# Automatic detection of fish and tracking of movement for ecology

Sebastian Lopez-Marcano[1], Eric Jinks[1], Christina Buelow[1], Christopher J Brown[1], Dadong Wang[2], Branislav Kusy[2], Ellen Ditria[1], and Rod Connolly[1]

[1]Griffith University Faculty of Environmental Sciences
[2]CSIRO

March 8, 2021

**Abstract**

1. Animal movement studies are conducted to monitor ecosystem health, understand ecological dynamics and address management and conservation questions. In marine environments, traditional sampling and monitoring methods to measure animal movement are invasive, labour intensive, costly, and measuring movement of many individuals is challenging. Automated detection and tracking of small-scale movements of many animals through cameras are possible. However, automated techniques are largely untested in field conditions, and this is hampering applications to ecological questions. 2. Here, we aimed to test the ability of computer vision algorithms to track small-scale movement of many individuals in videos. We apply the method to track fish movement in the field and characterize movement behaviour. First, we automated the detection of a common fisheries species (yellowfin bream, Acanthopagrus australis) from underwater videos of individuals swimming along a known movement corridor. We then tracked fish movement with three types of tracking algorithms (MOSSE, Seq-NMS and SiamMask), and evaluated their accuracy at characterizing movement. 3. We successfully detected yellowfin bream in a multi-species assemblage (F1 score = 91%). At least 120 of the 169 individual bream present in videos were correctly identified and tracked. The accuracies among the three tracking architectures varied, with MOSSE and SiamMask achieving an accuracy of 78%, and Seq-NMS 84%. 4. By employing these emerging computer vision technologies, we demonstrated a non-invasive and reliable approach to studying fish behaviour by tracking their movement under field conditions. These cost-effective technologies potentially will allow future studies to scale-up analysis of movement across many underwater visual monitoring systems.

**Automatic detection of fish and tracking of movement for ecology**

Sebastian Lopez-Marcano[1,2*], Eric Jinks[1], Christina A. Buelow[1], Christopher J. Brown[1], Dadong Wang[2], Branislav Kusy[3], Ellen Ditria[1], Rod M. Connolly[1]

[1]Coastal and Marine Research Centre, Australian Rivers Institute, School of Environment and Science, Griffith University, Gold Coast, QLD 4222, Australia

[2]Quantitative Imaging Research Team, Data61, CSIRO, Marsfield, NSW 2122, Australia

[3] Data61, CSIRO, QLD 4069, Australia

*corresponding author: sebastian.lopez-marcano@griffithuni.edu.au

## Abstract

1. Animal movement studies are conducted to monitor ecosystem health, understand ecological dynamics and address management and conservation questions. In marine environments, traditional sampling and monitoring methods to measure animal movement are invasive, labour intensive, costly, and measuring movement of many individuals is challenging. Automated detection and tracking of small-scale

movements of many animals through cameras are possible. However, automated techniques are largely untested in field conditions, and this is hampering applications to ecological questions.
2. Here, we aimed to test the ability of computer vision algorithms to track small-scale movement of many individuals in videos. We apply the method to track fish movement in the field and characterize movement behaviour. First, we automated the detection of a common fisheries species (yellowfin bream, *Acanthopagrus australis)* from underwater videos of individuals swimming along a known movement corridor. We then tracked fish movement with three types of tracking algorithms (MOSSE, Seq-NMS and SiamMask), and evaluated their accuracy at characterizing movement.
3. We successfully detected yellowfin bream in a multi-species assemblage (F1 score = 91%). At least 120 of the 169 individual bream present in videos were correctly identified *and* tracked. The accuracies among the three tracking architectures varied, with MOSSE and SiamMask achieving an accuracy of 78%, and Seq-NMS 84%.
4. By employing these emerging computer vision technologies, we demonstrated a non-invasive and reliable approach to studying fish behaviour by tracking their movement under field conditions. These cost-effective technologies potentially will allow future studies to scale-up analysis of movement across many underwater visual monitoring systems.

*Keywords*

Artificial Intelligence, connectivity, deep learning, dispersal, machine learning, object tracking, underwater video

## 1. Introduction

Computer vision (CV), the research field that explores the use of computer algorithms to automate the interpretation of digital images or videos, is revolutionising data collection in science (Waldchen & Mader, 2018; Beyan & Browman, 2020). The use of remote camera imagery, such as underwater stations, camera traps and stereography, has driven the uptake of CV because it has the capacity to process and analyse imagery quickly and accurately (Bicknell et al., 2016; Schneider et al., 2019). In ecological studies, advances in CV have led to an increase in sampling accuracy and repeatability (Waldchen & Mader, 2018). For example, drones are being used to track grassland animals (van Gemert et al., 2015) and estimate tree defoliation (Kälin et al., 2019), underwater observatories with CV are monitoring deep sea ecosystems (Aguzzi et al., 2019), and CV-capable dive scooters are being used to monitor coral reefs at large spatial and temporal scales (González-Rivero et al., 2020; Kennedy et al., 2020)

In the past few years, we have seen an increase in the uptake of CV to study and monitor marine ecosystems. These applications are related to the two main CV tasks – object detection (OD) and object tracking (OT). OD and OT automate the task of gathering information about the type, location and movement of objects of interest. OD has received the most attention as OD models can count and identify species of interest in underwater video footage (Christin, Hervet & Lecomte, 2019). For example, OD models have been applied to detect seals (Salberg, 2015), identify whale hotspots (Guirado et al., 2019), monitor fish populations (Xiu et al., 2015; Salman et al., 2016; Villon et al., 2016; Marini et al., 2018; Villon et al., 2018; Ditria et al., 2020b; Jalal et al., 2020; Villon et al., 2020) and quantify floating debris on the ocean surface (Watanabe, Shao & Miura, 2019). By comparison, the application of OT is less advanced in marine ecosystems. Previous work has shown that OT models can successfully track on-surface objects (see topios.org) and underwater objects such as fish, sea turtles, dolphins, and whales (Spampinato et al., 2008; Chuang et al., 2017; Xu & Cheng, 2017; Arvind et al., 2019; Kezebou et al., 2019). There is also evidence that automated monitoring of fish in underwater ecosystems through the combination of OD and OT is reliable and accurate (Spampinato et al., 2008; Lantsova et al., 2016; Mohamed et al., 2020). However, no studies have jointly applied OD and OT for animal movement studies. OD can help advance the automatic collection of traditional presence/absence data of different species (Xiu et al., 2015; Marini et al., 2018) and OT can subsequently track multiple individuals and provide fine-scale tracking data to assess behavioural and animal movement patterns across a range of environments (Francisco, Nührenberg & Jordan, 2020). With a single and non-invasive automated

approach, two types of ecological information can be obtained, which will provide individual level information of different species and that enhances our ability to quantify the environmental drivers of species abundance and behaviour.

