Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 54-16 (2021) 450–457

CIRAL: a hybrid active learning framework for plankon taxa labeling

Martin Lund Haug ** **Aya Saad** *** **Annette Stahl** ****

Dept. of Engineering Cybernetics, Norwegian University of Science and Technology, NTNU, Trondheim, Norway
** *(e-mail: martlh@stud.ntnu.no)*
*** *(e-mail: aya.saad@ntnu.no)*
**** *(e-mail: annette.stahl@ntnu.no)*

Abstract: With the complex structure of planktonic species and an immense amount of data captured from autonomous underwater vehicles (AUVs), a large burden is placed on the domain experts for plankton taxa labeling. At the same time, the most prominent machine learning (ML) methods for classification rely heavily on a massive amount of labeled datasets to create and train neural network classifier models that perform their tasks accurately. Active Learning (AL) is an ML paradigm that reduces this manual effort by proposing algorithms that support the construction of the training datasets, thus enlarging the sets while minimizing human involvement. To build the training set, AL methods apply heuristics to select a subset of images, i.e., samples, from the entire data. The selected samples that capture the common statistical patterns or feature space are likely to include all the information needed for the training and the learning processes. In addition, the algorithm should prioritize samples that are likely belonging to multiple classes, i.e., having close inter-class boundaries, and might lead to model confusion. Many of the current AL approaches fail to incorporate both types of samples representing the statistical pattern and the samples in which the particular machine learning model is uncertain about.

In this paper, we extend our framework which addresses these challenges with an augmentation module to increase the robustness of the model and ensure its adaptability to the planktonic domain. We compare the framework with existing hybrid AL techniques and test an adaption of our extended framework on the planktonic domain. The empirical results from the experiments exerted in this paper confirm higher accuracy achieved by the new extended framework.

Keywords: image analysis, deep learning, plankton taxa distribution, active learning, computer vision

1. INTRODUCTION

Planktonic species are critically important to the oceanic ecological structure as they are the basis of the aquatic food web. Hence, by studying temporal variations in plankton taxa distributions, one can achieve a proxy for the development of the oceanic ecosystems.

Progress in the development of autonomous underwater vehicles (AUV) and robotic visual sensing enables the possibility of capturing large amounts of planktonic image data. Further, Convolutional Neural Network (CNN) models have proved competent at solving computer vision problems in the supervised Machine Learning (ML) paradigm. Embedding CNN models into AUV enables the identification of plankton taxa distributions in-situ. However, modern CNNs require an immense amount of pre-classified labeled input in order to achieve satisfactory classification performance. Since plankton biomass appears in many different species, forms, and stages depending on the geographical environment and season, pre-classified training data has to be constructed for each different geographical environment, season, and image-acquiring system. Consequently, much effort is needed in the manual plankton taxa labeling with a constrained budget that requires domain expertise, i.e., biologists, to identify the complex structure of planktonic organisms.

Active Learning (AL) is a semi-supervised machine learning approach that aims at mitigating this burden placed on domain experts. The key idea of AL is to capture the data distribution of the full dataset with only a fraction of the samples. This is possible from the fact that not all images bring equal amounts of information to the image classifier (Vodrahalli et al., 2018).

Existing AL models in the literature can be classified based on the unlabeled data readiness and the sampling pool chosen. In other words, when data arrives in streams, the AL model is considered as a stream-based model, (Krishnamurthy, 2002), while it is pool-based otherwise (Lewis and Gale, 1994). Further, the AL models' mode of sampling varies between batch-mode (Ash et al., 2019) or single-mode (Lewis and Gale, 1994) depending on the number of data samples presented and chosen at each labeling round. With the recent developments of CNNs, batch-mode sampling has become increasingly relevant as

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2021.10.130

it is not computationally feasible to update a large network with single data points.

The most important distinction between the different sampling modes aforementioned is in their prioritization between informative and representative samples. While the former aims to prioritize samples that are at the proximity of the inter-class decision boundaries, the latter exploits the feature space of the data points to best capture the statistical patterns of the data. There exists a broad literature on Active Learning. The reader can refer to the survey presented in (Settles, 2009), and more recently, the survey on deep learning version of AL techniques is elaborated in (Ren et al., 2020).

The promise of removing the bottleneck of manual labeling in machine learning pipelines in addition to progress in the development of deep learning models has brought a surge in AL research. AL has been proven to be an efficient method of querying informative samples from an unlabeled pool of data points (Gal et al., 2017; Yoo and Kweon, 2019). Further, other approaches focusing on exploiting the latent-space structure of unlabeled samples have also been successfully proposed (Sener and Savarese, 2018). Furthermore, hybrid methods combining the informative and representative metric have become increasingly popular among researchers over the later years (Hsu and Lin, 2015). Still, much of the existing AL methods lack efficient utilization of the latent-space structure and often suffer from high correlation among queried samples. Moreover, by only incorporating model-based query methods, many existing AL approaches lack transferability to other deep learning models. In (Vodrahalli et al., 2018), the authors investigated how different datasets had unequal amounts of information distributed among the images. In some cases a few samples were enough to represent the full distribution of the dataset yet in other cases this proved not to be true. The success of active learning often depends on the information distribution of the dataset; hence, it is rarely possible to rely on either representative or informative sampling.

