A combined informative and representative active learning approach for plankton taxa labeling

Haug, Martin Lund, Saad, Aya, Stahl, Annette

Martin Lund Haug, Aya Saad, Annette Stahl, "A combined informative and representative active learning approach for plankton taxa labeling," Proc. SPIE 11878, Thirteenth International Conference on Digital Image Processing (ICDIP 2021), 118781Q (30 June 2021); doi: 10.1117/12.2601096

SPIE.

Event: Thirteenth International Conference on Digital Image Processing, 2021, Singapore, Singapore

A combined informative and representative active learning approach for plankton taxa labeling

Martin Lund Haug[1], Aya Saad[1], Annette Stahl[1]

[1]Dept. of Engineering Cybernetics, Norwegian University of Science and Technology, NTNU, Trondheim, Norway

ABSTRACT

With an ever-increasing amount of image data, the manual labeling process has become the bottleneck in many machine learning applications. Plankton taxa labeling is especially a challenge due to its complex nature, and the manual labeling effort places a large burden on the domain experts. The Active Learning (AL) paradigm is a promising research direction adopted in the literature to minimize the manual labeling effort exerted by domain experts. Many approaches for AL have been proposed over the recent years to improve the labeling task by supporting the construction of large datasets suitable to train machine learning models while minimizing human involvement in the process. Our empirical study suggests that many modern active learning methods fail to incorporate both the samples that represent the statistical pattern of the data and the samples in which the machine learning model is not confident about.

Inspired by these limitations, we propose an algorithm that combines these two types of sampling in order to capture the data distribution of the whole feature space, prevent redundant sampling from correlated uncertainty queries and fine-tune the inter-class decision boundary. Our experiments show that the proposed method outperforms each of the methods separately. Further, it also proves to be efficient on both the CIFAR dataset and the more complex Kaggle plankton dataset.

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

1. INTRODUCTION

Convolutional Neural Networks (CNN) models have proved competent at solving computer vision problems in the paradigm of the Supervised Machine Learning (ML) approaches. However, to make such models reliable, an immense amount of pre-classified input datasets is required in the training process. Constructing such large datasets requires an extensive manual effort for labeling, which requires a massive amount of time. Nevertheless, the resulting manual classification is imperfect and prone to errors.

Active learning (AL) is a promising research direction of machine learning that aims at mitigating the burden of human experts on labeling training instances. They do so by exploiting the fact that not all samples bring equally much information to an image classifier.[1] Therefore, by finding only the most informative samples and query them for manual labeling, the classifier can be trained to achieve equal performance as if it was trained on the whole dataset.[2] 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,[3] while it is pool-based[4] otherwise. Further, the AL models' mode of sampling varies between batch-mode[5] or single-mode[4] depending on the number of data samples presented and chosen at each labeling round. With the recent developments of convolutional neural networks (CNN), batch-mode sampling has become increasingly relevant as it is not computationally feasible to update a large network with single data points.

For the sampling modes aforementioned, there are differences among them in how samples are queried for labeling. Mainly, the most important difference is in choosing between informative and representative samples. While the former aims to find samples which the image classifier finds most informative, the latter exploits the feature space of the data points to best capture the data statistical patterns. There exists a broad literature on active learning. The reader can refer to a survey[6] on active learning based on traditional methods of ML, and more recently, the survey[7] on deep learning versions of AL techniques.

Previous work has shown that active learning has proven to be an effective way of choosing informative samples from a large number of unlabeled samples.[8–10] Although hybrid methods that combine informativeness and representativeness are increasingly popular among researchers, much of the existing methods only incorporate either informativeness or representativeness. Existing AL models lack efficient utilization of the feature space of the dataset under consideration.

Thirteenth International Conference on Digital Image Processing (ICDIP 2021), edited by Xudong Jiang, Hiroshi Fujita
Proc. of SPIE Vol. 11878, 118781Q · © 2021 SPIE · CCC code: 0277-786X/21/$21 · doi: 10.1117/12.2601096

They sometimes select samples for the training that fail to fit the different classes' representation specifically when the boundaries between the classes have some overlap. The resulting proposed samples suffer over-fitting or under-fitting the dataset. Therefore they affect the performance of the classifier. Furthermore, in [1], the authors investigated how different datasets had various amounts of information incorporated in the images. The study found that for some datasets,[11] a few representative samples were enough to capture the data statistical distribution. However, for other datasets,[12] this proved not to be true. Thus, the success in employing a stand-alone uncertainty or representative sampling mode is dependent on the dataset.

