Diving in Deep: Deep learning for the identification of yellow tang (*Zebrasoma flavescens*) grazing behaviors in Hawai'i

Trevor Nishida

Marine Science Department, University of Hawai'i at Hilo

200 W. Kawili St. Hilo, HI 96720

Advisors

Dr. John Burns and Dr. Drew Gray

May 11, 2023

Abstract

Advancements in underwater camera technologies have allowed various sectors to expand the breadth and depth of their studies into underwater environments. The use of these cameras, however, results in copious amounts of video data being collected. For researchers to extract information of interest from these data, they must be processed manually, costing valuable resources spent in time and labor. The use of artificial intelligence to automate video data processing provides a solution to this portion of the problem which can be cost effective and save substantial amounts of time. My project explores the possibility of using machine learning algorithms to track grazing behavior of a common Hawaiian reef fish, the yellow tang (*Zebrasoma flavescens*). Footage of yellow tang feeding behavior was collected from an underwater camera located in Kona, Hawai'i, U.S.A. This footage was split into still images which were annotated by a human. A machine learning model was trained on an 87:9:5 ratio split of training:validation:test datasets from a total of 3600 annotated still frames. The model was able to successfully detect feeding behaviors on videos it had not seen during training (Recall: 96.92%, Precision: 80.81%). These results demonstrate that it is possible to apply machine learning to detect behaviors in aquatic organisms. Further analysis of the behavioral data extracted from videos can be used as a tool to quantify health in underwater environments, providing policymakers with the information required to make informed decisions for the preservation of these environments.

Introduction

In recent years, the establishment of underwater video cameras in scientific, industrial, and military fields has allowed for more in-depth exploration of underwater environments

(Tamou et al. 2018). Underwater cameras can also be used in conjunction with other established ecological methods such as visual surveys, net surveys, and tagging of individuals to augment the study with video data (Ditria et al. 2020). However, video cameras often collect data which are not pertinent to the study for which they were designed. The overabundance of data can cause problems for researchers as manual processing can be cumbersome and time consuming. Artificial intelligence presents a unique solution to this problem through automatic processing using artificial intelligence vision.

Artificial intelligence is a subfield in machine learning which employs complex mathematical models that allow a machine to make predictions about the outcome of a process and make a decision based on this prediction (Ghahramani 2015, Mahesh 2018). This complex computation process results in a program which is not so different from the process of a human brain. Artificial intelligence vision, or AI vision, is an application of artificial intelligence to analyze image data. AI vision uses the pixel structure and color of an image to make it possible for a computer to recognize and label objects, facial features, and even estimate body poses in an image (Voulodimos et al 2018, Ditria et al 2020). In combination with recent advances in computational processing power using graphical processing units (GPUs), AI vision has been pushed to the forefront of machine learning applications.

Machine learning is a field which combines computer science and mathematics by leveraging the large amounts of data available online to create computational algorithms which are able to independently function without direct human supervision (Jordan & Mitchell 2015, Mahesh 2018). The purpose of these algorithms is to learn from the copious amounts of data and self-improve without needing to be explicitly programmed (Mahesh 2018). These algorithms can be applied in many different fields for many different uses such as self-driving cars, financial investment risk analysis, and image object detection. The specific application of machine learning for these types of applications is called deep learning.

Deep learning employs artificial neural networks to analyze and interpret various kinds of data. These neural networks are made up of a number of different layers containing interconnected nodes which function similar to the neurons found in human brains. For image analysis, the most common type of neural network used are convolutional neural networks (CNNs) (Villon et al. 2018, Valueva et al. 2020). This type of network is specialized to process data which comes into the network in a grid-like topology (Goodfellow et al. 2016). The network accomplishes this work through the use of specialized layers which perform a filtering operation on the input data, allowing the network to recognize a specific feature in the data (Goodfellow et al. 2016, Valueva et al. 2020). The filtering operation is an application of a mathematical operation which computes a weighted average that gives higher weights to a certain type of input (Villon et al. 2018). In CNNs used for image analysis, this weight is given to pixels closer together to one another on the grid (Goodfellow et al. 2016). The output from these layers then allows the network to create a set of feature maps which are used to capture the location of different features in an image. The use of these convolutional layers makes CNNs perform at a high level when identifying features from image and video data.

