

Investigation of Unsupervised Auto-segmentation for Weak Phytoplankton Annotations

Jason Deglint
Komal Chandiramani
Seoyeon Woo
James Vandersluis
Kathryn E. Thomas
Alexander Wong
Email: {jdeglint, a28wong}@uwaterloo.ca

University of Waterloo, Blue Lion Labs
University of Waterloo, Blue Lion Labs
University of Waterloo, Blue Lion Labs
University of Victoria, Blue Lion Labs
Blue Lion Labs
University of Waterloo

Abstract

When certain phytoplankton and algae bloom, they can be a threat to both freshwater and marine ecosystems. To overcome these challenges many industry players are looking for improved harmful algal bloom (HAB) monitoring. One such promising approach is the use of supervised deep learning to automatically identify and count plankton species in microscope images. However, the main drawback of supervised learning is that it requires a large labeled dataset, which requires significant time from subject matter experts (SMEs) to meticulously annotate images. In this paper we propose an unsupervised auto-segmentation approach to automatically generate noisy segmentations of four common marine phytoplankton species: *Prorocentrum lima*, *Alexandrium catenella*, *Heterosigma akashiwo*, and *Dolichospermum sp.* We demonstrate the efficacy of our method by measuring the Intersection over Union (IoU) metric between the unsupervised auto-segmentation and the ground truth annotations generated by a team of professional taxonomists. Specifically, the average IoU for the four species mentioned were 0.7619, 0.7191, 0.6978, and 0.6495, respectively. Therefore, the feasibility of unsupervised auto-segmentation is viable to create weak labels, which contributes towards building a more sustainable phytoplankton monitoring system.

1 Introduction

Algae are the base of the food chain and ubiquitous in aquatic ecosystems. However, some species of algae can be harmful, even at low concentrations (5 cells/mL), while others don't pose a threat until there are thousands of cells/mL [1]. The presence of harmful algae or harmful algae blooms (HABs) can have an adverse impact on human health, coastal ecosystems, public infrastructure, recreational aesthetic and overall water quality [2]. Further, as a result of climate change, warmer waters have accelerated the growth of HABs in areas that may have previously been under the optimal temperature to support such growth [3, 4].

Aquaculture, including farmed salmon, are susceptible to harmful algae and HABs as they are grown in pens with thousands of fish per pen and with limited mobility and management actions if an algae bloom occurs suddenly [5]. There are multiple ways in which harmful algae can effect finfish: (1) production of extracellular toxins, (2) physical damage to fish gills, and (3) reduction in oxygen in the water column [5]. While there are management actions that can be taken to alleviate some of the risk to farmed salmon, they are often reactive due to slow monitoring methods. The resulting loss or damage of stock has cost fish farmers million to billions of dollars from a single HAB event, and these losses are on the rise [6].

In order to mitigate HAB risk and reduce stock loss, it is important to consistently monitor harmful algae and the development of HABs. One of the approaches to accomplish this goal is to develop deep learning neural networks to quickly identify and count different type of phytoplankton in water sample [7, 8]. However, for the supervised learning paradigm, this requires large amounts of labeled data, which is often time-consuming and costly to generate. Furthermore, as seen in Figure 1 (top), there are many applications, such as medical imaging and biological samples, where obtaining ground truth labels requires highly qualified subject matter experts (SMEs), further exacerbating the problem. This results in datasets that have a small percentage of the total data labeled.

To alleviate this issue, there are several methods that can be applied, including active learning, transfer learning, auto-labeling, self training, weak learning, model-assisted labeling, etc [9–12]. In this paper, as seen in Figure 1 (bottom), we explore the feasibility of using unsupervised auto-segmentation of phytoplankton. To determine the efficacy of this approach, we quantify the difference between ground truth fine-grained polygon segmentation and the proposed unsupervised auto-segmentation.