Therefore in this paper, we propose an efficient algorithm that combines an informative metric with a representative metric approach for active learning. The proposed algorithm begins with a focus on the diverse feature space. It gradually focus more on samples located at the proximity of the classes' decision boundaries to further fine-tune the machine learning model. The aim behind this hybridization has three folds: 1) the novel model will have a good initialization from incorporating the full feature space in the early rounds of query and training. Inspired by the work in [13], a trade-off function gradually moves the focus from diverse samples to more uncertain samples during the training process in order to fine-tune the model with samples located at the boundaries of the classes representations. 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 have shown to be a bad proxy for the uncertainty of neural networks,[7,14] an adversarial active learning method is employed. This method has previously shown good results,[15] however it was not employed with sub-modular heuristics as is done in this paper.

Experiments are conducted on the plankton dataset from National Data Science Bowl[16] and the CIFAR dataset.[12] The ResNet-18[17] architecture is employed as the learning network model. It is worth noting that no data augmentation was performed on neither the CIFAR nor the plankton dataset, as is often done to enhance the performance on classification. The results demonstrate that the proposed algorithmic framework is more efficient compared to each of the strategies separately. Furthermore, they have shown that the novel proposed hybrid algorithm is more effective when dealing with difficult datasets such as the plankton.

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 active learning, and hybrid sampling methods in particular. Section 4 explains our proposed algorithmic framework. Section 5 presents the experimental results. Finally, in section 6, a conclusion is made on the the contributions of this paper and also future directions are presented.

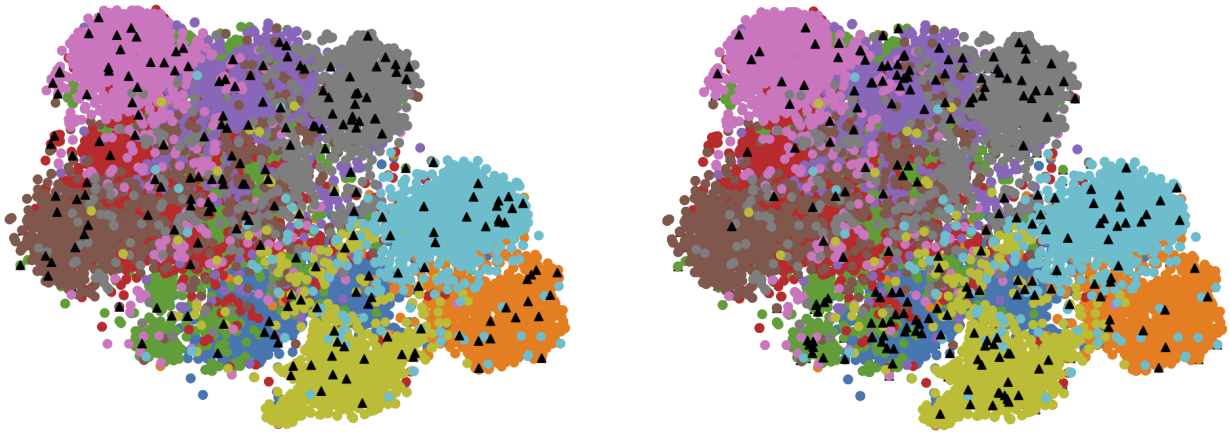
2. BACKGROUND

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 primarily be 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 finding 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 the 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 [18], who in addition pseudo-labeled high confidence samples for additional robustness. However, research[7,14] has shown that these softmax probabilities work as a bad proxy for the confidence of neural networks, 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. [8] 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 have also been studied in [9] where the authors employed an ensemble of different CNNs instead of MC dropout. A drawback with these ensemble methods is the computational effort that is increasing with the dimensions of the learning network and number of unlabeled samples.

