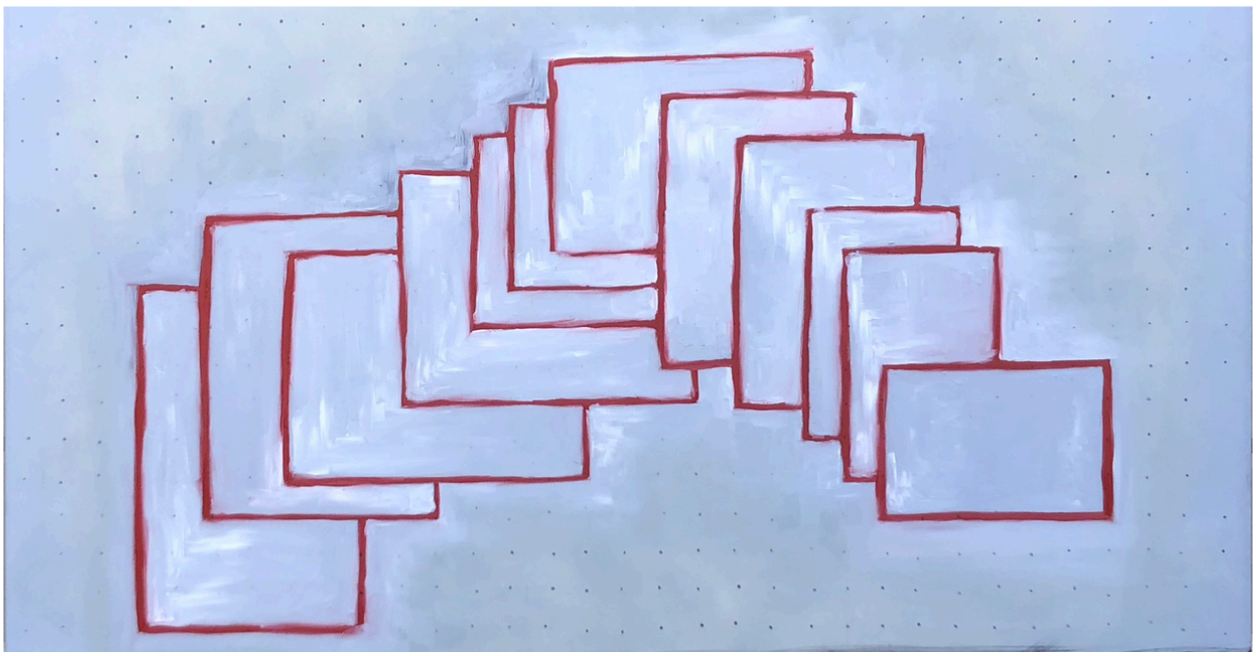
Emerging Technologies in Fisheries Science:
A Transdisciplinary Report



Alrik Firl, Electrical Engineering and Computer Science

Katlyn Haven, Marine Resource Management

Samm Newton, Environmental Arts and Humanities

Funded by the National Science Foundation

Table of Contents

I. Introduction

The Pacific Coast Groundfish Fishery harvests a diverse and large grouping of fishes, but it did not become heavily fished until around WWII. This makes the groundfish fishery a comparatively young fishery. Despite its youth, it is one of the largest and most lucrative fisheries in Oregon—with a current harvest value of approximately $48 million per year, which is exceeded only by the Dungeness crab fishery[1]. Northeastern Pacific Coast Groundfish species are also important for recreational and tribal purposes, although it is difficult to compare these to the commercial industry[2,3]. With over 90 different species to consider, this commercial fishery is complex, and there are many different stakeholder groups involved, each with their own goals, values, and perspectives[4].

Currently, this fishery is federally managed by the National Marine Fisheries Service (NMFS) and the Pacific Fisheries Management Council (PFMC) since Northeastern Pacific Coast groundfish reside within state and federal waters, as well as beyond into the high seas[3]. The primary law governing the management of this fishery is the Magnuson-Stevens Fishery Conservation and Management Act (MSA), which was first passed in 1976[5,6]. The MSA also established a network of regional fisheries management councils, of which the PFMC is a part. The PFMC and NMFS manage the groundfish fishery by developing and implementing (respectively) the rules and regulations laid out in the Pacific Coast Groundfish Fishery Management Plan (FMP)[7,8].

Part of the current fishery management plan includes estimating the fish populations and the future trends in those populations[9]. One tool used to achieve this requirement is a stock assessment model. Stock assessment models are made using information from multiple sources including landing records, hook-and-line surveys, and bottom trawl surveys[4,10]. There is inherent uncertainty associated with these stock assessments. If the fish population is underestimated, then this would cause fishing regulations to be stricter than necessary, thus harming the fishing industry and, ultimately, seafood consumers. However, if the fish population is overestimated, then the fishing regulations would be too lenient and this could result in overfishing or even a fishery collapse.

Fishing regulations greatly impact local stakeholders, some of whom rely on the fishery for their livelihoods. These local stakeholders are dependent on accurate stock assessment surveys and models so that the fishing regulations are appropriate. Some stakeholders feel that regulations tend to be overly cautious to compensate for the large amount of uncertainty involved with managing a fishery and estimating a fish population[5, 11,12]. To reduce this uncertainty and the need to err so heavily on the side of caution, stock assessment surveys could include innovative technologies and novel datasets. For example, these stock assessments do not currently use automated video surveillance on their bottom trawl surveys, an emerging form of machine learning.

As understood by the NSF-funded National Research Traineeship (NRT) training, there are three interwoven core concepts: 1) Big Data (BD), 2) Coupled Natural-Human (CNH) systems, and 3) Risk and Uncertainty (R&U) analysis and communication. Big Data refers to any high volume of data with high throughput. Coupled Natural-Human systems are the biological and human worlds, as well as their overlap and interaction. Risk is the potential and likelihood of an unfavorable event, and uncertainty refers to the unknowns of a likelihood, process, or analysis. This project chose to investigate these three concepts within the framework of emerging technologies and fisheries science. Emerging technologies are those dealing with BD, since this is a relatively new area of study, and this project specifically focused on computer vision within machine learning. This technology was applied to the realm of fisheries science and ultimately management, which is the study of a coupled natural-human system. Changing oceans conditions mean that Northeastern Pacific groundfish are at risk and their future is uncertain. Therefore, this project set out to determine how the influence of big data, machine learning, ecological inference, and environmental decision making overlap.

The story of the life and study of these fishes in a newly Americanized sea is ready for a closer examination. It is for these reasons combined that Pacific coast groundfish fishery science provides a robust platform in which to explore the autonomous capacity of technology and data production at the intersection of environmental science and decision making. More specifically, to what extent are large, ecological datasets informing the production and application of emerging technologies in fisheries science, and how are these new technologies and sampling methods being integrated into fisheries management frameworks? A case study in which to explore this concept can be found in the testimony of a flatfish, or rather, the complex, ecologically and economically important assemblages of numerous groundfish species in the northeastern Pacific Ocean where flatfish are found.

II. The Influence of Big Data

Datasets, as well as the technology used to acquire, assemble, and analyze those datasets, have had an influence over directions of knowledge inquiry, both within the advancement of that technology and the ecological questions that can be asked with it. However, datasets do not exist in a vacuum, and generally, datasets are created around problems that people find relevant. Often this comes from a company for a specific task important to their business interests, and there are numerous large proprietary datasets. However, datasets are not strictly created for commercial purposes; many notable datasets such as ImageNet and PASCAL VOC originated from academic institutions, as they filled gaps in areas of significant interest to the computer vision community. Thus, there is a cyclical relationship (Figure 1) between advancement of emerging technology like computer vision and the datasets in existence: problems that researchers and/or institutions are interested in will inevitably have datasets created about them, and people will work on problems that have pre-existing datasets. This in turn lowers the barrier to working on problems with datasets already, and raises the bar to working on problems that have no such datasets.

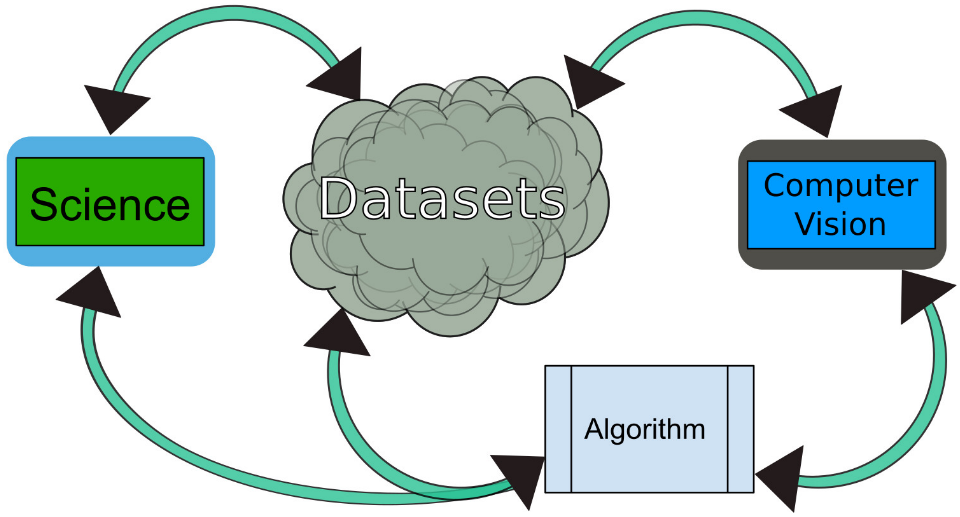


Figure 1: Tasks with one or more datasets have a lower barrier to working on them; tasks without datasets have a higher one. Problems that are important usually have datasets created for them, while datasets that are unimportant do not. Creating a dataset for a problem can encourage further research on the task. The sciences can encourage collaboration by creating datasets relevant to their problem domains.

Additionally, this cyclical relationship between datasets and knowledge inquiry has constructed a culture of path dependency in how ecological questions are asked, in the case of Northeastern Pacific groundfish, within fisheries science and the ability of scientists to predict fish behavior and abundance. Technology has played a key role in what data has been acquired and how it has been assembled and analyzed. The datasets that technology has provided in this way became powerful drivers behind institutional understandings of how fish populations have been measured and managed. This path dependency, combined with the powerful influence of technological choice and dataset production, has acted as both a limiting factor and common foundation for incorporating emerging technologies into fisheries science.

Datasets and Emerging Technologies

Throughout the history of machine learning, datasets have propelled innovation. Many recent advancements in machine learning have been fueled by the advent of deep learning. And, although prolific work on artificial neural networks progressed through 1980s and 90s, deep neural networks only became feasible and preeminent in 2012 with AlexNet and its impressive performance on ImageNet[13,14]. This can be attributed to increases in processing power and more notably, the availability of larger datasets.

Data lies at the heart of machine learning, as it is the information that the machine is meant to ‘learn’ from. However, in the supervised learning paradigm, raw data itself is insufficient; there must also be the desired result of any processing, or labeled datasets (Figure 2). This serves to both define the task to be performed, and provides the implicit relationships between the input and output data samples. While the network’s parameters are directly updated via gradient descent, which is a function of the loss, the loss is a function of the ground truth and the network predictions. Thus, the dataset ground truth lies at the foundation of the algorithm’s training.

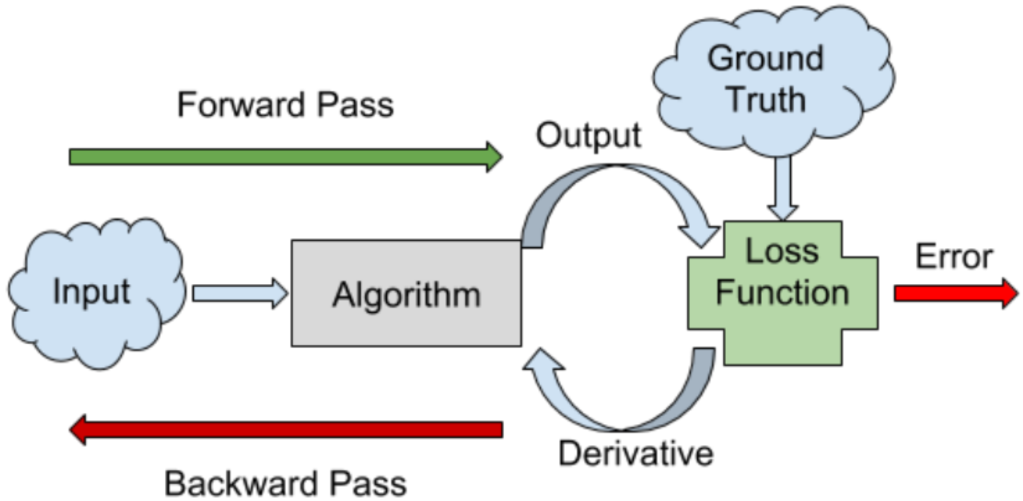


Figure 2: Diagram of prototypical supervised learning process

On a more philosophical level, datasets are a means to distill a specific task to be performed. Any number of tasks can be performed on a piece of data; within the infinite set of potential tasks, the dataset is what defines what the desired outcome should be. As an example, the labels of the dataset determine the type of task to be performed; classification would consist of categories per image, detection would have zero or more bounding boxes, segmentation would have pixel-wise mask images. Given a different set of labels, we can define a different task on the same input data. So, while the raw data is important, it is the ground truth labels that determine what in the data is considered important, and what the algorithm learns. And, the ground truth is labeled and discriminated by human mediaries.

This introduces the insidious problem of biased data. The use of machine learning is often considered a strength, opposed to fallible, imperfect humans, but humans are labeling the data. Humans have egos and biases, for example there are more favorable parole decisions given by judges early in the day and after lunch[15]. While algorithms themselves have no inherent biases, they can inherit the biases that have been encoded in the datasets they are trained on. For example, the Microsoft Common Objects in Context dataset (MS-COCO) was found to contain significant gender bias in visual semantic role labeling; even worse, they also found that training algorithms on such datasets can further amplify biases found in datasets[16].

There is increasing awareness around implicit biases being enshrined in datasets. As machine learning algorithms have entered the commercial sphere, there have been highly visible mistakes in the past years, such as Google's racist face recognition and Uber's self-driving car death[17,18]. There are also concerns regarding less visible algorithmic decision making, such as credit decisions, or COMPAS (predictive software used to decide which prisoners are eligible for early release) that may have built-in biases, but these are not available to be audited or examined due to intellectual property constraints. This raises ethical concerns around proprietary and/or automated algorithms that make decisions with minimal oversight. It also highlights the profound importance of datasets in the creation of valid models.

On a fundamental level, datasets are required for machine learning algorithms to learn. The "No Free Lunch" theorem (NFLT) (for supervised learning) states that no learning algorithm outperforms any other, when evaluated over all possible problems[19]. The NFLT implies that if there is no prior information or assumptions made about the data to evaluate an algorithm with, then no algorithm is better than any other; performing well on one task just means performing poorly on a different one. In practice, this means that algorithms need to impose assumptions regarding the data distributions to be successful on any given task. Often this is done by assuming structured data is more common (i.e., data is likely to have lower Kolmogorov complexity)[20].

In general, stronger assumptions are more informative, and such assumptions are informed by the data being used. In addition, since the dataset specifies what the 'correct' result is for every input (the ground truth), it is directly tied to how well the algorithm performs; if the dataset provides incorrect or noisy labels to the algorithm, then the algorithm will learn the incorrect model for the task at hand, and will likely generalize poorly.

Datasets and Fisheries Science

Despite lofty aspirations for potential applications, at its core, machine learning technology revolves around data. And in the world of ecological inferences, so does fisheries science. In Latin, the word *data* means "to give" or a "(given) thing." But scientific data is far from given. It is acquired by scientists through great effort, and in the case of monitoring the Pacific Coast groundfish fishery, the data has taken several forms. Historically, how scientists have acquired data on Northeastern Pacific coast groundfish has not been straightforward. At least one, and sometimes several, groundfish surveys and data collections processes have been in place in some form or fashion since 1956, but surveys were not as well documented or organized until 1977, just after the implementation of the original version of the Magnuson-Stevens Fisheries Conservation and Management act (MSA) of 1976.

In 1988, J.G. Pope said that "every important dataset should have someone to love and cherish it." Pope had encountered his fair share of disparate, abandoned, and obsolete datasets while putting together a collection of fisheries data—leading him to that conclusion. Catch and effort data for northeast Pacific groundfish from 1956 to 1980 was largely inconsistent, scattered, lost, damaged, orphaned, or erroneous and it has not been used for insights into historical populations of groundfish[21]. Although, much of the groundfish abundance information pre-1977 was gleaned from analyzing commercial fishing landings, using what is called fisheries-dependent datasets[22].

Fisheries-independent surveys are designed and carried out by scientists associated with the National Marine Fisheries Service (NMFS). Initial fisheries-independent surveys did not begin until the late 70's and were geared towards more efficient harvest, as the original MSA mandated that the abundance of the sea be harnessed as effectively as possible[23]. These early fishery-independent surveys were conducted every three years (the triennial surveys) and did not follow any standardized sampling methods, temporally, geo-spatially, species, or depth-wise. But, they did all use the same bottom trawl mechanism and were contracted with commercial fisherman for use of their trawls and vessels[24]. The consistent sampling mechanism, trawling, was the only thing connecting the years of patchy and variable data collected over that time-period.

The science behind fulfilling the requirements of fisheries management plans (FMP's) for Northeastern Pacific coast groundfish has come in great part from bottom trawl survey datasets completed by the regional fisheries science center of the NMFS.[3] Beam trawls, used to sample groundfish populations by the NMFS, are pieces of bottom trawl equipment that have been around since the 1300's. They were originally designed within the fishing industry to drag large open nets along the seafloor and could capture an unprecedented amount of fish. Even at the time of their inception centuries ago, beam trawls were contentious pieces of machinery that caused controversy among fishermen. For example, a petition from 1376 reflects an understanding that these wondyrechauns, as they were called, allowed fisherman to abuse common resources, promoted overfishing, destroyed habitat, and resulted in auxiliary loss of life[25].

Still, trawling endured as a fishing practice and arrived in the U.S. with early European settlers. Although bottom trawls have evolved to include other varieties of trawls, each with different adjustments to improve their fish harvesting abilities, when it comes to demersal fish, a beam trawl of some kind has remained the preferred method of acquiring data for scientists.

According to the National Research Council (NRC) report from the National Academy of Science (NAS), science needed to be relevant, inclusive, objective, transparent, open, timely, and peer-reviewed[26]. It is safe to say that the datasets pieced together from the 50's to the 90's of both fishery dependent and independent research does not necessarily fit into all those categories. But, it is the dataset they had to work with and the only set available when the first FMP was drafted in 1982[27].

In the mid to late 80's and 90's the process of trying to build better algorithms using improved datasets was expanding. For example, at Oregon State University, multiple proposals specifically about groundfish population modeling were funded through different arms of the National Oceanic and Atmospheric Association (NOAA) including the Sea Grant college program, the NMFS, and the NOAA Underwater Research Program (NURP)[28].

In 1995, the trawl survey started to become more formalized in conjunction with the establishment of the Fishery Resource Analysis and Monitoring Division (FRAM) of the Northwest Fisheries Science Center (NWFSC)[24]. This also coincided with the 1996 reauthorization of the MSA and incorporation of the Sustainable Fisheries Act. The 1996 reauthorization of the MSA also reiterated the second national standard first introduced in 1976 about using the "best scientific information available" (BSIA) to make decisions about managing the fishery. This "best science" was an elusive animal—which will be discussed more later—and continued to be a topic of contention for everyone involved in fisheries management through many versions of the MSA. Because best available science meant different things to different people in different fisheries and regions, it has often meant *more* science. The standardization of the WCGBTS could be interpreted as an attempt to provide *more* quantity and *more* quality datasets to analyze.

The poor datasets lacked predictive power, which was apparent when more research went into modeling, influencing investment into assembling better datasets. In time, these datasets and improved modeling impacted management as well and by 1997 there were changes made to the stock assessments alongside large cuts in groundfish quotas. NMFS seemed to shift in earnest to improving the predictive power of stock assessments models[29]. Which meant making innovative advancements in the acquisition of data.

A transition began into a more comprehensive and standardized single survey that cobbled together the disparate sampling methods from the previous thirty or so years. This move to standardization was an effort to create consistency and certainty over time and became the foundation for the main groundfish data survey, known today as the West Coast Groundfish Bottom Trawl Survey (WCGBTS)[24].

Datasets and Eastern Pacific groundfish

Predictive algorithms are based on the idea that you can predict the future, or even assume the present, considering past data inputs. They also assume the present datasets are reflective of reality and that the future will resemble the past datasets that it has learned from. This idea, compounded by dependence on the beam trawl, and the institutional structure of both research styles leading up to the turn of the century, could be a limiting factor in the ability to measure and manage Eastern Pacific groundfish. With the current state of changing oceans, fisheries scientists and management councils will inevitably have to deal with exceedingly expansive datasets gathered with emerging technologies, like computer vision and machine learning. Will the WCGBTS dataset become obsolete like its predecessors? And most importantly, what comes first? The datasets? Or the questions that can be answered by the datasets?

To explore this further we can look to William Pearcy and Waldo Wakefield. Both fisheries scientists and alumni of Oregon State University who worked together closely over their careers, both at OSU and when they were at other institutions or agencies[30]. The sampling Pearcy began in the 70's and 80's was also laying the groundwork for the Newport Hydrographic Line (NH Line) that would become the most sampled area off the Oregon Coast[31,32]. This invisible transect of water that emerges from Newport, Oregon was sampled east to west and near the shore, as opposed to north to south and in deeper water (like most of the previous data from early NMFS surveys). The choice to diverge from the historical lessons of the NMFS surveys is unclear, but sampling in nurseries, where fish grow up, is different than sampling for adults, which is most likely why he made that experimental design choice. It was also not the overt federal charge of the academicians to do work that would directly contribute to stock assessments and management. Pearcy's work on juvenile flatfish was collected and analyzed using a different yardstick (per say) than the NMFS research of the time. Incorporating the type of data that Pearcy was gathering was complicated. Nearshore nurseries for the groundfish fishery were not represented through fisheries-dependent or independent datasets.

