

EeLISA: Combating Global Warming Through the Rapid Analysis of Eelgrass Wasting Disease

Brendan H. Rappazzo,¹ Morgan E. Eisenlord,² Olivia J. Graham,²
Lillian R. Aoki,² Phoebe D. Dawkins,³ Drew Harvell,² Carla P. Gomes¹

¹ Department of Computer Science, Cornell University

² Department of Ecology and Evolutionary Biology, Cornell University

³ Department of Ecology and Evolutionary Biology, University of California, Irvine
{bhr54,me367,ojg5,lra53}@cornell.edu, dawkinsp@uci.edu, cdh5@cornell.edu, gomes@cs.cornell.edu

Abstract

Global warming is the greatest threat facing our planet, and is causing environmental disturbance at an unprecedented scale. We are strongly positioned to leverage the advancements of Artificial Intelligence (AI) and Machine Learning (ML) which provide humanity, for the first time in history, an analysis and decision making tool at massive scale. Strong evidence supports that global warming is contributing to marine ecosystem decline, including eelgrass habitat. Eelgrass is affected by an opportunistic marine pathogen and infections are likely exacerbated by rising ocean temperatures. The necessary disease analysis required to inform conservation priorities is incredibly laborious, and acts as a significant bottleneck for research. To this end, we developed EeLISA (Eelgrass Lesion Image Segmentation Application). EeLISA enables ecologist experts to train a segmentation module to perform this crucial analysis at human level accuracy, while minimizing their labeling time and integrating into their existing workflow. EeLISA has been deployed for over 16 months, and has facilitated the preparation of four manuscripts including a critical eelgrass study ranging from Southern California to Alaska. These studies, utilizing EeLISA, have led to scientific insight and discovery in marine disease ecology.

Introduction

Humanity has forged a new age known as the Anthropocene, marked by unprecedented influence of human activity on climate and the environment. Studying the damage to earth's ecosystems presents a massive challenge, as it requires collecting vast amounts of samples and fine-grained analysis. A major limitation of ecological studies is the expertise required to complete the necessary fieldwork and subsequent analysis. Consequently, the human bandwidth to do these crucial studies is very limited.

However, recent advances in AI have demonstrated a remarkable ability to perform analysis at or above human level [Huang et al. 2017; Chen et al. 2017], even when the task requires expert level knowledge [Lin, RoyChowdhury, and Maji 2015]. Furthermore, once an AI system has been trained to perform at expert level accuracy, it can be shared



Figure 1: Example of a healthy eelgrass bed.

and scaled to fit demand all across the world, allowing for huge leaps in the democratization of human knowledge and insight.

One application that can benefit from AI are seagrasses, one of the world's most productive ecosystems composed of the only flowering plants in the marine environment. These complex ecosystems provide essential services for our planet, such as the sequestration of consequential amounts of carbon, actively mitigating the effects of anthropogenic climate change [Waycott et al. 2009]. Eelgrass (*Zostera marina*) is one of the most widespread seagrass species, forming extensive meadows throughout the temperate Northern Hemisphere [Moore et al. 2006]. However, eelgrass is susceptible to Eelgrass Wasting Disease (EWD), which has historically caused precipitous decline, and is threatening to do the same now [Lefcheck et al. 2017; Groner et al. 2016].

In order to assess EWD, ecologists analyze relevant disease metrics of prevalence and severity. Prevalence is a basic measure of whether or not the plant is afflicted by EWD, and is determined by the absence or presence of EWD-associated lesions. Severity yields the extent of the infection, and is calculated by the ratio of lesioned tissue to total tissue. Using these two metrics, ecologists characterize eelgrass bed health. With auxiliary measurements of environmental conditions (i.e. temperature, salinity, pH), ecologists

Biome	Area (ha)	Value (USD/ha/year)
Open Ocean	33,200	660
Coastal	3,102	8,944
Seagrass/Algae Beds	234	28,916
Terrestrial Forests	4,261	3,800
Terrestrial Grass	4,418	4,166

Table 1: Summary of different ocean and terrestrial flora in terms of acreage as well as an evaluation of the economic value of the eco-services (including carbon sequestration) they provide on a per area basis, where ha is hectare. Table/Data from [Costanza et al. 2014] for year 2011. Note Seagrass/Algae Beds are a sub category to coastal.

can begin to identify the complex drivers of the disease in the field.

Despite the global significance of eelgrass ecosystems and the myriad services they provide, the methods by which EWD is studied are time intensive and not widely standardized between groups. These limitations in methods pose challenges for studying this disease on greater spatial scales and impede robust investigation of disease drivers. Overcoming these limitations and better understanding the dynamics of EWD will assist in informing management decisions to reduce eelgrass declines.

Specifically, the most standardized method to measure EWD involves the manual segmentation of diseased and healthy tissue in digital scans of eelgrass blades. This process of manual tissue annotation is extremely tedious and can take upwards of 30 minutes for a single eelgrass blade digital scan. This process is a massive bottleneck in the analysis pipeline, and effectively inhibits widespread analysis.

To address this issue, we developed the Eelgrass Lesion Image Segmentation App (EeLISA). This system uses deep learning methods to automatically annotate healthy and lesioned tissue, while minimizing the amount of time required by experts to do labeling, and integrating into their established workflow. This workflow provides an intuitive mechanism for the users to verify the output of our methods. This system was used by experts to simultaneously generate the training data needed to train large segmentation convolutional neural network (CNN) architectures, and generate the analysis data the ecologists needed to further their studies.

The EeLISA system has been deployed for over 16 months, and has been used to analyse thousands of eelgrass blades at human level accuracy. This has allowed ecologists to test hypotheses about drivers of climate change along a latitudinal gradient. Additionally, the EeLISA system has led to scientific discovery and insight by facilitating rapid disease analysis across multiple years and sites, and has led to four manuscripts currently in preparation/in review. Further, this system is currently being utilized in a large NSF funded study of EWD, with 32 sample sites spanning the entire West Coast of the United States, from Southern California to Alaska. Looking to the future, EeLISA has even greater application potential given the sheer extent of this essential marine habitat. Eelgrass dominates the coasts of

the Temperate North Atlantic, the Mediterranean, and the Temperate North Pacific, highlighting the untapped, global potential of this application and further legitimizing its practicality and future use [Moore et al. 2006].

