Active Learning Strategy Using Contrastive Learning and K-means for Aquatic Invasive Species Recognition

Shaif Chowdhury * ,Greg Hamerly * , Monica McGarrity †
*Baylor University, TX, USA †Texas Parks and Wildlife Department, TX, USA
shaif_chowdhury1@baylor.edu, greg_hamerly@baylor.edu,Monica.Mcgarrity@tpwd.texas.gov

Abstract

Aquatic invasive species like dreissenid mussels disrupt ecological balance and damage agricultural infrastructure. Machine vision tools can use plankton or water samples images for early detection of invasive dreissenid larvae. Supervised learning techniques require large amounts of labeled data, which is costly to acquire in the case of invasive species. Additionally, invasive species larvae can be rare among aquatic organisms, leading to the problem of data imbalance.

Active Learning (AL) reduces labeled data needs by iteratively selecting and labeling the most informative data for model training. In this paper, we propose an innovative active learning strategy for recognition of aquatic dreissenid larvae with minimal labeled data, while being robust to data imbalance. Our strategy is based on a combination of supervised contrastive training and k-means clustering. The key idea of our algorithm is to project the data into a smaller, more discriminative representation using contrastive learning, where we can apply clustering to select the most informative samples.

We evaluate our algorithm on invasive larvae data and compare with several state-of-the-art AL methods. Traditional AL methods face challenges in generalization, class bias, and low-budget effectiveness. Our method provides an efficient sampling process that is effective in the class-imbalanced, low budget setting. Starting with only 100 samples, after 100 additional active learning samples we get 78% balanced accuracy, which is a 27% improvement over random sampling and 22% over core-set.

1. Introduction

Zebra and quagga mussels are dreissenid mussels native to Eurasia but have been introduced and become invasive to the United States waterways [66]. These invasive species can cause ecological damage by out-competing native species for food as well as attach themselves to organisms, pipes, boats and other critical infrastructure [48]. Invasive mussels attach themselves to hard surfaces and concrete blocks, thereby restricting the flow of water through hydroelectric, irrigation, and fish facilities [29]. Recreational activities on Great Lakes are adversely affected as mussels accumulate on docks, boat hulls, anchors. Maritime archaeologists also note increased shipwreck deterioration caused by quagga mussels in the Great Lakes. The annual economic impact of invasive species specifically on power plants and municipal drinking water systems in North America has been estimated at between $ 267 million and $ 1 billion [11, 29].

A single female adult zebra mussel can produce over a millions of eggs in a year, which grow into veliger larvae that are free-floating in the water column until they settle and attach to hard surfaces. Dreissenid mussels attached to watercraft [1] can live outside of water for several days under the right conditions, thereby spreading to other water bodies [29]. Thus, early detection at the larvae stage and regular monitoring is key to response against invasive mussels. Traditionally, veliger/larvae presence is detected by examining water sample using cross-polarized light microscopy [29] or e-DNA [12]. These processes are generally time consuming, causing delays in ability to implement water infrastructure mitigation or conduct rapid response, and can be costly. A recent state-of-the art method is to use digital imaging-in-flow instruments with image recognition models to detect veliger presence in water samples [23, 38, 39]. This process aims to rapidly detect veliger presence using deep learning based models without any human supervision. A set of invasive and non-invasive images taken from a data set of water sample video are shown in Figure 1.

In the last few years, deep learning based models have made massive advancements in almost all areas of machine learning [41] especially in image classification [16, 24, 37, 67, 79]. However, one of the drawbacks of deep learning is that it is often dependent on the availability of large amounts of labeled data [14, 68, 75]. Most deep learning based methods like Vision Transformer are built to take advantage of

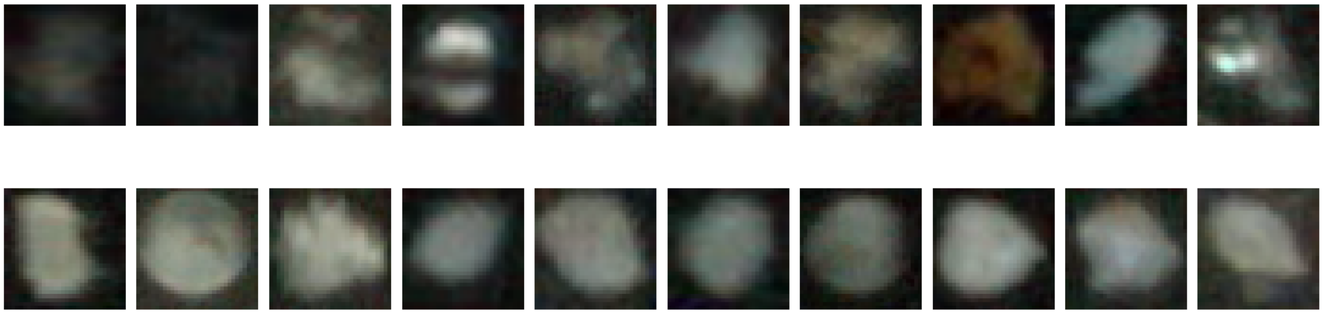

Figure 1. Example of invasive and non-invasive species larvae in our data set. The first row contains images of non-invasive larvae and the following row contains images of invasive larvae. The sample images demonstrate that the visual distinction between invasive and non-invasive species is quite subtle. It can be challenging for human annotators to discern invasive species unless they possess experience and expertise in analyzing larval images.

pre-training on large annotated data sets [16]. However, large annotated data sets are costly to build and at times impossible to get. There are many areas in Computer Vision like underwater imaging [28], medical imaging [49, 52, 68], microscopy imaging [7, 45] etc. where data annotation is very costly. In the case of invasive species, labelled examples are costly to acquire, as it requires previous experience and domain expertise to distinguish between invasive and non-invasive larvae. In the early introduction stage when detection is critical, dreissenid veligers are rare compared to non-invasive species and have a lot of seasonal variance too. So, there is an added problem of data imbalance in invasive larvae data sets. Active learning [44, 59] is a promising solution to this challenge, as it enables the development of models that require fewer labeled examples by iteratively training on the most informative instances of the data set.

Our aim is to build an active learning system that can maximize the accuracy of invasive species recognition model with minimal labels, while being robust to data imbalance issue. In active learning (AL) a model is learned with a small amount of training data and then a subset of unlabeled data is annotated to be used for incrementally training the model [65]. In order to maximize the accuracy of the model with fewer annotated data, AL methods generally rely on the latent space, data distribution, uncertainty or other heuristics based approaches to pick out the best samples [44, 60, 65].

