Open-Set Plankton Recognition Using Similarity Learning

Badreldeen Bdawy Mohamed Ola, Eerola Tuomas, Kraft Kaisa, Lensu Lasse, Kälviäinen Heikki

This is a Post-print version of a publication

published by Springer, Cham

in Advances in Visual Computing. ISVC 2022. Lecture Notes in Computer Science

DOI: 10.1007/978-3-031-20713-6_13

Please cite the publication as follows:

Badreldeen Bdawy Mohamed, O., Eerola, T., Kraft, K., Lensu, L., Kälviäinen, H. (2022). Open-Set Plankton Recognition Using Similarity Learning. In: , et al. Advances in Visual Computing. ISVC 2022. Lecture Notes in Computer Science, vol 13598. Springer, Cham. https://doi.org/10.1007/978-3-031-20713-6_13

Open-Set Plankton Recognition Using Similarity Learning

Ola Badreldeen Bdawy Mohamed[1], Tuomas Eerola[1][0000−0003−1352−0999], Kaisa Kraft[2][0000−0001−6290−3887], Lasse Lensu[1][0000−0002−7691−121X], and Heikki Kälviäinen[1][0000−0002−0790−6847]

[1] Computer Vision and Pattern Recognition Laboratory, School of Engineering Science, Lappeenranta-Lahti University of Technology LUT, Lappeenranta, Finland
[2] Finnish Environment Institute, Marine Research Centre, Helsinki, Finland

Abstract. Automatic plankton recognition provides new possibilities to study plankton populations and various environmental aspects related to them. Most of the existing recognition methods focus on individual datasets with a known set of classes limiting their wider applicability. Automated plankton imaging instruments capture images of unknown particles and the class (plankton species) composition varies between geographical regions and ecosystems. This calls for an open-set recognition method that is able to reject images from unknown classes and can be easily generalized to new classes. In this paper, we show that a flexible model capable of high classification accuracy can be obtained by utilizing similarity learning and a gallery set of known plankton species. The model is shown to generalize well for new plankton classes added in the gallery set without retraining the model. This provides a good basis for the wider utilization of plankton recognition methods in aquatic research.

Keywords: Plankton recognition · open-set classification · metric learning.

1 Introduction

Phytoplankton are microscopic organisms that grow at a rapid rate. Combined with their ability to produce organic compounds from inorganic material, phytoplankton are considered the foundation of the marine food web by supporting all other living organisms in the ocean. As a by-product of the photosynthetic operation, phytoplankton are one of the main producers of oxygen on the Earth. Because of the critical roles it plays both as a sustainer of marine ecosystems and as a regulator of a global climate change, monitoring phytoplankton populations over time and space is essential.

Recent technological advancements have resulted in the emergence of automated and semi-automated plankton imaging instruments with continuously improving image resolution and output rates. This has opened novel possibilities to study plankton communities. However, to fully utilize the large image volumes in plankton research automatic methods are needed to analyze the image data. The main image analysis task to be solved is plankton recognition, i.e., classifying the images based on the species they contain.

Convolutional neural networks (CNNs) have shown to reach close-to-human level accuracy in various image recognition tasks and plankton recognition is not an exception [12, 9]. However, they are known to struggle in open-set settings where the class composition of training data differs from the data for which the trained model is applied. Typical CNN-based classification models tend to classify the images from a new class to one of the known classes often with a high confidence, and to include new classes to the models, they need to be retrained. These are major problems for plankton recognition as the plankton species vary between different regions and seasons. Retraining a separate model for each dataset is not feasible. Therefore, there is a need for a recognition model that 1) is able to predict when the image contains a previously unknown plankton species and 2) can be generalized to new classes without retraining the whole model.

In this paper, we address these challenges by proposing a novel open-set plankton recognition method utilizing metric learning. The idea is to learn such image embeddings that the plankton images from the same species are close to each other and the images from the different species are far from each other in the feature space (see Fig. 1). The recognition method consists of a gallery set of known species and a learnt similarity metric allowing to compare query images to the gallery images. Similarity in this context corresponds to likelihood that the images belong to the same class. This further allows to define a threshold value for similarity enabling open-set classification: if no similar images are found in the gallery set, the query image is predicted to belong to an unknown class. Furthermore, new classes can be added by simply including them into gallery set as the model does not necessarily need to learn class-specific image features.