A different approach from using the classification results of the learning networks has been proposed by [19] 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 fine-tuning 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, [15] 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 the sample is from the decision boundary. This method of looking at the input to the network is somewhat the other way around of looking at the softmax layer as done in [18]. However, both of the latter methods queries the top most uncertain images. As can be seen in figure 1 and also stated in [20], uncertainty sampling tend 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



(a) Core-set representative metric. (b) DeepFool informative metric.

Figure 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,[21] the images are projected to the two-dimensional feature space.

figure 1 (a) one can observe that by employing a representative metric exploiting 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 tries to exploit the feature space and others that aims to maximize some performance metric. An example of the latter is, as proposed in [10], a method that approximates the complete data posterior of model parameters that produces diverse batches. By selecting sub samples, the method tries to lower the expected value of the loss function. An example of the former is, as proposed in [22], a diversity method that performs a farthest-first traversal to cover the feature space. A similar example is shown in [20] that proposes a core set method to find clusters based on the min max facility location problem and then optimize these clusters with mixed integer programming. The overall aim for most of the representativeness methods is to replicate the distribution of the complete unlabeled set. By regarding active learning as a binary classification task between the labeled and unlabeled sets, [23] aims to make the labeled dataset indistinguishable from the unlabeled dataset by capturing the statistical distribution of the unlabeled data.

3. RELATED WORK

This paper proposes a hybrid active learning framework combining the informative and representative sampling strategies described in section 2. This hybridization of active learning has become increasingly popular among researchers in later years. In [24] a method of combining predictive entropy based uncertainty sampling and a distance function on a learned feature space to optimize the selection of unlabeled samples was proposed. Their method was based on the assumption that the most informative samples are the ones where the model has the highest uncertainty and greatest distance to the existing training examples. In [13], the authors aim to fine-tune pre-trained networks with a combination of informative and representative samples. They are also employing a trade-off parameter to let the representative samples have high influence in the beginning, and gradually use more uncertain samples. Further, [25] proposed a hybrid method to deal with the imbalanced classification problem. Their uncertainty metric was based on the probability output from the neural network, while their diversity metric was based on distance from k-means clusters on already labeled data points. Similarly, [26] proposed a work of diversified subset selection that use classical methods of uncertainty, margin sampling and entropy

from the softmax probability distribution. To find diverse samples they used, similar to this paper, min-max facility location in addition to disparity minimum. An important finding in the paper suggested that similar data points within a class made disparity-min pick outliers and thus confuse the model. Moreover, instead of using the output layer probabilities directly, [5] proposed a hybrid method that uses the size of the backward gradient to incorporate the uncertainty metric. For the diversity the authors employed the *k-means++*[27] algorithm. Furthermore, [28] proposed a method to increase diversity in mini-batch active learning. Their experiments showed that diversity-enhancing approaches outperformed a baseline of uncertainty sampling methods. They combined informativeness with representativeness by using margin sampling from the softmax layer as uncertainty metric and the k-means algorithm as representative metric.

The aforementioned related work are often reliant on the output layer of the network employed as an uncertainty metric. Motivated by this, we build on the uncertainty metric proposed in [15] and employ it with a min max sub-modular heuristic to form a hybrid active learning method. Furthermore, similar to [13], we combine the metrics with a trade-off parameter. This is aligned with the findings of [5], who observed that it is advantageous to do representative sampling early in the training then in later rounds focus more on informative samples to fine-tune the model.

4. PROPOSED FRAMEWORK

The framework introduced in this paper combines the informative metric of an adversarial attack with the representative metric of the facility min max problem. Figure 2 illustrates how these methods are combined in the proposed framework. From the plot of the informative metric shown in figure 1 (b), one can observe that the queried samples have high correlation in some areas; this suggests that there exists some redundancy among the queried samples. By incorporating a representative metric to the active learning framework, one can choose the informative samples that also best represent the feature space. 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

