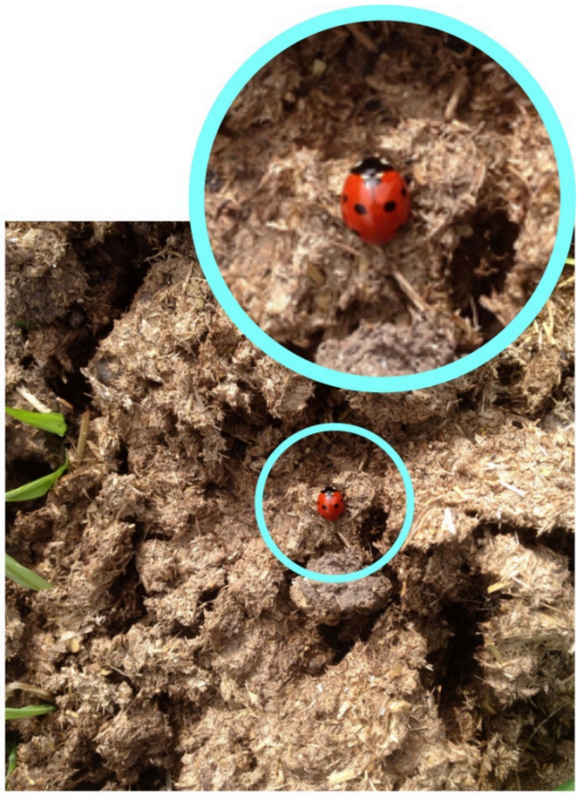


giant schnauzer

...



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*.

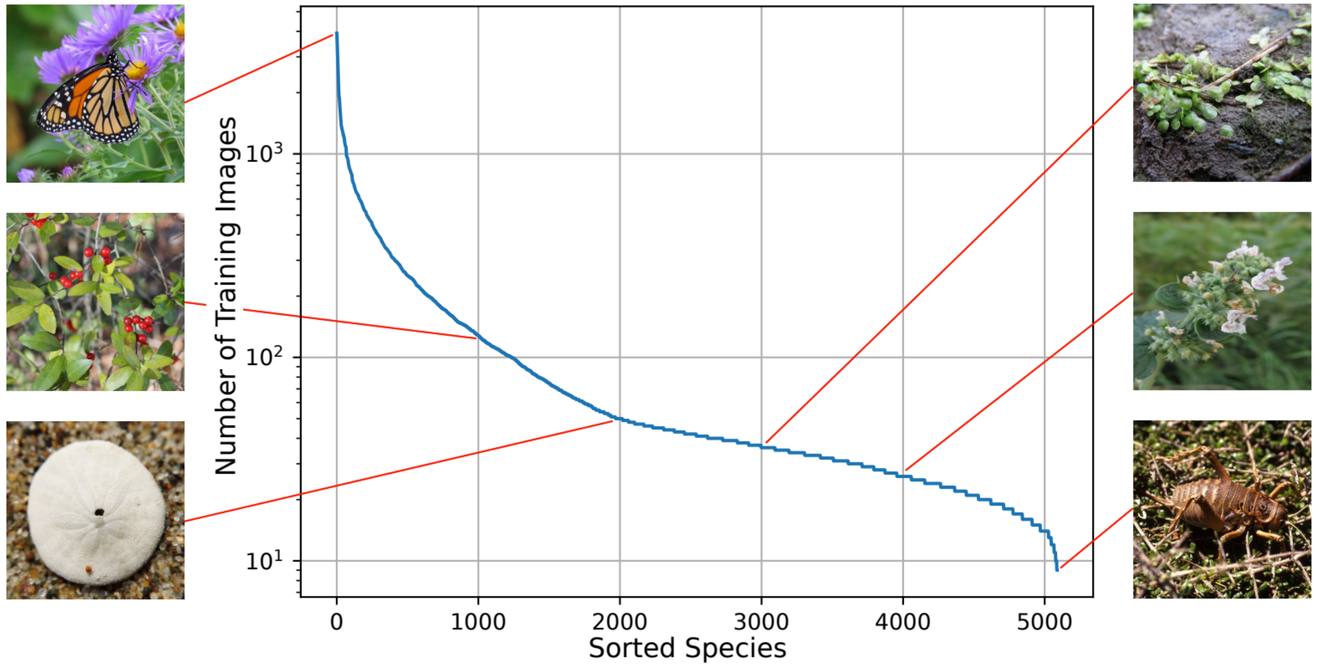
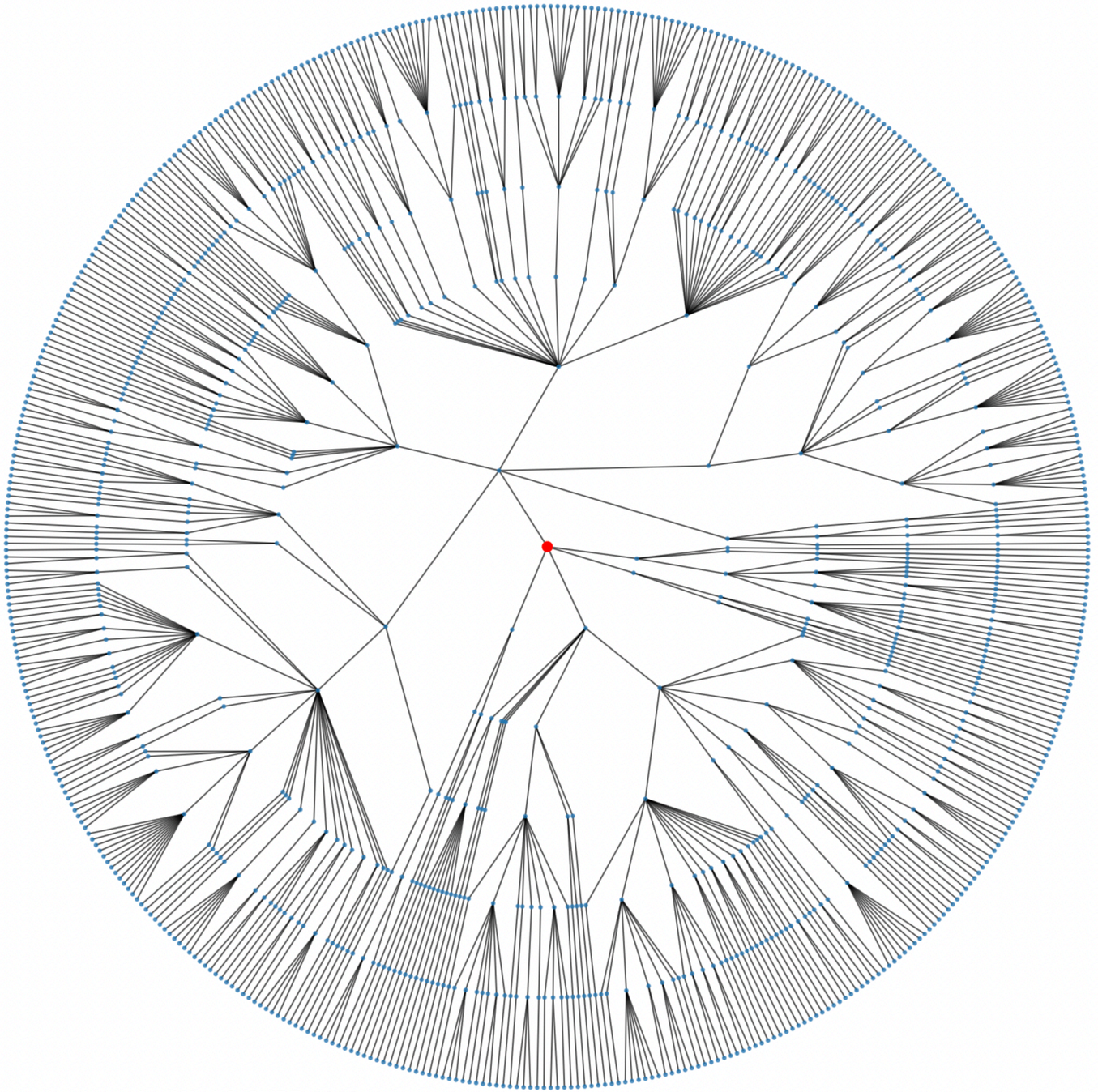
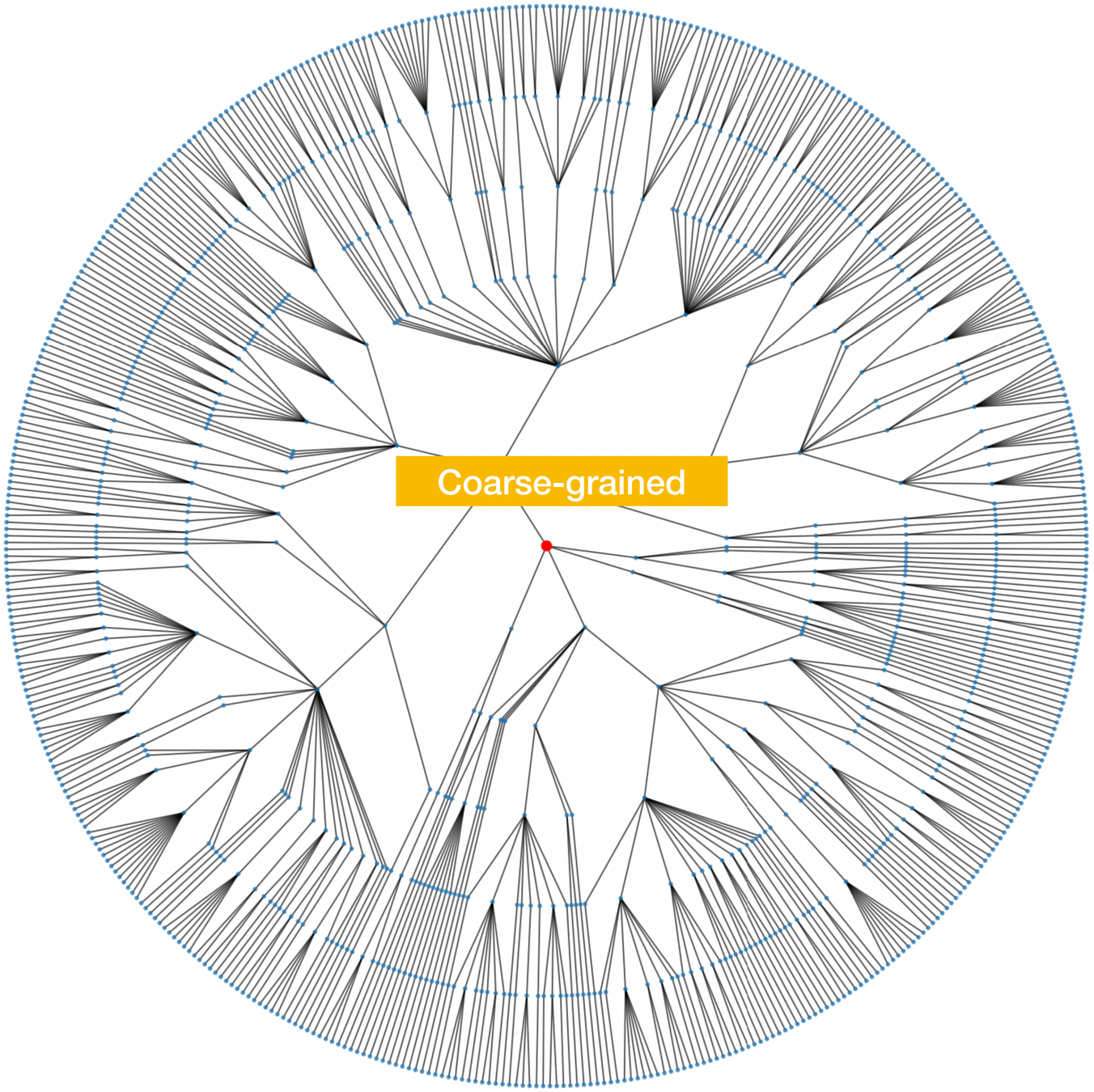


Figure 2. Distribution of training images per species. iNat2017 contains a large imbalance between classes, where the top 1% most populated classes contain over 16% of training images.

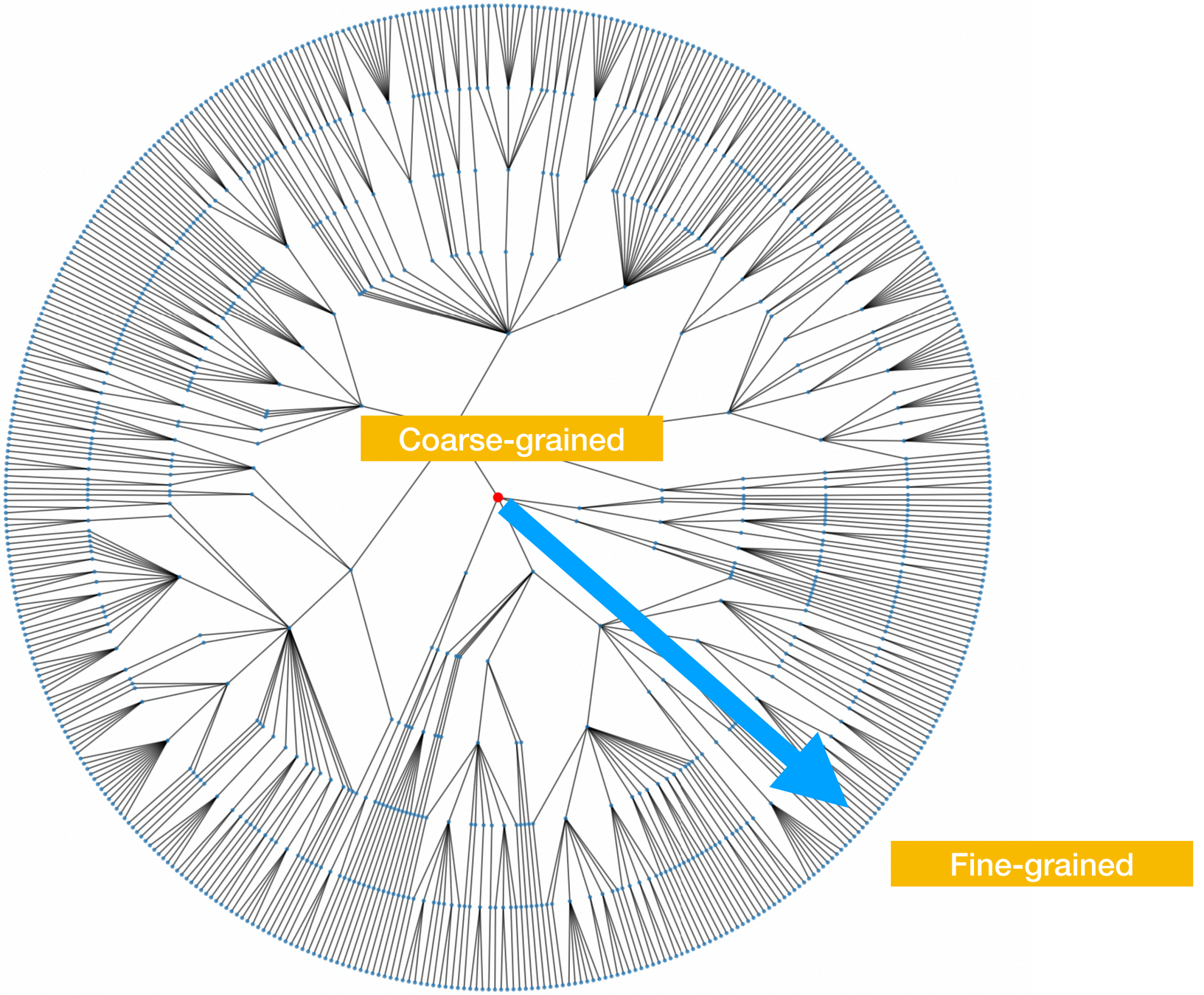


Cole et al., *When Does Contrastive Visual Representation Learning Work?*, CVPR 2022