In recent years there has been a lot of interest in active learning research [6, 10, 35, 54, 55, 63, 77, 78] due to its capability of learning from limited labeled data. However, traditional active learning approaches have several limitations [51]. Uncertainty sampling relies heavily on model architectures and might be difficult to generalize for different models, hyper-parameters and data sets [44, 51]. These methods can also suffer from efficiency problems due to continuous unbounded retraining and complex heuristics to select training data points. Therefore, there is a need for novel active learning approaches that can overcome these limitations. Our proposed Active learning algorithm is based on a combination of Supervised contrastive learning and k-means clustering which is efficient and robust against data imbalance.

Supervised contrastive learning (SCL) [34] allows us to apply self-supervised contrastive training in a supervised setting. The idea behind contrastive learning is to have an anchor and a group of positive and negative samples for each anchor. In the embedding space the distance between negative samples and anchor are maximized, while the distance between the positive samples and anchor are minimized. A positive sample is generally formed using data augmentation of the anchor and a negative sample is chosen from the examples in the rest of the batch. Unlike traditional supervised learning, contrastive training can learn a representation that captures the intrinsic structure of the data [69]. Contrastive learning has been effective in various downstream tasks in Computer Vision, especially in cases where there is a lack of labeled data [22, 27, 40, 47, 72, 76]. Due to its representation learning capability, contrastive learning can be really useful for active learning, especially in the process of sample selection.

k-means is a widely used unsupervised clustering technique, especially useful for low dimensional 1D vectors [43]. K-means starts with a set of cluster centers and it tries to minimize the distance between a center and a point within the same cluster. K-means is a well-studied algorithm and there are various improvements to support scalability and stability of clustering [2, 21, 43, 80]. In the context of active learning k-means can be used to guide the selection of samples that maximize the diversity of the data used for model training.

The combination of active learning and k-means presents a great opportunity for active learning, especially in the case of large scale deep learning models. SCL enables an encoder to learn an effective, discriminative data representa-

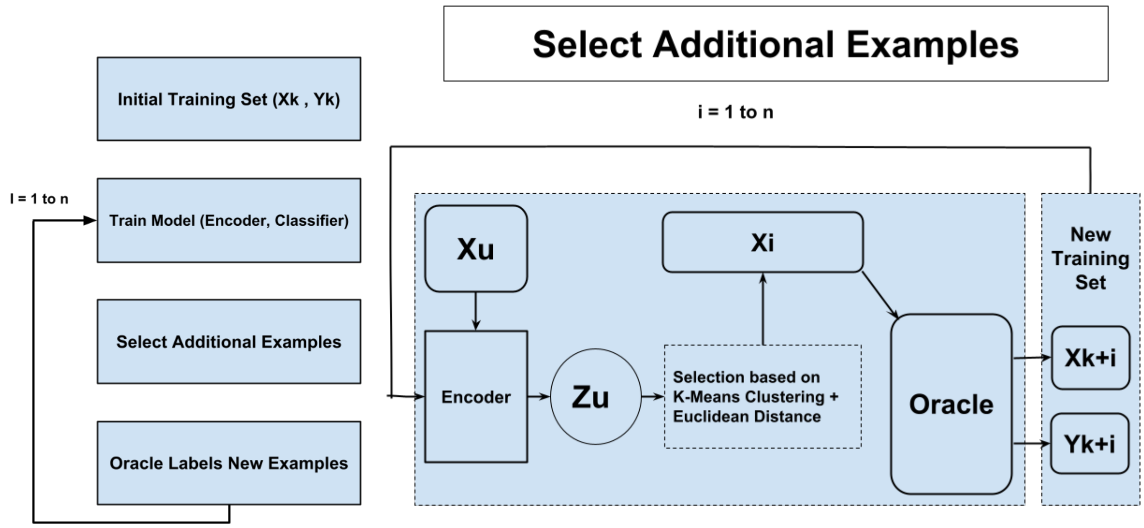

Figure 2. On left we have a flowchart for pool-based active learning. On the right we expand on the sampling process. Here X_k,Y_k are initial training set. X_u is the unlabeled set from which X_i data points are sleeted in an iteration. Finally, an oracle would give the annotation for those samples to create a new training setX_{k+i},Y_{k+i}

tion [34], which can serve as a foundation for effective clustering. K-means can cluster the learned representation into distinct groups. Now, we can use sample of data from each cluster to select a diverse batch of data. We use Euclidean distance [13] to maximize the information gained from samples within each cluster. K-means clustering is fast, effective and highly informative for a smaller representation of the data [2, 21, 43]. In short, the combination results in a highly scalable, efficient and stable active learning strategy that achieves strong performance across different ranges of data sets and experimental settings.

In this paper, we develop an active learning approach to address the high annotation cost of aquatic invasive species. The presence of invasive zebra and quagga mussels have a lot of seasonal variation as well as changes in life stages [31]. Larvae images also vary a lot based on time of the day, whether the samples are from surface or deep water, using kayak or pump [31]. Large part of our current zebra mussel data set [8, 9] is based on images collected from water sampling in 2019. Active learning can help fine-tune specialized models for different conditions and adapt to these changes [56, 58, 73]. We present a novel AL algorithm shown in Figure 2 that combines contrastive learning with k-means clustering to select highly informative, diverse samples. To the best of our knowledge the combination of contrastive learning and k-means clustering has never been used before for Active learning [44]. We primarily test our active learning approach on invasive species data set and conduct experiments in different settings. Our experiments show that the proposed AL algorithm is very effective in dealing with imbalanced data in the low budget regime, which is crucial for invasive larvae recognition. To show the robustness of our approach against imbalanced data we'd report balanced accuracy and F1-score on invasive species. To establish consistent improvement over random sampling we conduct experiment on other data sets like CIFAR10 [36] CIFAR100 [36].

2. Related Work

In many U.S. states, efforts are in place to mussel infestations and mitigate impacts on infrastructure [11, 46]. However, the economic impact of invasive species remains substantial [11, 26, 46]. It is estimated to be around \$219 billion in the U.S. and over \$4 trillion globally [11]. To mitigate both environmental and financial repercussions, an automated early detection system would play a crucial role [11].

