UNIVERSITY OF CALIFORNIA SAN DIEGO

Benchmarking the Automated Analysis of *In situ* Plankton Imaging and Recognition

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Electrical Engineering
(Machine Learning and Data Science)

by

Kevin T. Le

Committee in charge:

Nuno Vasconcelos, Chair
Jules S. Jaffe, Co-Chair
Kenneth Kreutz-Delgado

2021

The Thesis of Kevin T. Le is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

Chair

University of California San Diego

2021

EPIGRAPH

When I get knocked down, I'll get back up.

I may not be the smartest person in the room,

But I'll strive to be the grittiest.

Angela Duckworth

TABLE OF CONTENTS

Signature Page iii

Epigraph iv

Table of Contents v

List of Figures vii

List of Tables viii

Acknowledgements ix

Abstract of the Thesis xi

Chapter 1 Introduction 1

1.1 Motivation 1

1.2 Related Work 2

1.3 Thesis Overview 3

Chapter 2 Counting Methods and Data Acquisition 6

2.1 Data Acquisition 6

2.2 Traditional Microscopy Analysis: Lab-micro 6

2.3 Automated Imaging Systems: SPC-Pier and SPC-Lab 7

2.4 Species Selection and Manual Classification 8

2.5 Measurements 9

Chapter 3 Automated Imaging Classification 11

3.1 Convolutional Neural Networks 11

3.2 Network Training 11

3.3 Network Implementation 14

Chapter 4 Analyses 15

4.1 Classification Analyses 16

4.2 Counting Analyses 16

4.3 Volume Computation Analyses 17

Chapter 5 Results 18

5.1 Classification Performance 19

5.2 Comparison between Lab-micro and SPC+CNN counts 21

5.3 Volume Computation 25

5.4 Continuous Observation Data 25

Chapter 6 Discussion 27

Chapter 7 Conclusion 32

Appendix A Supplementary Information 33

A.1 Comparison of Lab-micro vs. SPC and SPC+CNN 33
References 40

LIST OF FIGURES

Figure 2.1: Imaging Systems. (a) SPC-Pier: SPC-MICRO Underwater Camera. (b) SPC-Lab: Benched laboratory configuration of SPC-MICRO. 8

Figure 2.2: Image examples taken by the SPC-Pier, SPC-Lab, and Lab-micro. Each row is organized by the SPC imaging system, while each column is the imaged species. We focus on studying the following 9 taxa (a - i): *Akashiwo sanguinea*, *Ceratium falcatiforme and fusus*, *Ceratium furca*, *Chattonella* spp., *Cochlodinium* spp., *Gyrodinium* spp., *Lingulodinium polyedra*, *Prorocentrum micans*, and *Pseudo-nitzschia* spp.. 9

Figure 2.3: Sampling Method Comparison Framework for Plankton Abundance. The diagram outlines where the samples are collected from their respective locations using 3 sampling methods: the Lab-micro, SPC-Lab, and SPC-Pier. Trained CNNs were then deployed onto the resulting data to produce CNN counts (SPC+CNN- Lab and Pier). 10

Figure 5.1: Species Enumerated. (a) Total count collected by each method time series. (b) Average count per day per species collected by each method. Throughout the collection, the species composition was dominated by abundant species, such as *Lingulodinium polyedra* and *Prorocentrum micans*, while some rarely appeared. 19

Figure 5.2: Quantification of the classification accuracy for SPC test sets. (a) Confusion Matrix. (b) Diagonal class accuracies of confusion matrix, sorted in a descending fashion from left to right 21

Figure 5.3: Pearson Correlation Coefficient Matrices. Each row is a compared setting, while each column is the corresponding species. Coefficient values are color coded with respect to the species correlation value of the compared setting, in an ascending fashion. (a) Correlation of Lab-micro vs. manually enumerated SPC counts. (b) Correlation of Lab-micro vs. SPC+CNN counts. 22

Figure 5.4: Relationships between counts of Lab-micro and SPC+CNN methods. Columns are the pair of counting methods, while the rows are organized by species. The solid line indicates the linearly fitted model and is coupled with multiple shaded areas indicating the 95% prediction (dark shade) and confidence interval (light shade). The slope and R^2 of the model are indicated in the panel 23

Figure 5.5: Time series of species abundances observed during a 3-month deployment via the SPC-pier and the SCOOS monitoring program during 2019. Automated image classification was used to produce counts on periods not sampled by the SCOOS program. Plots are shown for only the highly correlated abundant and rare species. 26

Figure A.1: Relationships between Lab-micro and SPC+CNN: Strongly linear-related species (*Akashiwo sanguinea, Cochlodinium* spp., *Lingulodinium polyedra, Prorocentrum micans*) 34

Figure A.2: Relationships between Lab-micro and SPC: Strongly linear-related species (*Akashiwo sanguinea, Cochlodinium* spp., *Lingulodinium polyedra, Prorocentrum micans*) 35

Figure A.3: Relationships between Lab-micro and SPC+CNN: Poorly linear-related species (*Ceratium falcatiforme or fusus, Ceratium furca, Chattonella* spp, *Pseudo-nitzschia* spp.) . . 37

Figure A.4: Relationships between Lab-micro and SPC: Poorly linear-related species (*Ceratium falcatiforme or fusus, Ceratium furca, Chattonella* spp, *Pseudo-nitzschia* spp.) 38

LIST OF TABLES

Table 3.1: Training, validation and test datasets to train SPC+CNN. Datasets are organized by the fine-tuning stage of the SPC+CNN because of the double fine tuning to adapt each network to the SPC-Pier and SPC-Lab after learning more general representations of plankton. 12

Table 5.1: Average classification results of a double fine-tuned model tested on independent, held-out samples collected by the SPC-Pier and SPC-Lab. Evaluation metrics used are ACC, MCA, and F1 Score. 19

Table 5.2: Calibrated SPC+CNN-Pier Sampling Volume Per Image. This calibration assumes a sampled Lab-micro volume of 1.76 mL and 2000 seconds of images at 8 Hz that were collected and classified for each species by the SPC+CNN-Pier. 25

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my co-advisors Nuno Vasconcelos and Jules S. Jaffe for their constant support and guidance throughout all these years. I am deeply indebted to them for the many opportunities they have provided me with since my undergraduate years and for their commitment to my growth in scientific research and communication. Their individual passion for science and teamwork is admiring, and I hope that I can one day be as inspirational as they have been to me. I would also like to thank Ken Kreutz-Delgado for serving on my committee. I hope that I can one day help advance the field of machine learning just as they all are doing right now. I am also thankful for all my past and present labmates in the SVCL and Jaffe Lab who have always been supportive of me, including Gautam Nain, Devin Ratelle, Pichaya Lertvilai, Joe Walker, and Paul Roberts. I am specifically thankful for my mentors Pedro Morgado and Eric Orenstein and their continuous guidance and patience ever since they took me under their wing. I attribute my growth as a researcher to their mentorship, and I could not have asked for a better mentor than them. I am also grateful for the dedicated efforts of my lab interns, Zhouyuan Yuan and Areeb Syed, as they made this project happen just as much as I did. It was my first experience of being a leader within a research setting when mentoring them, so I'm happy they stuck it out in the end.

I am also thankful for my parents, sister Tina Le, cousin Duy Tran and last but not least, girlfriend Jeanette Nguyen, for their unconditional love and unwavering support in pursuing my dreams. They have always been there for me during both the highs and the lows in my life, and my appreciation for them is something that words cannot express. My best friends Alvin Pham, Holly Zargham, and Brandon Nguyen also deserve my fullest love and gratitude for their amazing friendship. I am also in debt to my housemate, Marco Uytiepo, as I couldn't have had a truer friend throughout my MS. Finally, I am thankful for my BudEEs Billy Zeng, Jeanette Nguyen, Kenny Chen, Sam Vineyard, and Thomas An for their companionship and for all of our crazy times both in and out of the classroom. My success in the ECE M.S. program at UC San Diego would not have been possible without these people's love and support these past three years.

The chapters of this thesis consist of material currently being prepared for submission. The thesis author was the author of this material.

- Chapter 1-7 has material currently being prepared for submission as it may appear in K. Le, Z. Yuan, A. Syed, E. Orenstein, D. Ratelle, M. Carter, S. Strang, K. Kenitz, P. Morgado, P. Franks, N. Vasconcelos, and J. Jaffe, "Benchmarking the Automated Analysis of *In situ* Plankton Imaging and Recognition," *Methods in Ecol Evol*, 2021, in preparation.

ABSTRACT OF THE THESIS

Benchmarking the Automated Analysis of *In situ* Plankton Imaging and Recognition

by

Kevin T. Le

Master of Science in Electrical Engineering

(Machine Learning and Data Science)

University of California San Diego, 2021

Nuno Vasconcelos, Chair

Jules S. Jaffe, Co-Chair

To understand ocean health, it is crucial to carefully monitor and analyze marine plankton – the microorganisms that form the base of the marine food web and are responsible for the uptake of atmospheric carbon. With the recent development of *in situ* microscopes, that collect images of these organisms in vast quantities, the use of deep learning methods to taxonomically identify them has come to the forefront. Given this data, two questions arise: 1) How well do deep learning methods such as Convolutional Neural Networks (CNNs) identify these marine organisms using data from *in situ* microscopes? 2) How well do CNN derived estimates of abundance agree with established net- and bottle-based sampling?

Using images collected by the *in situ* Scripps Plankton Camera (SPC) system, we trained a CNN to recognize 10 species of phytoplankton that are identified as associated with Harmful Algal Blooms (HABs). The success of the CNNs is characterized using standard evaluation metrics. To compare abundance estimates, we fit a linear model between the number of organisms of each species counted in a known volume in the lab with the number of organisms collected by the *in situ* microscope sampling at the same time.

The CNNs evaluated on 26 independent natural samples collected at Scripps Pier achieved an averaged accuracy of 92%, with 7 of 10 target categories above 85%. The linear fit between lab and *in situ* counts of several key HAB species suggests in the case of these dinoflagellates there is good correspondence between the two methods. The linear relationship derived for other organisms that were not as abundant is not as conclusive as well as the failure of the SPC systems to successfully detect the diatom chain, *Pseudo-nitzschia*.

Given the excellent correlation between lab counts and *in situ* microscope counts on key species, the methodology proposed here provides a way to estimate an equivalent volume in which the employed microscope can identify in focus organisms and obtain reasonably confident estimates of abundance. Given the ease of collection and success of the method, we are hopeful that future systems can automatically monitor HAB formers and other plankton.

Chapter 1

Introduction

1.1 Motivation

Plankton are an extremely diverse group of underwater organisms with a profound effect on ocean health [1]: they form the foundation of the food web, contribute to the early developmental stages of commercially harvestable species and their abundance and composition are tightly related to hydro-climatic change [2]. Planktonic organisms can also adversely affect the marine ecosystem by forming dense toxic blooms, known as Harmful Algal Blooms (HABs), that can cause disease, or parasitize other organisms. The appearance and composition of these HABs is a topic of intense research since they have deleterious effects on human health, negatively affect fish stocks, and are linked to eutrophication that is likely to increase in the coming years [3]. These biological impacts have serious economic ramifications and there is urgent interest in developing inexpensive, automated ways to detect increased HAB abundance [4–7]. The main goal of this study is to examine the potential for *in situ* imaging microscopy, supported by automated deep learning algorithms, to provide reliable estimates of HAB species to make progress towards this goal.

1.2 Related Work

Most HAB monitoring programs use traditional plankton sampling techniques, such as net tows and bottle sampling [8]. These approaches require physically collecting the samples, chemically preserving the organisms, and manually enumerating species with a lab microscope. This laborious process is severely limited by a number of factors: net tows can damage delicate organisms during collection [9, 10]; certain organisms may dissolve in the preservation solution without proper treatment [11]; and critically, physical collection and subsequent analysis of the samples is expensive in terms of cost and human labor, resulting in less frequent sampling than is desirable.

Due to these limitations, there is an increasing interest in the use of imaging systems to monitor HABs. These systems have the capability to quantify organisms at very fine spatial and temporal resolution, therefore providing a more scalable solution for long term analysis [2, 12–16]. Imaging systems use underwater microscopes to continuously take images of plankton as they either freely flow through the camera's view [17, 18] or are transported via microfluidic system [19]. These systems do not require physically collecting or concentrating water, chemical treatment of samples, or the use of counting chambers. However, the major bottleneck for using *in situ* imaging instruments for monitoring is the sheer volume of data they collect. To speed analysis, scientists have begun using automated classification methods, such as Support Vector Machines and Convolutional Neural Networks (CNNs) that are capable of processing these large imaging libraries [20–22]. The results indicate that CNNs can be used successfully for the classification of marine organisms such as zooplankton, phytoplankton, coral, and fish [23–25].

Although the utilization of automated imaging and recognition systems for estimating plankton abundance promises to expand *in situ* observational capacity, the methodology has yet to be widely adopted for both scientific studies and monitoring programs. Several recent studies have been dedicated to comparing submerged instruments against traditional lab counting methods, but a critical difference is that the image data was manual, not automatically classified. Whitmore et al. [26] explicitly compared the Zooglider's

abundance estimates against MOCNESS net tows and acoustic data. Likewise, Sosik and Olson, [12] compared manual counts from the IFCB images to manual bench top counts.

Conversely, other related studies focused on validating the automated estimation of plankton abundance, but did not seek to compare the results to traditional methods. Gonzalez et al. [27] suggested that an automated classifier's performance can be improved by attempting to match the training set class distribution to the eventual target population. Orenstein and Beijbom, [22] describes the workflow for leveraging out-of-domain images to improve the CNN's capability to classify plankton images. Gonzalez et al. [23] proposed a number of automated quantification algorithms to improve plankton abundance estimates. Orenstein et al. [28] proposed similar methods to reduce human annotators' validation labor while reliably reproducing plankton distributions. However, there still remains a gap in demonstrating how automated workflows that employ imaging instruments paired with trained CNN classifiers agree with plankton population estimates that use the more traditional lab counting methods.

