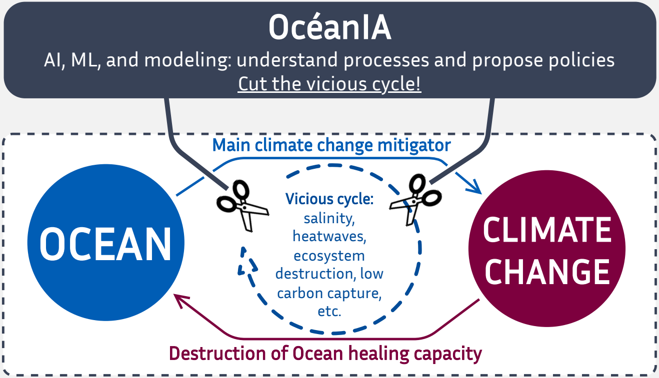
Explainable AI for Understanding Plankton Communities

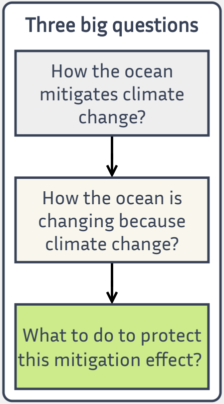
The case of plankton image identification

TAU and **Inria Chile**

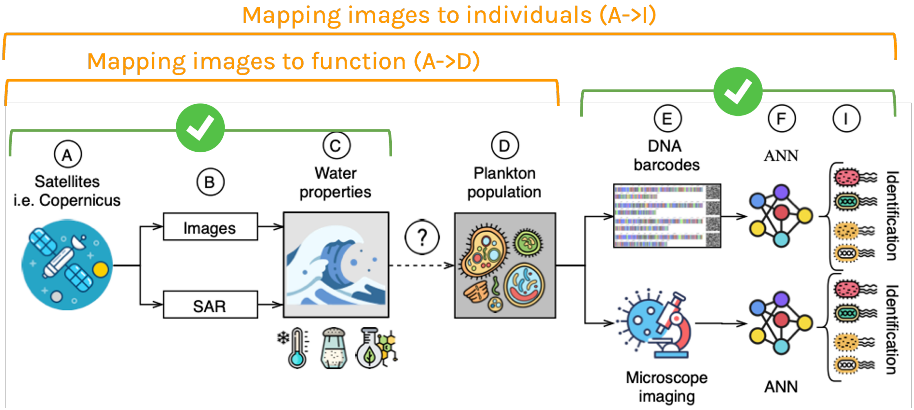
OcéanIA Annual Meeting
23 February 2024

Project OcéanIA: General goals





Towards remote identification of ocean ecosystems

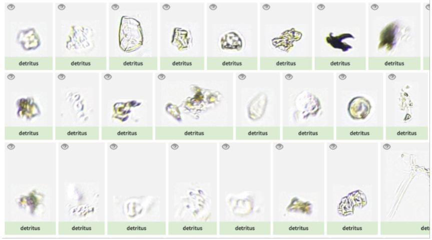


Identifying plankton species

- Plankton images are easier and cheaper to obtain than genetic sequences (DNA, RNA, etc.),
- but harder to annotate -> require experts.
- different cameras, expeditions, processing, etc.

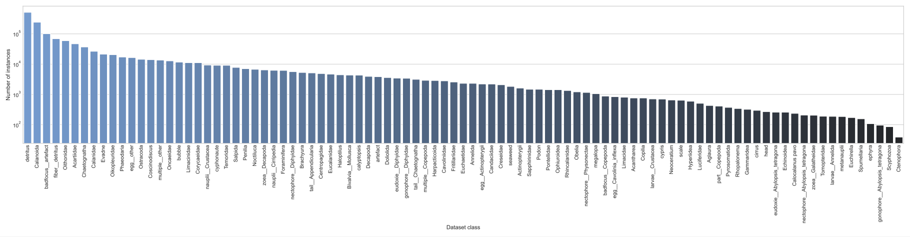
For instance, EcoTaxa:[1]

- **+320 million images**
- ≈ 4 million tagged as ‘living’
- ≈ 8 million as ‘non-living’

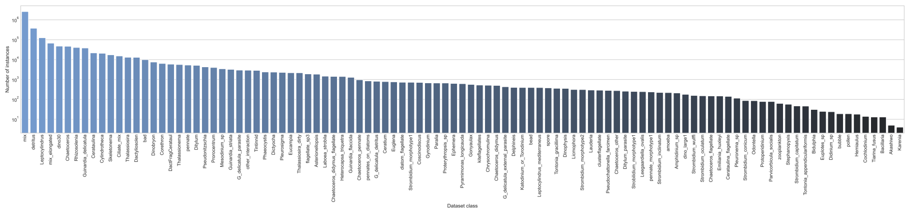


[1] Picheral, M., Colin, S., & J.-O., I. (2017). EcoTaxa, a tool for the taxonomic classification of images