Early detection methods for invasive mussels generally rely on microscope photography and eDNA-based methods [29], which is expensive as well as time-consuming. Deep learning based image recognition methods have been introduced before to recognise images of invasive species from water sample videos [23, 38, 39]. But annotation of dreissenid veligers requires a lot of domain expertise and it is costly to acquire. So, in this paper we're investigating the use of active learning to reduce the annotation burden while achieving good accuracy of invasive species recognition. In addition, we would look into some of the recent active learning approaches and discuss how we improve from that.

Compared to traditional machine learning methods Deep

Neural Networks (DNN) are able to take advantage of massive annotated data sets and have made a lot of progress in different areas like image recognition [16, 24, 37, 67, 79], natural language processing [33, 70, 74] etc. Active learning tries to maximize the performance of a model with fewer labeled data points. A Combination of DNN and AL has been used in areas like image recognition [17, 19, 20], text classification [61, 81] and object detection [18, 57] etc. Among different AL methods, (like pool-based, stream-based, membership query-based etc.) pool based active learning is most frequently used [10, 51, 60]. In this paper, we focus on pool-based active learning for image classification using Neural Networks. In pool-based active learning the most frequently used ones are uncertainty based [30,71], diversity based [19, 53, 63] and a few approaches that try to consider both diversity and uncertainty [3]. Here we establish the foundation of our method by discussing relevant work in the areas of uncertainty-based sampling, diversity sampling, and contrastive learning.

Uncertainty based methods try to select the most ambiguous samples based on confidence or entropy. The assumption is that adding samples that lie close to the decision boundary would lead to the highest information gain. These uncertainty based methods are hard to generalize for different models and are often biased towards certain classes leading to poor results especially in low budget regime [44, 51]. The difficulty with uncertainty sampling is defining the regions of uncertainty in the decision space and generalizing the sampling process across different data sets. There are several papers that try to define uncertain regions for a classifier by modifying the softmax layer [50], using a validation set with softmax [32] etc. An ensemble of classifiers have also been used to estimate uncertainty for active learning with encouraging results [5].

Among diversity based methods, Sener et al. [63] proposed a core-set based approach to enforce diversity of sampled labels on the unlabeled batch. Core-set tries to choose a subset of points for a batch that can represent the whole batch. This approach doesn't consider the informativeness of the samples and is difficult to use in case of large data set or batch size [10]. Another interesting approach to active learning is by Ash et al. [3] where the authors propose using gradients of the final layer of the network as a representation and use kmeans++ to select the samples. Here, the gradient vector size is dependent on the number of classes and can lead to inconsistent result [10].

In our method, we try to address some of these challenges using the combination of contrastive learning and k-means clustering. Contrastive learning tries to learn a representation of the data where similar data points are brought closer together, while dissimilar data points are pushed further apart. This is accomplished by optimizing a contrastive loss function. A classifier can be attached on top of it for supervised image classification. Due to the representation learning capability, contrastive learning is especially effective at learning from limited labeled data [47, 72, 76] and has been effective in different downstream tasks in Computer Vision [15,40,42,64]. Pool based Active learning can often depend on the initial labeled set. The ability of contrastive learning to generalize representation to unseen data makes it a good candidate for scalable active learning.

We use Supervised contrastive learning framework introduced by Khosla et al [34]. The encoder is used to learn a discriminative representation of the data and k-means clustering is used to select the samples iteratively. Kmeans++ based initialization introduced by Arthur et al. is used to speed up the sampling process. By building on the insights from these prior works, our method advances the state-of-the-art in active learning, providing improved results and scalability in variety of different data sets and experimental settings.

Active learning is a promising ML paradigm that can significantly reduce dependence on labeled data and speed up training process. But, there are several reproducibility related challenges in AL algorithms. To this end, Munjal et al. [51] have shown that the performance gains of different AL algorithms over random sampling are inconsistent over different experimental settings. The authors have given several recommendations and guidelines in order to improve the robustness and reproducibility of AL algorithms. Based on these guidelines we conducted our experiments under varying sizes of initial labeled data and annotation budget. We try to keep hyper-parameter settings consistent as much as possible and we share details of all our hyper-parameters and other experimental settings used in evaluation of our AL algorithm.

3. Method

Let's define data set X of size N as a set of all examples, where X_L = (x_i, y_i) is the set of labeled example and X_U = (x_j, y_j) is the set of unlabeled example. Initially a base model Ψ_0 is trained on the data set X_L. At every step of active learning, we need to select B examples from X_U for manual annotation and further training of the model. The selected example set $S \subseteq X_U$, $B = |S| \leq N$ is added to X_L to create a new training set X_L^i and train a new model Ψ_1. Generally, the cycle of training and annotation is repeated until the annotation budget is exhausted or a satisfying metric is achieved. Most active learning methods have unbounded training and annotation cycles, which can be computationally expensive. So, in our case, the number of iterations would be fixed and the amount of annotation budget would be varied for evaluation of the proposed algorithm. We train the model using supervised contrastive learning and use k-means with kmeans++ initialization for sample selection. In the next few paragraphs we would introduce contrastive learning based training strategy, describe our sample selection strategy with k-means in detail. Following that, we'd give the details of neural network architecture, hyper-parameters and different evaluation metrics.

3.1. Supervised Contrastive Learning

Our training method is similar to the supervised contrastive learning paper introduced by Khosla et al [34]. Our classification model is based on two steps 1. Training an encoder to create representation vector from an input image in a way that the vectors from the same class would be similar compared to the vectors from different class; 2. Training a classifier on top of a frozen encoder.

Given a batch of input data, we apply data augmentation twice and feed them to the encoder network. If x is the input image, we create $\hat{x}$ = Aug (x) where Aug() is a function that applies different data augmentation like random flips and rotations on the image. Followed bt that, the encoder would create a representation vector, r = Enc(x), r $\in \mathbb{R}$. In our case, Resnet50V2 architecture [24, 25] is used to build the encoder. A projection network maps r to f(z) where contrastive loss is applied, z = Proj(r). When the labels are not given the loss is calculated by taking the augmented sample as a positive example and the loss function for a sample in the batch looks like this

$$L = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k \in A(i)} \exp(\text{sim}(z_i, z_k)/\tau)} \quad (1)$$

Here, z_i is the anchor and z_j is the augmented positive sample and z_k are the negative samples in a mini-batch. sim() calculates the cosine similarity between two vectors and τ is a scalar temperature parameter. So, for each anchor there is one positive example and 2(N-1) negative examples. A(i) has a total of 2N-1 terms and is defined as $A(i) = \{k | k \in 1, 2, ..., 2N, k \neq i\}$.