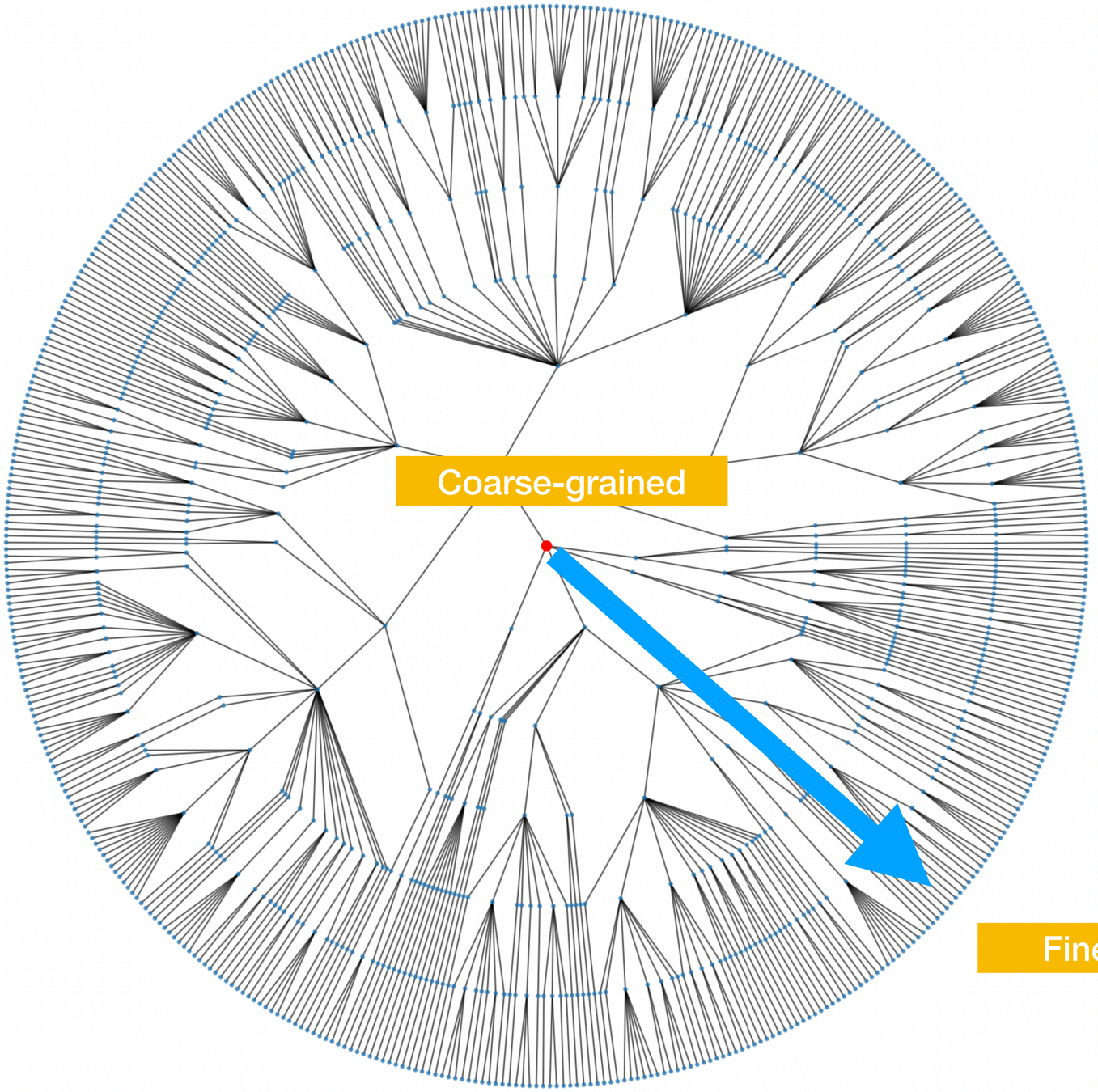
Slide: Elijah Cole



Slide: Elijah Cole



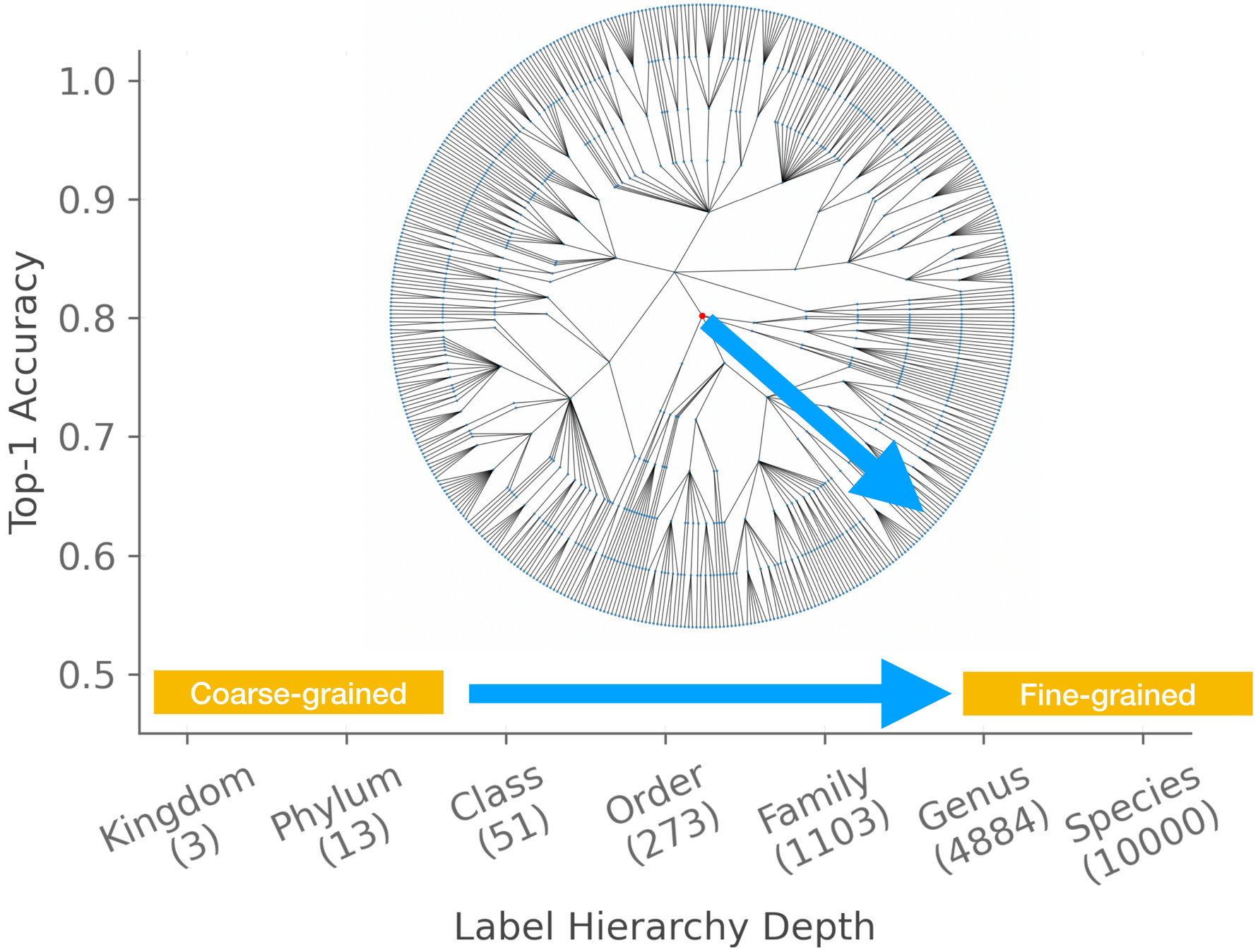
Slide: Elijah Cole



S. umbilicata

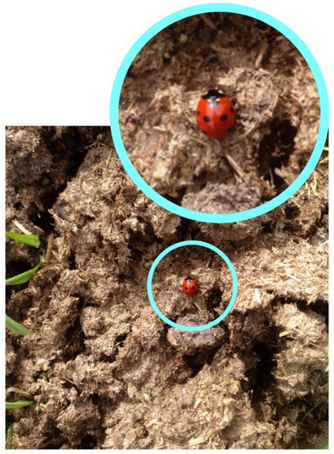
S. ornata

Slide: Elijah Cole



Slide: Elijah Cole

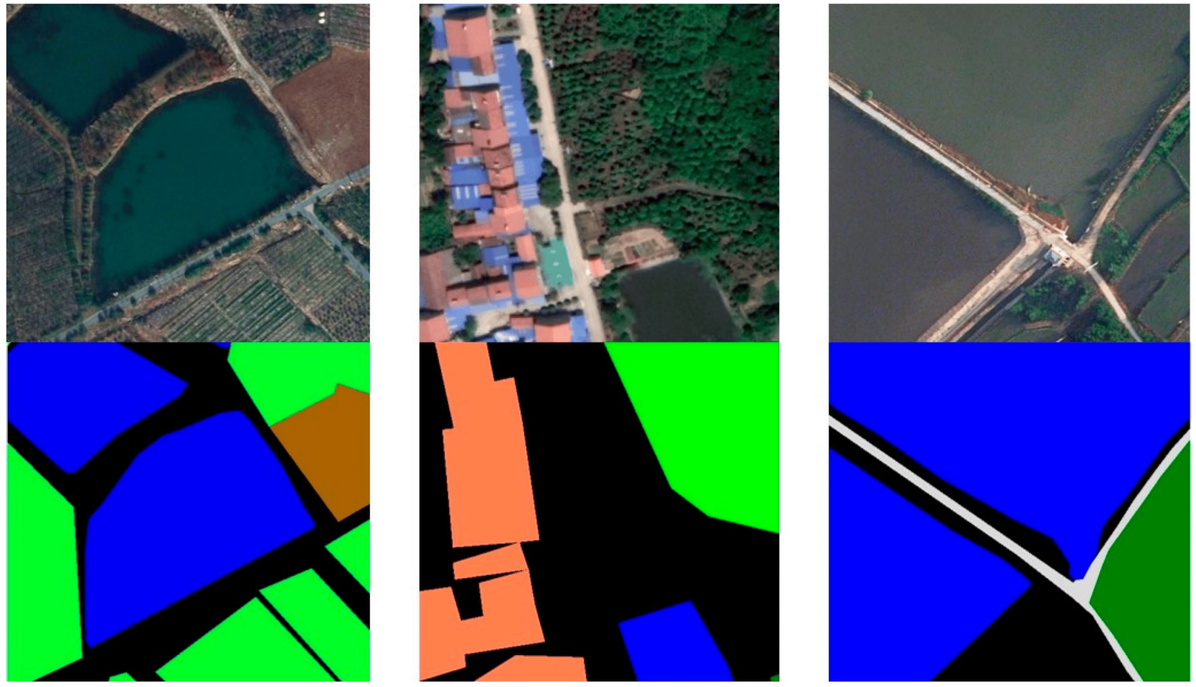
Two-spotted ladybug
Adalia bipunctata



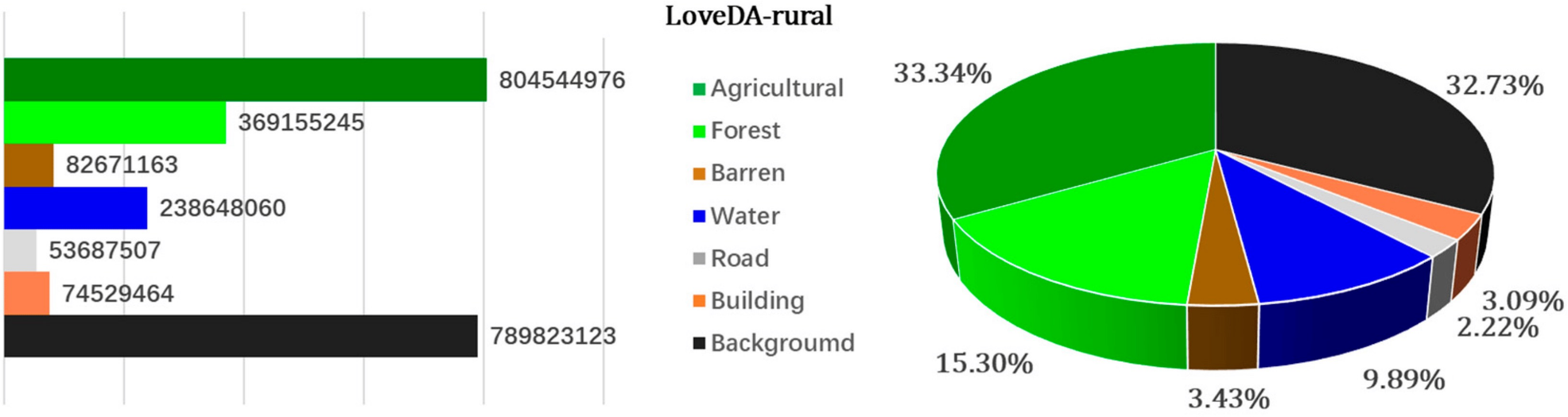
Seven-spotted ladybug
Coccinella septempunctata

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*.

Granularity spectrum



(a) (b) (c)


LoveDA-rural
804544976
369155245
82671163
238648060
53687507
74529464
789823123
Agricultural
Forest
Barren
Water
Road
Building
Backgroumd
33.34%
32.73%
3.09%
2.22%
9.89%
3.43%
15.30%

Other sources of imbalance

Not just class distribution!

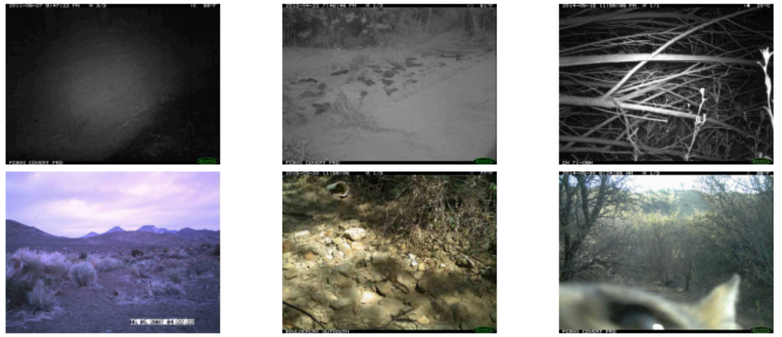


Figure 2. **Common data challenges**: (1) **Illumination**: Animals are not always salient. (2) **Motion blur**: common with poor illumination at night. (3) **Size of the region of interest** (ROI): Animals can be small or far from the camera. (4) **Occlusion**: e.g. by bushes or rocks. (5) **Camouflage**: decreases saliency in animals' natural habitat. (6) **Perspective**: Animals can be close to the camera, resulting in partial views of the body.

