# Zooplankton Classification Using Hierarchical Attention Branch Network

Koichi Ito[1], Kanta Miura[1], Takafumi Aoki[1], Yurie Otake[2], Wataru Makino[2], and Jotaro Urabe[2]

[1] Graduate School of Information Sciences, Tohoku University
6-6-05, Aramaki Aza Aoba, Sendai, 9808579, Japan.
ito@aoki.ecei.tohoku.ac.jp
[2] Graduate School of Life Sciences, Tohoku University
6-3, Aramaki Aza Aoba, Sendai, 9808578, Japan.

**Abstract.** Plankton is recognized as one of the most important indicators of the health of aquatic ecosystems and water quality. Surveys of plankton populations in oceans and lakes have been conducted manually. Plankton classification methods using deep learning have been developed to automatically classify plankton images. These methods do not sufficiently take into account the bias in the species included in the dataset or the similarity of their shapes. In this paper, we propose a hierarchical attention branch network (H-ABN) to utilize that plankton are hierarchically named according to their taxonomic ranks. We demonstrate the effectiveness of the proposed method through experiments using a zooplankton dataset collected from lakes and ponds in Japan.

**Keywords:** image classification · plankton · taxonomy · attention branch network · convolutional neural network.

## 1 Introduction

Zooplankton play a fundamental role in aquatic ecosystem services, for example, regulating water quality by eating algae and linking lower and higher trophic levels [18]. In recent years, aquatic ecosystems have become a serious environmental problem due to the decrease in biodiversity caused by the increase in alien species and the occurrence of nuisance algae. Therefore, plankton, including zooplankton, have been monitored periodically in economically important oceans, lakes, and reservoirs. On the other hand, it is difficult to conduct periodic and accurate plankton monitoring due to the limited number of experts who have the skills to identify the large amount of plankton contained in the collected samples. In addition, plankton communities need to be monitored frequently since they vary within a few days, however, accurate identification and enumeration at high frequencies is difficult even for experts.

To address the above problems, automated plankton monitoring systems using machine learning techniques have been investigated. The methods using hand-crafted features have been proposed as typical plankton image classification methods using machine learning [9,16,20]. Examples of typical hand-crafted features include Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG). Plankton images are classified from extracted features using discriminators such as Support Vector Machine (SVM) and Random Forest. Due to the generality of hand-crafted features, they are not necessarily suitable for plankton image classification, and the classification accuracy is low. With the rapid development of deep learning [8], Convolutional Neural Network (CNN)-based methods have recently been proposed for automated plankton image classification [3, 5, 7, 12–14]. Plankton images can be classified more accurately than methods using hand-crafted features since features are extracted from plankton images based on training using a large amount of data. These methods only utilize CNNs for image recognition to classify plankton images, and do not fully consider label bias or similarity of plankton shapes in the dataset. Although plankton are classified hierarchically based on taxonomic ranks, only one label is given to each plankton image in the available datasets.

In this paper, we consider the plankton image classification by taking into account the taxonomic ranks of the plankton. The taxonomic ranks of plankton are often determined based on their shapes. We expect that it is effective to classify images hierarchically according to the taxonomic ranks. Therefore, we propose a plankton image classification method using Attention Branch Network (ABN) [6], which can apply attention to regions of interests, in a hierarchical form to classify plankton images with high accuracy. We also construct a plankton image dataset by annotating labels based on the taxonomic ranks to samples collected in Japanese lakes and marshes to classify plankton images according to the taxonomic ranks. We demonstrate the effectiveness of the proposed method in plankton image classification through performance evaluation experiments using the constructed plankton image dataset.

## 2 Related Work

This section gives an overview of plankton image datasets and plankton image classification using CNNs.