In case of supervised contrastive learning there are more positive samples than the augmented image. So, a modification is used to incorporate the label information and it is given by

$$L = -\frac{1}{P(i)} \sum_{p \in P(i)} \log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k \in A(i)} \exp(\text{sim}(z_i, z_k)/\tau)} \quad (2)$$

Here $P(i)$ includes indices of all positive samples in the batch, including the augmented image. contrastive loss generally works better with a large batch size due to a greater number of negative images. Once the encoder is trained with a contrastive loss, a classifier is attached to make the final prediction and it is trained with a cross-entropy loss.

Algorithm 1: Active Learning using Contrastive Training and k-means Clustering

Input: Labeled set (X_L,Y_L), Unlabeled set X_U, annotation budget B, number of clusters k
Output: Updated labeled set X_L

for $i = 1$ *to* n **do**
  Train encoder, classifier on (X_L,Y_L) using supervised contrastive learning (SCL);
  Use the encoder to generate representations z_U from X_U;
  Perform k-means clustering on z_U to create k clusters;
  for *each cluster* C_j **do**
    Calculate centroid c_j;
    Select sample x^* in C_j that maximizes distance to c_j;
    if *number of labeled samples* $< B$ **then**
      Add (x^*, y^*) to X_L where y^* is the label for x^* given by the annotator;
      Remove x^* from X_U;
      Continue to train encoder and classifier on X_L using SLR;
    else
      break;
  if *number of labeled samples* $\geq B$ **then**
    break;

3.2. Sampling with k-means

As shown in algorithm 15, we start with a set of labeled examples X_L and train using contrastive learning. Once we have trained the encoder, it is used to create a set of representations z_u from the unlabeled set X_U. Given the representations learned by the CLR, we apply k-means clustering to partition the unlabeled set into distinct clusters based on the number of classes. Each cluster represents a specific class from the data and captures the diversity of the data. For each cluster, we select the samples that are most distant from cluster centers using euclidean distance. Let's assume that there are k clusters with k centroids μ_i where $i \in \{1, 2, ..., k\}$. For each cluster, there are S data points as z_s. We calculate the distance of each of those data points using euclidean distance, given by $d_s = \sqrt{(z_s - \mu_i)^2}$. The largest values of d_s are used to select samples from each cluster and these are used to continue training the model.

To improve the efficiency of our sampling algorithm we integrate Mini-Batch k-means algorithm into our active learning. Mini-Batch Kmeans, described by Sculley et al [62] operates on a random subset of data, which reduces computational cost while preserving cluster quality [4]. In addition to this, we use kmeans++ for selecting the initial centroids. Kmeans++ proposed by Arthur et al. [2] selects initial centers in a way that leads to improved clustering and faster convergence. kmeans++ selects one centroid randomly and each subsequent centroid is selected based on probability proportional to its distance squared to the nearest existing centroid. By integrating Mini-Batch and kmeans++, we improve the efficiency and effectiveness of our algorithm.

This selection is based on the distance of the samples to the cluster centroids, with the assumption that these samples are different from the initial labeled data which has been used to train the model already. By iteratively performing this selection process, our approach effectively reduces the volume of data that needs to be labeled, while training the model on diverse and highly informative samples. Selection from different clusters also ensure that the model is robust to data imbalance which is crucial in case of invasive species.

4. Experimental Results

We evaluate our algorithm on classification of aquatic invasive species data set. We seek to demonstrate improve the accuracy of our method with minimal annotation along with efficiency, robustness to data imbalance, robustness to changes in experimental setting and budget. For the encoder, we use ResNet50V2 architecture with ImageNet weights and a projector of feature size 2048. For active learning, we start with a base set of labeled examples and then choose a set of examples to annotate in each iteration. For all our experiments, we keep the number of AL iterations fixed to 10. We vary the amount of labeled data at start and also the budget available during AL iterations. The results are sensitive to the size of the base labeled set and budget. However, a good choice of budget depends on the data set size and number of clusters k. We compare our algorithm to different baseline AL methods listed here

1. **RAND:** This is the active learning baseline that randomly selects examples every iteration [51]. In this case, we train a ResNet50V2 in a supervised manner and select samples for annotation randomly.

2. **ENTROPY:** This is the active learning algorithm that selects samples with higher entropy. This method [30] tries to select instances for which the model's prediction confidence exhibits a greater level of variability.

3. **CONF:** This is the active learning algorithm that aims to acquire the most uncertain data samples. [71] In this case, the algorithm tries to find data samples that are challenging to classify or predict confidently and increase the robustness of the model.

4. **CORE-SET:** This is the active learning algorithm is based on core-set selection, which tries to select a small set of points for annotation that approximate the shape of the larger unlabeled data set. Core-set [63] is widely considered as one of the state-of-the-art methods for Active learning.

5. **RAND-CLR:** Here we compare random sampling based active learning with supervised contrastive learning for model training. The other baseline approaches mentioned above are based on simple supervised training of ResNet50V2. But, in case of RAND-CLR, we train using contrastive learning and make the selection of samples randomly.

4.1. Results on Invasive Species data set

Our invasive veliger data set is created from water sample videos. The larvae images are cropped using a proprietary algorithm that is based on a Kalman filter. The data set contains cropped images of 6,905 organisms, with a total of 178904 images. There are 42055 invasive images and 136849 non-invasive images. The data set is imbalanced towards non-invasive species. So, instead of accuracy, we have decided to use F1 Score of invasive species and Balanced Accuracy as evaluation criteria. The average image size is approximately 22×19 pixels. So, we resize images to $(32\times32\times3)$ with padding and cropping based on whether the original image is larger or smaller. We use ResNet50V2 with ImageNet weights for training with Adam Optimizer and cross-entropy loss. During Active learning process we start with an initial base set of images and then select a set of images to be annotated during 10 iterations. Our experiments are performed using varying range of base-set and annotation-set. We train the base-set for 10 epochs and then train the new training set for 10 epochs in each iteration. For contrastive learning we train the encoder for 10 epochs and the classifier for 10 epochs. We present the average results after 3 iterations of training and compare them with the baselines mentioned above.