Eelgrass

Importance of Eelgrass

Eelgrass is one of the most widespread seagrass species, dominating the Northern Hemisphere. Its meadows are ecologically and economically important habitat, and provide shoreline protection, high value nutrient cycling, and globally significant carbon sequestration [Waycott et al. 2009]. Seagrasses are one of the most significant carbon sinks on the planet, burying carbon 35x faster than tropical rainforests [McLeod et al. 2011; Macreadie et al. 2014]. Monetary valuations for seagrass eco-services are estimated at \$28,916/ha/yr [Costanza et al. 2014], much greater than that of coastal biomes, terrestrial forests, or grasses (see Table 1).

Eelgrass Wasting Disease

Eelgrass Wasting Disease (EWD) is caused by a pathogenic protist, *Labyrinthula zosterae* (Lz), and has been implicated in drastic declines of eelgrass habitat [Waycott et al. 2009]. In the 1930s, eelgrass beds along the Atlantic coasts of North America and Europe experienced a severe EWD outbreak that decimated over 90 percent of their population. This led to a major disruption of coastal food chains, and of global carbon sequestration [Renn 1936; Short, Ibelings, and Den Hartog 1988]. Unfortunately, eelgrass beds on both the Atlantic and Pacific Coasts of North America have shown declines in recent years [Lefcheck et al. 2017], and may be partially attributed to EWD.

Effect of Sea Temperature Rise

Strong evidence supports that marine heat waves – a direct result of global warming – threaten eelgrass beds [Smale et al. 2019]. Outbreaks of EWD in temperate systems are associated with increased temperatures, with highest disease prevalence recorded from surveys during the warmest months of the year [Bockelmann et al. 2013]. In experimental trials, the causative agent of EWD exhibited faster growth at elevated temperatures *in vitro*, while causing larger lesions *in vivo* [Dawkins et al. 2018]. Furthermore, a modelling study simulated eelgrass bed reductions mediated by EWD during warm periods [Bull, Kenyon, and Cook 2012].

Study of EWD and Analysis Bottleneck

There is still much to study about the mechanics and factors of EWD spread. With more studies, at larger scales, ecologists hope to build up a map of current disease levels, as well as understand what underlying factors contribute to higher disease spread, and higher disease severity. However, the measuring of eelgrass blades for disease metrics is an extremely tedious task that presents many challenges, and has gone through many iterations. To standardize EWD analysis researchers began to create digital images of their

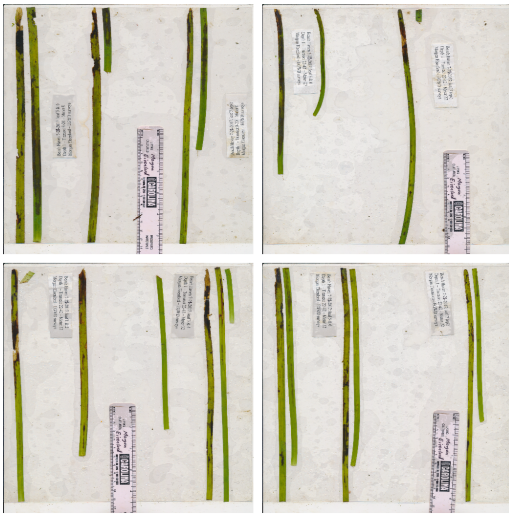


Figure 2: Four eelgrass scans from sites in San Juan Islands, WA.

eelgrass blades, placing them between clear, plastic transparency papers and scanning them (see Figure 2 for examples). With these digital scans they then use image processing software to carefully segment the amount of lesioned and healthy tissue to get precise measurements of prevalence and severity. This practice gave much more precise results, and also allowed for a way to track scans and review how they were measured at later dates. However, this process is disruptively laborious; completing disease analysis on a single scan containing 1-10 blades conservatively takes 30 minutes to complete. Further, there remained the issue of consistency amongst researchers in terms of measuring disease areas. A single eelgrass survey generates approximately 100 scans, requiring at least 50 hours of manual disease analysis. Further, the task is complicated enough that it must be done by domain experts and cannot be outsourced to the general public. Often 1-2 experts are left to do all 50 hours of analysis, meaning that the tedium and fatigue also led to inconsistencies in labeling methods.

The massive amount of time required to do the measurement of the disease metrics significantly limits research on larger spatial and temporal scales. Unfortunately, the data from larger scale studies is exactly what researchers need to make better predictive models of disease and its contributing factors, and to then put in place preventative measures.

Opportunity for AI

This bottleneck represents a perfect application for modern ML methods to provide automatic analysis of the disease metrics. A trained machine learning model would allow researchers a way to get consistent and accurate disease metrics at remarkable speed, and enable critical, large-scale research. To this end we develop an application that would ultimately allow the researchers to input eelgrass scans and receive as output all the relevant disease metrics they need, organized in their proper subdivisions. To do this, we first

needed to build up a labeled data set. However, given the extreme amount of time it takes to analyze even a single scan, we had to build a system that would allow researchers to produce the analysis data they still needed, while collecting the labels we needed to train our algorithms. We also wanted our system to aid in the labeling process and to allow ecologists a method to check the output of our models. The details of our system are discussed in the following section.

EeLISA System

Goal of System

The ultimate goal of the EeLISA system is to develop an AI system that can accurately, and at scale analyze eelgrass blades, segment the healthy and lesioned tissue, and then calculate and output disease metrics. However, because of the sensitivity of the data, it is necessary to have the output of our segmentation module be able to be checked and edited if necessary.

Further, because the annotation time can be upwards of 30 minutes for a single scan, it is necessary for our system to not get in the way of ongoing research. That is, our system can not increase labeling time, and has to output the data the ecology researchers require. To address this we developed the EeLISA system which accomplishes all of these tasks, while minimizing expert labeling time. The system is composed of three main components which are described below.