Wakefield, who had worked with Pearcy, has always been interested in fish habitat of all varieties. So, when Lorenzo Ciannelli, Oregon State's first official fisheries oceanographer, approached him in 2008 about a study of flatfish and hypoxia he knew the best way to design the study. Since Pearcy had already done work on juvenile flatfish in the 70's and 80's, Wakefield and Ciannelli modeled their data collection methods to mimic Pearcy's sampling. If they could build a consistent dataset, they may be able to incorporate past data and expect more predictive results. Pearcy's earlier studies used that familiar medieval piece of machinery mentioned earlier—the beam trawl. Studies of juvenile flatfish in sandy sediment were conducted many times using the beam trawl and Pearcy published several papers with that data set [33,34]. Wakefield and Ciannelli received funding from Sea Grant and moved forward with their flatfish hypoxia study, using Pearcy's beam trawl method, and expanded his earlier data set dramatically. In 2012, with additional funding from NOAA (the NMFS) where Wakefield was working at the time, a couple small adjustments were made to the classic beam trawl. Wakefield added a tickler chain, to encourage movement. And, he added a video camera.

There were several reasons for this. Wakefield was not only a seasoned ichthyologist, he had also spent a considerable portion of his career working for NURP, a program that focused on teaching researchers how to use emerging marine science technologies. He had also done early work using video to better understand habitat off the coast of California[30]. He brought to the equation a relationship with technology, and emerging ocean sampling methods, that was not well represented off the Oregon coast. His work in California was similar, but used a camera sled that did not produce the amount of dust that the tickler chain did in the Oregon studies. Using video cameras was also emerging in other areas of fisheries, but never in such turbulent waters. The videos were not great quality, and the water was not very clear.

Figure 3: "Murder Chain" Oil on paper. Samm Newton, 2017.

The use of video there marked a departure from older ways of sampling, which depended on counting physical fish in a net. Video added multiple dimensions to the data that could be collected, datasets that were assembled, and subsequently the kinds of ecological questions and inferences that could be developed expanded as well. By adding the camera to the beam trawl, science could ask questions of juvenile flatfish that they were not able to ask before. They could spatially correlate fish catch to certain types of habitat. They could begin classifying the microhabitats that make up nearshore groundfish nurseries. While the possibilities initially seemed endless, they eventually became less clear. The tickler chain was added to encourage the movement of fish, but the cloud of dust it created in the video greatly complicated efforts to analyze the data using any type of machine learning algorithm (Figure 3). The movement of the net and the ship also made analysis difficult. Going through the video frame by frame was incredibly labor and staff intensive. From five years of video data collected, one paper was published, and the data had all been sorted manually, covering only about 1% of the available footage[35]. The cumbersome nature of the amount of the data, paired with the challenging nature of the visual composition of the video had not been anticipated. The result was thousands of hours of data ripe for analysis, but without any clear direction as to what or how information could be gleaned from the video. It created a problem that could be further explored by the machine learning community.

Datasets and Machine Learning

In addition to being key components in the technical performance of algorithms, datasets also have deeper, more profound impacts on shaping the sorts of problems that the machine learning community work on. There is a fundamental interplay between datasets and the direction of technological innovation. Once a dataset is established, it becomes a benchmark that algorithms are evaluated against, and focus research efforts onto the problem(s) posed by the datasets. One of the most successful examples is ImageNet (Figure 4), which arguably launched the deep learning era, and presided over multiple consecutive and substantive gains in image classification accuracy and deep learning innovation. Datasets also offer an opportunity for establishing standardized benchmarks for comparing new methods to, and there are often leaderboards associated with different datasets that offer “objective measures of performance and therefore are important guides for research.” [36]

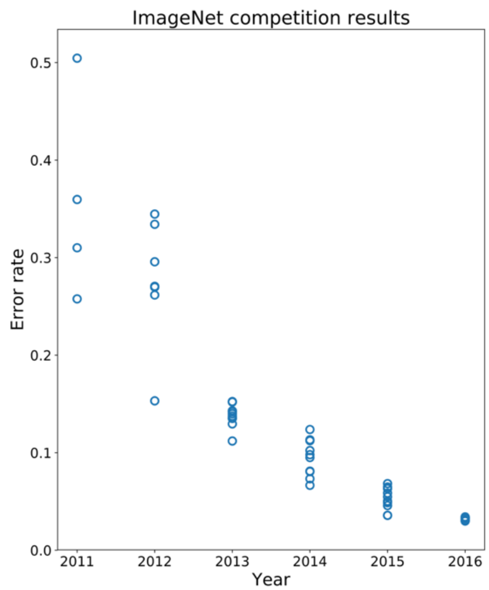


Figure 4: ImageNet classification accuracy error rates of 10 best teams per year (judged on top-5 classification accuracy). Figure from By Gkrusze - Own work, CC BY-SA 4.0, [23]

While datasets are of fundamental importance to machine learning, they can be time and resource intensive to create. ImageNet at its inception in 2009 consisted of 3.2 million labeled images, and as of 2014 had over 14 million images[14,13]. The JFT-300M dataset used consists of over 300 million (weakly annotated) images, and calls for larger datasets still[37]. The cost of creating modern, large datasets can be prohibitive, and is contingent on the task defined by the dataset. Classification tasks are comparatively simple, while video segmentation is notoriously difficult. For example, Cityscapes is a modern semantic segmentation dataset with high-quality pixel-wise instance-level semantic labels; in creating the dataset, "[a]nnotation and quality control required more than 1.5 h on average for a single image" (Cordts, et al., 2016). Cityscape has 5000 high quality frames, and 20000 frames with coarser annotations. Similarly, MS COCO is a modern object instance detection and segmentation dataset, consisting of roughly 2.5 million images[38]. The authors report that using Amazon Mechanical Turk, the image labeling process took roughly 85,000 worker hours to complete. As such, it can be prohibitively expensive to create datasets of sufficient scale for modern methods.

In the deep learning era, large labeled datasets are crucial for model performance. For limited budget or academic institutions, creating custom datasets can be infeasible. Thus, datasets can discourage inquiry in areas lacking suitable datasets, since only institutions of sufficient size or assets can hope to create new ones; everyone else is limited to the datasets on hand. This means that while the presence of a dataset can promote research in a specific domain, the lack of a suitable dataset in a specific domain can have a strong inhibitory effect. Conversely, if a suitable dataset does exist for a task (and made freely available), then people will be drawn towards working on it, since they are limited to working on problems with such a dataset to begin with. Thus, the landscape of datasets shapes the work being done. In the case of work on groundfish behavior and abundance, this influence can also be seen between the cyclical nature of measuring and managing fish discussed in the previous sections.

However, there are still many problems for which no large-scale dataset exists, and niches within existing datasets that are not filled. More concretely, the groundfish project is concerned (in part) with performing tracking-by-detection; however, we are examining Oregon nearshore benthos, and the beamtrawl environment is unique. Although our project falls within a common computer vision task, algorithms trained on existing tracking datasets will not generalize well to beamtrawl videos. As such, we had no recourse other than to build a dataset of our own, and leverage larger datasets via transfer learning for our models.

Recent advancements in deep learning have been fueled (in part) by increased dataset size and using even larger data in tandem with higher complexity models holds promise to continue improving representation learning[37]. Our proposed dataset, Newport Hydrographic Groundfish, or **NHFish**, is the latest in a long line of video tracking datasets. Video tracking is one of the fundamental tasks in video analysis pipelines, as it allows for forming temporally consistent sequence information between frames in a video. For our particular application, we had some unique challenges for our videos that make them distinct from existing datasets. Namely, all our data is captured underwater with specific, fixed scene elements. More specifically, the lighting is supplied by a spotlight mounted on the camera rig, making the center regions bright, and the corners comparatively dark. Likewise, shadows are cast in a consistent manner, and the camera is a relatively consistent distance from the ocean floor. The Oregon benthic habitats sampled are all sandy-bottomed, with colors largely in hues of brown and green. Finally, our targets for tracking in the video are flatfish, which can both burrow into the sand, and have evolved to be well camouflaged against the ocean floor. This makes appearance features less informative, and motion cues more important. Similarly, there are frequent and disruptive visual distractors, since the sampling method relies on dragging a chain to disturb the fish (to induce motion). These factors make our video collection distinctive in the video tracking dataset landscape.

III. Machine Learning and Ecological Inference

Eastern Pacific groundfish

Previous research has been done analyzing the communities of juvenile flatfishes and other small demersal fishes in Oregon estuaries and off the central Oregon coast[39–41]. However, nearshore juvenile demersal fish assemblages are still not fully understood and there is a knowledge gap about the abundance and distribution of fish between the larval and adult stages of these commercially-important species[42–44]. In order to better understand these fish communities, the Northwest Fisheries Science Center, in collaboration with the Pacific States Marine Fisheries Commission and Oregon State University, have been conducting an ongoing study since the summer of 2008[44–46]. This study has been sampling and analyzing newly-settled young-of-the-year groundfishes and other small demersal fishes in the nearshore nursery habitat along the central Oregon coast, specifically along the Newport Hydrographic (NH) line. This sampling area was chosen because nearshore habitats are known to be important nursery grounds for settling larval fish since they act as a “first stop” for these fish as they leave the larval pelagic stage and enter into an estuarine or nearshore open ocean nursery habitat as juveniles[40,47]. By better understanding the juvenile demersal fish communities, the knowledge gap that exists about the abundance and distribution of the early life history stages of these species can be filled. This knowledge about the early life history of these fishes can help with the estimation of year-class strength, which is important for fisheries stock assessments and management decisions[44,48,49].

In addition to sampling these young-of-the-year groundfish, this study has attached an high-definition underwater video camera to the sampling gear in order to see the seafloor and supplement the catch data with habitat and behavioral data[46,50]. These videos provide a method of addressing ecological questions about newly-settled groundfish and other small demersal fishes, such as microhabitat usage, net avoidance, and fish behavior, which are inaccessible otherwise[46,50–54]. Understanding microhabitat usage is important since microhabitats such as biogenic depressions, sand ripple/wave crests, and shell debris have been found to provide locations for demersal fish to avoidance predators, capture prey, and forage[51]. Additionally, evaluating net avoidance is important for bycatch reduction and for evaluating the catchability of some fish species, thus the accuracy or bias of catch surveys[53]. Furthermore, examining fish behavior is important for many reasons such as understanding how fish will respond to changing environmental conditions or to new underwater technology[46,54]. In addition to these ecological questions, these videos could be beneficial to fishery stock assessments by supplementing their available data of fish abundance.

Automating the Video Analysis Process

Addressing these ecological questions and supplementing stock assessments by manually analyzing the video material has proven to require excessive man hours. The high level of manual effort required to process the videos has been a limiting factor in their use in scientific research and inquiry (i.e. after 2 decades of collecting NH-line beam trawl video, there has been just 1 published paper[46] using its data, since it involved manually measuring characteristics of the fish from the video frames). This sort of activity can only be done by skilled marine ecologists, so the labor burden is especially onerous. Being able to perform various video analysis tasks without manual work by scientists would greatly expedite the process of gleaning information from the videos and would open the door to expanded video collection and analysis. Automating fisheries video data would "transform marine science"[55]. Therefore, for the sake of efficiency and a more thorough analysis, an automated process using computer vision and machine learning has been sought out. Although there is a wealth of potential questions to be asked of the beam trawl video, we have settled on the task of fish counting for this work. The reason for that is twofold:

1. Computing accurate fish counts is of vital importance to fisheries science and management, as it is tantamount to performing stock assessments.
2. The technical bedrock laid for performing fish counting is easily transferable to a wide array of analysis techniques pertinent to many other salient ecological questions.

Fisheries science makes heavy usage of predictive models, and sound fisheries management is heavily reliant on accurate population dynamics predictions[56]. However, counting the number of fish in a given area is a difficult, expensive, and error-prone undertaking. Where this project can supplement stock assessments is by providing new and more accessible data about fish abundance by counting and tracking fish in the videos using computer vision. Since available data is one restricting factor of the performance of stock assessments as a management tool, this project could reduce the risk and uncertainty associated with fisheries management. Risk and uncertainty can never be fully eliminated from fisheries management, but incremental reductions can still have a large impact on the social, ecological, and economic components of the fishery.

During this process of using computer vision to count and track fish, we will also be making our own new dataset. We hope that by making the video dataset available, fisheries science and management will benefit by having greater attention from the computer vision community. Although there is considerable interest in performing automated video analysis in unconstrained underwater environments, the state-of-the-art methods employed in ecology often lag behind those proposed in the computer vision community. By having a dataset that straddles both, we hope that new and novel CV approaches to its challenging data will be proposed and the barrier between integration of pure CV and applied ecological work will become closer.

Mathematical Models and Stock Assessment

With the purpose of fisheries management being to ensure sustainable fishing practices, stock assessment is a quantification tool used to predict how fish populations will react to different management decisions[57]. Fisheries stock assessments are mathematical models that are used to simulate the population dynamics of a fishery, make predictions about trends in the population through time based on policy decisions and expected environmental or biological changes, and take into account the fishermen's and industries' response to management decisions[57]. When these models, and fisheries science as a whole, were in their infancy, stock assessments usually focused on two questions: 1) "what is the optimum [fishing] effort?", and 2) "what is the maximum sustainable yield?"[57] There are two main issues with using stock assessments to only address the questions of "best" fishing effort and "maximum sustainable yield". The first is that the "maximum sustainable yield" can only be determined after it has been exceeded because it is the apex of an inverted U-curve, and therefore, one can only know that the maximum has been reached once the yield has begun to decrease (Figure 5)[57]. The second issue is that reducing fishing effort is one of the most difficult tasks to accomplish in fisheries management since it requires either the reduction in fishermen's catches or the reduction in the amount of fishermen, which translates to either the reduction of income or the amount of jobs available[57]. Although these two questions are still the main focus of some stock assessments, ideally a stock assessment would instead be used as a tool to focus on designing a fisheries management system that can adapt to changing ocean conditions, natural variable, and inherent uncertainty associated with fisheries and marine systems[57].

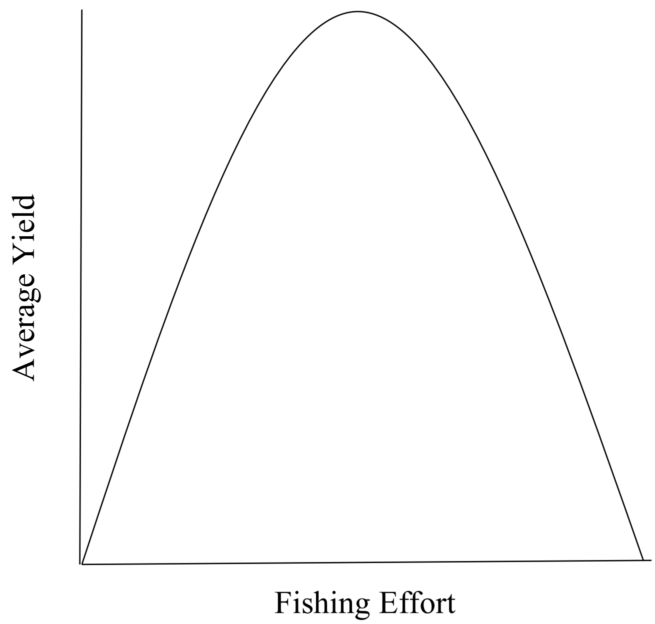


Figure 5: Theoretical relationship between the fishing effort and the average yield of a fishery. Adapted from Hilborn, R. & Walters, C. J. Role of Stock Assessment in Fisheries Management. in Quantitative fisheries stock assessment: choice, dynamics and uncertainty 3–21 (Kluwer Academic Publishers, 2001).

Before diving into how these tools of fisheries management (stock assessments) work, one must first understand the basics of mathematical models. A mathematical model is a mathematical equation with variables and parameters that attempts to represent a phenomenon or process[58]. The variables in a mathematical model represent something in nature that can be measured or defined, whereas the parameters are estimated values that both modify the impact of variables and establish the structure of the relationship between variables within the model[58]. These models come in many different types, and someone constructing a model (referred to as a modeler throughout) must make many choices about what type to use and what to include or exclude in the model[58]. The following are pairs of counterparts that each represent a characteristic of a model, and each is a choice that must be made by the modeler: descriptive or explanatory, realistic or idealistic, general or particular, deterministic or stochastic, and continuous or discrete[58]. These different characteristics refer to what the model can tell someone about the phenomenon or process it is modelling, what it is modeling, the scope of the model, and what kind of output the model delivers based on the types of variables or parameters it uses as well as its structure[58]. The purpose of the model is usually a driving factor in determining which characteristics it should possess.

Stock assessment models have the underlying purpose of using available data on population dynamics, fishing effort, and catch to predict future fish populations based on the fisheries response to natural dynamic processes as well as to fishing pressure[58]. Therefore, usually a modern stock assessment model is realistic, stochastic, and discrete; whether it is descriptive, explanatory, general, or particular depends on the data available and the application of the assessment[58]. However, these models are only simulations, so they will always some level of inaccuracy and uncertainty associated with them[57,58]. Additionally, they can only predict future possibilities based on the data available, so they do not perform well for new, unseen management strategies or unprecedented natural variations[57]. Again, stock assessments are only a tool and they have limitations[57,58]. They should not be used to decide policy and management decisions such as catch limits directly, but instead to make predictions about future fish populations and their response to management choices and natural variation[57]. These predictions can then be used as one piece to consider in the policy decision making process, along with other important factors affecting the final decision, such as stakeholder values[57].

Video Analysis with Computer Vision

Visual tracking is defined by as "the analysis of video sequences for the purpose of establishing the location of the target over a sequence of frames (time)."[59] Visual tracking is an essential component in video analysis, as it is the primary method for integrating temporal information. Algorithms devoid of tracking treat each frame as being independent; while videos themselves are simply collections of frames, the frame contents generally exhibit some degree of temporal continuity. That is, assuming a standard video frame rate, objects generally move through a video in a consistent manner with smooth spatial and temporal continuity. In the case of the beam trawl videos, frames are captured at 24 frames per second (FPS), meaning there are roughly 42.67ms between each frame. Although fish can move very quickly, 43ms is a very short duration. As such, due to the high capture rate of video frames, objects in the scene appear relatively regularly in the video, with (approximately) piecewise linear movements. This allows using information inferred in a frame to inform inference in other frames within a temporal window. Therefore, treating each frame as independent in a video is an incorrect assumption. At best, the assumption is inefficient (since treating frames as independent discards useful information), and at worst it limits the information that can be gleaned from video (maintaining a consistent identity through time is important for accurate scene characterization, e.g. counting the number of distinct objects in the video sequence).

Tracking links detections through time into a construct called a 'track', consisting of detections that belong to a single object through a temporally sequential collection of frames (Figure 6). There are multiple formulations of tracking; one of the most common is tracking-by-detection, where one or more separate detection algorithms provides noisy bounding boxes of potential salient objects in the scene, and the tracker builds tracks from the proposed detections. This is the formulation we use for beam trawl video analysis problem, and we present the challenging beam trawl video dataset as a tracking-by-detection video dataset in a similar vein to MOT[60,61]. The beam trawl dataset fits into the multi-object tracking paradigm, but features some new and interesting challenges not seen in contemporary tracking datasets. Namely, the data has many visual distractors, the target objects (flatfish) are naturally camouflaged in benthic environments, there is non-linear camera and object motion, significant object appearance deformation, and relatively homogenous object appearance. As such, motion cues and temporal continuity become important for building accurate tracks, and appearance models may be less useful. In addition, unlike most tracking datasets, this one does not feature humans as a detection target, reducing the utility of trackers such as DPM and potentially motivating new and novel solutions[62]. More details regarding the dataset are presented in the *NHFish Video Dataset* section below. Despite the apparent difficulty of the beam trawl dataset, there is much benefit to be gleaned from automating its analysis, as was previously discussed.