Initially, we vary the number of images in the initial labeled set from 1500 to 7000 and we select a total of 500 images for annotation in 10 AL iterations. We plot the results on Figure 3 and Figure 4. Those figures compare the Invasive Species F1 score or Balanced Accuracy against the size of base labeled set. The results are based on three different experiments with the mean score on X-axis along with the standard deviation plotted using area. The results demonstrate that our model can achieve high accuracy with very minimal labeled data and outperform all other AL methods specially when the initial labeled set is really small. This is mostly due to the representation learning capability of contrastive learning. Selecting samples from different classes ensures the diversity and selection of most distant samples ensures that most informative samples among the unlabeled data is used for active learning. The data points that are

]

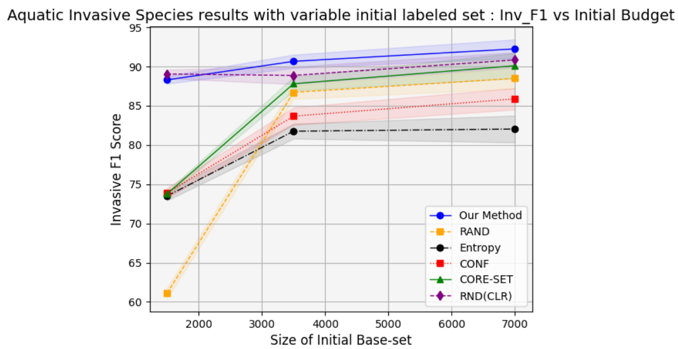

Figure 3. Invasive species results with different variations of initial labeled set. We plot the AL initial labeled set budget in X-Axis and F1-score on Y-Axis. The results are based on the average of three experiments and our algorithm clearly outperforms other approaches of active learning.

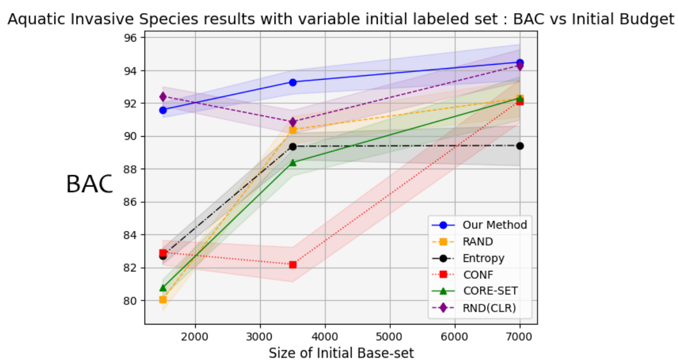

Figure 4. Invasive species results with different variations of initial labeled set. We plot the AL initial labeled set budget in X-Axis and Balanced Accuracy(BAC) on Y-Axis. Due to high data imbalance some of the comparing methods don't provide constant improvement with Active Learning. Compared to that our method performs well even with data imbalance and shows consistent improvement.

most distant from the cluster center are chosen using euclidean distance.

We conduct further experiments using a fixed amount of initial labeled set of 100 images and vary the AL annotation budget from 100 to 2000. We plot the results in Figure 5 and Figure 6. The results demonstrates the advantage of our algorithm over other AL methods especially in low budget regime.

4.2. Other data sets

To show the robustness of our model we performed some experiments on standard image recognition data sets like CIFAR10, CIFAR100. For random sampling, we train on the base labeled example for 50 epochs and then 10 epochs

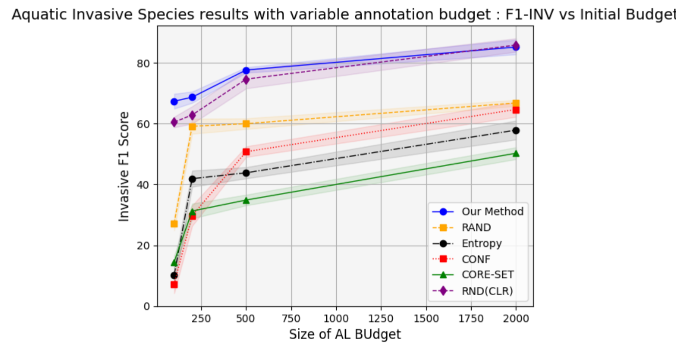

Figure 5. In this case, we start with a constant size of base set. From that, we vary the AL annotation budget and plot the F1-Score of invasive species.

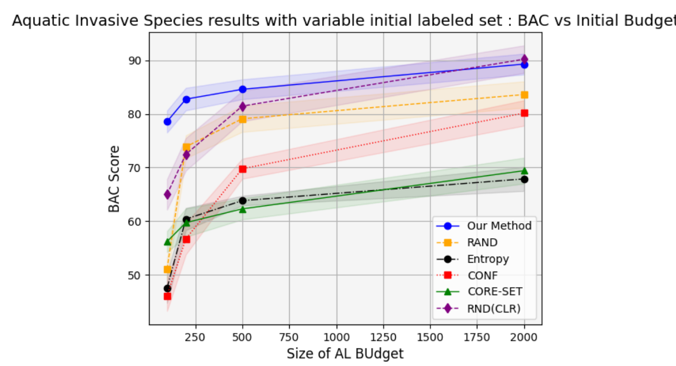

Figure 6. We also plot Balanced Accuracy (BAC) with variable AL annotation budget. Our approach provides huge improvement in low budget regime which is crucial for invasive species recognition.

in each iteration. For our contrastive learning based algorithm, we train the encoder for 50 epochs and the classifier for 10 epochs initially, and then 10 epochs of training for both encoder and classifier in each iteration. Because CIFAR10 and CIFAR 100 have larger number of classes, we decided to train for longer epochs on the initial labeled set.

CIFAR-10

CIFAR-10 has 50,000 images for training and 10,000 for testing. Each image is of size (32×32) and belongs to one of the 10 different classes. The images are evenly distributed between different classes. We use the ResNet50V2 architecture and vary the number of labeled examples during AL iterations. We plot the results in Figure 7 with accuracy on X-axis and annotation budget on Y-axis.

CIFAR-100

CIFAR-100 has 60,000 images of 100 different classes with each class having 60 images. Similar to CIFAR-10 the image size is (32×32) with 50,000 images for training and 10,000 images for test. We plot the results in Figure 8. The results show that our algorithm gives strong performance

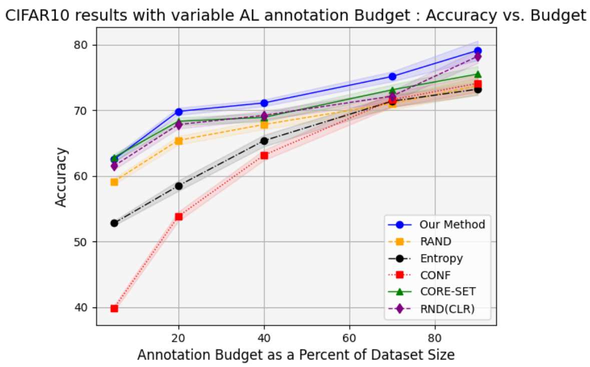

Figure 7. Comparative performance of our proposed active learning approach against random sampling on CIFAR-10 data set. For random sampling, we train on the base labeled example for 50 epochs and then 15 epochs in each iteration. For our contrastive learning based algorithm, we train the encoder for 50 epochs and the classifier for 10 epochs initially, and then 10 epochs of training for both encoder and classifier in each iteration. We plot the AL annotation budget in X-Axis and the accuracy on Y-Axis. As shown in the plot, our algorithm consistently outperforms other approaches of active learning. RND(CLR) is close to our method due to the quality of contrastive representation learning in in the small data regime.