To address this issue, we proposed in Haug et al. (2021) a combined representative and informative active learning (CIRAL) approach that incorporates the full feature space in the early cycles of querying and puts more weight on samples at the proximity of the inter-class decision boundaries at the later cycles. We compared the novel hybrid framework with informative and representative approaches. We proved that this hybridization outperforms the classical AL approaches under the two categories in terms of the overall model accuracy on the CIFAR dataset with minimal possible data presented to the model. The CIFAR dataset was the most utilized in the literature as a benchmark for performance comparison and as a proof of concept.

The aim behind the proposed hybridization is threefold: 1) the model will have a good initialization from incorporating the full feature space in the early rounds of querying and training. 2) Adding diversity sampling to the queried uncertainty samples prevents redundant labeling representation from the same area of uncertainty. 3) As the softmax layer on neural networks has shown to be a bad proxy for the uncertainty of neural networks (Ren et al., 2020), an adversarial active learning method is employed. This method has previously shown good results (Ducoffe and Precioso, 2018), however, it was not employed with sub-modular heuristics as is done in this work.

The contributions in this paper are twofold:

- First, we compare the performance of the novel hybrid framework with other well-known hybrid methods and show that it achieves better accuracy.
- Second, we extend the originally proposed framework with a data augmentation module to increase the robustness of the model and to ensure the adaptability of the proposed semi-supervised method to the plankton domain with the goal to minimize the burden on domain experts.

The experiments in this paper are conducted on subsets of the plankton dataset from National Data Science Bowl (kag, 2015) and the CIFAR dataset (Krizhevsky, 2009). The ResNet-18 architecture is employed as the learning network model (He et al., 2015) for the CIFAR, whereas a custom network is made for the plankton dataset.We further created a pre-processing module to adapt the images to the deep learning models employed in this paper and speed up the convergence of the training process. Pre-processing operations include normalization of pixel values and resizing of input images to a fixed dimension. Further, as opposed to many other AL studies (Mittal et al., 2019), we employ regularization techniques in order to enhance the classification performance of the AL models and improve their robustness. More specifically, we employ a random horizontal and vertical flip and a random affine transformation.

The rest of the paper is organized as follows. Section 2 introduces some preliminary knowledge related to this paper. Section 3 presents related work in the area of AL, emphasizing hybrid and plankton-specific AL methods in particular. Section 4 explains our proposed algorithmic framework. Section 5 presents the experimental results. In Section 6, a conclusion is made on the contributions of this paper and also future directions are presented.

2. BACKGROUND

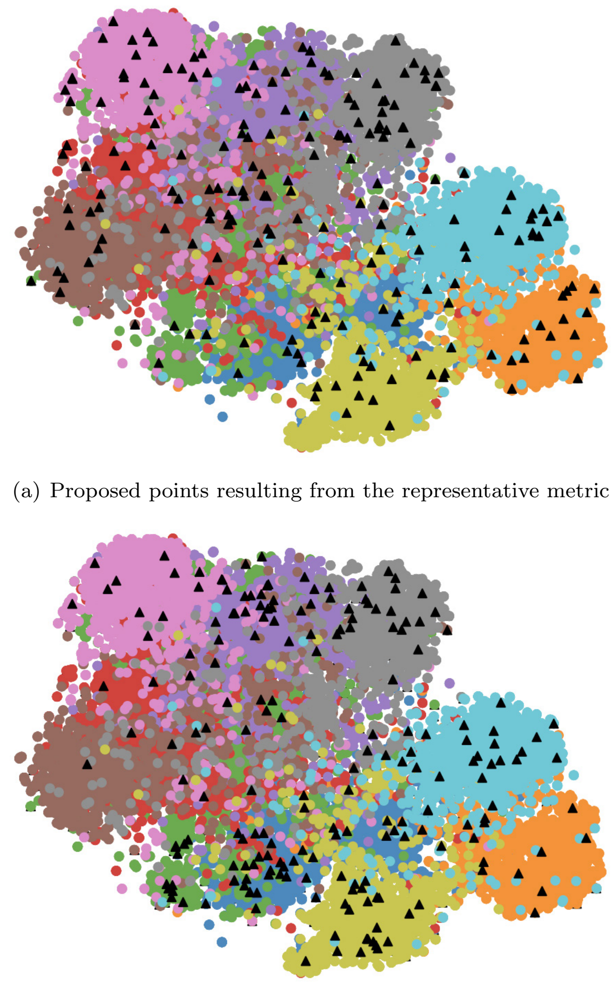
Active Learning is a type of semi-supervised learning that provides classification accuracy competitive with fully-supervised learning approaches while having the benefits of minimal human interaction from unsupervised learning. The main principle is to iteratively pick subsets from the available unlabeled data in order to build a training set for a machine learning model. As described in the previous section, the query methods of active learning can be primarily categorized into methods that exploit the feature of the data and methods that search for samples the machine learning model finds informative. A way of finding the latter has often been done by prioritizing samples the learning model is uncertain about, e.g. samples in the proximity of the inter-class decision boundaries.