Male and Female Blue Grosbeak ***(Passerina caerulea)***

Image credits: Ed Schneider/Shutterstock.com (left), Steve Byland/Shutterstock.com (right)

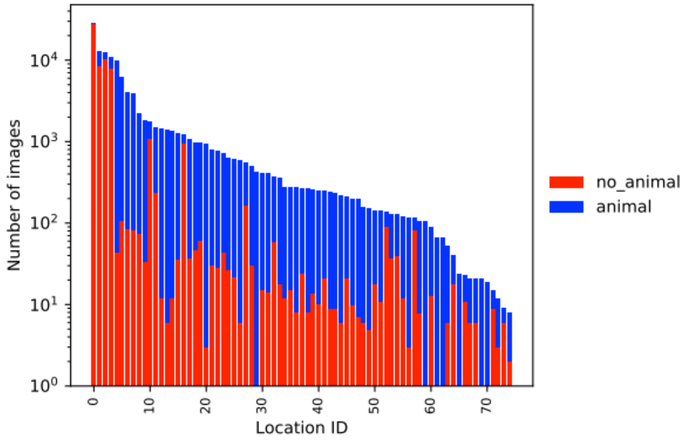
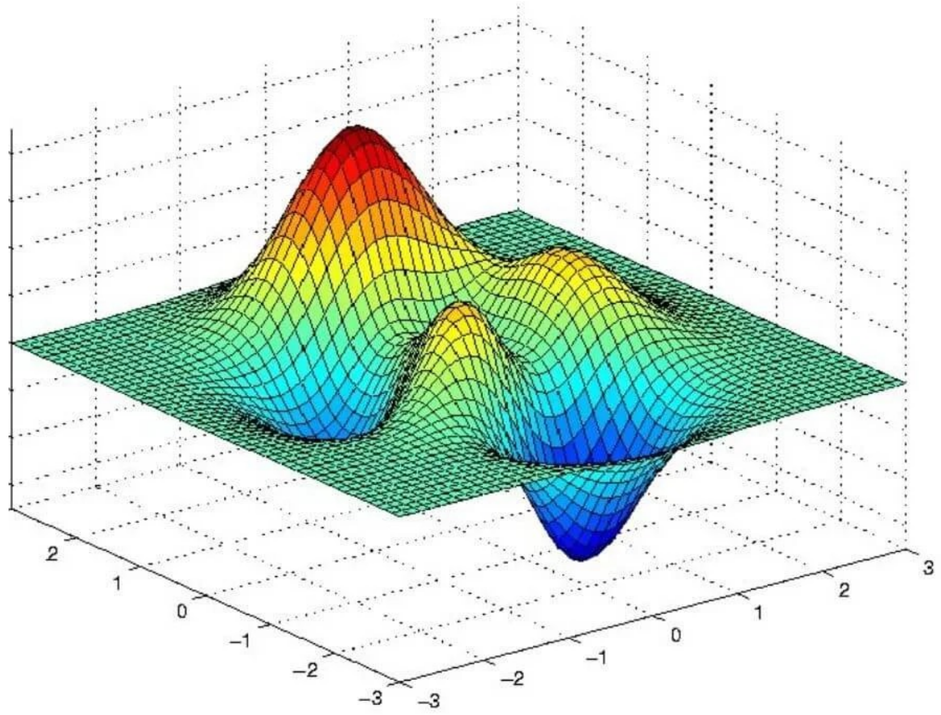
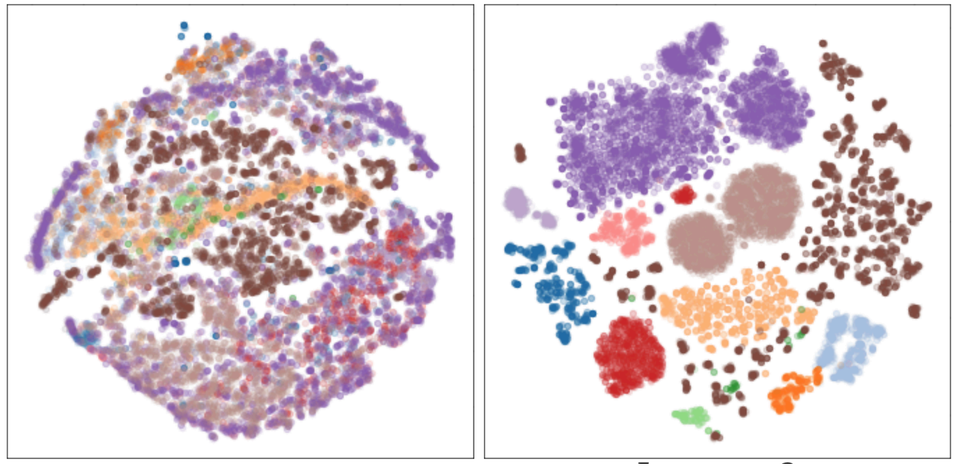


Figure 3. Number of annotations for each location, over the two classes. The distribution images per location is long-tailed, and each location has a different and peculiar class distribution.

Why are these hard problems in computer vision?