### 2.1 Plankton Image Datasets

Table 1 summarizes the plankton image dataset. WHOI-Plankton [2] is a dataset of 3,272,578 images consisting of 103 classes of zooplankton and phytoplankton collected at Martha's Vineyard. WHOI [17] is a dataset of 6,600 images consisting of 22 classes of zooplankton and phytoplankton collected at Woods Hole Harbor. The images in both datasets were taken with an underwater microscope (In Situ Ichthyoplankton Imaging System: ISIIS), and are therefore blurred. ZooScan [9] is a dataset of 3,771 images consisting of 20 classes of zooplankton collected at Villefranche-sur-Mer Bay. The images in this dataset were scanned using a

**Table 1.** Summary of plankton image datasets for plankton image classification, where "P" and "Z" indicate phytoplankton and zooplankton, respectively.

| Dataset | # of images | # of classes | Image size [px.] | Location | Type |
|---|---|---|---|---|---|
| WHOI-Plankton [2] | 3,272,578 | 103 | $54 \times 139$ | Sea | P, Z |
| WHOI [17] | 6,600 | 22 | $156 \times 367$ | Sea | P, Z |
| ZooScan [9] | 3,771 | 20 | $143 \times 154$ | Sea | Z |
| Kaggle [1] | 30,336 | 121 | $67 \times 73$ | Sea | P, Z |
| BearingSea [3] | 17,920 | 7 | $425 \times 411$ | Sea | Z |
| ZooLake [11] | 17,943 | 35 | $123 \times 121$ | Lake | Z |
| Ours | 35,820 | 76 | $191 \times 190$ | Lake | Z |

consumer product, ZooScan[1]. The background is clear due to the free of dust and other impurities, however, the resolution of the images is low. Kaggle [1] is a dataset of 30,336 images consisting of 121 classes of zooplankton and phytoplankton collected at Florida Strait. The images are blurred since they were taken by ISIIS as well as WHOI-Plankton and WHOI. BearingSea [3] is a dataset of 17,920 images consisting of 7 classes of zooplankton collected at Southeastern Bering Sea. The images in this dataset were taken using the zooplankton visualization and imaging system (ZOOVIS), which is a high-resolution digital imaging system that can acquire images of plankton underwater. ZooLake [11] is a dataset of 17,943 images consisting of 35 classes of zooplankton collected at Lake Greifensee. The images were taken using an underwater microscope (Dual-magnification Scripps Plankton Camera: DSPC), and the image quality is higher than that of other datasets. Although the plankton nomenclature is based on taxonomic ranks, only one label, such as species, is assigned to any of the datasets. Hierarchical labels are important for plankton image classification, since plankton taxonomic ranks are associated with their shapes. Therefore, we construct a new plankton image dataset with high resolution images using optical microscopy and hierarchical labels assigned according to the plankton taxonomic ranks. The details of our dataset are described in Sect. 4.

### 2.2 Plankton Image Classification Using CNNs

We provide an overview of plankton image classification methods using CNNs. Luo et al. classified plankton images extracted by k-means into 108 types of plankton using SparseConvNet [14]. Ellen et al. improved the classification accuracy by adding metadata obtained during data acquisition to the fully-connected layer for VGG16 [5]. Cheng et al. detected plankton using MSER and classified them with SVM using features extracted by CNN [3]. Lumini et al. propose a method to ensemble features extracted by multiple pre-trained CNNs [12, 13]. González et al. propose a method using ResNet [10] and analyze the counting plankton for time-series data. All of the methods only utilize CNNs used in image classification, and none of them have considered the characteristics of plankton.

[1] `http://www.zooscan.com`

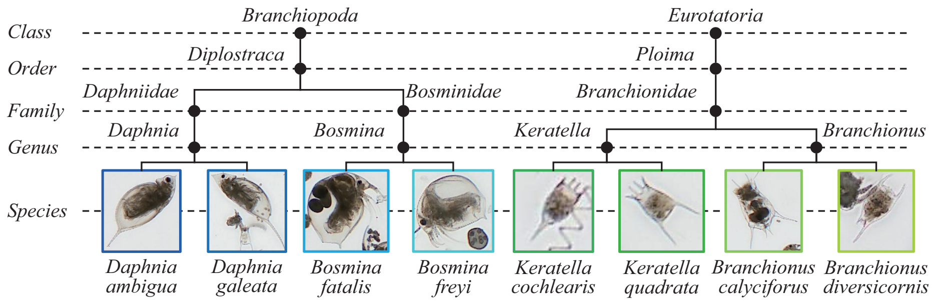


**Fig. 1.** Example of taxonomic ranks of plankton.

In addition, these methods use plankton data collected at the same location for training and test, and therefore, the generalizability of the classification methods has not been correctly evaluated.