The combination of OD and OT is particularly suited to the subfield of marine animal movement because these tasks can provide the volume of data required to quantify movement of numerous individuals (Lopez-Marcano et al., 2020).In marine environments, animal movement shapes predator-prey dynamics, nutrient dynamics and trophic functions (Olds et al., 2018). For example, the movement of herbivorous fish between seagrass and coral reefs helps maintain resilience by balancing fish abundances with algal growth rates that vary spatio-temporally (Pagès et al., 2014). The knowledge of animal movement is fundamental to many research objectives in marine science, and collecting movement data is challenging and requires substantial resources. Therefore, the development and applications of emerging technologies (i.e. computer vision) can help advance our understanding of animal movement across a broad range of spatio-temporal dimensions and ecological hierarchies (e.g. individuals, populations, communities).

In this study, we aimed to test the ability of deep learning algorithms to track small-scale animal movement of many individuals in underwater videos. We developed a CV pipeline consisting of two steps, OD and OT, and we used the pipeline to quantify underwater animal movement across habitats for ecological research. We tested and applied off-the-shelf OT architectures to determine the efficacy and capacity of these emerging techniques to be used for underwater ecological applications. To demonstrate the applications of OD and OT, we deployed a 6-camera network in a known coastal fish estuarine corridor and recorded the movement of a common fisheries species (yellowfin bream,*Acanthopagrus australis)* . The corridor, located in the Tweed River estuary, Australia, is located between a rockwall passage and a seagrass meadow. Multiple estuarine fish such as sand whiting (*Sillago ciliata),* river garfish (*Hyporhamphus regularis* ), luderick (*Girella tricuspidata)* , spotted scat (*Scatophagus argus* ), three-bar porcupinefish (*Dicotylichthys punctulatu* s) and yellowfin bream, frequently move back and forth with the tides through this corridor, representing a relatively challenging scenario (i.e. low visibility and with currents also carrying floating debris) to showcase the capacity of CV to detect the target species in a multi-species assemblage and quantify the direction of movement. Testing the method with fish tidal movement represents the ideal test, because of the common knowledge on how and where fish move with the tidal patterns. We expected the analysis of videos from cameras to detect and track bream moving in the corridor consistent with the direction of the tidal flow. For OD, we used an off-the-shelf model called Mask Regional Convolutional Neural Network (Mask R-CNN) (He et al., 2017) that has been shown to successfully and accurately detect and quantify fish in estuarine ecosystems (Ditria et al., 2020b). We also benchmarked three OT architectures: Minimum Output Sum of Squared Errors (MOSSE) (Bolme et al., 2010), Sequential Non-Maximum Suppression (Seq-NMS) (Han et al., 2016), and Siamese Mask (SiamMask)(Wang et al., 2019). Ultimately, we demonstrate that these technologies can complement the collection and analysis of animal movement data and potentially contribute to the data-driven management of ecosystems.

## 2. Methods

### 2.1 Object detection

Object detection is a field of CV that deals with detecting instances of objects in images and videos (Zhao et al., 2019). Methods for object detection generally include traditional image processing and analysis algorithms, and deep learning techniques (Zhao et al., 2019). Deep learning is a subset of machine learning that uses networks capable of learning from data that is unstructured, either labelled (supervised) or unlabelled (i.e. unsupervised) (Lecun, Bengio & Hinton, 2015). Studies show that deep learning models are robust and efficient for fish detection in underwater scenarios (Cui et al., 2020; Ditria et al., 2020b; Jalal et al., 2020; Villon et al., 2020). In this paper, we use deep learning, and more specifically Mask Regional Convolutional Neural Network (Mask R-CNN) for fish detection. Mask R-CNN is one of the most effective open-access deep learning models for locating and classifying objects of interest (He et al., 2017).

To develop and train the fish detection model, we collected video footage of the target species, yellowfin bream in the Tweed River estuary, Australia (-28.169438, 153.547594) between May and September 2019. We used six submerged action cameras (1080p Haldex Sports Action Cam HD) deployed for 1 hr over a variety of marine habitats (i.e. rocky reefs and seagrass meadows). We varied the angle and placement of the cameras to ensure the capture of diverse backgrounds and fish angles (Ditria et al., 2020a). We trimmed the original 1 hour videos into snippets where yellowfin bream was present using VLC media player 3.0.8. The snippets were then into still frames at 5 frames per second. The training videos included 8,700 fish annotated across the video sequences (Supplementary A). We used software developed at Griffith University for data preparation and annotation tasks (FishID - https://globalwetlandsproject.org/tools/fishid/). We trained the model using a ResNet50 architecture with a learning rate of 0.0025 (He et al., 2017). We used a randomly selected 90% sample of the annotated dataset for the training, with the remaining 10% for validation. To minimise overfitting, we used the early-stopping technique (Prechelt, 2012), where we assessed mAP50 on the validation set at intervals of 2,500 iterations and determined where the performance began to drop. We used a confidence threshold of 80%, meaning that we selected OD outputs where the model was 80% or more confident that it was a yellowfin bream. We developed the models and analysed the videos using a Microsoft Azure Data Science Virtual Machine powered with either NVIDIA V100 GPUs or Tesla K80 GPUs.

### 2.2 Object tracking

Tracking objects in underwater videos is challenging due to the 3D medium that aquatic animals move through, which creates greater variation in the shape and texture of the objects and their surroundings in a video (Sidhu, 2016). Additionally, underwater images are often obscured by floating objects, so automated tracking of fish is not a trivial task. Advances in object tracking have started to address these issues, and have managed to track objects consistently even with natural variations of the object's shape, size and location (Bolme et al., 2010; Cheng et al., 2018). We developed a pipeline where the OT architecture activates once the OD model detected a fish of the target species. This approach resulted in an automated detection and subsequent tracking of fish from the underwater videos. Additionally, we benchmarked the performance of three OT architectures (MOSSE, Seq-NMS and SiamMask) by using movement data gathered a month after the training dataset was collected from a different location in the Tweed River estuary, Australia. In this location, a 150 m long rocky wall restricts access to a seagrass-dominated harbour (Figure 1). The placement of the rock wall creates a 20 m wide passageway that fish use as a movement corridor to access a seagrass meadow.

We collected the fish movement data by submerging two sets of three action cameras (1080p Haldex Sports Action Cam HD) for one hour during a morning flood tide in October 2019. We placed the sets of cameras parallel to each other and separated by 20 m (Figure 1). Within each set, the cameras faced horizontally towards the fish corridor and parallel with the seafloor, and were separated by ˜ 3 m. The camera placement allowed us to calculate horizontal movement (left or right) of fish through the corridor. The distance between the cameras and between the sets ensured non-overlapping field of views. Set 1 cameras faced north and set 2 faced south (Figure 1). We placed the cameras in a continuous line starting at the harbour entrance and ending at the border of the seagrass meadow, deployed at a depth of 2-3 m. We manually trimmed each video using VLC media player 3.0.8 into video snippets with continuous yellowfin bream movement. The trimming resulted in 76 videos of varying durations (between 3 – 70 seconds) with each video transformed into still frames at 25 frames per second. All frames with bream present were manually annotated and these annotations used as groundtruth. We used the fish movement dataset to evaluate the object detection model and the OT architectures.

