

# Automatic detection of fish and tracking of movement for ecology

Sebastian Lopez-Marcano<sup>1,2</sup>  | Eric L. Jinks<sup>1</sup> | Christina A. Buelow<sup>1</sup> |  
Christopher J. Brown<sup>1</sup> | Dadong Wang<sup>2</sup> | Branislav Kusy<sup>3</sup> | Ellen M. Ditrta<sup>1</sup> |  
Rod M. Connolly<sup>1</sup>

<sup>1</sup>Coastal and Marine Research Centre, Australian Rivers Institute, School of Environment and Science, Griffith University, Gold Coast, QLD, Australia

<sup>2</sup>Quantitative Imaging Research Team, CSIRO, Marsfield, NSW, Australia

<sup>3</sup>Data61, CSIRO, Pullenvale, QLD, Australia

## Correspondence

Sebastian Lopez-Marcano, Coastal and Marine Research Centre, Australian Rivers Institute, School of Environment and Science, Griffith University, Gold Coast, QLD 4222, Australia.

Email: 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, labor intensive, costly, and limited in the number of individuals that can be feasibly tracked. Automated detection and tracking of small-scale movements of many animals through cameras are possible but are largely untested in field conditions, hampering applications to ecological questions.
2. Here, we aimed to test the ability of an automated object detection and object tracking pipeline to track small-scale movement of many individuals in videos. We applied the pipeline to track fish movement in the field and characterize movement behavior. We automated the detection of a common fisheries species (yellowfin bream, *Acanthopagrus australis*) along a known movement passageway from underwater videos. 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 multispecies assemblage ( $F_1$  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 this integrated object detection and tracking pipeline, we demonstrated a noninvasive and reliable approach to studying fish behavior by tracking their movement under field conditions. These cost-effective technologies provide a means for future studies to scale-up the analysis of movement across many visual monitoring systems.

## KEY WORDS

computer vision, connectivity, deep learning, dispersal, machine learning, object tracking, underwater video

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Ecology and Evolution published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

Computer vision, the research field that explores the use of computer algorithms to automate the interpretation of digital images or videos, is revolutionizing data collection in science (Beyan & Browman, 2020; Waldchen & Mader, 2018). The use of remote camera imagery, such as underwater stations, camera traps, and stereography, has driven the uptake of computer vision because it can process and analyze imagery quickly and accurately (Bicknell et al., 2016; Schneider et al., 2019). In ecological studies, advances in computer vision have led to increased 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 computer vision are monitoring deep-sea ecosystems (Aguzzi et al., 2019), and computer vision-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 computer vision to study and monitor marine ecosystems. These applications are related to the two main computer vision tasks: object detection and object tracking. Object detection and object tracking automate data collection, including gathering information about the type, location, and movement of objects of interest. Object detection algorithms can count and identify species of interest in underwater video footage (Christin et al., 2019) and have been applied to detect seals (Salberg, 2015), identify whale hotspots (Guirado et al., 2019), monitor fish populations (Ditria, Lopez-Marcano, et al., 2020; Jalal et al., 2020; Marini et al., 2018; Salman et al., 2016; Villon et al., 2016, 2018, 2020; Xiu et al., 2015), and quantify floating debris on the ocean surface (Watanabe et al., 2019). On the other hand, object tracking can locate and output the movement direction and speed of objects between video frames. In marine ecosystems, object tracking has been used to track on-surface objects (see [topios.org](http://topios.org)) and underwater objects such as fish, sea turtles, dolphins, and whales (Arvind et al., 2019; Chuang et al., 2017; Kezebou et al., 2019; Spampinato et al., 2008; Xu & Cheng, 2017).

There is evidence that automated monitoring of fish in underwater ecosystems through the combination of object detection and object tracking is reliable and accurate (Lantsova et al., 2016; Mohamed et al., 2020; Spampinato et al., 2008). However, no studies have jointly applied object detection and object tracking for animal movement studies. Object detection can automatically collect traditional presence/absence data of different species (Marini et al., 2018; Xiu et al., 2015) while object tracking simultaneously tracks individuals to provide fine-scale data to assess behavioral and animal movement patterns (Francisco et al., 2020). Combining object detection and tracking in a single and noninvasive automated approach enhances the amount of ecologically relevant information extracted from videos. This subsequently improves, for example, our ability to quantify and evaluate environmental drivers of species abundance, diversity, movement, and behavior.