Components

Segmentation CNN Module The role or task of the segmentation module is to take as input an eelgrass scan, and to produce a classification per pixel. That is, given an image $\mathbb{X}$ and mask $\mathbb{Y}$:

$$\mathbb{X} \in \mathbb{R}^{H \times W \times M} \quad \mathbb{Y} \in \mathbb{R}^{+H \times W \times N}$$

where M is the number of color channels in the image (in this case RGB), and N is the number of classes, and H and

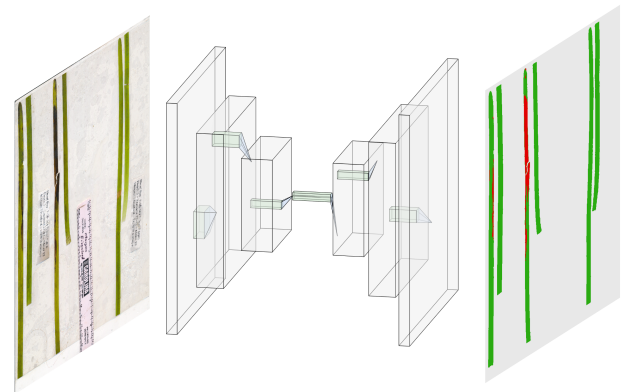


Figure 3: Schematic of CNN module. Using a fully convolution framework, our CNN module (center), takes as input an eelgrass scan (left) and produces a segmentation mask (right). Grey indicates background, green indicates healthy tissue, and red indicates diseased tissue. CNN schematic was generated via [LeNail 2019].

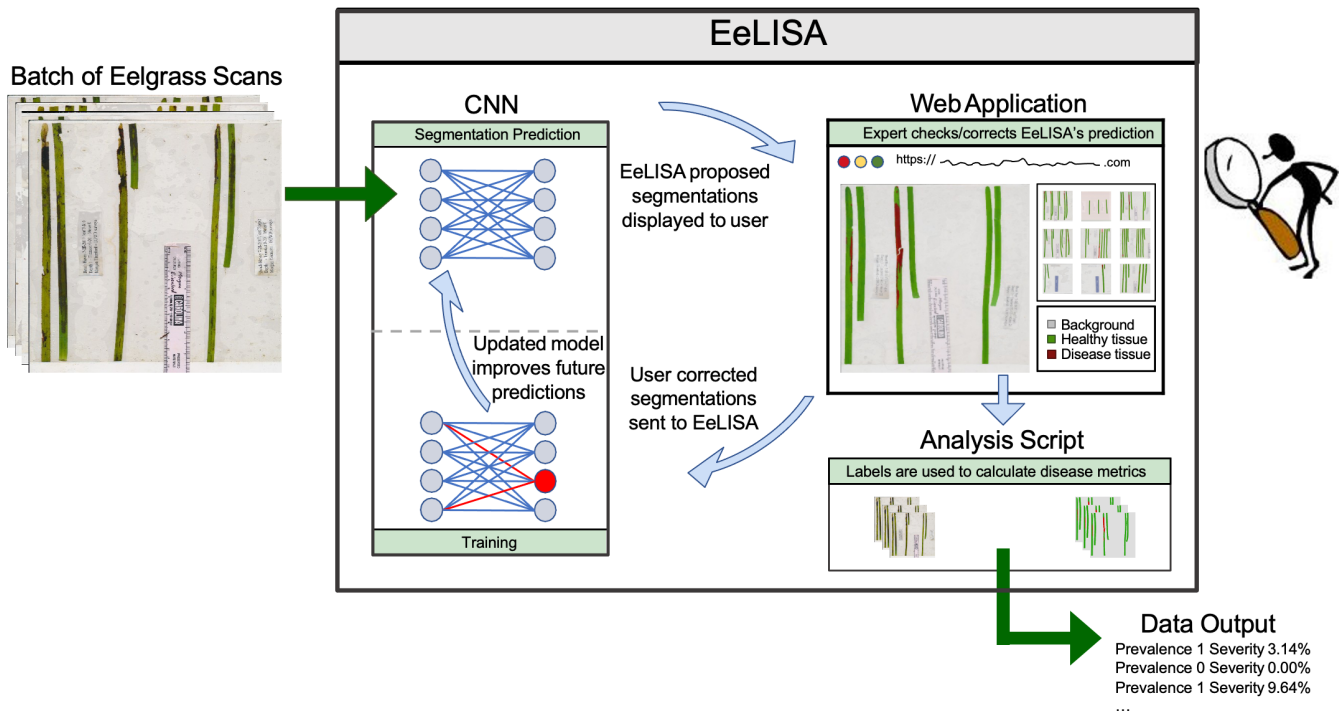


Figure 4: Schematic of EeLISA workflow. Scans are passed to EeLISA which generates predictions/segmentations of diseased and healthy area. Expert users can then check these predictions and correct them if necessary. Segmentations masks are then processed by an analysis script to export relevant data in an ordered manner for further study. Additionally, the corrected masks are then used to further train the segmentation model within EeLISA. In this way, EeLISA allows experts to obtain disease metrics in a timely fashion, while generating a labeled data set for a segmentation algorithm to train on. Once a data set of appropriate size is generated, the algorithm is able to predict with human level accuracy and users need only to verify potentially problematic scans.

W are the height and width of the image respectively, the segmentation module's task is to learn the following mapping:

$$f : \mathbb{X} \rightarrow \mathbb{Y}$$

The final prediction of the segmentation module is given by $\text{argmax } \mathbb{Y}$.

Specifically in the EeLISA system the segmentation module used was different variations of fully convolutional networks (FCNs) [Long, Shelhamer, and Darrell 2015]. The basic architecture is an encoding segment of convolutional and max-pooling layers that extracts features from the image and projects it into a low dimensional feature space. Then a decoding architecture that uses transpose convolutional layers to up-sample the embedding space back to the original image size, with as many channels as there are classes, providing a probability distribution over each class for each pixel. The final class is assigned by taking the class with the largest probability. A schematic of a FCN with an input eelgrass scan and output segmentation mask can be seen in Figure 3. More details about the specific architectures and methods are shared in the Machine Learning Methods section.