In the marine environment, deep learning algorithms can be used to identify species and their behaviors. The yellow tang (*Zebrasoma flavescens*) is a common member of the Family Acanthuridae, found in reef environments in the Indo-Pacific region, but they are most abundant in the coastal waters surrounding the Hawaiian archipelago (Jones 1968, Meyer & Holland 2004, Eble et al. 2009). In Hawaiʻi, this fish is known as *lauʻipala*, or yellow ti-leaf; deriving from its bright yellow coloration and ovular body shape. The scientific name also makes reference to their yellow coloration with *"flavescens"* deriving from the latin word "*flavus",* meaning yellow. Yellow tangs also have white scalpel-like spines on the caudal peduncle of the fish, a trait which is characteristic of the family Acanthuridae. These caudal spines are what the common name of "surgeonfish" for this fish family comes from. Juvenile yellow tang can be found taking shelter among branching corals in mid-depths of up to 30 m, while adults typically relocate to shallower reef flats up to 11 m deep (Ortiz & Tissot 2008).

Yellow tangs reproduce multiple times per year, with peak reproduction occurring during the spring and summer months (Bushnell et al. 2008). During peak season, females can produce up to 24,000 eggs per spawning session, resulting in pelagic larvae that are carried by currents for up to 62 days before settlement on a reef (Eble et al. 2009, Bushnell et al. 2010, Schemmel 2020). These larvae can be carried up to 184 km away from the original spawning site (Christie et al. 2010). The passive nature of larval dispersion and low survival rate of about 1% for juveniles (Claisse et al. 2009) results in a recruitment range between two and seventeen individuals per 100 m^2 (Williams et al. 2009). Juvenile survival rates increase in areas with suitable habitat and reduced competition for resources, suggesting that yellow tang's inherent high fecundity rate in combination with decreased population threats can result in rapid population growth (Claisse et al. 2009).

Yellow tang are one of the most commonly collected aquarium species in Hawai'i (Walsh et al. 2004, Williams et al. 2009). Between 1976 and 2003, over three million yellow tangs were caught for the aquarium trade in the state of Hawai'i. This number represents 37.2% of the total number of organisms caught for the aquarium trade during this time period (Walsh et al. 2004). The coastline along West Hawai'i Island, in particular, is an area of importance to the aquarium fish industry in the Hawaiian Islands. In 2003, 75% of the fish caught for the aquarium industry in the State of Hawai‘i were caught in the West Hawai‘i fishery (Walsh et al. 2004), this large amount of fish represents US$722,255 in commercial value for the state which is equivalent to 75% of the state’s total commercial marine landing value for aquarium fish (Walsh et al. 2004). Previous studies done in the West Hawai‘i fishery have also shown that this coastline is home to a larger than average population of yellow tangs (Bushnell 2007).

Yellow tangs belong to the Acanthuridae or surgeonfish family, one of the largest and most abundant groups of reef fish found in the nearshore Hawaiian island waters (Jones 1968), and are ecologically important because surgeonfish are herbivorous. The large amount of biomass that surgeonfish represent on the reef gives these fish a key role in the regulation of macroalgal growth and abundance (Wylie & Paul 1988, Meyer & Holland 2004, Marshall & Mumby 2015). The consumption of macroalgae by surgeonfish, including the yellow tang, helps maintain the health of the coral reefs they inhabit and prevent a shift from coral dominated reefs to algal dominated ones (Puk et al. 2016, Gove et al. 2022).

By combining footage from an underwater camera livestream established by the Multiscale Environmental Graphical Analysis Lab (MEGA Lab) and Fathom Ocean with a deep learning algorithm made for image analysis, this study was designed to successfully identify feeding behavior of a common and economically and ecologically valuable surgeonfish, *Zebrasoma flavescens.*

Methods

Archived livestream underwater footage from the multiscale environmental graphical analysis (MEGA) Lab Cam YouTube channel (www.youtube.com/c/MEGALabCam) was used to collect video footage on a Hawaiian reef located at 7 m depth at Keāhole Point (19.728°N, -

156.062°W) on the west coast of Hawai'i Island, Hawai'i, U.S.A. Livestream footage dates ranged from February 22, 2021 to April 26, 2021, and lasted from 5 min to 12 hrs in duration. The footage was reviewed by humans and timestamps in which yellow tang (*Zebrasoma flavescens)* feeding behavior was observed were collected. Footage was then downloaded from the MEGA Lab Cam Youtube channel and a Python script was employed to extract 2 seconds before and after the collected timestamps, saving the resulting 5 second clip.