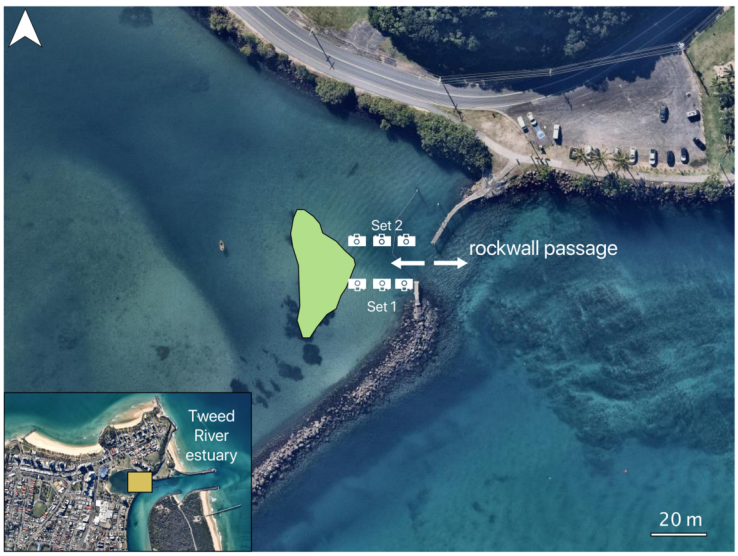



**FIGURE 1** The study location in Tweed River estuary, Australia, showing the camera array deployed in a fish corridor (two ended white arrow) between the rock wall channel and the seagrass meadow (green polygon). Each set of cameras consisted of three underwater cameras that recorded for 1 hr during a flood tide. Set 1 faced north and Set 2 faced south. The distance between cameras (˜3 m) and between sets (20 m) ensured non-overlapping field of views. Map data: NearMap 2020.

2.2.1 Minimum output sum of squared error (MOSSE)

The MOSSE algorithm produces adaptive correlation filters over target objects, and tracking is performed via convolutions (process of combining outputs to form more outputs). MOSSE was developed between 2010 and 2016 and it is robust to changes in lighting, scale, pose and shape of objects (Bolme et al., 2010; Sidhu, 2016). Here, we modified the MOSSE tracking process by activating the tracker with the OD output (Figure 2). The object detection model and the object tracking architecture interacted to maintain the consistency of the tracker on yellowfin bream individuals. When a fish was detected, the entry was used to initialise the tracker. MOSSE then tracked the fish for 4 frames and then a check was made on the subsequent frame to verify the accuracy of the tracker. In this check, if the detection bounding box overlapped by [?] 30% with the existing tracker bounding box, then the tracker continues on the same object. If the detection bounding box does not overlap with the existing tracker bounding box, then a new tracker entry starts. This interaction between the detection and tracking occurred for every yellowfin bream detected in a frame and stopped when no more detections were found.

2.2.2 Sequential non-maximum suppression (Seq-NMS)

Sequential non-maximum suppression (Seq-NMS) was developed in 2016 and is traditionally an algorithm developed to improve the classification results and consistency of deep learning outputs (Han et al., 2016). Seq-NMS works differently to the other trackers tested because it requires an OD output for every frame where there is a fish. Seq-NMS links detections of neighbouring frames, which means that a detection in the first frame can be connected with a detection in the second frame if there is an intersection above a defined threshold. In our case, we used the principles of Seq-NMS to create detection linkages for OT of fish when there was an overlap (intersection over union) of bounding box in subsequent frames of [?] 30%. In other words, if an object was detected in Frame 1 (Detection 1) and then another object detected in Frame 2 (Detection 2), then we checked if the bounding boxes of Detection 1 and Detection 2 overlapped

equal to or greater than 30% (Figure 2). If this is true, then the chain of detections continues. When the overlap is less than 30%, then a new detection link starts (i.e. the tracker will treat this detection as a new fish). We determined the movement direction by calculating the distance and angle (vector) between the two coordinates at the centres of bounding boxes of Detections 1 and 2.

#### 2.2.3 SiamMask

SiamMask is a tracking algorithm developed in October 2019 that uses outputs of deep learning models for estimating the rotation and location of objects (Wang et al., 2019). SiamMask is based on the concepts of Siamese network-based tracking. Similar to MOSSE, we slightly modified the tracking process by activating the tracker with the deep learning object detection model (Figure 2). The tracking with SiamMask started once a yellowfin bream was detected (Figure 4).

We have made available the OD annotations and images, movement dataset and annotations, and trackers and data wrangling codes (https://github.com/slopezmarcano/automated-fish-tracking).

### 2.3 Model evaluations and movement assessment

#### 2.3.1 Object detection evaluation

We evaluated the OD against the movement data (manually annotated and groundtruthed) described in section 2.2 and calculated precision, recall and F1. The precision is the rate of true positives relative to total detections and the recall is the rate of detection of true positives. We used the F1 score (the harmonic mean of the precision and recall) to assess the performance of our object detection model in answering ecological questions on abundance.

$$(1) \quad Precision = \frac{\text{TruePositives}}{TruePositives + FalsePositives}$$

$$(2) \quad Recall = \frac{\text{TruePositives}}{TruePositives + FalseNegatives}$$

$$(3) \quad F1 = 2 * \frac{precision * recall}{precision + recall}$$

Additionally, we determined the model's ability to fit a segmentation mask around the fish through the mean average precision value (mAP) (Everingham et al., 2010). We used the mAP50 value, which is the model's capacity to overlap a segmentation mask around 50% of the ground-truth outline of the fish. A high mAP50 value means that the model has high accuracy when overlapping a mask around the fish. We used the COCO evaluation python script to calculate mAP50 (Massa & Girshick, 2018).

#### 2.3.2 Object tracking evaluation

We evaluated the tracking architectures against the movement dataset by calculating precision, recall and a F1 score and by assessing the movement data (see Section 2.3.3). To calculate precision recall and a F1 score, we manually observed every second of video and determined if the OT architecture was correctly tracking the yellowfin bream individual (Supplementary B). We defined a true positive as a correct detection of yellowfin bream and then accurate tracking of the individual for [?] 50% of the time where yellowfin bream appeared on frame (Supplementary B). A false negative was when a bream was not detected and tracked or if the yellowfin bream was tracked $< 50\%$ of the time when the fish appeared on frame. Additionally, we classified a false positive when a non-yellowfin bream object was detected and tracked or when a yellowfin bream was detected but the tracking architecture failed by then tracking a non-yellowfin bream object.