Labeling/Web Application Module The role of the web application module is to display the predictions of the segmentation module to the expert user, and allow them to verify or edit the predictions. The labeling module was written as a web application to make it accessible from any personal device, and enable the user to complete their labeling work remotely. The model is hosted on Amazon Web Services and could be accessed via any web browser. When navigating to the module, the user would be shown the directory structure that was given to EeLISA, where they could then navigate to the specific scans they wanted to inspect. When opening a particular scan, the user would see the image of the original scan, overlaid with transparent, colored polygons indicating the predictions of the segmentation module, with the color of the polygons indicating the classification. They can then edit the polygons, delete them entirely or add new ones, and indicate when they are done editing or verifying all the scans in a particular batch. A screenshot of the web application can be seen in Figure 5. The web application was built on top of the open source software available on GitHub¹. Details of

¹<https://github.com/Hitachi-Automotive-And-Industry-Lab/semantic-segmentation-editor/releases/tag/1.2.0>

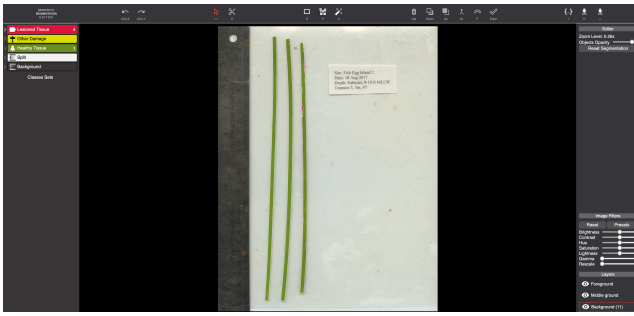


Figure 5: Screenshot of Web Application Module. A Sample scan is displayed with EeLISA's prediction overlaid as editable polygon masks.

the implementation are discussed in later sections.

Analysis Module The analysis module's task was to take the batch of images, and resulting expert verified or corrected segmentation masks, and produce a CSV file containing the disease metrics required by researchers. Specifically given a segmentation mask the module would calculate the diseased area and healthy area for each eelgrass blade, and report the resulting severity and prevalence disease metrics. The analysis module automatically organizes the output data according to relevant meta-data such as date, site location and further subdivisions. The ecologists can then easily use the meta-data tagged disease metrics as input to their own analysis.

Workflow

The general workflow of EeLISA is as follows. The segmentation module starts completely untrained and randomly initialized. The expert user selects a batch of eelgrass scans that they would like to analyze and calculate disease metrics for and uploads them to the EeLISA system. First, EeLISA runs the batch of scans through the segmentation module, and the segmentation masks are predicted. The scans and resulting masks are then uploaded to the server running the labeling module. The expert user will then go through each scan and corresponding label, and either verify that it is correct, or make the necessary edits to the segmentation masks. Initially, all predictions require editing since the segmentation module has not yet been trained. The corrected or verified labels are then downloaded, and sent to the analysis module, where the relevant disease metrics are calculated and sent to the expert user in the form of a CSV file. The corrected or labeled masks are also sent to the segmentation module, where they are then used to train the model.

Then the workflow repeats, with the user choosing the next batch of scans they would like to process. However, because the segmentation module has now been trained on some data, its predictions are more accurate, which in turn means the expert user will need less time to correct the predictions. Which means that they will be able to get their output analysis faster, but also that the segmentation module will have more data to train with faster. This cycle repeats as

Number Training Scans	789 ($\approx 75\%$)
Number Testing Scans	247 ($\approx 25\%$)
Scan Size	$512 \times 512 \times 3$
Label Size	$512 \times 512 \times 1$
Classes	3
Class 0	Background
Class 1	Healthy Tissue
Class 2	Lesioned Tissue

Table 2: Summary of eelgrass data set.

a positive feedback loop. A schematic of this workflow can be seen in Figure 4.

After working through many batches of scans (see detail in next section) the segmentation module had a sufficient data set to reach human level accuracy. In the next sections we describe the data set, as well as our ML methods on this data set.

Innovative Use of AI Technology

The AI technology used in EeLISA is deep segmentation neural networks and we believe EeLISA is an innovative use of this technology for two main reasons. The first is that, to our knowledge, this is the first AI system to provide rapid, fine-grained (per single blade) metrics of marine disease levels. That is, this is a novel and important use-case for AI segmentation methods. Secondly, the method by which the segmentation module is integrated into the existing workflow of the ecologists, and provides a positive feedback loop to lessen labeling time, while not necessarily conceptually novel, is certainly novel in a deployed application in this domain.

To further explain how the AI technology was developed, in this section we focus on the technical details for the training and testing of the deep learning based segmentation module, and the data used to train it.

Data Set

While the EeLISA system is an ongoing system, and more training data is being produced, the initial data set built with EeLISA will be the focus here. The data set consists of 1,036 scans. Of those scans we have designated 789 of them as the training set and 247 as the test set, representing an approximate 75/25 percent split for training and testing.

Each scan and mask has been scaled to be 512×512 pixels. The scans have the three standard RGB color channels, and the masks have a single channel, where each pixel is assigned an integer of value zero, one or two. Zero indicates the background class, one indicates healthy tissue, and two indicates lesioned tissue. A summary of the data set can be found in Table 2.

Data Pre-Processing

We augment the training set in the following way: for each image and mask pair we introduce two random rotations, flip the image on both the x and y axis, slide the image in two

directions with random magnitudes, zoom and crop two random portions of the image, and lastly do three random elastic deformations of varying random severity. Given the starting 789 labeled scans and these transformations we have a total of 9468 labeled scans. As this data set is relatively small, these augmentations are crucial to boost the model’s ability to generalize to new eelgrass scans, and in particular to generalize to eelgrass scans from new geographical locations. Following the augmentations, we then take the mean and standard deviation of each color channel for all 9468 scans, and before each image is passed to our model we subtract the mean values and divide by the standard deviation to ensure that each color channel has a zero mean and unit variance.

Of the 9468 labeled scans, we withhold 947 scans as a validation set. In summation, there are 8521 scans we use for training, 947 we use as a hold out validation set, and 247 scans that we use as a test set.

Machine Learning Methods

Models We use the following models: UNet [Ronneberger, P.Fischer, and Brox 2015], Deeplab V3 [Chen et al. 2017], and Tiramisu [Jégou et al. 2017]. In general, these models were all picked because they are extensions of the FCNs [Long, Shelhamer, and Darrell 2015] mentioned in previous sections. UNet is the most basic extension of the original FCN; UNet uses all convolutional layers, but extends the work by developing the down scaling and up scaling ‘U’ shape. The Tiramisu network is also an extension of the original FCN, as it is essentially a much deeper version of the traditional UNet, and uses ‘DenseBlocks’ borrowed from the DenseNet architecture [Huang et al. 2017]. DeepLab is also an extension of the FCN as it follows the basic architecture of UNet (except with an interchangeable backbone architecture), and further includes improvements like atrous convolutional layers which makes it a particularly powerful architecture. Specifically, the UNet model was picked because of its demonstrated success on medical data sets with few labels. DeepLab V3 was picked because of its atrous convolutional layers which we intuitively felt would be able to handle the different sizing scale of the lesions. Further, DeepLab V3 and Tiramisu were picked because of their demonstrated success on large scale data sets such as [Cordts et al. 2016].