A total of 720 individual 5 second clips were randomly selected and uploaded to the Roboflow online toolchain machine learning model-building platform (www.roboflow.com/). The platform extracted individual frames from the 5 second clips at a rate of 1 frame per second. Bounding box masks were manually drawn around the region of interest, in this case, yellow tangs which were feeding. If no feeding behavior could be observed in a given image, the image was marked as "NULL"; however the image was still added to the dataset. A total of 2167 instances of yellow tang feeding behaviors were annotated and 1433 images were marked as "NULL." Random images equating to 84% of the overall dataset were selected to be used for model training, another 10% was selected to create a validation dataset, and the remaining 6% was used to create a testing dataset, following one of the standardized methodologies used for supervised model training in machine learning (Alexandropoulos et al. 2019).

Model performance at identifying yellow tang feeding behaviors were examined using standardized machine learning model metrics. Detection performance was determined using the mean average precision 50% value (mAP50) (Ditria et al. 2020). The mAP50 value is the ability of the machine learning model to fit a prediction box which covers at least 50% of the region of interest (Voulodimos et al. 2018). This overlap is evaluated as the intersection over union value

(IoU), this is a percentage of the region of interest (ROI) overlapped by the machine learning prediction box which is calculated using the equation (Avazov et al. 2023):

$$IoU = \frac{ROI \cap Prediction}{ROI \cup Prediction}$$

The metrics of precision, recall, and F1 score were also used to evaluate the model's performance. Precision is the measure of the accuracy of the positive predictions made by the model, and recall is the fraction of the total positive labels the model correctly identified with a positive prediction. F1 score is the harmonic mean of precision and recall values, and measures how well the model can identify the feeding behavior (Ditria et al. 2021). Precision, recall and F1 score were calculated using the following formulas (Ditria et al. 2021, Avazov et al. 2023):

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

The model was also evaluated on its performance using precision, recall, F1 score, and mAP50 over the amount of epochs it was trained for. An epoch can be defined as one passing of the entire training dataset through the machine learning model. Evaluation over this parameter determines if the model is improving over many training iterations or if the model has reached the performance maxima and is beginning to overfit the data. Overfitting occurs when the model memorizes the training data instances and will cause the model to perform well on training data but poorly on the testing data.

The machine learning framework used in this study was an application of You Only Look Once, or YOLO, an object detection algorithm which is able to analyze images quicker and more efficiently than other existing frameworks (Redmon et al. 2015, Jocher et al. 2020). Once the annotation of the dataset was completed, images and annotations were exported to a Google Colaboratory notebook (colab.research.google.com) containing the backbone of YOLOv5 (Jocher et al. 2020). Model training and subsequent testing and analysis tasks were performed on the Google Colaboratory notebook using the included Nvidia Tesla K80 graphical processing unit.

Results

The machine learning model was highly effective at detecting grazing behavior on novel videos. The majority of frames that the model was tested on were correctly predicted with a total of 279 true positive predictions, 37 false positive predictions, and 41 false negative predictions (Figure 1). Precision, recall, and F1 score values were all greater than 80% on the testing data (Figure 2). Evaluating the model over all training epochs shows that the mAP50 reached 75% performance value after 31 training epochs, with an eventual maximum value of 93.11% at 190 epochs (Figure 3). Precision values reached a 75% performance value after 52 epochs, with a maximum value of 90.45% after 173 epochs; comparatively, recall performance reached 75% performance much quicker after only 10 epochs, and had a maximum value of 90.33% after 140 epochs (Figure 4).