#### 2.3.3 Movement assessment

The movement assessment was done to evaluate the accuracy of the directions provided by the tracker. We obtained one row of tracking data per frame when a fish was detected and subsequently tracked. For each

tracking output, the OT architecture provided a tracking angle of movement in 2 dimensions. We grouped tracking angles using reference angles into four directions: up, down, left and right (Supplementary B). Because the camera was facing horizontally towards the fish corridor, parallel with the seafloor, we were able to calculate horizontal movement of fish. Fish moving up meant that the fish movement had tracking angles between 44° and 315°. Fish moving right (east) had angles between 45° and 135°, whereas fish moving left (west) between 225° and 315°. Finally, fish moving down had tracking angles between 135° and 225°. The tracking angle for all OT architectures was obtained from the tracker vector that is generated within each tracker's bounding box (Supplementary B). By grouping the directions, we were able to count and group the number of movement angles per camera and per set. For each camera set, we then calculated the proportion of each tracking direction and determined net movement. We defined net movement as the movement angle with the highest proportion for a video. The data summary was generated in R (R Core Team, 2019) with the packages ggplot, tidyverse and sqldf (Wickham, 2009; Grothendieck, 2017; Wickham & Henry, 2019).

To groundtruth the tracking data, we manually observed all the videos and for each fish determined the direction of movement (groundtruth) (i.e. fish moving mainly right or left). The net movement of each video was determined (direction with the highest proportion for the video). We then compared the ground truth output with the net movement direction from the three OT architectures (Supplementary B).

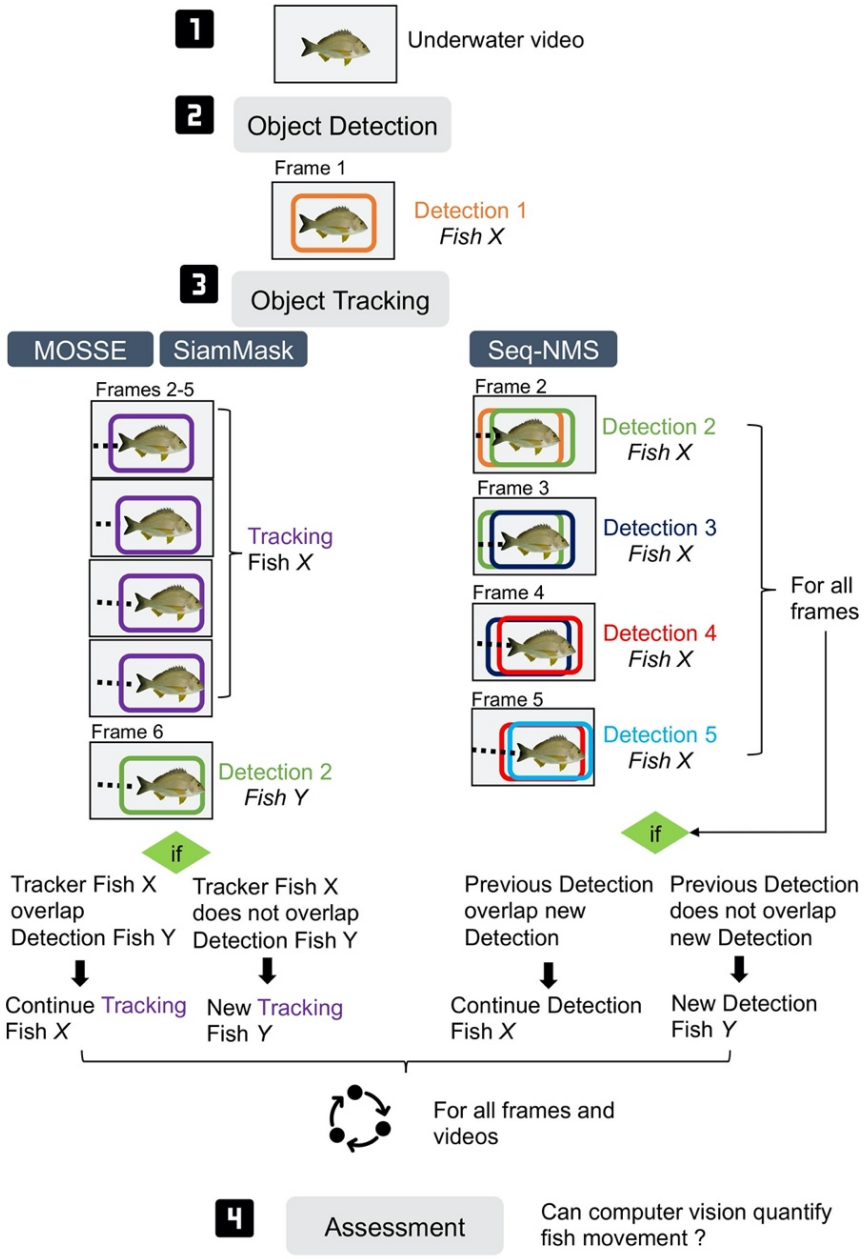



**FIGURE 2** Interaction between the object detection model and tracking architectures. The object detection model activates all three tracking architectures. For MOSSE and SiamMask the tracker continues for 4 frames after the initial detection. For Seq-NMS, the movement was determined by calculating the vector direction between two detections. For all architectures a check was made to determine if the tracker continued, stopped, or a new tracker started. For MOSSE and SiamMask the check was made after 4 tracking frames from the first detection. For SeqNMS the check was made for all frames after the first detection. The interaction between detections and tracker occurred through the whole length of a video where the object detection model detected a yellowfin bream and was carried for all frames, videos and cameras. All trackers provided a direction of movement for each frame where the interaction between the detection and tracking occurred successfully.

## 3. Results

### 3.1 Object detection

When using the Mask R-CNN framework for detecting yellowfin bream we obtained an 81% mAP50 value and an F1 score of 91% (Table 1). The OD model missed 21 fish (false negatives) and misidentified 8 objects (i.e. algae or other fish) as bream (false positives) out of the 169 fish (ground-truth) that were observed.

**TABLE 1** Object detection map50 and the evaluation results of the Mask R-CNN yellowfin bream model. The confusion matrix is shown as counts of individual fish, where the true positives were the correct detection of yellowfin bream. Yellowfin bream not detected were false negatives and misidentified objects were false positives.

| Task | mAP50 | Confusion matrix Ground-truth | Confusion matrix True Positives | Confusion matrix False Positives | Confusion matrix False Negatives | Average Precision | Average Recall | F1 |
|---|---|---|---|---|---|---|---|---|
| **Object detection** | 81% | 169 | 148 | 8 | 21 | 95% | 88% | 91% |

### 3.2 Object tracking

When comparing the performance of the OT architectures, we found that the three architectures detected and subsequently tracked more than 120 of the 169 individual fish that swam through the rockwall passage (Table 2). Average precision values for all architectures were above 80% with Seq-NMS being the most precise (93%) at detecting and tracking the yellowfin bream. Recall among architectures were very similar at around 73%. The architecture with the highest overall success at detecting and tracking bream was Seq-NMS (F1=84%) (Table 2).