Algorithm 1 CIRAL: Combined informative and representative active learning

Require: Unlabeled samples D_0^U
Require: Initially labeled samples D_0^L
Require: Query budget **B**
Require: Batch size β
Require: Set of hyper-parameters to train the network $\mathcal{H}$
Require: Trade-off constant $\mathbf{K}_0$
Require: Trade-off rate $\delta \in (0, 1)$
  $\mathbf{K}_k = \mathbf{K}_0$
  $D_k^L = D_0^L$
  $D_k^U = D_0^U$
  while $D_k^L - D_0^L \leq$ **B do**
    $\mathcal{A}_k = \textsc{Train}(\mathcal{H}, D_k^L)$
    for $x_i \in D_k^U$ **do**
      $r_i \leftarrow \textsc{DeepFool}(x_i, \mathcal{A}_k)$
    end for
    $b_i \longleftarrow \textsc{TradeOff}(r_i, \mathbf{K}_k)$
    $Q_k \longleftarrow \textsc{MinMax}(b_i, \beta)$
    $D_{k+1}^L \longleftarrow D_k^L \cup Q_k$
    $D_{k+1}^U \longleftarrow D_k^U \backslash Q_k$
    $\mathbf{K}_{k+1} \leftarrow \mathbf{K}_k \cdot \delta$
  end while

of samples going from the informative metric to the representative metric is lowering with a rate δ each round, indicating that the algorithm prioritize samples with high level of informativeness at the later AL cycles. The active learning cycle described is continued until a labeling budget **B** is exhausted.

As illustrated in figure 2, a neural network is trained on the labeled pool in each iteration, forming the decision boundaries used by the informative sampling method. However, as the training proceeds and the model becomes more confident, the decision boundaries become more static, thus it becomes increasingly important to put weight on the samples that are at 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 2. To find this distance, the informative metric employed uses the DEEPFOOL[29] algorithm to compute adversarial attacks in order to find a proxy for the distance to the decision boundary. The DEEPFOOL algorithm finds the closest hyperplane for each sample and then pushes the sample beyond it with a minimal possible perturbation.

Moreover, to find the representative samples in the next step, the min max facility location problem, well known from literature and described in [30], 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 new queried samples and existing pool of samples, 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 [20]. 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.[30] As described in our framework, the representative and informative metrics are combined through a trade-off function that gradually puts more focus on the informative samples at the cost of the representative samples. As the DEEPFOOL algorithm returns a list of samples based on their distance to the decision boundary, the trade-off function is only passing on a fraction $\mathbf{K}_k$ of the most informative samples to the representative function. 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

$$Q_k = \text{MINMAX}(\mathbf{K}_k \cdot \text{DEEPFOOL}(\mathbf{X}), \beta) \tag{3}$$

Where $\mathbf{K}_k$ is the trade-off constant, $\mathbf{X}$ is the input from the unlabeled samples and β is the number of samples to be queried.