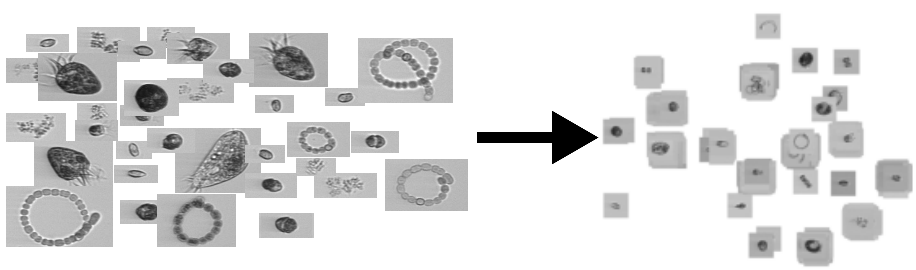

Fig. 1. Similarity metric learning for plankton images.

We propose to train the similarity metric using the angular margin loss (ArcFace) [5] combined with Generalised mean pooling (GeM) [20] allowing to aggregate of the deep activations to rotation and translation invariant representations.

ArcFace uses a similarity learning mechanism that allows distance metric learning to be solved in the classification task by introducing the Angular Margin Loss. This allows straightforward training of the model and only adds negligible computational complexity. In the experimental part of the work, we show that the proposed method obtains high plankton image classification accuracy and outperforms the previously proposed model utilizing OpenMax [1] layer in open-set classification. We further show that the method generalizes well to new classes added to the gallery set without retraining. This makes it straightforward to apply the model to new datasets with only partly overlapping plankton species composition.

2 Related work

2.1 Plankton recognition

In hope of mitigating the laborious task of manually classifying the plankton images, various automatic approaches have been proposed. Modern imaging devices often utilize flow cytometry and are able to produce separate images of individual particles rendering the plankton recognition task as an image classification problem. Traditional plankton recognition methods utilize hand-crafted image features such as shape and texture (see e.g. [2]). Recently, CNNs have replaced hand-crafted features and have shown recognition performance which is comparable to human experts [13, 9]. Such recognition models have already been implemented into operational phytoplankton recognition systems [11]. A typical approach utilizes common CNN architectures (e.g., ResNet), pre-trained models, and transfer learning [18, 12]. However, also custom architectures have been proposed to address the fine-grained nature of the classification problem [3, 4].

2.2 Open-set classification

Generic classifiers often fall under the false assumption that the model has already seen all the possible classes that it will encounter after the model has been deployed [7]. In a realistic setting, this assumption is typically not true. For example, continuous plankton imaging devices capture also non-plankton particles and rare plankton particles not present in the training data. This is even more evident when the classification model is applied to data collected from a new geographical location with only partially overlapping plankton species composition with the training data. Open-set classification aims to identify already known classes successfully and simultaneously reject unknown classes [7].

Bendale et al. [1] proposed the OpenMax which is an additional layer that allows deep neural networks to perform open-set recognition. The method utilizes meta-recognition to analyze activation scores and identify when the recognition model is likely to fail. Based on the distribution of the activation vector values, the OpenMax layer calculates the probability of an image being from an unknown class.

In the case of plankton recognition the open-set problem is often formulated as an anomaly detection problem where the model is trained to both correctly classify the known classes and to filter abnormal classes by training the model to produce high and low entropy distributions for the normal classes and abnormal classes respectively. Yuchun et al. [19] proposed a loss function which contains three loss terms to detect the anomalies and to maintain the classification accuracy for the images belonging to the normal classes by incorporating the expected cross-entropy loss, the expected Kullback-Leibler (KL) divergence, and the Anchor loss. The model was tested on classes of plankton images containing also bubbles or random suspending particles.

Walker et al. [22] utilized a large background set of images which do not belong to the target classes (classes to be recognized) and hard negative mining to find images that are more likely to cause false negatives. The training set was then complemented with these challenging images to improve the classifiers ability to recognize when the images are from novel classes. While promising results were obtained on open-set plankton recognition the method requires that a labeled background set is available which limits the usability of the method.