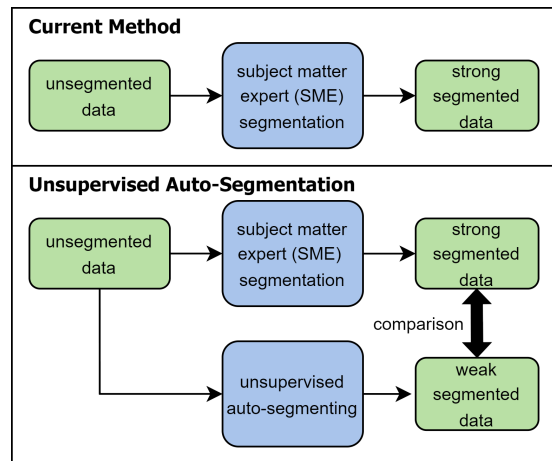


Fig. 1: (top) The current method to segment data requires subject matter experts (SMEs), such as highly trained phytoplankton taxonomists, which is time-consuming and tedious. (bottom) In this paper we quantify the performance of the proposed unsupervised auto-segmentation by comparing it to ground truth segmentations.

2 Background

2.1 Auto-labeling

Auto-labeling is the automatic labeling process which creates labels by aggregating and summarizing the content using a computational procedure [10]. However, labels created based on auto-labeling needed to be reviewed by SMEs for quality assessment and further manipulated to finalize the strong labeled data. Therefore, to further reduce the quality assessment process and enhance label quality, a semi-automated data labeling method could be applied, combining semi-supervised learning and active learning [11]. By providing little supervision to the machine predicted model, the predicted model learned towards where the machine and labeler agree on the final output [11]. Huxohl and Kummert demonstrated that re-training the model using the labels, which were generated with the model-assisted labeling method and modified by the human labeler, increases the model performance by 8.52% [9]. Moreover, in the medical field, auto-labeling method could significantly reduce the 4 hours of manual labeling per study to 8 minutes [13].

2.2 Benefits of Weak Segmented Data

In weak supervision, models are trained on incomplete, inaccurate or partial data. Opposed to a SME slowly annotating a dataset, data can be labeled using lower fidelity methods, which can include crowd sourcing, and using text metadata. First, weak labels can be sent to the SME biology taxonomy annotation team as an initial guess of the segmentation. Since segmentation is one of the more tedious and time-consuming tasks, having an approximate segmentation that one can improve saves significant time. This results in more ground truth strong labels being created, which in turn improves the downstream supervised learning task.

Second, these weakly labeled data can be used to pretrain a segmentation model, of whose model parameters and latent space better represent the final segmentation task compared to using an open-source standard pretrained segmentation model. This pretrained model can then be used as a starting point when training with the ground truth data produced by the annotation team. Intuitively, the closer the weak labels are to the strong labels, the more time is saved for the annotation team, and the closer the pretrained weak supervised model will be to the final strong supervised model, which in turn reduces expensive GPU training time.