Loss Functions We found empirically that a combined loss function of equal weighting yielded the best results. Specifically given the following definitions, where N is the number of classes, $\tilde{Y}$ is the prediction of our model and Y is the true label, Y_i and $\tilde{Y}_i$ represents a binary matrix that is 1 at each entry that is of class i and 0 elsewhere, the Dice loss is as follows:

$$L_D(\tilde{Y}, Y) = \sum_{i=0}^N \frac{2|\tilde{Y}_i \cap Y_i|}{|\tilde{Y}_i| + |Y_i|}$$

Cross Entropy is as follows:

$$L_{CE}(\tilde{Y}, Y) = - \sum_{i=0}^N Y_i \log \tilde{Y}_i$$

We used the following loss:

$$L_{Total} = L_D + L_{CE}$$

Optimizers We experimented with many different optimizers including: RMSProp, ADAM optimizer [Kingma and Ba 2015], and Stochastic Gradient Descent. We found empirically that the ADAM optimizer produced the best results, and all models were trained with a learning rate of .0001, a β_1 of .5 and a β_2 of .999, and the learning rate was lowered as the loss value of the validation set plateaued.

Batch Size Due to GPU memory limitations, the batch size of UNet and DeepLab are set to 8, whereas the batch size for Tiramisu is set to 2.

Implementation The training process was implemented using Python 3.7 and PyTorch, and was run on a Ubuntu based machine with two Intel Xeon Skylake processor with 36 cores, as well as 5 NVIDIA Tesla V100 GPU.

Metrics

While the Dice Score and Cross Entropy were the two loss functions used to optimize our models, we judged the criteria over seven different metrics. All metrics are listed below with a brief description for how they were calculated for a single scan. These single calculations were then averaged for each scan in the training or test set. Let N be the number of classes, $\tilde{Y}$ be the predicted segmentation mask, and Y be the true segmentation label and Y_i and $\tilde{Y}_i$ represents a binary matrix that is 1 at each entry that is of class i and 0 elsewhere. Additionally $\text{sum}(\dots)$ counts the number of instances of the expression being satisfied. For example $\text{sum}(\tilde{Y} = 2)$ is the number of pixels that have been classified as class two in the predicted segmentation mask.

Mean Dice Score Mean Dice Score is a standard metric for segmentation algorithms. It calculates the mean value for how much each class’ segmentation overlaps with the ground truth. It was calculated as follows:

$$\frac{1}{N} \sum_{i=0}^N \frac{2|\tilde{Y}_i \cap Y_i|}{|\tilde{Y}_i| + |Y_i|}$$

Mean Jaccard Score Mean Jaccard Score is a standard metric for segmentation algorithms. While it is very similar to the Dice Score, it can emphasize different kinds of losses. It was calculated as follows:

$$\frac{1}{N} \sum_{i=0}^N \frac{|\tilde{Y}_i \cap Y_i|}{|\tilde{Y}_i \cup Y_i|}$$

Mean Sensitivity Score Sensitivity is a measure of how likely a given model is to detect a class and we averaged this value over all classes. It was calculated as follows:

$$\frac{1}{N} \sum_{i=0}^N \frac{|\tilde{Y}_i = 1 \cap Y_i = 1|}{|Y_i = 1|}$$

Mean Specificity Score Specificity is a measure of how likely a given model is to correctly predict a class is not present, should it not be there, we averaged this value over all classes, as follows:

$$\frac{1}{N} \sum_{i=0}^N \frac{|\tilde{Y}_i = 0 \cap Y_i = 0|}{|Y_i = 0|}$$

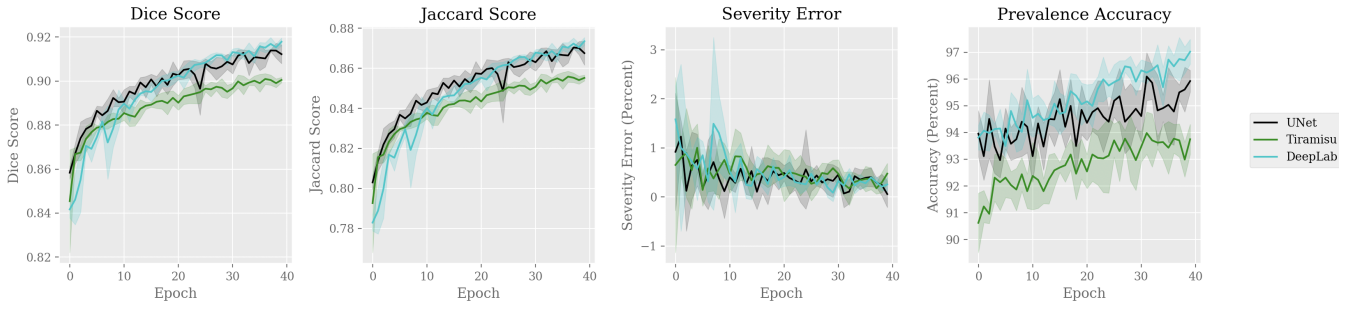


Figure 6: Results on validation set per epoch, line indicates mean value, whereas shaded region is standard deviation of runs.

Architecture	Dice Score	Jaccard Score	Sensitivity	Specificity	Severity	Absolute Severity	Prevalence
UNet	.917 +/-0.006	.874 +/-0.006	.937 +/-0.009	.992 +/-0.000	.075 +/-0.213	.554 +/-0.046	96.032 +/-0.865
DeepLab	.920 +/-0.001	.877 +/-0.001	.943 +/-0.006	.993 +/-0.001	.246 +/-0.132	.572 +/-0.038	97.328 +/-0.549
Tiramisu	.904 +/-0.000	.861 +/-0.001	.946 +/-0.004	.991 +/-0.000	.446 +/-0.171	.692 +/-0.075	93.441 +/-0.472

Table 3: Values of all metrics on test set. Reported as mean value of runs, +/- standard deviation of runs.

