Deep Active Learning for In Situ Plankton Classification

Erik Bochinski[1], Ghassen Bacha[1], Volker Eiselein[1], Tim J. W. Walles[2], Jens C. Nejstgaard[2], and Thomas Sikora[1]

[1] Communication System Group, Technische Universität Berlin
{bochinski,sikora}@nue.tu-berlin.de
[2] Leibniz-Institute of Freshwater Ecology and Inland Fisheries
{walles,nejstgaard}@igb-berlin.de

Abstract. Ecological studies of some of the most numerous organisms on the planet, zooplankton, have been limited by manual analysis for more than 100 years. With the development of high-throughput video systems, we argue that this critical bottle-neck can now be solved if paired with deep neural networks (DNN). To leverage their performance, large amounts of training samples are required that until now have been dependent on manually created labels. To minimize the effort of expensive human experts, we employ recent active learning approaches to select only the most informative samples for labelling. Thus training a CNN using a nearly unlimited amount of images while limiting the human labelling effort becomes possible by means of active learning. We show in several experiments that in practice, only a few thousand labels are required to train a CNN and achieve an accuracy-level comparable to manual routine analysis of zooplankton samples. Once trained, this CNN can be used to analyse any amount of image data, presenting the zooplankton community the opportunity to address key research questions on transformative scales, many orders of magnitude, in both time and space, basically only limited by video through-put and compute capacity.

Keywords: Classification · Zooplankton · Active Learning · automatic identification and sizing · Cost-Effective Active Learning · in situ

1 Introduction

The aquatic planktonic microorganisms are the basis of life in the largest part of our planet, the oceans and lakes, and thus of vital importance also for humanity by providing food, oxygen, as well as many other ecosystems services. However, the increasing pressure of humanity on the same ecosystems has created an urgent need to understand how these systems function and respond to the changing environment [2, 21]. To better understand this, it is critical to be able to measure fundamentals such as - who is where when and does what to whom.

T.J.W. Walles was supported by a Leibniz Competition grant ILES "Illuminating Lake Ecosystems", and technical investment was supported by IGB-startup funds to J.C. Nejstgaard.

For plankton this has been an enormous challenge due to a strong missmatch between available time-consuming manual analytic methods, and the need for high numbers of samples to cover the wide range of scales in time and space that the zooplankton operate in [6]. However, here we show that with effectively trained deep learning approaches to lab and in-situ image systems, this have a great promise to ease these key limitations.

In order to yield high quality data at high temporal and spatial scales, adequate tools for data acquisition as well as analysis are required. For the imagery, several camera-based systems have been developed in the recent past [1, 8] to gather considerable amounts of data in an automated way [15, 5]. Analysing these amounts of data by traditional methods, including manual annotation, are unfeasible. This raises the need for reliable automated labelling tools [14]. Otherwise it would not be possible to effectively explore this new scale of available imagery.

Recent advances in the computer vision domain, accelerated by the rediscovery of convolutional neural networks (CNNs) and deep learning methods in general [11], have greatly improved the possibilities for reliable automated labelling tools. One of the major challenges that prevents the direct application of such tools for successful plankton classification, is the general need for large, manually labelled training data. Since zooplankton appears in very many species, forms and stages, large training sets for each environment and image acquisition system would be needed. Furthermore, it is hard to decide which samples need to be manually annotated for the successful training of a CNN. This becomes even harder as there are commonly great class imbalances in plankton data [22, 12, 4], while CNN training generally benefits from balanced training data. It seems therefore natural to approach the problems of training data annotation and classification of large datasets in a joint way. This process is known as active learning [18]. It aims at training a classifier on a small initial training set and then selects the most informative samples from a larger unlabelled dataset to query them for labelling to e.g. a human expert for further training. Iteratively performed, this can greatly reduce the number of needed annotations till convergence. Furthermore, samples that were classified with high-confidence can also be included in the training set with a pseudo-label equally to the predicted class. This can further accelerate the training process and therefore reduce the need for manually created annotations.