The utility of combining object detection and tracking is particularly useful for studying animal movement, which typically requires large volumes of data for many individuals (Librán-Embid et al., 2020; Lopez-Marcano et al., 2020). Knowledge of animal movement is fundamental to many research objectives in marine science, as 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). Collecting movement data is, however, challenging and requires substantial resources. The development and applications of automated technologies (i.e., object detection and tracking pipelines) can overcome these restrictions and help advance our understanding of animal movement across a broad range of spatio-temporal scales 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 computer vision pipeline consisting of two steps, object detection and object tracking, and we used the pipeline to quantify underwater animal movement across habitats. To demonstrate the benefits of combining object detection and object tracking, we deployed cameras in a known coastal fish estuarine passageway and recorded the movement of a common fisheries species (yellowfin bream, *Acanthopagrus australis*). 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 computer vision 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 higher dimensional representations and detect patterns within unstructured data (Lecun et al., 2015; Schmidhuber, 2015). In this paper, we used deep learning, and more specifically Mask Regional Convolutional Neural Network (Mask R-CNN) for fish detection (Cui et al., 2020; Ditria, Lopez-Marcano, et al., 2020; Jalal et al., 2020; Villon et al., 2020). 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 bream in the Tweed River estuary, Australia ( $-28.169438, 153.547594$ ) between May and September 2019.

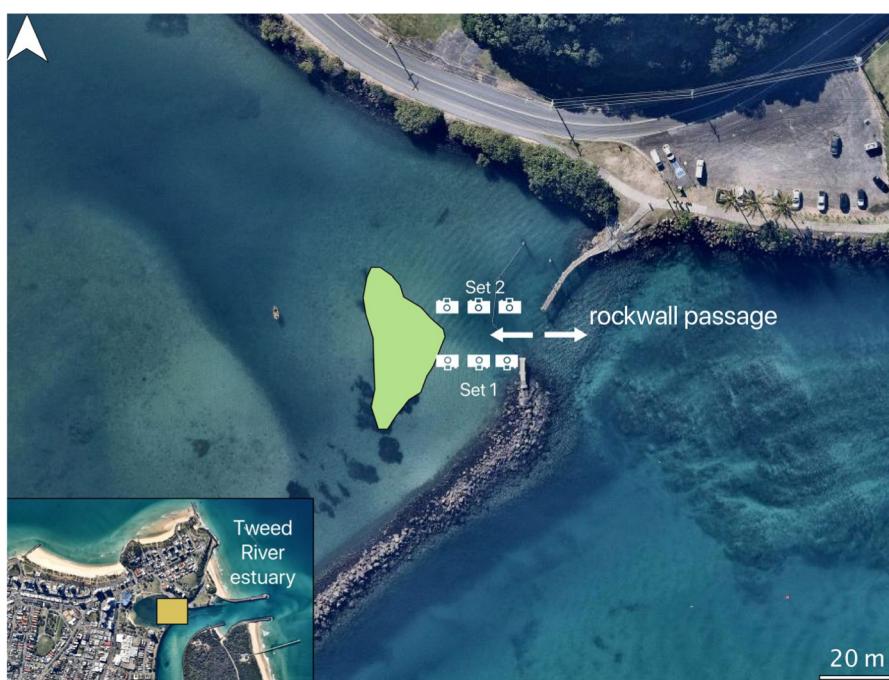
We used six submerged action cameras (1080p Haldex Sports Action Cam HD) deployed for 1 hr in a variety of marine habitats (e.g., rocky reefs and seagrass meadows). We varied the camera angle and placement to capture diverse backgrounds and fish angles (Ditria et al., 2020). We trimmed the original 1-hr videos into snippets where bream were present using VLC media player 3.0.8. The snippets were then converted 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 minimize 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. mAP50 is a measurement of the model's capacity to overlap a segmentation mask around 50% of the ground-truth outline of the fish (Everingham et al., 2010). We used a confidence threshold of 80%, meaning that we selected object detection outputs where the model was 80% or more confident that it was a bream. We developed the models and analyzed 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 can be obscured by floating objects, and which creates greater variation