1.3 Thesis Overview

In this article, we quantify the relationship between plankton population estimates derived from an *in situ* imaging system, the Scripps Plankton Camera (SPC), with those obtained from bottle-based sampling from surface samples followed by manual enumeration by a trained taxonomist. The SPC system consists of two underwater microscopes that image undisturbed volumes of water that can freely flow between a light source and a camera system. It has been operating nearly continuously for 5 years, resulting in the two cameras collecting more than a billion images of plankton, detritus, sand, and other objects. Using data from the SPC microscopes, CNNs have been trained to sort the resulting data and speed up ecological analyses [17,22,28,29]. The Scripps Pier is also a sampling location for the on-going Southern California Coastal Ocean Observing System (SCCOOS) HABMAP monitoring program (Kim et al., 2009) that has been enumerating HAB formers from weekly water samples since 2008. The methodology employs hand-drawn

water samples and a modern variant of the Utermöhl method to count a variety of plankton and estimate the abundance of HAB formers [30–32]. Here, we reference those lab-based abundance estimates as the "gold standard" that provides a baseline for comparing our automated methods that are based on automatically classified SPC data. If successful, the automated analysis workflow would provide an efficient, continuous monitoring system to detect and monitor phytoplankton and provide real-time, detailed, and reliable HAB warnings. The detection capability efficacy of both the imaging system itself and automated classification is evaluated in this study.

Our study compares the consistency of the automated workflow for plankton count estimates obtained via CNN classification of the SPC images (SPC+CNN-Pier) against those derived by a plankton taxonomist counting hand-drawn, preserved samples under a microscope in the lab (Lab-micro). As a bridge between the two methods, a subsample of the hand-drawn bottle samples was imaged by a benchtop version of the SPC (SPC-Lab) and classified with an identically trained CNN (SPC+CNN-Lab). A substantial issue here in using the SPC data to estimate abundance is to quantify the "effective" imaging volume of the SPC systems. The complication arises as they employ a dark field method of illumination [17] that we have found to produce optimal contrast to aid in identification. Unfortunately, this produces a range and focus dependent set of images from an unknown, and likely species-dependent volume, given the various morphologies and contrast inherent in the plankton images. To address this issue, we decided to use the raw counts of the various species as computed by the CNNs over a given time interval in comparison to the raw counts obtained by the Lab-micro system. Here, a simple linear model was used to relate the output of each method that then allowed us to determine an "effective" sample volume for the SPC. Using these species-specific volumes, we constrain the concentration estimates to be consistent between the three imaging modalities, SPC-CNN-Lab, SPC-CNN-Pier and Lab-micro. Details of the implementation and the calculation of the species-dependent effective volume for the two most abundant species are described in this article.

The remainder of this thesis consists of five chapters, and is organized as follows:

In Chapter 2, we introduce our counting methods for measuring plankton abundances. We outline how our deployment was taken and how each method operates. We also detail which species are selected for the count comparisons.

In Chapter 3, we provide a brief overview on automated imaging classification. Specifically, we go over CNN network training and implementation.

In Chapter 4, we present our analyses framework for evaluating classification performance, count comparisons, and volume computation.

In Chapter 5, we present analysis on the previously mentioned items. Each item is intended to convince the reader that the SPC+CNN can generate population estimates comparable to those produced via established benchtop methods.

In Chapter 6, we summarize the contributions made in this thesis and discuss the trade-offs between *in situ* imaging microscopy paired with CNNs and traditional sampling techniques.

In Chapter 7, we present our conclusion to this thesis. We also consider several potential directions of future work and refer the reader to the appendix for supplementary information.

Chapter 2

Counting Methods and Data Acquisition

2.1 Data Acquisition

Data for this study were obtained from three methods: traditional, lab-based microscopy from collected samples, *in situ* automated imagery, and lab-based automated imagery. All samples were collected from the end of the Ellen Browning Scripps Memorial Pier in La Jolla, CA ($32°52.02'N, 117°15.300'W$) from May through October 2019. Hand drawn samples were collected by SCCOOS personnel twice per week in the morning by lowering a 2 L bucket five times to collect 10 L of water at a depth of approximately 0.5 m. 2 L was then allocated for the traditional microscopy studies with the remaining 8 L imaged by the benchtop version of the SPC. The three sampling methods were then compared on 26 days during this time interval.

2.2 Traditional Microscopy Analysis: Lab-micro

Plankton were enumerated using the Utermöhl method for quantitative phytoplankton analysis via the routine monitoring program carried out by SCCOOS. This method will be referred to as Lab-micro throughout this paper. Seawater was concentrated in sedimentation chambers after being fixed in a 4% formaldehyde solution prior to manual counting. Once the sample settles, the upper chamber is removed

and replaced with a covered with a glass cover slip and placed under an inverted microscope. Cells are then classified to the lowest possible taxonomic level at 200x magnification and counted by a human operator [30–32]. SCCOOS technicians typically examine the organisms from settling 10 or 50 mL of seawater. However, the sample volume enumerated generally ranges from 1.25 mL to 12 mL based on the abundance of phytoplankton. Thus, for the purposes of this study, all of the volumes were scaled to a consistent volume of 1.76 mL volume. Although SCCOOS monitors additional species, we focus on the following 9 taxa: *Akashiwo sanguinea*, *Ceratium falcatiforme or fusus*, *Ceratium furca*, *Chattonella* spp., *Cochlodinium* spp., *Gyrodinium* spp., *Lingulodinium polyedra*, *Prorocentrum micans*, and *Pseudo-nitzschia* spp. Here, the presence of each species is reported in absolute counts from the calibrated sample volume.

2.3 Automated Imaging Systems: SPC-Pier and SPC-Lab

The SPC system is a set of two *in situ* underwater microscopes that are designed to observe undisturbed plankton populations [17]. The system does not employ filters, pumps, or nets; objects are imaged as they drift through the field of view. An onboard embedded computer segments the raw frames, saving each foreground object as individual Regions of Interest (ROIs). To restrict our interest to species that are less than 60 μm in body size, we used the higher resolution SPC-MICRO that has a 5x objective and employs dark field illumination to achieve a 40% contrast transmittance at a resolution of 5.0 μm. Pixel size in the image plane is .74 μm over a 2.5 mm x 2.0 mm field of view. We considered this to be adequate to observe organisms that are 10's of microns in size without sacrificing too much "in focus volume" that drastically decreases as the resolution is increased.

Two versions of the SPC-MICRO were used in this work: the SPC-Pier system, installed *in situ* at the Scripps Pier and the SPC-Lab system, a lab based version for benchtop imaging. The SPC-Pier system was moored at a tidally dependent average depth of 3 meters (Figure 2.1a). The SPC-Pier collected data at a rate of 8 frames per second throughout the study period of May to October 2019, with a brief pause

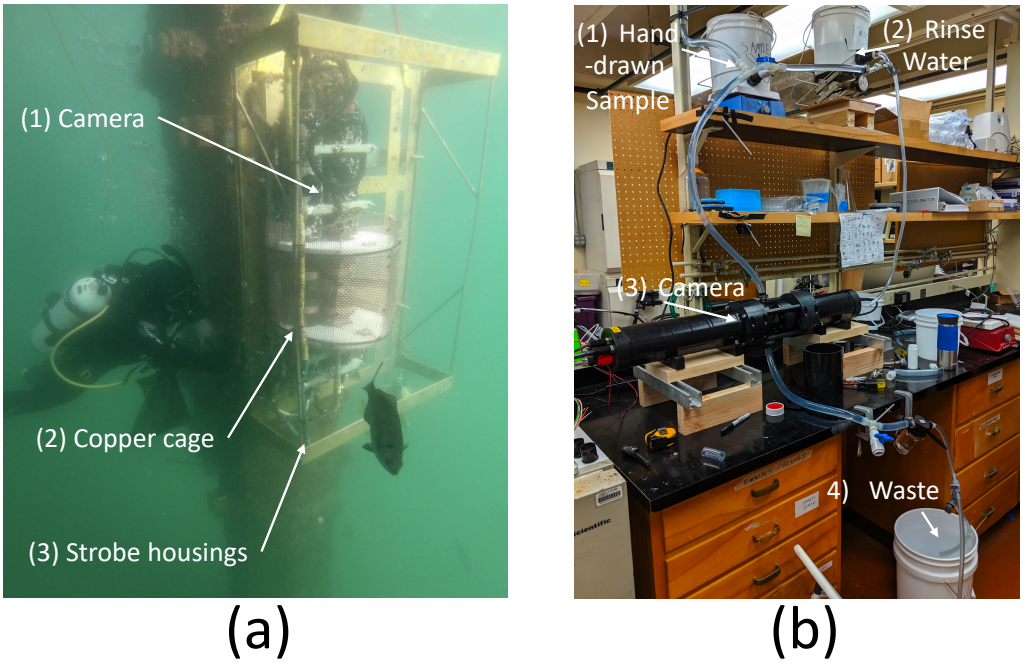

Figure 2.1: Imaging Systems. (a) SPC-Pier: SPC-MICRO Underwater Camera. (b) SPC-Lab: Benched laboratory configuration of SPC-MICRO.

in September due to heavy biofouling. In order to enumerate "counts" an arbitrary temporal window of $\pm$ 1000 seconds was chosen for evaluation that was centered around the exact time of the hand sampling.

The SPC-Lab is a reconfigured benchtop version of the SPC-Pier. To support the imaging of hand drawn samples, it was augmented with a gravity flow water system so that samples collected by SCCOOS operators could be passed through a clear acrylic chamber positioned in the field of view of the system (Figure 2.1b). This provided counts from the 8 L of water whose flow rate was kept constant by routinely replenishing the elevated water bucket with more seawater to maintain a minimum of 2 L of fluid. The sample chamber was flushed with filtered seawater between each run of seawater.

2.4 Species Selection and Manual Classification

In order to form a basis for comparing the observed image counts from the two SPC systems with those of the Lab-micro, a team of 3 taxonomists sorted all images collected by both SPCs into 10 classes

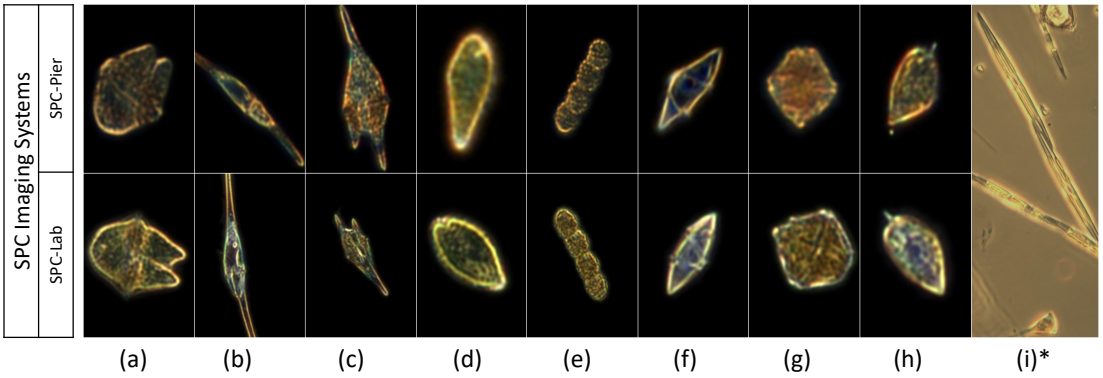

Figure 2.2: Image examples taken by the SPC-Pier, SPC-Lab, and Lab-micro. Each row is organized by the SPC imaging system, while each column is the imaged species. We focus on studying the following 9 taxa (a - i): *Akashiwo sanguinea*, *Ceratium falcatiforme and fusus*, *Ceratium furca*, *Chattonella* spp., *Cochlodinium* spp., *Gyrodinium* spp., *Lingulodinium polyedra*, *Prorocentrum micans*, and *Pseudo-nitzschia* spp..

where 9 were considered to be the target organisms shown in Figure 2.2 and where a tenth was classified as other that was used for the comparative count analyses. The other class is necessary to give the automated classifiers a place to put ambiguous objects and avoid high false-positive rates [33]. We note that the annotators only examined ROIs that had a projected major axis length between 30μm to 1000 μm as smaller objects are not well resolved and larger ones were of extremely low counts.