The additional use of high-confidence samples in the training of CNNs for image classification in an active learning manner was first introduced in [23] as Cost-Effective Active Learning (CEAL). This approach was adapted later for face identification [13], melanoma segmentation [7] and cancerous tissue recognition [20]. The authors of [4] proposed using a CNN to classify microscopic grayscale images of plankton. Several different network architectures were explored and superior results over traditional approaches reported. A Generative Adversarial Network is used in [22] to generate additional training samples to handle the class imbalance often found in plankton datasets caused by the uneven distribution of plankton presence in nature. The same problem is tackled in [12] using CNNs and transfer learning. [9] analyzes different methods for segmenting the plankton from the background and its impact on the classification accuracy.

The contribution of this paper is manifold: We propose a novel CNN-based classification system that is robust against the background noise in images and therefore does not require any instance segmentation for background removal. To our knowledge, we are the first to report the successful application of Cost-Effective Active Learning to the task of zooplankton classification. Our experiments show that the proposed system is able to solve the classification problem with accuracy levels comparable to manual routine analysis of zooplankton samples, after just a few active learning iterations. We further apply the method to a second dataset captured with a different camera in a different lake environment with the presence of a new prominent phytoplankton distractor class. The previously trained system is able to successfully classify this new data. With just one additional active learning update the same performance level as of the original dataset is reached, showing the validity of the system. As the numbers of organisms counted and sized increases with orders of magnitudes this has several critical advances: 1) numbers of rare organisms, typically larger sized than the numerous smaller organisms, can now be better estimated. 2) the estimate of total biomass and size distributions will now be much more accurate due to much higher and more accurate automatic size estimates, 3) the statistical power will increase due to higher numbers, 4) the high throughput enables sampling and in situ analyses on a much higher spatio-temporal scale, offering completely new options for future (zoo)plankton studies like real-time in situ profiles.

2 Method

2.1 Active Learning

Given a dataset D, the aim of active learning for training a classifier's parameters $\mathcal{W}$ is to minimize the number of required annotated training samples $D^L \subseteq D$ from a larger, initially unlabeled dataset $D^U \subseteq D$ while still obtaining satisfactory training results. This becomes possible due to the assumption that not all samples are equally informative for training and can be exploited by an iterative process of training the classifier using already annotated samples D^L and then classifying all samples of the unlabeled set D^U with it. Samples with the lowest classification confidence are considered the most informative for the future training step. Therefore, the top K least confident predictions are queried to e.g. a human annotator for labeling and moved to D^L. This is repeated until a termination criteria is fulfilled, e.g. the training loss is converged or a certain accuracy of the classifier is reached.

To measure the confidence of a predicted sample, several confidence criteria were proposed in literature. We considered the 3 most common criteria in our approach:

Least confidence [18]:
$$x^*_{LC} = \max_j \left(P(y_i = j | x_i; \mathcal{W})\right) \tag{1}$$

with $P(y_i = j|x_i; \mathcal{W})$ denoting the probability of x_i belonging to the j-th class which translates to the most probable softmax classification score for the j-th category. The lower this value the higher the uncertainty.

Margin sampling [17]:
$$x^*_{MS} = P(y_i = j_1|x_i; \mathcal{W}) - P(y_i = j_2|x_i; \mathcal{W}) \tag{2}$$

The margin between the best ($P(y_i = j_1|x_i; \mathcal{W})$) and second best ($P(y_i = j_2|x_i; \mathcal{W})$) classification score is used as confidence value. Lower values represent a higher grade of uncertainty.

Entropy [19]:
$$x^*_{EN} = -\sum_i P(y_i = j|x_i; \mathcal{W}) \log P(y_i = j|x_i; \mathcal{W}) \tag{3}$$

The entropy criterion inspired by information theory takes all predicted class labels into account and is accordingly defined by the entropy of the set of all classification scores $P(y_i = j|x_i; \mathcal{W})$. Higher values denote higher uncertainties.