in the shape and texture of the objects and their surroundings in a video (Sidhu, 2016). Advances in object tracking are addressing these issues, and objects can now be tracked consistently despite natural variations of the object's shape, size, and location (Bolme et al., 2010; Cheng et al., 2018). We developed a pipeline where the object tracking architecture activated once the object detection model detected a bream. This approach resulted in an automated detection and subsequent tracking of fish from the underwater videos. Additionally, we benchmarked the performance of three object tracking architectures: minimum output sum of squared errors (MOSSE) (Bolme et al., 2010), sequential nonmaximum suppression (Seq-NMS) (Han et al., 2016), and Siamese mask (SiamMask) (Wang et al., 2019) by using movement data gathered one 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 harbor (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. Multiple species of estuarine fish such as sand whiting (*Sillago ciliata*), river garfish (*Hyporhamphus regularis*), luderick (*Girella tricuspidata*), spotted scat (*Scatophagus argus*), three-bar porcupinefish (*Dicotylichthys punctulatus*), and bream, move back and forth with the tides through this passageway. This environment suffers from frequent low visibility and currents that bring floating debris, presenting a relatively challenging scenario in which to showcase the capacity of computer vision to detect the target species in a multispecies assemblage and quantify the direction of movement.

We collected fish movement data by submerging two sets of three action cameras (1080p Haldex Sports Action Cam HD) for 1 hr 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 toward the fish corridor and