2.5 Measurements

As illustrated in Figure 2.3, the entire data set consists of measuring plankton "counts" using 5 methods. The first method consists of SCCOOS's traditional microscopy counts (Lab-micro). The next two use the SPC-lab system. In one case, the images are manually classified (SPC-Lab). In the other, they are automatically classified by a CNN (SPC+CNN) trained as discussed below. The remaining two measurements use in situ SPC data (SPC-Pier). Two settings are again considered: manual (SPC-Pier) and CNN-based (SPC+CNN-Pier) counts. For both SPC-Lab and SPC-Pier, the comparison to the automatic counts (SPC+CNN-Lab and SPC+CNN-Pier, respectively) enables a quantification of human vs. classifier performance on the same images. These comparisons also serve as a baseline for understanding the similar-

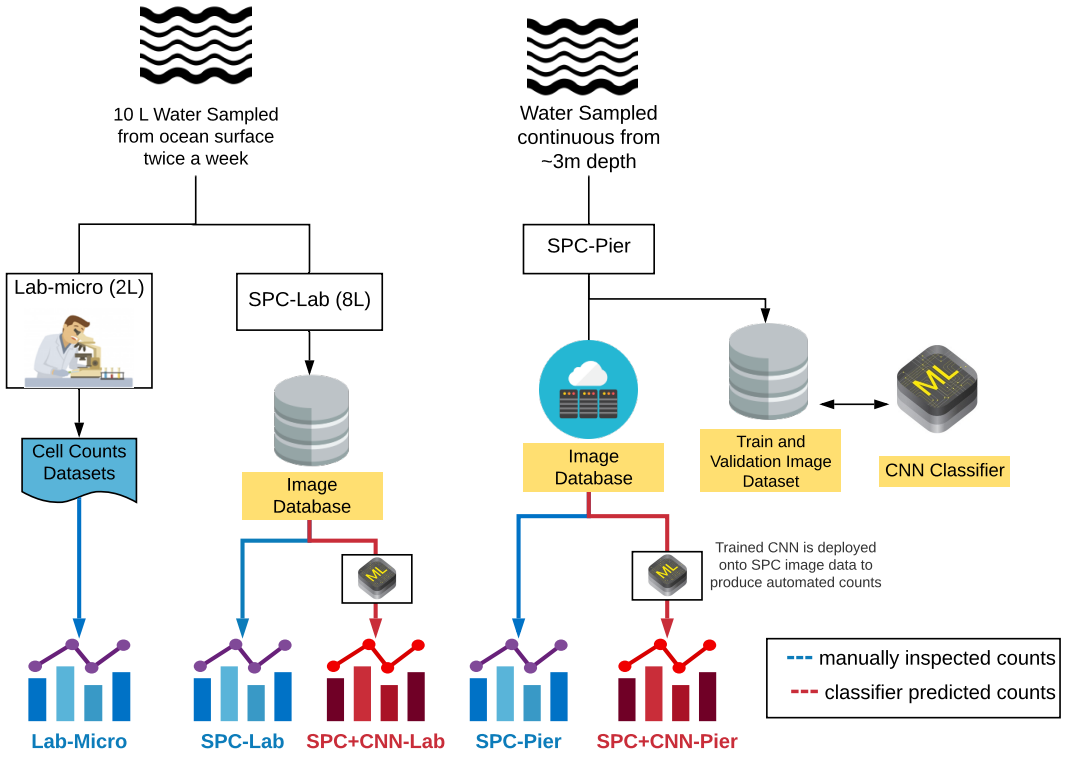

Figure 2.3: Sampling Method Comparison Framework for Plankton Abundance. The diagram outlines where the samples are collected from their respective locations using 3 sampling methods: the Lab-micro, SPC-Lab, and SPC-Pier. Trained CNNs were then deployed onto the resulting data to produce CNN counts (SPC+CNN-Lab and Pier).

ities and differences between the classical lab analysis (Lab-micro), laboratory microscopy (SPC-Lab), and *in situ* microscopy (SPC-Pier).

Chapter 3

Automated Imaging Classification

3.1 Convolutional Neural Networks

To test the accuracy of automatically generated species counts, we trained a collection of convolutional neural networks (SPC+CNN) to test on SPC-Lab and SPC-Pier images. A CNN builds an image representation for classification through a sequence of non-linear operations, including convolutions, non-linear rectification units, and pooling. The parameters in and between each layer are learned from training data. The representation is hierarchical, in the sense that early layers detect low-level features (such as colors or oriented edges), while deeper layers learn more abstract, high-level features (such as flagella or horns of plankton). Once trained, CNNs can rapidly process large amounts of image data and have been shown to outperform other methods, such as ensemble or margin-based classifiers [34]. We use the ResNet-18 model, a shallow variant of the ResNet [35] architecture with a small parameter space, that is quick to train and less likely to overfit to the relatively small training sets we collected [36].

3.2 Network Training

The SPC+CNN was trained in two fine tuning stages: (i) We fine tuned a ResNet-18 pre-trained on the ImageNet database [37] with SPC phytoplankton images. (ii) The resulting network was again fine tuned

Table 3.1: Training, validation and test datasets to train SPC+CNN. Datasets are organized by the fine-tuning stage of the SPC+CNN because of the double fine tuning to adapt each network to the SPC-Pier and SPC-Lab after learning more general representations of plankton.

Dataset	Fine Tuning Stage	Data	Number of Classes	Number of Images
Phytoplankton-Train	1	Train	30	29,196
Phytoplankton-Val	1	Validation	30	7,750
SPC-Pier (n=25 dates)	2	Train	10	avg ∼39,000
SPC-Pier (n=1 date)	2	Train	10	778
SPC-Lab (n=1 date)	2	Train	10	745

on just the nine classes of interest from either SPC-Pier or SPC-Lab. Fine tuning repurposes the weights of a network trained for a particular task to a different target. The procedure reduces training time and improves accuracy when training with small datasets [38]. Double fine tuning further adapts each network to subtle differences between the SPC-Pier and SPC-Lab data after learning more general representations of plankton [22].

The first fine tuning step uses a labeled phytoplankton training set from the SPC-Pier system comprised of 37,147 images spanning 51 classes. This dataset was produced by 17 expert taxonomists from the US West Coast during a three-day workshop geared toward encouraging interactions between ecologists and engineers. The annotated ROIs in the workshop dataset came from an earlier portion of the SPC-Pier time series and has no temporal overlap with the data acquired in our experiment. Experts sorted the annotated ROIs into 45 taxonomic classes and 6 noise categories, which included the 9 HAB species of interest. The workshop dataset was minorly adjusted by combining categories of the same species tagged with semantic descriptors such as the number of cells (e.g. *Ceratium furca* pair vs. single) and eliminating categories with less than 300 images. This resulted in a total of 30 classes, 24 identifiable species and 6 noise categories. 80% of the 36,496 images were then randomly chosen for training (Phytoplankton-Train) and the remaining 20% used for validation (Phytoplankton-Val) (Table 3.1).

The second fine tuning step has two objectives: (i) force the network to recognize only the 9 species of interest and the background class Other in the SPC-Pier and SPC-Lab dataset; and (ii) account for dataset shift, the well-known property that classifiers can be sensitive to changes in the input data, both the appearance of the images and the relative distribution of the classes between training and testing [23, 28, 39, 40]. There are several examples of dataset shift between our training sets, notably the slight variations in illumination between images captured by the SPC-Pier and SPC-Lab systems (Fig 2.1). In addition, for the second fine tuning, the classifier is fine tuned to the collected SPC-Pier dataset, which was partitioned in a leave one-out cross-validation manner for training and testing. Specifically, the model is trained on data from all dates from the SPC-Pier except for one, which is used as a held-out test set (Table 3.1). The same procedure is repeated several times with each sampled date being used as a held-out set once, and performance metrics are averaged across all 26 days. The training sets for each cross-validation iteration contain approximately 39,000 images, and test sets respectively hold 745 and 778. The restriction to fine tuning to only the SPC-Pier image dataset is specifically meant to examine the potential effects of dataset shift when the classifier is deployed on a new target domain, in our case the SPC-Lab. Training on SPC-Pier and testing on SPC-Lab data is a proxy for the more general transfer of a classifier trained on an *in situ* imaging system to an *in vitro* imaging system.

To prevent overfitting, images were subject to random affine transformations – rotations and translations. Data augmentation enables the creation of additional training examples without the burden of additional data collection and is, thus, frequently used in CNN training. Prior to the random affine transformations, images are padded into a square image and resized into 224 x 224 pixels.

Throughout each phase of the training procedure, the cross-entropy loss was weighted inversely proportionally to the class distribution of the corresponding training dataset, to mitigate potential class imbalance problems [41]. Note, this also includes recomputing the weight of the loss for each cross-validated model training.

3.3 Network Implementation

In implementing the first stage, the base ResNet-18 model pretrained on ImageNet was fine tuned for 50 epochs on the 30-class phytoplankton taxonomy workshop dataset. Model weights that achieved the lowest loss on the validation set during the 50 epochs were utilized. In this stage, the model achieved an accuracy of 95.5% on the Phytoplankton-Train set and accuracy of 95.2% on the Phytoplankton-Val set. The second stage was initialized with the model weights learned in the first stage, where the final layer was replaced with a layer of 10 outputs (9 categories of interest plus *Other*). Fine tuning to the leave-one-out cross-validation training datasets was performed for an additional 50 epochs with model weight selection corresponding to the lowest training loss. This resulted in a collection of 26 trained models, where each model is tested on an independent date from the SPC-Pier and SPC-Lab dataset. Counts produced by the CNN on all independent test samples respective to the system refer to this collection of models (SPC+CNN). All models were trained with an initial learning rate of 0.001 and a batch size of 16 using Adam optimizer [42]. Models were trained on a NVIDIA Titan Xp GPU. Python code used to train and evaluate the models is available at https://github.com/hab-spc/hab-ml.

Chapter 4

Analyses

To compare the 3 sampling methods, we used the total number of counts for each species collected on each of the 26 independent days. Although relative abundance is a widely used measure of plankton distributions, we use raw counts of each species as the volume of interrogation of the SPC systems is unknown in detail and comparisons of relative abundance are corrupted by the numerical instability caused by frequent counts of 0 or 1. Additionally, Lab-micro counts were scaled to a consistent sampling volume of 1.76 mL due to the previously mentioned fluctuating sample volume sizes. This then facilitated the use of a linear model to compute the effective sampling volume of the SPC+CNN systems as considered in the following sections.

With the species-specific counts, we performed 1) An assessment of the classifier performance and 2) a comparison between the Lab-micro counts and SPC+CNN counts. A baseline comparison between the Lab-micro counts and manually enumerated SPC counts is also included to compare to the SPC+CNN counts. To relate the SPC-CNN counts of both the Pier and lab implementation, we explored the use of a linear model to relate those counts to those of the Lab-micro. The relationship is based on the hypothesis that a fraction of the organisms present, α, are detected and classified using the SPC system when coupled with the CNNs. Hence,

$$C_{SPC+CNN} = \alpha C_{Lab-micro} \tag{4.1}$$

that relates the SPC+CNN counts to the Lab-micro counts. The multiplicative scaling factor α takes into account the differences in contrast, aspect dependence, additional noise from out of focus objects (in the SPC+CNN) and any other exclusive *in situ* imaging factors. It was estimated by computing a linear regression between each pair of counting methods. A separate model is fit for each of the 9 species across all 3 pairs: SPC+CNN-Lab vs. Lab-micro; SPC+CNN-Pier vs. Lab-micro; and SPC+CNN-Pier vs. SPC+CNN-Lab.

4.1 Classification Analyses

Our collection of double fine-tuned classifiers is applied to the 26 held out test sets collected by SPC-Lab and SPC-Pier to evaluate performance. Thus, performance results from our CNN are averaged across the test sets. Classification performance is assessed by 1) Accuracy (ACC), the fraction of correct predictions, 2) Mean class accuracy (MCA), the average correct predictions over each individual class, and 3) the F1 score, a metric for scoring class-imbalanced problems. Together these metrics capture both model generalization ability and bias towards highly populated classes – ACC characterizes the overall classifier performance while MCA and F1 scores assess how well the system does on a per class basis. Significant differences between the three chosen metrics might indicate in the results that it favors common classes while underperforming for rare ones.

4.2 Counting Analyses

Counts are compared across the 3 sampling methods under both cases of using the manually enumerated SPC counts and automated SPC+CNN counts. Results of these comparisons are measured by computing the Pearson Correlation Coefficient, a measure of linear correlation between two variables, and the estimation of both factor α and R^2, a percentage of the variance explained by the model relative to the

total variance, from the computed linear regression between each pair of counting methods. In conjunction, these measurements express how related the counting methods are [43].

To estimate the factor α, we simply took the slope from the proportional equation 4.1 that results from comparing the SPC-CNNs to the Lab-micro counts. The model was fit with a linear-linear least squares estimator, assuming zero intercept. However, for display purposes, the data was transformed to a log scale

4.3 Volume Computation Analyses

An important aspect of our work is the resultant estimation "effective sampling volume" for each species that is a result of the focus dependent dark field imaging scheme employed by the microscope. Assuming that the concentration of the species that were counted by each method is the same, we can divide equation 4.1 by an effective volume on the right-hand side ($V_{SPC+CNN}$) and a true volume taken from the lab procedure on the left-hand side ($V_{Lab-micro}$) to equate the two. Equating first the two concentrations and then making the substitution, we obtain

$$\frac{C_{SPC+CNN}}{V_{SPC+CNN}} = \frac{C_{Lab-micro}}{V_{Lab-micro}} \tag{4.2}$$

$$\frac{\alpha C_{Lab-micro}}{V_{SPC+CNN}} = \frac{C_{Lab-micro}}{V_{Lab-micro}}$$

Which implies that:

$$V_{SPC+CNN} = \alpha V_{Lab-micro} \tag{4.3}$$

We note that this volume was normalized to be 1.76 mL for the Lab-micro counts and was integrated over the 2000 seconds of images taken at 8 Hz for the SPC-Pier. Of the possible abundant and rare species for the SPC+CNN volume computation, we selected only abundant species that were significantly correlated via the Pearson correlation with the Lab-micro observations, as opposed to rare species of frequently low counts.

Chapter 5

Results

Over the course of 5 months, 43 independent plankton samples were acquired via Lab-micro, SPC-Lab, and the SPC-Pier. After common quality assurance checking, 26 days were deemed suitable for analysis as several days were inconsistent in the sampling times across the 3 methods mainly due to events such as the heavy biofouling of the SPC-Pier or unaccounted manual sample collection and imaging. 21,211 images from the SPC-Lab and 20,148 images from the SPC-Pier were manually classified in 10 categories – 9 species of interest and all remaining images as the 10th category, called "other". The 26 independent samples from both datasets were largely dominated by the "other" category (83% of the SPC-Pier total and 92% of the SPC-Lab total). The resulting manual counts are denoted as SPC counts. CNN-produced counts on the same dataset are denoted SPC+CNN counts. Lab-micro counts were produced by the human annotator.