## 3 Hierarchical Plankton Image Classification

We present the proposed method for classifying plankton images with labels based on the taxonomic ranks. The following describes the plankton taxonomic ranks, ABN, which is the fundamental of the proposed method, and the hierarchical ABN.

### 3.1 Taxonomic Ranks of Plankton

A hierarchical nomenclature based on taxonomy such as *class*, *order*, *family*, *genus*, and *species* is used for plankton. An example of the hierarchical structure of plankton taxonomic ranks is shown in Fig. 1. Plankton of the same *genus* have very similar shapes, even though they belong to different *species*. Thus, the taxonomic rank of plankton is based on its shape, which may be useful for plankton image classification. In this paper, we use five taxonomic ranks of plankton: *class*, *order*, *family*, *genus*, and *species* .

### 3.2 Attention Branch Network

The proposed method utilizes Attention Branck Network (ABN) [6], which provides feedback on the attention map generated from the feature map, to classify plankton images considering the shape of the plankton. ABN is an attention mechanism based on a heat map of the region of interest in CNN, consisting of the attention branch and perception branch. The attention branch performs class classification based on the input feature map and generates an attention map that represents the region of interest in CNN. The perception branch performs class classification based on the features that emphasize the region of interest by multiplying the feature map and the attention map. ABN not only improves the

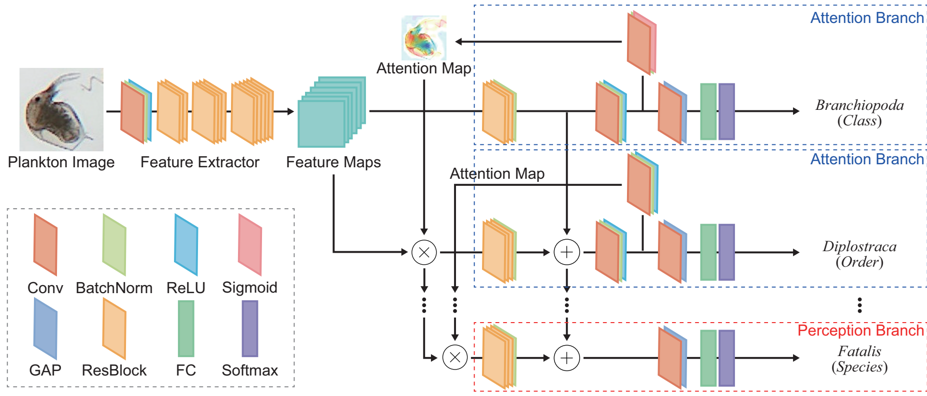


**Fig. 2.** Overview of the proposed method consisting of ResNet and H-ABN, which is the case of using 5 ranks from *class* to *species*.

accuracy of general image classification, but has also been demonstrated to be effective in Fine-Grained Recognition for CompCars [19]. Since plankton images often have a uniform background, the attention map indicates the shape of the plankton (or a local region that is useful for classification). By applying ABN from the top to the bottom according to the hierarchical labels of plankton, we expect that plankton images can be classified in consideration of the taxonomic ranks.

### 3.3 Hierarchical Attention Branch Network

In this paper, we propose Hierarchical Attention Branch Network (H-ABN) that hierarchically applies ABN to the feature maps extracted by CNN according to the hierarchical labels given to the plankton images. Fig. 2 shows an overview of the proposed method. The proposed method first extracts feature maps from plankton images using CNN. In this paper, we empirically use ResNet [10] as a feature extractor. Next, the feature maps extracted by CNN are input to H-ABN to classify plankton images. H-ABN consists of multiple attention and perception branches. The number of attention branches is determined by the number of ranks in the taxonomic ranks to be considered. Each attention branch performs class classification for the labels of each taxonomic rank and generates an attention map. Note that the attention branch of the proposed method differs from the original ABN in the following two points. In the proposed method, the attention map is multiplied by the feature map and input to the attention branch one layer below. The features extracted in the upper attention branch are added to the features extracted in the lower attention branch. This is because the features used in the hierarchical classification of plankton images are expected to be effective in the classification of the lower layer as well. The classification is performed sequentially from the highest attention branch, and finally the classification of the target label is performed in the perception branch. H-ABN