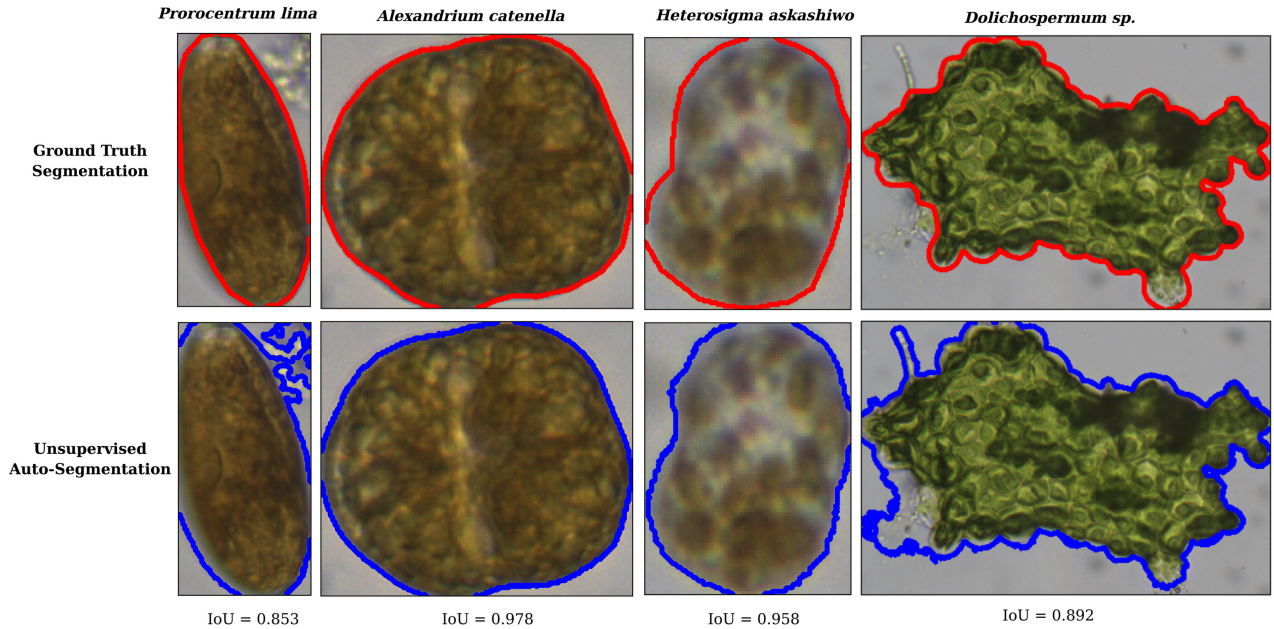


Fig. 2: The four species that imaged to evaluate our proposed method were *Prorocentrum lima*, *Alexandrium catenella*, *Heterosigma askashiwo*, and *Dolichospermum sp.* For each of these species the original microscopy image can be seen, along with the both the ground truth segmentation (red) and unsupervised auto-segmentation (blue). The intersection over union (IoU) metric, listed in Table 1 was used to compare these two masks.

2.3 Contributions

Given the importance of these weak annotations in reducing human labeling efforts and use in weak learning, the first and most critical step in evaluating this approach is to evaluate the quality of these noisy labels, as seen in Figure 1 (bottom). Therefore, in this paper we generate noisy annotations using unsupervised segmentation and evaluate the quality of these noisy labels by comparing them to strong annotations created by a team of biologists that specialize in phytoplankton taxonomy.

3 Experimental Setup

3.1 Dataset

The dataset used in this research consists of marine cultures of four distinct phytoplankton species, including *Alexandrium catenella*, *Dolichospermum sp.*, *Heterosigma askashiwo*, and *Prorocentrum lima*. The size and morphology of the four species chosen varied from 5-50 μm in diameter; oval, round, single-celled, and colonial. They were considered to be indicative of their natural morphologies found within the environment, with the exception of *Dolichospermum sp.* which deviated significantly from its natural form. Initially, the dataset had 636 microscope specimen images at a resolution of either 5320x3032 pixels or 3208x2200 pixels and 2247 individual annotations. Annotations include ground truth segmentation masks with their area, bounding boxes, and species labels. A total of 636 images were further manipulated by cropping them with their corresponding segmentation masks' bounding boxes. Total ground truth segmentation masks for *Alexandrium catenella*, *Dolichospermum sp.*, *Heterosigma askashiwo*, and *Prorocentrum lima* were 271, 861, 767, and 348, respectively. As a result, the final dataset had 2247 distinct images and 2247 corresponding masks.

3.2 Unsupervised Segmentation

Image Pre-processing: For each of the four plankton species, a normalization factor was imperially derived by observing the downstream unsupervised segmentation and post-processing tasks.

K-means Clustering:

Unsupervised segmentation was done using K-means clustering in our experiment. K-means clustering is an unsupervised machine learning algorithm used to divide the data into k clusters. By cropping the images according to the bounding boxes of an instance of an organism, we ensure that only two possible clusters

Table 1: The mean and variance of the intersection over union (IoU), which ranges from zero to one, of the four species. Overall *Prorocentrum lima* performed best, followed by *Alexandrium catenella*, *Heterosigma askashiwo*, and *Dolichospermum sp.*, respectively.

Species	IOU Average	IOU Variance
<i>Prorocentrum lima</i>	0.7619	0.1037
<i>Alexandrium catenella</i>	0.7191	0.0986
<i>Heterosigma askashiwo</i>	0.6978	0.1125
<i>Dolichospermum sp.</i>	0.6495	0.0839

would be formed, hence k is chosen as two. The algorithm assigns the data points to one of the k centroids based on a metric. A commonly used metric is the Euclidean distance between the data point and the centroid. For some point, (x, y) , let $p(x, y)$ be the input pixels, and c_k be the cluster centers the euclidean distance is $d = \|(x, y) - c_k\|$.