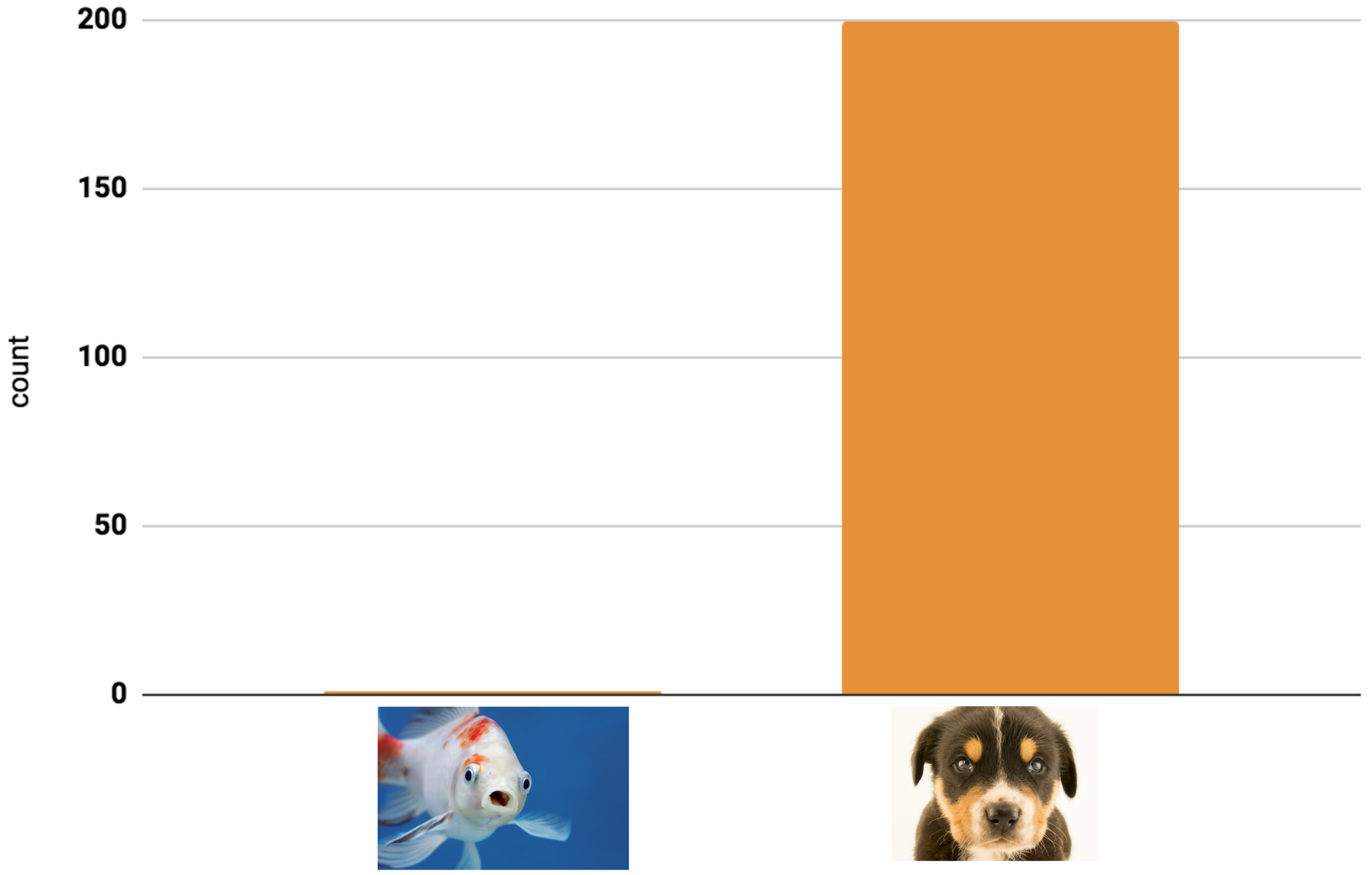


Optimization

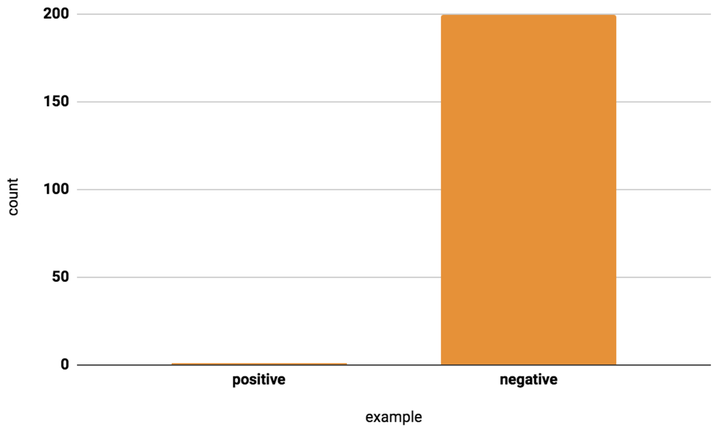


Representation

Data Imbalance

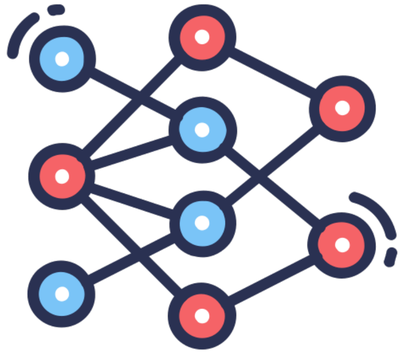

200
150
100
50
0
count

Optimization challenges

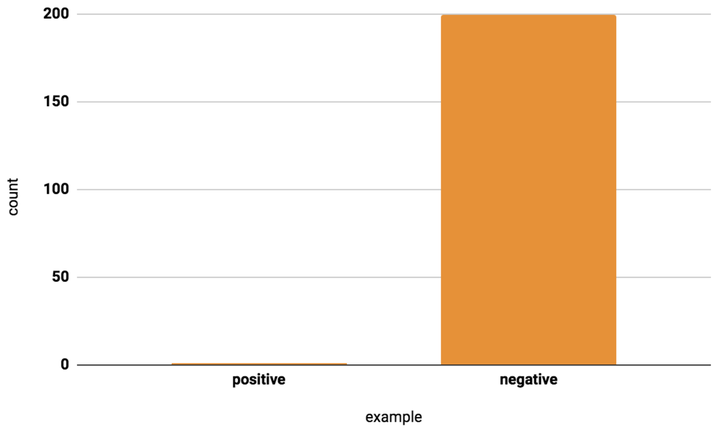


Mini batch

Loss

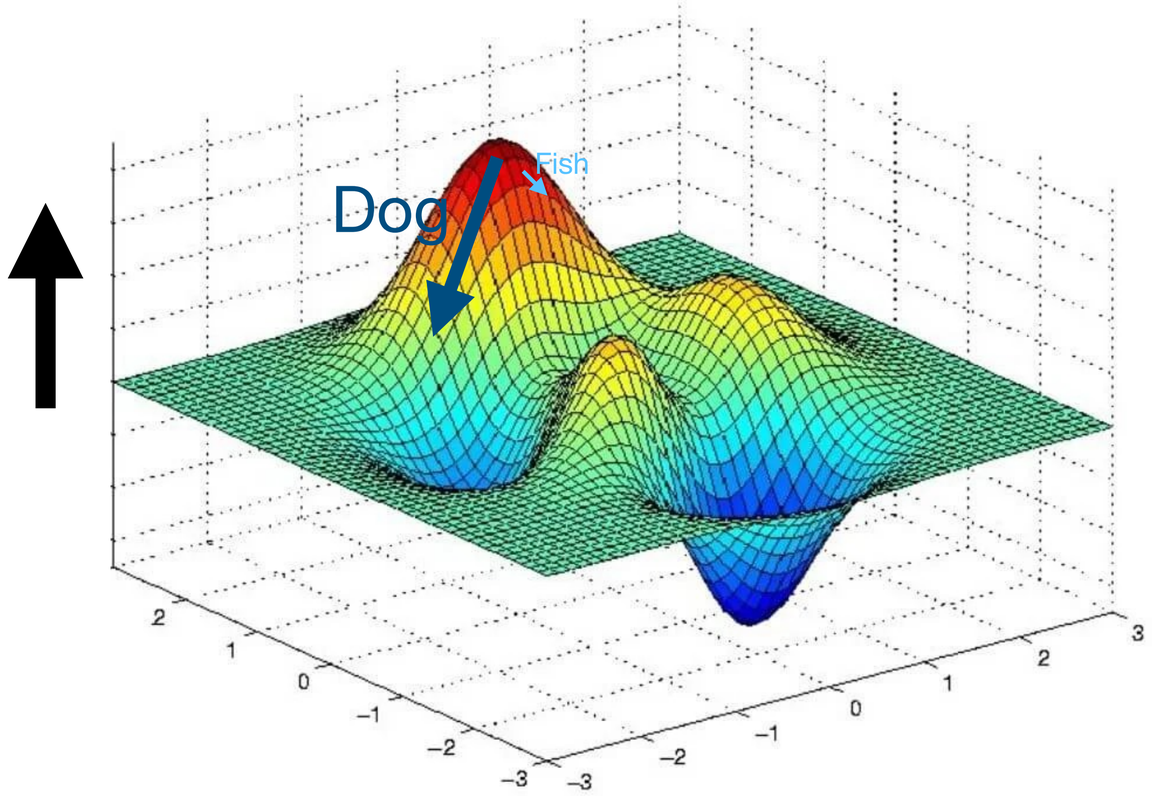


Optimization challenges

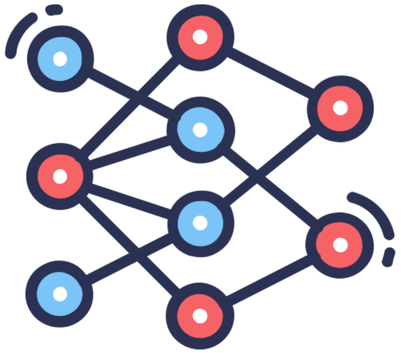


Mini batch

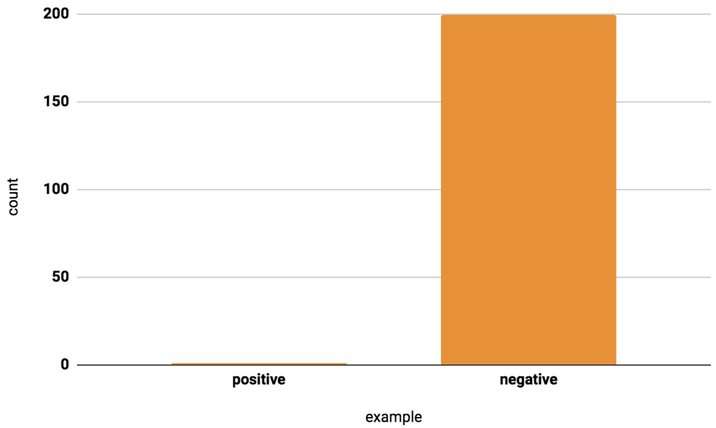
Loss

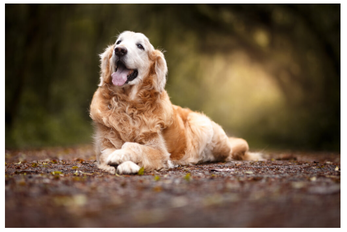


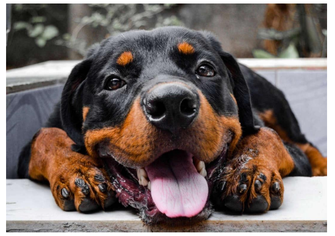
Loss



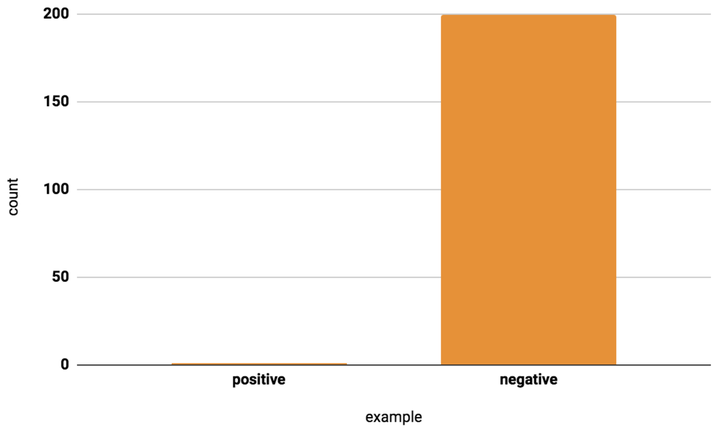
Representation challenges

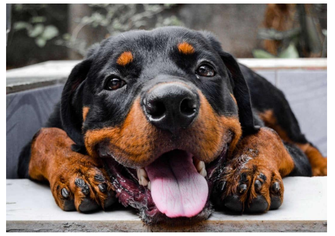


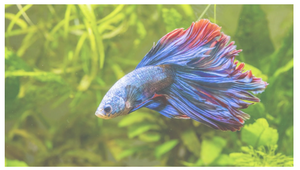




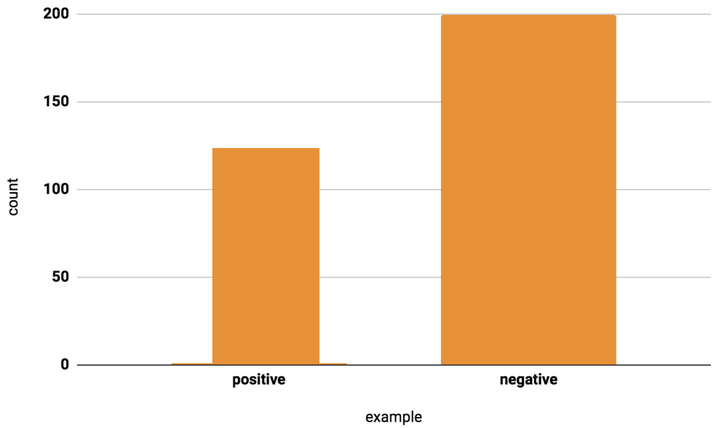
Representation challenges

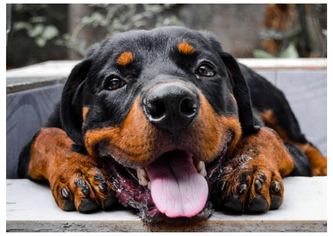


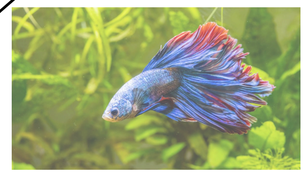




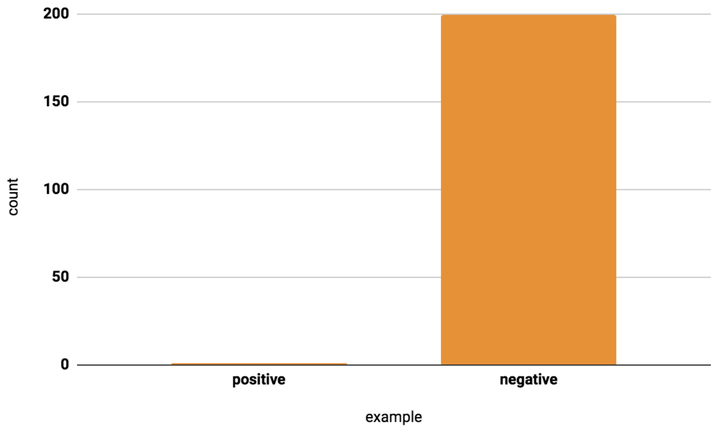
Representation challenges



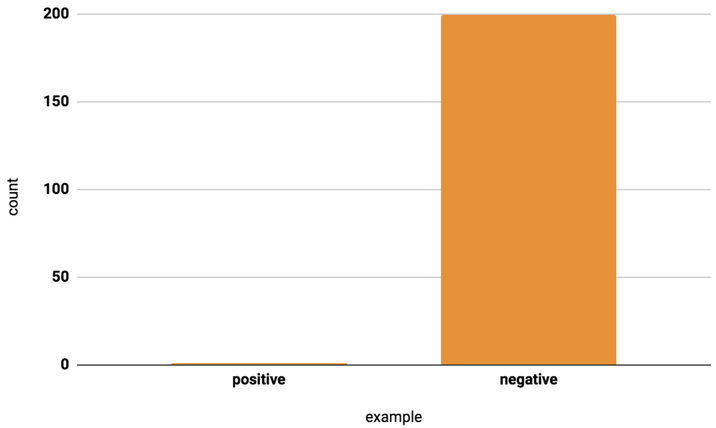


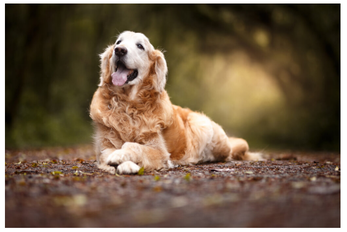


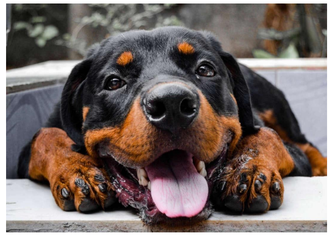
Representation challenges

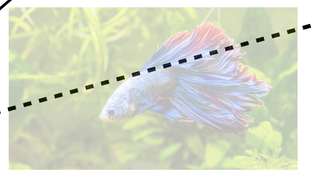


Representation challenges

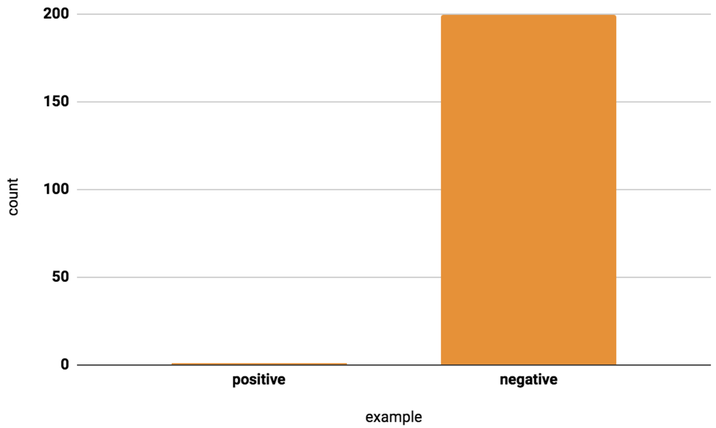