2.3 Classification by metric learning

The aim of deep metric learning is to obtain image embedding vectors that model the similarity between images. It is commonly utilized in person [23] and animal re-identification [15], as well as, content based image retrieval [6], but has been also successfully applied to more traditional image recognition problems such as vehicle attribute recognition [16]. The main benefit of metric learning is that training with the full set of target classes is not needed which makes metric learning more suitable for open-set recognition than traditional classification models.

The most common approaches for deep metric learning include triplet-based learning strategies and classification-based metric learning. The first approach learns the metric by sampling image triplets with and anchor, positive, and negative examples [10]. The loss function is defined in such a way that the distance (similarity) from the embeddings of the anchors to the positive samples are minimized, and the distance from the anchors to the negative samples are maximized. The second approach approximates the classes using learnt proxies [14] or class centers [5] that provide the global information needed to learn the metric. This makes it possible to formulate the loss function based on the softmax loss and allows to avoid the challenging triplet mining step.

Recently, metric learning has been utilized also in plankton classification. Teigen et al. [21] studied the viability of few-shot learners in correctly classifying plankton images. A Siamese network was trained using the triplet loss and used to determine the class of a query image. Two scenarios were tested: the multi-class classification and the novel class detection. A model trained to distinguish between five classes of plankton using five reference images from each class was able to achieve a reasonable accuracy. In the novel class detection, however, the model was able to filter out only 57 images out of 500 unknowns. Furthermore, the used triplet loss approach suffers from the high cost of the triplets mining and exponentially increasing computations as the number of classes increases.

3 Proposed method

The proposed method for plankton recognition is based on similarity metric and a gallery set of known classes. To obtain the similarity metric, a CNN model is trained using the Angular Margin loss (ArcFace) [5]. Given an image as input, the trained CNN model outputs an embedding vector and a similarity of two images is quantified by computing the cosine distance between the image embeddings as

$$d_{cos}(\boldsymbol{v_1}, \boldsymbol{v_2}) = \frac{\boldsymbol{v_1} \cdot \boldsymbol{v_2}}{||\boldsymbol{v_1}||\ ||\boldsymbol{v_2}||}, \tag{1}$$

where $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ are the embedding vectors. The embedding vector will be discussed further below.

To perform the plankton recognition for a query image, the embedding vector is first computed using the trained model. Then the distances to the embedding vectors of gallery set images are computed and the label is given based on the most similar image. See Fig. 2 for the overview of the method. It should be noted that since the image embeddings for gallery set can be computed and stored beforehand the query image recognition can be done efficiently by computing the cosine similarities between the vectors (simple dot product if the vectors are L^2 normalised). If the similarity between the query image and the most similar gallery set image exceeds the predetermined threshold the query image is labelled as unknown providing the basis for the open-set recognition. The metric learning approach increases inter-class separability while decreasing intra-class variation making the recognition less sensitive to selected threshold values when compared to a traditional classification approach with class probability thresholding. The threshold values can be tuned by minimizing the amount of misclassifications in the validation set.

Since the method learns to quantify the similarity (likelyhood that the images originate from the same class) instead of representations for individual classes,

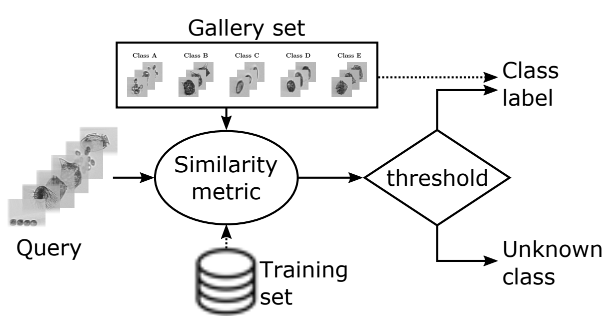

Fig. 2. The proposed method.

the set of plankton species for which the method is applied can differ from the set of classes in the training set. Therefore, to utilize the trained model on a new dataset with different set of classes due to, for example, different geographical region or ecosystem, one must only select and label a new gallery set. The gallery set requires considerably less labeled images per class than model training. Technically, even just one gallery image per class is enough to apply the method if intra-class variation is very small. However, a very small amount of gallery images may lead to a subpar recognition performance.