Severity Score Severity score is a metric specific to this task, and it is useful for ecologists to determine how severe an EWD infection is. It is calculated as the ratio of diseased tissue to total tissue. The closer this value is to 1, the more advanced the EWD infection is on the eelgrass blade, and vice versa. It was calculated as follows:

$$\frac{\text{sum}(\tilde{Y} = 2)}{\text{sum}(\tilde{Y} = 1) + \text{sum}(\tilde{Y} = 2)} - \frac{\text{sum}(Y = 2)}{\text{sum}(Y = 1) + \text{sum}(Y = 2)}$$

Absolute Severity Score Because the Severity score is averaged over all samples in a particular set, it is possible for over estimations and under estimations of disease to cancel out. Therefore it is also useful to look at the absolute severity error, calculated as follows:

$$\left| \frac{\text{sum}(\tilde{Y} = 2)}{\text{sum}(\tilde{Y} = 1) + \text{sum}(\tilde{Y} = 2)} - \frac{\text{sum}(Y = 2)}{\text{sum}(Y = 1) + \text{sum}(Y = 2)} \right|$$

Prevalence Score This metric is also specific to this task, and is based on whether particular blade has any diseased tissue at all. It is thus a binary classification and is calculated as follows:

$$\text{sum}(\min(1, \text{sum}(\tilde{Y} = 2))) = \min(1, \text{sum}(Y = 2))$$

Experimental Setup

We train each network for a total of 40 epochs, where one epoch consists of training on the entire training set once, and accuracy metrics are reported on the validation set. At the end of 40 epochs, the version of the model that gives the best results on the validation set is then used to predict once on the test set. For each network the same experiment is run 5 times.

Results

The results of the models on the validation set for four of our seven metrics are shown in Figure 6: Dice Score, Jaccard Score, Severity Error and Prevalence Accuracy. The opaque line for each represents the mean value of the five runs, whereas the shaded region is the standard deviation of the runs. For the validation set it can be seen that UNet and DeepLab outperformed Tiramisu on the Dice and Jaccard metrics, and the eelgrass specific metric of prevalence accuracy. However for severity error, there was no significant difference between the three.

In Table 3 we have listed the results on the test set for these models for each of the seven metrics we calculated. The number listed represents the mean value of the 5 runs, with the $\pm$ indicating the standard deviation of the 5 runs, to 3 significant digits, with the best performance highlighted.

As can be seen in the table, DeepLab demonstrated the best results in 4/7 of the metrics. However, UNet demonstrated the best results in the two most important metrics for this domain Severity and Absolute Severity.

EeLISA as a Deployed AI System

Application Use

Given that the UNet model gave the best performance on the most important metrics for this domain, Severity and Absolute Severity, we use the trained UNet model in our active deployment of the EeLISA system. The EeLISA system has been live and been in continuous use by ecology researchers for over 16 months. During this period EeLISA has been fully available as a web service, and hosts thousands of eelgrass scans. The segmentation and analysis modules are mostly used in bursts when new eelgrass scans are made and need to be processed. The web application is used more frequently to visualize and inspect uploaded data. Additionally, the web application is frequently used as a more

intuitive way to visualize and organize all of the data that has been collected.

Payoff

Enabling Scientific Discovery and Insight Thanks to the rapid and consistent analysis of EeLISA, ecologists were able to quantify disease and demography before, during, and after a marine heat wave, and have determined that EWD is indeed sensitive to warming temperatures. These findings are pivotal in understanding mechanisms underlying eelgrass declines, and prove that EeLISA has practical and important ecological applications. The use of EeLISA has led to four manuscripts, three currently in preparation, one that has been submitted. All four manuscripts are based on large scale studies, either temporally or geographically or both, that are only possible given the massive speedup provided by EeLISA.

Enabling New Research Since the deployment of the EeLISA system, ecologists have been able to collect and analyze data on a massive scale. The EeLISA system was used to analyze the data of multi-year, multi-site study of EWD along the San Juan Islands in northwestern Washington State. Further, EeLISA is the sole tool being used to analyze eelgrass scans in a new multi-university NSF study. The study is being done over the course of three years, and each year collects over 3500 eelgrass samples from 32 sites ranging from Southern California to Alaska. An EWD study of this size is completely unprecedented and is only made possible by the rapid analysis facilitated by the EeLISA system.

Massive Speed Up It is a conservative estimate that earlier methods to measure disease metrics would take 30 minutes per each scan. Just the first year of data for the NSF project contains 694 scans. **It would require approximately 347 hours of manual analysis to label this entire set.** Further, this task cannot be outsourced to non-domain experts or community scientists, because of its complexity; it had to be done by ecological experts. **The same data was analyzed, measured, and cataloged in an output CSV file by EeLISA in roughly 5 minutes. This means EeLISA can perform the analysis more than 5000 times faster than traditional methods.** Additionally, with more GPUs available this analysis speed could improve even further.

Consistency Additionally, while the training process for the EeLISA system is stochastic, once the model is trained all internal model parameters are frozen and the system becomes deterministic. This means for the same input, EeLISA will always produce precisely the same output. This is a significant benefit over human labeling as two different experts might label the scan slightly differently due to personal judgement or labeling error, and even the same expert may label differently depending on how long they have been labeling. This makes it more difficult to compare disease metrics that are coming from multiple different experts, and from a large range of time. However, with EeLISA each measurement can be trusted to be perfectly consistent.

Deployment Architecture and Maintenance

The web server is based off open source code¹. It is written using the Meteor and Node.js frameworks, and is written in JavaScript. It is hosted via an Amazon Web Services (AWS) T2 Large instance with two cores, 8 GB of memory and 200 GB of expandable storage. The segmentation and analysis module are written in Python 3.7, and primarily use the modules PyTorch, Scipy, and Numpy to provide the ML and analysis functionality. The segmentation and analysis modules are run on a private Ubuntu based machine with two Intel Xeon Skylake processors with a total of 36 cores with a total of 376 GB of memory, as well as 5 NVIDIA Tesla V100 GPU with a combined GPU memory pool of 80 GB.