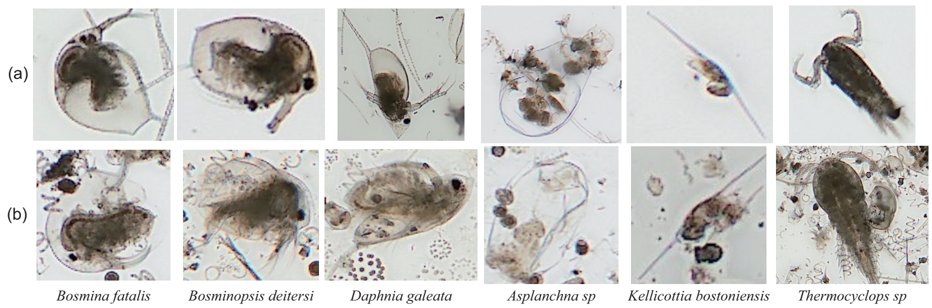


**Fig. 3.** Examples of plankton images in our dataset: (a) DB1 and (b) DB2, where images of the same *species* are presented.

**Table 2.** Configuration of our dataset used in the experiments, where all *species* in DB2 are included in DB1.

| DB | Train | Val | Test | # of *species* | # of sites |
|---|---|---|---|---|---|
| DB1 | 15,886 | 2,961 | 3,799 | 76 | 26 |
| DB2 | — | — | 13,174 | 44 | 6 |

takes into account not only the shape of the plankton but also the hierarchical structure of the labels. The features used in the classification of the upper ranks, which are less biased, can be applied to the feature maps of the lower ranks as an attention map, reducing the degradation of the classification accuracy due to label bias. H-ABN is trained using the sum of the cross-entropy losses computed for each branch as the overall loss.

## 4 Plankton Image Dataset

This section describes the plankton image dataset with labels based on taxonomic ranks used in the experiments. The dataset consists of zooplankton images collected from 32 lakes and marshes in Japan between 2006 and 2022. The plankton community images were obtained by scanning samples with an upright fluorescence microscope (OLYMPUS BX63) at a magnification of 40x. We define a boundary box for each plankton using the Computer Vision Annotation Tool (CVAT)[2] and annotated each plankton with a label of its five taxonomic ranks (*order*, *class*, *family*, *genus*, and *species*). Each plankton was cropped based on the bounding box to obtain 35,820 plankton images. For evaluating the classification accuracy, we divide the dataset into DB1, which consists of plankton images collected at 26 sites, and DB2, which consists of plankton images collected at the other 6 sites, as shown in Table 2. The number of taxonomic labels in DB1 is 3 for *order*, 7 for *class*, 28 for *family*, 45 for *genus*, and 76 for *species*, while those in DB2 are 3 for *order*, 7 for *class*, 20 for *family*, 28 for *genus*, and 44 for *species*.

[2] `https://github.com/opencv/cvat`

Note that all 44 plankton *species* in DB2 are included in the 76 *species* in DB1. Fig. 3 shows examples of plankton images in DB1 and DB2. As shown in Fig. 3 (a), DB1 contains clear images with little background debris. On the other hand, as shown in Fig. 3 (b), DB2 contains images with a lot of background debris and images partially overlapped with other plankton. Furthermore, the shape of the plankton differs depending on the environment, even if they are the same type of plankton, since the sampling sites are different. The generalizability of the plankton image classification method can be evaluated by training with DB1 and testing with DB2. The dataset in this paper will be available to the public under a research-use license.

## 5 Experiments and Discussion

This section describes experiments using our dataset to demonstrate the effectiveness of the proposed method in plankton image classification.