Zooscan dataset (≈ 1.4M EcoTaxa subset, high quality)[2]



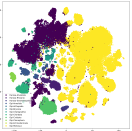
Woods Hole Oceanographic Inst. (WHOI) (≈ 3.6M, B&W/low quality)[3]



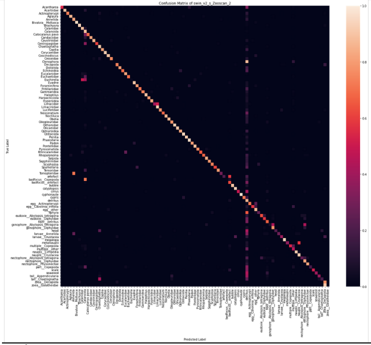
[2] Amanda, E., Corinne, D., Laetitia, J., Marion, O., Jean-Baptiste, R., Manoela, C. B., Fabien, L., , Natalia, L., Justine, C., Louis, C.-C., Bruno, S., Jean-Olivier, I., Gaby, P. M. G., & Lars, S. (2017). Zooscannet: plankton images captured with the zooscan. *SEANOE*

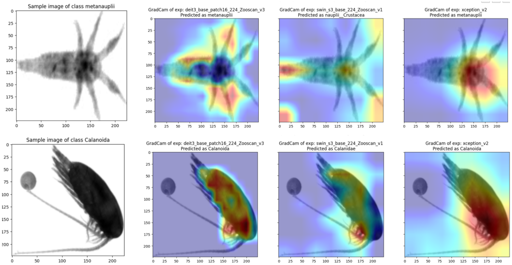
[3] Orenstein, E. C., Beijbom, O., Peacock, E. E., & Sosik, H. M. (2015). Whoi-plankton- A large scale fine grained visual recognition benchmark dataset for plankton classification. *CoRR*, abs/1510.00745

First step: the supervised way



Xception,[4] Efficientnet v2,[5] ConvNeXt[6], Inception,[7] SWIN s3,[8] and DeiT[9].
Data augment, adapt weighted resampling, focal loss, etc.





[4] Chollet, F. (2016). Xception: Deep learning with depthwise separable convolutions. *CoRR*, abs/1610.02357

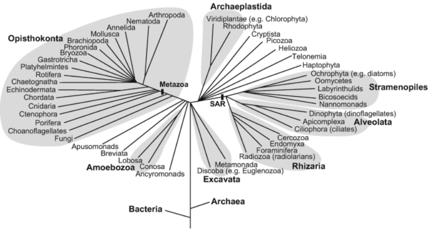
[5] Tan, M. & Le, Q. V. (2021). Efficientnetv2: Smaller models and faster training. *CoRR*, abs/2104.00298

[6] Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. *CoRR*, abs/2201.03545

[7] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S. E., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014). Going deeper with convolutions. *CoRR*, abs/1409.4842

[8] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *CoRR*, abs/2103.14030

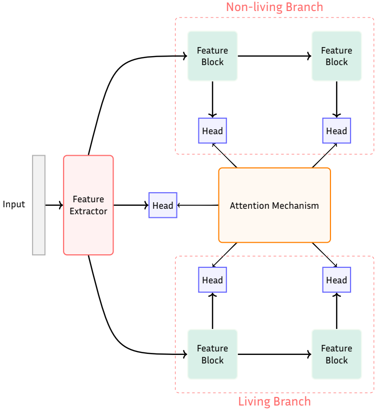
Step 2: Hierarchical classification



Intuitively, it makes sense to exploit the taxonomy information.

- Branch CNN[10]
- Hierarchical bilinear CNN[11]
- Branch-attention CNN[12]

Segregated BA-CNN model: branches for *living* and *non-living*.



[10] Zhu, X. & Bain, M. (2017). B-CNN: branch convolutional neural network for hierarchical classification. *CoRR*, abs/1709.09890

[11] Zhang, X., Tang, L., Luo, H., Zhong, S., Guan, Z., Chen, L., Zhao, C., Peng, J., & Fan, J. (2021). Hierarchical bilinear convolutional neural network for image classification. *IET Computer Vision*, 15

[12] Pizarro, I., Ñanculef, R., & Valle, C. (2023). An attention-based architecture for hierarchical classification with cnns. *IEEE Access*, 11, 32972–32995