The method can be used with any backbone architecture, but ResNet-18 [8] has been found to produce a high classification accuracy on plankton image data with low computation cost [11]. We further propose to use Generalised mean pooling (GeM) [20] to aggregate the deep activations and to construct a representation that is invariant to both rotation and translation of the plankton. The embedding vector $\boldsymbol{v}$ aggregated through GeM can be written as

$$\boldsymbol{v} = [v_1 \dots v_k \dots v_C]^\top, \quad v_k = \left(\frac{1}{|\mathcal{X}_k|} \sum_{x \in \mathcal{X}_k} x^{p_k} \right)^{\frac{1}{p_k}}, \quad k \in \{1 \dots C\}, \tag{2}$$

where $\mathcal{X}_k$ is a set of elements of the feature map k and C is the number of channels. The greater the power parameter p_k, the more the network values strong features. One of the major benefits of GeM is that p_k is also learnable so it can be optimized during the learning process.

3.1 Angular Margin Loss

ArcFace [5] utilizes the Angular Margin Loss to learn a distance metric for the classification task. The idea behind the method is to consider the weights of the last fully-connected layer as class centers. Normalization is used to distribute embeddings on a hypersphere with predefined radius which makes it possible to utilize geodesic distance. The loss is formulated as:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq y_i}^{n} e^{s \cos \theta_j}}, \tag{3}$$

where s is the the feature scale (hypersphere radius), θ_{y_i} is an angle between embedding and the class center (vector of weights) of the correct class y_i, θ_j is an angle between weight vector for class j and the predicted embedding vector, N and n are the batch size and number of classes, respectively. m is a predefined additive margin that is used to increase inter-class separability while decreasing intra-class variation. The most notable benefits of the ArcFace method include the lack of need for triplet mining and a better class separability.

4 Experiments

4.1 Data

The data was collected from the Baltic Sea using an Imaging FlowCytobot (IFCB) [17] that capture grayscale images of individual phytoplankton (see

Fig. 3). The SYKE-plankton_IFCB_2022 dataset consists of 63 074 images representing 50 different classes of phytoplankton manually labeled by an expert. Due to the varying rarity of plankton species, the dataset is highly imbalanced and the number of images per class varies from 19 to 12 280 images. For detailed description of the data, see [11]. The dataset has been made publicly available[3].

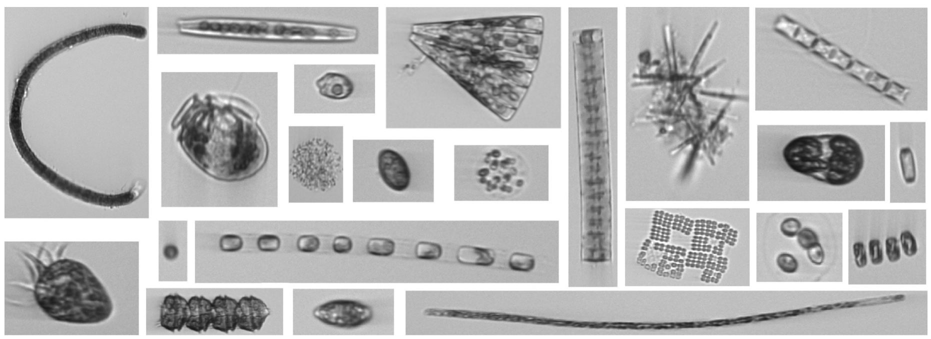

Fig. 3. Example images from the dataset.

To prepare the data for the training phase, several preprocessing steps were done. The images were resized to have a standard dimension [224,224]. Resizing was done using bicubic interpolation and the aspect ratio was maintained by padding with the background color. The dataset was split into the training, validation and test subsets with a ratio 6:2:2. To address the large class imbalance, undersampling was utilized for large classes and data augmentation with random affine transformations for small classes in order to create a balanced training set with 2 000 images per class.

4.2 Description of experiments

To evaluate the open-set classification accuracy 10 classes were selected as unknowns and excluded from the training set. The remaining 40 classes were used for training. The gallery set was constructed by randomly selecting 100 images per class from the the training set. The experiment was repeated 5 times in such a way that each class was selected as unknown once.