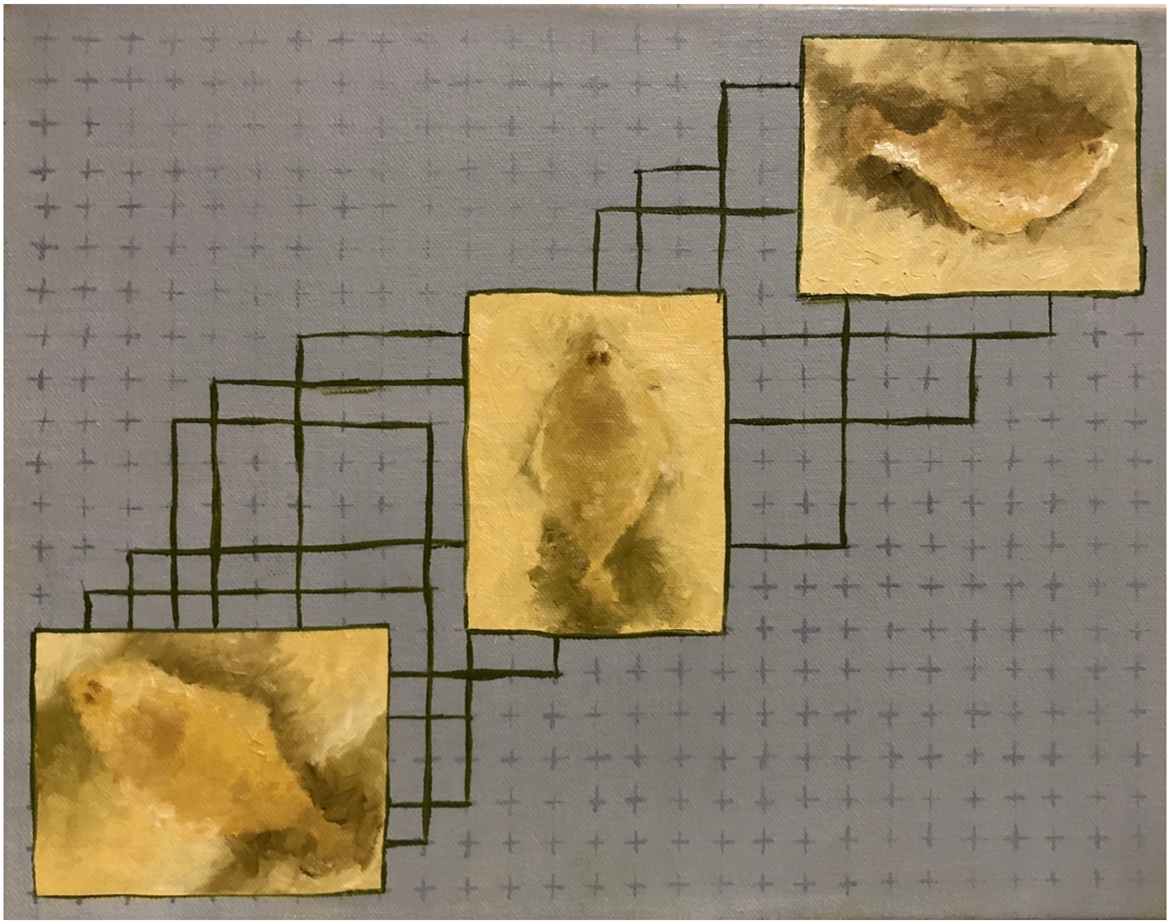


Figure 6: "Track" Oil on canvas. Samm Newton, 2018.

Not Reinventing the Wheel with Computer Vision

Computer vision has a long history of applied work in the natural sciences; Waldchen et al. provides a comprehensive survey of computer vision for plant species identification, Pun et al. gives an overview of computer vision for medical imaging, Ko et al. describes computer vision use in natural disaster warning and detection systems[63,64,65]. There are multiple science-oriented computer vision tool boxes, such as WildBox for animal counting and identification, TMARKER for cell counting and staining analysis, and BioTracker and trackdem for visual tracking of animal populations, among others[66,67,68,69.] In general, computer vision has the potential to reduce manual workloads and both vertically and horizontally scale out the scope and power of natural science inquiry when repetitive visual tasks are required. Visual tracking is especially important for many ecological tasks, as it enables making ecological inferences, such as population counting, individual behaviors, group dynamics, etc. There are numerous specific works of applied visual tracking, such as honey bees in controlled environments[70] and in the wild[71], elephant detection and tracking[72,73] and most pertinent to this project, analyzing fish in underwater conditions. For example, Siddiquie et al. used pre-trained CNNs (convolutional neural networks) for fish classification on underwater images from Australia, and Hossain et al.

used GMMs for detection, SVMs for classification, and Kalman filters for tracking[74,75]. Rodriguez et al. uses a 3D optical stereo matching approach for fish segmentation[76]. Chuang et al. used a deformable parts kernel matching tracker for underwater fish tracking[77]. In addition, there is interest from the regulatory agencies, such as NOAA, regarding applications of automated recognition systems in fishery services[78]. Despite the strong interest, the primary existing dataset is ImageCLEF, but it primarily consists of images of tropical fish in less turbid waters, and is a poor fit for our specific task[79]. The Fish4Knowledge project started in 2010, and examined detection and tracking in a large corpus of underwater tropical coral reef fish[80]. Unfortunately, the project is since defunct, the cameras were stationary, and the underwater environment differed significantly from the beam trawl dataset. In keeping, despite the corpus of work on underwater fish video analysis, none of the existing approaches are suitable to our NH-line beam trawl video data. Hossain's reliance on background subtraction is only feasible for stationary cameras, Rodriguez requires a different multi-camera arrangement and was performed on data from a controlled (pool) environment[75,76]. Chuang heavily relies on appearance features, which will likely be unsuited to Oregon benthos[77]. The closest approach is detailed in Wäldchen's technique, but transfer learning is heavily reliant on testing and training datasets being sufficiently close; none of the existing datasets are remotely close to the beam trawl video, so transfer learning performance would suffer[63]. Shafait and Jäger are two recent works dealing with video detection and tracking in unconstrained underwater environments[55,81]. Shafait performs species recognition in video sequences, but does not address detection and tracking, and uses ImageCLEF for training and evaluations[55]. Jäger uses a two-stage graph-based tracking formulation over learned CNN features, calculating tracklets in a shortest-path problem formulation, then grouping tracklets with an affinity metric based on L2-distance of their CNN features[81]. In all, existing work on underwater fish video analysis is promising, but taken by itself, insufficient for our dataset.

However, more recent works in detection and tracking have significantly improved over the past state-of-the-art methods. PASCAL VOC[82] and MS COCO[38] are recent datasets focused on image detection methods, and have helped drive the field of object detection forward in recent years. Broadly, modern convolutional object detection algorithms can be divided into groups, one-stage and two-stage detectors.

Some of the most successful and popular algorithms in recent years have stemmed from R-CNN[83]. Proposed in 2014, it has helped to inspire numerous subsequent research, and the lineage of R-CNN, Fast R-CNN[84], Faster R-CNN[85], and Mask R-CNN[86] have all been highly successful as two-stage detection algorithms. R-CNN uses selective search to propose ~2K differently-sized bounding boxes on the frame, resizing each to a common resolution and running them through a CNN sequentially, and classifying each region as containing an object or not[87]. Although state-of-the-art in 2014, it was also extremely slow, due to passing each ROIs through the CNN individually. Fast R-CNN improved upon R-CNN by computing features over the entire frame once, then proposing bounding boxes in feature space with selective search. Each feature ROI is then pooled to a common resolution, and two outputs are generated from a network, one from a softmax classification layer and the other from a bounding box regressor to classify whether any given box contains an object, and to adjust the bounding boxes to accurately localize in image space respectively. Faster R-CNN builds on insights from Fast R-CNN and does away with selective search, and instead uses a Region Proposal Network (RPN) to propose bounding boxes over the image CNN features of varying aspect ratios and scales. Faster R-CNN proved to be extremely successful and highly influential; as of November 2016, "half of the submissions to the COCO object detection server [...] are reported to be based on the Faster R-CNN system in some way."[88] One of the detection proposal methods used in this work is a Faster R-CNN implementation. Mask R-CNN uses Faster R-CNN, but it adds another output head to perform segmentation (pixel-wise labeling) rather than just bounding boxes, and it also adds a more accurate feature pooling method than the one Fast R-CNN and Faster R-CNN used (Figure 7). These additions pushed Mask R-CNN to be among the most accurate methods for detection (and instance semantic segmentation) on COCO through 2017, and it remains highly competitive currently. Unfortunately, Mask R-CNN requires segmentation annotations for offline training; segmentation masks are expensive and laborious to create, and our beam trawl video dataset lacks them, so evaluating Mask R-CNN as a detection proposal method was not feasible.

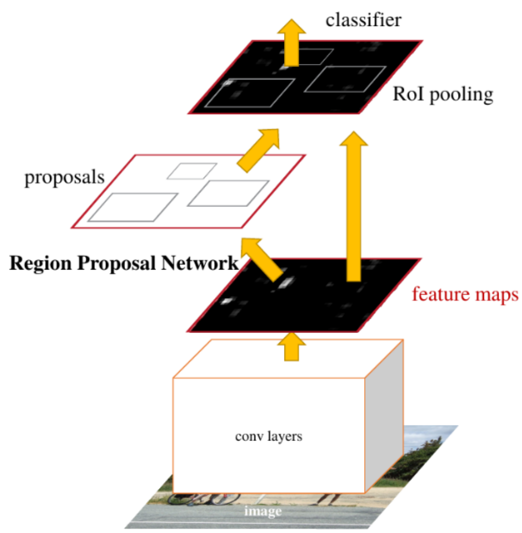


Figure 7: Faster R-CNN system architecture diagram[85]

Although two-stage detection methods have enjoyed considerable success in recent years, there have also been considerable effort in developing 1-stage detectors. 1-stage detectors are often faster, albeit at the cost of lower performance on average. Overfeat was an early (2013) deep-learning approach for detection, using a CNN to perform classification, localization and detection using multiscale sliding windows, in which the predicted bounding boxes are accumulated to build detection confidence[89]. YOLO was another impressive 1-stage detector, formulating detection as a regression task from image space to bounding boxes, enjoying reasonable performance at real-time speeds[90]. SSD improved on YOLO's runtime and accuracy performance by defining a fixed-size set of bounding boxes, predicting likelihoods of each box containing an object, and refining probable boxes for more accurate localizations, while doing so at multiple scales[91]. Retinanet formulates a novel loss function that rectifies the foreground versus background pixel imbalance (i.e. on average, the majority of pixels in an image belong to scene elements not of interest)[92]. The authors couple this loss (termed 'focal loss') with a relatively simple detection network architecture and achieve state-of-the-art runtime and accuracy performance. Due to its impressive performance, RetinaNet is the other detection proposal method used in this work.

Computer vision historically has been deployed in controlled environments, limiting its use to recognizing dead fish (e.g. Strachan et al.).[93] More recent work has attempted to tackle the problem of live fish in uncontrolled environments (e.g. Shafait et al.), a much harder problem[55]. In the context of computer vision problems, fish analysis tasks fall under detection and tracking (as done in Shafait et al. using a particle filter for tracking, or Wang et al.[94] which used keypoint descriptor tracking for movement pattern analysis), or classification done on a per-frame basis (as in Rova et al.[95], which used an SVM classifier with deformable template matching, and Qin et al.[96] which uses sparse, low-rank matrix decomposition for foreground extraction, computes deep features over the foreground, and classifies fish species per-frame in video with an SVM). Spampinato et al. performed both classification and tracking, using shape and texture features[97]. Spampinato also proposed a system for detection, tracking, and counting fish in unconstrained underwater video[98]. Although highly related to our task, their work is over a decade old, evaluated their algorithm on tropical fish as part of the now-defunct EcoGrid project, and used CamShift as the tracking algorithm of choice, which has long since been surpassed by other multi-object tracking algorithms[99].

Among computer vision tasks, detection and tracking is more challenging than classification, as they involve performing localization (for detection) and integrating temporal information (for tracking), both of which are absent in classification. In keeping, detection and tracking can reveal more information regarding fish behavior and are important for properly utilizing video data rather than still images. In addition, another fundamental computer vision task is segmentation; this has even more stringent localization requirements than detection, as it involves per-pixel labeling, rather than bounding boxes. Each task can allow asking different ecological questions from any given data.

For a (non-exhaustive) list:

- Classification:
  - Identifying the presence of specific fish species in an image, identifying habitats.
- Detection:
  - Localizing different fish in an image, counting the number of fish. Prerequisite for performing tracking
- Tracking:
  - Counting the number of fish in a video sequence, analyzing fish trajectories or response times
- Segmentation:
  - Analyzing specific fish behaviors, detailed fish sizes and potential interactions

Tracking is a fundamental operation for processing video data, as it unifies per-frame information along the temporal dimension. Namely, detection and classification tasks are performed on a per-frame basis, but only provide information for that given frame. Each frame is implicitly treated as being independent. As such, inferring temporal continuity between frames is not supported. In contrast, tracking is the process of associating information computed between frames within a temporal window. As such, without utilizing tracking, the types of information that can be gleaned from video data is limited. For example, computing accurate counts of the number of fish in a video requires detection and tracking. Detection will provide the number of fish in any given frame, but to compute an accurate count along the entire video, it must be possible to discern new fish from old ones. Concretely, frame t can have 1 fish, and frame t+1 can also have 1 fish, but if the fish from frame t exits the scene, and a new fish enter the scene in t+1, then the sequence contains 2 distinct fish. If simply performing per-frame detection, it appears there is only 1 fish in the sequence. Differentiating between new fish and old fish can only be done if the frames are not assumed to be independent, which in turn necessitates tracking. There are also more insidious issues, such as a fish exiting the field of view, then re-entering at a later date. Identifying these occurrences would require computing per-fish classifiers, building a bank of 'known fish' within a video sequence, and attempting to match any fish entering the scene against the list of 'known fish'. However, this itself can be error prone, especially in cases where the fish are not visually distinctive, are within the field of view for a short period of time (some fish in the beam trawl videos may be in the scene for less than 5 frames, which is equivalent to 200ms), or are only seen from limited angles and exhibit significant appearance variation based on viewpoint.

Sampling Procedure

Sampling of the NH line was conducted monthly from July 2012 to July 2018 with a few months omitted due to poor oceanographic conditions. Sampling sites consisted of six stations categorized by depth (Figure 8). These stations included two off Moolack Beach at depths of 30 and 40m (MB-30m and MB-40m), three along the NH line at NH-3 (~50m), NH-5 (~60m), NH-10 (~80m), and a station (~100m) north of Stonewall Bank used as a replacement for NH-15 which is too rocky to trawl. The Moolack Beach stations were used instead of NH-1 due to rocky outcrops. The stations were chosen based on depth (~30, 40, 50, 60, 80 m). Commercial fishermen operating the vessel were also consulted when determining the locations of the stations in order to find the best soft-sediment areas to perform beam trawl sampling. At each station sampled, a beam trawl tow of approximately ten minutes was conducted. A Conductivity, Temperature, and Depth profile (CTD; Seabird SBE 19 CTD) was also taken that included sensors for dissolved oxygen, light scattering transmission, and chlorophyll fluorescence.

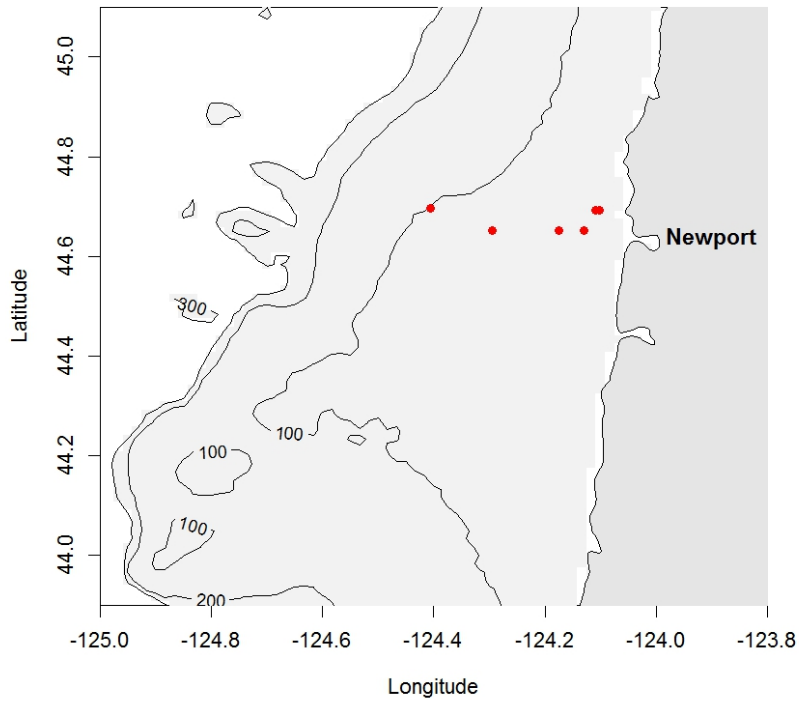


Figure 8: Locations of sampling stations along the Newport Hydrographic (NH) Line designated by the red points. Black lines represent the 100, 200, and 300-meter bathymetric boundaries.

The beam trawl was comprised of a galvanized steel frame with two 0.5-meter tall sled-like runners equipped with a paired odometer wheel system for measuring the distance sampled, a 2-meter long beam separating the two runners, a net attached to the two trailing edges of the runners, and a tickler chain attached to the front of the runners (Figure 9). The net of the beam trawl was composed of shrimp trawl webbing, which was lined throughout with a 2.5 X 3mm mesh liner. This beam trawl gear was constructed in such a way as to target small demersal fish and avoid sampling large adults. The beam trawl was also equipped with a high-definition video camera system with scaling lasers and two L.E.D. lights that allowed for a video image of the seafloor to be taken while sampling.

The catch from each beam trawl tow was sorted at sea, with fish 151mm or more in length being identified to species, measured to the nearest mm in length (standard length, SL), and discarded (Figure 10). All other fish less than 150 mm SL were flash frozen at sea using dry ice and then stored in the lab at -80°C until processed. These fish were then identified back at the lab to the lowest taxonomic level (species in most cases), and their total length and wet weight were recorded (Figure 10). This catch data was not used for this project due to the scope of our transdisciplinary goals.

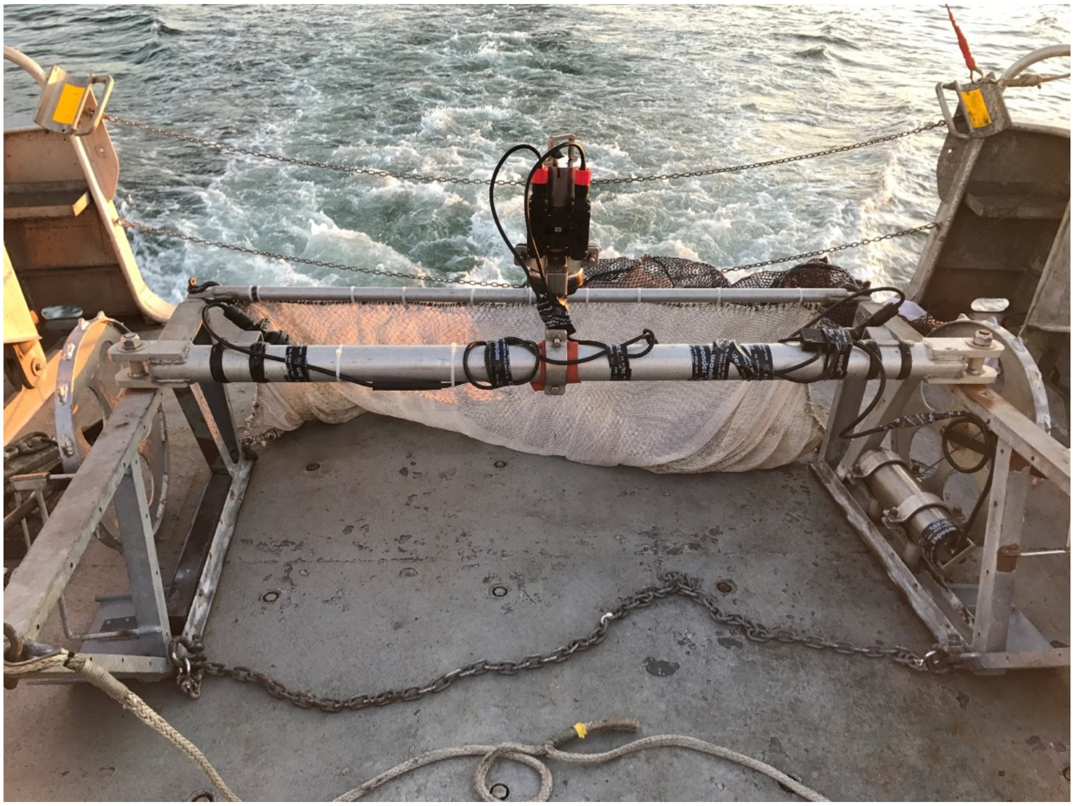


Figure 9: Galvanized steel beam trawl used for sampling along the Newport Hydrographic (NH) line from July 2012 - July 2018. Essential components shown include: 0.5-meter tall sled-like runners, paired odometer wheel system, 2-meter long beam, shrimp trawl webbing net with 2.5x3 mm mesh liner, tickler chain, high-definition video camera within waterproof housing, two L.E.D. lights, and scaling lasers.

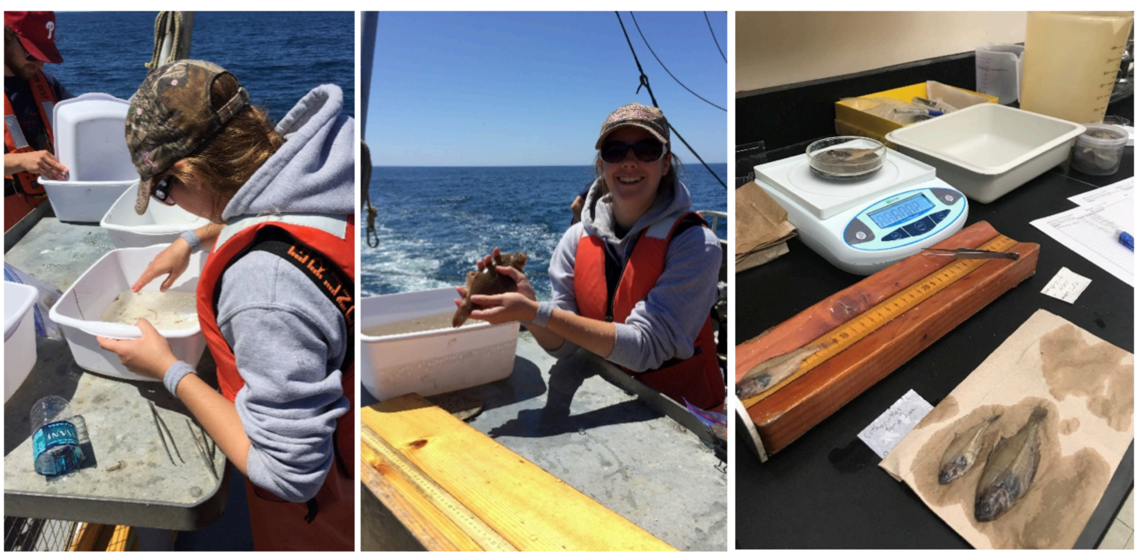


Figure 10: Sorting catch at sea (left) with fish 151 mm or longer identified and measured before being discarded (middle). Fish less than 151 mm in length were flash frozen and processed in the lab, where they were identified to lowest possible taxonomic level, weighed, and measured (right).