A large number of methods for finding uncertainty samples have been proposed in recent years due to their simplicity and comprehensiveness. Many of these have been based on the softmax layers of CNNs as a proxy for the networks' uncertainty. Such an approach was proposed by

Wang et al. (2017), who in addition pseudo-labeled high confidence samples for additional robustness. However, research has shown that these softmax probabilities work as a bad proxy for the confidence of neural networks (Ren et al., 2020), and will often lead to worse performance than random benchmark sampling. Consequently, other ways of measuring the uncertainty of neural networks have been proposed in the later years. Gal et al. (2017) proposed a way of creating an ensemble of network architectures by using Monte Carlo dropout and measure the disagreement in prediction among the networks. A conceptually equal method has also been studied in (Beluch et al., 2018), where the authors employed an ensemble of different CNNs instead of the Monte Carlo dropout. A drawback with the latter ensemble method is the computational effort that is increasing with the dimensions of the learning network and the number of unlabeled samples.

A different approach from using the classification results of the learning networks has been proposed by Tong and Koller (2001) to calculate the distance to the inter-class decision boundary. Samples lying close to the decision boundary are considered to be informative for the machine learning model as they can help to fine-tune the model parameters. However, as it is feasible for support vector machines (SVM), it is a more complex operation for CNNs. Nevertheless, to transfer this approach to CNNs, Ducoffe and Precioso (2018) proposed a way of measuring the distance by making adversarial attacks and find which of the images change the classification. By ranking the size of the perturbation needed to change the sample classification, one can get a proxy on how far a given sample is from the decision boundary. This method looks at the input to the network rather than the soft-max layer as done in (Wang et al., 2017). However, both of the latter methods query the topmost uncertain images. As can be seen in figure 1 (b) and also stated in (Sener and Savarese, 2018), uncertainty sampling tends to lead to high correlation among the samples leading to a lack of utilization of the data distribution and also the labeling of redundant samples.

From figure 1 (a) one can observe that by employing a representative metric to exploit the full feature space of the available data points, this problem can be overcome. A large number of methods for finding such representative samples have been investigated over the later years. They can be roughly divided into categories that try to exploit the feature space and others that aim to maximize some performance metric. An example of the latter is, as proposed in (Pinsler et al., 2021), a method that approximates the complete data posterior of model parameters that produce diverse batches. By selecting subsamples, the method tries to lower the expected value of the loss function. An example of the former is, as proposed in (Geifman and El-Yaniv, 2017), a diversity method that performs a farthest-first traversal to cover the feature space. A similar example is shown in (Sener and Savarese, 2018) proposing a core set method to find clusters based on the min-max facility location problem and then optimizing these clusters with mixed-integer programming.



(a) Proposed points resulting from the representative metric

(b) Proposed points resulting from the informative metric

Fig. 1. T-SNE plot of 200 samples queried with a representative metric and an informative metric. The different colored data points represent the images of the 10 different classes from the CIFAR dataset. With the T-SNE algorithm (van der Maaten and Hinton, 2008), the images are projected onto the two-dimensional feature space.

3. RELATED WORK

Two areas of active learning are related to our work. Firstly, other methods of hybrid active learning have been increasingly popular among researchers in later years. Kaushal et al. (2018) proposed a work of diversified subset selection that utilizes methods of least confidence, smallest margin, and highest entropy from the softmax probability distribution to find informative samples. To incorporate representative samples they used, similar to this work, min-max facility location in addition to disparity minimum. A similar approach was proposed by Zhdanov (2019) to increase diversity in mini-batch Active Learning. Their experiments reported that diversity-enhancing approaches outperformed a baseline of uncertainty sampling methods. They combined informative sampling with representative sampling by using the smallest margin sampling from the softmax layer as uncertainty metric and the k-means algorithm as a representative metric. In (Huang et al., 2018), the authors aim to fine-tune pre-trained networks with a

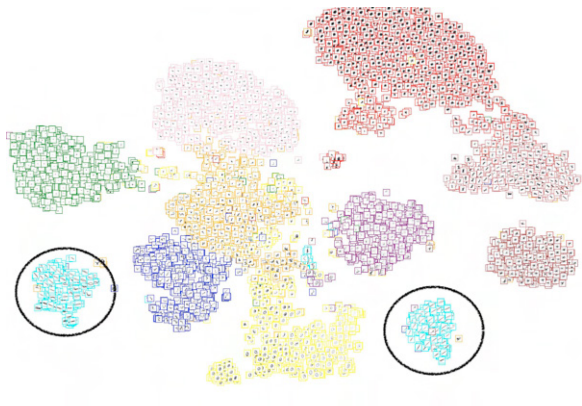


Fig. 2. Visualization of the plankton classes show how the Chaetognath Sagitta class is separated into two groups based on its orientation. The plankton images are projected onto the two-dimensional feature space using the T-SNE algorithm (van der Maaten and Hinton, 2008)

combination of informative and representative samples. Further, by employing a trade-off parameter, they can let the representative samples have high influence in the beginning, and gradually put more weight on informative samples.