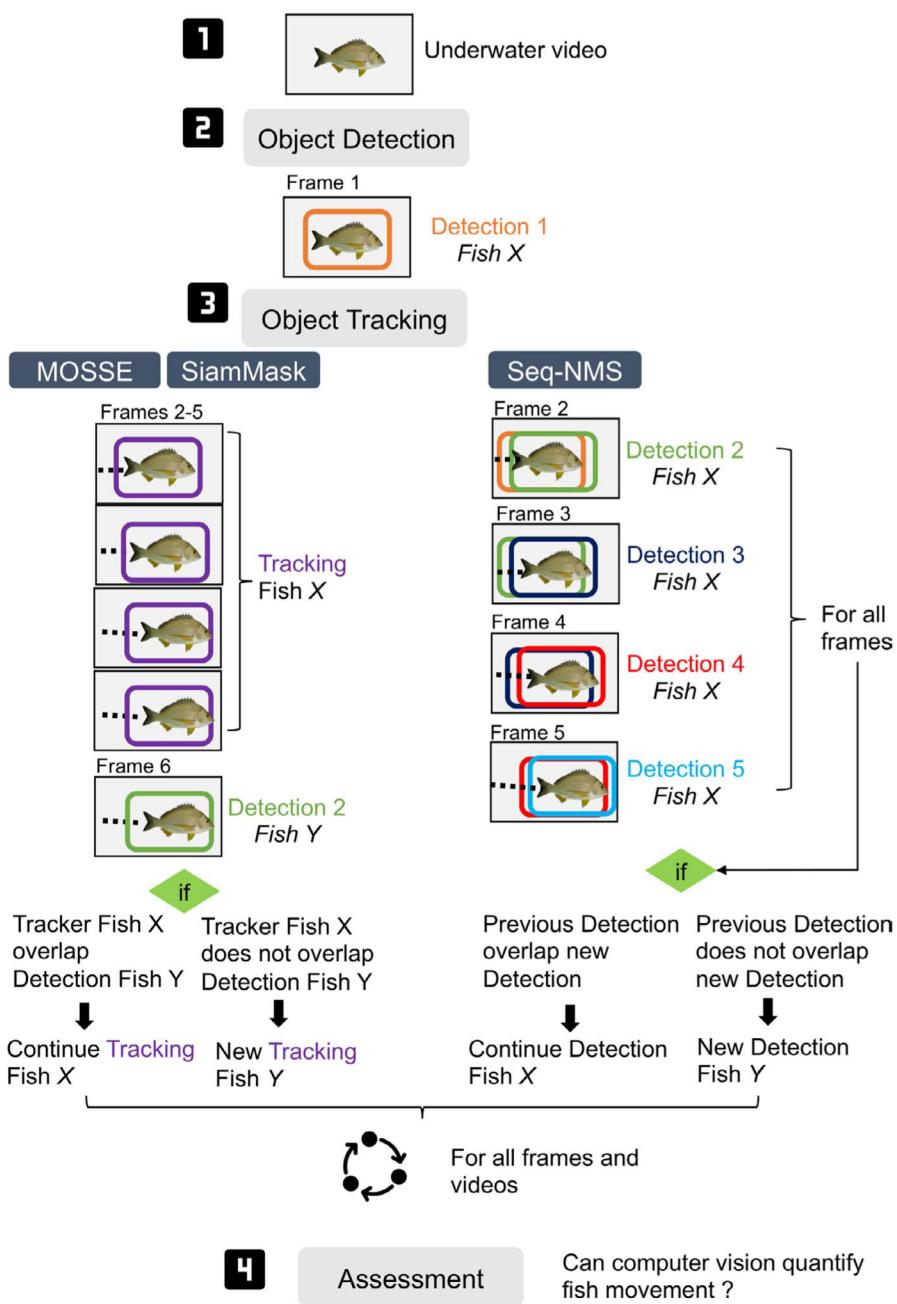


**FIGURE 1** The study location in Tweed River Estuary, Australia, showing the camera array deployed in a fish passageway (two ended white arrow) between the rock wall channel and the seagrass meadow (green polygon). Each set of cameras consisted of three underwater cameras that recorded for 1 hr during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

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 nonoverlapping 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 harbor 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 bream movement, resulting in 76 videos of varying durations (between 3 and 70 s), which we converted into still frames at 25 frames per second. All frames with bream were manually annotated, and these annotations were used as ground-truth. We used the fish movement dataset to evaluate the object detection model and the object tracking architectures.

## 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 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 object detection output (Figure 2). The object detection model and the object tracking architecture interacted to maintain the consistency of the tracker on bream individuals. When a fish was detected, the entry was used to initialize the tracker. MOSSE tracked the fish for four frames, and 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  $\geq 30\%$  with the existing tracker bounding box, the tracker continues on the same object. If the detection bounding box does not overlap with the existing tracker bounding box, a new tracker entry starts. This interaction between the detection and tracking occurred for every fish detected in a frame and stopped when no more detections were found.

### 2.2.2 | Sequential nonmaximum suppression (Seq-NMS)

Sequential nonmaximum suppression (Seq-NMS) was developed in 2016 traditionally 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 object detection output for every frame containing a fish. Seq-NMS links detections of neighboring 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 object tracking of fish when there was an overlap (intersection over union) of bounding boxes in subsequent frames of  $\geq 30\%$  (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).

### 2.2.3 | SiamMask

SiamMask is a tracking algorithm developed in 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 bream was detected (Figure 2).

We have made all object detection annotations, images, trackers, and data wrangling codes, as well as the movement dataset openly available (<https://doi.org/10.5281/zenodo.4571760>).

## 2.3 | Model evaluations and movement assessment

### 2.3.1 | Object detection evaluation

We evaluated the object detection against the movement data (manually annotated and ground-truthed) 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.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (1)$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (2)$$

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

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 an F1 score and by assessing the movement data. To calculate precision recall and an F1 score, we manually observed every second of video and determined if the object tracking architecture was correctly tracking the bream individual (Supplementary B). We defined a true positive as a correct detection and accurate tracking of the individual for  $\geq 50\%$  of the time where bream appeared on frame (Supplementary B). A false negative occurred when a bream was not detected and tracked or if it was tracked  $< 50\%$  of the time when the fish appeared on frame. Additionally, we classified a false positive when a nonbream object was detected and tracked, or when a bream was detected but the tracking architecture tracked a nonbream object.

### 2.3.3 | Movement assessment

We conducted a movement assessment to evaluate the accuracy of the directions provided by the tracker. From all trackers, we obtained the bounding boxes and centroids of the boxes for fish that were detected and subsequently tracked. For each tracking output, the object tracking architecture provided a tracking angle of movement in 2 dimensions relative to the camera frame. Depending on the camera set (Figure 1), we summarized angles for north-facing cameras (Set 1) and south-facing cameras (Set 2). We grouped tracking angles using reference angles into four directions: up, down, left, and right (Supplementary B). Because the cameras were facing horizontally toward the fish passageway parallel with the seafloor, we calculated horizontal movement of fish. Fish moving up meant that the fish movement had tracking angles between  $0^\circ$ – $44^\circ$  and  $315^\circ$ – $360^\circ$ . Fish moving right had angles

between 45° and 135°, whereas fish moving left between 225° and 315°. Finally, fish moving down had tracking angles between 135° and 225°. The tracking angle for all object tracking architectures was obtained from the tracker vector generated within each tracker's bounding box (Supplementary B). By grouping the directions, we can 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 with the packages ggplot and sqldf (Grothendieck, 2017; Wickham, 2009).