NHFish Video Dataset

DAVIS-2016 is a state-of-the-art video object segmentation dataset and consists of 50 high-resolution video sequences with densely annotated ground truth segmentations. DAVIS-2017 expanded on DAVIS-2016, subsuming the sequences from DAVIS-2016 and adding another 100 sequences, and introducing instance annotations. DAVIS-2017 has a total of 10459 frames and 376 different instance objects. Due to the DAVIS being a successful and relatively new video dataset, we borrow some of the analysis performed in DAVIS for our dataset below. For comparison, the DAVIS 2016 dataset's has a compendium of challenging factors; the beam trawl data exhibits nearly all of them (except HO; rather, the general lack of color variation in the beam trawl videos is actually a problem, and MB, since the video is captured at a sufficiently high frame rate)[100]. For reference, a (lightly edited) list of the attributes from DAVIS are duplicated here:

- (BC) Background Clutter. The back- and foreground regions around the object boundaries have similar colors (χ 2 over histograms).
- (DEF) Deformation. Object undergoes complex, non-rigid deformations.
- (MB) Motion Blur. Object has fuzzy boundaries due to fast motion.
- (FM) Fast-Motion. The average, per-frame object motion, computed as centroids Euclidean distance, is larger than τ f m = 20 pixels.
- (LR) Low Resolution. The ratio between the average object bounding-box area and the image area is smaller than t lr = 0.1.
- (OCC) Occlusion. Object becomes partially or fully occluded.
- (OV) Out-of-view. Object is partially clipped by the image boundaries.
- (SV) Scale-Variation. The area ratio among any pair of bounding-boxes enclosing the target object is smaller than τ sv = 0.5.
- (AC) Appearance Change. Noticeable appearance variation, due to illumination changes and relative camera-object rotation.
- (EA) Edge Ambiguity. Unreliable edge detection. The average ground-truth edge probability (using [48]) is smaller than τ e = 0.5.
- (CS) Camera-Shake. Footage displays non-negligible vibrations.
- (HO) Heterogeneus Object. Object regions have distinct colors.
- (IO) Interacting Objects. The target object is an ensemble of multiple, spatially-connected objects (e.g. mother with stroller).
- (DB) Dynamic Background. Background regions move or deform.
- (SC) Shape Complexity. The object has complex boundaries such as thin parts and holes.

In addition to exhibiting the above difficulties, the beam trawl dataset introduces some new challenges of its own. Namely, the video sequences are often characterized by a lack of color variation; that is, everything along the ocean floor are shades of brown, green and grey. See Figure 11 for a sampling of frames from the NHFish dataset sequences. In addition, we are primarily surveying flatfish, which have evolved natural camouflage with respect to the ocean floor and are themselves not as colorful or visually distinctive as, e.g., tropical fish. As such, appearance-based features, which are typically the most discriminative, are of limited use. In addition, there are some other problems that do not show up in DAVIS: 1) Non-linear Camera and object motion (related to CS, but concerns camera motion), and 2) Environmental camera lens occlusion / obscuring: in some video samples, water enters the camera housing, creating water droplets on the camera lens and distorts parts of the frame.

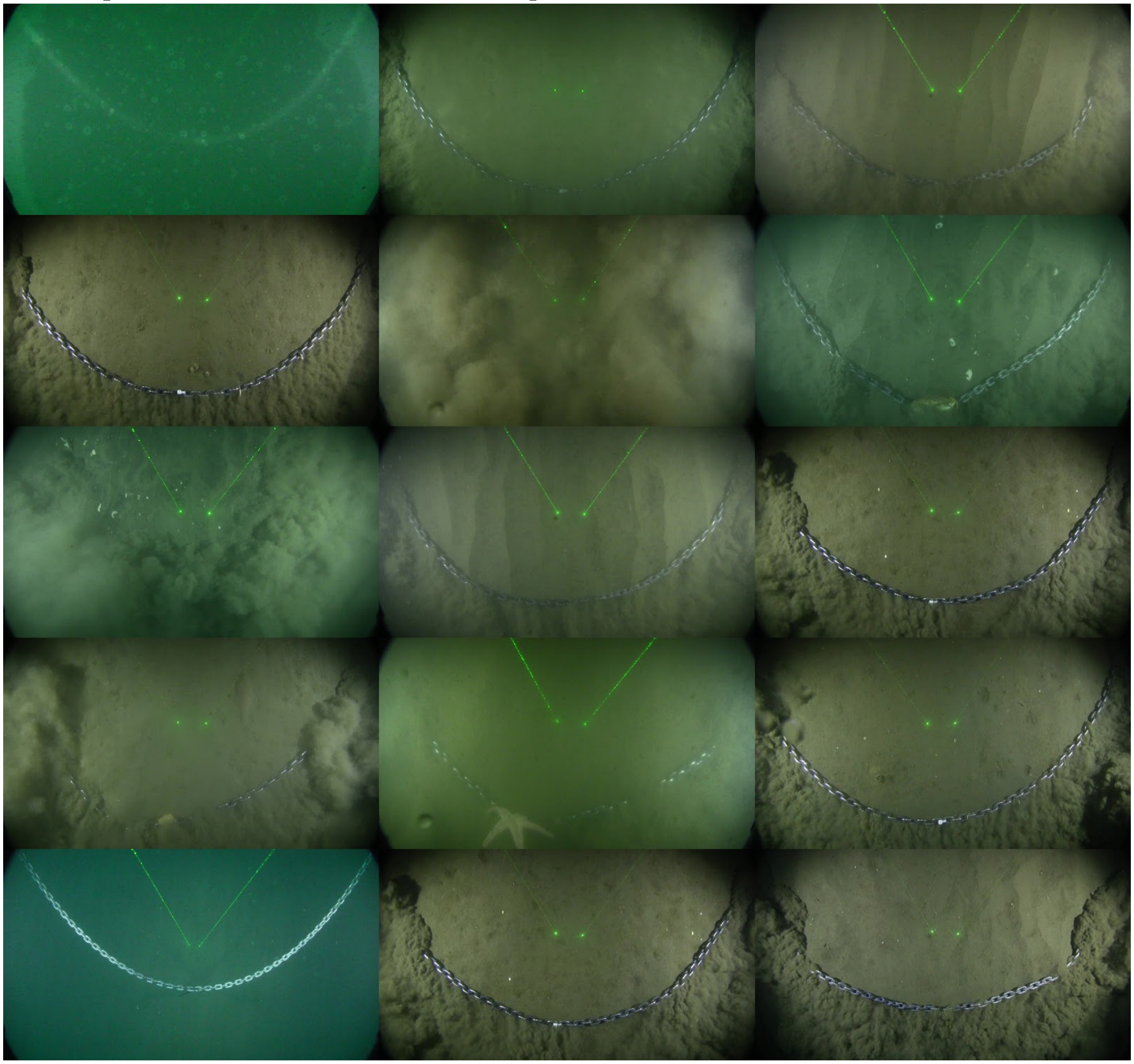


Figure 11: Sampling of frames from different sequences used in the training and validation splits of NHFish

Even within the various categories shared with DAVIS, the beam trawl video data is significantly more difficult than what is present in DAVIS (most notably, BC, OCC, OV, AC, and DB) which is the current SOA video object segmentation dataset. While segmentation would allow semantically richer ecological inferences, in light of the difficulties posed by the beam trawl video data, we elected to opt for a slightly simpler CV task, of performing detection and tracking. Detection and tracking allow us to still infer information about fish trajectories, coarse behavior information, and compute population counts, all of which are both valuable and, if performed manually by a human, work-intensive undertakings.

However, our dataset is relatively unique in the multi-object tracking space. Most efforts in tracking center around detecting and tracking humans, such as MOT[60,61], cityscapes[101], and PETS[102]. This is likely in part because most practical applications of video tracking deal with humans, from self-driving cars to surveillance applications. As such, there has been considerable interest in human-specific problems, with tertiary semantic categories, such as cars and bicycles. However, these are all a far cry from flatfish in turbid waters. Similarly, many contemporary MOT datasets consist of many objects; in contrast, the beam trawl video has relatively few objects in the scene concurrently, and indeed has long stretches of video in which there are no flatfish at all. However, there are many persistent visual distractors (notably, the dust cloud produced by the trawl chain), which can often resemble salient objects. As such, there is a strong class imbalance in the dataset, with relatively few true positives, and many potential (and convincing) false positives. In addition, flatfish are naturally camouflaged, and often have a predilection for partially burying themselves in the sand. This makes detection a very challenging task, as the target objects are, by evolution, visually non-distinctive at best, and not visible at worst. This is why the trawl chain is of importance; the chain induces motion in the fish, and fish motion plays a large role in the ability to identify them in the scene. By the same token, the trawl chain is the cause of the dust cloud, which is the primary visual distractor. In addition, an unfortunate side-effect is that there is a strong likelihood of the fish becoming visible at the dust cloud. This is because the fish will often be partially buried in the sand until it makes contact with the chain, and only begins swimming at that point. As such, fish are probabilistically more likely to appear in or around the visual distractors in the scene, which further complicates accurately detecting and tracking them (Figure 12).

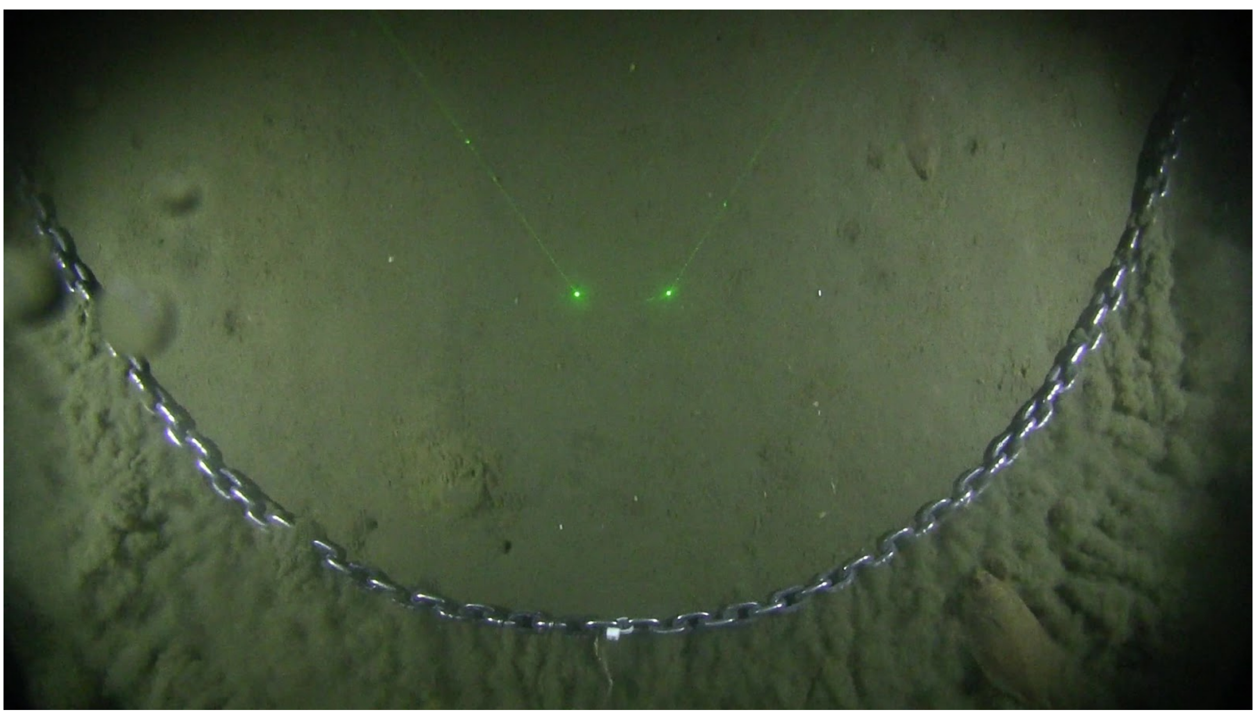


Figure 12: Screen grab from NHFish dataset displaying a fish in the dust cloud, a visual distractor, and another fish partially buried in the sand before the chain makes contact.

Dataset Technical Aspects

Although our main contribution is the labeled beam trawl video dataset, we also present preliminary results of applying automated analytic methods to the dataset, in order to assess the feasibility of current methods in the beam trawl domain. We adopt a tracking-by-detection paradigm for analyzing the beam trawl video, using two standard state-of-the-art detectors, Faster R-CNN and RetinaNet[85,92]. We use the official open source implementations provided at the Facebook AI repository https://github.com/facebookresearch/Detectron. We use base network models trained on MS COCO detection, and fine-tune the networks with the labeled beam trawl video datasets. The feature extraction backbone networks are resnet-101 and resnet-50 for Faster R-CNN and RetinaNet respectively[103]. Detection proposals are extracted for all the labeled video frames and stored to disk as part of the dataset. We format the beam trawl dataset into the COCO json format, as it has become a de-facto standard for detection tasks and integrates easily with many existing codebases. Since we believe the dataset would primarily be of interest to the computer vision community as a tracking dataset, we frame it primarily from a tracking-by-detection perspective. We drew inspiration from the MOT dataset, which is the current primary multi-object tracking dataset in pre-computing detection proposals and including them as part of the dataset[60,61.]

We formulate our analysis as a multi-target tracking-by-detection problem and are primarily looking to solve the multi-object tracking problem in the beam trawl video. Although our main contribution in this domain is not a new or novel processing algorithm, we are still looking to assess the viability of current computer vision techniques in performing ecological inference on the dataset. As such, we apply well known and accurate detection methods for generating detection proposals, and tried using MHT-DAM, a state-of-the-art multi-target tracking algorithm for tracking.

Results and Evaluation

We first present the hand-annotated ground truth data first, then the experimental detection proposals accuracy. There are a total of 19,901 annotated bounding boxes and nearly 218K frames across the train and validation splits of the NHFish dataset. Specifically, each split has:

- Training Set: 9361 annotations, across nearly 143K frames
- Validation Set: 10540 annotations, across nearly 75000 frames.

Table 1 Descriptive statistics about ground truth bounding box annotations

Ground Truth Metrics	Value
Mean Area	39929 pixels
Median Area	35534 pixels
Mode Area	32760 pixels
Median Height, Width	171, 206 pixels
Height, Width Std.	75.46, 74.31 pixels
Skew, Kurtosis	4.62, 38.11

There are varying weather and ocean conditions across the sequences, and the data is drawn from trawls performed in different months in 2013 and 2014 along the NH Line in Oregon. This provides a representative sample of the data produced by the beam trawls. Example frames from different sequences used in the dataset are given in Figure 13. Statistics regarding dataset ground truth bounding boxes are given in Table 1. The training and validation splits have ground truth data, while applying algorithms in the wild would involve evaluating on unlabeled, previously un-seen sequences.

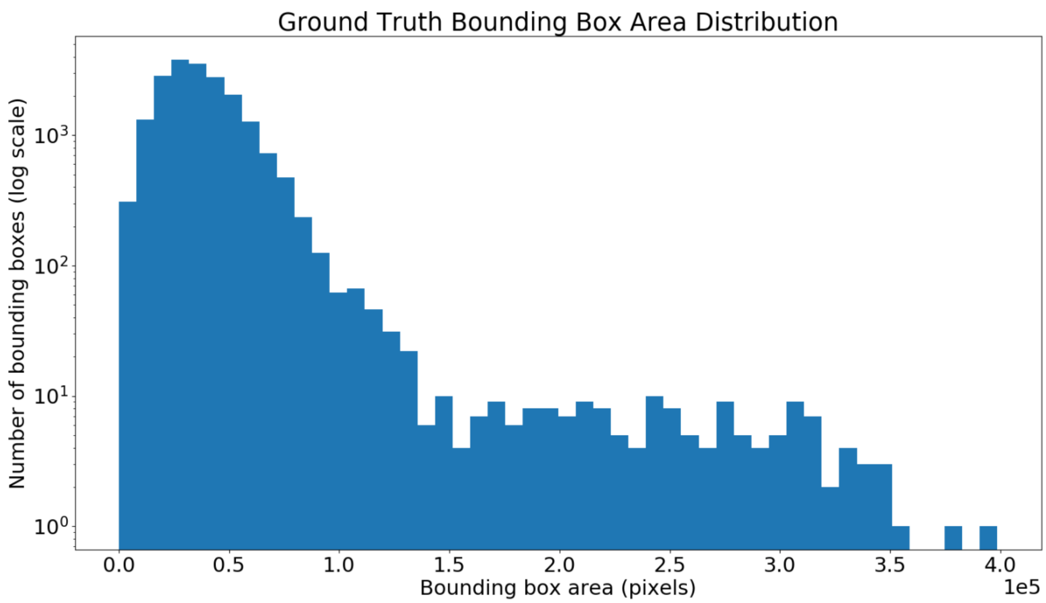


Figure 13: Ground truth bounding box size histogram, in terms of pixels using 50 bins. Note that the majority are comparatively small, with the smallest bounding box at 30 pixels.

Detection Proposals:

The detection proposals from RetinaNet and Faster R-CNN are included in the dataset; since proposals were generated from fine-tuned networks, we performed two training and evaluation cycles for the detector networks, starting from their base COCO-detection trained models:

1. Train on the NHFish training set, generate proposals for the validation set
2. Train on the NHFish validation set, generate proposals for the training set

In this way, we generate detection proposals for all frames in the dataset. In total, there are 70,588 bounding box proposals that have confidence scores > 0.5, and 457,104 proposals total without thresholding generated from RetinaNet. There are 60,726 Faster R-CNN proposals, 47,409 of which have confidence scores > 0.5. The detection proposal performance, as given by the Detectron project's utilities[83], are given in Table 2 and Table 3.

Table 2: RetinaNet ResNet-50 detection performance on the training and validation detection proposal experiments

RetinaNet	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
train-val	0.2506	0.4932	0.2215	0.0	0.0022	0.2546
val-train	0.30095	0.5584	0.2930	0.0	0.0091	0.3098

Table 3: Faster R-CNN ResNet-101 detection performance on the training and validation detection proposal experiments

Faster R-CNN	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
train-val	0.2793	0.5116	0.2747	0.0	0.0080	0.2838
val-train	0.2909	0.5233	0.2930	0.0	0.0203	0.2991

The evaluation metrics in Table 2 and Table 3 are used by COCO for evaluating detection performance. The evaluation descriptions from the COCO website are replicated below:
Average Precision (AP):

- AP: AP is averaged over 10 discrete thresholds (specifically, IoU = .50 : .05 : .95]) rather than a single one. This provides progressively higher AP for bounding boxes with more accurate object localizations, and is a good general metric for overall performance
- AP50: AP with a threshold of IoU=.50 for determining bounding box matching. Concretely, if two bounding boxes have < 0.5 IoU, they do not match; otherwise they do.
- AP75: AP with a threshold of IoU=.75 for determining bounding box matching. This is a more stringent localization requirement than AP50.
- AP Across Scales:
- APs: AP for small objects: area < 322
- APm: AP for medium objects: 322 < area < 962
- APl: AP for large objects: area > 962

RetinaNet and Faster R-CNN are both successful state-of-the-art object detection networks. Their relatively poor performance on NHFish, as shown in Table 2 and Table 3, is indicative of how challenging the dataset can be for existing methods. Additionally, successful tracking approaches on NHFish will have to be robust to imperfect detection proposals (Table 4). Depending on the confidence threshold used for filtering which detection proposals, there may be too few true positives, or too many false positives. In both cases, there can be poorly localized bounding boxes, which will lead to inaccurate appearance and motion features. This sort of noise in the detection proposals adds an additional layer of difficulty for tracking methods on NHFish, and likely account for the lack of meaningful experimental tracking results. Erroneous detections, whether false positives or false negatives, will be detrimental to tracker performance (Figure 14).

Table 4: Descriptive statistics about the set of RetinaNet and Faster R-CNN bounding box proposals across the train and validation splits.