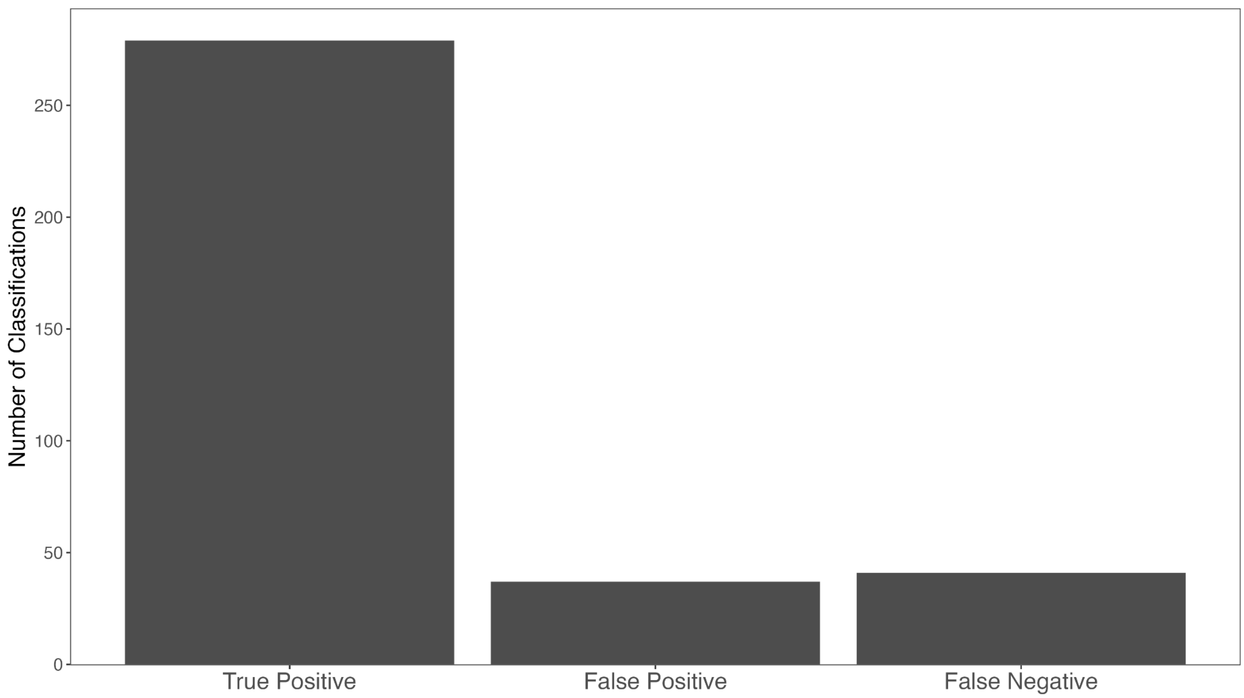

Figure 1: Classification performance of the machine learning model on images it had not seen during the training process.

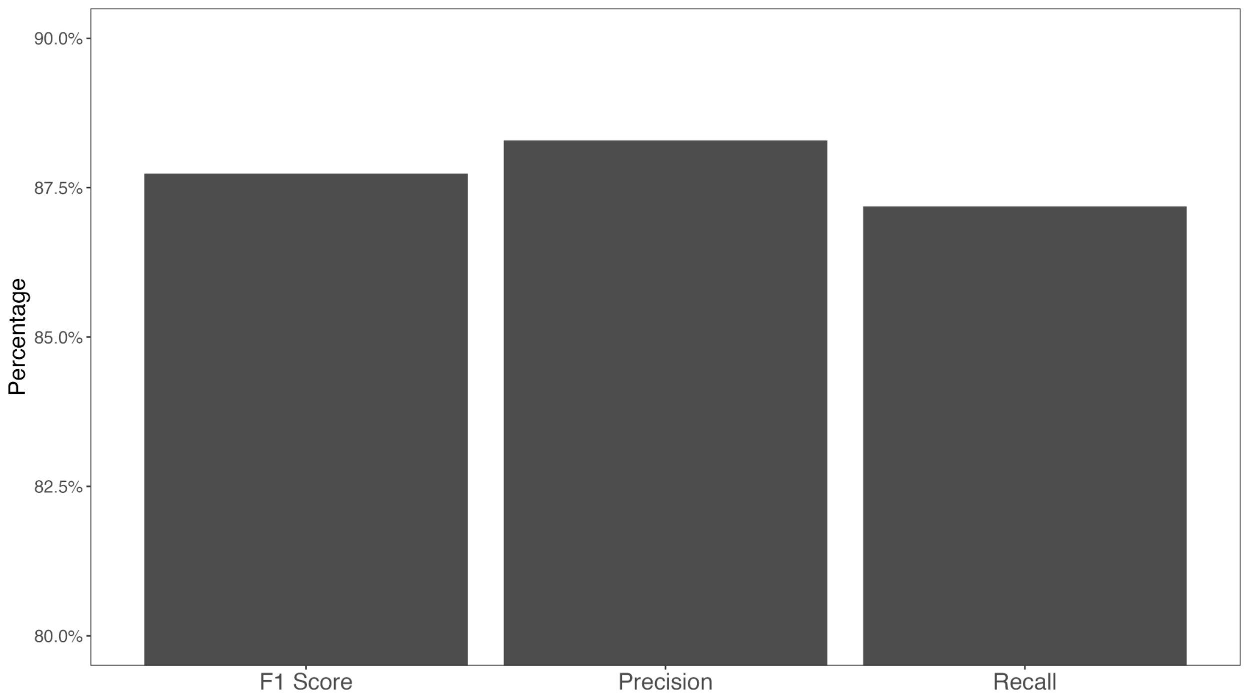

Figure 2: Machine learning model performance on images it had not seen during the training process.