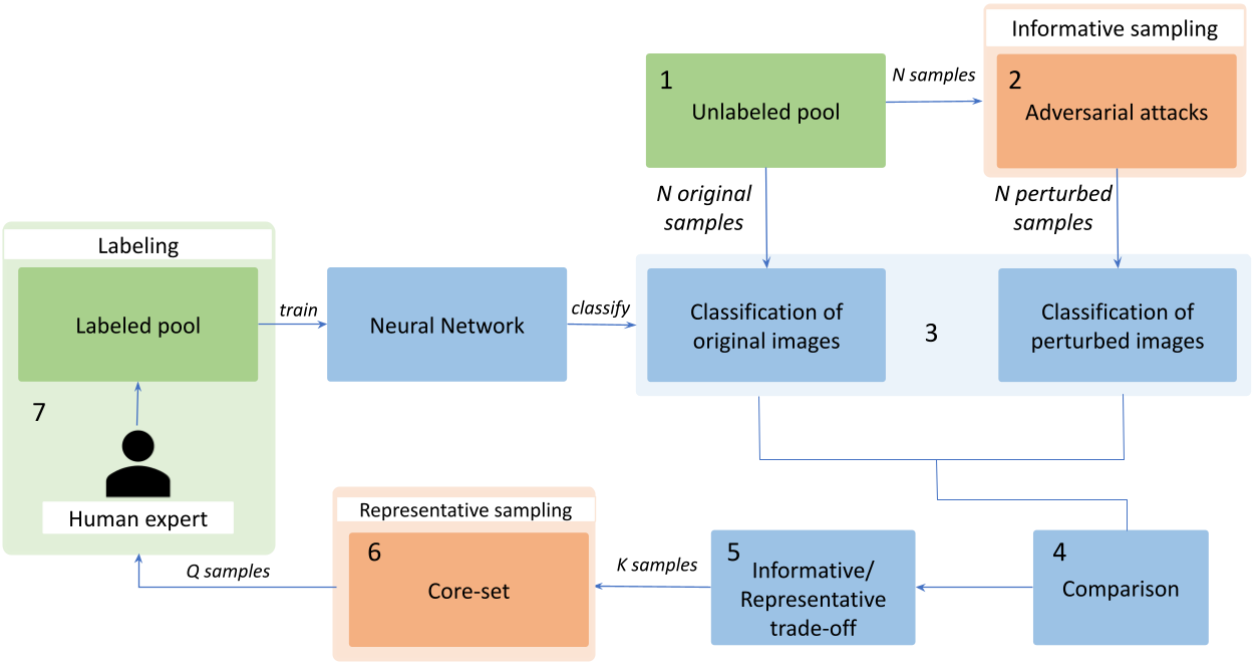


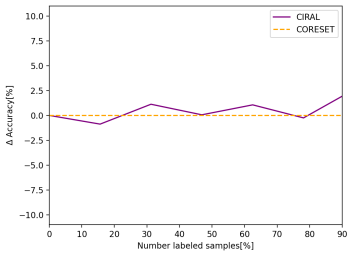
Figure 2: Block diagram of the active learning framework showing the relevant modules and the flow of samples from the unlabeled pool to the labeled pool.

5. EXPERIMENTAL RESULTS

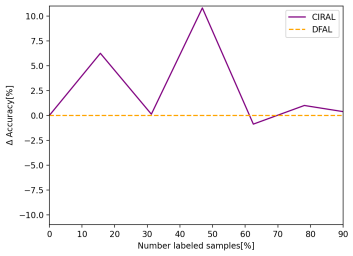
The experiments were performed on subsets of the CIFAR[12] and Kaggle National Data Science Bowl[16] (plankton) datasets, both containing 10 different classes. After each round of querying, the neural network got trained on the labeled pool until convergence of validation accuracy. A prediction was then performed on a separate testing set.

In figure 3 (a-c), results from the proposed hybrid method tested on the plankton dataset are presented. Further, in figure 3 (d-f), the results on the CIFAR dataset are presented. For both datasets, the result for the hybrid method is plotted relatively to the accuracy of the DFAL,[15] core-set[20] and random strategy. Results are also presented in figure 3 (g-h) for the CIFAR and plankton datasets, respectively. One can observe that the hybridization is performing steadily better than the other methods separately. As can be seen in figure 3 (a-c), the proposed method gains advantage on combining representative and informative metrics on the more complex plankton dataset. Looking at figure 3 (a), the method is performing better than the core-set method because it is better at choosing samples that fine-tune the decision boundaries. Further, looking at figure 3 (b) the hybrid method is clearly better in the early rounds of training suggesting that the incorporated representative samples in the early rounds are beneficial for the model.

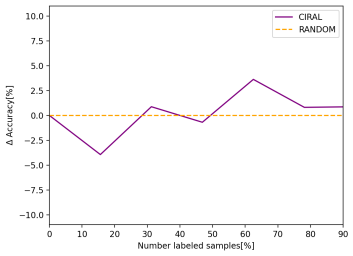
Looking at the graphs in figure 3 (g-h), the proposed method is consistently having high accuracy compared to the other methods. Observing from the results in table 1 and 2 that the proposed method is achieving good results overall, it is especially prominent on the CIFAR dataset.



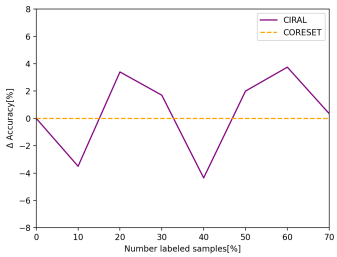
(a) Comparison of CIRAL and a representative method.