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

### 3 | RESULTS

#### 3.1 | Object detection

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

#### 3.2 | Object tracking

We simultaneously detected and tracked 1 to 30 individual bream per video. All three architectures detected and subsequently tracked more than 120 of the 169 individual fish that swam through the passageway (Table 2). Average precision values for all architectures were above 80%, with Seq-NMS the most precise at detecting and tracking the bream (93%). Recall among architectures was 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 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				Average precision	Average recall	F1
		Ground-truth	True positives	False positives	False negatives			
Object detection	81%	169	148	8	21	95%	88%	91%

#### 3.3 | Movement assessment

We expected the cameras to detect and track fish moving in the passageway 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 analyzing the videos (ground-truth). The movement direction with the highest proportion for all tracking architectures was left (Set 1) and right (Set 2) (Figure 3). For Set 1, Seq-NMS (0.53) was the closest to the ground-truth (0.65), and for Set 2, MOSSE (0.49) and Seq-NMS (0.41) were the closest to the ground-truth (0.71) (Figure 3).

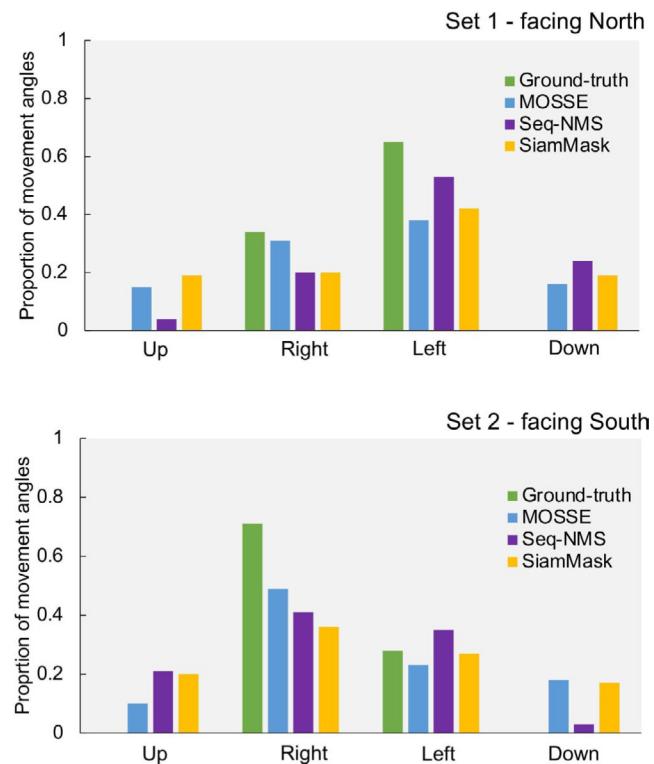
### 4 | DISCUSSION

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

Previous ecological work has tracked fish in controlled environments (Bingshan et al., 2018; Papadakis et al., 2014; Qian et al., 2016; Sridhar et al., 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 et al., 2020). Automated approaches tested in "real-world" scenarios provide the best indication and evidence that computer vision is a robust technique for fish monitoring in aquatic ecosystems. When evaluating the object tracking architectures, Seq-NMS had the best performance and was able to quantify the net movement of multiple individuals. The number of fish that were simultaneously detected and tracked by our framework ranged from 1 to 30 individuals per video. While the movement dataset did not contain videos with a very large number of individuals (e.g., >50 fish in a single frame), previous research has shown that occlusion can influence the accuracy of both the object detection algorithms and Seq-NMS (Connolly et al., 2021). Moreover, Seq-NMS is not an object tracking algorithm and it requires a high-performing object

Architecture	Confusion matrix					
	True positives	False positives	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%

**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  $\geq 50\%$  of the time when it appeared on a video frame, otherwise, it was false negative. False positives were misidentified objects (e.g., algae or other fish) that were detected and tracked



**FIGURE 3** Proportion of the movement angles (up, down, right, left) 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 horizontal yellowfin bream movement in two dimensions

detection model because it uses the object detection outputs of every frame to create the detection links and track the movement direction.