Moreover, instead of using the output layer probabilities directly, Ash et al. (2019) computed a gradient of the predicted category with respect to the parameters of the last layer in the network. To measure the uncertainty of the model, they used this gradients magnitude. Further, to find diverse samples, they collected a batch of samples with the *k-means++* algorithm (Arthur and Vassilvitskii, 2006) to find gradients that span a diverse set of directions. Furthermore, another way of combining informative and representative sampling was proposed by Hsu and Lin (2015). Their method, inspired by the multi-armed bandit problem, would for each iteration explore the performance of different sampling methods and exploit the one with the best performance.

Another field of related work is plankton-specific active learning. Luo et al. (2005) proposed an AL method using multi-class support vector machines (SVM). They used least confidence sampling and margin sampling based on the SVMs decision function to decide which samples to query. Following the developments of CNNs, Bochinski et al. (2018) proposed a deep active learning approach conceptually similar to the aforementioned method by Wang et al. (2017). Another approach for minimizing human labeling effort in plankton taxa labeling was proposed by Pastore et al. (2020). Their method utilized fuzzy k-means clustering on extracted features, and a supervised model trained using the k-means clustering labels. Further, they also employed an SVM to do anomaly detection and detect unseen species of plankton.

The above-mentioned related work on hybrid AL is often reliant on the output layer probability distribution to work as an uncertainty metric. Additionally, a majority of the proposed hybrid approaches make no use of modern data augmentation, making it difficult to assess their validity on real applications. Motivated by this, we employ in this paper a data augmentation module as an extension to our original work in Haug et al. (2021) and assess the applicability of the framework to the plankton domain. Further, we compare the results of the novel framework with other well-known hybrid AL methods on both datasets the CIFAR and the plankton datasets.

4. PROPOSED FRAMEWORK

The framework introduced in this work builds on the active learning hybridization proposed in Haug et al. (2021). Figure 3 illustrates how the informative and a representative metric are combined. A data augmentation is added to this framework to increase the robustness of the model and enhance the performance of the informative metric. The reason behind extending the framework with this module is that captured planktonic species have complex structures compared to other datasets; moreover, we found that planktonic organisms from the same class but captured with different orientations are usually split by the models into separate groups as shown by the visualization tool in figure 2.

The data augmentation module, illustrated by module 8 in figure 3, consists of two steps. The first step is the flipping function which randomly generates images horizontally or vertically flipped with 50% probability. The flipping function allows the model to be more invariant to 90°image rotation; The second step is an affine transformation function that is applied with a rotation angle of 7° and with a horizontal and vertical translation of 0.1. This step is used to keep the images center-invariant, thus making the dataset dynamic rather than static which is particularly beneficial for tasks with small amounts of labeled data where overfitting is an issue. This set of augmentation techniques are summarized as $\mathcal{T}$ in Algorithm 1.

Figure 1 shows that the batch of samples queried with an informative metric has a high correlation in some areas; this suggests that there exists some redundancy among the queried samples. Based on this inefficiency in sample querying, a representative metric is integrated into the active learning framework. This hybridization enables the algorithm to choose the informative samples that also best represent the feature space of the unlabeled data. Moreover, with a trade-off function initially incorporating all samples, the learning network will gain an overview of the whole feature space. As the training proceeds and general decision boundaries are formed, more focus is put on samples on the inter-class decision boundaries. By switching focus to these samples, the learning model is able to fine-tune the decision boundaries to handle examples that are difficult to classify. As described in algorithm 1, the number of samples going from the informative metric to the representative metric is lowering with a rate δ each round, indicating that more of the informative samples are chosen at the end of the training. After the representative sampling, a number Q_k of samples are queried to a human expert for labeling. This active learning process continues until a labeling budget $\mathbf{B}$ is exhausted.

As illustrated in figure 3, a neural network is trained on an augmented labeled pool in each round. For the CIFAR dataset, the ResNet-18 architecture is employed as the learning network. However, for the plankton dataset, a custom network architecture consisting of 3 convolutional layers, 2 max-pooling layers, and 2 fully connected layers is employed to avoid overfitting and increase generalization.

Algorithm 1 CIRAL: Combined informative and representative active learning extended with the augmentation module

Require: Unlabeled samples D_0^U
Require: Initially labeled samples D_0^L
Require: Query budget $\mathbf{B}$
Require: Batch size β
Require: Set of hyper-parameters to train the network $\mathcal{H}$
Require: Set of data augmentation techniques $\mathcal{T}$
Require: Trade-off constant K_0
Require: Trade-off rate $\delta \in (0, 1)$

```
K_k = K_0
D_k^L = D_0^L
D_k^U = D_0^U
while D_k^L − D_0^L ≤ B do
    A_k = Train(D_k^L, H, T)
    for x_i ∈ D_k^U do
        r_i ← DeepFool(x_i, A_k)
    end for
    b_i ⟵ TradeOff(r_i, K_k)
    Q_k ⟵ MinMax(b_i, β)
    D_{k+1}^L ⟵ D_k^L ∪ Q_k
    D_{k+1}^U ⟵ D_k^U \ Q_k
    K_{k+1} ← K_k · δ
end while
```

By increasing the labeled pool with queried samples and updating the parameters of the neural network at each iteration, the inter-class decision boundaries are changing for each round. However, as the training proceeds and the model become more confident, the decision boundaries become more static, thus it is becoming increasingly important to put weight on the samples that are in the proximity of the boundary rather than samples far away from it. This is done by filtering out the samples with the largest distance result from the informative sampling, illustrated with module 5 in figure 3. To find this distance, the informative metric employed uses the Deep-Fool (Moosavi-Dezfooli et al., 2016) algorithm to compute adversarial attacks in order to find a proxy for the distance to the decision boundary. The Deep-Fool algorithm finds the closest hyperplane for each sample and then pushes the sample beyond it with a minimal possible perturbation. By adding the aforementioned data augmentation module to the framework, the network will improve its decision boundaries from training on more samples, and resultingly improve the accuracy of the boundary distance proxy provided by the informative metric.