2.2 Cost-Effective Active Learning

The plain active learning approach only takes advantage of the most informative samples and their retrieved labels. For cost-effective active learning [23], the class predictions of the high-confidence samples are used as pseudo-labels and added temporarily to D^L for the next training step of the classifier as they can still contribute to the training process. In [23] it was proposed to use the entropy criteria for the selection of the high-confidence samples. We consider all three as possible criteria in our experiments for pseudo-labeling:

$$j^* = \arg\max_j \left(P(y_i = j|x_i; \mathcal{W})\right) \tag{4}$$

$$y_i = \begin{cases} j^*, & x^*_{LC,MS} > \delta \text{ or } x^*_{EN} < \delta \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

with:
$$\delta = \begin{cases} \delta_0, & t = 0 \\ \delta + dr \cdot t, & t > 0 \end{cases} \tag{6}$$

as a threshold for the selection of the high-confidence samples with δ_0 as the start value and dr as the decay rate. The decaying threshold allows for the selection of more high-confidence samples at the beginning and enforces a higher confidence of the samples towards the end of the active learning process.

2.3 Classifier

As classifier, a Convolutional Neural Network (CNN) as shown in Figure 1 is employed. The architecture is based on the popular AlexNet [11] and was carefully adapted to fit the requirements of the used plankton image data. ReLU activation is performed after each layer except for the last one which consists of

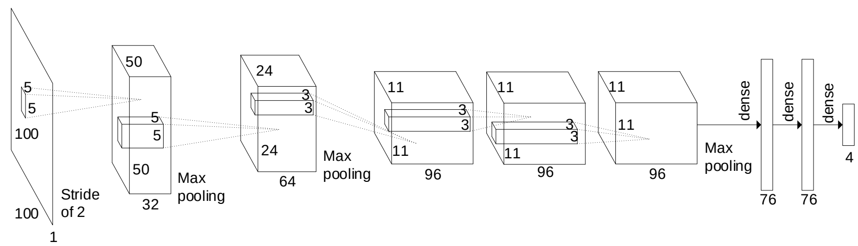


Fig. 1: Employed CNN architecture

C softmax activated outputs where C denotes the number of possible classes. A dropout of 50% was used while training the dense layers to prevent overfitting. As loss function the cross-entropy is employed:

$$\mathcal{L}(\mathbf{x}, y, \mathcal{W}) = -\sum_{c=1}^{C} \mathbb{1}_{y=c} \log P(y = j|\mathbf{x}; \mathcal{W}) \tag{7}$$

where $\mathbb{1}_{y=c}$ is 1 if y equals c or otherwise 0 and $P(y = j|\mathbf{x}; \mathcal{W})$ the softmax output of the CNN for the jth class. The task is then to find a parameter set $\mathcal{W}$ of the CNN satisfying the following minimization problem in each active learning iteration T:

$$\min_{\{\mathcal{W}, n=|D|\}} = \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(\mathbf{x}_i, y_i, \mathcal{W}) \tag{8}$$

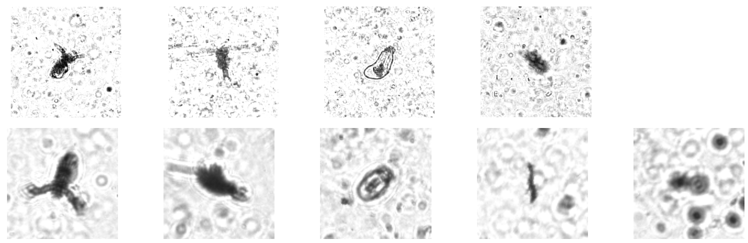
The non-optimal optimization can then be performed using ordinary gradient descent methods.

3 Experiments

Several experiments were conducted in order to assess the effectiveness of cost-effective active learning for training a CNN for zooplankton classification. Two datasets from different biological environments were captured and analysed. The first dataset is used to analyse the achievable accuracy of the CNN and how the cost-effective active learning can be used to minimize the number of required annotations. The second dataset is used to examine the generalization ability of the CNN and if the CEAL method can be used to fine-tune the system to adapt to the characteristics of this new data.

3.1 Datasets

All images used in the experiments were captured using the in situ imager described in [3]. The first dataset ILES contains about 840K images. They were



(a) Cladocera (b) Copepod (c) Rotifer (d) Junk (e) Distractors

Fig. 2: Example images of the ILES (top row) and CZECH (bottom row) datasets. Different scales, rotations and background clutter pose challenges for the classification system.

divided into 4 classes: *cladocera, copepod, rotifer* and *junk* whereas the latter one comprises everything which can not be certainly assigned to one of the other classes. The dataset is further split into two subsets A and B. Subset A comprises about 60K fully annotated images in order to evaluate the cost-effective active learning framework with 80% of the data as training and the remaining 20% as testing data. Subset B contains the other 780K images and was used to validate the approach and further investigation concerning suitability of the results for analyses from the biological perspective. The second dataset CZECH contains about $167K$ all unlabelled images. It was collected by the same imager but equipped with a different camera and is used to further validate the approach using data captured in a different environment. The major challenge is the prominent presence of a new phytoplankton distractor class shown in Figure 2e. A total of $10K/5K$ randomly selected images of ILES B/CZECH were labelled for performance analysis.