In general, Lab-micro collected more total counts of the 9 target species, over the set of images, than the SPC systems (Figure 5.1a). Averaged over all 26 independent samples, Lab-micro count data was predominantly composed of 3 common species: *Pseudo-nitzschia* spp., *Lingulodinium polyedra*, and *Prorocentrum micans* (Figure 5.1b). The latter two also dominated both SPCs' counts. However, in the case of the SPCs, the *Pseudo-nitzschia* spp. counts were notably less. Possible reasons for this are potentially due to either the inability of the dark field imaging to successfully image Pseudo-nitzschia spp. cells or whether

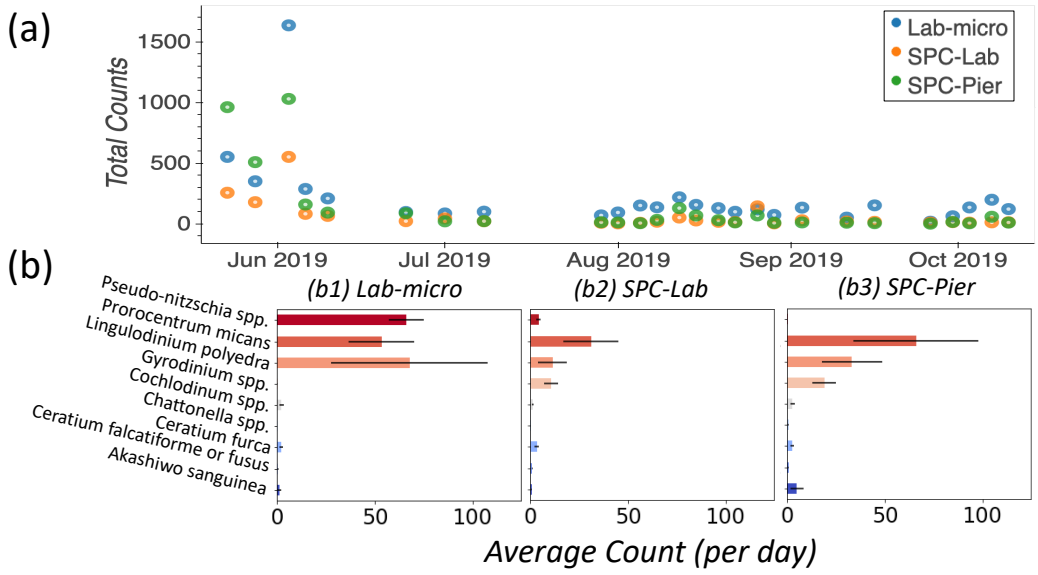

Figure 5.1: Species Enumerated. (a) Total count collected by each method time series. (b) Average count per day per species collected by each method. Throughout the collection, the species composition was dominated by abundant species, such as *Lingulodinium polyedra* and *Prorocentrum micans*, while some rarely appeared.

the projection of the organism from 3-dimensions into a 2-dimensional image created lower detection abilities. We do note that in Kenitz et al. [29] many diatoms were well detected. Rare dinoflagellate species (namely *Akashiwo sanguinea*, *Ceratium falcatiforme or fusus*, *Ceratium furca*, *Chattonella* spp., *Cochlodinium* spp., and *Gyrodinium* spp.), were more often observed by the SPCs than the Lab-micro suggesting the methodology has some taxonomic dependence.

5.1 Classification Performance

Table 5.1: Average classification results of a double fine-tuned model tested on independent, held-out samples collected by the SPC-Pier and SPC-Lab. Evaluation metrics used are ACC, MCA, and F1 Score.

Dataset	ACC	MCA	F1 Score
SPC-Lab	0.92	0.68	0.47
SPC-Pier	0.92	0.74	0.64

The averaged ACC, MCA, and F1 Score performance of a CNN trained on our data were mea-

sured from a CNN tested on independent samples from the 26 SPC-Pier and SPC-Lab image datasets. Our CNN achieved averaged test accuracies of 92% with high 95% on both the SPC-Lab and SPC-Pier data (Table 5.1). The MCAs were lower (68 and 74%) suggesting an unbalanced performance across classes. This discrepancy between the metrics is due to species that are relatively rare to the SPC (e.g. *Ceratium falcatiforme or fusus*, *Chattonella* spp., and *Pseudo-nitzschia* spp.) and would then result in less training data to effectively learn the species' morphology. The F1 scores were the lowest of the three (47 and 64%), due to the CNNs' frequent overestimation of the number of HAB species which is penalized in the F1 score for poor precision (Supplemental Figure 8a). Such class imbalance in the training dataset has a large effect on the learned model and is a well-established feature of training CNNs on natural populations.

Inspection of the confusion matrices for the CNN performance on the SPC-Pier versus SPC-Lab is shown in Figure 5.2a. As illustrated, the CNN performed significantly better on the SPC-Pier than on the SPC-Lab based on the MCA and F1 score difference. For half of the tested species (*Akashiwo sanguinea, Ceratium furca, Cochlodinium* spp., *Lingulodinium polyedra, Prorocentrum micans*) the accuracy dropped more than 10% from SPC+CNN-Pier to SPC+CNN-Lab, especially *Lingulodinium polydra* (Figure 5.2b). This is a manifestation of the domain shift between the SPC-Pier and SPC-Lab image methods, namely the laboratory conditions from using a flow-through system that appeared to result in more out-of-focus images that rendered the species differentiation more difficult. This was not unexpected, given that the model is only trained on pier data, however, it does illustrate that deployment of the same imaging system under settings may require fine-tuning of the classifier.

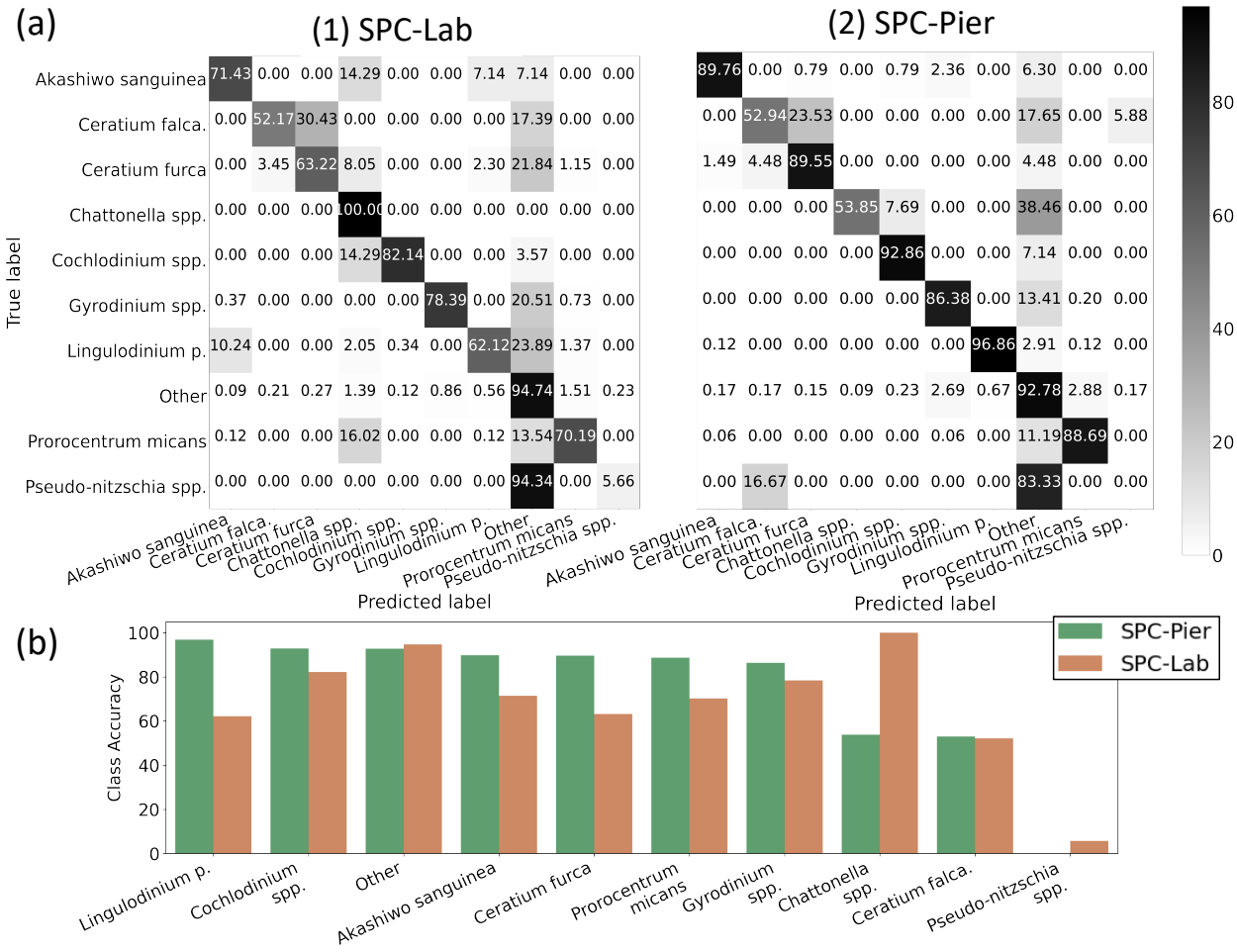

Figure 5.2: Quantification of the classification accuracy for SPC test sets. (a) Confusion Matrix. (b) Diagonal class accuracies of confusion matrix, sorted in a descending fashion from left to right

5.2 Comparison between Lab-micro and SPC+CNN counts

The Pearson correlation analysis on the intermediary comparison of the Lab-micro and manually enumerated SPC counts reveals high-to-very high correlations between the sampling methods on 4 out of the 9 species– *Akashiwo sanguinea, Cochlodinium* spp., *Lingulodnium polyedra*, and *Prorocentrum micans*– representing a mix of abundant and rare organisms (Figure 5.3a). The other 4 species; *Ceratium falcatiforme, C. fusus, Chattonella* spp., *Gyrodinium* spp., and *Pseudo-nitzschia* spp., demonstrated a pattern of scoring low correlation scores for two out of the three pairs. The last of the 9 species, *Ceratium furca*, was different of this trend, showing moderate correlation between both SPC methods and the Lab-

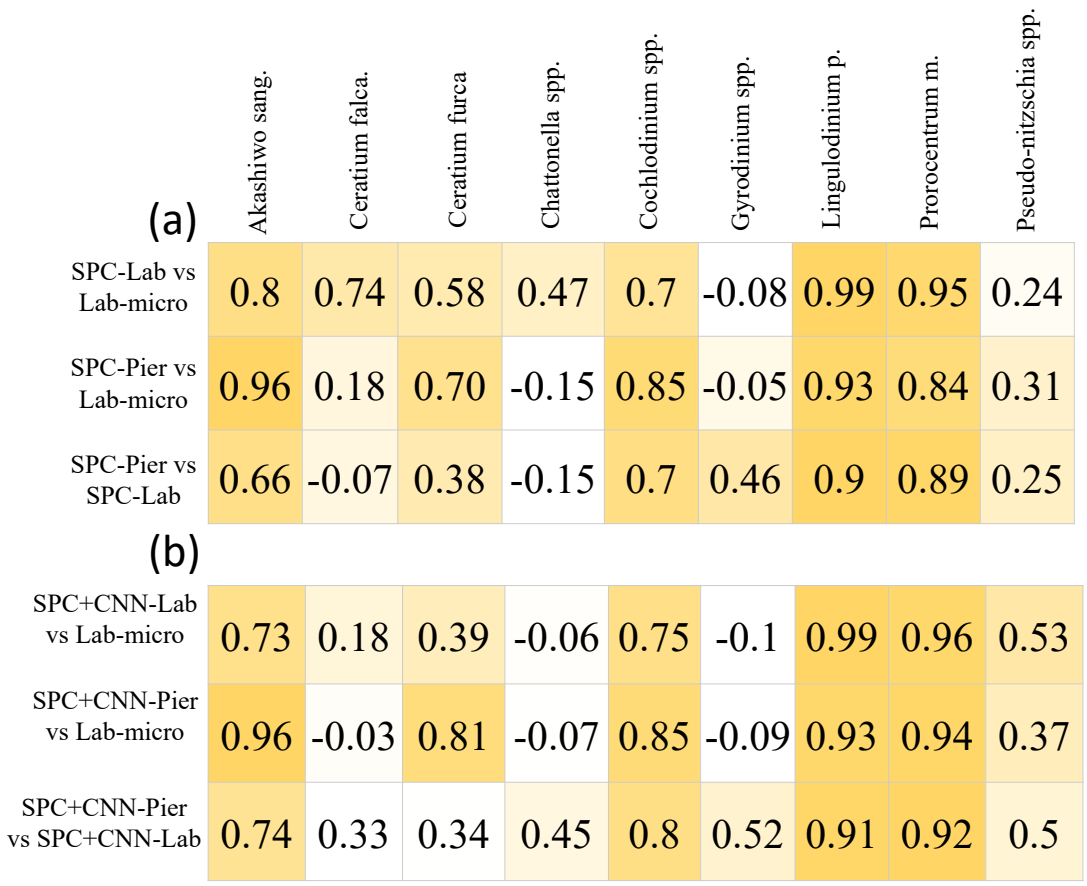

Figure 5.3: Pearson Correlation Coefficient Matrices. Each row is a compared setting, while each column is the corresponding species. Coefficient values are color coded with respect to the species correlation value of the compared setting, in an ascending fashion. (a) Correlation of Lab-micro vs. manually enumerated SPC counts. (b) Correlation of Lab-micro vs. SPC+CNN counts.

micro (0.58 and 0.70). We believe that it is likely these high variances are due to low counting number of the rarer species.