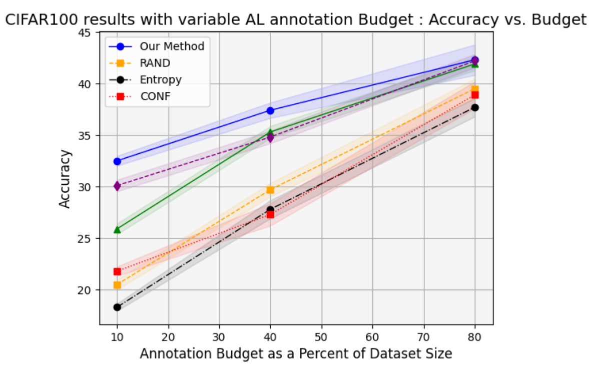

Figure 8. Comparative performance of our proposed active learning approach against random sampling on CIFAR-100 data set. We plot the AL annotation budget in X-Axis and the accuracy on Y-Axis. Our algorithm provides strong improvement over other approaches of active learning.

across different data sets and experimental settings.

For invasive species, we also plot some of the examples selected by our AL method during different iterations on Figure 9. Due to k-means clustering, our method selects specific types of examples during the different iterations and thereby targets different distributions of the data set. Compared to that, Random sampling doesn't have any clear pattern and doesn't lead to a lot of improvement.

The results show that, active learning can be used to build an invasive species recognition model using minimal labeled data. Our algorithm is robust to data imbalance and outperforms baseline solutions across different data sets and experimental settings. However, the effectiveness of our approach depends a lot on the quality of the representation learning. Due to this, starting with pre-trained ImageNet weights might be helpful. Contrastive learning is also affected by batch size and that might make our approach computationally expensive in some cases. Our active learning experiments are based on fixed number of iteration and the number of iterations might have an impact on the results. This should ideally be chosen based on the time and computational resources available. Our experiments also assume noiseless human annotation during AL iterations, which might not be true in every case, especially for application in invasive species recognition.

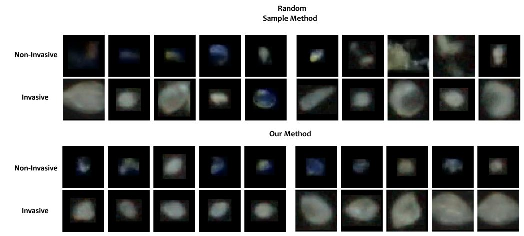

Figure 9. Comparison of samples selected by our algorithm and Random Sampling

5. Conclusion

The application of deep learning for recognition of invasive veligers is potentially rewarding but difficult task due to the cost of acquiring labeled data. We have proposed a framework that utilises active learning to overcome these difficulties. Our approach is based on contrastive learning to iteratively train an encoder and a classifier. The encoder projects the data to a smaller representation space, where we apply k-means to select the most informative samples. Selection of most distant samples from different clusters ensure that our algorithm is robust against data imbalance while getting high accuracy. With 100 initial labeled samples and 100 during active learning, we achieve 78% BAC, rising to 85% with 2000 AL samples. Our algorithm consistently outperforms state-of-the-art on invasive larvae as well as datasets like CIFAR10, CIFAR100. However, there is still much work needed to have working application that can automate invasive veliger recognition.

In this paper, we concentrate on using active learning to recognize zebra and quagga mussel larvae. As more invasive species are anticipated, the focus of research needs to shift to new species. Nonetheless, our paper provides an active learning framework that can reduce the annotation cost of invasive larvae, facilitating the construction of efficient and robust deep learning recognition models.

References

[1] Josef Daniel Ackerman, Blair Sim, S Jerrine Nichols, and Renata Claudi. A review of the early life history of zebra mussels (dreissena polymorpha): comparisons with marine bivalves. *Canadian Journal of Zoology*, 72(7):1169–1179, 1994. 1

[2] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. Technical report, Stanford, 2006. 2, 3, 6

[3] Jordan T Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. Deep batch active learning by diverse, uncertain gradient lower bounds. *arXiv preprint arXiv:1906.03671*, 2019. 4

[4] Javier Béjar Alonso. K-means vs mini batch k-means: a comparison. 2013. 5

[5] William H Beluch, Tim Genewein, Andreas Nürnberger, and Jan M Köhler. The power of ensembles for active learning in image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9368–9377, 2018. 4

[6] Mélanie Bernhardt, Daniel C Castro, Ryutaro Tanno, Anton Schwaighofer, Kerem C Tezcan, Miguel Monteiro, Shruthi Bannur, Matthew P Lungren, Aditya Nori, Ben Glocker, et al. Active label cleaning for improved dataset quality under resource constraints. *Nature communications*, 13(1):1161, 2022. 2

[7] Yuan Chen, Shangpeng Liu, Peiran Tong, Ying Huang, He Tian, and Fang Lin. A small-dataset-trained deep learning framework for identifying atoms on transmission electron microscopy images. *Scientific Reports*, 13(1):2631, 2023. 2

[8] Shaif Chowdhury and Greg Hamerly. Recognition of aquatic invasive species larvae using autoencoder-based feature averaging. In *International Symposium on Visual Computing*, pages 145–161. Springer, 2022. 3

[9] Shaif Chowdhury, Sadia Nasrin Tisha, Monica McGarrity, and Greg Hamerly. Video-based recognition of aquatic invasive species larvae using attention-lstm transformer. Technical report, EasyChair, 2023. 3

[10] Gui Citovsky, Giulia DeSalvo, Claudio Gentile, Lazaros Karydas, Anand Rajagopalan, Afshin Rostamizadeh, and Sanjiv Kumar. Batch active learning at scale. *Advances in Neural Information Processing Systems*, 34:11933–11944, 2021. 2, 4