3.2 Preprocessing

Only a two dimensional projection of the three dimensional object with an unknown spacial orientation serves as input image for classifying the object. This greatly differs to most of the well-researched computer vision applications as they are focused to non-microscopic and non-aquatic images. To support the CNN to deal with the induced rotation invariance, we use a pre-processing pipeline as sketched in Figure 3. First, the approximate shape of the object is determined by otsu thresholding [16]. The main axis is then calculated using principal component analysis. The image is rotated in a way that the rotation of this axis is strictly vertical. Finally, the center crop of the resulting image is used. Note that the segmentation using otsu thresholding is only used for rotation but not for cropping or removal of the background (as eg. in [9]) as we find that no current segmentation approach is able to perform this task reliably enough to not remove fine-grained details belonging to the planktonic object. Instead we rely on the CNN to learn the differentiation between background and foreground features during training. The ILES A training set consists of about 20K samples for each

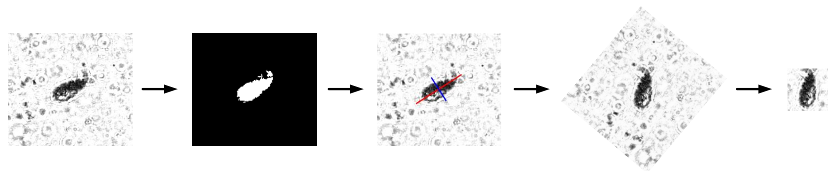


Fig. 3: Preprocessing pipeline for the images.

of the *cladocera* and *junk* classes, about 5K samples for *copepod* and only $1.3K$ samples belonging to the $rotifer$ class. To handle this imbalance during training, the *copepod* images were mirrored horizontally and vertically, quadrupling the amount of training samples to a total of $20K$. Also $20K$ samples for *rotifer* were created by mirroring and rotating to produce 15 samples per image including the original. The rotations render the normalization regarding the orientation useless, however there is no practical impact to the training since the $rotifer$ individuals have a distinct round shape (see Fig. 2c) which still causes that class to be the most reliable during classification.

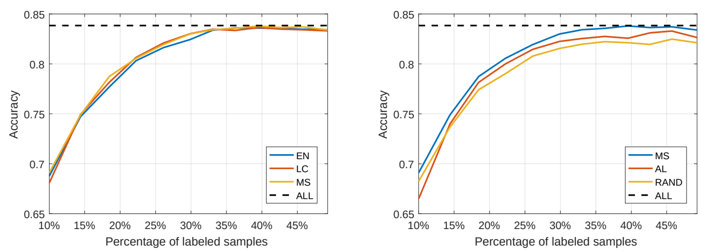
3.3 Training procedure

The first training evaluation was performed using the ILES A dataset. Since this dataset is fully annotated, the active learning strategies can be simulated.

As initial labelled training set D^L, 10% of the augmented training data was used. Training is performed for 11 active learning iterations T with $K = 2200$. An initial threshold $\delta_0 = 0.995$ and a decay rate of $dr = 0.33 \cdot 10^{-5}$ was chosen for least confidence (LS) and margin sampling (MS) and $\delta_0 = 0.4 \cdot 10^{-4}$, $dr = -0.2 \cdot 10^{-5}$ for entropy-based (EN) high-confidence sample selection. In each iteration, the CNN was trained for 10 epochs with a batch size of 32 on D^L including the temporarily added high-confidence samples. Training was performed using the adam algorithm [10] with a learning rate of 10^{-4}. The remaining parameters are set to the default values of $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$ in all experiments.