Once the data points have been assigned to a centroid, the centroid is re-calculated based on the clusters created:

$$c_k = \frac{1}{k} \sum_{y \in c_k} \sum_{x \in c_k} (x, y) \quad (1)$$

This is followed by another re-calculation of the evaluation metric between the new centroids and the data points. The data points are assigned to the centroids in a manner that optimizes the evaluation metric [14].

Image Post-processing: Each binary mask was post-processed by only retaining the largest connected component and filling any holes of that largest connected component.

3.3 Performance Metric

The Intersection over Union (IoU) is a method of comparing the similarity of two shapes, $A, B \subseteq \mathbb{U}$ calculated by [15]:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

We use this metric to compare the similarity of the algorithmically generated cell annotations relative to the ground truth annotations created manually by a team of biologists. The range of this metric is bound between zero and one.

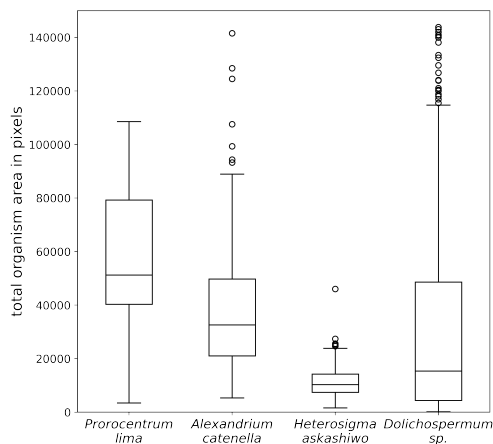


Fig. 3: Box plots of total ground truth segmentation mask area in pixels per species. Bottom, medium, top line of the box, and circles represent 3rd quartile, median, 1st quartile, and outliers, respectively. *Dolichospermum sp.* has the highest variability with significant amount of outliers, and as observed in Table 1 has the lowest overall IoU score.

4 Results & Discussion

Table 1 shows the average IoU as well as the variance IoU between the ground truth segmentation masks and masks created by our unsupervised segmentation algorithm. *Prorocentrum lima* has the highest average IoU of 0.7619 ± 0.1037 , while *Alexandrium catenella* and *Heterosigma askashiwo* show similar performance with average IoUs of 0.7191 ± 0.0986 and 0.6978 ± 0.1125 respectively, and *Dolichospermum sp.* has the lowest average IoU of 0.6495 ± 0.0839 .

Figure 2 shows the original microscope image with both the the ground truth segmentation and the unsupervised segmentation for all four organisms. Overall, the outcome of our proposed segmentation matches the ground truth segmentation, as seen by the IoU metrics of 0.978, 0.958, 0.892, 0.853 for *Alexandrium catenella*, *Heterosigma askashiwo*, *Dolichospermum sp.* and *Prorocentrum lima* respectively. The masks produced for *Prorocentrum lima* and *Dolichospermum sp.* include debris and other surrounding detritus. As seen in the prediction masks, our approach includes this detritus in it's mask, while the ground truth segmentation does not.

Figure 3 shows box plots for the total area (in pixels) of the four from the the ground truth segmentations. From Table 1, recall that *Dolichospermum sp.* had the lowest performance. Here, in Figure 3, *Dolichospermum sp.* is observed to have higher variability compared to the other three species, with an extreme number of outliers (circles). Therefore the lower IoU performance of *Dolichospermum sp.*, as see in Table 1, in comparison to other species, is likely a consequence of this high variance in organisms size, as it is difficult to generalize the unsupervised segmentation algorithm.

5 Conclusion & Future Work

Harmful algal blooms (HABs) are damaging environmental ecosystems and causing significant losses for a variety of industries such as drinking water and marine aquaculture. Automated image analysis via deep learning is a promising approach to improve monitoring of these HABs, which in turn will allow better management decisions to be made. However, the main drawback of supervised deep learning is that it requires a large amount of labeled data, which is time consuming and tedious to produce, especially when polygon segmentations are required.