Hierarchical classification results

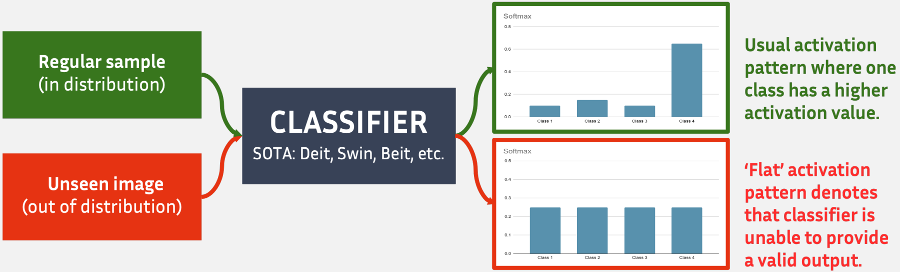
Hierarchical Experiments												
Experiment Names	First Level				Second Level				Third Level			
	Acc	F1	Pre	Rec	Acc	F1	Pre	Rec	Acc	F1	Pre	Rec
Flat Baseline	X	X	X	X	X	X	X	X	88.5	89.6	91.2	88.5
B-CNN	97.1	97.1	97.1	97.1	95.8	95.8	95.1	94.7	92.3	92.3	92.6	90.6
HB-CNN	97.2	97.2	97.2	97.2	95.6	95.5	94.8	94.0	92.0	91.9	91.8	89.0
BA-CNN	98.7	98.7	98.7	98.7	95.9	95.8	95.5	94.9	92.2	92.1	92.3	89.5
SBA-CNN	97.2	97.2	97.2	97.2	90.2	90.8	89.1	90.2	83.7	88.4	89.7	83.7

Hierarchical results for individual branches										
Experiment Names	Specific Branch		Second Level				Third Level			
	Relevant	Irrelevant	Acc	F1	Pre	Rec	Acc	F1	Pre	Rec
B-CNN	X		98.9	99.6	94.3	98.9	94.5	96.5	95.3	94.5
HB-CNN	X		98.9	99.5	94.0	98.9	93.2	96.4	95.0	93.1
BA-CNN	X		99.2	99.7	96.8	99.2	93.1	96.5	95.3	93.1
SBA-CNN	X		99.2	99.7	99.2	99.2	93.1	96.2	91.4	93.1
B-CNN		X	91.9	94.2	87.7	91.9	83.8	90.1	87.0	83.1
HB-CNN		X	91.1	94.1	87.8	91.1	80.7	89.8	87.1	80.7
BA-CNN		X	92.0	94.4	91.7	92.0	83.1	90.1	88.0	83.1
SBA-CNN		X	89.8	93.9	92.6	89.8	80.0	89.5	85.3	80.0

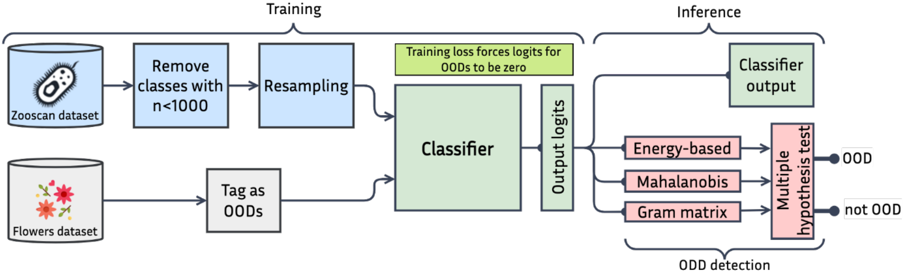
Step 3: What about the masses of unlabeled data?

- Out-of-distribution detection methods analyze the output of a classifier to determine if the input is OOD or ‘unknown’.
- The definition of OOD-ness -as outliers- is somewhat intuitive and based on context.
- Use existing annotated datasets to evaluate the quality of anomaly detection.

OOD patterns in classifiers

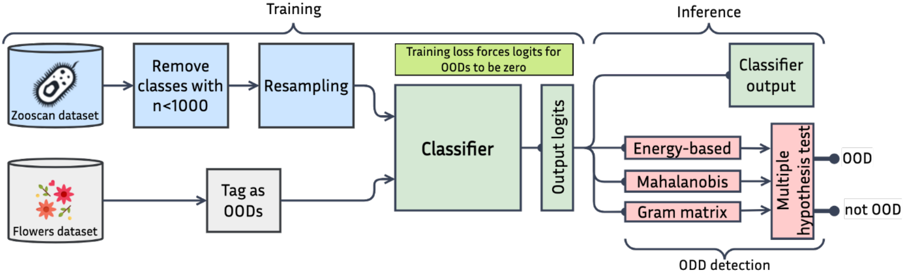


Using adversarial examples for OOD detection



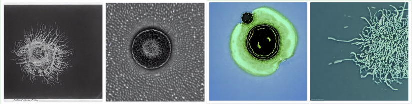
- Use ‘known OODs’ and induce ‘flat logits’ behavior via loss function.
- Leave one class out and use it for OOD evaluation (for all classes).
- OOD detection bumped from 58% to 93% retaining classifier performance!