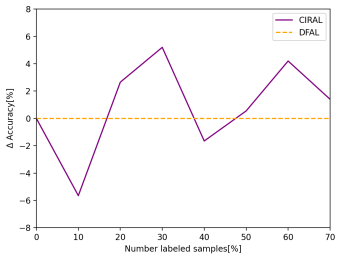
(b) Comparison of CIRAL and an informative method.



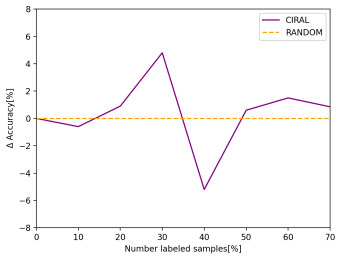
(c) Comparison of CIRAL and the random benchmark.



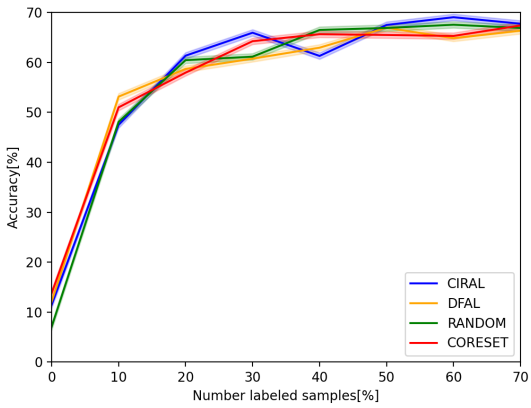
(d) Comparison of CIRAL and a representative method.



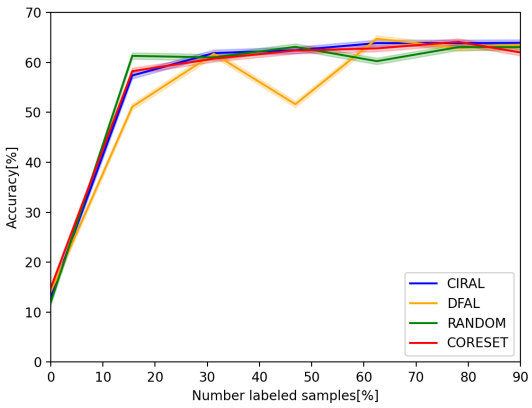
(e) Comparison of CIRAL and an informative method.



(f) Comparison of CIRAL and the random benchmark.



(g) Comparison of CIRAL and other hybrid methods.



(h) Comparison of CIRAL and other informative/representative methods.

Figure 3: Results from experiments with our proposed framework (CIRAL). Results in plot (a-c, g) are from experiments on the plankton dataset. Results in plot (d-f, h) are from experiments on the CIFAR dataset.

Method/Round	1	2	3	4	5	6	7
CIRAL	0.1	0.574	**0.62**	0.624	0.64	**0.64**	**0.64**
DFAL	0.1	0.511	0.62	0.52	**0.647**	0.63	0.638
CORE-SET	0.1	0.573	0.61	0.62	0.63	0.64	0.61
RANDOM	0.1	**0.61**	0.61	**0.63**	0.60	0.63	0.62

Table 1: Classification accuracy from the experiments with our proposed framework on the plankton dataset.

Method/Round	1	2	3	4	5	6	7	8
CIRAL	0.1	0.48	**0.614**	**0.66**	0.613	**0.68**	**0.69**	**0.68**
DFAL	0.1	**0.532**	0.59	0.608	0.63	0.67	0.65	0.66
CORE-SET	0.1	0.51	0.58	0.64	0.65	0.655	0.653	0.67
RANDOM	0.1	0.48	0.605	0.612	**0.66**	0.67	0.676	0.67

Table 2: Classification accuracy from the experiments with our proposed framework on the CIFAR dataset.

6. CONCLUSION

This paper presents a new framework furthering the field planktonic image analysis. Manual labeling of planktonic data is time consuming and puts a large burden on the domain experts. The proposed active learning method is able to minimize this effort while still achieving satisfactory classification results. The framework presented in this paper combines metrics for representative and informative sampling and achieve 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. The informative metric employed in the proposed framework is dependent on good decision boundaries to get full utilization. An interesting future direction would therefore be to investigate how other representative functions affect the performance of the classifier. In particular, looking at combining Bayesian-based representative metrics with the informative metrics employed in this framework is an interesting direction.