In general, the SPC+CNN vs. Lab-micro correlations produced similar results to the baseline correlation values between the manually enumerated SPC vs. Lab-micro counts (Figure 5.3b). The same 4 species that previously produced high-to-very high correlations were consistent when using SPC+CNN counts, with correlation value differences up to 10%. The correlation differences were due to the previously mentioned unbalanced performance across the classes from the SPC+CNN, that comes from using imbalanced training data. In the case of the SPC+CNN-Lab vs. Lab-micro, we can observe many correlation scores drop which

we can attribute to the domain shift problem. This is yet another demonstration of the need for network fine tuning in the domain where the network is applied.

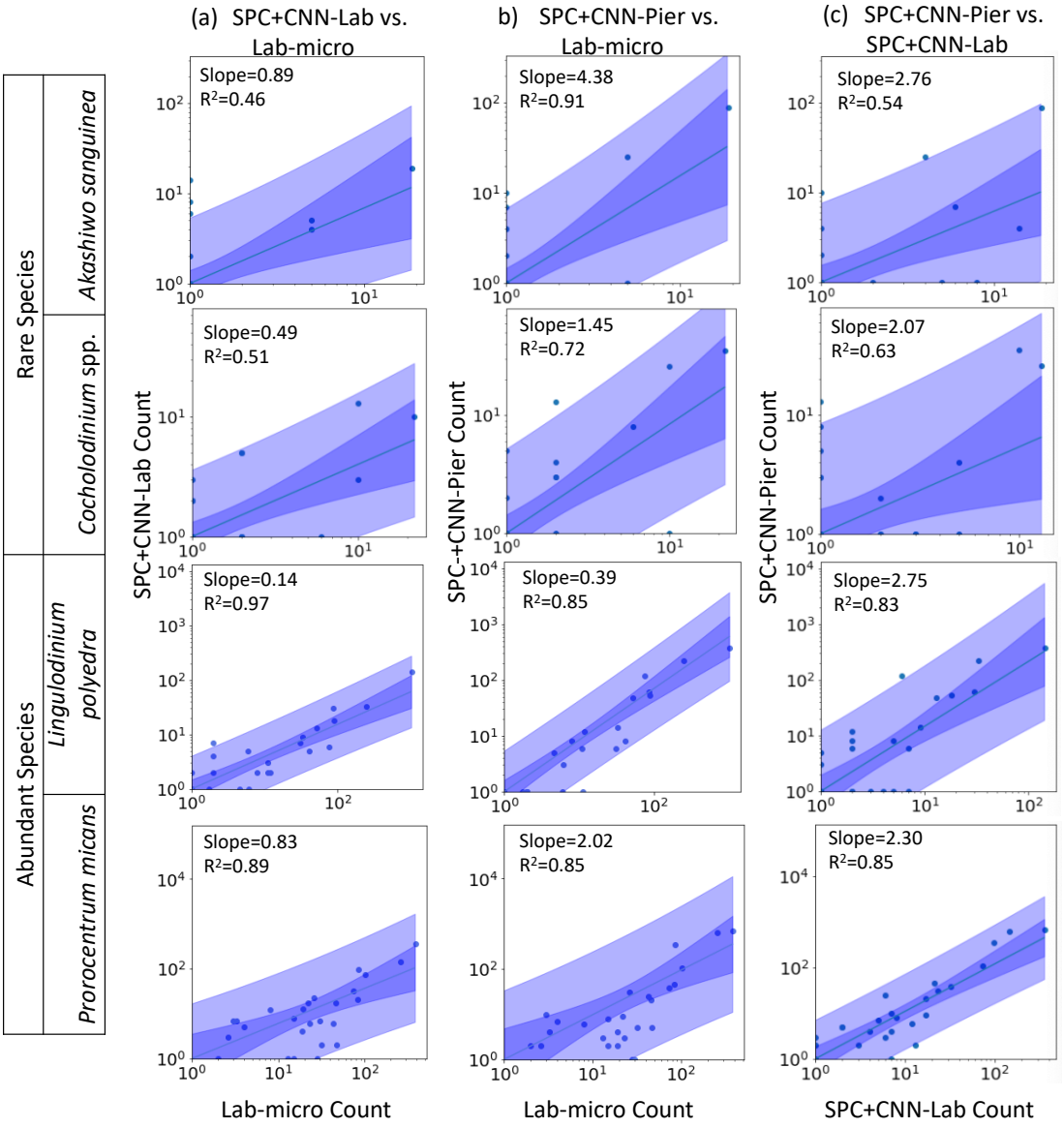

Figure 5.4: Relationships between counts of Lab-micro and SPC+CNN methods. Columns are the pair of counting methods, while the rows are organized by species. The solid line indicates the linearly fitted model and is coupled with multiple shaded areas indicating the 95% prediction (dark shade) and confidence interval (light shade). The slope and R^2 of the model are indicated in the panel

To further evaluate the relationship between counts of the Lab-micro and SPC+CNN, we computed a least squares regression between the species-specific counts acquired by each method. Figure 5.4 dis-

plays results of the linear fitting on a log-scaled graph for 4 of the species, while the other 5 species can be found in the Supplemental Figure A.3. Similar to the results of the Pearson correlation analysis for the pair of SPC+CNN-Pier vs. Lab-micro, we found that the less abundant species (*Akashiwo sanguinea, Cochlodinium* spp.) produced high R^2 values, as well as the more abundant species (*Lingulodinium polyedra, Prorocentrum micans*). We also observed a pattern of the size of the prediction and confidence bands being related to the occurring frequency of the species. The two abundant species showed much narrower prediction and confidence bands, in contrast to the two rare species, which exhibited wider bands. Discrepancy of the size of the bands could be due to the sparse signal that comes from sampling rare species. The comparability observations of the 4 species could also be validated by the linearly fitted models between these species counts of the Lab-micro and manually enumerated SPCs (Supplementals A.2), which only showed marginally different R^2 values.

In a general comparison of the linear fit computed by the model across the 3 possible pairs, we can observe the strongest linear relationship amongst the SPC+CNN-Pier and SPC+CNN-Lab. The results, specifically the proportionality approximation, convey that the SPC+CNN-Pier's sampling of an aggregate volume over the 2000 seconds was nearly 2x that of the SPC+CNN-Lab. Regarding the other 5 species, a majority of these species showed non-existing-to-poor linear relationships between the Lab-micro and SPC+CNN counts (Figure A.3). The linear fit for the *Pseudo-nitzschia* spp. showed little ability to model the relationship between the SPC+CNN and Lab-micro, as the SPC itself poorly detected the diatom chains. *Gyrodinium* spp. was mostly absent from the Lab-micro, preventing a linear regression between the compared counts. Species that had previously demonstrated low classification performance resulted in poorer relationships when computing the linear regression for the CNN-based pairs of counting methods. Comparatively to its manually enumerated-based linear regressions (Supplemental A.4), *Ceratium falcatiforme or fusus* and *Chattonella* spp. showed degraded R^2 values and slopes across all 3 possible pairs, suggesting the effect of the classification performance upon the linearly modeled relationship. *Ceratium furca* also showed

some fluctuations when comparing automated vs manual regressions, but generally showed only a moderate relationship.

5.3 Volume Computation

Table 5.2: Calibrated SPC+CNN-Pier Sampling Volume Per Image. This calibration assumes a sampled Lab-micro volume of 1.76 mL and 2000 seconds of images at 8 Hz that were collected and classified for each species by the SPC+CNN-Pier.

Species	Proportionality	Volume In-Focus (mL)
Lingulodinium polyedra	0.39	4.29 x 10^{-5}
Prorocentrum micans	2.02	2.22 x 10^{-4}

A significant feature of the work reported here is the computation of the "effective sampling volume" of the SPC-CNN results that permits the estimate of organism abundance. Considering the most abundant and highly correlated species (*Lingulodinium polyedra* and *Prorocentrum micans*) equation 4.3 can be used to compute this volume using the slope of the fit as displayed in Figure 5.4. Given that this slope is (.39, 2.02) and that the reported Lab-micro samples a 1.76 mL sample volume, our cumulative sampling volume for 2000 seconds of images at 8 Hz is (.69, 3.56) mL. Then, the "effective sampling volume" per image is estimated as (0.043, 0.22) μL after dividing by the 16000 frames.

5.4 Continuous Observation Data

One major advantage of the SPC-Pier system is that it can observe plankton continuously with a variable integration time to aggregate totals for the species, we viewed the systems' output over our 2000 seconds integration time over the total period of operation from the end of May until October 2019. The data are displayed in Figure 5.5. Here, the continuous grey line indicates counts of the 4 species from the SPC+CNN during both the lab sampling occurrences as well as other times where there were not manually

collected samples. Here we note the increase in the *Akashiwo sanguinea* and *Cochlodinium* spp. during and in between the lab samples as well as the absence of increased abundance for the *Lingulodinium polyedra* as well as the *Prorocentrum micans*.

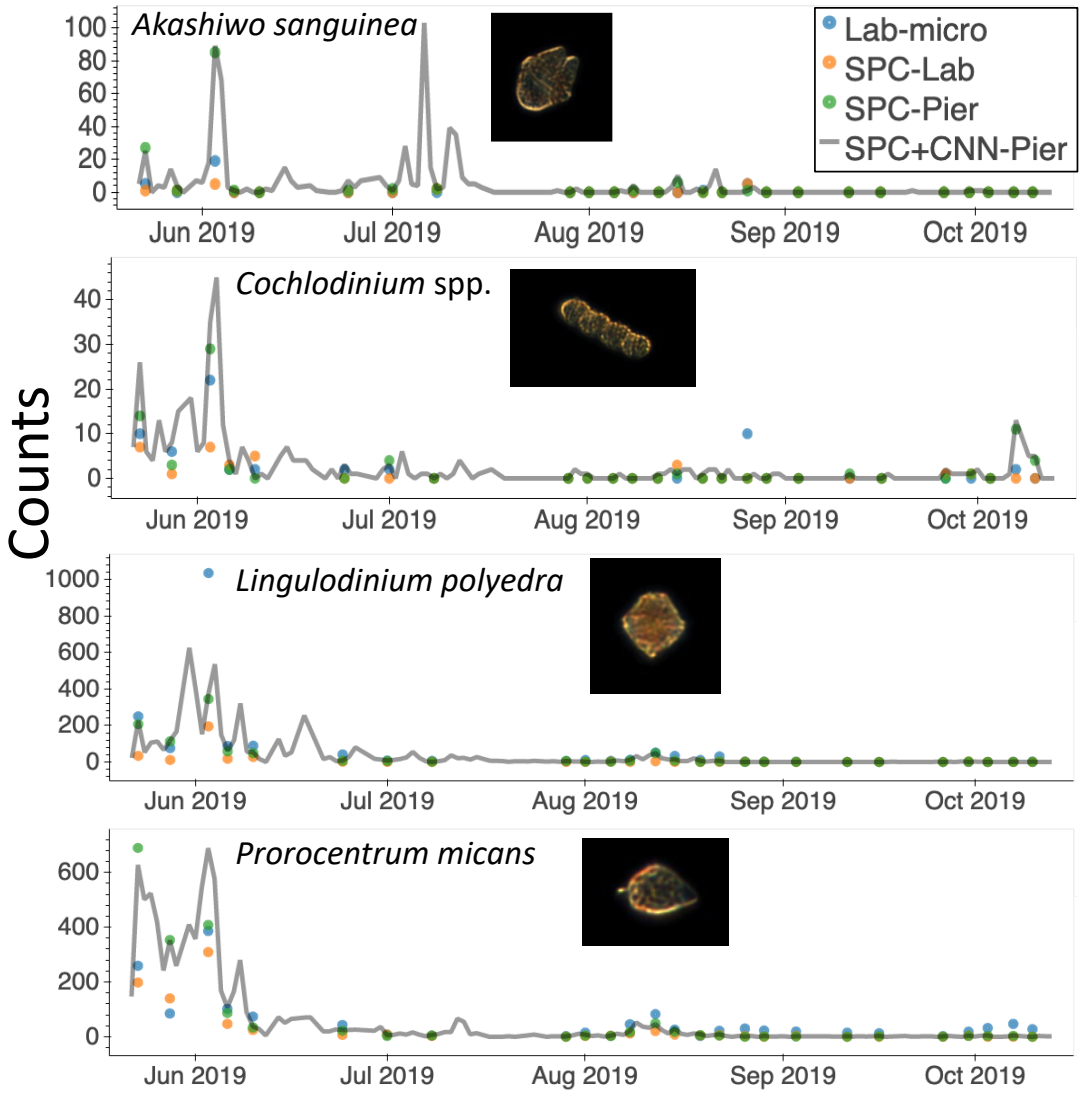

Figure 5.5: Time series of species abundances observed during a 3-month deployment via the SPC-pier and the SCOOS monitoring program during 2019. Automated image classification was used to produce counts on periods not sampled by the SCOOS program. Plots are shown for only the highly correlated abundant and rare species.

Chapter 6

Discussion

In recounting the goals of the work reported here we first sought to explore the ability of CNNs to correctly classify the images that were recorded from the SPC systems. Although lab-based taxonomy of the phytoplankton species is well established, the correspondence between those species and our dark field microscopes has not been established in in the context of the lab work. In examining the potential differences that could contribute between the two methods, Lab-micro vs. SPC+CNN, there are several factors to consider. As described, the samples observed by the SPC microscopes are present in a ranged based defocus that is a necessary consequence of the dark field illumination. In addition, since the SPC microscopes image organisms that are freely drifting in the field of view of the system, a natural assumption is that their orientation, relative to the viewpoint of the camera, is uniformly random. In contrast to larger zooplankton, such as copepods, our organisms of interest have smaller morphological differences that are also confounded by the aspect dependent views that we observe. This makes the identification more difficult for automated systems as well as taxonomists viewing the resultant SPC images.