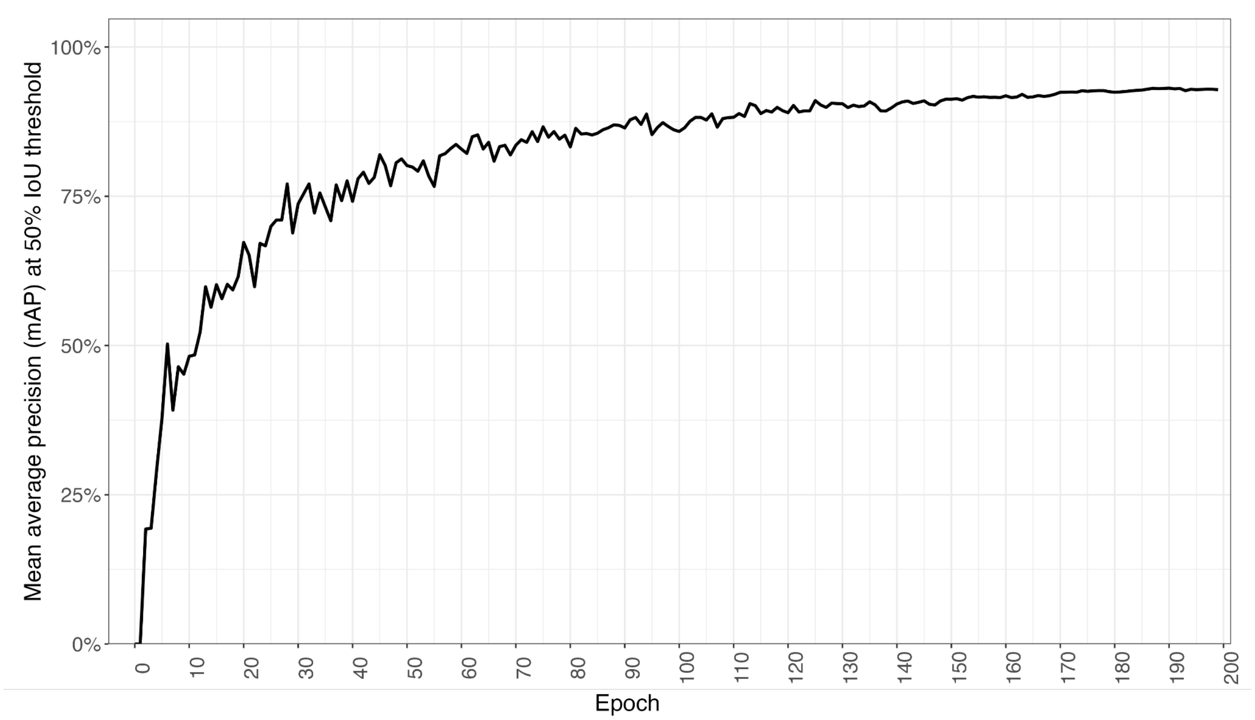

Figure 3: Machine learning model performance as measured by the mAP50 score across training epochs.