Using adversarial examples for OOD detection

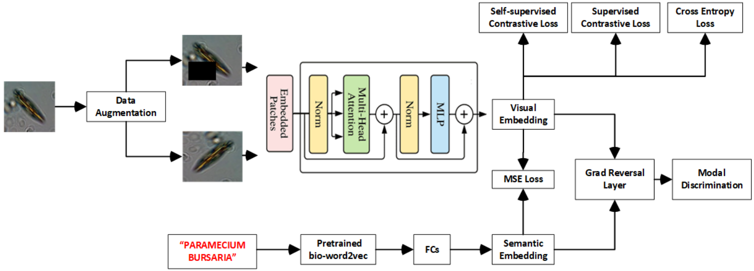


- Use 'known OODs' and induce 'flat logits' behavior via loss function.
- Leave one class out and use it for OOD evaluation (for all classes).
- OOD detection bumped from 58% to 93% retaining classifier performance!

Current: Hey, AI, generate some (fake/adversarial) plankton examples to use as OODs!



Multi-modal/multi-task approach



- **Classification:** known plankton species in training dataset.
- **Zero-shot classification:** known plankton species <u>not</u> in dataset.
- **Anomaly detection:** beyond OOD with contrastive learning.

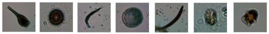


Figure 1: Examples from Lensless dataset.


Figure 2: Examples from the newly added chimeras images.


Figure 3: Examples from the newly added coco images.

Multi-modal/multi-task learning: zero-shot results

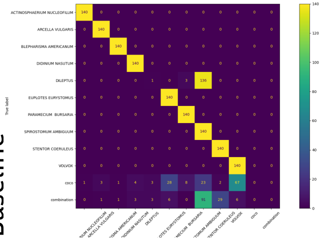
- We then report the results of zero-shot learning for novel plankton species detection of different methods.
- We use the Top-k accuracy to evaluate the zero-shot learning.

$$\text{top-k acc}(y, f) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{k}\mathbf{1}(f_{i,j} == y_i) \tag{5}$$

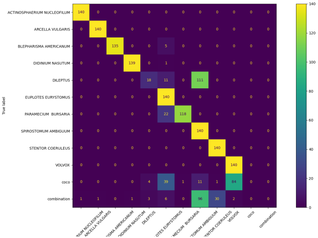
where $f_{i,j}$ is the predicted class for the i-th sample corresponding to the j-th largest predicted score. y_i is the corresponding true label. N is the number of samples, k is the number of gueses, and **1** is the indicator function.

Methods	Top-1	Top-2	Top-5	Top-10
Proto-net + Constrasive	0.71	96.43	100.0	100.0
w/o Supervised Contrastive	1.43	96.43	100.0	100.0
Full model	12.86	100.0	100.0	100.0

Baseline



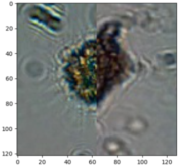
Constrative lrng



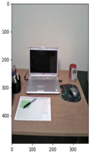
Multi-modal/multi-task learning: Anomaly detection results

Considering "Chimera" and "COCO" as the anomalies:

Methods	AUC (Overall)
w/o Supervised Contrastive	71.64
w/o Gromov-Wasserstein	78.77
Full model	96.25

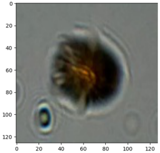


Anomaly score: 0.9574

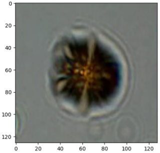


Anomaly score: 0.9757

True Positives



Anomaly score: 0.9523



Anomaly score: 0.9515

False Positives

What's next?

- Publish papers, papers, papers,
- (Re)using pretrained multi-modal models like CLIP[13]: human-readable explanations and more,
- imbalance learning by mixing augmentation and resampling,
- domain adaptation respecting domain constraints (causality),
- invariant representation learning,[14] and
- aggregation of images: What can we deduce relying on geographical distance and time between images?

- OcéanIA `deep-plankton`.
- plankton datasets, +700 models, all bells and whistles, scalable multi-host/multi-gpus, sharing models, etc.! Demo link.
- Best partner: Ecotaxa (model distillation, scalable inference).

[13] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision

[14] Kwon, S., Choi, J. Y., & Ryu, E. K. (2023). Rotation and Translation Invariant Representation Learning with Implicit Neural Representations

Merci ! *¡Gracias!* Thank you!

Questions?