In this paper we showed that noisy prelabels can be automatically generated using unsupervised segmentation for four common marine phytoplankton species. These noisy prelabels were compared to the ground truth segmentation masks created by a team of phytoplankton taxonomists. Given these encouraging results, future work involves (1) exploring other unsupervised segmentation approaches such as contour-based segmentation and edge-based segmentation, (2) quantifying the time saved for the annotation team, and (3) using these prelabels in a weakly supervised manner to create a pretrained model, which can be used in the downstream task of supervised learning with strong labels.

Acknowledgments

This work was funded by the Waterloo AI Institute, Mitacs, and Blue Lion Labs. The dataset was provided by Blue Lion Labs. The computing resources were provided by the Vision and Image Processing (VIP) Lab at the University of Waterloo and Blue Lion Labs.

References

- [1] R. M. Montes, X. Rojas, P. Artacho, A. Tello, and R. A. Quinones, "Quantifying harmful algal bloom thresholds for farmed salmon in southern Chile," *Harmful Algae*, vol. 77, pp. 55–65, 2018.
- [2] K. G. Sellner, G. J. Doucette, and G. J. Kirkpatrick, "Harmful algal blooms: causes, impacts and detection," *Journal of Industrial Microbiology and Biotechnology*, vol. 30, no. 7, pp. 383–406, 2003.
- [3] C. J. Gobler, "Climate change and harmful algal blooms: Insights and perspective," *Harmful Algae*, vol. 91, p. 101731, 2020, climate change and harmful algal blooms.
- [4] M. L. Wells, V. L. Trainer, T. J. Smayda, B. S. O. Karlson, C. G. Trick, R. M. Kudela, A. Ishikawa, S. Bernard, A. Wulff, D. M. Anderson, and W. P. Cochlan, "Harmful algal blooms and climate change: learning from the past and present to forecast the future," *Harmful Algae*, vol. 49, pp. 68–93, 2015.
- [5] G. Hallegraeff, "Harmful algal blooms: a global overview," *Manual on harmful marine microalgae*, vol. 33, pp. 1–22, 2003.
- [6] D. M. Anderson, A. D. Cembella, and G. M. Hallegraeff, "Progress in understanding harmful algal blooms: paradigm shifts and new technologies for research, monitoring, and management," *Annual Review of Marine Science*, pp. 143–176, 2012.
- [7] J. L. Deglint, C. Jin, and A. Wong, "Investigating the automatic classification of algae using the spectral and morphological characteristics via deep residual learning," in *International Conference on Image Analysis and Recognition*. Springer, 2019, pp. 269–280.
- [8] N. Bamra, V. Voleti, A. Wong, and J. Deglint, "Towards generating large synthetic phytoplankton datasets for efficient monitoring of harmful algal blooms," *arXiv preprint arXiv:2208.02332*, 2022.
- [9] T. Huxohl and F. Kummert, "Model-assisted labeling and self-training for label noise reduction in the detection of stains on images of laundry," *Mathematics*, vol. 9, no. 19, p. 2498, 2021.
- [10] X. Tang, H. Mou, J. Liu, and X. Du, "Research on automatic labeling of imbalanced texts of customer complaints based on text enhancement and layer-by-layer semantic matching," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.
- [11] M. Desmond, E. Duesterwald, K. Brimijoin, M. Brachman, and Q. Pan, "Semi-automated data labeling," in *NeurIPS 2020 Competition and Demonstration Track*. PMLR, 2021, pp. 156–169.
- [12] J. Robinson, S. Jegelka, and S. Sra, "Strength from weakness: Fast learning using weak supervision," in *International Conference on Machine Learning*. PMLR, 2020, pp. 8127–8136.
- [13] C. Sun, J. K. Udupa, Y. Tong, C. Wu, S. Guo, J. M. McDonough, D. A. Torigian, and R. M. Campbell, "Auto-labeling of respiratory time points in free-breathing thoracic dynamic mr image acquisitions for 4d image construction," in *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 10953. SPIE, 2019, pp. 277–283.
- [14] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image segmentation using k-means clustering algorithm and subtractive clustering algorithm," *Procedia Computer Science*, vol. 54, pp. 764–771, 2015.
- [15] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," 2019.