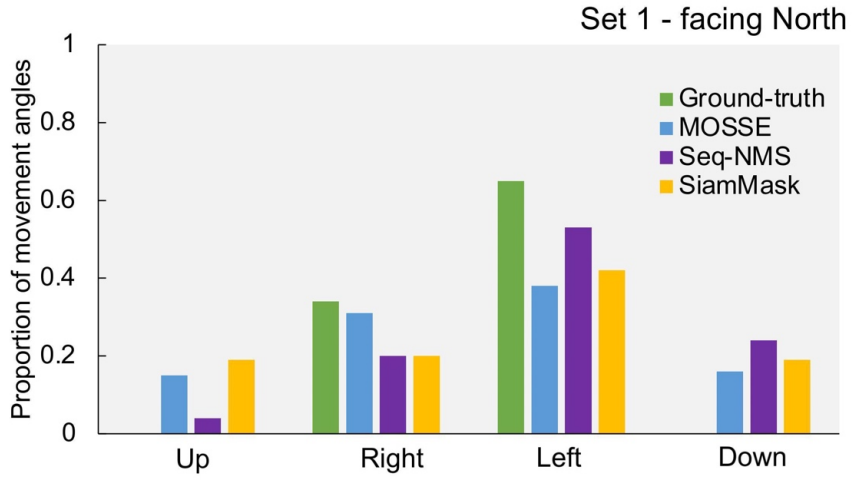
**TABLE 2** Confusion matrix for the three object tracking architectures (MOSSE, Seq-NMS and SiamMask) are shown as counts of individual fish, where the true positive means a bream was detected and tracked correctly for [?] 50% of the time when it appeared on a video frame, otherwise, it was false negative. False positives were misidentified objects (i.e. algae or other fish) that were detected and tracked.

| Architecture | Confusion matrix True Positives | Confusion matrix False Positives | Confusion matrix False Negatives | Average Precision | Average Recall | F1 |
|---|---|---|---|---|---|---|
| **MOSSE** | 123 | 23 | 46 | 84% | 73% | 78% |
| **Seq-NMS** | 129 | 9 | 40 | 93% | 76% | 84% |
| **SiamMask** | 121 | 19 | 48 | 86% | 72% | 78% |

### 3.3 Movement assessment

We expected the cameras to detect and track fish moving in the corridor consistent with the direction of the tidal flow (i.e. bream moving to seagrass). The expected results were that bream would mostly move to the left (Set 1) and to the right (Set 2) and these patterns were observed when manually analysing the videos (ground-truth). We found that the movement direction with the highest proportion for all tracking architectures were left (Set 1) and right (Set 2) (Figure 3). For Set 1, SeqNMS (0.53) was the closest to the ground-truth (0.65) and for Set 2, MOSSE (0.49) and SeqNMS (0.41) were the closest to the ground-truth

(0.71) (Figure 3).



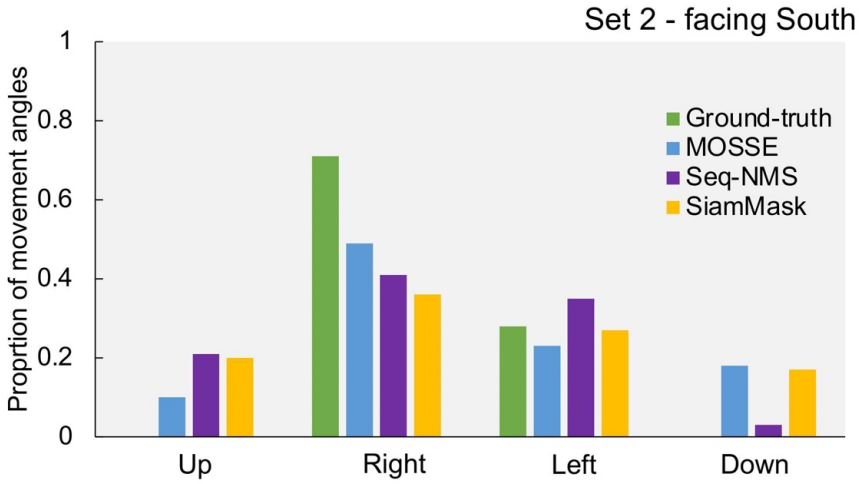



**FIGURE 3** Proportion of the movement angles (up, down, right=east, left=west) for the ground-truth and the three tracking architectures and for the two camera sets (Set 1: facing North and Set 2: facing South). The movement angles are spatial angles of yellowfin bream movement in two dimensions.

## 4. Discussion

We demonstrate a computer vision-based method for detecting and tracking individual fish in underwater footage. Our study incorporates open-source CV methods into a pipeline that allows scientists to assess animal movement in marine ecosystems. This method quantified animal behaviour and detected the expected tidal movement in our case study. The experimental results show that the proposed method is an effective and non-invasive way to detect and track small-scale movement of many fish in aquatic environments.

Previous ecological work has tracked fish in controlled environments (Papadakis, Glaropoulos & Kentouri, 2014; Qian et al., 2016; Bingshan et al., 2018; Sridhar, Roche & Gingins, 2019), used automated detections and counts as proxies for movement (Marini et al., 2018) and, most recently, used automated movement tracking algorithms to quantify movement (Francisco, Nührenberg & Jordan, 2020). Automated approaches tested in 'real-world' scenarios provide the best indication and evidence that CV is a robust technique for fish monitoring in aquatic ecosystems. In this paper, we propose an easily replicable and non-invasive method to measure fish movement in aquatic ecosystems by combining OD and OT algorithms. The object detection framework used in our study (Mask R-CNN) was recently shown to be robust and accurate enough to detect fish in a variety of aquatic conditions (Ditria et al., 2020b; Francisco, Nührenberg & Jordan, 2020). While other more recent OD frameworks have been developed since Mask R-CNN was published, our study further demonstrates that Mask R-CNN is capable of detecting fish in underwater footage. When evaluating the OT architectures, Seq-NMS had the best performance and was able to quantify the net movement of multiple individuals. While Seq-NMS is not an OT algorithm, it does require a high-performing OD model because it uses the OD outputs of every frame to create the detection links and track the movement direction. Additionally, for both OD and OT, we used frameworks that were not initially designed to detect fish in underwater footage. Our results add to the growing evidence that the learning capabilities and adaptability of CV methods can aid in the data collection and analysis of fish detection and tracking in aquatic ecosystems (Xiu et al., 2015; Villon et al., 2016; Marini et al., 2018).