In this experiment, we evaluate the fundamental performance using DB1 and the generalizability using DB2, as shown in Table 2. We use 15,886 images of DB1 for training and 3,799 images of DB1 for validation. The plankton image is padded to be square, resized to $224 \times 224$ pixels, enhanced by histogram equalization, and used as input for each method. Random flip, random rotation, color jitter, random erasing [21], cutmix [22] are used as data augmentation in training. Random erasing masks rectangular regions of random position and size in the image with black color. The ratio of the masked area to the entire image is 0.002 to 0.2, and the aspect ratio is 0.3 to 3. Cutmix swaps a rectangular region of random position and size between two images and weights the labels in proportion to the size of the rectangular region for each label. The probability of cutmix is set to 0.5 and the parameters $\alpha$ and $\beta$ are set to 1. In this experiment, we use ResNet-50 pre-trained on ImageNet as the feature extractor of the proposed method. In the proposed method, the perception branch of H-ABN is fixed to *species* and the attention branches of H-ABN are varied from it class to *genus*. Nesterov Accelerated Gradient (NAG) [15] is used as the optimizer, the initial learning rate is 0.001, the batch size is 64, and the number of epochs is 200. We compare the classification accuracy of the proposed method with ResNet-50 [10], ABN [6], and Marginalization Classifier (MC) [4] to demonstrate the effectiveness of the proposed method. MC is an image classification method using hierarchical labels that performs classification in the upper layers based on the predicted probability of each class in the lowest layer estimated by CNN. Experiments on the ETH Entomological Collection dataset have demonstrated that classification accuracy is improved when the hierarchical structure of labels is taken into account. In this experiment, the two ranks (*genus* and *species*) with empirically highest accuracy are used in MC. We use the accuracy, which is the percentage of correct classifications, and the F1-score, which is the harmonic mean of precision and recall, as the evaluation metrics. Note that we use only accuracy for evaluating the results for DB2, since each method uses a model for 76 classifications to perform 44 classifications on DB2.

Table 3 shows the experimental results for each method for DB1. Focusing on the accuracy of *species*, all the methods exceed 90%. Focusing on the F1-score, the proposed methods have a higher F1-score than the conventional methods. In particular, the proposed method with four levels of labels (*order*, *family*, *genus*, and *species*) achieves the highest accuracy. When plankton from the same sampling sites are included in the training and test, all methods can classify plankton with high accuracy.

Table 4 shows the experimental results of each method for DB2. The accuracy of DB2 is lower than that of DB1. The classification accuracy is decreased due to changes in the plankton shape depending on the environment at the sampling site and the large amount of debris in the sample. Focusing on conventional methods, MC is more accurate than ResNet-50 and ABN, since MC classifies images using a hierarchical structure of labels. The proposed method has the highest classification accuracy when all labels in the five ranks are used, since H-ABN takes into account the shape of the plankton step by step from *class* to *species*.

Fig. 4 shows examples of plankton images misclassified by the proposed method. Fig. 4 (a) shows the misclassified plankton image, the estimated label and the ground-truth label, and Fig. 4 (b) shows the plankton image corresponding to the estimated label of (a). The proposed method correctly classifies up to *genus* and incorrectly in *species* since the global shape is the same. These plankton images are misclassified even by experts since the species differ due to differences in the length of the plankton beard and the shape of the organ. Hierarchical image classification by H-ABN allows us to fully take into account the shape of the plankton. On the other hand, it is not always possible to distinguish local structural differences, so we are planning to develop a mechanism that can take into account detailed differences in plankton shape.

## 6 Conclusion

We proposed the plankton image classification method using Hierarchical Attention Branch Network (H-ABN) to take into account the taxonomic ranks of the plankton. We also constructed a plankton image dataset by annotating labels based on the taxonomic ranks to samples collected in Japanese lakes and marshes to classify plankton images according to the taxonomic ranks, We demonstrated the effectiveness of the proposed method in plankton image classification through performance evaluation experiments using the constructed plankton image dataset. In the future, we plan to investigate a classification method that focuses on the fine structure of plankton and an automatic plankton detection method from plankton community images.

## 7 Acknowledgments

We thank Ryotaro Yoshida for his contribution to this study. This work was supported by ERCA ERTDF Program Grant Number JPMEERF20214003.

**Table 3.** Results of the experiments using DB1, where the values in bold indicate the highest values of accuracy [%] and F1 score [%] at each rank.