ResNet-18 was used as the backbone architecture for all experiments. The network was trained from scratch using Adam optimizer. A fixed learning rate of 1e-5 was used to train 200 epoch with a batch size of 64. The main two hyperparameters related to ArcFace are the hypersphere radius s and the additive angular margin penalty m. s and m were set to 2.39 and 0.95, respectively. The threshold values for open-set classification were defined for each class separately

[3] http://doi.org/10.23728/b2share.abf913e5a6ad47e6baa273ae0ed6617a

based on the validation set. The thresholds were found based on the distance between the query image and all the images in the gallery set. For OpenMax a pretrained ResNet-18 was used as a backbone.

4.3 Results

Table 1 shows the comparison between the proposed metric learning based method and the OpenMax method. The classification of the knowns presents the results in traditional closed-set setting with the same 40 classes included in both the training and test sets. The classification accuracy with the proposed method varied between 92.5% and 95.4%. These are comparable accuracies with baseline CNN classifiers obtained with the similar datasets (96% accuracy with 32 classes of phytoplankton [3] and 97% accuracy with 50 classes [11]). The classification of the knowns with the threshold shows the results when the test set contains only images from the known classes, but the threshold is applied to filter out predicted unknowns. As it can be seen, the accuracy decreases only little, which indicates that the known classes are only rarely classified as unknowns. The classification of the unknowns shows how many percentage of images from the previously unseen classes were correctly classified as the unknowns and the open-set classification shows results with 41 classes (40 known classes + unknowns). As it can be seen, the proposed method outperforms OpenMax in both recognition accuracy and ability to reject images from previously unseen classes.

Table 1. Mean classification accuracies and standard deviations over all 5 subexperiments.

	Classification of knowns	Classification of knowns+threshold	Classification of unknowns	Open-set recognition
OpenMax [1]	93.85±0.84%	91.96±0.68%	41.80±8.10%	90.65±0.39%
Proposed	94.60±1.05%	93.27±0.95%	65.20±6.43%	92.33±0.90%

One benefit of the proposed similarity learning approach is that by including images to the gallery set it allows to generalize the method to new classes without retraining the model itself. To study the method's ability to generalize, example images from the 10 unknown classes were included into the gallery set. Two experiments were carried out: 1) the gallery set and the query set contained images from all 50 classes (40 classes used to train the similarity model and 10 unknown classes), and 2) the gallery set and the query set contained images only from the 10 classes that were not included in the training. The results are shown in Table 2. While a drop in accuracy can be observed due to considerably more challenging tasks and a mismatch between training and test set distributions, a reasonably high accuracy was obtained.

Table 2. Capability to generalize to previously unseen classes.

	Top-1	Top-2	Top-3	Top-4	Top-5
50 classes (10 new)	84.48±1.90%	91.96±0.99%	94.69±0.70%	95.88±0.41%	96.57±0.38%
10 classes (all new)	74.07±7.08%	90.21±3.29%	95.68±2.03%	97.78±1.60%	98.91±0.70%

5 Conclusions

In this paper, a similarity learning approach to tackle the open-set plankton recognition problem was proposed. The method consists of a similarity metric learned using angular margin loss and a gallery set of known plankton species. The feature embeddings produced by the similarity learning model allow to compute the similarities between images and to find the most similar image (species) in the gallery set of known plankton species. Moreover, by setting similarity thresholds, the method is able to recognize when the query image contains a plankton species not present in the gallery set, enabling open-set recognition. The proposed method was shown to accurately recognize plankton species and it outperformed OpenMax in the open-set recognition task. Furthermore, we showed that the proposed method can adapt to new classes added to the gallery set without retraining the similarity learning model. This is a promising step towards a general-purpose plankton recognition method applicable to different datasets with varying class compositions, promoting the wider utilization of automatic plankton recognition for aquatic research.

Acknowledgements The research was carried out in the FASTVISION and FASTVISION-plus projects funded by the Academy of Finland (Decision numbers 321980, 321991, 339612, and 339355).