A key benefit of camera-CV applications to animal movement research, and science more broadly, is that it can complement traditional data collection techniques (Lopez-Marcano et al., 2020). Cameras and CV can be deployed at many sites and cover large spatial extents but are limited by environmental factors and incapable of detecting and classifying complex ecological parameters such predatory interactions or the identification of morphologically similar, but taxonomically different species (Christin, Hervet & Lecomte, 2019). Traditional approaches (i.e. netting or in-water diver assessments) are still more capable at collecting the highest variety and complexity of ecological variables and parameters, but by combining cameras, automation and traditional approaches the spatial and temporal scope of monitoring can be increased. Moreover, camera-CV approaches do not require specialised equipment to study animal movement and the rapid analysis of imagery can provide movement data that is accurate, valid and consistent (Weinstein, 2018; Francisco, Nührenberg & Jordan, 2020).

CV techniques can enhance animal movement ecology through the streamlined collection of several sets of ecological information (Botella et al., 2018; Christin, Hervet & Lecomte, 2019), and this new data may revolutionize ecological studies. Traditional presence/absence data is used to understand the environmental drivers of a species' geographic distribution, and the collection of presence/absence data from videos can easily be automated (Schneider, Taylor & Kremer, 2018; Schneider et al., 2019; González-Rivero et al., 2020; Kennedy et al., 2020). However, presence/absence data by themselves cannot inform us about how multiple ecological processes interact, and presence/absence data conflates movement of individuals with mortality (Zurell, Pollock & Thuiller, 2018). Future studies could use our combined OD and OT approach to simultaneously quantify species distributions and movement. The integration of movement data into species distribution models means that the models could accurately predict how the ranges of mobile species

respond dynamically to environmental change through individual movement decisions and population level parameters like mortality (Bruneel et al., 2018).

The capacity to use our CV approach for monitoring fish populations is dependent on the ability to obtain and deploy several underwater cameras across the desired seascape. In this study, we deployed a six camera array in a fish corridor to maximise the chances to obtain movement data. However, each set and camera obtained unequal amounts of data and the array also resulted in repeated tracking of fish. Therefore, a major task when using camera-based technologies is to design and deploy an appropriate camera system to monitor animal interactions (Wearn & Glover-Kapfer, 2019). A recent global survey suggested that methodological improvements in the quality and accessibility of methods and analytical tools for camera-based technologies are still required (Glover-Kapfer, Soto-Navarro & Wearn, 2019). While our study demonstrates that fish can be detected and tracked automatically in aquatic ecosystems, further research into methodological designs (i.e. the optimal number of cameras needed to detect movement) are still required. The development of standardised camera-based methodologies, such as methodological guides for baited remote underwater surveys (Langlois et al., 2020) or for camera traps (Rovero et al., 2013), but specific to ecological camera-CV applications will help advance the applications of CV into movement ecology. Furthermore, the combination of both traditional and emerging techniques can provide data that can increase our understanding of complex movement behaviours in marine ecosystems (Christin, Hervet & Lecomte, 2019; Lopez-Marcano et al., 2020).

Remote camera systems and CV techniques can help provide robust, reliable and automatic tools to monitor and observe fish movement in marine ecosystems (Rowcliffe et al., 2016; Francisco, Nührenberg & Jordan, 2020). Technological advances have allowed us to better understand the complexities of animal movement, and our study shows that these techniques can be successfully applied in complex marine scenarios (Weinstein, 2018). By utilising a combination of CV frameworks, we demonstrated that automated tracking of fish movement between distinct seascapes (e.g. artificial and natural) is possible. We suggest that these methods are transferable to other types of fish corridors and other habitats, such as the mangrove, seagrass and coral reef continuum (Spampinato et al., 2008; Olds et al., 2018; Francisco, Nührenberg & Jordan, 2020). Further development of these models and architectures, such as integrated OD and OT with stereo video (Huo et al., 2018) and pairwise comparisons of detections (Guo et al., 2020), will likely lead to improvements in accuracy. Continual improvements in accuracy will provide a rigorous framework to study and quantify fish connectivity in the wild.

## 5. Conclusion

Computer vision and automated techniques offer a new generation of methods for collecting and analysing movement data. Our approach complements, rather than replaces, traditional techniques. Although current CV techniques have certain limitations, we demonstrated that CV can monitor small-scale movement of many individuals from underwater footage. CV has the capacity to provide several streams of ecological information that can inform data-driven decision that directly influence the health and productivity of marine ecosystems.

## 6. Acknowledgments

The authors acknowledge Adam Shand, Mia Turner and Mischa Turschwell for help in the field and for comments on the manuscript. Funding was provided by the Microsoft AI for Earth program. The work was also supported by the Global Wetlands Project, with support by a charitable organization which neither seeks nor permits publicity for its efforts.

## 7. Author Contributions

SL-M, CJB, RMC and DW designed the study. SL-M and EJ designed the models and computational frameworks with guidance of DW and BK. SL-M and ED collected the data. SL-M and CAB analysed the data. SL-M and RMC wrote the paper with input from all authors.

## 8. Data Availability

The training images and annotations, movement dataset annotations, images and videos, and the tracking and data wrangling scripts have been made available at (https://github.com/slopezmarcano/automated-fish-tracking).

## 9. References

Aguzzi, J., Chatzievangelou, D., Marini, S., Fanelli, E., Danovaro, R., Flogel, S., . . . Company, J.B. (2019). New high-tech flexible networks for the monitoring of deep-sea ecosystems. Environmental Science & Technology, 53, 6616-6631. http://dx.doi.org/10.1021/acs.est.9b00409

Arvind, C.S., Prajwal, R., Bhat, P.N., Sreedevi, A., Prabhudeva, K.N. & Ieee (2019). Fish detection and tracking in pisciculture environment using deep instance segmentation. In *Proceedings of the 2019 ieee region 10 conference* , (pp. 778-783). *New York:* IEEE.

Beyan, C. & Browman, H.I. (2020). Setting the stage for the machine intelligence era in marine science. ICES Journal of Marine Science, http://dx.doi.org/10.1093/icesjms/fsaa084

Bicknell, A.W.J., Godley, B.J., Sheehan, E.V., Votier, S.C. & Witt, M.J. (2016). Camera technology for monitoring marine biodiversity and human impact. Frontiers in Ecology and the Environment, 14, 424-432. http://dx.doi.org/10.1002/fee.1322

Bingshan, N., Guangyao, L., Fang, P., Jing, W., Long, Z. & Li, Z. (2018). Survey of fish behavior analysis by computer vision. Journal of Aquaculture Research and Development, 9, 15. http://dx.doi.org/10.4172/2155-9546.1000534

Bolme, D.S., Beveridge, J.R., Draper, B.A. & Lui, Y.M. (2010) Visual object tracking using adaptive correlation filters. *IEEE Conference on Computer Vision and Pattern Recognition* , pp. 2544-2550.