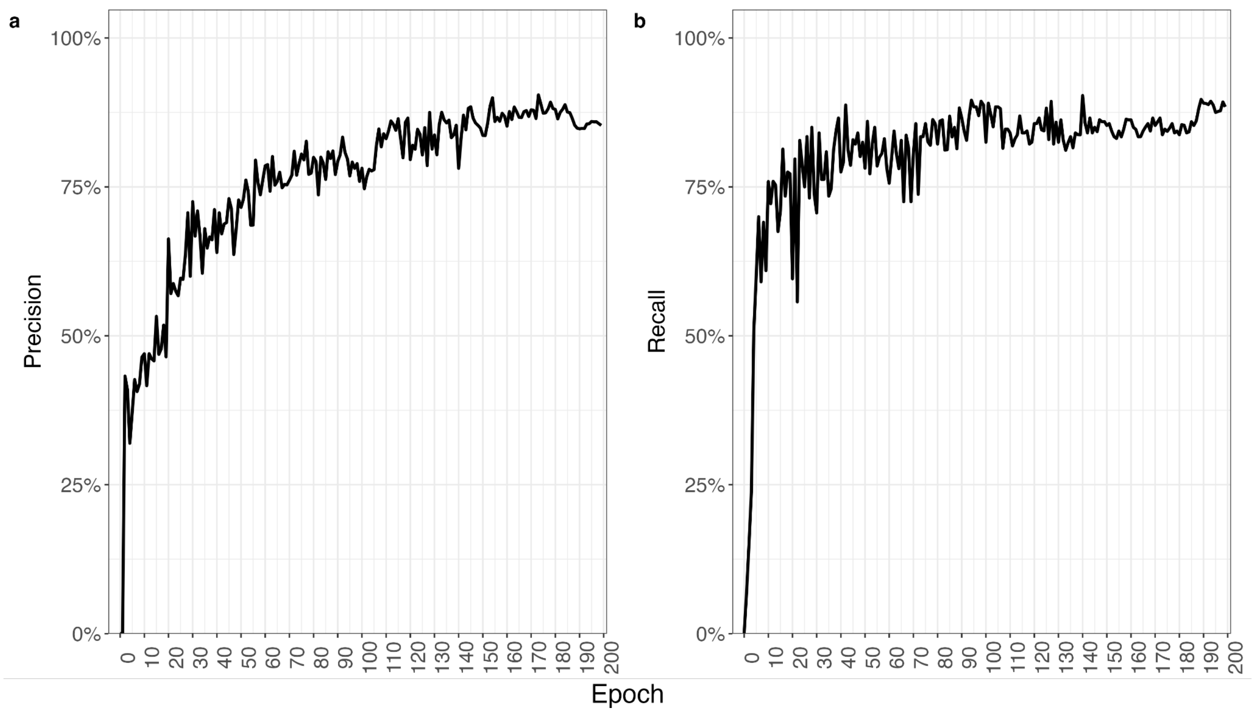

Figure 4: Machine learning model performance as measured by the precision (a) and recall (b) values across training epochs.

Discussion

Results from this study are evidence that machine learning algorithms can be implemented to successfully detect and identify grazing behaviors in yellow tang individuals. Combining machine learning tools with other technological advancements, such as the underwater livestream camera, can help to improve scientific research and the conservation of coral reefs. These tools can help enhance our understanding of coral reef systems without the need to deploy divers to complete traditional monitoring methods, such as snorkel or SCUBA surveys. This key difference allows scientists to collect large amounts of data in a completely remote setting which can give researchers a more realistic picture of what occurs *in situ* without behavioral changes due to human presence. Data can then be analyzed by a machine learning algorithm to quickly and efficiently identify areas which require immediate attention and develop corrective actions to mitigate damage to the ecosystem.

While the machine learning algorithm was successful in the detection and identification of grazing behaviors of yellow tang individuals, it was only able to correctly classify 78% of the grazing behaviors that it had not seen during the training process. Behavioral detection in other habitats with other animals are often able to collect data through instruments directly tagged onto the animal, such as accelerometers (Valletta et al. 2017, Kleanthous et al. 2022). The additional data provides more insight into various aspects about the animals' behaviors that are otherwise unobtainable through image data alone (Kleanthous et al. 2022). Across all machine learning algorithms however, performance drops between the training dataset and novel instances are not unusual. Per the "No Free Lunch" theorem: any elevated performance for one class of problems is offset by a loss of performance for another task (Wolpert & Macready 1997). In this case, for machine learning models to improve accuracy for image data, performance would need to be given up in other areas; or improvements would need to be made to the algorithm. Various improvements have been proposed in previous publications, however they were not implemented in this study due to computational limitations(Valletta et al. 2017, Ditria et al. 2021).

One such method proposed by Ditria et al. (2021) suggests the use of a motion estimation algorithm to generate optical flow data. These data are then used by the object detection algorithm to estimate the pixel motion between sequential frames. This addition would allow the machine learning algorithm to recognize movement between frames. Since the feeding behavior occurs across multiple frames, being able to contextualize a single frame using the frames before and after becomes a powerful tool for detection. The application of optical flow can then be used to filter out observations which otherwise would have been incorrectly classified. For example, if the model identifies a frame as containing grazing behaviors but the surrounding frames were classified as non-grazing, it is likely that the single frame classification was false positive. In previous studies, the addition of this process to the machine learning algorithm resulted in a 20% increase in the number of correctly identified grazing behavior instances (Ditria et al. 2021).