References

1. Bendale, A., Boult, T.: Towards open set deep networks. In: Conference on Computer Vision and Pattern Recognition. pp. 1563–1572 (2016)
2. Bueno, G., Deniz, O., Pedraza, A., Ruiz-Santaquiteria, J., Salido, J., Cristóbal, G., Borrego-Ramos, M., Blanco, S.: Automated diatom classification (Part A): handcrafted feature approaches. Applied Sciences **7**(8), 753 (2017)
3. Bureš, J., Eerola, T., Lensu, L., Kälviäinen, H., Zemčík, P.: Plankton recognition in images with varying size. In: International Conference on Pattern Recognition Workshops. pp. 110–120 (2021)
4. Dai, J., Wang, R., Zheng, H., Ji, G., Qiao, X.: ZooplanktoNet: Deep convolutional network for zooplankton classification. In: OCEANS Conference. pp. 1–6 (2016)
5. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: ArcFace: Additive angular margin loss for deep face recognition. In: Conference on Computer Vision and Pattern Recognition. pp. 4690–4699 (2019)

6. Dubey, S.R.: A decade survey of content based image retrieval using deep learning. IEEE Transactions on Circuits and Systems for Video Technology **32**(5), 2687–2704 (2021)
7. Geng, C., Huang, S.j., Chen, S.: Recent advances in open set recognition: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(10), 3614–3631 (2020)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
9. Henrichs, D.W., Anglès, S., Gaonkar, C.C., Campbell, L.: Application of a convolutional neural network to improve automated early warning of harmful algal blooms. Environmental Science and Pollution Research **28**(22), 28544–28555 (2021)
10. Hoffer, E., Ailon, N.: Deep metric learning using triplet network. In: International Workshop on Similarity-based Pattern Recognition. pp. 84–92 (2015)
11. Kraft, K., Velhonoja, O., Eerola, T., Suikkanen, S., Tamminen, T., Haraguchi, L., Ylöstalo, P., Kielosto, S., Johansson, M., Lensu, L., Kälviäinen, H., Haario, H., Seppälä, J.: Towards operational phytoplankton recognition with automated high-throughput imaging, near real-time data processing, and convolutional neural networks. Frontiers in Marine Science **9** (2022)
12. Lumini, A., Nanni, L.: Deep learning and transfer learning features for plankton classification. Ecological Informatics **51**, 33–43 (2019)
13. Mitra, R., Marchitto, T., Ge, Q., Zhong, B., Kanakiya, B., Cook, M., Fehrenbacher, J., Ortiz, J., Tripati, A., Lobaton, E.: Automated species-level identification of planktic foraminifera using convolutional neural networks, with comparison to human performance. Marine Micropaleontology **147**, 16–24 (2019)
14. Movshovitz-Attias, Y., Toshev, A., Leung, T.K., Ioffe, S., Singh, S.: No fuss distance metric learning using proxies. In: International Conference on Computer Vision. pp. 360–368 (2017)
15. Nepovinnykh, E., Eerola, T., Kalviainen, H.: Siamese network based pelage pattern matching for ringed seal re-identification. In: Winter Conference on Applications of Computer Vision Workshops. pp. 25–34 (2020)
16. Ni, X., Huttunen, H.: Vehicle attribute recognition by appearance: computer vision methods for vehicle type, make and model classification. Journal of Signal Processing Systems **93**(4), 357–368 (2021)
17. Olson, R., Sosik, H.: A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging flowcytobot. Limnology and Oceanography: Methods **5**(6), 195–203 (2007)
18. Orenstein, E., Beijbom, O.: Transfer learning and deep feature extraction for planktonic image data sets. In: Winter Conference on Applications of Computer Vision. pp. 1082–1088 (2017)
19. Pu, Y., Feng, Z., Wang, Z., Yang, Z., Li, J.: Anomaly detection for in situ marine plankton images. In: International Conference on Computer Vision. pp. 3661–3671 (2021)
20. Radenović, F., Tolias, G., Chum, O.: Fine-tuning cnn image retrieval with no human annotation. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(7), 1655–1668 (2018)
21. Teigen, A.L., Saad, A., Stahl, A.: Leveraging similarity metrics to in-situ discover planktonic interspecies variations or mutations. In: Global Oceans 2020: Singapore–US Gulf Coast. pp. 1–8 (2020)

22. Walker, J., Orenstein, E.: Improving rare-class recognition of marine plankton with hard negative mining. In: International Conference on Computer Vision. pp. 3672–3682 (2021)
23. Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., Hoi, S.C.: Deep learning for person re-identification: A survey and outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(6), 2872–2893 (2021)