Botella, C., Joly, A., Bonnet, P., Monestiez, P. & Munoz, F. (2018). Species distribution modeling based on the automated identification of citizen observations. Applications in plant sciences, 6, e1029-e1029. http://dx.doi.org/10.1002/aps3.1029

Bruneel, S., Gobeyn, S., Verhelst, P., Reubens, J., Moens, T. & Goethals, P. (2018). Implications of movement for species distribution models - rethinking environmental data tools. Science of the Total Environment, 628-629, 893-905. http://dx.doi.org/10.1016/j.scitotenv.2018.02.026

Cheng, J.C., Tsai, Y.H., Hung, W.C., Wang, S.J. & Yang, M.H. (2018) Fast and accurate online video object segmentation via tracking parts.*IEEE Conference on Computer Vision and Pattern Recognition* , pp. 7415-7424.

Christin, S., Hervet, E. & Lecomte, N. (2019). Applications for deep learning in ecology. Methods in Ecology and Evolution, 10, http://dx.doi.org/10.1111/2041-210x.13256

Chuang, M., Hwang, J., Ye, J., Huang, S. & Williams, K. (2017). Underwater fish tracking for moving cameras based on deformable multiple kernels. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 47, 2467-2477. http://dx.doi.org/10.1109/TSMC.2016.2523943

Cui, S., Zhou, Y., Wang, Y. & Zhai, L. (2020). Fish detection using deep learning. Applied Computational Intelligence and Soft Computing, 2020, 3738108. http://dx.doi.org/10.1155/2020/3738108

Ditria, E., Sievers, M., Lopez-Marcano, S., Jinks, E.L. & Connolly, R.M. (2020a). Deep learning for automated analysis of fish abundance: The benefits of training across multiple habitats. bioRxiv, http://dx.doi.org/10.1101/2020.05.19.105056

Ditria, E.M., Lopez-Marcano, S., Sievers, M., Jinks, E.L., Brown, C.J. & Connolly, R.M. (2020b). Automating the analysis of fish abundance using object detection: Optimizing animal ecology with deep learning. Frontiers in Marine Science, 7, http://dx.doi.org/10.3389/fmars.2020.00429

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J. & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International Journal of Computer Vision, 88, 303-338. http://dx.doi.org/10.1007/s11263-009-0275-4

Francisco, F.A., Nührenberg, P. & Jordan, A. (2020). High-resolution, non-invasive animal tracking and reconstruction of local environment in aquatic ecosystems. Movement Ecology, 8, 27. http://dx.doi.org/10.1186/s40462-020-00214-w

Glover-Kapfer, P., Soto-Navarro, C.A. & Wearn, O.R. (2019). Camera-trapping version 3.0: Current constraints and future priorities for development. Remote Sensing in Ecology and Conservation, 5, 209-223. http://dx.doi.org/10.1002/rse2.106

González-Rivero, M., Beijbom, O., Rodriguez-Ramirez, A., Bryant, E.P.D., Ganase, A., Gonzalez-Marrero, Y., . . . Hoegh-Guldberg, O. (2020). Monitoring of coral reefs using artificial intelligence: A feasible and cost-effective approach. Remote Sensing, 12, http://dx.doi.org/10.3390/rs12030489

Grothendieck, G. (2017) Sqldf: Manipulate r data frames using sql. R package version 0.4-11.

Guirado, E., Tabik, S., Rivas, M.L., Alcaraz-Segura, D. & Herrera, F. (2019). Whale counting in satellite and aerial images with deep learning. Scientific Reports, 9, 14259. http://dx.doi.org/10.1038/s41598-019-50795-9

Guo, S., Xu, P., Miao, Q., Shao, G., Chapman, C.A., Chen, X., . . . Li, B. (2020). Automatic identification of individual primates with deep learning techniques. iScience, 23, 101412. http://dx.doi.org/https://doi.org/10.1016/j.isci.2020.101412

Han, W., Khorrami, P., Le Paine, T., Ramachandran, P., Babaeizadeh, M., Shi, H., . . . Huang, T. (2016). Seq-nms for video object detection. ArXiv, 3, http://dx.doi.org/arXiv:1602.08465

He, K.M., Gkioxari, G., Dollar, P. & Girshick, R. (2017). Mask r-cnn. IEEE International Conference on Computer Vision, 2980-2988. http://dx.doi.org/10.1109/Iccv.2017.322

Huo, G., Wu, Z., Li, J. & Li, S. (2018). Underwater target detection and 3d reconstruction system based on binocular vision. Sensors, 18, 3570. http://dx.doi.org/10.3390/s18103570

Jalal, A., Salman, A., Mian, A., Shortis, M. & Shafait, F. (2020). Fish detection and species classification in underwater environments using deep learning with temporal information. Ecological Informatics, 57, 101088. http://dx.doi.org/10.1016/j.ecoinf.2020.101088

Kälin, U., Lang, N., Hug, C., Gessler, A. & Wegner, J.D. (2019). Defoliation estimation of forest trees from ground-level images. Remote Sensing of Environment, 223, 143-153. http://dx.doi.org/10.1016/j.rse.2018.12.021

Kennedy, E.V., Vercelloni, J., Neal, B.P., Ambariyanto, Bryant, D.E.P., Ganase, A., . . . Hoegh-Guldberg, O. (2020). Coral reef community changes in karimunjawa national park, indonesia: Assessing the efficacy of management in the face of local and global stressors. Journal of Marine Science and Engineering, 8, 760. http://dx.doi.org/10.3390/jmse8100760

Kezebou, L., Oludare, V., Panetta, K. & Agaian, S.S. (2019) Underwater object tracking benchmark and dataset. *2019 IEEE International Symposium on Technologies for Homeland Security (HST)* , pp. 1-6.

Langlois, T., Goetze, J., Bond, T., Monk, J., Abesamis, R.A., Asher, J., . . . Harvey, E.S. (2020). A field and video annotation guide for baited remote underwater stereo-video surveys of demersal fish assemblages. Methods in Ecology and Evolution, http://dx.doi.org/10.1111/2041-210X.13470

Lantsova, E., Voitiuk, T., Zudilova, T. & Kaarna, A. (2016) Using low-quality video sequences for fish detection and tracking. *2016 SAI Computing Conference (SAI)* , pp. 426-433.

Lecun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. Nature, 521, 436-444. http://dx.doi.org/10.1038/nature14539

Lopez-Marcano, S., Brown, C.J., Sievers, M. & Connolly, R.M. (2020). The slow rise of technology: Computer vision techniques in fish population connectivity. Aquatic Conservation: Marine and Freshwater Ecosystems, http://dx.doi.org/10.1002/aqc.3432

Marini, S., Fanelli, E., Sbragaglia, V., Azzurro, E., Fernandez, J.D. & Aguzzi, J. (2018). Tracking fish abundance by underwater image recognition. Scientific Reports, 8, 13748. http://dx.doi.org/10.1038/s41598-018-32089-8

Massa, F. & Girshick, R. (2018) Maskrcnn-benchmark: Fast, modular reference implementation of instance segmentation and object detection algorithms in pytorch. Retrieved 29 October 2020, from https://github.com/facebookresearch/maskrcnn-benchmark

Mohamed, H.E.-D., Fadl, A., Anas, O., Wageeh, Y., ElMasry, N., Nabil, A. & Atia, A. (2020). Msr-yolo: Method to enhance fish detection and tracking in fish farms. Procedia Computer Science, 170, 539-546. http://dx.doi.org/10.1016/j.procs.2020.03.123

Olds, A.D., Nagelkerken, I., M Huijbers, C., Gilby, B., Pittman, S. & Schlacher, T. (2018). Connectivity in coastal seascapes. In S.J. Pittman, *Seascape ecology* , (pp. 261-291). *Hoboken, NJ:*John Wiley & Sons Ltd.