A key benefit of camera-assisted applications and computer vision analysis to animal movement research, and science more broadly, is that these approaches can complement traditional data collection techniques (Lopez-Marcano et al., 2020). Cameras and computer vision can be deployed at many sites and cover large spatial extents, but are limited by environmental factors and are incapable of detecting and classifying complex ecological parameters such as predatory interactions or the identification of morphologically similar, but taxonomically different, species (Christin et al., 2019). Traditional approaches (e.g., netting or in-water diver assessments) are superior 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, computer vision approaches do not require specialized equipment to study animal movement and the rapid analysis of imagery can provide movement data that is accurate, valid, and consistent (Francisco et al., 2020; Weinstein, 2018).

The combination of object detection and object tracking can enhance animal movement ecology through the streamlined collection of several sets of ecological information (Botella et al., 2018; Christin et al., 2019), and this new data may revolutionize ecological studies. Traditional presence/absence data can be used, for example, to understand the environmental drivers of a species' geographic distribution, and the collection of presence/absence data from videos can easily be automated (González-Rivero et al., 2020; Kennedy et al., 2020; Schneider et al., 2018, 2019). However, presence/absence data alone cannot inform about how multiple ecological processes interact, and presence/absence data conflate movement of individuals with mortality (Zurell et al., 2018). Future studies could use our combined object detection and object tracking 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 of our computer vision approach for monitoring fish populations is dependent on the underwater camera setup within the desired seascape. We deployed an array of cameras in a fish passageway to maximize the collection of movement data. However, each set and camera obtained unequal amounts of data and the array also resulted in repeated tracking of fish. Therefore, an important consideration when using camera-based technologies is to design and deploy an appropriate camera system to monitor animal interactions (Glover-Kapfer et al., 2019; Wearn & Glover-Kapfer, 2019). While we demonstrate that the detection and tracking of fish can be automated in aquatic ecosystems, further research into methodological designs (e.g., the optimal number of cameras needed to detect movement) is still required. The development of standardized 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 computer

vision-ecology applications will help advance the applications of computer vision into movement ecology.

By utilizing a combination of computer vision frameworks, we demonstrated that automated tracking of fish movement between distinct seascapes (i.e., artificial and natural) is possible. We suggest that these methods are transferable to other types of fish passageways and other habitats, such as the mangrove, seagrass, and coral reef continuum (Francisco et al., 2020; Olds et al., 2018; Spampinato et al., 2008). Further development of these models and architectures, for example, integrated object detection and object tracking with stereo video (Huo et al., 2018) and pairwise comparisons of detections (Guo et al., 2020), will likely lead to improvements in accuracy and for 3D triangulation of detections. 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 analyzing movement data. Our combined object detection and object tracking approach complements, rather than replaces, traditional techniques. Although current computer vision techniques have limitations, we demonstrated that object detection and object tracking can monitor small-scale movement of many individuals from underwater footage. The combination of object detection and object tracking 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.

## 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.

## CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest.

## AUTHOR CONTRIBUTIONS

**Sebastian Lopez-Marcano:** Conceptualization (lead); Data curation (lead); Formal analysis (equal); Funding acquisition (lead); Investigation (lead); Methodology (equal); Project administration (lead); Resources (equal); Software (supporting); Supervision (equal); Validation (lead); Visualization (equal); Writing-original draft (lead); Writing-review & editing (lead). **Eric Jinks:** Conceptualization (supporting); Data curation (supporting); Formal analysis (equal); Funding acquisition (supporting); Investigation (supporting); Methodology (equal); Project administration (supporting); Resources (supporting); Software (lead); Supervision (supporting); Validation (supporting);