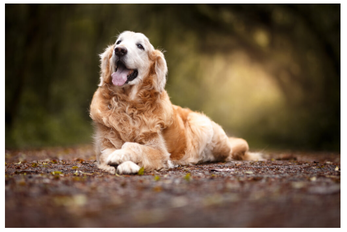


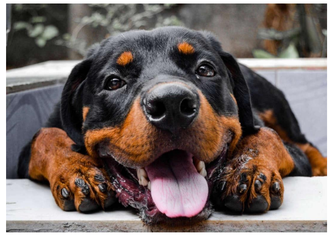




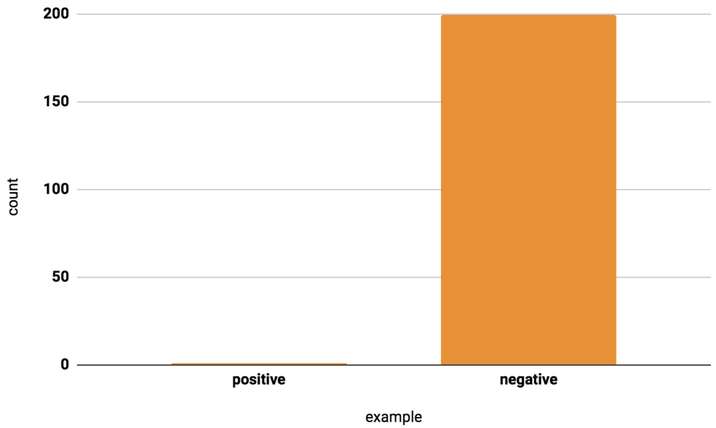
Representation challenges

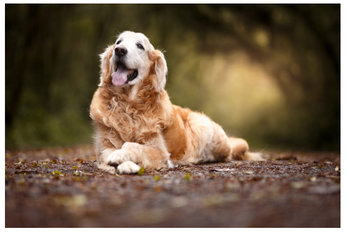


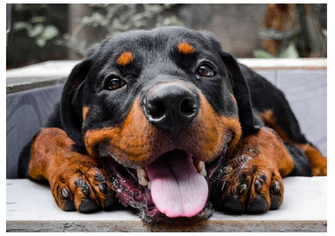


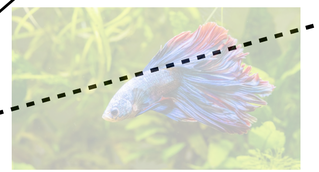


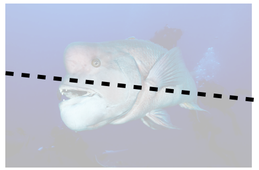
Representation challenges



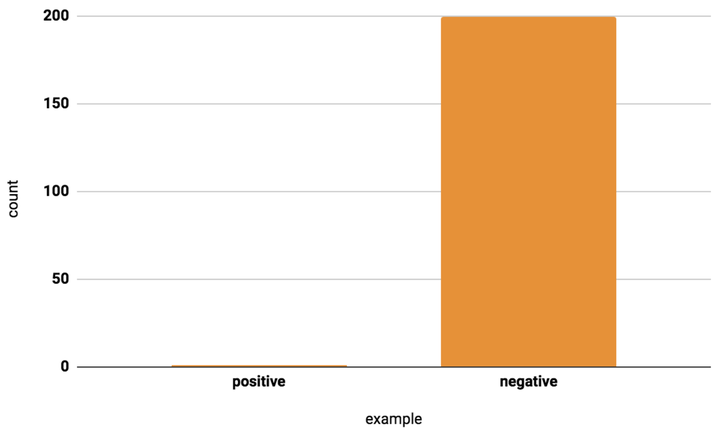


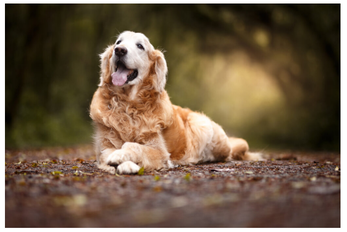


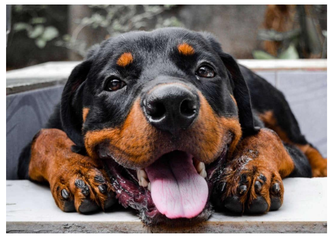


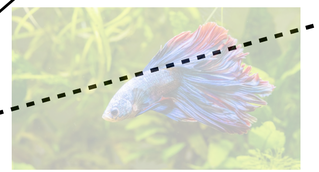


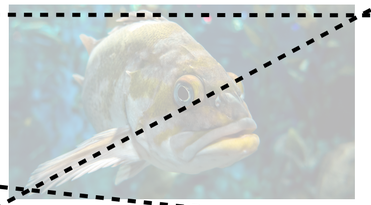
Representation challenges

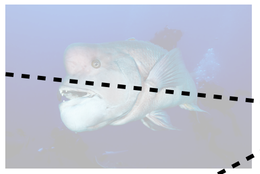


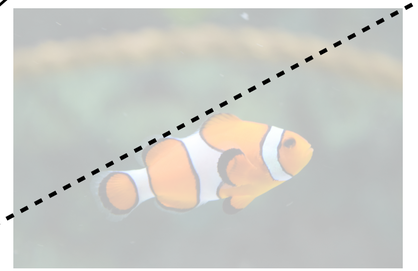








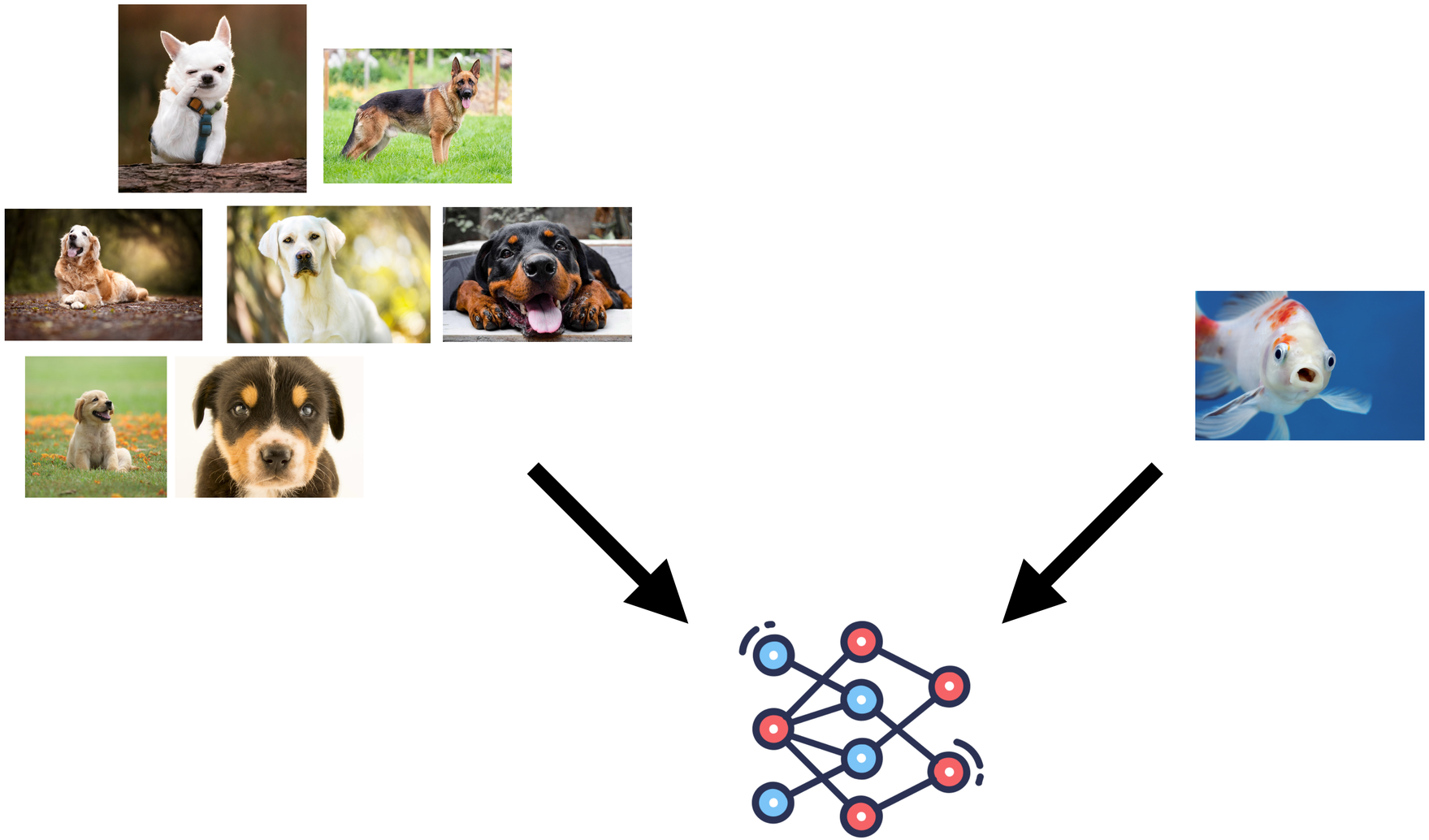




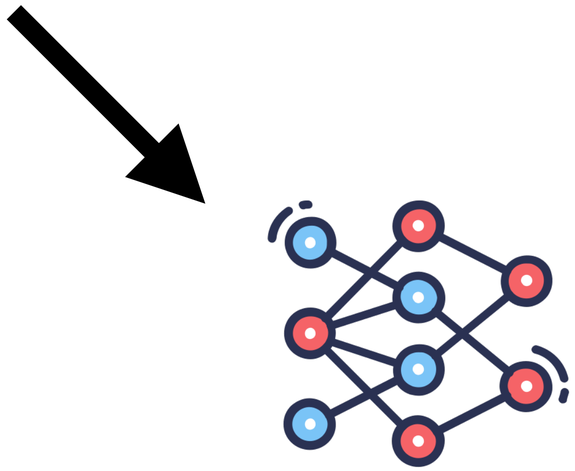
Imbalance

A few strategies

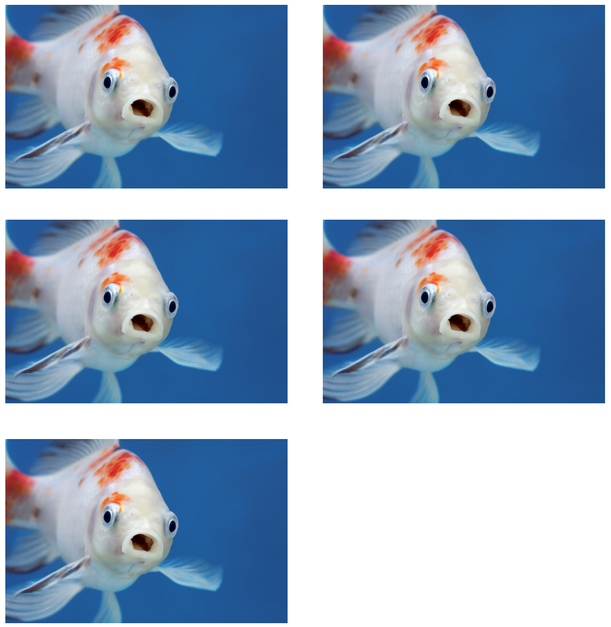
Over/under-sampling

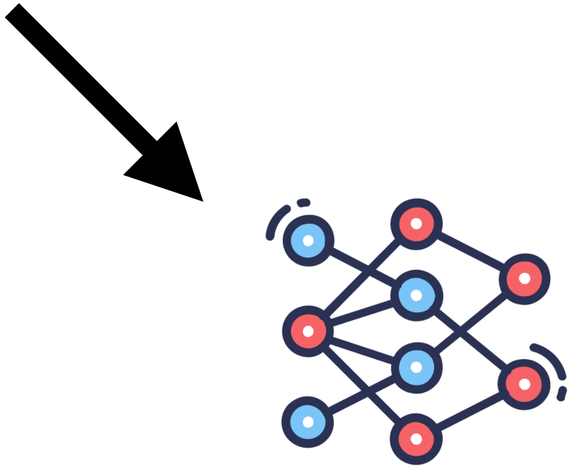


Over/under-sampling



Over/under-sampling

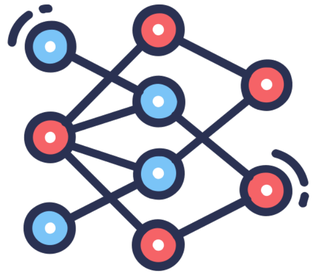




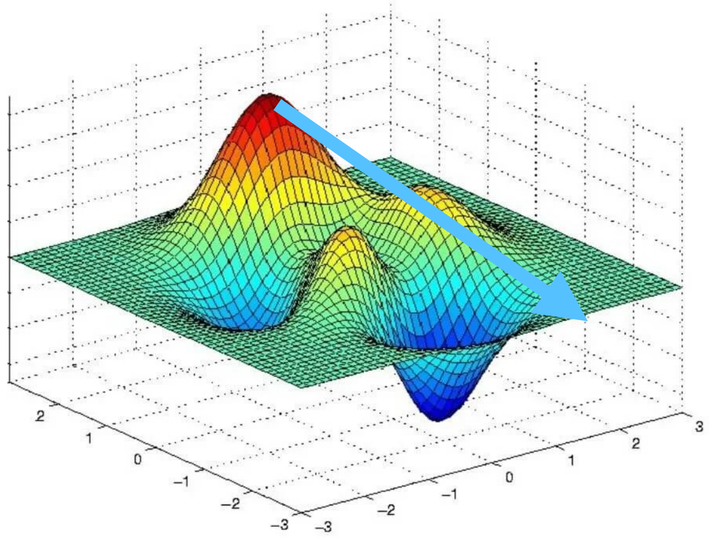
Loss weighting

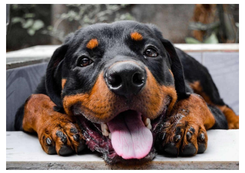
Loss x 1

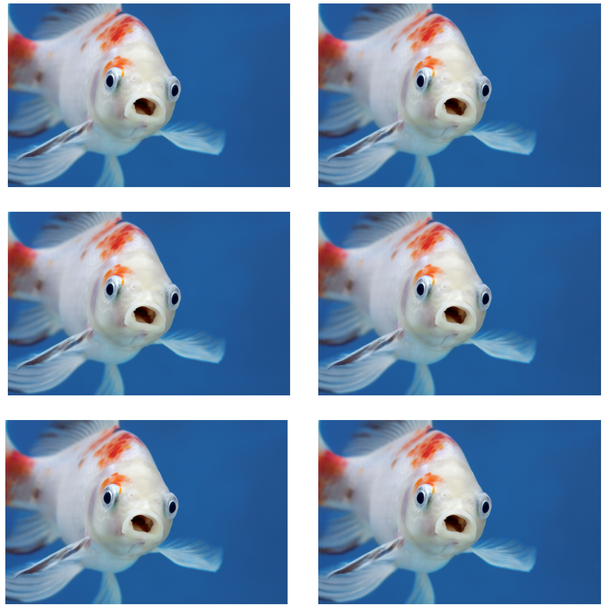
Loss x 10



Loss weighting

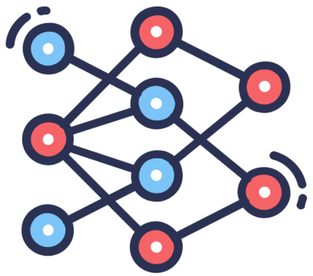