Pagès, J.F., Gera, A., Romero, J. & Alcoverro, T. (2014). Matrix composition and patch edges influence plant–herbivore interactions in marine landscapes. Functional Ecology, 28, 1440-1448. http://dx.doi.org/10.1111/1365-2435.12286

Papadakis, V.M., Glaropoulos, A. & Kentouri, M. (2014). Sub-second analysis of fish behavior using a novel computer-vision system. Aquacultural Engineering, 62, 36-41. http://dx.doi.org/10.1016/j.aquaeng.2014.06.003

Prechelt, L. (2012). Early stopping — but when? In G. Montavon, G.B. Orr & K.-R. Müller, *Neural networks: Tricks of the trade: Second edition* , (pp. 53-67). *Berlin, Heidelberg:* Springer Berlin Heidelberg.

Qian, Z.-M., Wang, S.H., Cheng, X.E. & Chen, Y.Q. (2016). An effective and robust method for tracking multiple fish in video image based on fish head detection. BMC Bioinformatics, 17, 251. http://dx.doi.org/10.1186/s12859-016-1138-y

Rovero, F., Zimmermann, F., Berzi, D. & Meek, P. (2013). "Which camera trap type and how many do i need?" A review of camera features and study designs for a range of wildlife research applications. Hystrix, the Italian Journal of Mammalogy, 24, 148-156. http://dx.doi.org/10.4404/hystrix-24.2-8789

Rowcliffe, J.M., Jansen, P.A., Kays, R., Kranstauber, B. & Carbone, C. (2016). Wildlife speed cameras: Measuring animal travel speed and day range using camera traps. Remote Sensing in Ecology and Conservation, 2, 84-94. http://dx.doi.org/10.1002/rse2.17

Salberg, A. (2015) Detection of seals in remote sensing images using features extracted from deep convolutional neural networks. *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* , pp. 1893-1896.

Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J. & Harvey, E. (2016). Fish species classification in unconstrained underwater environments based on deep learning. Limnology and Oceanography: Methods, 14, 570-585. http://dx.doi.org/10.1002/lom3.10113

Schneider, S., Taylor, G.W. & Kremer, S.C. (2018). Deep learning object detection methods for ecological camera trap data. 2018 15th Conference on Computer and Robot Vision (Crv), 321-328. http://dx.doi.org/10.1109/Crv.2018.00052

Schneider, S., Taylor, G.W., Linquist, S. & Kremer, S.C. (2019). Past, present and future approaches using computer vision for animal re-identification from camera trap data. Methods in Ecology and Evolution, 10, 461-470. http://dx.doi.org/10.1111/2041-210X.13133

Sidhu, R. (2016). Tutorial on minimum output sum of squared error filter. Degree of Master of Science, Colorado State University.

Spampinato, C., Chen-Burger, Y.H., Nadarajan, G. & Fisher, R.B. (2008) Detecting, tracking and counting fish in low quality unconstrained underwater videos. *VISAPP Proceedings of the Third International Conference on Computer Vision Theory and Applications* , pp. 514-+.

Sridhar, V.H., Roche, D.G. & Gingins, S. (2019). Tracktor: Image-based automated tracking of animal movement and behaviour. Methods in Ecology and Evolution, 10, 815-820. http://dx.doi.org/10.1111/2041-210X.13166

van Gemert, J.C., Verschoor, C.R., Mettes, P., Epema, K., Koh, L.P. & Wich, S. (2015) Nature conservation drones for automatic localization and counting of animals. *Computer Vision - ECCV 2014 Workshops*(eds L. Agapito, M.M. Bronstein & C. Rother), pp. 255-270. Springer International Publishing, Cham.

Villon, S., Chaumont, M., Subsol, G., Villéger, S., Claverie, T. & Mouillot, D. (2016). Coral reef fish detection and recognition in underwater videos by supervised machine learning: Comparison between deep learning and hog+svm methods. In, (pp. 160-171). Springer International Publishing.

Villon, S., Mouillot, D., Chaumont, M., Darling, E.S., 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. http://dx.doi.org/10.1016/j.ecoinf.2018.09.007

Villon, S., Mouillot, D., Chaumont, M., Subsol, G., Claverie, T. & Villéger, S. (2020). A new method to control error rates in automated species identification with deep learning algorithms. Scientific Reports, 10, 10972. http://dx.doi.org/10.1038/s41598-020-67573-7

Waldchen, J. & Mader, P. (2018). Machine learning for image based species identification. Methods in Ecology and Evolution, 9, 2216-2225. http://dx.doi.org/10.1111/2041-210x.13075

Wang, Q., Zhang, L., Bertinetto, L., Hu, W. & Torr, P. (2019). Fast online object tracking and segmentation: A unifying approach. ArXiv, http://dx.doi.org/arxiv:1812.05050

Watanabe, J.-I., Shao, Y. & Miura, N. (2019). Underwater and airborne monitoring of marine ecosystems and debris. Journal of Applied Remote Sensing, 13, 044509. http://dx.doi.org/10.1117/1.JRS.13.044509

Wearn, O.R. & Glover-Kapfer, P. (2019). Snap happy: Camera traps are an effective sampling tool when compared with alternative methods. Royal Society Open Science, 6, 181748. http://dx.doi.org/10.1098/rsos.181748

Weinstein, B.G. (2018). A computer vision for animal ecology. Journal of Animal Ecology, 87, 533-545. http://dx.doi.org/10.1111/1365-2656.12780

Wickham, H. (2009) Ggplot2: Elegant graphics for data analysis.*Ggplot2: Elegant Graphics for Data Analysis* , pp. 1-212.

Wickham, H. & Henry, L. (2019) Tidyr: Tidy messy data

Xiu, L., Min, S., Qin, H. & Liansheng, C. (2015) Fast accurate fish detection and recognition of underwater images with fast r-cnn. IEEE.

Xu, Z. & Cheng, X.E. (2017). Zebrafish tracking using convolutional neural networks. Scientific Reports, 7, 42815. http://dx.doi.org/10.1038/srep42815

Zhao, Z.-Q., Zheng, P., Xu, S.-t. & Wu, X. (2019). Object detection with deep learning: A review. ArXiv, http://dx.doi.org/arxiv:1807.05511

Zurell, D., Pollock, L.J. & Thuiller, W. (2018). Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments? Ecography, 41, 1812-1819. http://dx.doi.org/10.1111/ecog.03315