3.4 Results

First, the maximum achievable CNN accuracy was determined by training on dataset ILES A without employing any active learning strategy. Otherwise, the training procedure previously described in 3.3 was followed. An accuracy of 83.84% was achieved and serves as baseline for the subsequent experiments using the active learning approaches. Figure 4a shows the development of the CNN accuracy depending on the percentage of labelled training samples for the EN, LC and MS cost-effective active learning strategies. It can be seen that there is no significant difference in the performance of the different confidence metrics. Margin sampling however was slightly better than the other two strategies and was therefore selected for the subsequent experiments. In Figure 4b, a comparison between CEAL with margin sampling (MS), plain active learning (AL) and a random selection of samples queried for labelling is shown. In addition, ALL denotes the baseline accuracy without any active learning. It can be observed that CEAL performs favourably and is the only method reliably reaching the peak accuracy of 83.84%. This is achieved while only requiring labels for one third of the available training samples which translates to about $16K$ required labels in total.



(a) Different confidence metrics (b) Different selection methods

Fig. 4: Comparison of different sample selection methods and confidence metrics.

3.5 Generalization

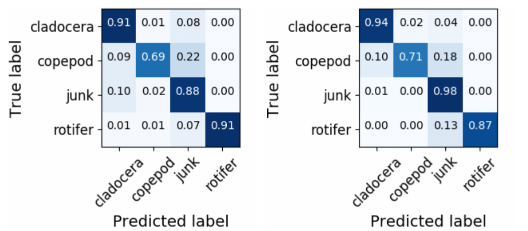
In addition to the experiments on the fully labelled dataset, experiments under real conditions were conducted. The final model from the previous experiments trained on the dataset ILES A is evaluated on the ILES B and CZECH datasets. Initially, a drop in the accuracy for the ILES B dataset is noticed. For the CZECH dataset, a high accuracy of about 91% is achieved. The reason is the prominent presence of a new phytoplankton class which is mostly classified correctly as *junk*, which makes up about 88% of the whole test set. Following [15], the predictions were also evaluated using the unweighted F1 score to get more meaningful results. This reveals that the performance on the CZECH dataset is with an F1 score of 0.55 indeed considerably worse than on the ILES A set with 0.85.

To adopt the CNN classifier to the two new unlabelled datasets, a final cost-effective active learning iteration was performed. In order to do so, 5.35% of the ILES B and 3.37% of the CZECH dataset were labelled by a human expert. The samples with the least margin sampling confidence were selected. Additionally, the high-confidence samples using the same confidence metric were pseudo-labelled as in the previous experiment. With these new samples, the CNN was fine-tuned. The results are presented in Table 1.

The accuracy as well as the F1 score improved. In the case of the ILES B, the performance compared to the initial ILES A was even slightly increased. For the more challenging CZECH dataset, the accuracy was raised to 96% and the F1 score is with 0.80 only slightly behind the ILES datasets. Still, this result shows that the approach is able to adapt not only to the different lake environment of the CZECH dataset, but to a different camera setup with different visual properties (see Fig. 2) as well. In Figure 5, confusion matrices for both datasets are presented. It shows that the *cladocera* and *rotifer* classes are predicted most robustly while the performance for *rotifer* is slightly decreased for the CZECH dataset. Most likely this is due to the visual similarity to the new phytoplankton distractor only present in this dataset. Otherwise, the *copepod* and *junk* classes are confused sometimes. The most probable reason for this is that there are many *copepod* samples which are not focused properly, rendering the antennas as most distinguishing visual feature invisible.

Dataset	Accuracy	F1
ILES A	83.84%	0.85
ILES B (no fine-tuning)	72.78%	0.75
ILES B (with fine-tuning)	86.23%	0.86
Czech (no fine-tuning)	90.87%	0.55
Czech (with fine-tuning)	96.08%	0.80

Table 1: Comparison of the classification performance for all datasets.



(a) *ILES B* (b) *CZECH*

Fig. 5: Confusion matrices for both validation datasets.