| Method | *Class* Acc. [%] F1 [%] | *Order* Acc. [%] F1 [%] | *Family* Acc. [%] F1 [%] | *Genus* Acc. [%] F1 [%] | *Species* Acc. [%] F1 [%] |
|---|---|---|---|---|---|
| ResNet-50 [10] | 99.3 | 98.6 | 94.2 | 93.1 | 90.5 |
| | 99.2 | 95.3 | 83.2 | 79.1 | 74.9 |
| ABN [6] | 99.5 | 98.8 | 94.2 | 92.9 | 90.5 |
| | 99.5 | 98.1 | 85.2 | 80.0 | 76.6 |
| MC [4] | 99.5 | 98.6 | 93.9 | 93.0 | 90.5 |
| | 99.5 | 94.5 | 80.8 | 78.5 | 74.4 |
| Proposed (G,S) | **99.6** | 98.8 | 94.5 | 93.6 | 91.2 |
| | **99.6** | 96.6 | **83.1** | 79.4 | 77.5 |
| Proposed (F,G,S) | 99.4 | 98.7 | 94.2 | 93.0 | 90.4 |
| | 99.3 | 97.7 | 86.1 | 83.1 | **79.2** |
| Proposed (O,F,G,S) | 99.5 | **98.9** | **94.9** | **94.1** | **91.8** |
| | 99.5 | **97.8** | 83.0 | 80.1 | 77.8 |
| Proposed (C,O,F,G,S) | 99.5 | 98.8 | 94.5 | 93.4 | 90.8 |
| | 99.5 | 96.9 | 84.1 | 80.0 | 76.9 |

**Table 4.** Results of the experiments using DB2, where the values in bold indicate the highest values of accuracy [%] at each rank.

| Method | *Class* Acc. [%] | *Order* Acc. [%] | *Family* Acc. [%] | *Genus* Acc. [%] | *Species* Acc. [%] |
|---|---|---|---|---|---|
| ResNet-50 [10] | 88.4 | 82.2 | 74.2 | 64.3 | 59.1 |
| ABN [6] | 88.4 | 82.5 | 73.2 | 64.8 | 59.5 |
| MC [4] | 88.5 | 83.1 | 74.8 | 68.4 | 63.0 |
| Proposed (G,S) | 88.8 | 84.0 | 75.8 | 65.6 | 60.2 |
| Proposed (F,G,S) | 90.3 | 84.8 | 77.5 | 68.6 | 61.9 |
| Proposed (O,F,G,S) | 92.6 | 86.4 | 79.0 | 69.9 | 64.0 |
| Proposed (C,O,F,G,S) | **93.4** | **88.8** | **80.7** | **71.6** | **65.4** |

## References


1. Kaggle. `https://www.kaggle.com/c/datasciencebowl`
2. WHOI-Plankton. `https://hdl.handle.net/1912/7342`
3. Cheng, K., Cheng, X., Wang, Y., Bi, H., Benfield, M.C.: Enhanced convolutional neural network for plankton identification and enumeration. PLOS ONE **14**(7), e0219570–1–17 (Jul 2019)
4. Dhall, A., Makarova, A., Ganea, O., Pavllo, D., Greeff, M., Krause, A.: Hierarchical image classification using entailment cone embeddings. Proc. IEEE Conf. Computer Vision and Pattern Recognition pp. 836–837 (Jun 2020)
5. Ellen, J.S., Graff, C.A., Ohman, M.D.: Improving plankton image classification using context metadata. Limnology and Oceanography: Methods **17**(8), 439–461 (Aug 2019)