ACKNOWLEDGMENTS

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

[1] K. Vodrahalli, K. Li, and J. Malik, "Are all training examples created equal? an empirical study," CoRR **abs/1811.12569** (2018).

[2] K. Chitta, J. M. Alvarez, E. Haussmann, and C. Farabet, "Less is more: An exploration of data redundancy with active dataset subsampling," CoRR **abs/1905.12737** (2019).

[3] V. Krishnamurthy, "Algorithms for optimal scheduling and management of hidden markov model sensors," IEEE Transactions on Signal Processing **50**(6), 1382–1397 (2002).

[4] D. D. Lewis and W. A. Gale, "A sequential algorithm for training text classifiers," CoRR **abs/cmp-lg/9407020** (1994).

[5] J. T. Ash, C. Zhang, A. Krishnamurthy, J. Langford, and A. Agarwal, "Deep batch active learning by diverse, uncertain gradient lower bounds," CoRR **abs/1906.03671** (2019).

[6] B. Settles, "Active learning literature survey," (2009).

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

[8] Y. Gal, R. Islam, and Z. Ghahramani, "Deep bayesian active learning with image data," CoRR **abs/1703.02910** (2017).

[9] W. H. Beluch, T. Genewein, A. Nurnberger, and J. M. Kohler, "The power of ensembles for active learning in image classification," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9368–9377 (2018).
[10] R. Pinsler, J. Gordon, E. Nalisnick, and J. M. Hernández-Lobato, "Bayesian batch active learning as sparse subset approximation," (2021).
[11] Y. LeCun and C. Cortes, "MNIST handwritten digit database," (2010).
[12] A. Krizhevsky, "Learning multiple layers of features from tiny images," (2009).
[13] S. Huang, J. Zhao, and Z. Liu, "Cost-effective training of deep cnns with active model adaptation," CoRR **abs/1802.05394** (2018).
[14] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," CoRR **abs/1706.04599** (2017).
[15] M. Ducoffe and F. Precioso, "Adversarial active learning for deep networks: a margin based approach," CoRR **abs/1802.09841** (2018).
[16] "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)," (2015). Access: 2020-16-12.
[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," CoRR **abs/1512.03385** (2015).
[18] K. Wang, D. Zhang, Y. Li, R. Zhang, and L. Lin, "Cost-effective active learning for deep image classification," CoRR **abs/1701.03551** (2017).
[19] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," Journal of machine learning research **2**(Nov), 45–66 (2001).
[20] O. Sener and S. Savarese, "Active learning for convolutional neural networks: A core-set approach," (2018).
[21] L. van der Maaten and G. Hinton, "Viualizing data using t-sne," Journal of Machine Learning Research **9**, 2579–2605 (11 2008).
[22] Y. Geifman and R. El-Yaniv, "Deep active learning over the long tail," CoRR **abs/1711.00941** (2017).
[23] D. Gissin and S. Shalev-Shwartz, "Discriminative active learning," CoRR **abs/1907.06347** (2019).
[24] A. Smailagic, P. Costa, H. Young Noh, D. Walawalkar, K. Khandelwal, A. Galdran, M. Mirshekari, J. Fagert, S. Xu, P. Zhang, and A. Campilho, "Medal: Accurate and robust deep active learning for medical image analysis," in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 481–488 (2018).
[25] H. Dong, B. Zhu, and J. Zhang, "A cost-sensitive active learning for imbalance data with uncertainty and diversity combination," 218–224 (02 2020).
[26] V. Kaushal, A. Sahoo, K. Doctor, N. R. Uppalapati, S. Shetty, P. Singh, R. K. Iyer, and G. Ramakrishnan, "Learning from less data: Diversified subset selection and active learning in image classification tasks," CoRR **abs/1805.11191** (2018).
[27] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," tech. rep., Stanford (2006).
[28] F. Zhdanov, "Diverse mini-batch active learning," CoRR **abs/1901.05954** (2019).
[29] S.-M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "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 (2016).
[30] D. S. Hochbaum and D. B. Shmoys, "A best possible heuristic for the k-center problem," Mathematics of operations research **10**(2), 180–184 (1985).