The segmentation module requires maintenance in the form of re-training, which happens every time there is a batch of new data. During this process the new data is added in the same 75/25 train/test split to the main training and testing data, and the model is re-trained. The newly trained weights are check-pointed and saved for future use. The web application and analysis module require very sporadic maintenance, and are only updated when a new feature is requested.

Lessons Learned

We quickly learned to leverage on available labeling tools¹, as originally we developed our own, but it lacked certain key functionality. Additionally, we learned how critical it was to integrate our solution into the established workflow of the ecologists, in order for our solution to really be practical. We also learned the importance of considering the success of our trained models through multiple objective/metrics. Lastly, we learned how powerful AI/ML can be in these domains and how it can empower researchers to do studies at unmatched scale.

Related Works

EWD Monitoring

In [Lefcheck et al. 2017] the authors used 31 years of manually annotated aerial imagery to track the area of eelgrass beds in the Chesapeake Bay. In [Groner et al. 2014, 2016] the authors collected eelgrass blade samples in a field study, whereby samples were measured and visually inspected for lesions indicating EWD. If a lesion was visually identified it was measured to quantify disease levels. In [Dawkins et al. 2018] the authors would quantify the disease using images and image processing software.

Annotation Systems

Many segmentation annotation frameworks exist [Russell et al. 2008; Lin et al. 2014]. Additionally, there has been much work in using trained models to perform semi-automatic annotation to speed up labeling [Castrejon et al. 2017; Yu et al. 2020].

Computer Vision in Ecology

Most efforts utilizing ML to study eelgrass is based on detection from aerial or satellite imagery [Islam et al. 2019;

Traganos and Reinartz 2018]. Additionally, [Moniruzzaman et al. 2019] released a public detection data set for seagrass in an aquarium setting. Outside of eelgrass there are several public segmentation and classification data sets for plants [Minervini et al. 2016; Scharr et al. 2014; Horn et al. 2017].

Semi-Supervised Learning

The feedback process of this method (i.e. the network producing outputs that are then used as training data) is a similar setting to semi-supervised learning, which has been well studied [Grandvalet and Bengio 2004; Lee 2013; Chapelle, Scholkopf, and Zien, Eds. 2009]. While our feedback process is most similar to the setting in [Lee 2013], it differs in that we consider the ideal situation where an expert annotator corrects the miss-classified pixels, and we leverage on the verified correct predictions. In this way, we use semi-supervised learning to limit labeling time, instead of to leverage on unlabeled data.

Conclusion

Global warming is a massive problem facing humanity, and it is affecting nearly every aspect of our lives. While it can feel overwhelming, ML is an important and powerful tool to combat it. While stopping the causes of global warming is extremely important, it is also extremely important to be able to carefully measure the effects it has already had on our ecosystems so that we can introduce preventative measures. Eelgrass is a critical marine habitat that is being affected by EWD, and is likely accelerated by global warming. In this work we have introduced the EeLISA system that allows ecologists to rapidly calculate disease metrics for eelgrass scans. We are pleased that EeLISA has been in use by ecological researchers for over 16 months, and has led to scientific discovery and insight highlighted in four separate manuscripts, and to the ability to do EWD research at unprecedented scales. We feel EeLISA presents a great example of how AI can be used symbiotically with human researchers to allow humanity to study and understand the natural world like never before.

Acknowledgements

This material is supported by NSF awards OCE-1829921 and CNS-0832782 (Expeditions).

References

Bockelmann, A.-C.; Tams, V.; Ploog, J.; Schubert, P.; and Reusch, T. 2013. Quantitative PCR Reveals Strong Spatial and Temporal Variation of the Wasting Disease Pathogen, *Labyrinthula zosterae* in Northern European Eelgrass (*Zostera marina*) Beds. *PLoS one* 8: e62169. doi: 10.1371/journal.pone.0062169.

Bull, J. C.; Kenyon, E. J.; and Cook, K. J. 2012. Wasting disease regulates long-term population dynamics in a threatened seagrass. *Oecologia* 169(1): 135–142. doi: 10.1007/s00442-011-2187-6.

Castrejon, L.; Kundu, K.; Urtasun, R.; and Fidler, S. 2017. Annotating object instances with a polygon-rnn. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5230–5238.

Chapelle, O.; Scholkopf, B.; and Zien, Eds., A. 2009. Semi-Supervised Learning (Chapelle, O. et al., Eds.; 2006) [Book reviews]. *IEEE Transactions on Neural Networks* 20(3): 542–542. doi:10.1109/TNN.2009.2015974.

Chen, L.; Papandreou, G.; Schroff, F.; and Adam, H. 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. *CoRR* abs/1706.05587. URL <http://arxiv.org/abs/1706.05587>.

Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; and Schiele, B. 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Costanza, R.; De Groot, R.; Sutton, P.; Van der Ploeg, S.; Anderson, S. J.; Kubiszewski, I.; Farber, S.; and Turner, R. K. 2014. Changes in the global value of ecosystem services. *Global environmental change* 26: 152–158.

Dawkins, P.; Eisenlord, M.; Yoshioka, R.; Fiorenza, E.; Fruchter, S.; Giammona, F.; Winningham, M.; and Harvell, C. 2018. Environment, dosage, and pathogen isolate moderate virulence in eelgrass wasting disease. *Diseases of Aquatic Organisms* 130. doi:10.3354/dao03263.

Grandvalet, Y.; and Bengio, Y. 2004. Semi-supervised learning by entropy minimization. *Advances in neural information processing systems* 17: 529–536.

Groner, M.; Burge, C.; Couch, C.; Kim, C.; Siegmund, G.-F.; Singhal, S.; Smoot, S.; Jarrell, A.; Gaydos, J.; Harvell, C.; and Wyllie-Echeverria, S. 2014. Host demography influences the prevalence and severity of eelgrass wasting disease. *Diseases of Aquatic Organisms* 108: 165–175. doi: 10.3354/dao02709.

Groner, M. L.; Burge, C. A.; Kim, C. J.; Rees, E.; Van Alstyne, K. L.; Yang, S.; Wyllie-Echeverria, S.; and Harvell, C. D. 2016. Plant characteristics associated with widespread variation in eelgrass wasting disease. *Diseases of aquatic organisms* 118(2): 159–168.

Horn, G. V.; Aodha, O. M.; Song, Y.; Shepard, A.; Adam, H.; Perona, P.; and Belongie, S. J. 2017. The iNaturalist Challenge 2017 Dataset. *CoRR* abs/1707.06642. URL <http://arxiv.org/abs/1707.06642>.