Proposal Bounding Box Metric	RetinaNet ResNet-50	Faster R-CNN ResNet-101
Mean Area (pixels)	36486	35195
Median Area (pixels)	29439	30294
Mode Area (pixels)	2877	1890
Median Height (pixels)	178	176
Median Width (pixels)	158	160
Height Std. (pixels)	77.91	61.76
Width Std. (pixels)	64.24	79.89
Skew, Kurtosis	2.34, 13.03	2.81, 25.12

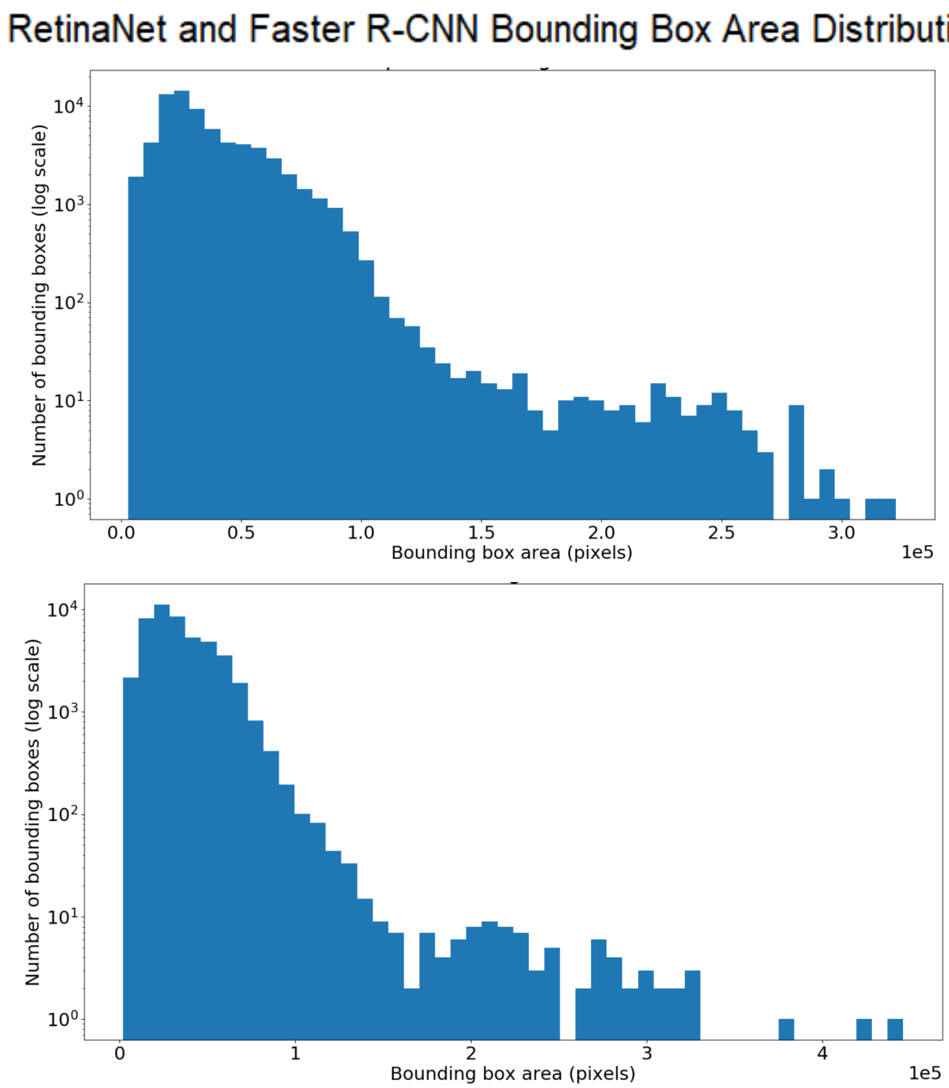


Figure 14: Detection proposal bounding box size histograms, for RetinaNet and Faster R-CNN detectors in terms of pixels using 50 bins. Note that the overall distributions generally mirror that of the ground truth bounding box area distribution.

IV. Ecological Inference and Environmental Decision Making

Studies of groundfish abundance, habitat, and more complex biological interactions became prevalent beginning in the 80's and 90's, evidenced by funded proposals within academia[28]. During that time there was also what has been described as a turn inward leading up to the turn of the century; an attempt to better understand the life history and ecological connections of fish populations[29]. In 1996, Essential Fish Habitat (EFH) was inserted into the reauthorization of the MSA as well, making it more important to understand the complex relationships present in groundfish ecosystems[104]. EFH was further defined in the 2006 MSA reauthorization as "those waters and substrate necessary to fish for spawning, breeding, feeding or growth to maturity"[105]. It was also around this time between 2006 and 2008 when OSU posted their first official position for a fisheries oceanographer within the college of oceanography—this is when they hired Lorenzo Ciannelli[106]. The beam trawl surveys done by Ciannelli and Wakefield discussed previously were born. The video aspect was added in hopes that it would offer a deeper understanding of flatfish as well as more data about their behavior and possible responses to changing oceans[30]. But, if these massive visual datasets are to be used to predict fish abundance, they need to be applicable in the context of environmental decision making.

Management Framework

To understand the feasibility of adding automated video surveillance to the stock assessment surveys, one must first understand the current framework of the management system. As previously stated, the Pacific groundfish fishery is managed by the National Marine Fisheries Service (NMFS) and the Pacific Fisheries Management Council (PFMC). The PFMC is one of eight regional fishery management councils established by the MSA in 1977 [102]. The region that the PFMC oversees includes the waters within the U.S. exclusive economic zone (EEZ) of Washington, Oregon, and California, which spans up to 200 nautical miles offshore. Idaho is also included within the region covered by the PMFC due to the inclusion of salmon which migrate into Idaho's rivers, lakes, and streams. There are 19 people on the council, 14 are voting members, and the remaining 5 are non-voting members. The voting members include: state representatives, state obligatory members, at-large members, a tribal representative, and a NMFS representative[107,108]. The obligatory, at-large, and tribal members are appointed by the U.S. Secretary of Commerce, while the remaining members are appointed by their respective agencies. The non-voting members represent the state of Alaska, the Pacific States Marine Fisheries Commission, the U.S. State Department, the U.S. Fish and Wildlife Service, and the U.S. Coast Guard[107,108]. The NMFS regional general council also provides legal advice to the PFMC. The voting members are tasked with developing four fisheries management plans, or updating old ones, that encompass over 120 different fish species plus 13 fish families and 8 krill species that are under their management, and they make decisions based on majority vote[102]. These fisheries management plans are then sent to the NMFS Regional Administrator and NMFS Headquarters whom must approve, partially approve, or disapprove them[108]. If approved, the Secretary of Commerce promulgates the fishery management plan or the amendment, and then NMFS is responsible for implementing, administering, and enforcing these newly established plans[108,109].

In addition to the voting and non-voting members on the council, there are also groups that advise the council, and are therefore called advisory bodies. These advisory bodies include: Advisory Subpanels, Enforcement Consultants, a Habitat Committee, a Groundfish Allocation Committee, the Scientific and Statistical Committee, as well as various Plan Development, Technical, and Management Teams. The Advisory Subpanels are made up of stakeholder representatives such as commercial and recreational fishing industry members, tribal members, conservation groups, and the public. The Enforcement Consultants represent the Coast Guard, state fish and wildlife agencies, state police agencies, and the National Marine Fisheries Service. The Habitat Committee is composed of representatives from state and federal management agencies, tribes, conservation groups, and fishing industries. This committee collaborates with other advisory bodies on issues related to habitat of council-managed fish species. The Groundfish Allocation Committee is composed of representatives from state management agencies, the National Marine Fisheries Service, and the council. It is also provided legal advice by NOAA's regional office since it oversees allocating groundfish harvestable surplus among the fishing industry. Finally, the Scientific and Statistical Committee is composed of agency and academic scientists, and they are tasked with reviewing the scientific content of the council and other advisory bodies to confirm that management decisions are based on best available science as according to the MSA. Any new scientific methodologies, technology, or general innovation must be approved by the Scientific and Statistical Committee to maintain the robustness of the Fisheries Management Plans. To incorporate automated video surveillance into the current management framework, it would first need to be considered best available science[102].

Best Available Science

This term "best available science" has come up in multiple pieces of United States' legislature within the last 50 years. It first appeared in the Endangered Species Conservation Act of 1969, and it was brought over to the current Endangered Species Act of 1973 [110,111]. Since then, it has been seen in the Environmental Protection Agency's Clean Water Act of 1997, and in the MSA reauthorizations of 1996 and 2007 [112,113]. Despite this term having been in use since 1969, it is still difficult to determine what is considered "best" [112,113]. The 1996 MSA reauthorization called for the use of "best scientific information available" (BSIA), specifically to better inform the setting and enforcing of maximum sustainable yield (MSY), as well as to focus on essential fish habitat (EFH), and cooperative research[114]. BSIA appears in National Standard 2, one of the ten National Standards established in the 1996 reauthorization, which states "[fishery] conservation and management measures shall be based upon the best scientific information available[115]." In 2004, a national report on what "best scientific information available" meant was published by the National Academy of Sciences (NAS) as mandated by the 2002 MSA reauthorization[116]. It stressed the need to improve scientific information and reduce uncertainty, but ultimately recommended that a universal definition or application of best available scientific information was impossible and problematic.[26] Still, the 2006 MSA reauthorization called for better best science, once again[105].

The concept of BSIA has been an evolving one, not a static definition that can be painted broadly over the fisheries councils and the scientists they work with. Due to the vague and ambiguous nature of it, National Standard 2 was revised in 2013 in order to better define what BSIA meant in the context of conservation and management for federal fisheries and Fishery Management

Plans (FMPs)[117]. Since science is dynamic and the amount and availability of quality data varies, the revised National Standard 2 did not give a prescriptive definition of BSIA. Instead, the revised National Standard 2 stated that BSIA should follow a credible scientific method process and be evaluated according to: relevance, inclusiveness, objectivity, transparency and openness, timeliness, verification and validation, and peer review[116,117]. Each of these terms were further defined within the revision. It was also stated that the role of peer review is to evaluate the quality and credibility of the scientific information and not to provide advice to the Regional Fisheries Councils since that is the role of the Scientific and Statistical Committee[117].

Incorporating Innovation

Therefore, for automated video surveillance to be considered the BSIA, it must be evaluated based on the criteria laid out in the newly-revised National Standard 2. One main concern with evaluating the machine learning required for automated video surveillance is that it uses deep networks. These deep networks are difficult to interpret, and are usually deemed a “black box” method, referring to the process of feeding data into an algorithm and receiving an output from this algorithm without knowing what it did to obtain the output. The question becomes how can this method be considered BSIA if it cannot be fully explained, and therefore, cannot be fully evaluated for quality, validity and reliability—which was the purpose behind revising National Standard 2 of the MSA[117]. Additionally, this inability to interpret or explain the generation of output associated with machine learning has led to the new “right to explanation” regulation within the European Union, and has outlawed some deep learning methodologies since they cannot be explained[118]. Similar concerns could come up if these deep networks were used for automated video surveillance that would ultimately be used to supplement fisheries stock assessments, especially since these stock assessments are used to set annual catch limits. This issue would need to be overcome for this method to confidently be considered the BSIA and to be incorporated into the fishery management framework.

Even if this issue of interpretability was resolved and this new technology was used to supplement fisheries stock assessments, there are additional sources of information that the regional fisheries councils use to set annual catch limits[108,119]. Science is not the sole driver of policy decisions since these decisions are innately laden with human values and perspectives[112,119,120]. Policy decisions such as determining an annual catch limit for a fishery are mandated to be driven by the BSIA, but science can only tell the decision makers information about the fish stocks, it is up to the council to decide what to do with this information[108,119]. Even best case scenario, with the BSIA, science is only one piece in the decision-making process as discussed in section III[112,119].

More realistically, the information that science can tell decision makers is either not complete or not of high quality, especially for fisheries where there is little data about them, known as “data poor” fisheries[112,117]. Science can be used as a tool, but it can only take the decision-maker so far, and the council must set annual catch limits for all federally-managed fisheries, even those with incomplete or poor quality data[108,117]. These decisions that the council must make are known as “decisions made under uncertainty” because there are many unknowns associated with them[112]. There are four main types of uncertainty associated with fisheries, some of which are due to lack of data (such as those associated with “data poor” fisheries), and others that are unavoidable and will never be known[112]. These four sources of uncertainty in fisheries science include: 1) information about the biology of the fish stock, 2) information about how the fish stock interacts with other biological populations or with environmental factors, 3) unpredictable events such as natural disasters, and 4) variability in parameter estimates that are necessary to determine the size of the current or past stock and predict the future stock[112]. Making decisions in spite of these uncertainties is a main source of risk in fisheries management[112,117]. However, it is an unavoidable risk because—as previously stated—regional fisheries councils are mandated to set an annual catch limit for all federally-managed fisheries regardless of limitations in scientific information[117].

Figure 15: "Sole Soul" Oil on paper. Samm Newton, 2017

Perceived Risk

The original 1976 MSA gave regional fisheries councils discretionary authority on what fish species would need to be managed and regulated. The councils were mandated to develop an FMP for all of these fishes that they deemed in need of management and regulation[117]. Their discretionary authority lead to the first FMP for Salmon being implemented in 1977.[121] Pacific salmon, as well as groundfish, have a riddled history of population vacillations—successes, and failures. But salmon received more attention, which equaled institutional support (from the PFMC, and both conservation and fishing industries), and financial investment (as evidenced from university funding)[29,28]. There was not even an FMP drafted for Pacific groundfish until 1982, after peak landings of the most popular species, and over exploitation had already become painfully apparent. Despite the surveys and the FMPs, the groundfish population continued to decline for the next twenty years[122].

In 1984 an annual slope survey, with an adjusted trawl system, was added to the overall groundfish sampling on top of the ongoing triennial surveys that had started in '77. But, the annual survey, like the surveys before it, had variable sampling methods until the early 2000's (i.e. variable parameter estimates).[120] Unfortunately, this meant that meaningful data for predicting and/or assessing stocks was unavailable. Questions assessing the stocks were asked of the data that it could not provide due to its inconsistent nature[29]. The groundfish datasets gathered from the beginning of the 1970's through to the 90's were reflective of national and economic desires of the late nineteenth century to find new populations to put into the market and more efficiently commodify the ones that existed, not to save a diminishing stock[123].

> "By 2000, Oregon's catch of groundfish had dropped from a 20-year average of 74,000 tons to just 27,000 tons. In 2002, PFMC declared nine species of groundfish overfished. Faced with an extremely slow growth rate and a high degree of scientific uncertainty, the PFMC decided to close the entire continental shelf to bottom trawling."[122]

In 2018, seven of those nine stocks are still in the process of being rebuilt[118]. It was an ecological and socioeconomic disaster. In the case of the West Coast groundfish crash, all four types of uncertainty were present: 1) science lacked information on the life history of groundfish, 2) the science concerning groundfish populations was relatively new, 3) there had been several unforeseen El Nino events, and 4) there was a history of variability in population sampling[122]. The stocks were reduced to a fraction of their originally estimated stock and have been slow to recover. They have been, and still are, ecologically and socioeconomically important to the Pacific Northwest. They are at risk and their future is uncertain. Additionally the future of every fishery is uncertain, as nationwide stocks have been on the decline[124]. As of 2001:

> "Perhaps 60% of the world's major fish stocks are now overexploited, in the sense that stock sizes have been driven to lower levels than would produce the largest annual biological surplus or net economic value. The 60% is a very rough guess."[57]

Why then, do some fisheries, like the groundfish fishery, lack the institutional support and financial investment needed to incorporate and nurture innovation and emerging technologies that could reduce uncertainty?

One reason could be the idea of perceived risk. Even though the experiences of the fisheries are essentially the same, they are looked at differently and therefore have different influence in management and decision making—usually because of socio-cultural factors[125]. Salmon have had a perceived value and place in society that perhaps groundfish have not. They were visible, easily accessible, arguably more beautiful (but so are Groundfish, Figure 15), and Salmon continue to be a historical and cultural icon of the Pacific Northwest[126]. Not to mention, they are an anadromous species that can be seen and caught in rivers, so they have a greater presence in non-coastal areas. Another risk perception issue could be in the concept of an identifiable victim effect, or when people are more connected to the well-being of an individual versus a vague group with similar needs. The salmon fishery is composed of five fish, while the groundfish fishery has several different types of fish and over ninety different species[127]. And this comparison is not specific to salmon, it refers to the way fisheries are valued and prioritized in general, due to perception gaps. The NMFS prides itself on innovation[124]. Yet, for innovation to be incorporated into the management framework, there must be institutional support and financial investment. Considering the role of perceived risk in determining those two things could remove some roadblocks to incorporating emerging technologies.

Roadblocks and Moving Forward

If it was decided that automated video surveillance would be a helpful supplement to fisheries stock assessments, there are still roadblocks for this technology to be implemented into the fishery management framework. One major roadblock is that NOAA's science budget is being cut for the 2019 fiscal year[128]. This means that current stock assessment surveys are going to need to be cut and new surveys or expansions to surveys are currently not feasible[128]. One potential solution is to have another agency, such as Oregon Department of Fish and Wildlife (ODFW) or Washington Department of Fish and Wildlife (WDFW), partner with NOAA and contribute to the stock assessments with this new technology. This would require man-hours and funding from the state agencies, which is still difficult to find when budgets are tight. Additionally, this would require communication and partnership between the agencies, which can be facilitated through the Pacific Fisheries Management Council process[102]. However, another roadblock would be the practicality of using underwater video footage from the Eastern Pacific Ocean along the Washington, Oregon, California coasts. As previously discussed in section III, this can be much more difficult in the Pacific Northwest area when compared to other areas such as the tropics where there is high water clarity and fish are brightly colored and distinct. Off of the coasts of Washington and Oregon in the Northern California Current there is seasonal spring upwelling and winter storms that cause high water turbidity and low water clarity[129,130]. This causes fish detection to be difficult, especially fish such as flatfish that use camouflage to blend in with their sandy habitat. However, if these roadblocks were overcome, automated video surveillance could cut down on the amount of time required for scientists to analyze video, and it could provide useful supplemental information about fish stocks with minimal man-hours.

In Carmel Finley's, *All the Fish in the Sea,* Finley explores the "shape of science" involved in fisheries management.[131] She argues that policy implications paralyzed scientists and created roadblocks to both scientific progress and new paths of knowledge generation about the nature and management of marine fishes. However, politics and economics are not the only actors on stage in science-informed, industry-driven, environmental policy and management dynamics. In the same way that policy and economy acted as constraining frameworks in fisheries, so too can emerging technologies and perceptions of risk and worth. The datasets needed to move models forward were not available, so the modelers had no incentive to develop further. In turn, if the models could not incorporate new datasets (like the ones collected with emerging technology for example), there was no incentive to make new, possibly better datasets available. This has caused an iterative process of mutual stagnation between innovation, technological choice, and environmental decision making.

The dependency of predictive models on consistent methods and datasets has validity, but has also been limiting. The inability of the management system to accommodate new varieties of data, obtained using emerging technologies, has created a path dependency that limits robust understandings of fish populations before, during, and after they are considered under risk. Large, novel datasets have the potential to provide population abundance data that could be incorporated into existing models, or used to develop more predictive models before another fisheries disaster—not in reaction to one. Although, there ought to also be more open and critical analysis of socioeconomic perceptions of risk. Additionally, the feasibility of automated video surveillance becoming BSIA depends on institutional support and financial investment. The current management framework needs fuel to move forward and the permission to embrace innovative approaches to measuring and managing fish.

V. Synopsis

Discussion

The technical feasibility of performing fully automated fish analysis is likely still a few years out in the future. The primary issues include:

1. Manual labor involved in labeling data for training: modern successful ML and CV techniques are primarily within the supervised domain, meaning that a training dataset with ground truth is still required for informing the algorithm of what its effective transfer function should be. As such, generating larger and/or more varied datasets requires human effort to label and curate new data. Our experiences in creating the NHFish dataset have led us to conclude that this is a large and tedious burden, and it is unlikely that we can easily find willing participants. In addition, there is a degree of familiarity with the target benthos required to make informed annotations, which further limits the pool of potential labelers.
2. The beamtrawl video data itself is extremely challenging. The visual homogeneity of the environment with respect to the fish, the small, fast-moving targets, the significant dynamic appearance deformation fish undergo while swimming, and the abundance of visual distractors (notably, the trawl chain and resulting dust cloud) mean that the data environment is highly divergent from other commonly used datasets in the field of CV, and confounds existing algorithmic approaches. (see sections II and III for more specific details). Certain classes of algorithms perform very poorly with this data, such as optical flow, due to the highly non-linear camera motion and high magnitude of complicated, per-pixel motion and occlusion from the dust cloud (we applied an implementation of DCFlow, a state-of-the-art optical flow algorithm optimizing over a 4D cost volume based on the author's open source implementation at https://github.com/IntelVCL/dcflow -- the results were not usable).
3. The data is collected in an uncontrolled environment, with occasional unexpected events and widely varying degrees of visual quality. In some trawls the camera rig is obscured by crab pots, effectively rendering the data from that trawl useless. Similarly, seasonal environmental conditions can significantly degrade visual quality of the videos, reducing visual sight range and obscuring scene details. The unconstrained nature of the data requires more robust algorithms and software, which in turn is more challenging for any potential algorithms and limits the application of existing software systems to the dataset.

While #1 is a surmountable problem, #2 and #3 are more challenging. Our experiments in running two modern object detection deep networks on the dataset show that contemporary CV state-of-the-art methods are unable to adequately handle the data. The lack of suitable detection proposals in turn hinder tracking-by-detection efforts; trackers are used to link detections temporally and filter out false-positives; as such, they have a high reliance on the detection proposals used. If there are too many missing detections, the tracker will be unable to link them into temporally consistent tracks. Conversely, if there are too many false positives, the tracker will suffer from high degrees of ID-switches and will have difficulty constructing valid tracks. Unfortunately, the results from running RetinaNet and Faster R-CNN detectors on the dataset provided low-quality detection proposals and performed especially poorly on smaller targets in the scene. As-is, the data annotation process eliminated any fish under 10cm (as per the trawl rig's laser range finder) from the ground truth, so the lackluster performance (e.g. APs was 0 for both detectors, and APm averaged 0.0056 and 0.014 for RetinaNet and Faster R-CNN respectively) on small targets bodes ill for future use cases expanding further into juvenile / smaller fish domains.

The lack of reliable detections in turn hinder any attempts at running tracking on the data, which is required for inferring longer-term temporal information and making inferences regarding fish dynamics and populations estimates. Our experiments using a modern multi-object tracking algorithm, MHT-DAM, with the detection proposals and deep visual features to date have not yielded promising or usable results. MHT-DAM is a multi-hypothesis tracker that has enjoyed considerable success on the MOT challenge, and which has an open source implementation.