[11] Nancy A Connelly, Charles R O'Neill, Barbara A Knuth, and Tommy L Brown. Economic impacts of zebra mussels on drinking water treatment and electric power generation facilities. *Environmental management*, 40:105–112, 2007. 1, 3

[12] Dominique A Cowart, Koen GH Breedveld, Maria J Ellis, Joshua M Hull, and Eric R Larson. Environmental dna (edna) applications for the conservation of imperiled crayfish (decapoda: Astacidea) through monitoring of invasive species barriers and relocated populations. *Journal of Crustacean Biology*, 38(3):257–266, 2018. 1

[13] Per-Erik Danielsson. Euclidean distance mapping. *Computer Graphics and image processing*, 14(3):227–248, 1980. 3

[14] Harriet L Dawson, Olivier Dubrule, and Cédric M John. Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification. *Computers & Geosciences*, 171:105284, 2023. 1

[15] Ali Diba, Vivek Sharma, Reza Safdari, Dariush Lotfi, Saquib Sarfraz, Rainer Stiefelhagen, and Luc Van Gool. Vi2clr: Video and image for visual contrastive learning of representation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1502–1512, 2021. 4

[16] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR*, 2020. 1, 2, 4

[17] Xuefeng Du, Dexing Zhong, and Huikai Shao. Building an active palmprint recognition system. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1685–1689. IEEE, 2019. 4

[18] Di Feng, Xiao Wei, Lars Rosenbaum, Atsuto Maki, and Klaus Dietmayer. Deep active learning for efficient training of a lidar 3d object detector. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 667–674. IEEE, 2019. 4

[19] Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In *International conference on machine learning*, pages 1183–1192. PMLR, 2017. 4

[20] Denis Gudovskiy, Alec Hodgkinson, Takuya Yamaguchi, and Sotaro Tsukizawa. Deep active learning for biased datasets via fisher kernel self-supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9041–9049, 2020. 4

[21] Greg Hamerly. Making k-means even faster. In *Proceedings of the 2010 SIAM international conference on data mining*, pages 130–140. SIAM, 2010. 2, 3

[22] Hongfeng Han, Nanyi Fei, Zhiwu Lu, and Ji-Rong Wen. Supervised contrastive learning for few-shot action classification. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part III*, pages 512–528. Springer, 2023. 2

[23] Whitney Hassett, Julie Zimmerman, Gretchen Rollwagen-Bollens, Stephen M Bollens, and Timothy D Counihan. An experimental evaluation of the efficacy of imaging flow cytometry (flowcam) for detecting invasive dreissened and corbiculid bivalve veligers. *Lake and Reservoir Management*, 37(4):406–417, 2021. 1, 3

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 4, 5

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 630–645. Springer, 2016. 5

[26] Amy Lee Hetherington, Lars G Rudstam, Rebecca L Schneider, Kristen T Holeck, Christopher W Hotaling, John E

Cooper, and James R Jackson. Invader invaded: population dynamics of zebra mussels (dreissena polymorpha) and quagga mussels (dreissena rostriformis bugensis) in polymictic oneida lake, ny, usa (1992–2013). *Biological Invasions*, 21(5):1529–1544, 2019. 3

[27] Xinrong Hu, Dewen Zeng, Xiaowei Xu, and Yiyu Shi. Semi-supervised contrastive learning for label-efficient medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24*, pages 481–490. Springer, 2021. 2

[28] Leilei Jin and Hong Liang. Deep learning for underwater image recognition in small sample size situations. In *OCEANS 2017-Aberdeen*, pages 1–4. IEEE, 2017. 2

[29] Ladd E Johnson. Enhanced early detection and enumeration of zebra mussel (dreissena spp.) veligers using cross-polarized light microscopy. *Hydrobiologia*, 312(2):139–146, 1995. 1, 3

[30] Ajay J Joshi, Fatih Porikli, and Nikolaos Papanikolopoulos. Multi-class active learning for image classification. In *2009 ieee conference on computer vision and pattern recognition*, pages 2372–2379. IEEE, 2009. 4, 6

[31] Alexander Y Karatayev, Lyubov E Burlakova, and Dianna K Padilla. Zebra versus quagga mussels: a review of their spread, population dynamics, and ecosystem impacts. *Hydrobiologia*, 746:97–112, 2015. 3

[32] Nicholas Kashani Motlagh, Jim Davis, Tim Anderson, and Jeremy Gwinnup. Learning when to say “i don’t know. In *Advances in Visual Computing: 17th International Symposium, ISVC 2022, San Diego, CA, USA, October 3–5, 2022, Proceedings, Part I*, pages 196–210. Springer, 2022. 4

[33] Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT*, pages 4171–4186, 2019. 4

[34] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020. 2, 3, 4, 5

[35] Andreas Kirsch, Joost Van Amersfoort, and Yarin Gal. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. *Advances in neural information processing systems*, 32, 2019. 2

[36] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 3

[37] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017. 1, 4

[38] FirstName LastName. Recognition of aquatic invasive species larvae using autoencoder-based feature averaging. International Symposium on Visual Computing:inv.pdf, 2022. 1, 3

[39] FirstName LastName. Video-based recognition of aquatic invasive species larvae using attention-lstm transformer. To appear in International Symposium on Visual Computing 2023:inv.pdf, 2023. 1, 3

[40] Phuc H Le-Khac, Graham Healy, and Alan F Smeaton. Contrastive representation learning: A framework and review. *Ieee Access*, 8:193907–193934, 2020. 2, 4

[41] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. 1

[42] Rui Li, Yiheng Zhang, Zhaofan Qiu, Ting Yao, Dong Liu, and Tao Mei. Motion-focused contrastive learning of video representations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2105–2114, 2021. 4

[43] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. The global k-means clustering algorithm. *Pattern recognition*, 36(2):451–461, 2003. 2, 3

[44] Peng Liu, Lizhe Wang, Rajiv Ranjan, Guojin He, and Lei Zhao. A survey on active deep learning: From model driven to data driven. *ACM Computing Surveys (CSUR)*, 54(10s):1–34, 2022. 2, 3, 4

[45] Zhichao Liu, Luhong Jin, Jincheng Chen, Qiuyu Fang, Sergey Ablameyko, Zhaozheng Yin, and Yingke Xu. A survey on applications of deep learning in microscopy image analysis. *Computers in Biology and Medicine*, 134:104523, 2021. 2

[46] Frances Lucy and Elizabeth Muckle-Jeffs. History of the zebra mussel/icais conference series. *Aquatic Invasions*, 2010. 3