In considering the success of the CNNs to classify the species present in the images, we found that the imbalanced nature of datasets collected in the wild influenced the performance of the system significantly. We note that class imbalance is a well-studied problem that exists in many real-world ocean ecosystem datasets (e.g. WHOI-plankton [24], EILAT and RAMAS coral dataset [44]) in which rare species

have far fewer images than abundant species. To combat this problem, we applied transfer learning from a less-imbalanced and filtered dataset to a more-imbalanced and unfiltered one. We also applied cost-sensitive learning, one of the techniques commonly used to improve the performance of class imbalanced classification [41]. However, the results of comparatively low performance on rare classes suggest a limited capability of our techniques to mitigate the class imbalance problem. As a future direction, to further improve our model, it would be worth experimenting with other methods, such as an ensemble of CNN models [45]; transferring parallel CNNs [46]); or applying transfer learning by pre-training with class-normalized data [47].

Compared to the class imbalance problem, domain shift is less discussed in deep learning applications on ecosystem literatures. However, our results suggest that this problem deserves critical consideration when deep learning systems are to be deployed on a testing environment different from that used for training. Many zooplankton detection systems, such as ZooplanktoNet [48], Zooglider [26], did not explicitly address and investigate their deep learning models' capability to transfer across domains. When purely trained on SPC-Pier image data, our model was not able to replicate its high performance on SPC-Lab data. Compared to the test result of SPC-Pier, half of tested classes in SPC-Lab have noticeably lower class accuracies. The class accuracies of *Prorocentrum micans* and *Lingulodinium polyedra* dropped the most from SPC-Pier to SPC-Lab. Compared to SPC-Pier data, the *Lingulodinium polyedra* appeared to be dimmer and more transparent in the SPC-Lab data, whereas *Prorocentrum micans* tend to be lighter colored (shown in appendix). In future research, experimenting with other domain adaptation techniques, such as similarity learning [49], image-to-image translation [50], can help further improve our model. Solving the domain shift problem is essential to ensuring the reliability of deep learning automated systems in different environments.

Considering the (9) species investigated here, the significant correlation between the Lab-micro counts and the SPC+CNN-Pier data for *Prorocentrum micans* and *Lingulodinium polyedra* indicates that,

under the environmental and lab identification procedures developed here, the *in situ* system counts can be transformed into estimates of concentration that are "consistent" with traditional microscopy observations. These correlation results were also consistent when using the manually enumerated SPC counts instead of the SPC+CNN. This illustrates that any potential limitations could be due to the *in situ* imaging system sampling factors which our multiplicative scaling factor α seeks to address in our volume computation analysis.

Both the SPC methods and Lab-micro show gaps in their ability to detect certain species. Firstly, Lab-micro only detected *Gyrodinium* on one day, while both SPC methods detected it on more than 20 of the 26 days. We found this to be due to the formaldehyde treatment used during the Lab-micro's sample sedimentation process. Specifically, fixing samples with formaldehyde solution can lead to dissolvement and subsequent misidentification of "naked" species like *Gyrodinium* spp. [51]. On the other hand, the SPC methods show difficulty detecting *Pseudo-nitzschia* spp.. Whether this is due to the inefficiency of the dark field imaging technique or, rather, some effects related to their chain like structure when viewed in 3D is unknown. We do note, however, that there may be some advantages to observing settled samples.

As the other major goal of this research, we sought to estimate the "effective sampling volumes" so that abundance could be estimated, an important metric. Here, we note that, as reported on the web site, (spc.ucsd.edu) the SPCP2 camera, which is the one being used here, has a "high-resolution image volume" of .1 μL and a "Blob detection volume" of 10 μL. The sample volumes reported in Table 5.2 of .043 μL and .22 μL for *Lingulodinium polyedra* and *Prorocentrum micans*, respectively, for the SPC+CNN-Pier are marginally less, likely due to the fact that the system has only one view angle that can result in ambiguities that prevent the unique identification of the species. We note also that in comparing the SPC+CNN-Lab values of the slope vs Lab-micro, the proportionalities are approximately half of these values. The discrepancy may be because the SPC-Lab samples were taken from the near-surface of the ocean (0.5m), whereas the SPC-Pier samples from a tidally dependent depth of 3 meters. Alternatively, the

difference may be due to orientation-dependent effects that were a result of "flowing" the collected water past the SPC-Lab or rather some differences in the two optical systems, such as illumination intensity. We also note that less abundant species (e.g. *Akashiwo sanguinea* and *Cochlodinium* spp.) had a reasonable fit between the SPC-Pier and the Lab-micro, with the SPC-Pier having a larger slope and hence, an estimated larger sampling volume. However, in these cases, the certainty of these values is less due to the small number of samples.

A distinguishing feature of this analysis is that the "effective sampling volumes" as computed via comparison with the Lab-micro calibrations are different for each of these species (e.g. *Lingulodinium polyedra* and *Prorocentrum micans*). The difference in sampling volume amongst each species was not entirely unanticipated as our dark-field illumination setup measures an orientation varying image of these organisms, thereby causing CNNs and expert taxonomists to likely be less capable of measuring the exact identity for each species. Consequently, our linear fit for each of the species has a different slope, leading to different effective sampling volumes that are species dependent.

An important aspect of *in situ* sampling is that it is capable of detecting organisms on a 24/7 basis. As demonstrated in Figure 5.5, the *in situ* microscope can provide continuous, real-time sampling during periods when there was no manual data collection. The period from May to October 2019 provided roughly 128k images during this period via the automated sampling. The values obtained for the *Lingulodinium polyedra* and *Prorocentrum micans* study showed realistically gradual abundance increases and decreases of both, that occur before and after a detected bloom. Rarer species, such as *Akashiwo sanguinea* and *Gyrodinium* spp., also showed similar trends but demonstrated increases in abundance were missed by the manual collection. Although a more detailed analysis would be needed to estimate the confidence in these observations, it seems that these cell estimates are simply undetected because of the less frequent sampling by the Lab-micro. This, in turn, highlights the need of real-time continuous monitoring with less human effort. With reference to Figure 5.5, we note that the lower estimates of the SPC systems during the times

between July and October indicated that there were no significant blooms. We also note that for the other species, the *Akashiwo sanguinea* and *Cochlodinium* spp., there were significant peaks in their number of counts that were outside of the sampling times for the Lab-micro studies here, however, given the low numbers of counts during the intervals where there were both Lab-micro and SPC samples, the confidence in predicting abundance from the SPC-Pier counts could be suspect. We do, however, note that given the continuous nature of the SPC data stream that a set of algorithms could be implemented that would then motivate adaptive sampling that would result in lab quantification and clarification.

One advantage of systems, like SPC+CNN, that produce real time data is their potential use as an early detection system. The early, data-driven insights would inform decision making in monitoring programs, such as SCCOOS, for which shore station leaders have limited information on the daily abundance level of the HAB species. For example, previous studies show that it can be advantageous to know the initial and final periods of a bloom [52]. Stroming showed the socioeconomic benefit of early HAB detection and estimated a saving of $370,000 following the early warning of a 2017 cyanoHAB event in Utah Lake. Given the strong validity found in the present study for estimating HAB abundances, the recommended next steps would be to explore the use of the SPC for supporting decision-making in such settings.

Chapter 7

Conclusion

In summary, the SPC+CNN workflow has shown its capability to provide real-time, high accuracy detection of HABs on certain species, such as *Akashiwo sanguinea, Cochlodinium* spp., *Lingulodinium polyedra* and *Prorocentrum micans*. Although its performance has been species-dependent, it has shown a high correlation with the Lab-micro counts. Moreover, this automated workflow can detect rare species more frequently than the manual method. It also minimizes manual labor and can provide continuous sampling at a high spatial and temporal resolution. All of these benefits make the SPC+CNN a potentially important tool that has the capability to advance the study of imaging, recognition, and monitoring of HAB-related phytoplankton. The results suggest that image-based monitoring systems, supported by high-throughput automated classifiers, can be a reliable alternative to time-consuming manual sampling campaigns. Moreover, our experimental techniques and analyses provide a framework for future intercalibration studies of innovative new plankton sampling modalities.

Appendix A

Supplementary Information

A.1 Comparison of Lab-micro vs. SPC and SPC+CNN

Below are supplementals to the analysis of the linearly fitted models between the SPC+CNN vs. Lab-micro counts along with our previous analysis referenced in Section 5.2.

To further evaluate the relationship between counts of the Lab-micro and SPC+CNN, we computed a least squares regression between the species-specific counts acquired by each method. Figure 5.4 displays results of the linear fitting on a log-scaled graph for 4 of the species, while the other 5 species can be found in the Supplemental Figure A.2. Similar to the results of the Pearson correlation analysis for the pair of SPC+CNN-Pier vs. Lab-micro, we found that the less abundant species (*Akashiwo, Cochlodinium*) produced high R^2 values, as well as the more abundant species (*Lingulodinium polyedra, Prorocentrum micans*). We also observed a pattern of the size of the prediction and confidence bands being related to the occurring frequency of the species. The two abundant species showed much narrower prediction and confidence bands, in contrast to the two rare species, which exhibited wider bands. Discrepancy of the size of the bands could be due to the sparse signal that comes from sampling rare species. The comparability observations of the 4 species could also be validated by the linearly fitted models between these species counts of the Lab-micro and manually enumerated SPCs (Supplementals A.2), which only showed marginally different

R^2 values.

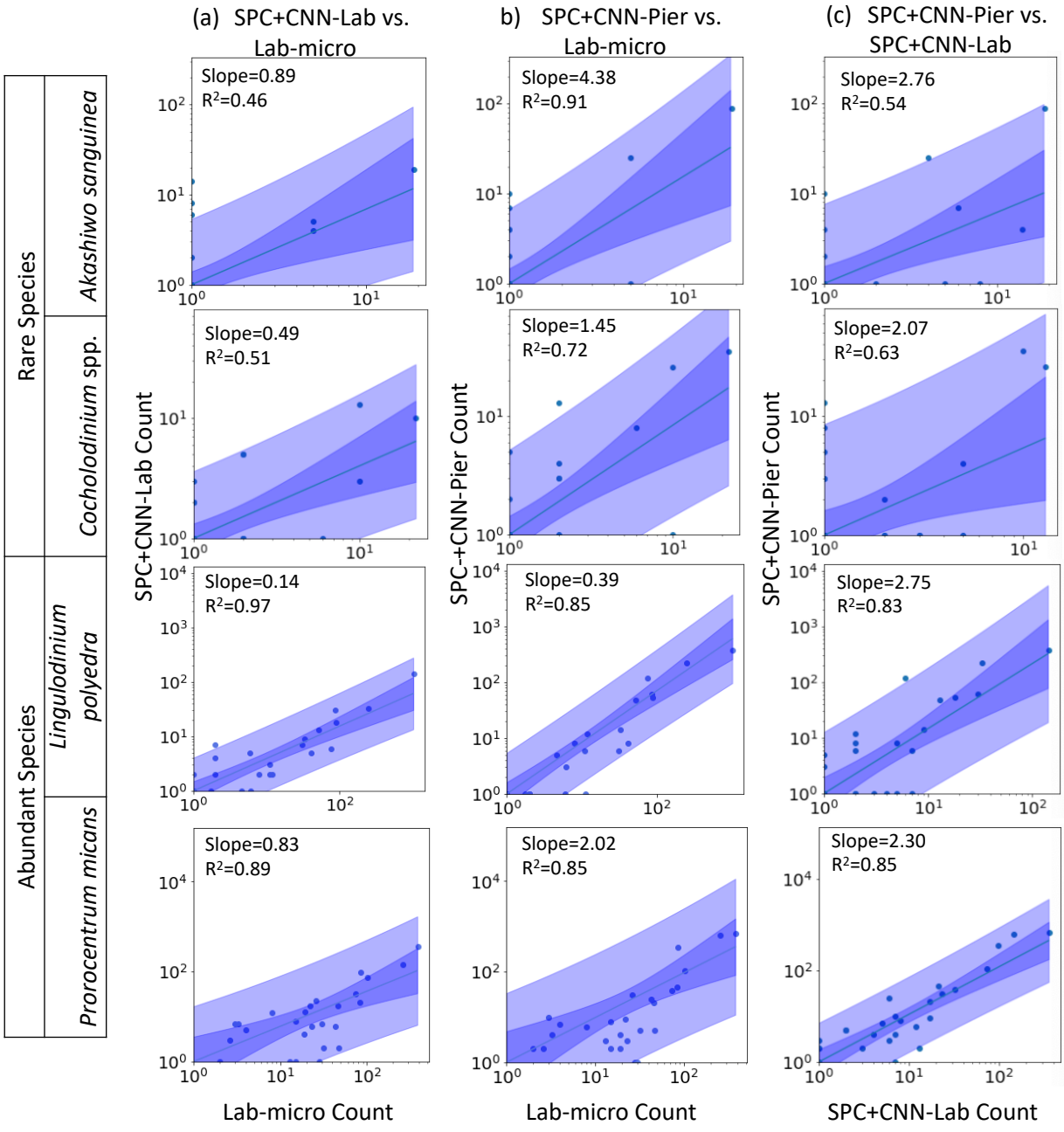

Figure A.1: Relationships between Lab-micro and SPC+CNN: Strongly linear-related species (*Akashiwo sanguinea, Cochlodinium* spp., *Lingulodinium polyedra, Prorocentrum micans*)