Moreover, to find the representative samples in the next step, the min max facility location problem, well known from literature and described in (Hochbaum and Shmoys, 1985), is employed. It can be formally described as

$$\min_{s^1 : s^1 \leq b} \max_i \min_{j \in s^1 \cup s^0} \triangle(x_i, x_j) \tag{1}$$

Where $\Delta(x_i, x_j)$ represents the Euclidean distance between the data points x_i and x_j. Further, s^1 and s^0 is the pool of labeled and unlabeled data points, respectively. The optimization problem in (1) can be understood as choosing b cluster centers such that the largest distance from any single point to its nearest cluster center is minimized. As this problem is NP-hard, a sub-optimal solution is found by a greedy algorithmic approach as described in (Sener and Savarese, 2018). This method is proven to have a solution such that

$$\max_i \min_{j \in s^1 \cup s^0} \triangle(x_i, x_j) \leq 2\; X\; OPT \tag{2}$$

is satisfied, where OPT is the optimal solution to the optimization problem in 1 (Hochbaum and Shmoys, 1985). As described in our framework, the representative and informative metrics are combined through a trade-off function that only passes on the top K_k samples closest to the decision boundary. Thus, the algorithm will eventually ignore samples found at large distances away from the decision boundary. Formally, this trade-off method can be described as

$$Q_k = \textsc{MinMax}(K_k \cdot \textsc{DeepFool}(\mathbf{X})) \tag{3}$$

Where K_k is the trade-off constant and $\mathbf{X}$ is the input from the unlabeled samples.

5. EXPERIMENTAL RESULTS

The experiments were performed on the CIFAR dataset (Krizhevsky, 2009) and a subset of the plankton dataset of the Kaggle national data science bowl (kag, 2015), both containing 10 different classes. After each round of querying, a neural network got trained on the labeled pool until convergence of accuracy on a held-out validation set. A prediction was then performed on a separate testing set. We repeated this process until a pre-defined labeling budget was exhausted. All our results report an average of 3 complete trials. In figure 6, results from the proposed hybrid method and other state-of-the-art AL methods tested on the plankton dataset are presented. The figure presents both the results with and without the data augmentation module described in section 4. In figure 6 (a), the accuracy of our method is compared to other hybrid methods. Further, in figure 6 (b), our method is compared to informative and representative methods. Random benchmark sampling is included in both (a) and (b) for reference. One can observe from these results that our proposed method(CIRAL) is performing steadily in terms of classification accuracy and is outperforming the random sampling benchmark by a large margin. Random sampling needs approximately twice as many samples to reach the same level of accuracy as our proposed method. This result is valid for the other methods as well, suggesting that active learning is effective on the plankton dataset. That is, the information provided in the images is not uniformly distributed in the dataset, making it possible to strategically select images for labeling. Furthermore, the results can be studied in more detail in figure 4, where the classification accuracy of our method is presented relative to the other methods. In each plot, our method is compared with another AL method. Similar results can also be found in figure 5, where our method has been applied to the CIFAR dataset. In the latter plot, one can observe that the hybridization benefits from combining informative and representative methods in that it outperforms each of them