[47] Orchid Majumder, Avinash Ravichandran, Subhransu Maji, Alessandro Achille, Marzia Polito, and Stefano Soatto. Supervised momentum contrastive learning for few-shot classification. *arXiv preprint arXiv:2101.11058*, 2021. 2, 4

[48] André L Martel and Jacqueline Baylin Madill. Twenty-six years (1990–2015) of monitoring annual recruitment of the invasive zebra mussel (dreissena polymorpha) in the rideau river, a small river system in eastern ontario, canada. *Canadian Journal of Zoology*, 96(10):1071–1079, 2018. 1

[49] Axel H Masquelin, Nicholas Cheney, C Matthew Kinsey, and Jason HT Bates. Wavelet decomposition facilitates training on small datasets for medical image classification by deep learning. *Histochemistry and cell biology*, 155:309–317, 2021. 2

[50] Marcin Możejko, Mateusz Susik, and Rafał Karczewski. Inhibited softmax for uncertainty estimation in neural networks. *arXiv preprint arXiv:1810.01861*, 2018. 4

[51] Prateek Munjal, Nasir Hayat, Munawar Hayat, Jamshid Sourati, and Shadab Khan. Towards robust and reproducible active learning using neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 223–232, 2022. 2, 4, 6

[52] Dianwen Ng, Xiang Lan, Melissa Min-Szu Yao, Wing P Chan, and Mengling Feng. Federated learning: a collaborative effort to achieve better medical imaging models for individual sites that have small labelled datasets. *Quantitative Imaging in Medicine and Surgery*, 11(2):852, 2021. 2

[53] Hieu T Nguyen and Arnold Smeulders. Active learning using pre-clustering. In *Proceedings of the twenty-first international conference on Machine learning*, page 79, 2004. 4

[54] Amin Parvaneh, Ehsan Abbasnejad, Damien Teney, Gholamreza Reza Haffari, Anton Van Den Hengel, and Javen Qinfeng Shi. Active learning by feature mixing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12237–12246, 2022. 2

[55] Fengchao Peng, Chao Wang, Jianzhuang Liu, and Zhen Yang. Active learning for lane detection: a knowledge distillation approach. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15152–15161, 2021. 2

[56] Claudio Persello and Lorenzo Bruzzone. Active learning for domain adaptation in the supervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):4468–4483, 2012. 3

[57] Zhenshen Qu, Jingda Du, Yong Cao, Qiuyu Guan, and Pengbo Zhao. Deep active learning for remote sensing object detection. *arXiv preprint arXiv:2003.08793*, 2020. 4

[58] Piyush Rai, Avishek Saha, Hal Daumé III, and Suresh Venkatasubramanian. Domain adaptation meets active learning. In *Proceedings of the NAACL HLT 2010 Workshop on Active Learning for Natural Language Processing*, pages 27–32, 2010. 3

[59] Hiranmayi Ranganathan, Hemanth Venkateswara, Shayok Chakraborty, and Sethuraman Panchanathan. Deep active learning for image classification. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3934–3938. IEEE, 2017. 2

[60] Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B Gupta, Xiaojiang Chen, and Xin Wang. A survey of deep active learning. *ACM computing surveys (CSUR)*, 54(9):1–40, 2021. 2, 4

[61] Christopher Schröder and Andreas Niekler. A survey of active learning for text classification using deep neural networks. *arXiv preprint arXiv:2008.07267*, 2020. 4

[62] David Sculley. Web-scale k-means clustering. In *Proceedings of the 19th international conference on World wide web*, pages 1177–1178, 2010. 5

[63] Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017. 2, 4, 6

[64] Pierre Sermanet, Corey Lynch, Jasmine Hsu, and Sergey Levine. Time-contrastive networks: Self-supervised learning from multi-view observation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 486–487. IEEE, 2017. 4

[65] Burr Settles. Active learning literature survey. 2009. 2

[66] Erik Stokstad. Feared quagga mussel turns up in western united states, 2007. 1

[67] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016. 1, 4

[68] Nima Tajbakhsh, Laura Jeyaseelan, Qian Li, Jeffrey N Chiang, Zhihao Wu, and Xiaowei Ding. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, 63:101693, 2020. 1, 2

[69] Yonglong Tian, Chen Sun, Ben Poole, Dilip Krishnan, Cordelia Schmid, and Phillip Isola. What makes for good views for contrastive learning? *Advances in neural information processing systems*, 33:6827–6839, 2020. 2

[70] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 4

[71] Dan Wang and Yi Shang. A new active labeling method for deep learning. In *2014 International joint conference on neural networks (IJCNN)*, pages 112–119. IEEE, 2014. 4, 6

[72] Peng Wang, Kai Han, Xiu-Shen Wei, Lei Zhang, and Lei Wang. Contrastive learning based hybrid networks for long-tailed image classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 943–952, 2021. 2, 4

[73] Binhui Xie, Longhui Yuan, Shuang Li, Chi Harold Liu, Xinjing Cheng, and Guoren Wang. Active learning for domain adaptation: An energy-based approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8708–8716, 2022. 3

[74] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, pages 2048–2057. PMLR, 2015. 4

[75] Peng Xu, Dhruv Kumar, Wei Yang, Wenjie Zi, Keyi Tang, Chenyang Huang, Jackie Chi Kit Cheung, Simon JD Prince, and Yanshuai Cao. Optimizing deeper transformers on small datasets. *arXiv preprint arXiv:2012.15355*, 2020. 1

[76] Zhanyuan Yang, Jinghua Wang, and Yingying Zhu. Few-shot classification with contrastive learning. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XX*, pages 293–309. Springer, 2022. 2, 4

[77] Ofer Yehuda, Avihu Dekel, Guy Hacohen, and Daphna Weinshall. Active learning through a covering lens. *arXiv preprint arXiv:2205.11320*, 2022. 2

[78] Donggeun Yoo and In So Kweon. Learning loss for active learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 93–102, 2019. 2

[79] Xiaohua Zhai, Alexander Kolesnikov, Neil Houlsby, and Lucas Beyer. Scaling vision transformers. In *CVPR*, pages 12104–12113, 2022. 1, 4

[80] Alibek Zhakubayev and Greg Hamerly. Clustering faster and better with projected data. In *Proceedings of the 6th International Conference on Information System and Data Mining*, pages 1–6, 2022. 2

[81] Ye Zhang, Matthew Lease, and Byron Wallace. Active discriminative text representation learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017. 4