Visualization (supporting); Writing-original draft (equal); Writing-review & editing (equal). **Christina A Buelow:** Conceptualization (supporting); Data curation (equal); Formal analysis (equal); Funding acquisition (supporting); Investigation (supporting); Methodology (supporting); Software (supporting); Supervision (supporting); Visualization (equal); Writing-original draft (equal); Writing-review & editing (equal). **Christopher J Brown:** Conceptualization (supporting); Data curation (supporting); Formal analysis (supporting); Methodology (supporting); Supervision (lead); Validation (supporting); Visualization (equal); Writing-original draft (equal); Writing-review & editing (equal). **Dadong Wang:** Conceptualization (supporting); Methodology (equal); Resources (equal); Software (equal); Supervision (lead); Writing-original draft (equal); Writing-review & editing (equal). **Branislav Kusy:** Conceptualization (supporting); Investigation (supporting); Methodology (supporting); Software (supporting); Supervision (lead); Writing-original draft (equal); Writing-review & editing (equal). **Ellen Dritia:** Investigation (supporting); Methodology (supporting); Resources (supporting); Software (supporting); Writing-original draft (supporting); Writing-review & editing (equal). **Rod Connolly:** Conceptualization (equal); Data curation (supporting); Formal analysis (supporting); Funding acquisition (supporting); Investigation (supporting); Methodology (supporting); Project administration (supporting); Resources (supporting); Software (supporting); Supervision (lead); Writing-original draft (equal); Writing-review & editing (equal).

## OPEN RESEARCH BADGES

This article has earned an Open Data and Open Materials Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at 10.5281/zenodo.4571757 and <https://github.com/slopezmarcano/automated-fishtracking/tree/fish-track2>.

## DATA AVAILABILITY STATEMENT

The training images and annotations, movement dataset annotations, images and videos, and the tracking and data wrangling scripts have been made available at (<https://doi.org/10.5281/zenodo.4571760>).

## ORCID

Sebastian Lopez-Marcano  <https://orcid.org/0000-0002-0814-2906>

## REFERENCES

- Aguzzi, J., Chatzievangelou, D., Marini, S., Fanelli, E., Danovaro, R., Flögel, S., Lebris, N., Juanes, F., De Leo, F. C., Del Rio, J., Thomsen, L., Costa, C., Riccobene, G., Tamburini, C., Lefevre, D., Gojak, C., Poulain, P.-M., Favali, P., Griffa, A., ... Company, J. B. (2019). New high-tech flexible networks for the monitoring of deep-sea ecosystems. *Environmental Science & Technology*, 53, 6616–6631. <https://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 10 Conference* (pp. 778–783). <https://doi.org/10.1109/TENCON.2019.8929613>.