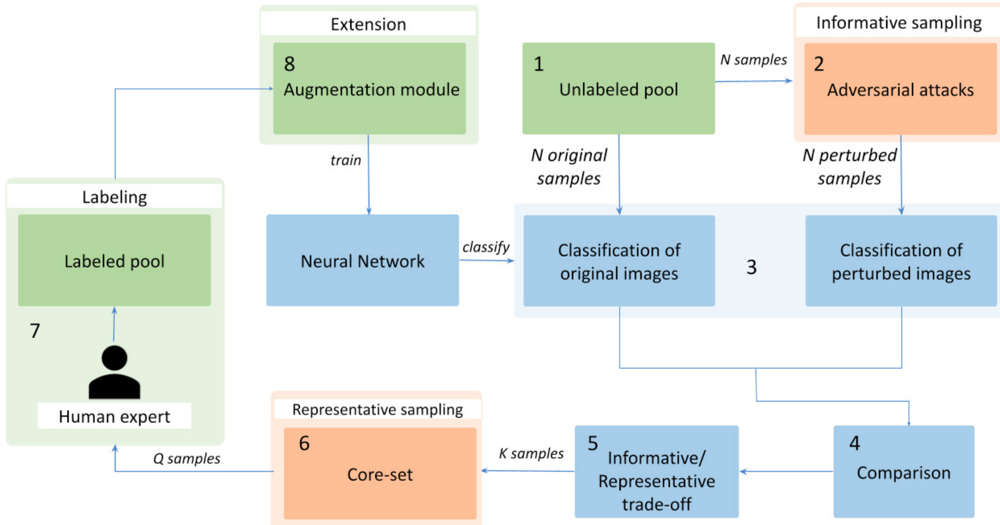


Fig. 3. Block diagram of the Active Learning framework. The process is initiated with an unlabeled pool of N images. An adversarial attack is performed on the unlabeled instances and they are sorted by how much perturbation is needed in order for the neural network to change its classification. This adversarial attack works as a proxy on how far each sample is from the decision boundary and is an uncertainty metric for the model. Based on the trade-off function, a set of K uncertainty samples are sent to the representative sampling method. Lastly, Q samples with combined informative and representative values are queried to a human expert for manual labeling.

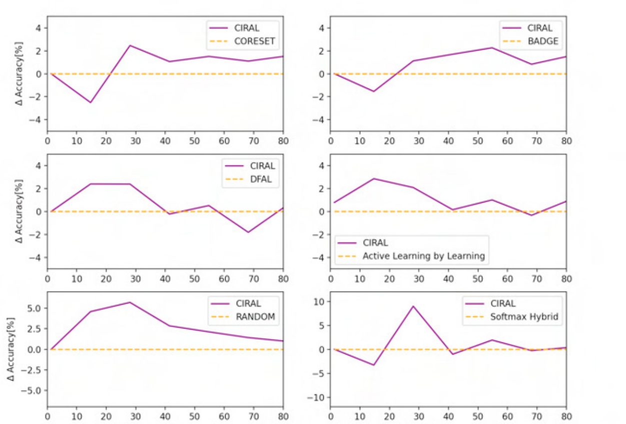


Fig. 4. (LHS) The proposed AL method compared to informative, representative and random methods. (RHS) The proposed AL method compared to other hybrid methods (BADGE, Active Learning by Learning, Softmax Hybrid). All experiments in this figure are performed on the plankton dataset.

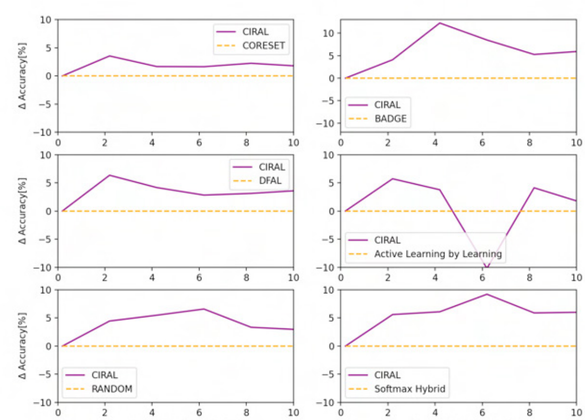
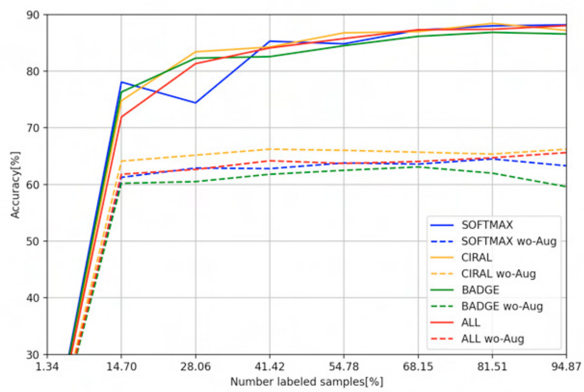
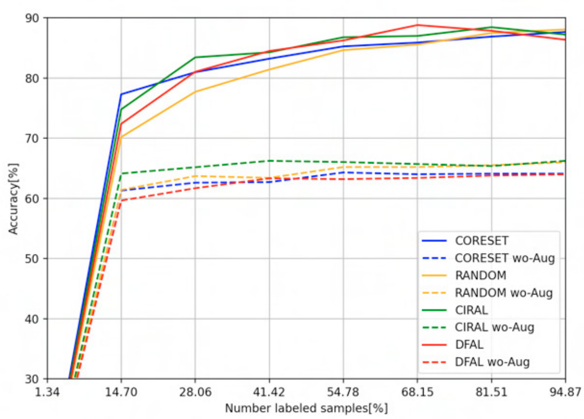


Fig. 5. (LHS) The proposed AL method compared to informative, representative and random methods. (RHS) The poposed AL method compared to other hybrid methods (BADGE, Active Learning by Learning, Softmax Hybrid). All experiments in this figure are performed on the CIFAR dataset.