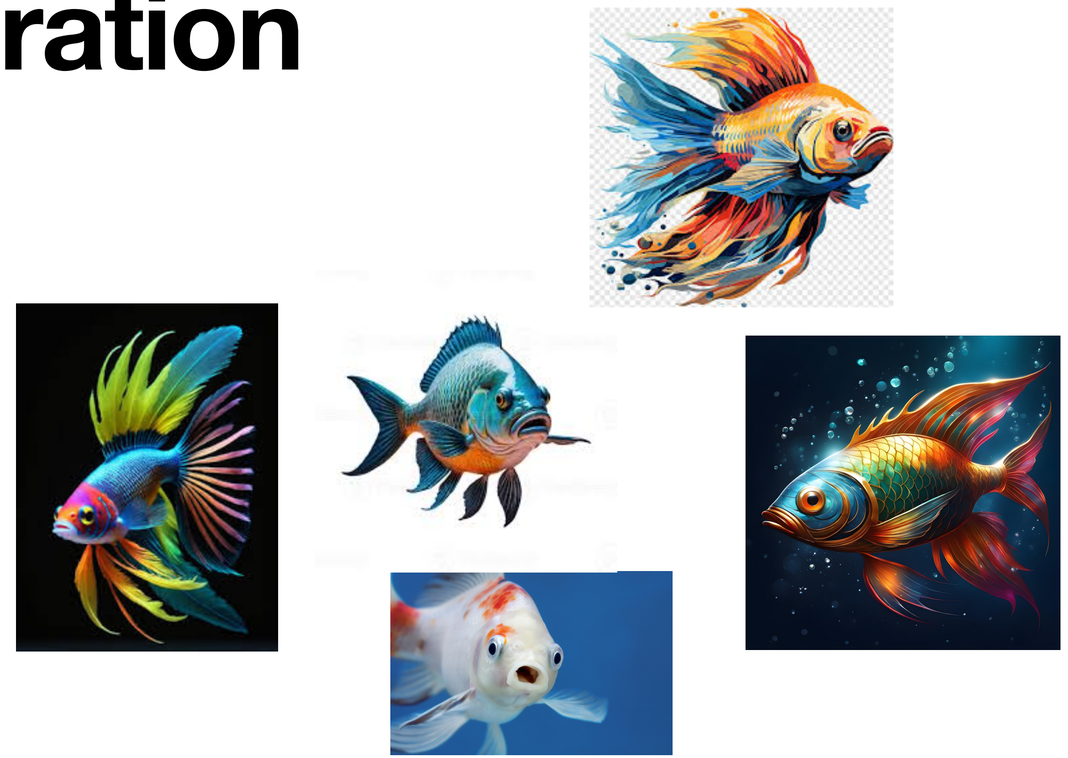


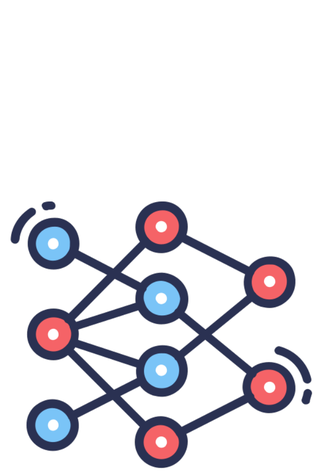
Loss x 1

Loss x 10

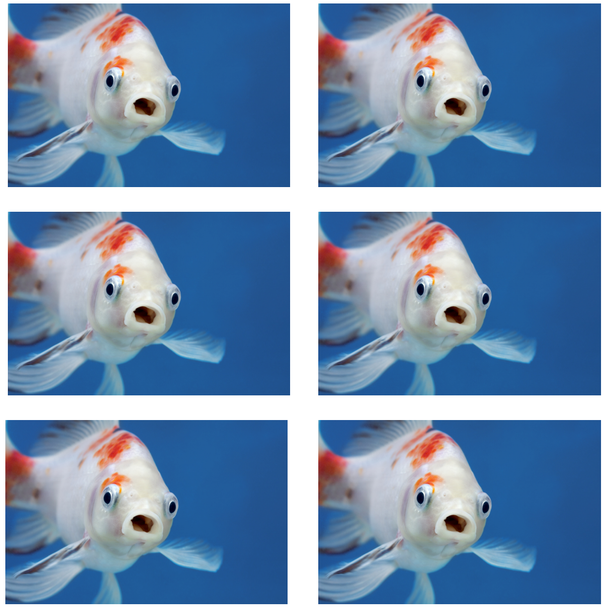


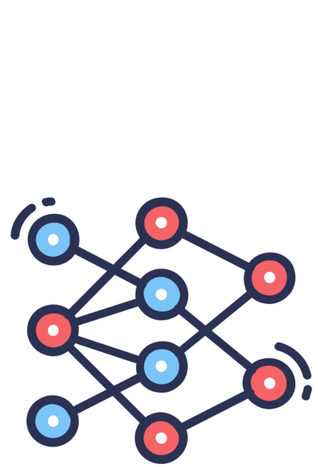
Data augmentation/generation



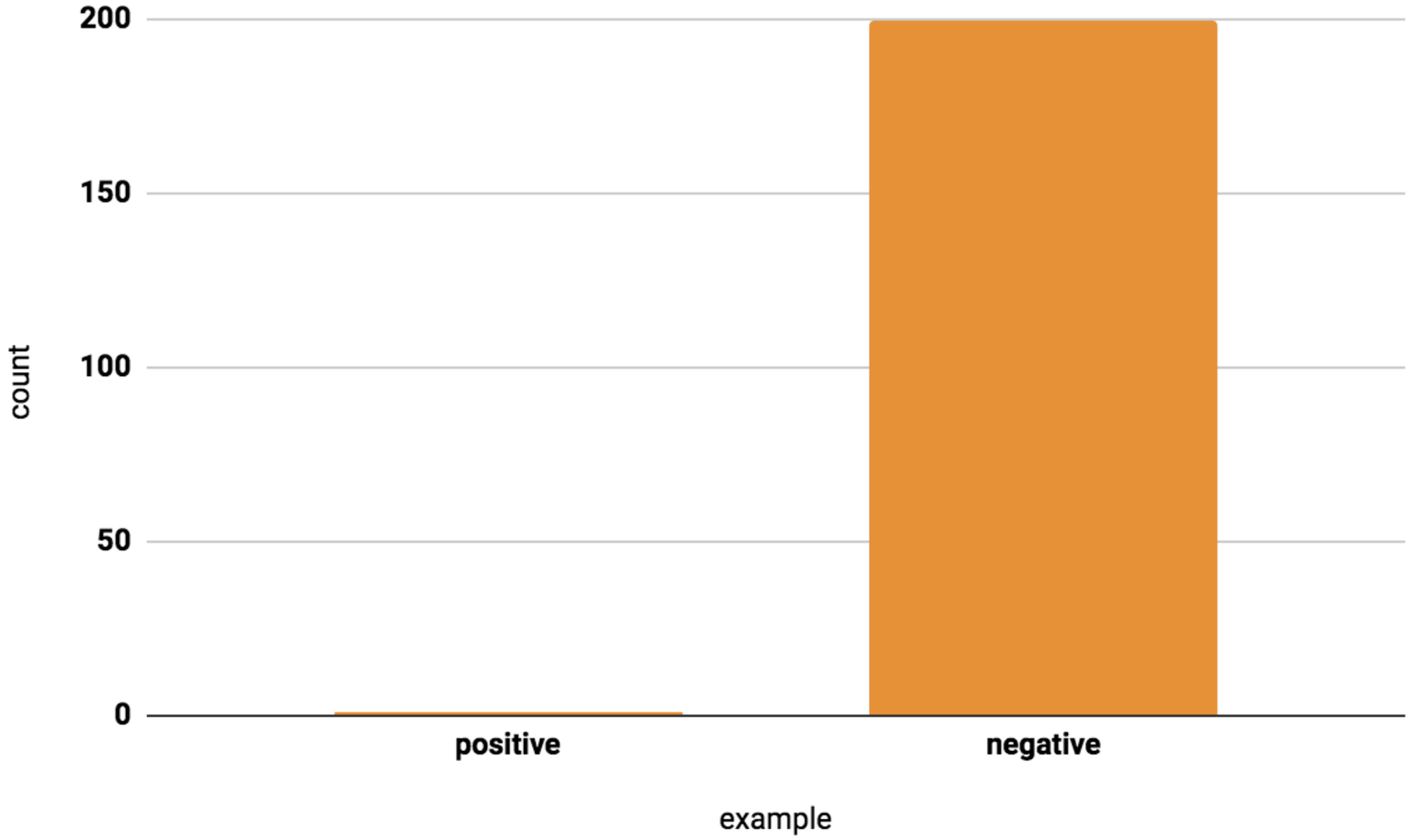


Data augmentation/generation

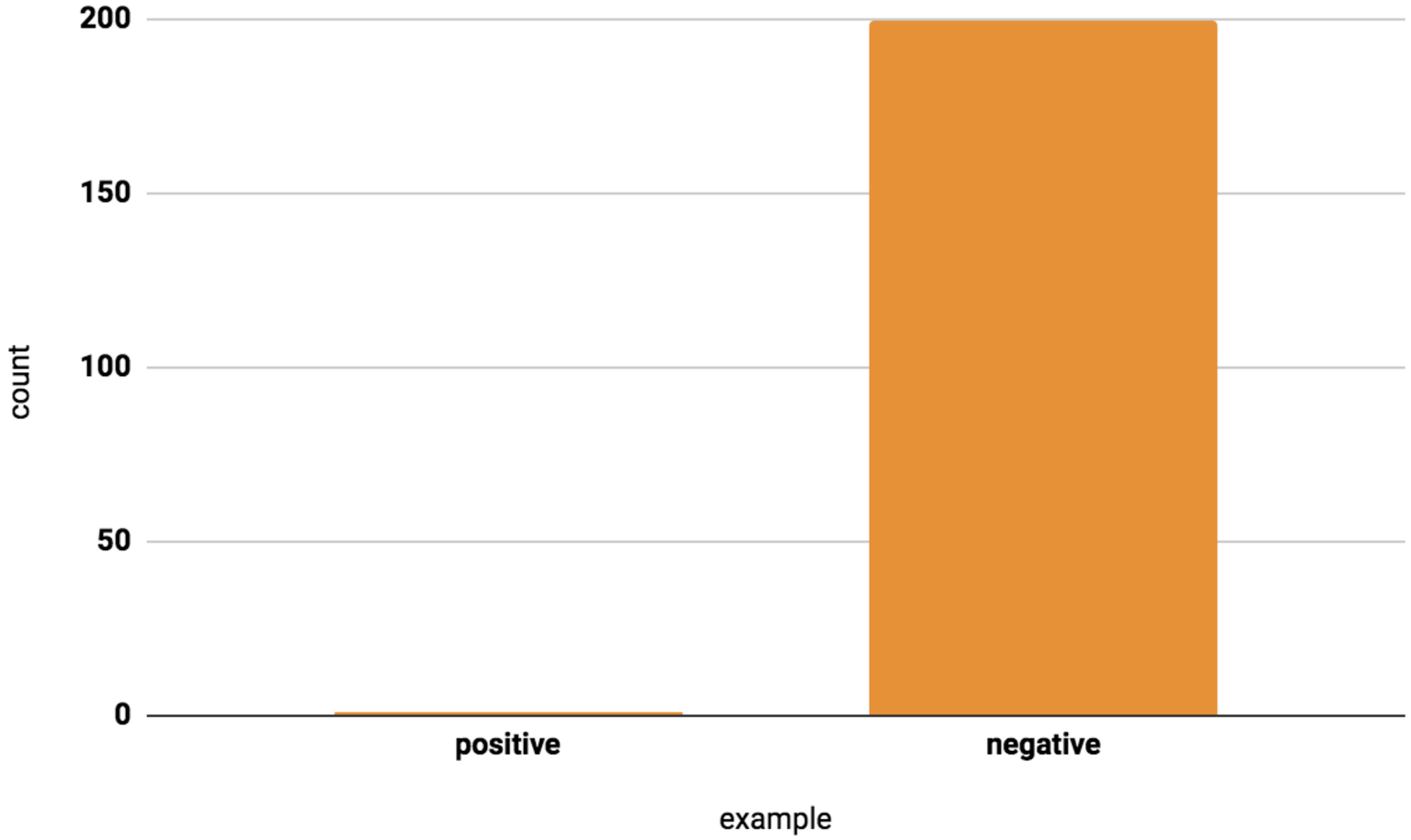




In practice: What is “good”?



In practice: What is “good”?

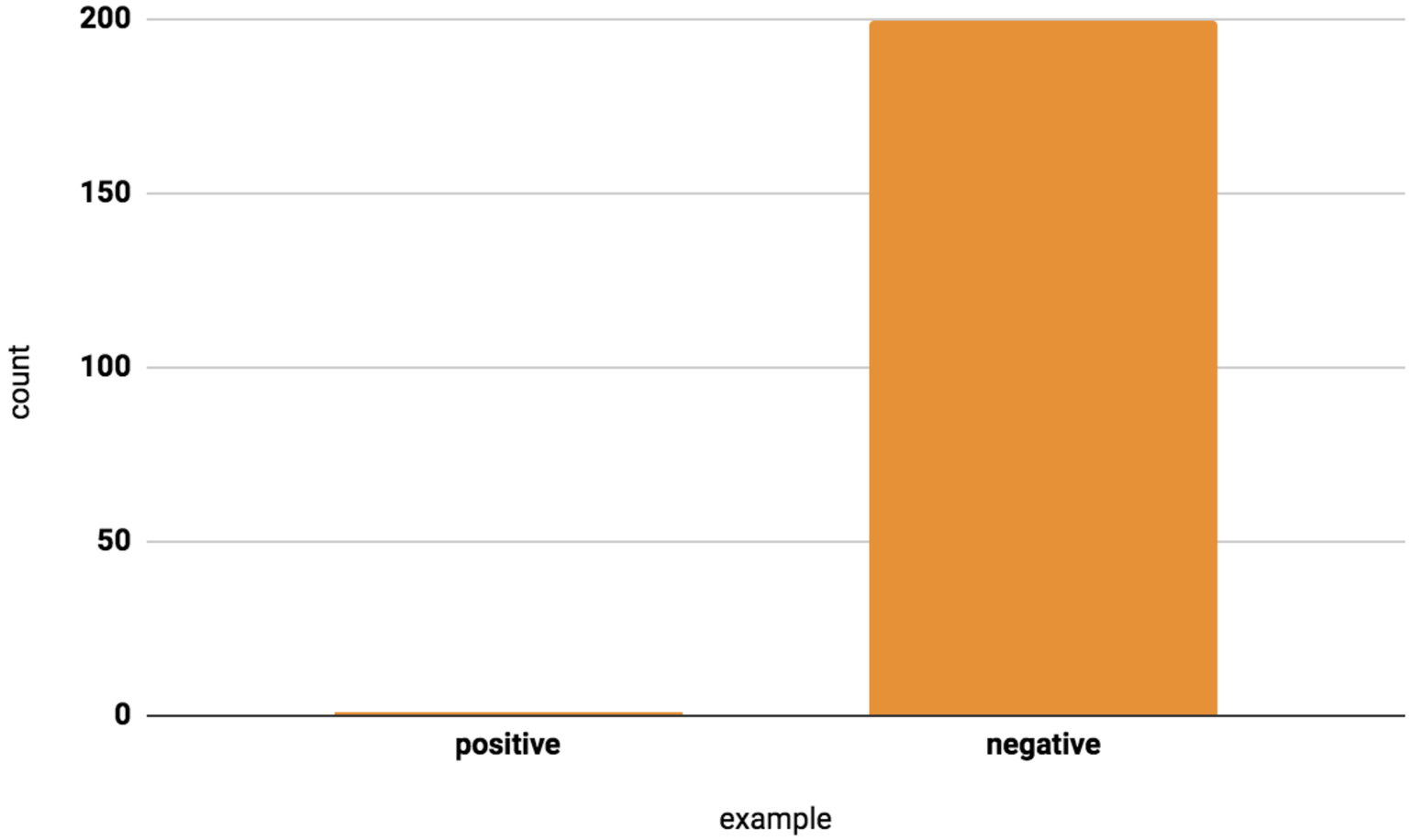


$$\text{Accuracy} = \frac{\text{Correct predictions}}{\text{All predictions}}$$

```
def my_classifier():
    return 0
```

99.5% accuracy!

In practice: What is “good”?



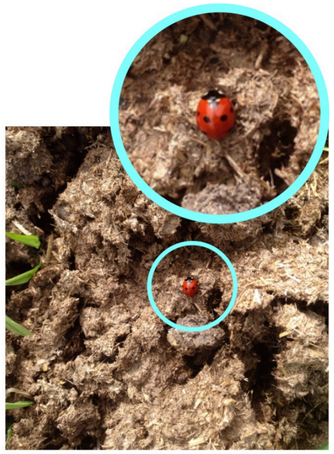
$$\text{Macro Accuracy} = \frac{\frac{\text{Correct predictions}}{\text{All predictions}}_{\text{class 1}} + \frac{\text{Correct predictions}}{\text{All predictions}}_{\text{class 0}}}{\text{Num classes}}$$

```
def my_classifier():
    return 0
```

50.25% accuracy

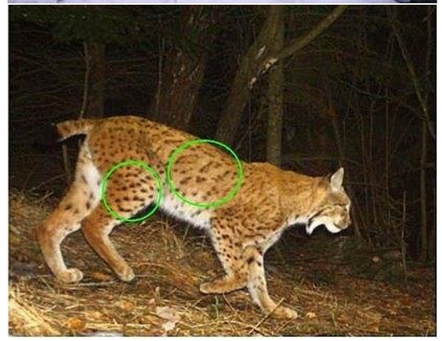
Fine-Grained Classification

Two-spotted ladybug
Adalia bipunctata



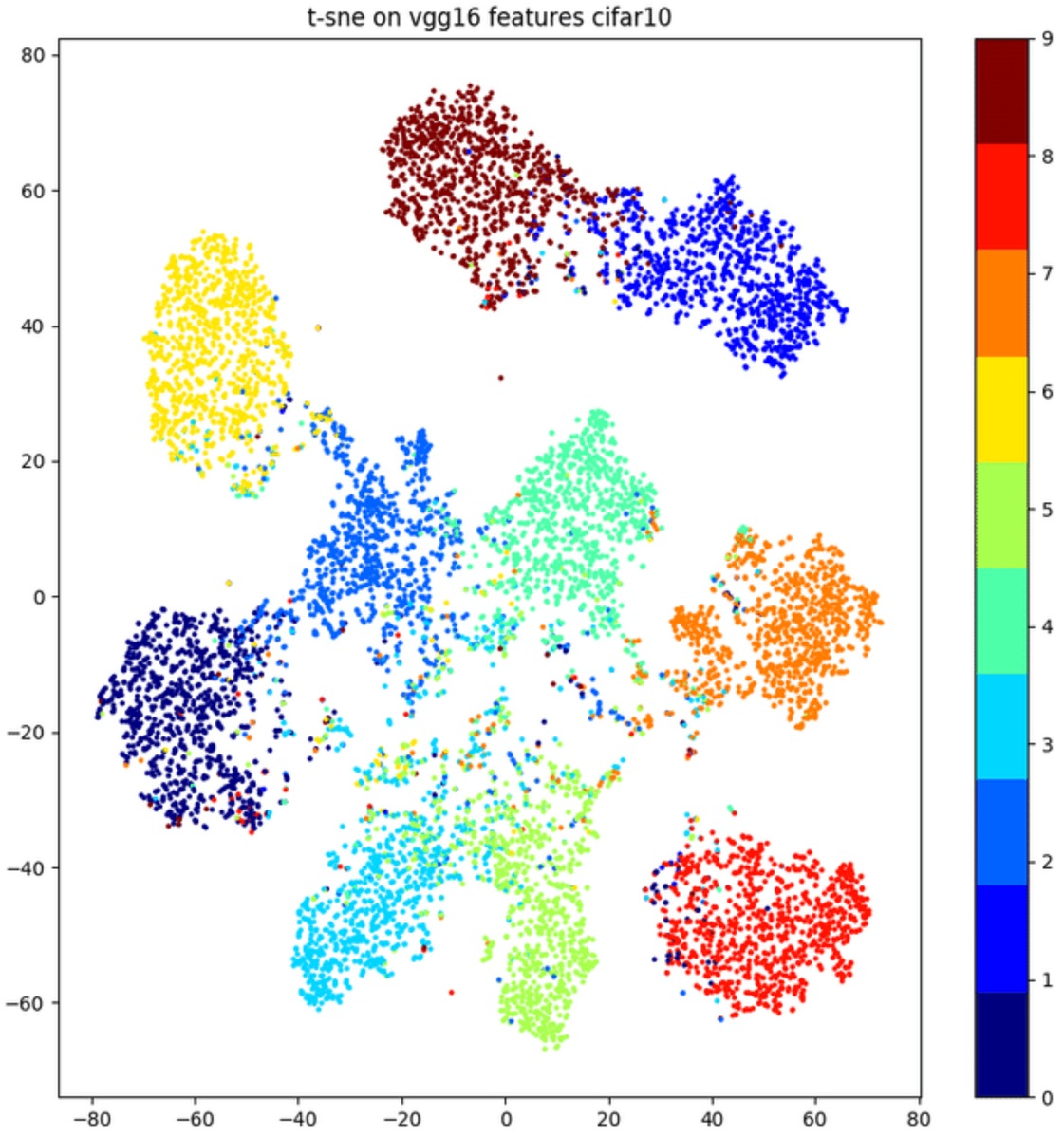
Seven-spotted ladybug
Coccinella septempunctata

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*.