4 Conclusion

In this paper, we investigated the adaptability of the cost-effective active learning (CEAL) approach to train a Convolutional Neural Network (CNN) for the task of zooplankton classification. Various experiments showed that CNNs are indeed suitable for this task. With CEAL, just a fraction of the samples need to be annotated by a human expert to reach the maximum possible accuracy of the CNN. It was shown that the system is further capable of adapting to different camera setups, lake environments and is even robust to a new, unseen distractor class. Hence, the proposed approach contributes to close the gap between automated, large-scale image data acquisition systems and the actual interpretation of the data from the biological application side by efficiently inferring the required annotations automatically. This contribution reduces the effort of training and using such systems significantly. When adapted in the ecological research community together with the necessary optical equipment this will yield possibilities to acquire key zooplankton data at an unprecedented magnitude in both temporal and spatial scales, in its turn expected to create transformative changes in plankton ecology in the near future.

References

1. Benfield, M.C., Grosjean, P., Culverhouse, P.F., Irigoien, X., Sieracki, M.E., Lopez-Urrutia, A., Dam, H.G., Hu, Q., Davis, C.S., Hansen, A., et al.: Rapid: research on automated plankton identification. Oceanography (2007)
2. Connor, R.: The United Nations world water development report 2015: water for a sustainable world. UNESCO Publishing (2015)

3. Cowen, R.K., Guigand, C.M.: In situ ichthyoplankton imaging system (isiis): system design and preliminary results. Limnology and Oceanography: Methods (2008)
4. Dai, J., Wang, R., Zheng, H., Ji, G., Qiao, X.: Zooplanktonet: Deep convolutional network for zooplankton classification. In: OCEANS (2016)
5. Douglas, K., Juniper, S., Jenkyns, R., Hoeberechts, M., Macoun, P., Hillier, J.: Developing spatial management tools for offshore marine protected areas. In: OCEANS–Anchorage (2017)
6. Eileen E. Hofmann, J.M.K.: Future marine zooplankton researcha perspective. Marine Ecology Progress Series (2001)
7. Gorriz, M., Carlier, A., Faure, E., Giro-i Nieto, X.: Cost-effective active learning for melanoma segmentation. arXiv preprint arXiv:1711.09168 (2017)
8. Gorsky, G., Ohman, M.D., Picheral, M., Gasparini, S., Stemmann, L., Romagnan, J.B., Cawood, A., Pesant, S., García-Comas, C., Prejger, F.: Digital zooplankton image analysis using the zooscan integrated system. Journal of plankton research (2010)
9. Hirata, N.S., Fernandez, M.A., Lopes, R.M.: Plankton image classification based on multiple segmentations. In: ICPR CVAUI Workshop (2016)
10. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: International Conference for Learning Representations (2015)
11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS (2012)
12. Lee, H., Park, M., Kim, J.: Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning. In: IEEE International Conference on Image Processing (2016)
13. Lin, L., Wang, K., Meng, D., Zuo, W., Zhang, L.: Active self-paced learning for cost-effective and progressive face identification. IEEE Transactions on Pattern Analysis and Machine Intelligence (2018)
14. MacLeod, N., Benfield, M., Culverhouse, P.: Time to automate identification. Nature (2010)
15. Orenstein, E.C., Beijbom, O., Peacock, E.E., Sosik, H.M.: Whoi-plankton-a large scale fine grained visual recognition benchmark dataset for plankton classification. arXiv preprint arXiv:1510.00745 (2015)
16. Otsu, N.: A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics (1979)
17. Scheffer, T., Decomain, C., Wrobel, S.: Active hidden markov models for information extraction. In: International Symposium on Intelligent Data Analysis (2001)
18. Settles, B.: Active learning literature survey. 2010. Computer Sciences Technical Report 1648
19. Shannon, C.E.: A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review (2001)
20. Stanitsas, P., Cherian, A., Truskinovsky, A., Morellas, V., Papanikolopoulos, N.: Active convolutional neural networks for cancerous tissue recognition. In: IEEE International Conference on Image Processing (2017)
21. Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., et al.: Global threats to human water security and river biodiversity. Nature (2010)
22. Wang, C., Yu, Z., Zheng, H., Wang, N., Zheng, B.: Cgan-plankton: Towards large-scale imbalanced class generation and fine-grained classification. In: IEEE International Conference on Image Processing (2017)
23. Wang, K., Zhang, D., Li, Y., Zhang, R., Lin, L.: Cost-effective active learning for deep image classification. IEEE TCSVT (2017)