individually. This performance enhancement compared to the other strategies is a result of incorporating the full feature space while also taking samples close to the inter-class decision boundaries into account. The samples obtained in the latter case help fine-tune the model to gain additional performance. This is particularly evident in figure 4 where one can observe how our proposed method outperforms the coreset representative method when 20% of the samples have been labeled. Both our proposed method and the coreset method perform well in the beginning from incorporating the full dataset. The hybrid method does, however, eventually put more weight on the informative samples and thereby gain an advantage over the pure representative method as illustrated in the top left plot in figure 4. The opposite is true in the mid-left plot in figure 4 where the hybridization performs better than the informative method in the beginning but on par in the later rounds. Moreover, the proposed CIRAL method is also showing promising results compared to the BADGE (Ash et al., 2019), Active Learning by Learning (Hsu and Lin, 2015) and Softmax Hybrid (Kaushal et al., 2018) methods. Comparing with the performance of the Softmax Hybrid, it can be observed that the proposed hybridization is significantly better on the CIFAR dataset, as illustrated in figure 5. This observation may suggest that the proposed method is better at identifying informative samples when



(a) Our method(CIRAL) vs hybrid methods with and without data augmentation on the plankton dataset. Results without augmentation are denoted as 'wo-Aug'



(b) Our method(CIRAL) vs informative/representative with and without data augmentation. Results without augmentation are denoted as 'wo-Aug'

Fig. 6. Comparison of the experimental results with and without data augmentation during training. (a) Performance comparison between our method and other hybrid AL methods with and with out data augmentation. (b) Performance comparison between our hybrid method and informative and representative methods with and without data augmentation.

the classes are more intertwined, such as in the CIFAR dataset. The different classes in the plankton dataset, as illustrated in figure 2, are less intertwined compared to the classes from the CIFAR dataset illustrated in figure 1. This information about the latent space is relevant for the efficiency of the hybrid framework. Comparing the results of figure 4 and figure 5, one can observe that the hybrid framework, in general, performs better when the classes are more intertwined. This suggests that the hybrid framework is able to overcome redundant sampling, and select images that brings much information to the classifier.

Furthermore, from figure 6 one can observe how the data augmentation module significantly increases the classification accuracy of the active learning methods. The hybridization is performing best when no augmentation is applied, however, with the augmentation module, the difference in the performance of the methods becomes less significant suggesting that all the methods are fully enabled to utilize the information provided in the dataset. However, an interesting observation can be made in figure 6 (b) when comparing the DFAL method with and without the augmentation module. It can be observed that, without the augmentation module, it performs worst of the compared methods, however, with the applied augmentation module, it surpasses most of the other AL methods. This observation suggests that the data augmentation improves the decision boundaries in the learning model from which the DFAL and CIRAL methods benefit. Hence, the application of a data augmentation module is justified in terms of a general performance increase and the increased performance of the decision boundary-dependent informative metric.

6. CONCLUSION AND FUTURE WORK

This paper presents a novel framework furthering the field of in-situ underwater planktonic image analysis (Saad et al., 2020, 2021). Manual labeling of planktonic data is time-consuming and puts a large burden on the domain experts. The proposed active learning method can minimize this effort while achieving satisfactory classification results and outperform random sampling. The framework presented in this paper combines metrics for representative and informative sampling and achieves better performance than each of them separately. The method has proven to be efficient on both the benchmark CIFAR dataset and the more complex plankton dataset, suggesting that these metrics should be considered in combination when applying active learning. Furthermore, empirical results show that our proposed framework outperforms other state-of-the-art hybrid AL methods.

The informative metric employed in the proposed framework is dependent on good decision boundaries to get full utilization. The augmentation algorithm which is added as an extension to the originally proposed CIRAL framework (Haug et al., 2021), further allowed the model to create better decision boundaries on complex data structures that exist in the plankton and CIFAR datasets and increased its general classification performance. It was seen from the performance enhancement of the DFAL method that the applied data augmentation module increased the performance of the information metric in the hybridization.

An interesting future direction would be to investigate how other representative functions affect the performance of the classifier. In particular, looking at combining Bayesian-based representative metrics with informative metrics is an interesting direction. Another interesting future direction is to construct, from this novel hybrid AL framework, classifier models that require a minimum amount of labeled datasets for training and embedding those created models into AUV platforms for in-situ plankton classification.

ACKNOWLEDGEMENTS

This research is funded by the Research Council of Norway (RCN) IKTPLUSS program (project number 262741) and supported by NTNU AMOS (RCN project number 223254).

REFERENCES

(2015). Plankton imagery data collected from f.g. walton smith in straits of florida from 2014-06-03 to 2014-06-06 and used in the 2015 national data science bowl (nodc accession 0127422). Access: 2020-16-12.

Arthur, D. and Vassilvitskii, S. (2006). k-means++: The advantages of careful seeding. Technical report, Stanford.

Ash, J.T., Zhang, C., Krishnamurthy, A., Langford, J., and Agarwal, A. (2019). Deep batch active learning by diverse, uncertain gradient lower bounds. *CoRR*, abs/1906.03671. URL `http://arxiv.org/abs/1906.03671`.

Beluch, W.H., Genewein, T., Nurnberger, A., and Kohler, J.M. (2018). The power of ensembles for active learning in image classification. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9368–9377. doi:10.1109/CVPR.2018.00976.

Bochinski, E., Bacha, G., Eiselein, V., Walles, T.J., Nejstgaard, J.C., and Sikora, T. (2018). Deep active learning for in situ plankton classification. In *International Conference on Pattern Recognition*, 5–15. Springer.