Studies have also shown that factors such as turbidity, light attenuation, and fishes' natural camouflage with the environment can have a profound effect on the efficacy of the detection algorithm (Ditria et al. 2020, Salman et al. 2020). Limited information from the camera's field of view and resolution will also impact how thoroughly the model is able to classify instances (Ditria et al. 2020). The effects these factors have on the algorithm can be reduced through a process called background modeling (Salman et al. 2020). Background modeling employs a Gausiann mixed model (GMM) to determine the distribution of pixel data in an image; this distribution can then be separated into background and foreground features and consequently learned by the model (Salman et al. 2020). Consequently any sudden shift in pixel distribution, from a fish entering the frame for example, can be recognized by the model even in low visibility environmental conditions. Alternatively, another study found that training the model using video collected from a variety of different environments and environmental conditions yielded a higher accuracy and consistency than models trained on a single type of environment (Ditria et al. 2020).

Automated video processing has the potential to revolutionize researchers' understanding of the marine environment and improve their ability to manage and preserve marine ecosystems.

As demonstrated by my project, artificial intelligence allows researchers to gather and analyze vast amounts of data about the underwater world without worrying about time-consuming manual processing. The video data can be inputted into a machine learning algorithm to rapidly detect and identify both marine organisms and their behaviors. Information derived from the output of the algorithm can then be used to inform conservation efforts, guide sustainable management decisions, and provide the information needed by policymakers. Additionally, as the use of artificial intelligence to augment current research methods becomes more mainstream, the usage of machine learning tools can aid in the development of new methods for underwater exploration and data collection. These new methods can lead to new discoveries and a deeper understanding of the oceans and marine organisms. While revolutionary, the use of artificial intelligence must take into consideration the ethical and social implications of widespread use in underwater environments; such as concerns around the privacy of people who use the ocean for recreation, the efficacy of artificial intelligence over long time scales, the ecological impacts of increased surveillance on marine life, and the possibility of misuse by humans to exploit desirable species. The use of automated video processing in underwater environments will change the way scientists research the oceans; however, researchers must ensure that these tools are used in a way which mitigates detrimental impacts on the environment.

References

Alexandropoulos SN, Aridas CK, Kotsiantis SB, Vrahatis MN (2019) Multi-objective evolutionary optimization algorithms for machine learning: a recent survey. Approximation and Optimization. 145:35-55

Avazov K, Hyun AE, Sami SAA, Khaitov A, Abdusalomov AB, Cho YI (2023) Forest fire detection and notification method based on AI and IoT approaches. Future Internet. 15(6):61

Bushnell M (2007) Reproduction of *Zebrasoma flavescens:* Oocyte maturation, spawning patterns, and an estimate of reproductive potential for female yellow tang in Hawaiʻi. Master's thesis, University of Hawaiʻi at Mānoa, Honolulu, HI.

Bushnell ME, Claisse JT, Laidley CW (2010) Lunar and seasonal patterns in fecundity of an indeterminate, multiple-spawning surgeonfish, the yellow tang *Zebrasoma flavescens*. Journal of Fish Biology 76: 1343–1361.

Christie MR, Tissot BN, Albins MA, Beets JP, Jia YL, Ortiz DM, Thompson SE, Hixon MA (2010) Larval connectivity in an effective network of marine protected areas. PLoS ONE. 5: e15715

Claisse JT, Mctee SA, Parrish JD (2009) Effects of age, size, and density on natural survival for an important coral reef fishery species, yellow tang, *Zebrasoma flavescens*. Coral Reefs 28: 95–105.

Ditria EM, Lopez-Marcano S, Sievers M, Jinks EL, Brown CJ, Connolly RM (2020) Automating the analysis of fish abundance using object detection: optimizing animal ecology with deep learning. Front. Mar. Sci. 7:249

Ditria EM, Sievers M, Lopez-Marcano S, Jinks EL, Connolly RM (2020) Deep learning for automated analysis of fish abundance: the benefits of training across multiple habitats. Environmental Monitoring and Assessment. 192:698

Ditria EM, Jinks E, Connolly RM (2021) Automating the analysis of fish grazing behaviour from videos using image classification and optical flow. Animal Behavior. 177:31-37

Eble JA, Toonen RJ, Bowen BW (2009) Endemism and dispersal: comparative phylogeography of three surgeonfishes across the Hawaiian Archipelago. Marine Biology. 156:689-698

Jordan MI, Mitchell TM (2015) Machine learning: Trends, perspectives, and prospects. Science. 349:255-260

Ghahramani Z (2015) Probabilistic machine learning and artificial intelligence. Nature. 521:452-459

Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press. Cambridge, MA.