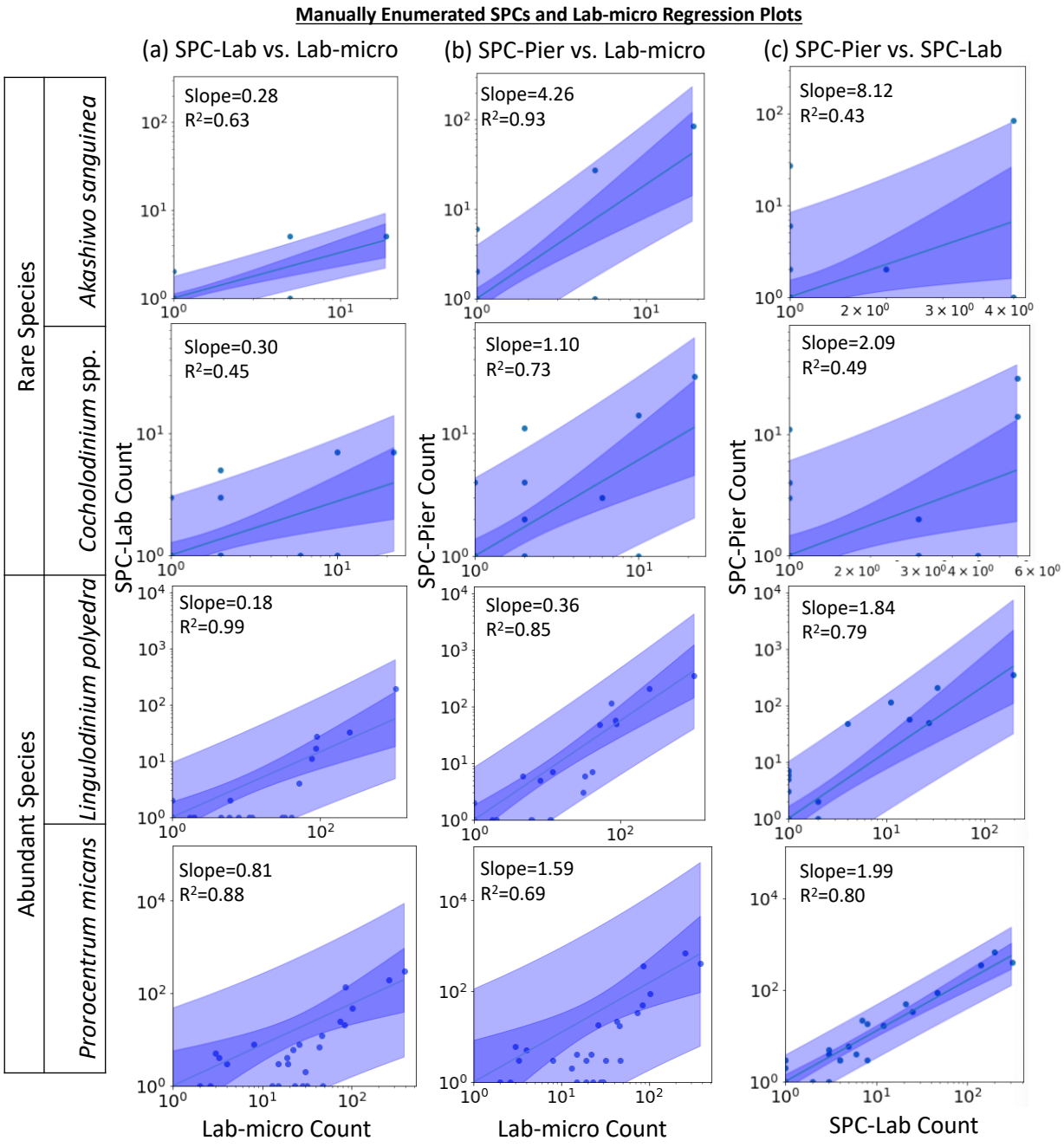

Figure A.2: Relationships between Lab-micro and SPC: Strongly linear-related species (*Akashiwo sanguinea, Cochlodinium* spp., *Lingulodinium polyedra, Prorocentrum micans*)

In a general comparison of the linear fit computed by the model across the 3 possible pairs, we can observe the strongest linear relationship amongst the SPC+CNN-Pier and SPC+CNN-Lab. The results, specifically the proportionality approximation, convey that the SPC+CNN-Pier's sampling of an aggregate volume over the 2000 seconds was nearly 2x that of the SPC+CNN-Lab. Regarding the other 5 species, a majority of these species showed non-existing-to-poor linear relationships between the Lab-micro and SPC+CNN counts. The linear fit for the *Pseudo-nitzschia* spp. showed little ability to model the relationship between the SPC+CNN and Lab-micro, as the SPC itself poorly detected the diatom chains. *Gyrodinium* spp. was mostly absent from the Lab-micro, preventing a linear regression between the compared counts. Species that had previously demonstrated low classification performance resulted in poorer relationships when computing the linear regression for the CNN-based pairs of counting methods. Comparatively to its manually enumerated-based linear regressions (Supplemental A.2), *Ceratium falcatiforme or fusus* and *Chattonella* spp. showed degraded R^2 values and slopes across all 3 possible pairs, suggesting the effect of the classification performance upon the linearly modeled relationship. *Ceratium furca* also showed some fluctuations when comparing automated vs manual regressions, but generally showed only a moderate relationship.

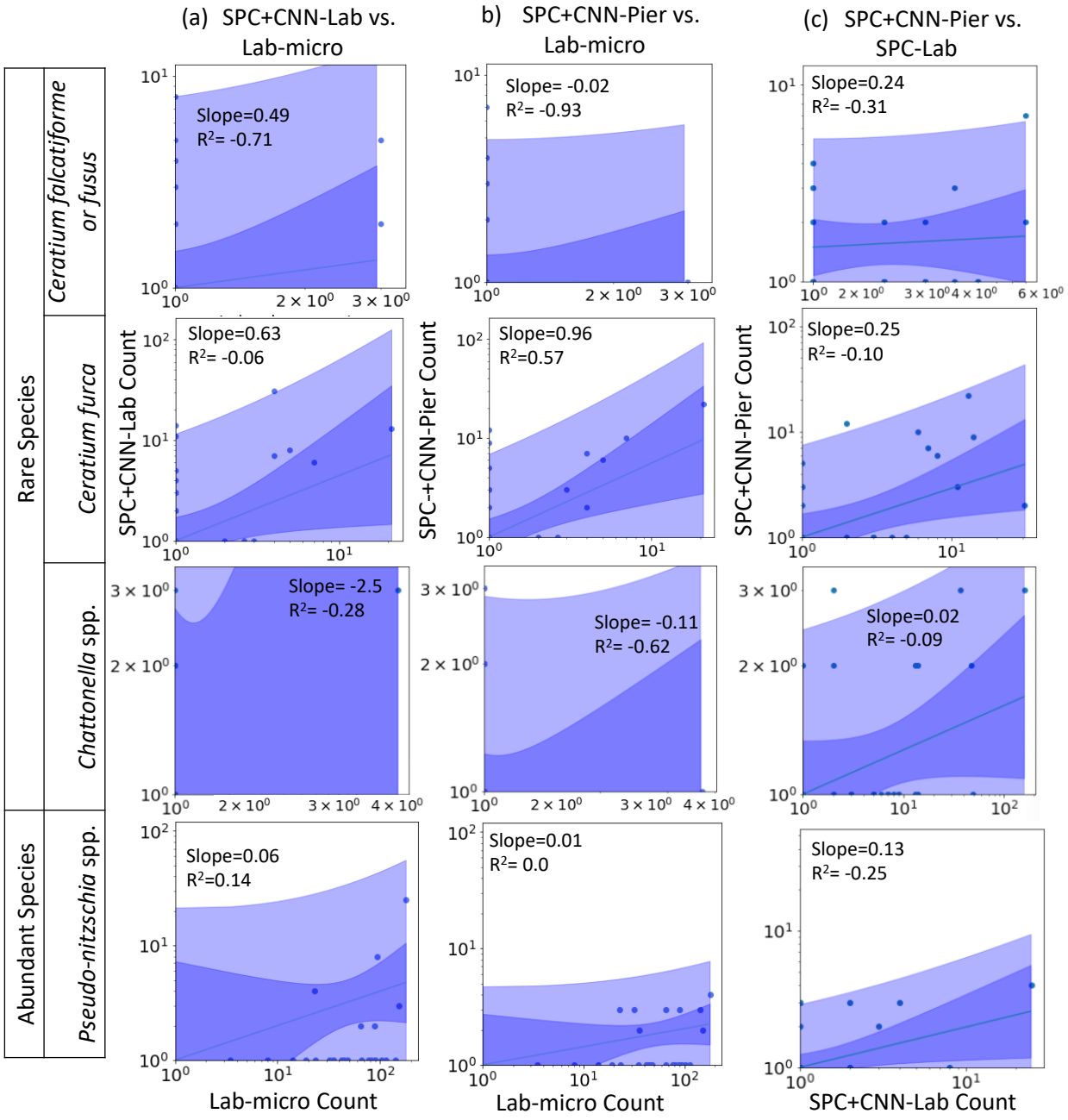

Figure A.3: Relationships between Lab-micro and SPC+CNN: Poorly linear-related species (*Ceratium falcatiforme or fusus, Ceratium furca, Chattonella* spp, *Pseudo-nitzschia* spp.)

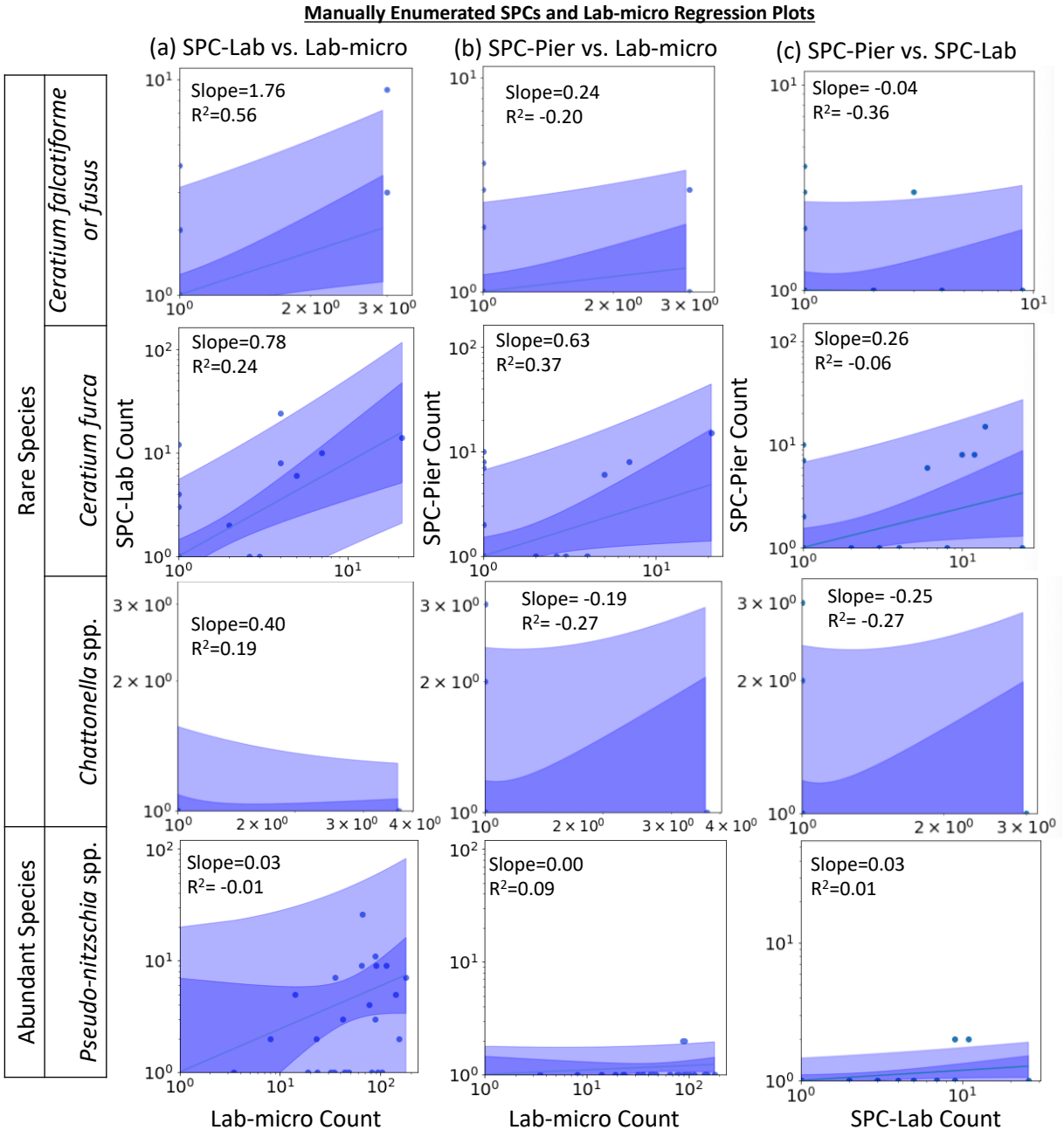

Figure A.4: Relationships between Lab-micro and SPC: Poorly linear-related species (*Ceratium falcatiforme or fusus, Ceratium furca, Chattonella* spp, *Pseudo-nitzschia* spp.)

The chapters of this thesis consist of material currently being prepared for submission. The thesis author was the author of this material.

- Chapter 1-7 has material currently being prepared for submission as it may appear in K. Le, Z. Yuan, A. Syed, E. Orenstein, D. Ratelle, M. Carter, S. Strang, K. Kenitz, P. Morgado, P. Franks, N. Vasconcelos, and J. Jaffe, "Benchmarking the Automated Analysis of *In situ* Plankton Imaging and Recognition," *Methods in Ecol Evol*, 2021, in preparation.