Ducoffe, M. and Precioso, F. (2018). Adversarial active learning for deep networks: a margin based approach. *CoRR*, abs/1802.09841. URL `http://arxiv.org/abs/1802.09841`.

Gal, Y., Islam, R., and Ghahramani, Z. (2017). Deep bayesian active learning with image data. *CoRR*, abs/1703.02910. URL `http://arxiv.org/abs/1703.02910`.

Geifman, Y. and El-Yaniv, R. (2017). Deep active learning over the long tail. *CoRR*, abs/1711.00941. URL `http://arxiv.org/abs/1711.00941`.

Haug, M.L., Saad, A., and Stahl, A. (2021). A combined informative and representative active learning approach for plankton taxa labeling. In *Thirteenth International Conference on Digital Image Processing (ICDIP 2021)*, volume 11878, 118781Q. International Society for Optics and Photonics.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, abs/1512.03385. URL `http://arxiv.org/abs/1512.03385`.

Hochbaum, D.S. and Shmoys, D.B. (1985). A best possible heuristic for the k-center problem. *Mathematics of operations research*, 10(2), 180–184.

Hsu, W.N. and Lin, H.T. (2015). Active learning by learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29.

Huang, S., Zhao, J., and Liu, Z. (2018). Cost-effective training of deep cnns with active model adaptation. *CoRR*, abs/1802.05394. URL `http://arxiv.org/abs/1802.05394`.

Kaushal, V., Sahoo, A., Doctor, K., Uppalapati, N.R., Shetty, S., Singh, P., Iyer, R.K., and Ramakrishnan, G. (2018). Learning from less data: Diversified subset selection and active learning in image classification tasks. *CoRR*, abs/1805.11191. URL `http://arxiv.org/abs/1805.11191`.

Krishnamurthy, V. (2002). Algorithms for optimal scheduling and management of hidden markov model sensors. *IEEE Transactions on Signal Processing*, 50(6), 1382–1397. doi:10.1109/TSP.2002.1003062.

Krizhevsky, A. (2009). Learning multiple layers of features from tiny images.

Lewis, D.D. and Gale, W.A. (1994). A sequential algorithm for training text classifiers. *CoRR*, abs/cmp-lg/9407020. URL `http://arxiv.org/abs/cmp-lg/9407020`.

Luo, T., Kramer, K., Goldgof, D.B., Hall, L.O., Samson, S., Remsen, A., Hopkins, T., and Cohn, D. (2005). Active learning to recognize multiple types of plankton. *Journal of Machine Learning Research*, 6(4).

Mittal, S., Tatarchenko, M., Çiçek, Ö., and Brox, T. (2019). Parting with illusions about deep active learning. *arXiv preprint arXiv:1912.05361*.

Moosavi-Dezfooli, S.M., Fawzi, A., and Frossard, P. (2016). Deepfool: a simple and accurate method to fool deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2574–2582.

Pastore, V.P., Zimmerman, T.G., Biswas, S.K., and Bianco, S. (2020). Annotation-free learning of plankton for classification and anomaly detection. *Scientific reports*, 10(1), 1–15.

Pinsler, R., Gordon, J., Nalisnick, E., and Hernández-Lobato, J.M. (2021). Bayesian batch active learning as sparse subset approximation.

Ren, P., Xiao, Y., Chang, X., Huang, P.Y., Li, Z., Chen, X., and Wang, X. (2020). A survey of deep active learning.

Saad, A., Bergrum, S., and Stahl, A. (2021). An instance segmentation framework for in-situ plankton taxa assessment. In *Thirteenth International Conference on Machine Vision*, volume 11605, 1160511. International Society for Optics and Photonics.

Saad, A., Stahl, A., Våge, A., Davies, E., Nordam, T., Aberle, N., Ludvigsen, M., Johnsen, G., Sousa, J., and Rajan, K. (2020). Advancing ocean observation with an ai-driven mobile robotic explorer. *Oceanography*, 33(3), 50–59.

Sener, O. and Savarese, S. (2018). Active learning for convolutional neural networks: A core-set approach.

Settles, B. (2009). Active learning literature survey.

Tong, S. and Koller, D. (2001). Support vector machine active learning with applications to text classification. *Journal of machine learning research*, 2(Nov), 45–66.

van der Maaten, L. and Hinton, G. (2008). Viualizing data using t-sne. *Journal of Machine Learning Research*, 9, 2579–2605.

Vodrahalli, K., Li, K., and Malik, J. (2018). Are all training examples created equal? an empirical study. *CoRR*, abs/1811.12569. URL `http://arxiv.org/abs/1811.12569`.

Wang, K., Zhang, D., Li, Y., Zhang, R., and Lin, L. (2017). Cost-effective active learning for deep image classification. *CoRR*, abs/1701.03551. URL `http://arxiv.org/abs/1701.03551`.

Yoo, D. and Kweon, I.S. (2019). Learning loss for active learning. *CoRR*, abs/1905.03677. URL `http://arxiv.org/abs/1905.03677`.

Zhdanov, F. (2019). Diverse mini-batch active learning. *CoRR*, abs/1901.05954. URL `http://arxiv.org/abs/1901.05954`.