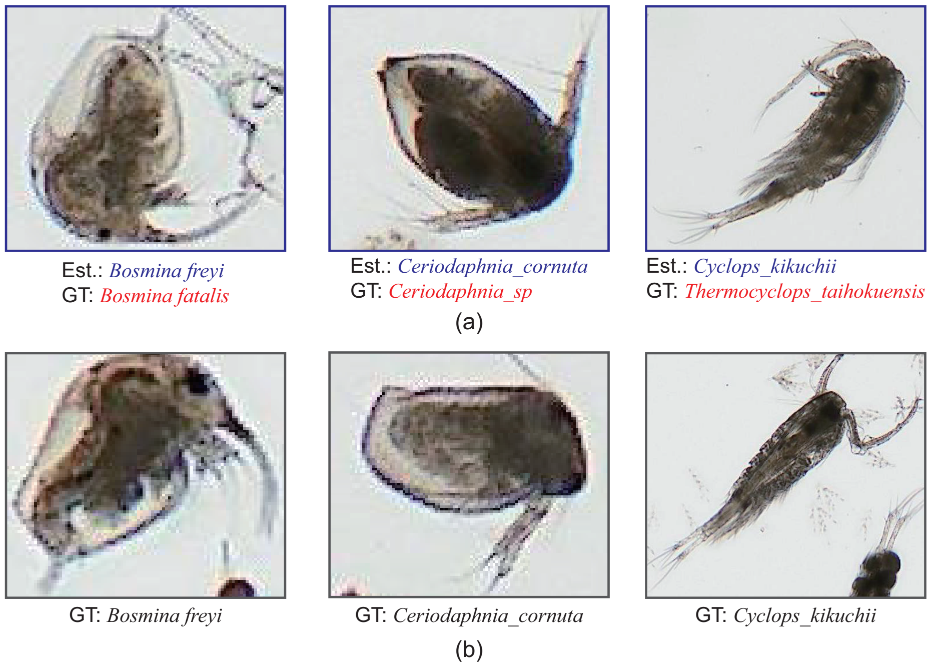


**Fig. 4.** Examples of plankton images misclassified by the proposed method: (a) misclassified plankton image, estimated label, and ground-truth label and (b) plankton image corresponding to the label estimated in (a).


6. Fukui, H., Hirakawa, T., Yamashita, T., Fujiyoshi, H.: Attention branch network: Learning of a attention mechanism for visual explanation. Proc. IEEE Conf. Computer Vision and Pattern Recognition pp. 10705–10714 (Jun 2019)
7. González, P., Castaño, A., Peacock, E.E., Díez, J., Del Coz, J.J., Sosik, H.M.: Automatic plankton quantification using deep features. J. Plankton Research **41**(4), 449–463 (Jul 2019)
8. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016)
9. Gorsky, G., Ohman, M.D., Picheral, M., Gasparini, S., Stemmann, L., Romagnan, J., Cawood, A., Pesant, S., Garcíacomass, G., F., P.: Digital zooplankton image analysis using the zooscan integrated system. J. plankton research **32**(3), 285–303 (Mar 2010)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. Proc. IEEE Conf. Computer Vision and Pattern Recognition pp. 770–778 (Jun 2016)
11. Kyathanahally, S.P., Hardeman, T., Merz, E., Bulas, T., Reyes, M., Isles, P., Pomati, F., Baity-Jesi, M.: Deep learning classification of lake zooplankton. Frontiers in microbiology **12**, 3226–3238 (Nov 2021)
12. Lumini, A., Nanni, L.: Deep learning and transfer learning features for plankton classification. Ecological Informatics **51**, 33–43 (May 2019)
13. Lumini, A., Nanni, L., Maguolo, G.: Deep learning for plankton and coral classification. Applied Computing and Informatics pp. 1–19 (Aug 2020)

14. Luo, J.Y., Irisson, J.O., Graham, B., Guigand, C., Sarafraz, A., Mader, C., Cowen, R.K.: Automated plankton image analysis using convolutional neural networks. Limnology and Oceanography: Methods **16**(12), 814–827 (Dec 2018)
15. Nesterov, Y.: A method of solving a convex programming problem with convergence rate $O(1/k2)$. Soviet Mathematics Doklady **27**(2), 372–376 (1983)
16. Silva, N.L., Marcolin, C.R., R., S.: Using image analysis to assess the contributions of plankton and particles to tropical coastal ecosystems. Estuarine, Coastal and Shelf Science **219**, 252–261 (Apr 2019)
17. Sosik, H.M., Olson, R.J.: Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. Limnol. Oceanogr. Methods **5**(6), 204–216 (Jun 2007)
18. Suthers, I., Rissik, D., Richardson, A.: Plankton: A guide to their ecology and monitoring for water quality. CSIRO Publishing (2019)
19. Yang, L., Luo, P., Loy, C.C., Tang, X.: A large-scale car dataset for fine-grained categorization and verification. Proc. IEEE Conf. Computer Vision and Pattern Recognition pp. 3973–3981 (Jun 2015)
20. Zheng, H., Wang, R., Yu, Z., Wang, N., Gu, Z., Zheng, B.: Automatic plankton image classification combining multiple view features via multiple kernel learning. BMC Bioinformatics **18**(16), 1–18 (Dec 2017)
21. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. Proc. AAAI Conf. Artificial Intelligence **34**(7), 13001–13008 (Feb 2020)
22. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Cutmix: Regularization strategy to train strong classifiers with localizable features. Proc. IEEE Conf. Computer Vision and Pattern Recognition pp. 6023–6032 (Jun 2019)