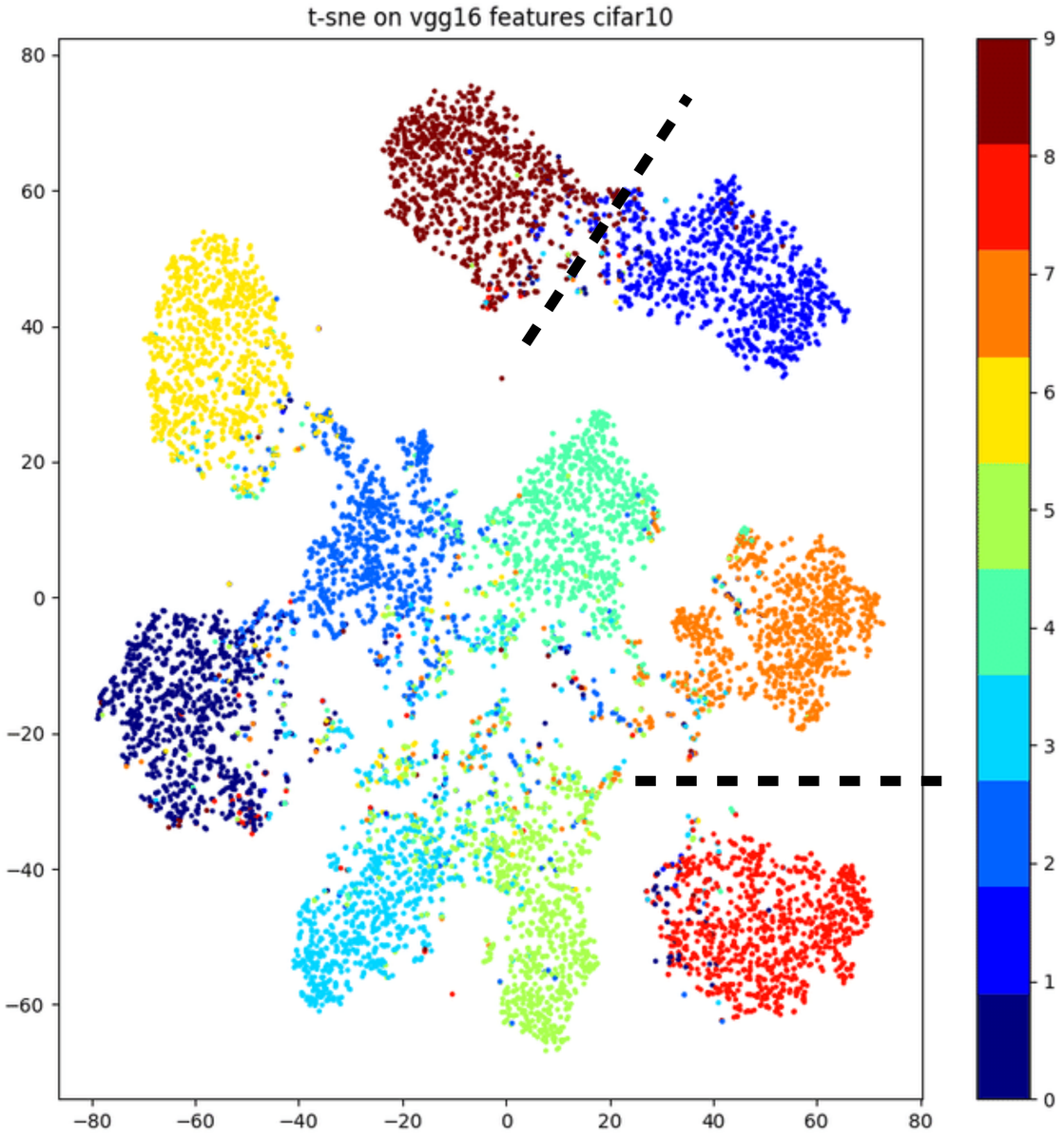


Granularity spectrum

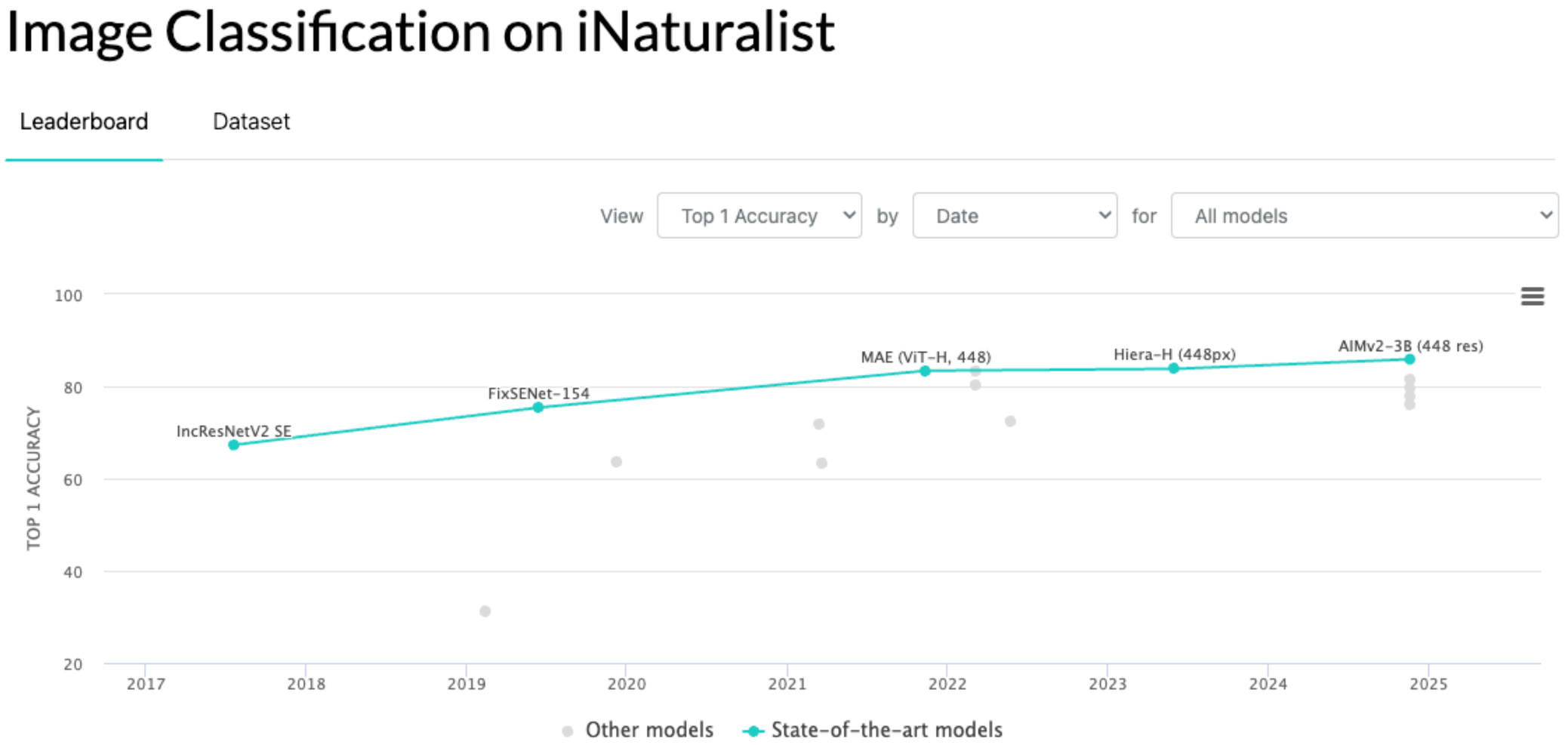
Representation challenges



Representation challenges

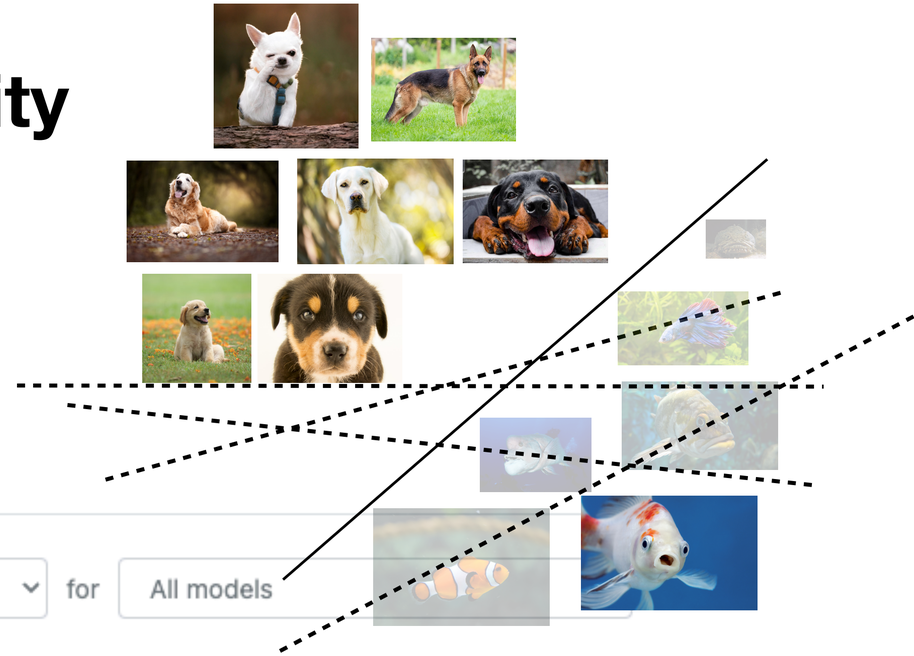


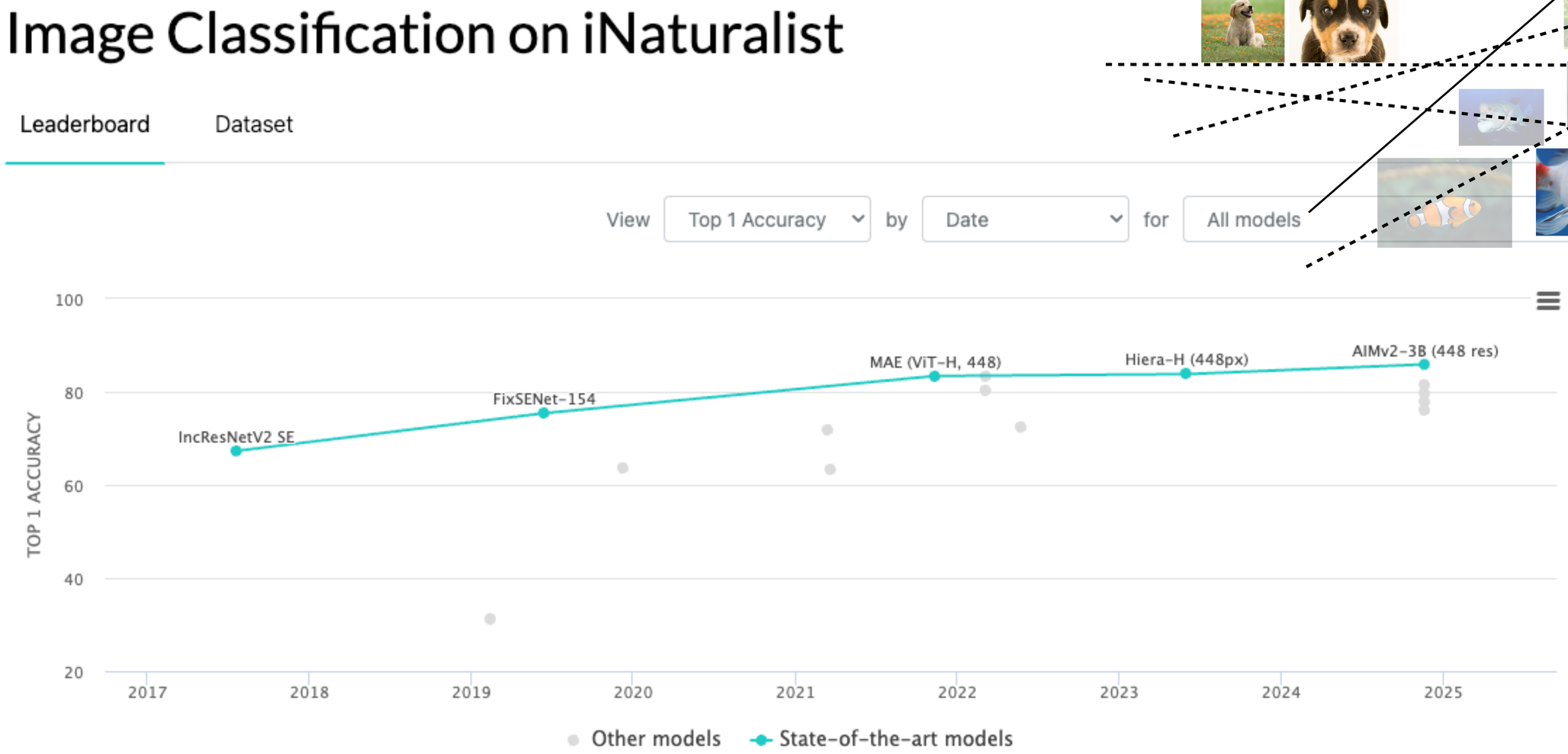
Simple solution: More representation capacity



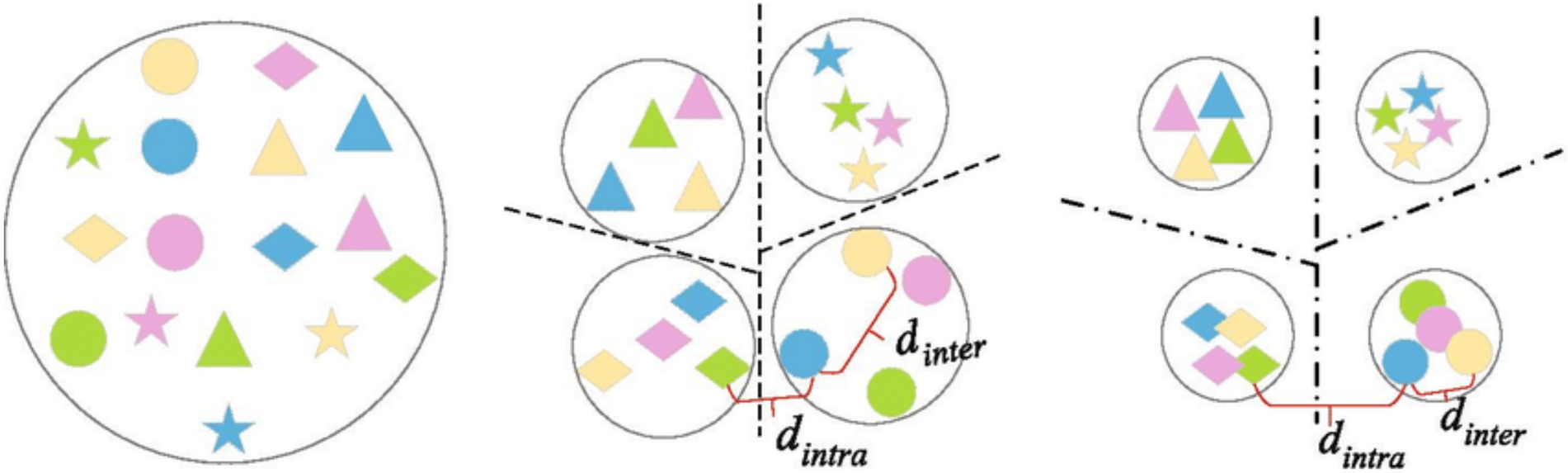
Simple solution: More representation capacity

Caveat: need enough data to separate the classes





How do we get here?



(a) Initial Data

(b) Cross-Entropy Loss

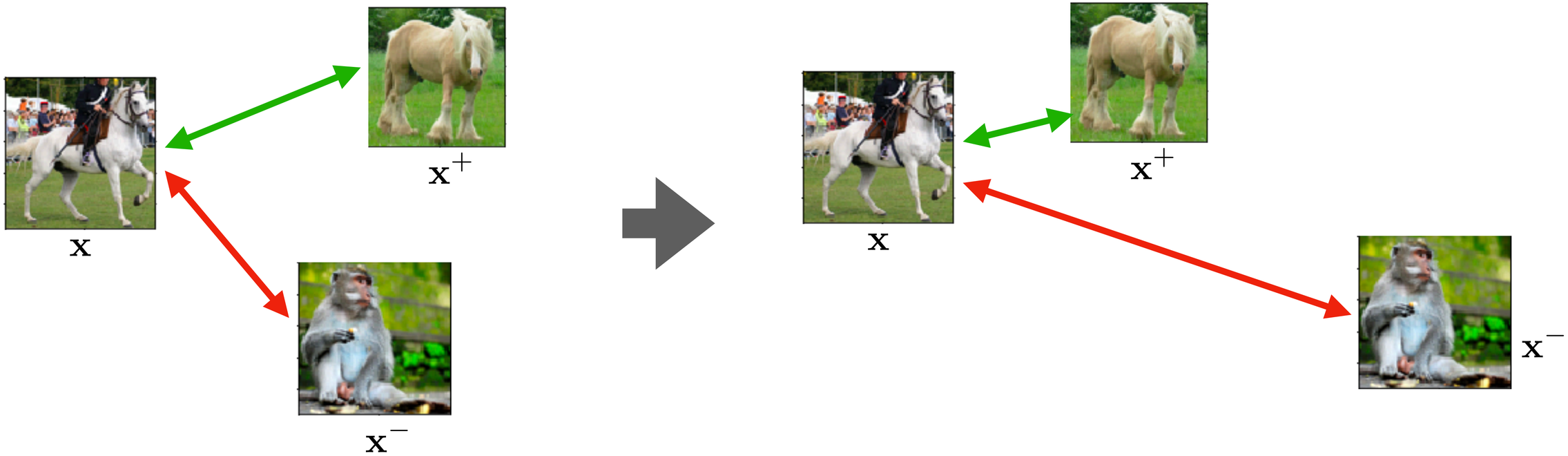
Metric / Contrastive Learning

distance of dissimilar pair(s) ≫ distance of similar pair(s)