The institutional feasibility of incorporating video monitoring into ecological inference is not as insurmountable as it is in environmental decision-making. While individual scientists may choose to embrace innovation, and explore emerging technologies (like video, computer vision, and deep learning), the ability of current management frameworks to incorporate them is another question.

Changing ocean conditions present continuing uncertainties for Pacific groundfish and magnify the concept of what it means for a fishery to be perceived as under risk. Unless these issues are faced head on, the iterative cycle of innovative stagnation and arrested momentum between fisheries science and management will continue. We have discussed the importance of datasets in influencing both the direction and quality of fisheries science. And we have presented how institutional support and financial investment are key factors in moving fisheries forward. It is surprising then to learn that the WCGBTS will be having its funding downsized soon, even though some species of groundfish are still rebuilding after the recent west coast groundfish disaster. This fishery also contains over 90 species of fish that will be affected by the decision to reduce population abundance data. There also may be stocks that become commercially viable in the future, and if we are to manage them better than we managed petrale sole (and others) we need to have robust datasets on their environment, behavior, abundance and life history.

> "If fisheries science is to be successful we must learn from and avoid the mistakes of the past. We must help managers make *choices* about *dynamic* fishery systems in the face of *uncertainty*… we are fortunate that our predecessors have made lots of mistakes… We cannot avoid making mistakes, but we do our predecessors a great disservice if we do not take advantage of what has been learned and try not to repeat the same mistakes… building on historical experience involves a recognition that stock assessment does not consist of making static predictions about optimum efforts and sustainable yields, but concerns the assessment of time trajectories of fish and fishermen in response to management and other

> changes. It also involves a recognition that stock assessment biologists must educate managers and decision makers to ask appropriate questions and to think of the dynamic response of fisheries to change." [132]

Conclusion

Computer vision and machine learning techniques offer a tantalizing opportunity for performing automated analysis of beamtrawl videos of Pacific groundfish off the Oregon coast. This could be transformative for the fields of fisheries management and fishery science, provided such methods gave sufficient performance. After extensive experimentation and the creation of a new, large-scale annotated dataset of beamtrawl videos, we have concluded that while such a thing is certainly desirable, it is not feasible given current day algorithmic techniques and management frameworks (Figure 16).

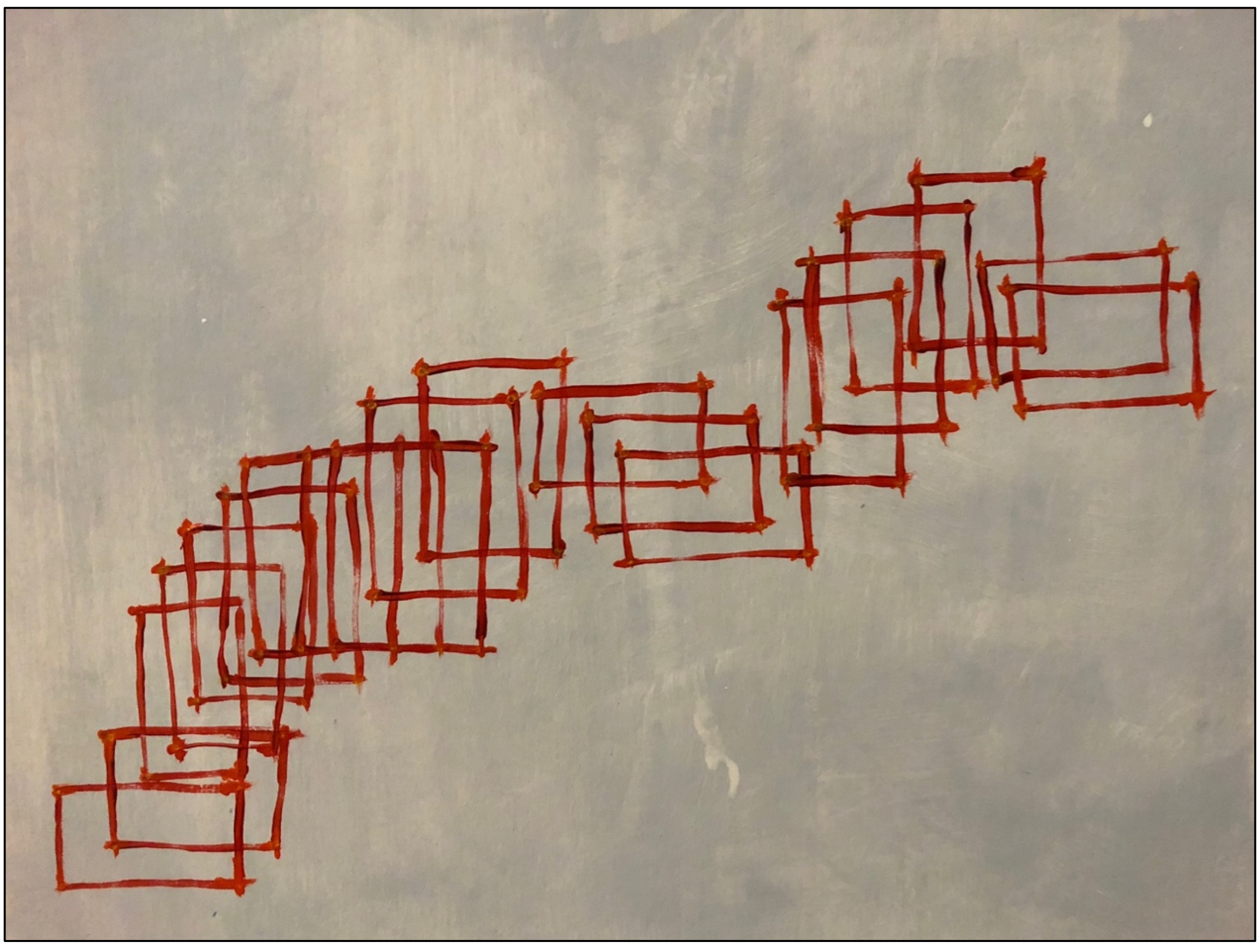


Figure 16: "Feasiblity" Oil on paper. Samm Newton, 2018.

Considering the lack of success with modern state-of-the-art detection and tracking approaches, the current capabilities of CV are insufficient for reliable usage in Oregon benthic automated fish analysis applications. Although rough heuristics could be used, it would not necessarily offer any tangible benefit over existing techniques. Recent advances in CV techniques brought about by DL have yielded significant improvements over relatively recent methods (e.g. even just 5 years ago), and we expect to continue to enjoy improved performance as time progresses. Although not sufficiently accurate in the present, future progress in CV will iteratively improve performance, until it achieves a more useable state for the NHFish data, closing the gap in desired and actual functionality. As a result, the coming years will see renewed interest in CV techniques for automated ecological inference. Initial results from our experimentation indicate that CV holds much promise for innovation in automated underwater video processing, and given both the time for technologies to mature and the decision-making frameworks to adapt, CV for analyzing trawl videos could become a viable approach in the coming years.

Due to the critical influence of ecological datasets in shaping both technical and institutional aspects of knowledge inquiry, we expect that the newly introduced NHFish dataset (a challenging, high-resolution collection of annotated beam trawl video) will encourage further work in the area, and provide novel and successful approaches to measuring and managing fisheries. We hope the release of this labeled dataset will encourage and direct the future development of algorithms tailored to the specific challenges of the beamtrawl video, as well as the direction of further fisheries science and management. We also hope that the value of Oregon's fisheries will be more deeply explored and possibly reevaluated with perceived risk in mind. It is not to say that we should reduce fishing pressure full stop, because that is a policy decision associated with people's livelihoods, and datasets and fisheries science are only a piece of the full framework. It is saying, however, that we need not only more robust datasets about the fish we do plan to pursue, but also management in place that is adaptive enough to incorporate them.

VI. Methods

The relationship between culture and science is a tangled knot that exemplifies the inseparable realms of the natural and the social, the human and the nonhuman. Consequently, cross-disciplinarity tends to be the best approach to study the associations that have created such a complex world.

Cross-disciplinarity has become an increasingly prevalent method of knowledge inquiry and generation. Terms like 'interdisciplinary' and 'transdisciplinary' (forms of cross-disciplinarity) have frequently been used in funding calls and research proposals, especially in the study of marine systems. These terms have also more recently—to broach complex ecological and social issues—looked to combine the humanities with the natural and social sciences. Although, transdisciplinarity, like interdisciplinarity, lacks consensus in both meaning and execution[133]. There is an extensive discourse on what the ideal cross-disciplinary process might look like, but less on how it plays out in practice.

To more closely examine the opportunities and challenges associated with cross-disciplinary approaches to knowledge generation this section will consider that very issue. How is transdisciplinary work done? What are practical methods for producing meaningful work as a team, especially with researchers from different areas? Through this project, our group found that being open and critical, embracing disparity between disciplines, strong communication, and a process-driven perspective led to a successful team dynamic and overall research outcome.

For specific author contributions and discipline specific methodologies please refer to the Appendix B.

Methods of Transdisciplinarity

In “A Philosophical Framework for an Open and Critical Transdisciplinary Inquiry,” Jacqueline Y. Russell argues in favor of a methodological shift to ethically oriented transdisciplinary inquiry that can better broach environmental issues, especially complex problems frequently referred to as wicked problems. Overall, to do this, inquirers must embrace three aspects of knowledge generation: epistemological, ontological, and ethical. The author adopts her own advice by framing her argument around how we come to know things (epistemology), the limits of what we can know (ontology), and if this knowledge can offer some type of improvement in the world (ethical). She explores the reliability of knowledge, knowledge communities, and the influence of positivism on contemporary knowledge generation. G. R. Midgley’s systemic interventionist account of knowledge, which Russell focuses on, provides a convincing example of how systems of inquiry that intervene can bring ontological, epistemological and ethical structures together (systems of intervention i.e. work that intervenes in the world, which the problem of measuring and managing fish fits into). Russell also stresses how and why that novel composition allows for a more robust consideration of wicked problems, like the future of predicting fish populations. By being open (expanding knowledge of the world to include not just physical, but also social and cultural) and critical (accounting for the uncertainty of what we can know) knowledge generation organically becomes transdisciplinary and is better poised to improve relations among humans and our scientific, social, cultural (and other) interactions with the environment.

Therefore, while working on this project each participating researcher strove to be both open and critical. This was achieved first by maintaining a curious attitude about one another’s approaches to the world. The group spent a considerable amount of time at the beginning of the project learning about each other’s personal, professional, and academic backgrounds. Because of the time constraints posed on this project, it sometimes seemed like too much time was spent on that aspect, but it was a crucial step that built a foundation for being open and critical in the future. Taking the time to truly understand and build respect for each other’s disciplines and approaches to research was key. As the work progressed, the group continued to listen and seriously consider everyone’s perspectives. While on the surface this may sound like a straightforward and simple step, it can be challenging when you combine different ages, education, experience, and areas of knowledge. Computer Science, Marine Resource Management, and the Environmental Arts and Humanities are disparate disciplines that do not approach research or academic work similarly nor do they intuitively fit together at first glance. But, the three areas of expertise represented strongly correlate with Russel’s theories combining physical, social, and cultural knowledge with critical. This transdisciplinary inquiry allowed us to not only generate knowledge about the world, but to incorporate an understanding of how we know that world, and what we ought to do about it. Even still, it took many hours of working, communicating, struggling, and socializing to make it work.

Habermas wrote that scientific knowledge was not the only type of rational knowledge, and that methods from instrumental, ethical, and aesthetic knowledges could be integrated[134]. Because of this, cross disciplinary processes are successful when they value collaborations among individuals who approach knowledge differently, not just from different disciplines (Figure 17).

Much cross-disciplinary work might have researchers from different disciplines, but that still approach knowledge generation in more or less similar ways.

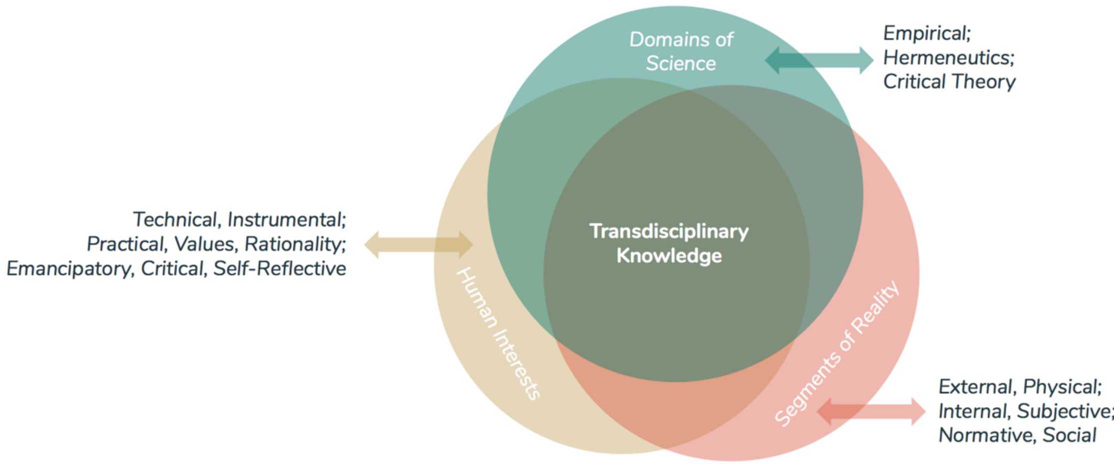


Figure 17 Russel's argument for transdisciplinary knowledge generation based on Habermas and others.

This project contained individual understandings of marine biology, fisheries science, environmental decision making, resource management, computer science, machine learning, computer vision, environmental history, philosophy, science and technology studies, and the incorporation of art as method. With a diverse group like that we had to move away from personal expertise and blur the lines between disciplines to find a question that we could ask together that we could not ask individually. At the same time, we had to constantly navigate the vacillating process that was letting go of personal expertise to foster that sense of openness, while at the same time contributing individual research and knowledge to remain critical. Russel argues that transdisciplinarity is uniquely poised to ask what should be, what is, what could be, and what can be. Together, our group attempted to do bring together those questions and make our research about the topic, and less about discipline. Abandoning pure disciplinarity, which seeks unity of knowledge, and embracing deconstruction, which seeks coherence, allows for a novel style of hybrid scholarship that can study complex problems in a more holistic way (Figure 18) [135].

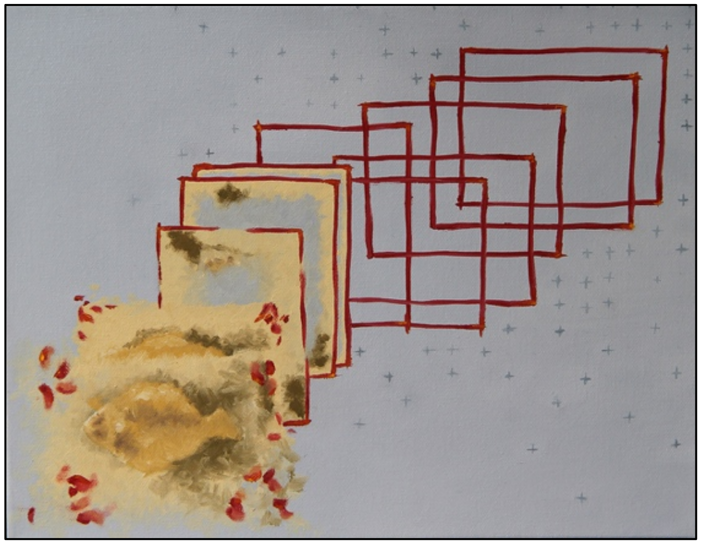


Figure 18 "Deconstruction & Coherence" Oil on canvas. Samm Newton, 2018.

Strong communication was also an important method for this transdisciplinary project. Communicating effectively and practicing positive interpersonal skills are what most theorists focus on when writing about how to do inter and trans disciplinary work. For example, our influences came from Michael O'Rourke[136], Kendra Cheruvelil[137], Sanford Eigenbrode[138], and Patrick Hughes[139] to name a few, all whom stressed the importance of communication and offered several tools and advice on how to work together effectively. What we found to work best, in practice, was the use of boundary objects, creating and translating language, discussing ambiguity, having backstage conversation, and being kind and considerate of one another. For example, if we got lost or unsure of the direction of the research we would return to what brought the group together—our boundary object—which in this case was the groundfish videos. Also, because there were many technical semantics and concepts in the represented disciplines, defining terms and understanding the meaning behind everything discussed was also an important and regular task. The group would discuss and define ambiguity, explain concepts in detail, and sometimes create new terms or definitions to hold the group together. Being open with one another about things outside of academia also helped to foster mutual respect and understanding between group members. There were often jokes about forced friendship, but in the end even if we were not best friends, we did allow for backstage (not work related) conversations that allowed us to stay connected on a social as well as scholarly level.

Another important aspect of this project's success was its focus on process, rather than product. This is another method of transdisciplinary work that is difficult in practice. So often the emphasis of research and scholarship in the academic setting is on immediate, tangible deliverables. We chose instead to focus on the novelty of the situation and to capitalize on the unique opportunity to create and present something different than we would be able to do through traditional disciplinary based scholarship. After much deliberation on the question "what can we ask together that we cannot ask alone," it was decided that it would be best to approach our topic with a three-pronged approach that was based on the idea of dyads collaborating more efficiently than a triad.

The group split into three pairs focused on different parts of our main topic, and that explored the topic in different ways. Firl and Haven looked at the feasibility of using machine learning to make ecological inferences; Haven and Newton looked at the intersections of ecological inference and environmental decision making; Newton and Firl looked at the influence of datasets on machine learning and ecological inference (Figure 19). Our influence on one another went far behind the words written on these pages, but that success is less tangible, and less deliverable. Although, it does have the potential to shift dominant paradigms of what scholarship looks like in the future and possibly change how we measure research outcomes, especially when it comes to cross-disciplinarity. What can be argued is that we greatly influenced one another, our thought process about this topic evolved and transformed as we continued to work together, and will most likely have lasting impacts on how we continue our own personal research.

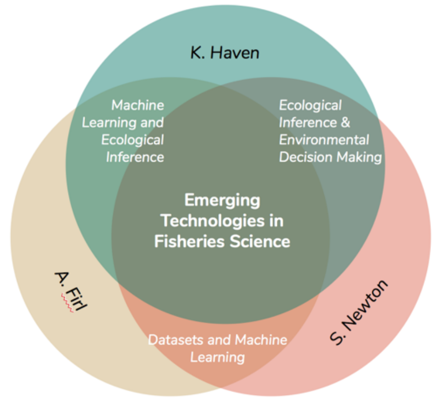


Figure 19: Composition of the transdisciplinary team and how the dyads contributed to the overall triad.

We live in an uncertain time of changing ocean conditions where the problems of measuring and managing fish have serious ecological, societal, and political implications. Together we tackled this issue and questioned how big data and emerging technologies might influence and be incorporated into natural resource management, and ultimately the future of fisheries science. The outcome was a transdisciplinary report written in one untied voice that was developed and executed based on the unique strengths and perspectives of each of its members; further confirming that there is no one recipe or toolbox that works for every group. Rather, transdisciplinary teams must frontload a foundation of respect for one another and develop unique methods and work plans that are specific to the composition of their group.

VII. References

1. The Research Group, LLC. *Oregon Commercial Fishing Industry Year 2016 Economic Activity Summary*. 1–13 (Oregon Department of Fish & Wildlife, Marine Resource Program, 2017).
2. ODFW, Economic Impact. (2018). Available at: http://www.dfw.state.or.us/agency/economic_impact.asp. (Accessed: 22nd January 2018)
3. Background | Pacific Fishery Management Council. Available at: https://www.pcouncil.org/groundfish/background/. (Accessed: 31st August 2018)
4. Pacific Fishery Management Council. Available at: https://www.pcouncil.org/. (Accessed: 17th February 2018)
5. Bethesda, A. F. S. 425 B. P. & Md 20814897-8616. House Natural Resources Subcommittee Holds Hearing on Magnuson-Stevens Act. *American Fisheries Society* (2017). Available at: https://fisheries.org/2017/07/house-natural-resources-subcommittee-holds-hearing-on-magnuson-stevens-act/. (Accessed: 5th February 2018)
6. Fisheries, N. Laws & Policies | NOAA Fisheries. (2018). Available at: https://www.fisheries.noaa.gov/topic/laws-policies. (Accessed: 21st February 2018)
7. PFMC. *PACIFIC COAST GROUNDFISH FISHERY MANAGEMENT PLAN*. 1–160 (Pacific Fishery Management Council, 2016).
8. Fishery Management Plan and Amendments | Pacific Fishery Management Council. (2018). Available at: https://www.pcouncil.org/groundfish/fishery-management-plan/. (Accessed: 21st February 2018)
9. Keller, A. A., J. R. Wallace, and R. D. Methot. *The Northwest Fisheries Science Center's West Coast Groundfish Bottom Trawl Survey: History, Design, and Description. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-NWFSC-136. DOI: 10.7289/V5/TM-NWFSC-136.* (2017).
10. Fishery Resource Analysis and Monitoring Division - Northwest Fisheries Science Center. (2018). Available at: https://www.nwfsc.noaa.gov/research/divisions/fram/index.cfm. (Accessed: 7th February 2018)
11. Webmaster. Debate Continues Over Magnuson-Stevens Act Reauthorization | Ocean Leadership. *Consortium for Ocean Leadership* (2017). Available at: http://oceanleadership.org/debate-continues-magnuson-stevens-act-reauthorization/. (Accessed: 5th February 2018)
12. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60,** 84–90 (2017).
13. Jia Deng, R. *et al.* ImageNet: A large-scale hierarchical image database. in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* 248–255 (IEEE, 2009). doi:10.1109/CVPR.2009.5206848
14. Extraneous factors in judicial decisions | PNAS. Available at: http://www.pnas.org/content/108/17/6889. (Accessed: 31st August 2018)

15. Zhao, J., Wang, T., Yatskar, M., Ordonez, V. & Chang, K.-W. Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints. (2017).