References

[1] Christopher B. Field, Michael J. Behrenfeld, James T. Randerson, and Paul Falkowski. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, 281(5374):237, July 1998.

[2] Fabien Lombard, Emmanuel Boss, Anya M. Waite, Meike Vogt, Julia Uitz, Lars Stemmann, Heidi M. Sosik, Jan Schulz, Jean-Baptiste Romagnan, Marc Picheral, Jay Pearlman, Mark D. Ohman, Barbara Niehoff, Klas O. Möller, Patricia Miloslavich, Ana Lara-Lpez, Raphael Kudela, Rubens M. Lopes, Rainer Kiko, Lee Karp-Boss, Jules S. Jaffe, Morten H. Iversen, Jean-Olivier Irisson, Katja Fennel, Helena Hauss, Lionel Guidi, Gaby Gorsky, Sarah L. C. Giering, Peter Gaube, Scott Gallager, George Dubelaar, Robert K. Cowen, François Carlotti, Christian Briseño-Avena, Léo Berline, Kelly Benoit-Bird, Nicholas Bax, Sonia Batten, Sakina Dorothée Ayata, Luis Felipe Artigas, and Ward Appeltans. Globally Consistent Quantitative Observations of Planktonic Ecosystems. *Frontiers in Marine Science*, 6, 2019. Publisher: Frontiers.

[3] E. Sinha, A. M. Michalak, and V. Balaji. Eutrophication will increase during the 21st century as a result of precipitation changes. *Science*, 357(6349):405–408, July 2017. Publisher: American Association for the Advancement of Science Section: Report.

[4] Hey-Jin Kim, Arthur J. Miller, John McGowan, and Melissa L. Carter. Coastal phytoplankton blooms in the Southern California Bight. *Progress in Oceanography*, 82(2):137–147, August 2009.

[5] K. A. Lefebvre, C. L. Powell, M. Busman, G. J. Doucette, P. D. Moeller, J. B. Silver, P. E. Miller, M. P. Hughes, S. Singaram, M. W. Silver, and R. S. Tjeerdema. Detection of domoic acid in northern anchovies and California sea lions associated with an unusual mortality event. *Natural Toxins*, 7(3):85–92, 1999.

[6] C. A. Scholin, F. Gulland, G. J. Doucette, S. Benson, M. Busman, F. P. Chavez, J. Cordaro, R. DeLong, A. De Vogelaere, J. Harvey, M. Haulena, K. Lefebvre, T. Lipscomb, S. Loscutoff, L. J. Lowenstine, R. Marin, P. E. Miller, W. A. McLellan, P. D. Moeller, C. L. Powell, T. Rowles, P. Silvagni, M. Silver, T. Spraker, V. Trainer, and F. M. Van Dolah. Mortality of sea lions along the central California coast linked to a toxic diatom bloom. *Nature*, 403(6765):80–84, January 2000.

[7] Jayme Smith, Paige Connell, Richard H. Evans, Alyssa G. Gellene, Meredith D. A. Howard, Burton H. Jones, Susan Kaveggia, Lauren Palmer, Astrid Schnetzer, Bridget N. Seegers, Erica L. Seubert, Avery O. Tatters, and David A. Caron. A decade and a half of Pseudo-nitzschia spp. and domoic acid along the coast of southern California. *Harmful Algae*, 79:87–104, November 2018.

[8] Claudia Castellani. Plankton: A Guide to their Ecology and Monitoring for Water Quality. *Journal of Plankton Research*, 32(2):261–262, February 2010. Publisher: Oxford Academic.

[9] W. M. Hamner, L. P. Madin, A. L. Alldredge, R. W. Gilmer, and P. P. Hamner. Underwater observations of gelatinous zooplankton: Sampling problems, feeding biology, and behavior1. *Limnology and Oceanography*, 20(6):907–917, 1975. _eprint: https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.4319/lo.1975.20.6.0907.

[10] M. Omori and W. M. Hamner. Patchy distribution of zooplankton: Behavior, population assessment and sampling problems. *Marine Biology*, 72(2):193–200, December 1982.

[11] John R. Beers and Gene L. Stewart. The Preservation of Acantharians in Fixed Plankton Samples1. *Limnology and Oceanography*, 15(5):825–827, 1970. _eprint: https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.4319/lo.1970.15.5.0825.

[12] Robert J. Olson and Heidi M. Sosik. A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging FlowCytobot: In situ imaging of nano- and microplankton. *Limnology and Oceanography: Methods*, 5(6):195–203, June 2007.

[13] Mark C. Benfield, Cabell S. Davis, Peter H. Wiebe, Scott M. Gallager, R. Gregory Lough, and Nancy J. Copley. Video Plankton Recorder estimates of copepod, pteropod and larvacean distributions from a stratified region of Georges Bank with comparative measurements from a MOCNESS sampler. *Deep Sea Research Part II: Topical Studies in Oceanography*, 43(7):1925–1945, January 1996.

[14] Neeraj Iyer. Machine Vision Assisted In Situ Ichthyoplankton Imaging System. July 2013. Accepted: 2013-07-12T17:23:59Z.

[15] Robert K. Cowen, Adam T. Greer, Cedric M. Guigand, Jonathan A. Hare, David E. Richardson, and Harvey J. Walsh. Evaluation of the In Situ Ichthyoplankton Imaging System (ISIIS): comparison with the traditional (bongo net) sampler. *Fishery Bulletin*, 111(1), December 2013.

[16] Philip F. Culverhouse, Norman Macleod, Robert Williams, Mark C. Benfield, Rubens M. Lopes, and Marc Picheral. An empirical assessment of the consistency of taxonomic identifications. *Marine Biology Research*, 10(1):73–84, January 2014. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17451000.2013.810762.

[17] Eric C. Orenstein, Devin Ratelle, Christian Briseño-Avena, Melissa L. Carter, Peter J. S. Franks, Jules S. Jaffe, and Paul L. D. Roberts. The Scripps Plankton Camera system: A framework and platform for in situ microscopy. *Limnology and Oceanography: Methods*, 18(11):681–695, 2020. _eprint: https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.1002/lom3.10394.

[18] Marc Picheral, Lionel Guidi, Lars Stemmann, David M. Karl, Ghizlaine Iddaoud, and Gabriel Gorsky. The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnology and Oceanography: Methods*, 8(9):462–473, 2010. _eprint: https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.4319/lom.2010.8.462.

[19] Heidi M. Sosik and Robert J. Olson. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. *Limnology and Oceanography: Methods*, 5(6):204–216, 2007. _eprint: https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.4319/lom.2007.5.204.

[20] Jeffrey S. Ellen, Casey A. Graff, and Mark D. Ohman. Improving plankton image classification using context metadata. *Limnology and Oceanography: Methods*, 17(8):439–461, 2019.

[21] Jessica Y. Luo, Jean-Olivier Irisson, Benjamin Graham, Cedric Guigand, Amin Sarafraz, Christopher Mader, and Robert K. Cowen. Automated plankton image analysis using convolutional neural networks. *Limnology and Oceanography: Methods*, 16(12):814–827, 2018.

[22] Eric C. Orenstein and Oscar Beijbom. Transfer Learning and Deep Feature Extraction for Planktonic Image Data Sets. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1082–1088, March 2017.

[23] Pablo González, Alberto Castaño, Emily E Peacock, Jorge Díez, Juan José Del Coz, and Heidi M Sosik. Automatic plankton quantification using deep features. *Journal of Plankton Research*, 41(4):449–463, July 2019.

[24] Eric C Orenstein, Oscar Beijbom, Emily E Peacock, and Heidi M Sosik. WHOI-Plankton- A Large Scale Fine Grained Visual Recognition Benchmark Dataset for Plankton Classification. page 2, 2015.

[25] Ahmad Salman, Ahsan Jalal, Faisal Shafait, Ajmal Mian, Mark Shortis, James Seager, and Euan Harvey. Fish species classification in unconstrained underwater environments based on deep learning. *Limnology and Oceanography: Methods*, 14(9):570–585, 2016. _eprint: https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.1002/lom3.10113.

[26] Benjamin M Whitmore, Catherine F Nickels, and Mark D Ohman. A comparison between Zooglider and shipboard net and acoustic mesozooplankton sensing systems. *Journal of Plankton Research*, 41(4):521–533, July 2019.

[27] Pablo González, Eva Álvarez, Jorge Díez, Ángel López-Urrutia, and Juan José del Coz. Validation methods for plankton image classification systems. *Limnology and Oceanography: Methods*, 15(3):221–237, 2017. _eprint: https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.1002/lom3.10151.

[28] Eric C. Orenstein, Kasia M. Kenitz, Paul L. D. Roberts, Peter J. S. Franks, Jules S. Jaffe, and Andrew D. Barton. Semi- and fully supervised quantification techniques to improve population estimates from machine classifiers. *Limnology and Oceanography: Methods*, n/a(n/a), 2020. _eprint: https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.1002/lom3.10399.

[29] Kasia M. Kenitz, Eric C. Orenstein, Paul L. D. Roberts, Peter J. S. Franks, Jules S. Jaffe, Melissa L. Carter, and Andrew D. Barton. Environmental drivers of population variability in colony-forming marine diatoms. *Limnology and Oceanography*, n/a(n/a), May 2020. Publisher: John Wiley & Sons, Ltd.

[30] B. Karlson, C. Cusack, and E. Bresnan. Microscopic and molecular methods for quantitative phytoplankton analysis. Report, UNESCO, 2010. Accepted: 2017-03-05T10:49:24Z.

[31] H. Utermöhl. Neue Wege in der quantitativen Erfassung des Plankton.(Mit besonderer Berücksichtigung des Ultraplanktons.). *SIL Proceedings, 1922-2010*, 5(2):567–596, January 1931. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/03680770.1931.11898492.

[32] Hans Utermöhl. *Zur Vervollkommnung der quantitativen Phytoplankton-Methodik*. Schweizerbart, Stuttgart, 1958. OCLC: 310079977.

[33] Akshay Raj Dhamija, Manuel Günther, and Terrance E. Boult. Reducing Network Agnostophobia. *arXiv:1811.04110 [cs]*, December 2018. arXiv: 1811.04110.

[34] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015. Number: 7553 Publisher: Nature Publishing Group.

[35] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]*, December 2015. arXiv: 1512.03385.

[36] Igor V. Tetko, David J. Livingstone, and Alexander I. Luik. Neural network studies. 1. Comparison of overfitting and overtraining. *Journal of Chemical Information and Modeling*, 35(5):826–833, September 1995.

[37] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. page 8.

[38] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? page 9.

[39] Jose G. Moreno-Torres, Troy Raeder, Rocío Alaiz-Rodríguez, Nitesh V. Chawla, and Francisco Herrera. A unifying view on dataset shift in classification. *Pattern Recognition*, 45(1):521–530, January 2012.

[40] Baochen Sun, Jiashi Feng, and Kate Saenko. Return of Frustratingly Easy Domain Adaptation. *arXiv:1511.05547 [cs]*, December 2015. arXiv: 1511.05547.

[41] Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. Learning to Model the Tail. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 7029–7039. Curran Associates, Inc., 2017.

[42] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, January 2017. arXiv: 1412.6980.

[43] MM Mukaka. A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal : The Journal of Medical Association of Malawi*, 24(3):69–71, September 2012.

[44] Asm Shihavuddin. Coral reef dataset. 2, July 2017. Publisher: Mendeley.

[45] Alessandra Lumini and Loris Nanni. Deep learning and transfer learning features for plankton classification. *Ecological Informatics*, 51:33–43, May 2019.

[46] C. Wang, X. Zheng, C. Guo, Z. Yu, J. Yu, H. Zheng, and B. Zheng. Transferred Parallel Convolutional Neural Network for Large Imbalanced Plankton Database Classification. In *2018 OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO)*, pages 1–5, May 2018.

[47] H. Lee, M. Park, and J. Kim. Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3713–3717, September 2016. ISSN: 2381-8549.

[48] Jialun Dai, Ruchen Wang, Haiyong Zheng, Guangrong Ji, and Xiaoyan Qiao. ZooplanktoNet: Deep convolutional network for zooplankton classification. In *OCEANS 2016 - Shanghai*, pages 1–6, Shanghai, China, April 2016. IEEE.

[49] Pedro O. Pinheiro. Unsupervised Domain Adaptation with Similarity Learning. *arXiv:1711.08995 [cs]*, April 2018. arXiv: 1711.08995.

[50] Zak Murez, Soheil Kolouri, David Kriegman, Ravi Ramamoorthi, and Kyungnam Kim. Image to Image Translation for Domain Adaptation. *arXiv:1712.00479 [cs]*, December 2017. arXiv: 1712.00479.

[51] Eduardo Costas, Rafael Zardoya, Jose Bautista, Amando Garrido, Carmen Rojo, and Victoria López-Rodas. Morphospecies Vs. Genospecies in Toxic Marine Dinoflagellates: An Analysis of

Gymnodznium Catenatum/Gyrodinium Impudicum and Alexandrium Minutum/a. Lusitanicum Using Antibodies, Lectins, and Gene Sequences1. *Journal of Phycology*, 31(5):801–807, 1995. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0022-3646.1995.00801.x.

[52] Signe Stroming, Molly Robertson, Bethany Mabee, Yusuke Kuwayama, and Blake Schaeffer. Quantifying the Human Health Benefits of Using Satellite Information to Detect Cyanobacterial Harmful Algal Blooms and Manage Recreational Advisories in U.S. Lakes. *GeoHealth*, 4(9):e2020GH000254, 2020. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020GH000254.