Gove JM, Maynard JA, Lecky J, Tracey DP, Allen ME, Asner GP, Conklin C, Couch C, Hum K, Ingram RJ, Kindinger TL, Leong K, Oleson KLL, Towle EK, van Hooidonk R, Williams GJ, Hospital J. 2022. 2022 Ecosystem Status Report for Hawai‘i. Pacific Islands Fisheries Science Center, PIFSC Special Publication, SP-23-01, 91p.

Jocher, G. Chaurasia A, Stoken A, Borovec J, Nanocode012, Kwon Y, Michael K, TaoXie, Fang J, imyhxy, Lorna, Yifu Z, Wong C, Abhiram V, Montes D, Wang Z, Fati C, Nadar J, Laughing, UnglvKitDe, Sonck V, tkianai, yxNONG, Skalski P, Hogan A, Nair D, Strobel M, Jain M (2020) ultralytics/yolov5: v7.0 - YOLOv5 SOTA realtime instance segmentation. Zenodo. 7347926

Jones RS (1968) Ecological Relationships in Hawaiian and Johnston Island *Acanthuridae* (Surgeonfishes). Micronesica 4:309-361.

Marshall A, Mumby P (2015) The role of surgeonfish (*Acanthuridae*) in maintaining algal turf biomass on coral reefs. Journal of Experimental Marine Biology and Ecology. 473:152-160

Mahesh B (2018) Machine learning algorithms - A review. International Journal of Science and Research. 9(1):381-386

Meyer CG, Holland KN (2005) Movement patterns, home range size and habitat utilization of the bluespine unicornfish, *Naso unicornis* (*Acanthuridae)* in a Hawaiian marine reserve. Environmental Biology of Fishes. 73:201-210

Ongsulee P (2017) Artificial intelligence, machine learning, and deep learning. 2017 Fifteenth International Conference on ICT and Knowledge Engineering.

Ortiz DM, Tissot BN (2008) Ontogenetic patterns of habitat use by reef-fish in a Marine Protected Area network: a multi-scaled remote sensing and in situ approach. Marine Ecology Progress Series 365:217–232.

Puk LD, Ferse SCA, Wild C (2016) Patterns and trends in coral reef macroalgae browsing: a review of browsing herbivorous fishes of the Indo-Pacific. Rev. Fish Biol. Fisheries. 26:53-70

Redmon J, Divvala S, Girshick R, Farhadi A (2015) You Only Look Once: Unified, real-time object detection. IEEE Conference on Computer Vision and Pattern Recognition. 2016:779-788

Salman A, Siddiqui SA, Shafait F, Mian A, Shortis MR, Khurshid K, Ulges A, Schwanecke U (2020) Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system. ICES Journal of Marine Science. 77(4):1295-1307

Valueva MV, Nagornov NN, Lyakhov PA, Valuev GV, Chervyakov NI (2020) Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. Mathematics and Computers in Simulation. 177:232-243

Villon S, Mouillot D, Chaumont M, Darling ES, Subsol G, Claverie T, Villeger S (2018) A Deep learning method for accurate and fast identification of coral reef fishes in underwater images. Ecological Informatics. 48:238-244

Voulodimos A, Dolulamis N, Dolamis A, Protopapapdakis E (2018) Deep learning for computer vision: a brief review. Computational Intelligence and Neuroscience 2018:7068349

Walsh WJ, Cotton SSP, Dierking J, Williams ID (2004) The commercial marine aquarium fishery in Hawaiʻi 1976-2003. Status of Hawaiʻi's Coastal Fisheries in the New Millennium. In: Proceedings of a Symposium sponsored by the American Fisheries Society. 132-159

Williams ID, Walsh WJ, Claisse JT, Tissot BN, Stamoulis KA (2009) Impacts of a Hawaiian marine protected area network on the abundance and fishery sustainability of the yellow tang, *Zebrasoma flavescens*. Biological Conservation. 142:1066-1073