- Beyan, C., & Brownman, H. I. (2020). Setting the stage for the machine intelligence era in marine science. *ICES Journal of Marine Science*, 77(4), 1267–1273. <https://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. <https://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. <https://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*, 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. <https://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. <https://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(10), 1632–1644. <https://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. <https://doi.org/10.1109/TSMC.2016.2523943>
- Connolly, R., Fairclough, D., Jinks, E., Ditría, E., Jackson, G., Lopez-Marcano, S., & Jinks, K. (2021). Improved accuracy for automated counting of a fish in baited underwater videos for stock assessment. Cold Spring Harbor Laboratory.
- Cui, S., Zhou, Y., Wang, Y., & Zhai, L. (2020). Fish detection using deep learning. *Applied Computational Intelligence and Soft Computing*, 2020, 3738108. <https://doi.org/10.1155/2020/3738108>
- Ditría, E. M., Lopez-Marcano, S., Sievers, M., Jinks, E. L., Brown, C. J., & Connolly, R. M. (2020). Automating the analysis of fish abundance using object detection: Optimizing animal ecology with deep learning. *Frontiers in Marine Science*, 7. <https://doi.org/10.3389/fmars.2020.00429>
- Ditría, 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*. <https://doi.org/10.1101/2020.05.19.105056>
- 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. <https://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. <https://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. <https://doi.org/10.1002/rse.2106>
- González-Rivero, M., Beijbom, O., Rodriguez-Ramirez, A., Bryant, D. E. P., Ganase, A., Gonzalez-Marrero, Y., Herrera-Reveles, A., Kennedy, E. V., Kim, C. J. S., Lopez-Marcano, S., Markey, K., Neal, B. P., Osborne, K., Reyes-Nivia, C., Sampayo, E. M., Stolberg, K., Taylor, A., Vercelloni, J., Wyatt, M., & Hoegh-Guldberg, O. (2020). Monitoring of coral reefs using artificial intelligence: A feasible and cost-effective approach. *Remote Sensing*, 12(3), 489. <https://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. <https://doi.org/10.1038/s41598-019-50795-9>
- Guo, S., Xu, P., Miao, Q., Shao, G., Chapman, C. A., Chen, X., He, G., Fang, D., Zhang, H. E., Sun, Y., Shi, Z., & Li, B. (2020). Automatic identification of individual primates with deep learning techniques. *iScience*, 23(8), 101412. <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.
- He, K. M., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask r-cnn. *IEEE International Conference on Computer Vision*, 2980–2988. <https://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. <https://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. <https://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. <https://doi.org/10.1016/j.rse.2018.12.021>
- Kennedy, E. V., Vercelloni, J., Neal, B. P., Ambariyanto, A., Bryant, D. E. P., Ganase, A., Gartrell, P., Brown, K., Kim, C. J. S., Hudatwi, M., Hadi, A., Prabowo, A., Prihatinningsih, P., Haryanta, S., Markey, K., Green, S., Dalton, P., Lopez-Marcano, S., Rodriguez-Ramirez, 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. <https://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., Barrett, N., Bernard, A. T. F., Bouchet, P. J., Birt, M. J., Cappo, M., Currey-Randall, L. M., Driessens, D., Fairclough, D. V., Fullwood, L. A. F., Gibbons, B. A., Harasti, D., Heupel, M. R., Hicks, 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*, 11(11), 1401–1409. <https://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. <https://doi.org/10.1038/nature14539>
- Librán-Embid, F., Klaus, F., Tscharntke, T., & Grass, I. (2020). Unmanned aerial vehicles for biodiversity-friendly agricultural landscapes - a systematic review. *Science of the Total Environment*, 732, 139204. <https://doi.org/10.1016/j.scitotenv.2020.139204>
- 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*, 31(1), 210–217. <https://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. <https://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 from <https://github.com/facebookresearch/maskrcnn-benchmark>. Accessed 29 October 2020
- 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. <https://doi.org/10.1016/j.procs.2020.03.123>
- Olds, A. D., Nagelkerken, I., Huijbers, M., Gilby, B., Pittman, S., & Schlacher, T. (2018). Connectivity in coastal seascapes. In S. J. Pittman (Ed.), *Seascape ecology* (pp. 261–291). 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. <https://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. <https://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 (Eds.), *Neural networks: Tricks of the trade*, 2nd ed. (pp. 53–67). 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. <https://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. <https://doi.org/10.4404/hystrix-24.2-8789>
- 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. <https://doi.org/10.1002/lom3.10113>
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- 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. <https://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. <https://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. <https://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. In L. Agapito, M. M. Bronstein, & C. Rother (Eds.), *Computer vision - ECCV 2014 workshops* (pp. 255–270). Springer International Publishing.
- 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: J Blanc-Talon & C Distante (eds.), *Advanced Concepts for Intelligent Vision Systems* (pp. 160–171). International Conference on Advanced Concepts for Intelligent Vision Systems. 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. <https://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. <https://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. <https://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*.
- Watanabe, J.-I., Shao, Y., & Miura, N. (2019). Underwater and airborne monitoring of marine ecosystems and debris. *Journal of Applied Remote Sensing*, 13, 44509. <https://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. <https://doi.org/10.1098/rsos.181748>
- Weinstein, B. G. (2018). A computer vision for animal ecology. *Journal of Animal Ecology*, 87, 533–545. <https://doi.org/10.1111/1365-2656.12780>
- Wickham, H. (2009). Ggplot2: Elegant graphics for data analysis. In *Ggplot2: Elegant graphics for data analysis* (pp. 1, 1–212). New York: Springer-Verlag.
- Xiu, L., Min, S., Qin, H., & Liansheng, C. (2015). Fast accurate fish detection and recognition of underwater images with Fast R-CNN. *IEEE*. <https://doi.org/10.23919/OCEANS.2015.7404464>
- Xu, Z., & Cheng, X. E. (2017). Zebrafish tracking using convolutional neural networks. *Scientific Reports*, 7, 42815. <https://doi.org/10.1038/srep42815>
- Zhao, Z.-Q., Zheng, P., Xu, S.-T., & Wu, X. (2019). Object detection with deep learning: A review. *ArXiv*.
- 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. <https://doi.org/10.1111/ecog.03315>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Lopez-Marcano S, Jinks E, Buelow CA, et al. Automatic detection of fish and tracking of movement for ecology. *Ecol Evol*. 2021;11:8254–8263. <https://doi.org/10.1002/ece3.7656>