16. Google Photos labels black people as 'gorillas' - Telegraph. Available at: https://www.telegraph.co.uk/technology/google/11710136/Google-Photos-assigns-gorilla-tag-to-photos-of-black-people.html. (Accessed: 31st August 2018)

17. Wakabayashi, D. Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam. *The New York Times* (2018).

18. Wolpert, D. H. & Macready, W. G. No free lunch theorems for optimization. *Evol. Comput. IEEE Trans. On* **1,** 67–82 (1997).

19. Lattimore, T. & Hutter, M. No Free Lunch versus Occam's Razor in Supervised Learning. (20111116).

20. Sampson, D. B., Crone, P. R. & Northwest Fisheries Science Center U. S. *Commercial fisheries data collection procedures for U.S. Pacific coast groundfish*. (USDeptof Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northwest Fisheries Science Center ; Available from National Technical Information Service, 1997).

21. *Fish population dynamics*. (Wiley, 1977).

22. *Fishery Conservation and Management Act of 1976*. *331* (1976).

23. Keller, A. A., Wallace, John R. (John Robert), 1959- & Methot, Richard Donald, 1953-. *The Northwest Fisheries Science Center's West Coast Groundfish Bottom Trawl Survey : history, design, and description*. (U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northwest Fisheries Science Center, 20). doi:10.7289/V5/TM-NWFSC-136

24. Bolster, W. J. *The mortal sea: fishing the Atlantic in the age of sail*. (2014).

25. Committee on Defining the Best Scientific Information Available for Fisheries Management, Ocean Studies Board, Division on Earth and Life Studies, National Research Council & National Academy of Sciences. *Improving the Use of the 'Best Scientific Information Available' Standard in Fisheries Management*. (National Academies Press, 2004).

26. Warlick, A., Steiner, E. & Guldin, M. History of the West Coast groundfish trawl fishery: Tracking socioeconomic characteristics across different management policies in a multispecies fishery. *Mar. Policy* **93,** 9–21 (2018).

27. Research and Accounting Office Records, 1935 - 2010, RG026. Special collections and Archives, Oregon State University.

28. Susan Hannah. History of Groundfish in the Pacific Northwest. (1998).

29. Waldo Wakefield. Interviewed by Samm Newton. *Waldo Wakefield*. (2018).

30. Pearcy, W., Hosie, M. & Richardson, S. Distribution and duration of pelagic life of larvae of Dover sole, Microstomus pacificus; rex sole, Glyptocephalus zachirus; and petrale sole, Eopsetta jordani, in waters off Oregon. *Fish Bull NMFSNOAA* **75,** 173–183 (1977).

31. Richardson, S. & Pearcy, W. Coastal and oceanic fish larvae in an area of upwelling off Yaquina Bay, Oregon. *Fish Bull NMFSNOAA* **75,** 125–145 (1977).

32. Pearcy, W. G. Distribution and abundance of small flatfishes and other demersal fishes in a region of diverse sediments and bathymetry off Oregon. (1978).

33. Krygier, E. E. & Pearcy, W. The role of estuarine and offshore nursery areas for young English sole, Parophrys vetulus Girard, of Oregon. *Fish. Bull.* **84,** 119–132 (1986).

34. Stinton, A., Ciannelli, L., Reese, D. & Wakefield, W. USING IN SITU VIDEO ANALYSIS TO ASSESS JUVENILE FLATFISH BEHAVIOR ALONG THE OREGON CENTRAL COAST. *Rep. Calif. Coop. Ocean. Fish. Investig.* **55,** 158–168 (2014).

35. Leal-Taixé, L. *et al.* Tracking the Trackers: An Analysis of the State of the Art in Multiple Object Tracking. *ArXiv170402781 Cs* (2017).

36. Sun, C., Shrivastava, A., Singh, S. & Gupta, A. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. *ArXiv170702968 Cs* (2017).

37. Lin, T.-Y. *et al.* Microsoft COCO: Common Objects in Context. *ArXiv14050312 Cs* (2014).

38. Pearcy, W. Distribution and abundance of small flatfishes and other demersal fishes in a region of diverse sediments and bathymetry off Oregon. *Fish Bull 763 629-640 1978* (1978).

39. Krygier, E. E. & Pearcy, W. The role of estuarine and offshore nursery areas for young English sole, Parophrys vetulus Girard, of Oregon. *Fish. Bull.* **84,** 119–132 (1986).

40. Toole, C., Brodeur, R. D., Donohoe, C. J. & Markle, D. Seasonal and interannual variability in the community structure of small demersal fishes off the central Oregon coast. *Mar. Ecol. Prog. Ser.* **428,** 201–217 (2011).

41. McClatchie, S. *Regional Fisheries Oceanography of the California Current System The CalCOFI program*. (Springer Netherlands : Imprint: Springer, 2014).

42. Keller, A. A., Wallace, J. R., Horness, B. H., Hamel, O. S. & Stewart, I. J. Variations in eastern North Pacific demersal fish biomass based on the U.S. west coast groundfish bottom trawl survey (2003-2010).(Report). *Fish. Bull.* **110,** 205–222 (2012).

43. Sobocinski, K. L., Ciannelli, L., Wakefield, W. W., Yergey, M. E. & Johnson-Colegrove, A. Distribution and abundance of juvenile demersal fishes in relation to summer hypoxia and other environmental variables in coastal Oregon, USA. *Estuar. Coast. Shelf Sci.* **205,** 75–90 (2018).

44. Johnson, A. M. *An investigation of the distribution and abundance of ichthyoplankton and juvenile benthic fishes in relation to nearshore hypoxia within the Northern California Current system*. (Oregon State University, 2012).

45. Stinton, A., Ciannelli, L., Reese, D. C. & Wakefield, W. Using in situ video analysis to access juvenile flatfish behavior along the oregon central coast. *Calif. Coop. Ocean. Fish. Investig. Rep.* **55,** 158–168 (2014).

46. Hughes, B. B. *et al.* Climate mediates hypoxic stress on fish diversity and nursery function at the land-sea interface. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 8025–30 (2015).

47. Thanassekos, S., Latour, R. J. & Fabrizio, M. C. An individual-based approach to year-class strength estimation. *ICES J. Mar. Sci.* **73,** 2252–2266 (2016).

48. Lee A. Fuiman & Robert G. Werner. *Fishery science: the unique contributions of early life stages*.

49. Mueter, F. & Norcross, B. Linking community structure of small demersal fishes around Kodiak Island, Alaska, to environmental variables. *Mar. Ecol. Prog. Ser.* **190,** 37–51 (1999).

50. Auster, P., Malatesta, R. & Larosa, S. Patterns of microhabitat utilization by mobile megafauna on the southern New England (USA) continental shelf and slope. *Mar. Ecol. Prog. Ser. Oldendorf* **127,** 77–85 (1995).

51. Cailliet, G. M., Andrews, A. H., Wakefield, W. W., Moreno, G. & Rhodes, K. L. Fish faunal and habitat analyses using trawls, camera sleds and submersibles in benthic deep-sea habitats off central California. *Oceanol. Acta* **22,** 579–592 (1999).

52. Hannah, R. W., Jones, S. A., Lomeli, M. J. M. & Wakefield, W. W. Trawl net modifications to reduce the bycatch of eulachon (Thaleichthys pacificus) in the ocean shrimp (Pandalus jordani) fishery. *Fish. Res.* **110,** 277–282 (2011).

53. Wakefield, W. W. Evaluating the role of fish behavior in surveys conducted with underwater vehicles. *Can. J. Fish. Aquat. Sci.* **65,** 1230–1243 (2008).

54. Shafait, F. *et al.* Fish identification from videos captured in uncontrolled underwater environments. *ICES J. Mar. Sci.* **73,** 2737–2746 (2016).

55. Olden, J. D., Jackson, D. A. & Peres-Neto, P. R. Predictive Models of Fish Species Distributions: A Note on Proper Validation and Chance Predictions. *Trans. Am. Fish. Soc.* **131,** 329–336 (2002).

56. Hilborn, R. & Walters, C. J. Role of Stock Assessment in Fisheries Management. in *Quantitative fisheries stock assessment: choice, dynamics and uncertainty* 3–21 (Kluwer Academic Publishers, 2001).

57. Haddon, M. Fisheries, Population Dynamics, and Modelling. in *Modelling and quantitative methods in fisheries* 1–18 (CRC Press, 2011).

58. Smeulders, A. W. M. *et al.* Visual Tracking: An Experimental Survey. *Pattern Anal. Mach. Intell. IEEE Trans. On* **36,** 1442–1468 (2014).

59. Milan, A., Leal-Taixe, L., Reid, I., Roth, S. & Schindler, K. MOT16: A Benchmark for Multi-Object Tracking. (20160302).

60. Leal-Taixé, L., Milan, A., Reid, I., Roth, S. & Schindler, K. MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking. (20150408).

61. Felzenszwalb, P. F., Girshick, R. B., Mcallester, D. & Ramanan, D. Object Detection with Discriminatively Trained Part-Based Models. *Pattern Anal. Mach. Intell. IEEE Trans. On* **32,** 1627–1645 (2010).

62. Wäldchen, J. & Mäder, P. Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review. *Arch. Comput. Methods Eng.* **25,** 507–543 (2018).

63. Pun, T., Gerig, G. & Ratib, O. Image analysis and computer vision in medicine. *Comput. Med. Imaging Graph.* **18,** 85–96 (1994).

64. Ko, B. & Kwak, S. Survey of computer vision-based natural disaster warning systems. *Opt. Eng.* **51,** <xocs:firstpage xmlns:xocs=""/> (2012).

65. Berger-Wolf, T. Y. *et al.* Wildbook: Crowdsourcing, computer vision, and data science for conservation. (20171024).

66. Schü Ffler, P. *et al.* TMARKER: A free software toolkit for histopathological cell counting and staining estimation.(Symposium - Original Research)(Report). *J. Pathol. Inform.* **4,** 2–2 (2013).

67. Mönck, H. J. *et al.* BioTracker: An Open-Source Computer Vision Framework for Visual Animal Tracking. (20180321).

68. Bruijning, M., Visser, M. D., Hallmann, C. A. & Jongejans, E. trackdem: Automated particle tracking to obtain population counts and size distributions from videos in r. *Methods Ecol. Evol.* **9,** 965–973 (2018).

69. Kimura, T. *et al.* Development of a New Method to Track Multiple Honey Bees with Complex Behaviors on a Flat Laboratory Arena.(Research Article). *PLoS ONE* **9,** e84656 (2014).

70. Boenisch, F. *et al.* Tracking all members of a honey bee colony over their lifetime. (20180209).

71. Mittal, A., Dua, I., Shukla, P. & Raman, B. A Computer Vision Framework for Detecting and Preventing Human-Elephant Collisions. in *Computer Vision Workshop (ICCVW), 2017 IEEE International Conference on* **2018-,** 2883–2890 (IEEE, 2017).

72. Dabarera, R. & Rodrigo, R. Vision based elephant recognition for management and conservation. in *Information and Automation for Sustainability (ICIAFs), 2010 5th International Conference on* 163–166 (IEEE Publishing, 2010). doi:10.1109/ICIAFS.2010.5715653

73. Siddiqui, S. A. *et al.* Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data. *ICES J. Mar. Sci.* **75,** 374–389 (2018).

74. Hossain, E., Alam, S. M. S., Ali, A. A. & Amin, M. A. Fish activity tracking and species identification in underwater video. in *Informatics, Electronics and Vision (ICIEV), 2016 5th International Conference on* 62–66 (IEEE, 2016). doi:10.1109/ICIEV.2016.7760189

75. Rodriguez, Alvara et. al. Fish monitoring and sizing using computer vision. in (Springer Netherlands : Imprint: Springer, 2015).

76. Meng-Che Chuang, K., Jenq-Neng Hwang, K., Jian-Hui Ye, K., Shih-Chia Huang, K. & Williams, K. Underwater Fish Tracking for Moving Cameras Based on Deformable Multiple Kernels. *Syst. Man Cybern. Syst. IEEE Trans. On* **47,** 2467–2477 (2017).

77. Williams, K., Rooper, C. & Harms, J. Report of the National Marine Fisheries Service Automated Image Processing Workshop. 56

78. Fish task | ImageCLEF / LifeCLEF - Multimedia Retrieval in CLEF. Available at: https://www.imageclef.org/2014/lifeclef/fish. (Accessed: 1st September 2018)

79. Fisher, R. B., Chen-Burger, Y.-H., Giordano, D. C. teacher, Hardman, L. & Lin, F.-P. *Fish4Knowledge: collecting and analyzing massive coral reef fish video data*. (Springer, 2016).

80. Jager, J., Wolff, V., Fricke-Neuderth, K., Mothes, O. & Denzler, J. Visual fish tracking: Combining a two-stage graph approach with CNN-features. in *OCEANS 2017 - Aberdeen* **2017-,** 1–6 (IEEE, 2017).

81. Everingham, M., Gool, L., Williams, C., Winn, J. & Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **88,** 303–338 (2010).

82. Girshick, R., Donahue, J., Darrell, T. & Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. (20131111).

83. Girshick, R. Fast R-CNN. (20150430).

84. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Pattern Anal. Mach. Intell. IEEE Trans. On* **39,** 1137–1149 (2017).

85. He, K., Gkioxari, G., Dollar, P. & Girshick, R. Mask R-CNN. *Pattern Anal. Mach. Intell. IEEE Trans. On* **PP,** 1–1

86. Uijlings, J., Sande, K., Gevers, T. & Smeulders, A. Selective Search for Object Recognition. *Int. J. Comput. Vis.* **104,** 154–171 (2013).

87. Huang, J. *et al.* Speed/accuracy trade-offs for modern convolutional object detectors. (20161130).

88. Sermanet, P. *et al.* OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. (20131221).

89. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. (20150608).

90. Liu, W. *et al.* SSD: Single Shot MultiBox Detector. in **9905,** 21–37 (20151207).

91. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollar, P. Focal loss for dense object detection. *Pattern Anal. Mach. Intell. IEEE Trans. On* **PP,** 1–1

92. Strachan, N. J. C., Nesvadba, P. & Allen, A. R. Fish species recognition by shape analysis of images. *Pattern Recognit.* **23,** 539–544 (1990).

93. Wang, N. X. R., Cullis-Suzuki, S. & Branzan Albu, A. Automated Analysis of Wild Fish Behavior in a Natural Habitat. in 21–26 (ACM, 2015). doi:10.1145/2764873.2764875

94. Rova, A., Mori, G. & Dill, L. M. One fish, two fish, butterfish, trumpeter: Recognizing fish in underwater video. in *In IAPR Conference on Machine Vision Applications* (2007).

95. Qin, H., Li, X., Liang, J., Peng, Y. & Zhang, C. DeepFish: Accurate underwater live fish recognition with a deep architecture. *Neurocomputing* **187,** 49–58 (2016).

96. Spampinato, C. *et al.* Automatic fish classification for underwater species behavior understanding. in 45–50 (ACM, 2010). doi:10.1145/1877868.1877881

97. Spampinato, C., Chen-Burger, Y.-H., Nadarajan, G. & Fisher, R. Detecting, Tracking and Counting Fish in Low Quality Unconstrained Underwater Videos. *Proc 3rd Int Conf Comput. Vis. Theory Appl. VISAPP* 514–519 (2008).

98. Allen, J., Xu, R. & Jin, J. Object tracking using CamShift algorithm and multiple quantized feature spaces. in **100,** 3–7 (2004).

99. Perazzi, F. *et al.* A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation. in *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on* 724–732 (IEEE, 2016). doi:10.1109/CVPR.2016.85

100. Cordts, M. *et al.* The Cityscapes Dataset for Semantic Urban Scene Understanding. (20160406).

101. Patino, L., Nawaz, T., Cane, T. & Ferryman, J. PETS 2017: Dataset and Challenge. in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* 2126–2132 (2017). doi:10.1109/CVPRW.2017.264

102. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. (20151210).

103. *Sustainable Fisheries Act*. (1996).

104. *Magnuson-Stevens Fishery Conservation and Management Reauthorization Act of 2006*.

105. Lorenzo Ciannelli. Interviewed by Samm Newton. *Lorenzo Ciannelli*. (2017).

106. Council Staff | Pacific Fishery Management Council. Available at: https://www.pcouncil.org/council-operations/meet-the-council-staff/. (Accessed: 6th February 2018)

107. Eagle, J., Goldberg, M. & Sterne, J. Domestic Fishery Management. in *Ocean and Coastal Law and Policy* 305–329 (American Bar Association, Section of Environment, Energy, and Resources, 2015).

108. Regional Operating Agreement between the PFMC and NMFS. (2017).

109. Kuhn, E. Science And Deference: The “Best Available Science” Mandate is A Fiction in the Ninth Circuit. *Environmental Law Review Syndicate* (2016).

110. *Endangered Species Conservation Act of 1969*. *Stat.* **83,** 275–278 (1969).

111. Sullivan, P. *et al.* Defining and Implementing Best Available Science for Fisheries and Environmental Science, Policy, and Management. *Am. Fish. Soc.* **31,** 460–460 (2006).

112. Schwaab, E. Taking Stock: The Magnuson-Stevens Act Revisited: The Magnuson Act Thirty-Five Years Later. *Roger Williams Univ. Law Rev.* **17,** 14–20 (2012).

113. *Ocean and coastal law and policy*. (American Bar Association, Section of Environment, Energy, and Resources, 2015).

114. *National standards for fishery conservation and management*. *U.S. Code* **16,** (2007).

115. National Research Council. *Improving the Use of the ‘Best Scientific Information Available’ Standard in Fisheries Management*. (The National Academies Press, 2004). doi:10.17226/11045

116. Magnuson-Stevens Act Provisions; National Standard 2-Scientific Information. *Federal Register* (2013). Available at: https://www.federalregister.gov/documents/2013/07/19/2013-17422/magnuson-stevens-act-provisions-national-standard-2-scientific-information. (Accessed: 4th May 2018)
117. Samek, W., Wiegand, T. & Müller, K.-R. Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. *ArXiv170808296 Cs Stat* (2017).
118. Lackey, R. Perspective: Ecological Policy. *Fisheries* **31,** 286–286 (2006).
119. Lackey, R. Keep science and scientists credible: avoid stealth policy advocacy. *Bull. Ecol. Soc. Aust.* **46,** 14–15 (2016).
120. Pacific Fishery Management Council (PFMC). *Pacific Coast Salmon Fishery Management Plan for Commercial and recreational Salmon Fisheries off the Coasts of WA, OR, and CA as revised through Amendment 19*. (PFMC).
121. Shaw, W. & Conway, F. *Responses to the West Coast Groundfish Disaster: Lessons Learned for Communities and Decision Makers*. (2007).
122. *Pacific Coast groundfish plan (CA,OR,WA) :environmental impact statement.* (1982).
123. *Science and its role in the National Marine Fisheries Service*. (National Academy Press, 2002).
124. Slovic, P. Perception of Risk. *Science* 280–285 (1987).
125. *Upstream: salmon and society in the Pacific Northwest*. (National Academy Press, 1996).
126. JENNI, K. & LOEWENSTEIN, G. Explaining the Identifiable Victim Effect. *J. Risk Uncertain.* **14,** 235–257 (1997).
127. Daniels, B. Come Hell and High Water: Climate Change Policy in the Age of Trump. *FIU Law Rev.* **13,** (2018).
128. Juan-Jorda, M., Barth, J., Clarke, M. E. & Wakefield, W. Groundfish species associations with distinct oceanographic habitats in the Northern California Current (vol 18, pg 1, 2009). *Fish. Oceanogr.* **26,** 705–705 (2017).
129. Huyer, A. Coastal upwelling in the California current system. *Prog. Oceanogr.* **12,** 259–284 (1983).
130. Finley, C. *All the fish in the sea: maximum sustainable yield and the failure of fisheries management*. (University of Chicago Press, 2011).
131. Hilborn, R. & Walters, C. J. Role of Stock Assessment in Fisheries Management. in *Quantitative fisheries stock assessment: choice, dynamics and uncertainty* 3–21 (Kluwer Academic Publishers, 2001).
132. Brown, V. A., Harris, J. A. & Russell, J. *Tackling Wicked Problems: Through the Transdisciplinary Imagination*. (Routledge, 2010).
133. Habermas, J. *Communication and the Evolution of Society*. (Beacon Press, 1979).
134. Roderick J. Lawrence. Deciphering Interdisciplinary and Transdisciplinary Contributions. *Transdiscipl. J. Eng. Sci.* **1,** 125–130 (2010).

135. *Enhancing Communication & Collaboration in Interdisciplinary Research*. (SAGE Publications, Inc, 2013).

136. Cheruvelil, K. S. *et al.* Creating and maintaining high-performing collaborative research teams: the importance of diversity and interpersonal skills. *Front. Ecol. Environ.* **12,** 31–38 (2014).

137. Eigenbrode, S. D. *et al.* Employing Philosophical Dialogue in Collaborative Science. *BioScience* **57,** 55–64 (2007).

138. *Perspectives in interdisciplinary and integrative studies*. (Texas Tech University Press, 2015).