Huang, G.; Liu, Z.; Van Der Maaten, L.; and Weinberger, K. Q. 2017. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269.

Islam, K. A.; Hill, V.; Schaeffer, B.; Zimmerman, R.; and Li, J. 2019. Semi-supervised Adversarial Domain Adaptation for Seagrass Detection in Multispectral Images. In *2019 IEEE International Conference on Data Mining (ICDM)*, 1120–1125. IEEE.

- Jégou, S.; Drozdal, M.; Vazquez, D.; Romero, A.; and Bengio, Y. 2017. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 11–19.
- Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In Bengio, Y.; and LeCun, Y., eds., *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. URL <http://arxiv.org/abs/1412.6980>.
- Lee, D.-H. 2013. Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. *ICML 2013 Workshop : Challenges in Representation Learning (WREPL)*.
- Lefcheck, J. S.; Wilcox, D. J.; Murphy, R. R.; Marion, S. R.; and Orth, R. J. 2017. Multiple stressors threaten the imperiled coastal foundation species eelgrass (*Zostera marina*) in Chesapeake Bay, USA. *Global Change Biology* 23(9): 3474–3483. doi:10.1111/gcb.13623.
- LeNail, A. 2019. NN-SVG: Publication-Ready Neural Network Architecture Schematics. *Journal of Open Source Software* 4(33): 747. doi:10.21105/joss.00747. URL <https://doi.org/10.21105/joss.00747>.
- Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft COCO: Common Objects in Context. In Fleet, D.; Pajdla, T.; Schiele, B.; and Tuytelaars, T., eds., *Computer Vision – ECCV 2014*, 740–755. Cham: Springer International Publishing. ISBN 978-3-319-10602-1.
- Lin, T.-Y.; RoyChowdhury, A.; and Maji, S. 2015. Bilinear cnn models for fine-grained visual recognition. In *Proceedings of the IEEE international conference on computer vision*, 1449–1457.
- Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440.
- Macreadie, P.; Baird, M.; Trevathan-Tackett, S.; Larkum, A.; and Ralph, P. 2014. Quantifying and modelling the carbon sequestration capacity of seagrass meadows – A critical assessment. *Marine Pollution Bulletin* 83(2): 430 – 439. ISSN 0025-326X. Seagrass meadows in a globally changing environment.
- McLeod, E.; Chmura, G. L.; Bouillon, S.; Salm, R.; Björk, M.; Duarte, C. M.; Lovelock, C. E.; Schlesinger, W. H.; and Silliman, B. R. 2011. A blueprint for blue carbon: toward an improved understanding of the role of vegetated coastal habitats in sequestering CO₂. *Frontiers in Ecology and the Environment* 9(10): 552–560.
- Minervini, M.; Fischbach, A.; Scharr, H.; and Tsafaris, S. A. 2016. Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognition Letters* 81: 80 – 89. ISSN 0167-8655. doi:<https://doi.org/10.1016/j.patrec.2015.10.013>.
- Moniruzzaman, M.; Islam, S. M. S.; Lavery, P.; and Ben-namoun, M. 2019. Faster R-CNN Based Deep Learning for Seagrass Detection from Underwater Digital Images. In *2019 Digital Image Computing: Techniques and Applications (DICTA)*, 1–7.
- Moore, K.; Short, F.; Larkum, A.; Orth, R.; and Duarte, C. 2006. *Zostera: Biology, Ecology, and Management*, 361–386. doi:10.1007/1-4020-2983-7_16.
- Renn, C. E. 1936. The Wasting Disease of *Zostera Marina*: I. A Phytological Investigation of the Diseased Plant. *The Biological Bulletin* 70(1): 148–158. doi:10.2307/1537320.
- Ronneberger, O.; P.Fischer; and Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of LNCS, 234–241. Springer. URL <http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a>. (available on arXiv:1505.04597 [cs.CV]).
- Russell, B. C.; Torralba, A.; Murphy, K. P.; and Freeman, W. T. 2008. LabelMe: a database and web-based tool for image annotation. *International journal of computer vision* 77(1-3): 157–173.
- Scharr, H.; Minervini, M.; Fischbach, A.; and Tsafaris, S. 2014. Annotated Image Datasets of Rosette Plants. "*European Conference on Computer Vision*".
- Short, F.; Ibelings, B.; and Den Hartog, C. 1988. Comparison of a current eelgrass disease to the wasting disease in the 1930s. *Aquatic Botany* 30(4): 295 – 304. ISSN 0304-3770. doi:[https://doi.org/10.1016/0304-3770\(88\)90062-9](https://doi.org/10.1016/0304-3770(88)90062-9).
- Smale, D.; Wernberg, T.; Oliver, E.; Thomsen, M.; Harvey, B.; Straub, S.; Burrows, M.; Alexander, L.; Benthuyssen, J.; Donat, M.; Feng, M.; Hobday, A.; Holbrook, N.; Perkins-Kirkpatrick, S.; Scannell, H.; Gupta, A.; Payne, B.; and Moore, P. 2019. Marine heatwaves threaten global biodiversity and the provision of ecosystem services. *Nature Climate Change* 9. doi:10.1038/s41558-019-0412-1.
- Traganos, D.; and Reinartz, P. 2018. Mapping Mediterranean seagrasses with Sentinel-2 imagery. *Marine Pollution Bulletin* 134: 197 – 209. ISSN 0025-326X. doi:<https://doi.org/10.1016/j.marpolbul.2017.06.075>. Securing a future for seagrass.
- Waycott, M.; Duarte, C. M.; Carruthers, T. J. B.; Orth, R. J.; Dennison, W. C.; Olyarnik, S.; Calladine, A.; Fourqurean, J. W.; Heck, K. L.; Hughes, A. R.; Kendrick, G. A.; Kenworthy, W. J.; Short, F. T.; and Williams, S. L. 2009. Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proceedings of the National Academy of Sciences* 106(30): 12377–12381. ISSN 0027-8424. doi:10.1073/pnas.0905620106.
- Yu, F.; Chen, H.; Wang, X.; Xian, W.; Chen, Y.; Liu, F.; Madhavan, V.; and Darrell, T. 2020. BDD100K: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2636–2645.