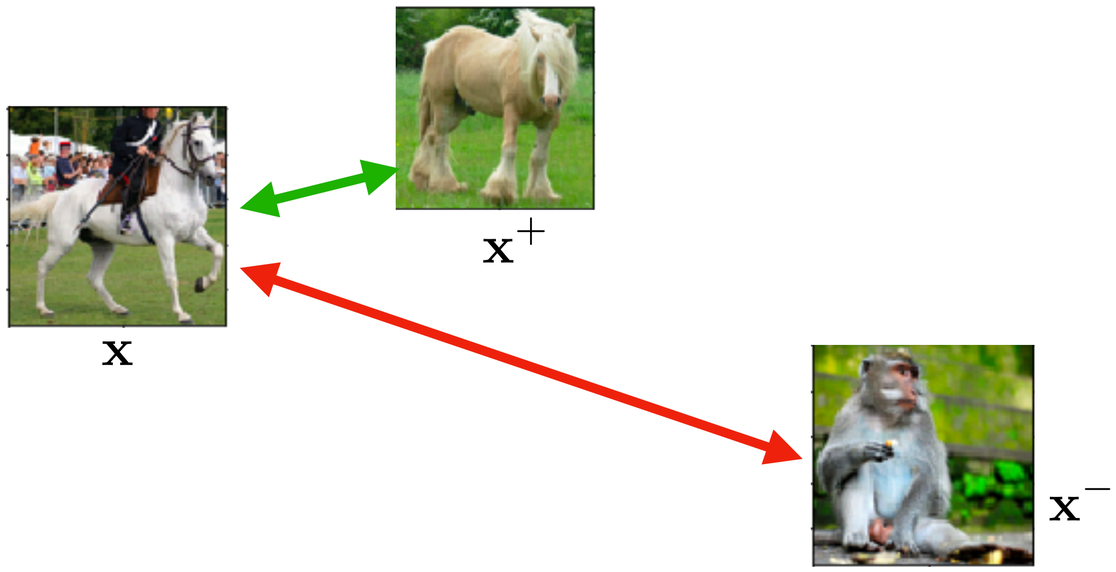
- Triplet loss *(Schroff et al 2015)*:

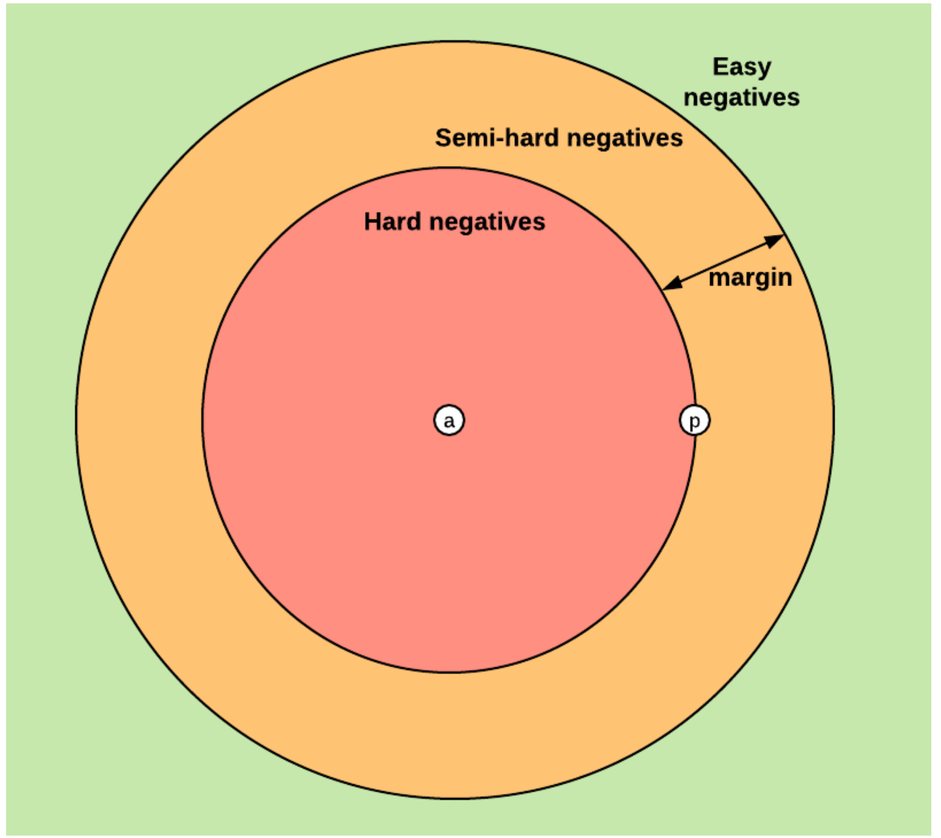
$$\mathcal{L}_{\text{triplet}}(\mathbf{x}, \mathbf{x}^+, \mathbf{x}^-) = \sum_{\mathbf{x} \in \mathcal{X}} \max\left(0, \|f(\mathbf{x}) - f(\mathbf{x}^+)\|_2^2 - \|f(\mathbf{x}) - f(\mathbf{x}^-)\|_2^2 + \epsilon\right)$$

margin



Importance of hard negatives





Beware the objective function

Cole et al., *When Does Contrastive Visual Representation Learning Work?*, CVPR 2022

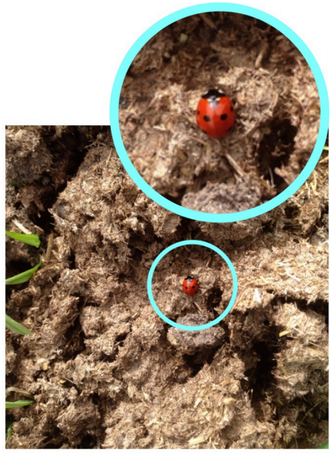
Slide: Elijah Cole

Beware the objective function

Cole et al., *When Does Contrastive Visual Representation Learning Work?*, CVPR 2022

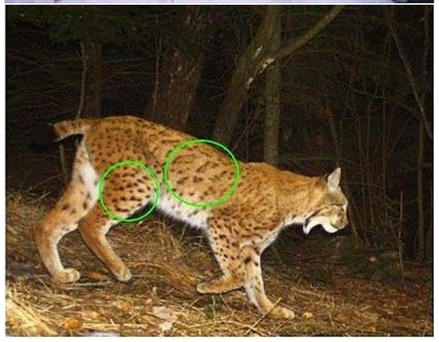
Slide: Elijah Cole

Two-spotted ladybug
Adalia bipunctata

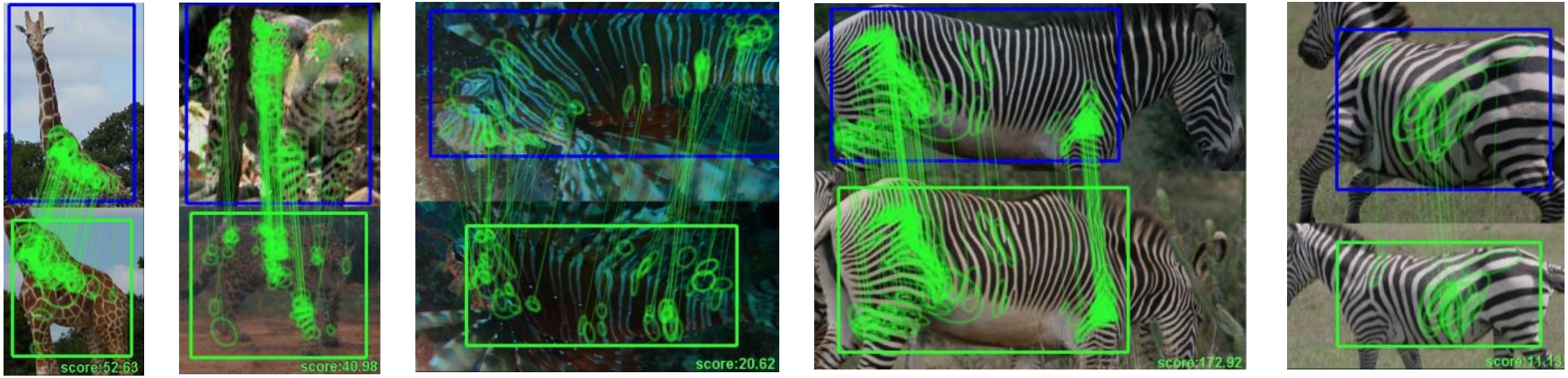


Seven-spotted ladybug
Coccinella septempunctata

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*.



Explicit feature matching



Triplet network

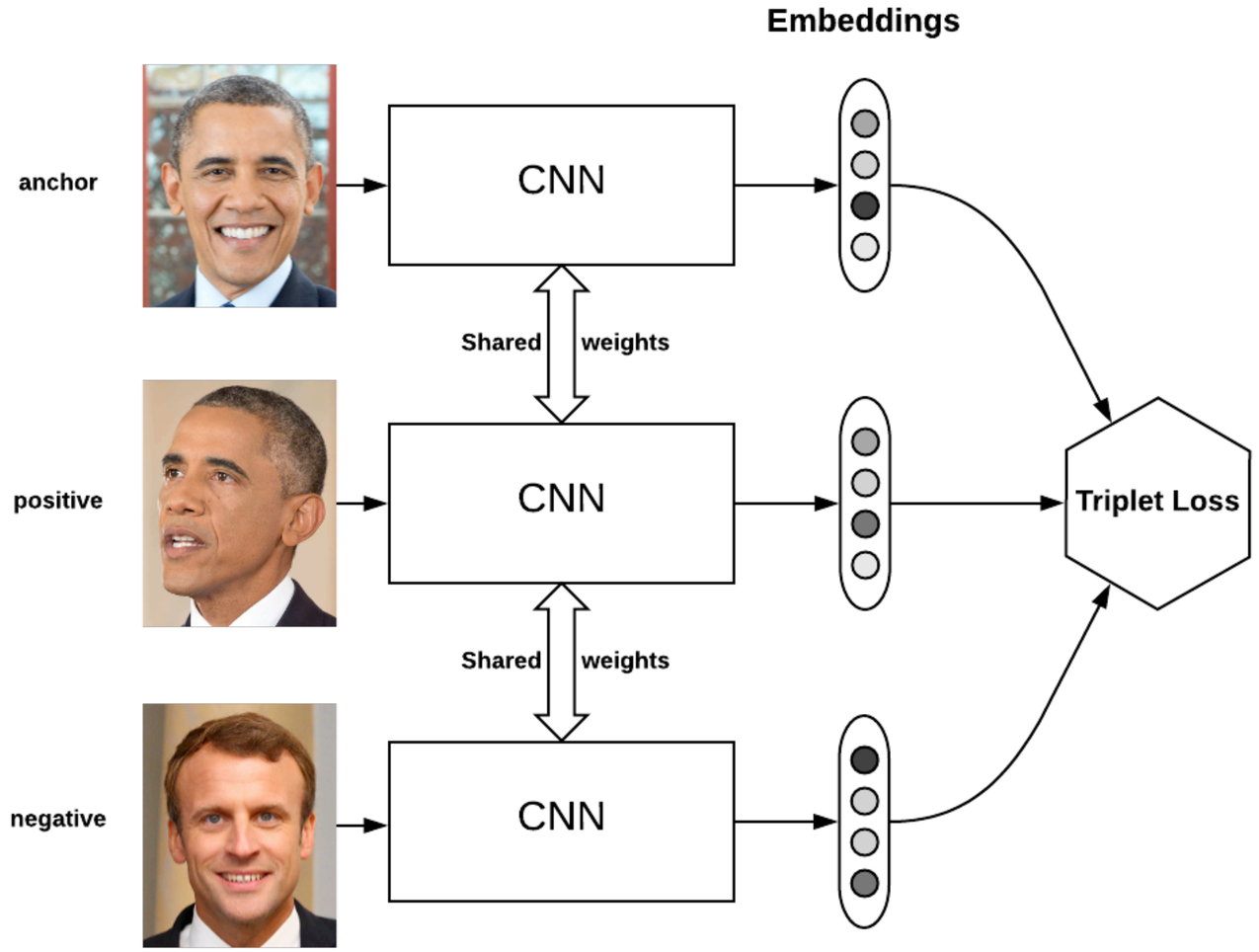


figure: https://omoindrot.github.io/triplet-loss

+ more data

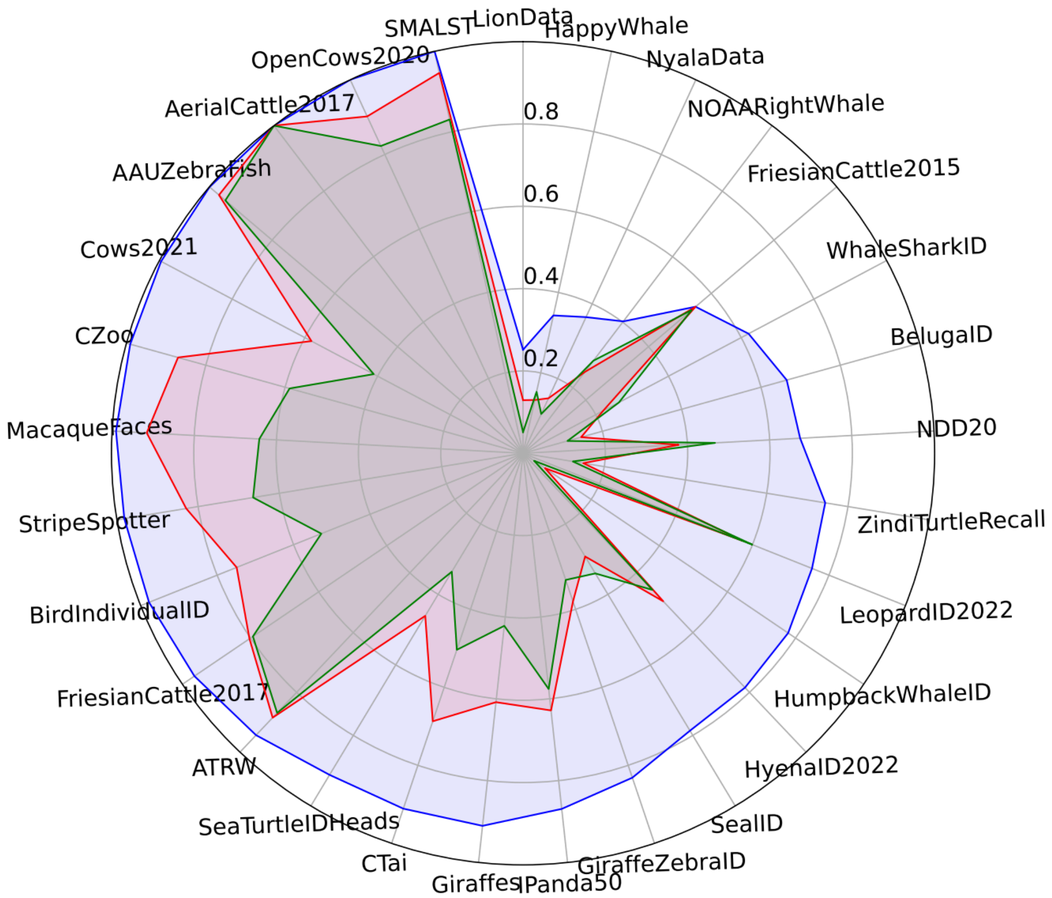


Figure 8. ***Pre-trained*** **models performance evaluation.** We compare DINOv2 (ViT-L/p14-518), CLIP (ViT-L/p14-336), and MegaDescriptor-L (Swin-L/p4-w12-384) on 29 selected datasets.

Data Imbalance and Fine-Grained Classification

6.S954 Computer Vision and Planetary Health

Justin Kay 02/20/25