Appendix A. FishLabeler Usage Instructions

Introduction

FishLabeler is a piece of software written to expedite the video labeling process. Specifically, it allows stepping through the video on a frame-by-frame basis, annotating the frames with pixel-wise segmentation labels or bounding boxes, and adding free-text annotations for frames. The resultant metadata (i.e. bounding box lists, segmentation masks, and free-text data) are written out to disk. To simplify distribution, the application is dockerized, and so in theory is relatively portable – the main challenge is getting the X-forwarding to work, since this is a GUI-based application. To date, it works on Linux, Windows, and Mac systems. You can find the docker image at the docker hub repository [1] and the source code at the git repository [2] (under the *dockerized* branch).

Before You Run

Check if your docker image is up to date: in the terminal, run docker pull alrikai/nrt to pull the latest image from the docker hub repository. You can see which docker images are on your system by running docker images. If docker is not running, then begin running it (if it's not open, then you should get an error message when running any docker commands).

For example, running docker images gives:

```
alrik@kai:~$ docker images
REPOSITORY  TAG      IMAGE ID       CREATED      SIZE
alrikai/nrt latest   a63170ce013c   9 days ago   1.33GB
```

Old docker images will clutter your system over time; old docker images can be deleted by running (on the command line):
docker system prune

Linux (Tested on Ubuntu 17.10 and 16.04)

Installation Requirements and Setup

1. Install Docker CE – follow instructions on the page
2. Follow instructions on running docker as non-root user. This is important for sharing the X11 socket.

Running Docker Image

1. Make sure your docker image is up-to-date section 2

[1] https://hub.docker.com/r/alrikai/nrt/
[2] https://bitbucket.org/alrikai/fishlabeler

2. Run the docker image from the command line:

*docker run -ti --rm -u $(id -u) -v <src data dir>:/data *
*-v <dst data dir>:/outdata -e DISPLAY=$DISPLAY *
-v /tmp/.X11-unix:/tmp/.X11-unix alrikai/nrt bash

Note that the −*u name* option sets the username within the docker container. On one system, I have it set to NRTFish (since this is the username that is set in the Dockerfile), but it seems to work with the current user (*$(id -u)*) as well. e.g.

*docker run -ti --rm -u $(id -u) -v /home/alrik/Data:/data *
*-v /home/alrik/Data:/outdata -e DISPLAY=$DISPLAY *
-v /tmp/.X11-unix:/tmp/.X11-unix alrikai/nrt bash

3. Run commands as detailed in section 6.

Windows

Installation Requirements and Setup

1. Install Docker (if Win10 Pro, install Docker CE, otherwise install Docker Toolbox).
2. Install XMimg using default options.
3. Open Docker ‘Settings’, select ‘C’ (or whatever drive has the video data) as a shared drive.

Running Docker Image

1. Make sure your docker image is up-to-date section 2
2. Run XLaunch if XMimg is not running, note the display number (defaults to 0), and select the ’No Access Control’ checkbox under the ’Additional Parameters’ dialog. Leave everything else as defaults. Or, you can save the profile to disk to expedite the process in the future.
3. Open the XMimg log in the toolbar (right click, ”view log”, if open), note the XMimg IP address (the one next to *newAddress*).
4. Open Powershell if Docker CE, or docker quickstart terminal if Docker Toolbox – see below for commands depending on if you’re using Docker toolbox or Docker CE.
5. Start docker if it’s not running.

Docker CE

1. *docker run -ti --rm -u NRTfish -v <src data dir>:/data \ -v <dst data dir>:/outdata *
   -e DISPLAY=<ip addr>:<display #> alrikai/nrt bash

The commands between <brackets> should be substituted with actual values (and the brackets omitted). As a concrete example,

*docker run -ti --rm -u NRTfish -v C:/:/data -v C:/:/outdata *
-e DISPLAY=192.168.56.1:0 alrikai/nrt bash

2. Run commands as detailed in section 6.

Docker Toolbox

1. *docker run -ti --rm -u NRTfish -v "<src data dir>":/data *
   *-v "<dst data dir>":/outdata *
   -e DISPLAY=<ip addr>:<display #> alrikai/nrt bash

   ... where the parts between <brackets> should be substituted with actual values (and the brackets omitted). As an example,

   *docker run -ti --rm -u NRTfish *
   *-v "/c/Users/Katlyn/Documents/NRT/NRTDATA/Source":/data *
   *-v "/c/Users/Katlyn/Documents/NRT/NRTDATA/Output":/outdata *
   -e DISPLAY=10.248.234.174:0 alrikai/nrt bash

2. Run commands as detailed in section 6.

Mac

Installation Requirements and Setup

1. Install homebrew (if needed)
2. Install Docker
3. Install XQuartz (or via homebrew →brew cask install xquartz)
4. Install socat (brew install socat)

Running Docker Image

1. Make sure your docker image is up-to-date section 2
2. Run socat TCP-LISTEN:6000,reuseaddr,fork UNIX-CLIENT:\"$DISPLAY\" in a terminal, leave that running
3. Run XQuartz in a new terminal open -a XQuartz
4. Determine IP address for X-forwarding – ifconfig | grep inet
5. Run the docker image from the command line:

*docker run -ti --rm -u NRTfish -v <src data dir>:/data *

-v <dst data dir>:/outdata -e DISPLAY=<ip addr>:<display #> \ -v /tmp/.X11-unix:/tmp/.X11-unix alrikai/nrt bash

Note that the *−u name* option sets the username within the docker container. On one system, I have it set to NRTFish (since this is the username that is set in the Dockerfile), but it seems to work with the current user (*$(id -u)*) as well. e.g.

*docker run -ti --rm -u NRTfish \
-v /Users/SamanthaNewton/Drive/NRTLabeling:/data \
-v /Users/SamanthaNewton/Drive/NRTLabeling/DockerOutput:/outdata \ -e DISPLAY=10.249.17.255:0 -v /tmp/.X11-unix:/tmp/.X11-unix alrikai/nrt bash*

6. Run commands as detailed in section 6.

Running the FishLabeler Application

In all cases, running the *docker run* command in the above sections will download (if required) and launch the docker container, build the application, and navigate to the fishlabeler directory. The next important step is to run the application – to do so, run the **LabelFish.sh** script, passing in 2 command-line arguments:

1. The path to the video to label (using the mounted path directory, i.e. starting from /data).
2. The path to the directory to write the video frames and labeling metadata(also using the mounted paths, i.e. starting from /outdata).

e.g. The following command would extract video 20130117144639.mts and output all the resultant data (and metadata) to /outdata/NRT/20130117144639. Where exactly this is on your host filesystem depends on your docker run command arguments, specifically the <src data dir> and <dst data dir> arguments. If, for example, <src data dir> is /home/alrik/Data, then the input file would be at /home/alrik/Data/NRT/20130117144639.mts. Similarly, if the <dst data dir> argument is /home/alrik/Data/NRTFish, then the results would end up at /home/alrik/Data/NRTFish/NRT/20130117144639.

$ bash LabelFish.sh /data/NRT/20130117144639.mts /outdata/NRT/20130117144639

The **LabelFish.sh** script will extract the individual frames of the video and write them to the specified path as jpeg images, extract some video metadata, and then launch the labeling application. If there are already frames in the specified output directory, then the script will skip the frame extraction step. If using the script, the Fishlabeler application will pop up with the specified (i.e. newly-extracted or already-existing) video sequence selected after the frame extraction completes. In the event that the application is run manually (i.e. running the application executable directly), then there are two options:

1. No command-line arguments are provided: You should see a dialog UI pop up, and use that to navigate to the directory of frames you want to label. Once you've selected your target directory, the main window will pop up, and you can begin labeling.

2. The user passes in the target directory, which already has its frames and video metadata extracted, as is done in the **LabelFish.sh** script. Then no dialog UI will be displayed, and the user-specified video sequence will be opened. e.g. ./FishLabeler /data/NRTFishAnnotations/20130117144639

Also, if you want to skip the frame extraction and metadata extraction step (i.e. if you've already gone through the extraction steps), then you can also just run the executable directly from the build directory.

Using the FishLabeler Application

Fishlabeler is supposed to be relatively bare-bones, so it should be straightforward to use. The default window will look similar to section 7. Different fish should be given different instance IDs, which correspond to different colored bounding boxes. The specific functionality provided by the Fishlabeler application is detailed in the following sections:

- Frame navigation: subsection 7.1
- Segmentation (per-pixel labeling): subsection 7.2
- Detection (bounding boxes): subsection 7.3
- Detection interpolation: subsection 7.4
- Hotkeys: subsection 7.5

When labeling, do notbother with fish < 10 cm, as (in my experience) they are too small to be easily seen by eye. The laser range-finder is calibrated to be roughly 10 cm on the ocean floor, so this is a good heuristic to use for determining which fish to label and which to ignore.

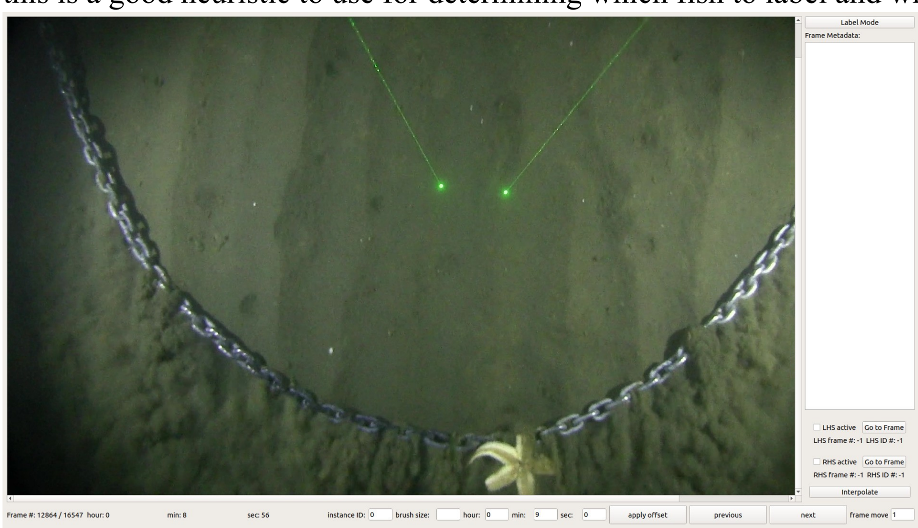


Figure 1: The Fishlabeler UI window of a beamtrawl video sequence frame (and starfish casualty).

Frame Navigation

The bottom panel has the frame navigation tools; the left side displays the current frame number and timestamp, the edit box for instance ID controls what ID the bounding box to be drawn is assigned, the timestamp edit fields allow jumping to a specific timestamp (enter the timestamp, click the apply offset button), while prev and next buttons move through frames according to a fixed frame offset, as set in the edit field frame move, which defaults to 1. Corresponding hotkeys can be found in subsection 7.5. The brush size edit field controls the annotation pixel size (i.e. for the bounding box border, or the number of pixels that each click corresponds to in segmentation).

Segmentation

Segmentation is the process of assigning labels to pixels, i.e. designating the labeled pixels as belonging to a groundfish, and the non-labeled pixels as being background (or anything not-groundfish). This differs from detection in that segmentation operates in pixel-space. As such, it is correspondingly more work intensive to label, since one has to assign labels to every pixel of interest in the scene. However, segmentation also gives more information about the scene than detection does, so it is valuable to do. To perform segmentation in Fishlabeler, click the Label Mode button in the upper right until it says Segmentation, then clicking on the displayed frame will produce labels for the affected pixels, contingent on the current brush size and instance ID. An example is shown in section 7.2. In the interest of full disclosure, segmentation annotations are extremely tedious to label, and despite having implemented the functionality for it in the application, I gave up on segmentation labeling after a few dozen frames.

Detection

Detection is the process of enclosing a target object within a (axis-aligned) rectangle, such that the entirety of the object is contained within the rectangle, and as few pixels that *do not*belong to the target object are included as possible. This is the weakest form of localization information, but is fast to draw and to process, and enables detection and tracking algorithms (which form the basis of many video analysis algorithms). The rectangle is commonly referred to as a bounding box. To draw detection annotations in Fishlabeler, ensure the annotation mode is not set to Segmentation, and then draw the bounding box by first clicking on the frame for the upper-left corner of the bounding box, dragging the mouse to the bottom right corner, then releasing the button. The bounding box will then be shown on the frame. Clicking on an existing bounding box outline and dragging it elsewhere will translate the bounding box according to the mouse movement rather than creating a new bounding box. An example labeled frame is shown in section 7.3.

Interpolation

The aim of the interpolation module is to automate some of the tedious detection labels; currently, only linear interpolation is implemented, so it works best when the apparent fish motion is linear. Note that since the camera is also moving (in a non-linear fashion), linearity assumptions are limited, and the apparent fish size is contingent on where the fish is in the scene and varies heavily based on the distance of the fish from the camera (if it swims closer to the camera, its apparent size increases). Thus, it is not a good idea to rely solely on interpolation, as it will often be inaccurate. However, it can give good results in certain situations (constant, linear motion), so it is still of use. I've found it to be best applied over relatively short time-scales, and/or when the fish is stationary, and the boat is moving at a relatively constant rate.

To use the interpolation, the starting and ending bounding boxes have to be provided, and the interpolation logic will 'guess' the bounding boxes in between. To provide the starting (LHS) and ending (RHS) boxes, check the corresponding checkbox in the lower-right side of the window, then draw the box on the frame. This will register that annotation to either the LHS or RHS, based on which checkbox was chosen. Once a box has been registered, the frame index that the box is at, and the instance ID of the box will be displayed. Clicking the Go to frame button will change the current frame to the associated bounding boxes' frame. Once the LHS and RHS boxes have been chosen, click the interpolate button to compute the intervening frame's bounding boxes. Navigating to these frames should display them, and the user can adjust the boxes accordingly (either by translating the boxes by dragging them by their frame, or by undoing them (using ctrl + z, see subsection 7.5 for hotkey details) and re-drawing, if necessary. The LHS and RHS instance IDs should match, since it is assumed that they refer to the same fish (and indeed, they should).

Hotkeys

- **n**: next frame, according to the current frame move value (hold the n-key down to move quickly)
- **p**: previous frame, according to the current frame move value (hold the p-key down to move quickly)
- **ctrl + z**: undo the last bounding box or pixel-label on the current frame. This can be used to erase existing annotations (e.g. hold down ctrl, then press z to remove detections / pixel labels)
- **ctrl + r**: redo the last bounding box or pixel-label on the current frame that was undone. This can be used to restore an undo operation; i.e. if you erroneously erased an existing annotations with ctrl + z, then ctrl + r will restore it. Once you navigate to a different frame, the re-do history will be lost however.
- **f**: toggle the frame display mode; either full-sized (which generally requires scrolling), or zoomed-to-fit (no scrolling required, but the image is smaller to fit the screen).

After Running the FishLabeler Application

The metadata results will be in the output folder based on how the docker command was run. There will be 3 folders with the labeling results, Annotations, Detections, Metadata for segmentation masks, detection bounding boxes, and text information respectively. These are not uploaded anywhere automagically, so (unless you really enjoy labeling), upload them somewhere before deleting anything. Currently, the de-facto location is at box.com under the NRT/Labeling Metadata folder. After labeling a sequence, you can safely delete the video frames that get extracted to your system, as they can always be regenerated from the video file (and after you've uploaded the metadata, you can delete those folders too).

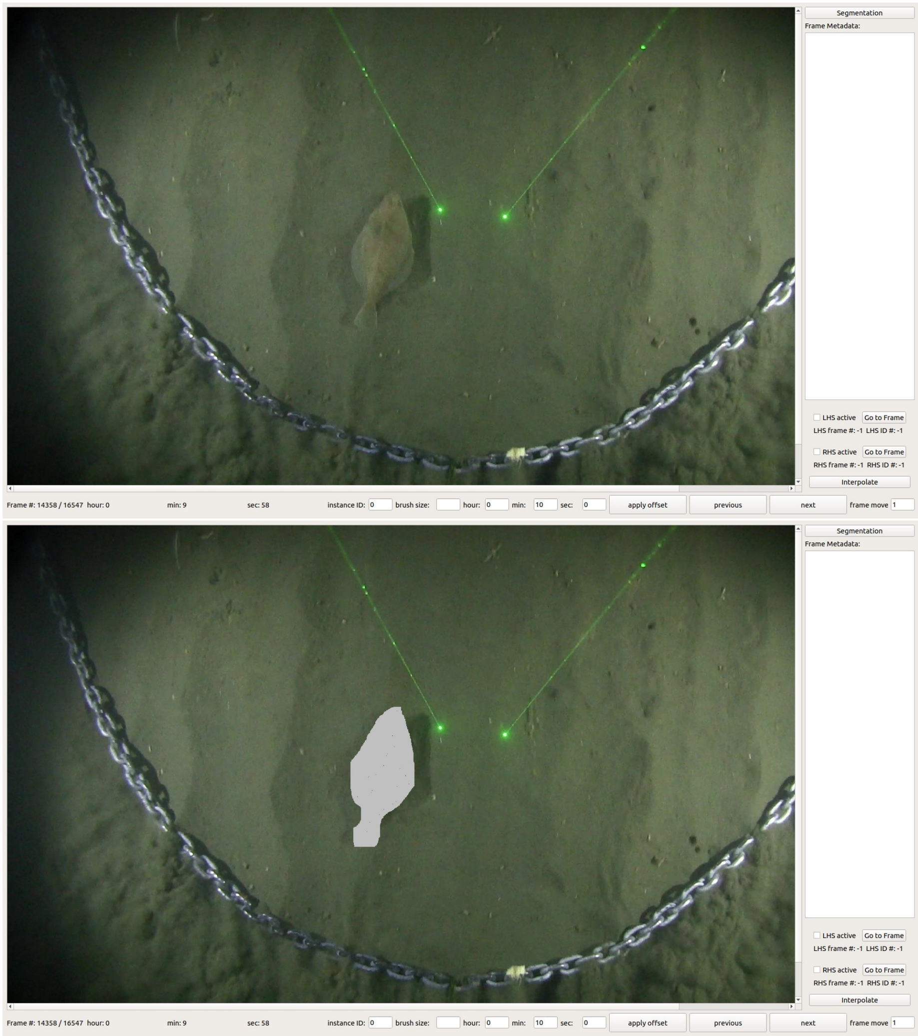


Figure 2: Before and after screenshots of a segmentation. Note that the upperright button is in Segmentation mode. The color of the segment corresponds to the instance ID (in this case, ID 0). The goal in creating a segmentation annotation is to cover the pixels belonging to the fish as accurately as possible.

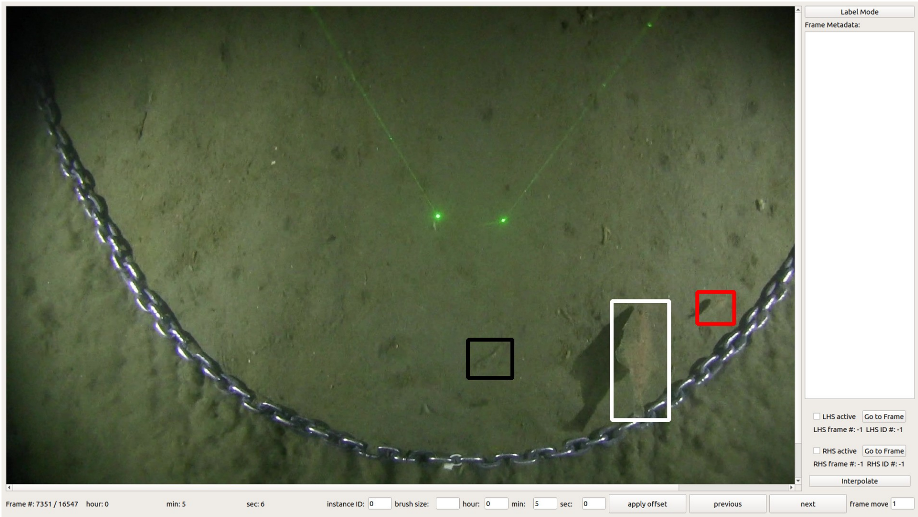


Figure 3: Screenshot of the Fishlabeler UI with three fish detections annotated. The different colored bounding boxes correspond to different instance IDs. The bounding boxes should cover the entirety of the fish, but be as tight as possible to the fish (i.e. not have extraneous non-fish background pixels with in the bounding box).

Appendix B: Statement of Contributions

Ongoing Activities		
Activity	*Team member*	*When*
Attend and Facilitate Group Meetings	All	Weekly
Attend and Present at Cluster Meetings	All	Each Term
Developed outline for TDR	All	Winter term
Developed Research Questions	All	Fall Term
Wrote TDR	All	Spring / Summer

TDR Process				
	Main Contributor	*First Reader*	*Second Reader*	*Third Reader*
Introduction	Katlyn	S/A	Alrik	Samm
Background	All	Katlyn	Samm	Alrik
Big Data	Alrik & Samm	Samm	Alrik	Katlyn
Environmental Inference	Alrik & Katlyn	Alrik	Katlyn	Samm
Decision Making	Katlyn & Samm	Katlyn	Samm	Alrik
Methods	Samm	S/A	Alrik	Katlyn
Discussion	Alrik	S/A	Katlyn	Samm
Conclusion	Alrik	S/A	Katlyn	Samm
References	All	Samm	N/A	Katlyn

Key	The readers in order basically go from the smasher who is the least refined, to the finisher who is the most refined
S/A	Means single author, so smashing is not necessary
Main Contributor	Who is delivering the main content
First Reader	The smasher, the person that takes the main content and put it together in a meaningful way
Second Reader	The content editor, cutting un-needed info, making sure it makes sense, making sure it flows etc
Third Reader	The finisher, this is editing for grammar, spelling, punctuation, correct in line citations, readability, etc